Algebra & Number Theory

Volume 19
2025
No. 6

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 19:6 (2025)
https://doi.org/10.2140/ant.2025.19.1049

msp

Semistable representations as limits of crystalline representations

Anand Chitrao, Eknath Ghate and Seidai Yasuda

We construct an explicit sequence V_{k_n,a_n} of crystalline representations of exceptional weights converging to a given irreducible two-dimensional semistable representation $V_{k,\mathcal{L}}$ of $\mathrm{Gal}(\overline{\mathbb{Q}}_p/\mathbb{Q}_p)$. The convergence takes place in the blow-up space of two-dimensional trianguline representations studied by Colmez and Chenevier. The process of blow-up is described in detail in the rigid-analytic setting and may be of independent interest. Also, we recover a formula of Stevens expressing the $\mathcal{L}$-invariant as a logarithmic derivative.

Our result can be used to compute the reduction of $V_{k,\mathcal{L}}$ in terms of the reductions of the V_{k_n,a_n}. For instance, using the zig-zag conjecture we recover (resp. extend) the work of Breuil and Mézard and Guerberoff and Park computing the reductions of the $V_{k,\mathcal{L}}$ for weights k at most $p-1$ (resp. $p+1$), at least on the inertia subgroup. In the cases where zig-zag is known, we are further able to obtain some new information about the reductions for small odd weights. Finally, we explain some apparent violations to local constancy in the weight of the reductions of crystalline representations of small weight.

1. Introduction

Let p be an odd prime. Let E be a finite extension of $\mathbb{Q}_p$ containing $\sqrt{p}$. Let D_{st} be Fontaine's functor inducing an equivalence of categories between semistable representations of the Galois group $\mathrm{Gal}(\overline{\mathbb{Q}}_p/\mathbb{Q}_p)$ over E and admissible filtered (φ, N)-modules over E. We introduce two kinds of representations using this functor.

MSC2020: 11F80, 14G22.
Keywords: Galois representations, (φ, Γ)-modules, $\mathcal{L}$-invariants, rigid geometry, blow-ups.

For every integer $k \geq 2$ and $a_p \in E$ of positive valuation, there is an irreducible two-dimensional *crystalline* representation V_{k,a_p} over E of $\mathrm{Gal}(\overline{\mathbb{Q}}_p/\mathbb{Q}_p)$ with Hodge–Tate weights $(0, k-1)$ and $D_{\mathrm{st}}(V^*_{k,a_p}) = D_{k,a_p}$, where $D_{k,a_p} = Ee_1 \oplus Ee_2$ is the filtered φ-module defined by

$$\begin{cases} \varphi(e_1) = p^{k-1}e_2, \\ \varphi(e_2) = -e_1 + a_p e_2 \end{cases} \qquad \text{and} \qquad \mathrm{Fil}^i D_{k,a_p} = \begin{cases} D_{k,a_p} & \text{if } i \leq 0, \\ Ee_1 & \text{if } 1 \leq i \leq k-1, \\ 0 & \text{if } i \geq k. \end{cases}$$

Similarly, for every integer $k \geq 2$ and $\mathcal{L} \in \mathbb{P}^1(E)$ (called the $\mathcal{L}$-invariant), there is a two-dimensional *semistable* representation $V_{k,\mathcal{L}}$ over E of $\mathrm{Gal}(\overline{\mathbb{Q}}_p/\mathbb{Q}_p)$ with Hodge–Tate weights $(0, k-1)$ and $D_{\mathrm{st}}(V^*_{k,\mathcal{L}}) = D_{k,\mathcal{L}}$, where $D_{k,\mathcal{L}} = Ee_1 \oplus Ee_2$ is the filtered (φ, N)-module defined by

$$\begin{cases} \varphi(e_1) = p^{\frac{1}{2}k}e_1, \\ \varphi(e_2) = p^{\frac{1}{2}(k-2)}e_2 \end{cases} \qquad \text{and} \qquad \mathrm{Fil}^i D_{k,\mathcal{L}} = \begin{cases} D_{k,\mathcal{L}} & \text{if } i \leq 0, \\ E(e_1 + \mathcal{L}e_2) & \text{if } 1 \leq i \leq k-1 \text{ and } \mathcal{L} \neq \infty, \\ E(e_1 + e_2) & \text{if } 1 \leq i \leq k-1 \text{ and } \mathcal{L} = \infty, \\ 0 & \text{if } i \geq k, \end{cases}$$

and

$$\begin{cases} N(e_1) = e_2, \\ N(e_2) = 0 \end{cases} \quad \text{if } \mathcal{L} \neq \infty, \quad \text{and} \quad N = 0 \quad \text{if } \mathcal{L} = \infty.$$

If $k \geq 3$, then the semistable representation $V_{k,\mathcal{L}}$ is irreducible and when $\mathcal{L} = \infty$ is isomorphic to the representation V_{k,a_p} with $a_p = p^{k/2} + p^{(k-2)/2}$.

This paper studies several relationships between the crystalline representations V_{k,a_p} and the semistable representations $V_{k,\mathcal{L}}$ for $k \geq 3$.

In particular, we show how information about the reductions of the former representations implies information about the reductions of the latter. In general, computing the reductions of Galois representations has applications to computing deformation rings, to the weight part of Serre's conjecture, to the Breuil–Mézard conjecture and to modularity lifting theorems.

1.1. *Notation.*

- p is an odd prime and $\sqrt{p}$ is a fixed square root of p.
- E is a p-adic number field, i.e., a finite extension of $\mathbb{Q}_p$.
- v_p is the p-adic valuation normalized such that $v_p(p) = 1$.
- ζ_{p-1} is a fixed primitive $(p-1)$-th root of unity in $\mathbb{Q}_p^*$.
- log is the p-adic logarithm, normalized by setting $\log(p) = 0$.
- $\mathcal{T}$ is the rigid-analytic space parametrizing continuous characters of $\mathbb{Q}_p^*$.
- $x^i \in \mathcal{T}(\mathbb{Q}_p)$ for $i \geq 0$ is the character $\mathbb{Q}_p^* \to \mathbb{Q}_p^*$ that sends an element to its i-th power.
- $\chi \in \mathcal{T}(\mathbb{Q}_p)$ is the p-adic cyclotomic character $\mathbb{Q}_p^* \to \mathbb{Q}_p^*$ which maps p to 1 and which is the identity on $\mathbb{Z}_p^*$.
- Characters of the form $x^i\chi$ for $i \geq 0$ are called exceptional characters.

- μ_λ is the character of $\mathbb{Q}_p^*$ sending p to $\lambda \in \bar{\mathbb{F}}_p^*$ or $\bar{\mathbb{Q}}_p^*$ and $\mathbb{Z}_p^*$ to 1.
- Normalize the map $\mathbb{Q}_p^* \to \mathrm{Gal}(\bar{\mathbb{Q}}_p/\mathbb{Q}_p)^{\mathrm{ab}}$ of class field theory by sending p to a geometric Frobenius. We sometimes think of χ and μ_λ as characters of $\mathrm{Gal}(\bar{\mathbb{Q}}_p/\mathbb{Q}_p)$.
- Let $\Gamma = \mathrm{Gal}(\mathbb{Q}_p(\mu_{p^\infty})/\mathbb{Q}_p)$. We also think of χ as a character $\Gamma \xrightarrow{\sim} \mathbb{Z}_p^*$. From Section 3.1, we fix a topological generator γ of Γ such that $\chi(\gamma) = \zeta_{p-1}^a(1+p)$ for a fixed integer a.
- $I_{\mathbb{Q}_p}$ is the inertia subgroup of $\mathrm{Gal}(\bar{\mathbb{Q}}_p/\mathbb{Q}_p)$.
- ω is the fundamental character of $I_{\mathbb{Q}_p}$ of level 1; it has a canonical extension to $\mathrm{Gal}(\bar{\mathbb{Q}}_p/\mathbb{Q}_p)$.
- ω_2 is the fundamental character of $I_{\mathbb{Q}_p}$ of level 2; choose an extension of ω_2 to $\mathrm{Gal}(\bar{\mathbb{Q}}_p/\mathbb{Q}_{p^2})$ so that for an integer c with $p+1 \nmid c$ the representation $\mathrm{ind}(\omega_2^c)$ obtained by inducing this extension from $\mathrm{Gal}(\bar{\mathbb{Q}}_p/\mathbb{Q}_{p^2})$ to $\mathrm{Gal}(\bar{\mathbb{Q}}_p/\mathbb{Q}_p)$ has determinant ω^c.
- $\mathcal{R}_A$ is the Robba ring with coefficients in A for A an affinoid algebra.
- $\mathcal{R}_A(\delta)$ for $\delta \in \mathcal{T}(A)$ is the (φ, Γ)-module of rank 1 over $\mathcal{R}_A$ with action of φ and Γ:

$$\varphi f(T) = \delta(p) f((1+T)^p - 1) \quad \text{and} \quad \gamma f(T) = \delta(\chi(\gamma)) f((1+T)^{\chi(\gamma)} - 1) \quad \text{for } \gamma \in \Gamma.$$

- k is an integer greater than or equal to 2 and $r = k-2$.
- v_- and v_+ are the largest and smallest nonnegative integers, respectively, such that $v_- < \frac{1}{2}(k-2) < v_+$ for $k \in [3, p+1]$.
- $H_0 = 0$ and $H_l = \sum_{i=1}^{l}(1/i)$ is the l-th partial harmonic sum for $l \geq 1$; write $H_\pm = H_{v_\pm}$.
- $v = v_p(\mathcal{L} - H_- - H_+)$ is the p-adic valuation of $\mathcal{L}$ in a finite extension of $\mathbb{Q}_p$ shifted by the partial harmonic sums H_- and H_+. Note v equals $v_p(\mathcal{L})$ if either quantity is negative.
- $\mathbb{P}(V)$ is the projectivization of a vector space V.
- ϕ is Euler's totient function.
- $\phi_n(T)$ for $n \geq 1$ is the p^n-th cyclotomic polynomial.

1.2. ***Limits of crystalline representations.*** Colmez [2008] and Chenevier [2013] have constructed a moduli space of nonsplit trianguline (φ, Γ)-modules of rank 2 over the Robba ring (assuming that the quotient of the characters occurring in the triangulation is not of a certain kind). The first goal of this paper is to construct for $k \geq 3$ an explicit sequence of crystalline representations converging in this space to the (dual of the) semistable representation $V_{k,\mathcal{L}}$ for a prescribed $\mathcal{L}$-invariant $\mathcal{L}$. We will use this in conjunction with a local constancy result to study the reductions of semistable representations.

In order to state our result, let us recall the definition of the rigid-analytic space constructed by Colmez and Chenevier. Let $\mathcal{T}$ be the parameter space for characters of $\mathbb{Q}_p^*$. Let $x : \mathbb{Q}_p^* \to \mathbb{Q}_p^*$ be the identity character. Let $\chi : \mathbb{Q}_p^* \to \mathbb{Z}_p^*$ is the p-adic cyclotomic character, sending p to 1 and such that $\chi|_{\mathbb{Z}_p^*}$ is the identity character. We call the characters $x^i\chi$ for $i \geq 0$ *exceptional*. For each $i \geq 0$, let F_i and F_i' be the closed analytic subvarieties of $\mathcal{T} \times \mathcal{T}$ such that, for every finite extension E of $\mathbb{Q}_p$, we have

$$F_i(E) = \{(\delta_1, \delta_2) \in \mathcal{T}(E) \times \mathcal{T}(E) \mid \delta_1\delta_2^{-1} = x^i\chi\},$$
$$F_i'(E) = \{(\delta_1, \delta_2) \in \mathcal{T}(E) \times \mathcal{T}(E) \mid \delta_1\delta_2^{-1} = x^{-i}\}.$$

Let $F = \bigcup_{i\geq 0} F_i$ and $F' = \bigcup_{i\geq 0} F_i'$. The Colmez–Chenevier space $\widetilde{\mathcal{T}}_2$ is the blow-up of $(\mathcal{T}\times\mathcal{T})\setminus F'$ along F in the category of rigid-analytic spaces. Our first main theorem is the following:

Theorem 1.1. *Let $k \geq 3$, $r = k-2$ and $\mathcal{L} \in \mathbb{P}^1(E)$. For $n \geq 1$, let*

$$(k_n, a_n) = \begin{cases} \big(k + p^n(p-1),\ p^{\frac{1}{2}r}\big(1 + \frac{1}{2}\mathcal{L}p^n(p-1)\big)\big) & \text{if } \mathcal{L} \neq \infty, \\ (k + p^{n^2}(p-1),\ p^{\frac{1}{2}r}(1+p^n)) & \text{if } \mathcal{L} = \infty. \end{cases} \tag{1}$$

Then

$$V^*_{k_n,a_n} \to V^*_{k,\mathcal{L}},$$

i.e., the sequence of points in $\widetilde{\mathcal{T}}_2$ associated to the crystalline representations $V^*_{k_n,a_n}$ converges to the point in $\widetilde{\mathcal{T}}_2$ associated to the semistable representation $V^*_{k,\mathcal{L}}$.

Remarks. (1) The weights k_n appearing in the theorem are for sufficiently large *n exceptional* in the sense that they are two more than twice the valuation $v_p(a_n)$ of a_n modulo $(p-1)$ [Ghate 2021]. This will be of key importance in what follows. In fact, if (k_n, a_n) is an arbitrary sequence of points such that $V^*_{k_n,a_n} \to V^*_{k,\mathcal{L}}$, then the k_n are eventually exceptional weights. Indeed, it is not hard to see that $(k_n, a_n) \to (k,\ p^{r/2})$ and $k_n \equiv k \bmod p-1$ for large n.

(2) In the case $\mathcal{L} \neq \infty$, the sequence of points (k_n, a_n) in Theorem 1.1 lies on the line $a_p(l) = p^{r/2}\big(1 + \frac{1}{2}\mathcal{L}(l-k)\big)$. Clearly

$$\mathcal{L} = 2\frac{a_p'(k)}{a_p(k)}. \tag{2}$$

In Section 6.2, we show that (2) holds more generally for an arbitrary sequence (k_n, a_n) of points on *any* smooth curve $a_p(l)$ such that $V^*_{k_n,a_n} \to V^*_{k,\mathcal{L}}$ in the blow-up space $\widetilde{\mathcal{T}}_2$. The formula (2) is a variant of a classical formula initially proved by Greenberg and Stevens [1993, Theorem 3.18] for elliptic curves with split multiplicative reduction (a weight-2, slope-0 case) and extended by Stevens [2010, Theorem B] (see also [Bertolini, Darmon, and Iovita 2010, Theorem 4]) to higher weights and slopes. Further generalizations were proved by Colmez [2010, Théorème 0.5, Corollaire 0.7], Benois [2010, Theorem 2] and others. This classical formula was a key local ingredient in the proof of the Mazur–Tate–Teitelbaum conjecture for elliptic curves due to Greenberg and Stevens. Our proof of (2) is essentially geometric. Recall the classical picture in algebraic geometry of the blow-up of $\mathbb{A}^2$ at a point shown in Figure 1.

The strict transform of the curve $y = f(x)$ passing through the origin $(0,0)$ in $\mathbb{A}^2$ passes through the exceptional divisor $\mathbb{P}^1$ above the origin at "height" the derivative of f at 0. The proofs of (2) and Theorem 1.1 are based on an analogous principle in a rigid-analytic setting.

(3) The techniques used to prove Theorem 1.1 can also be used to prove that the limit of a sequence of irreducible two-dimensional crystalline representations of $\mathrm{Gal}(\overline{\mathbb{Q}}_p/\mathbb{Q}_p)$ with Hodge–Tate weights belonging to an interval $[a, b]$ is also irreducible crystalline with Hodge–Tate weights in $[a, b]$, at least if the difference

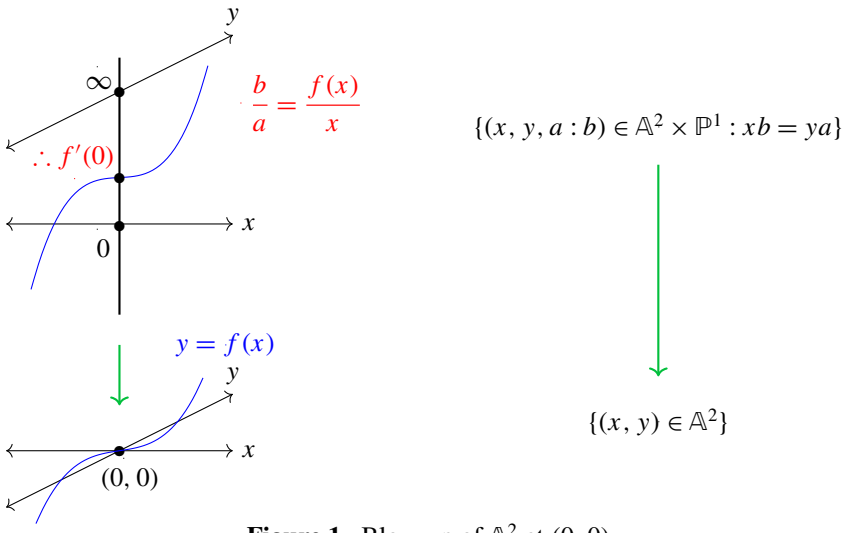


Figure 1. Blow-up of $\mathbb{A}^2$ at $(0, 0)$.

of the Hodge–Tate weights of the representations in the sequence is at least 2 infinitely often. This gives another (geometric) proof of a special case of a general result of Berger [2004, Théorème 1] (see Section 8).

1.3. ***Computing the $\mathcal{L}$-invariant.*** In this section, we now explain the techniques involved in proving Theorem 1.1. The discussion should also serve as an overview of the contents of Sections 2 to 6.

Let E be a finite extension of $\mathbb{Q}_p$. Let $\mathcal{R}_E$ be the Robba ring over E consisting of bidirectional power series having coefficients in E that converge on the elements of $\overline{\mathbb{Q}}_p$ with valuation in $]0, M]$ for some $M > 0$. For a more precise description, see Section 3.

The E-valued points of $\widetilde{\mathcal{T}}_2$ are tuples (δ_1, δ_2, L) where δ_1, δ_2 are E^*-valued characters of $\mathbb{Q}_p^*$ (with $\delta_1\delta_2^{-1} \neq x^{-j}$ for any $j \geq 0$) and $L \in \mathbb{P}^1(E)$ is the $\mathcal{L}$-invariant of the (φ, Γ)-module associated to the isomorphism class of the nonsplit extension

$$0 \to \mathcal{R}_E(\delta_1) \to * \to \mathcal{R}_E(\delta_2) \to 0,$$

when $\delta_1\delta_2^{-1} = x^i\chi$ for $i \geq 0$ is an exceptional character, and is taken to be ∞ otherwise (more precisely, in the former case, L is the $\mathcal{L}$-invariant defined by Colmez — see Section 3 — of this extension twisted by δ_2^{-1}).

The functor $\mathbf{D}_{\mathrm{rig}}$ sets up an equivalence of categories between the category of E-linear representations of $\mathrm{Gal}(\overline{\mathbb{Q}}_p/\mathbb{Q}_p)$ and the category of (φ, Γ)-modules over $\mathcal{R}_E$ of slope 0. Let $k \geq 3$ and $r = k - 2$, and, for $n \geq 1$, let (k_n, a_n) be as in (1). By [Berger 2012, Proposition 3.1], $\mathbf{D}_{\mathrm{rig}}(V^*_{k_n,a_n})$ is an extension

$$0 \to \mathcal{R}_E(\mu_{y_n}) \to \mathbf{D}_{\mathrm{rig}}(V^*_{k_n,a_n}) \to \mathcal{R}_E(\mu_{1/y_n}\chi^{1-k_n}) \to 0,$$

where

$$y_n = \tfrac{1}{2}(a_n + \sqrt{a_n^2 - 4p^{k_n-1}}). \tag{3}$$

This allows us to associate to (the dual of) V_{k_n,a_n} the point $(\mu_{y_n}, \mu_{1/y_n}\chi^{1-k_n}, \infty)$ of the blow-up $\widetilde{\mathcal{T}}_2$. We claim that for $n \geq 1$ this sequence of points converges in the blow-up to the point

$$(\mu_{p^{r/2}}, \mu_{1/p^{r/2}}\chi^{1-k}, -\mathcal{L}) \quad \text{if } \mathcal{L} \neq \infty,$$
$$(\mu_{p^{r/2}}, \mu_{1/p^{r/2}}\chi^{1-k}, \infty) \quad \text{if } \mathcal{L} = \infty.$$

It turns out that the corresponding (φ, Γ)-module is étale (for any $\mathcal{L}$) and that the corresponding Galois representation is the semistable representation $V^*_{k,\mathcal{L}}$ (see the end of Section 6.1). In the $\mathcal{L} = \infty$ case, the representation $V^*_{k,\mathcal{L}}$ is in fact crystalline. This proves Theorem 1.1.

It remains to prove the claim. Let $\widetilde{\mathcal{T}}$ be the blow-up of $\mathcal{T} \setminus \{x^{-j}\}_{j\geq 0}$ at $\{x^i\chi\}_{i\geq 0}$. The E-valued points of $\widetilde{\mathcal{T}}$ are tuples (δ, L), where δ is an E^*-valued character of $\mathbb{Q}_p^*$ (with $\delta \neq x^{-j}$ for any $j \geq 0$) and $L \in \mathbb{P}^1(E)$ is the $\mathcal{L}$-invariant (defined by Colmez, see Section 3) of the (φ, Γ)-module associated to the isomorphism class of the nonsplit extension

$$0 \to \mathcal{R}_E(\delta) \to * \to \mathcal{R}_E \to 0,$$

when $\delta = x^i\chi$ for $i \geq 0$ is an exceptional character, and is taken to be ∞ otherwise. Now a sequence of points $(\delta_{1,n}, \delta_{2,n}, L_n)$ in $\widetilde{\mathcal{T}}_2$ converges to a point (δ_1, δ_2, L) in $\widetilde{\mathcal{T}}_2$ if and only if

- $\delta_{1,n}$ and $\delta_{2,n}$ converge to δ_1 and δ_2, respectively, in $\mathcal{T}$, and
- $(\delta_{1,n}\delta_{2,n}^{-1}, L_n)$ converges to $(\delta_1\delta_2^{-1}, L)$ in $\widetilde{\mathcal{T}}$.

That is, with respect to the commutative diagram

$$\begin{array}{ccc} \widetilde{\mathcal{T}}_2 & \longrightarrow & \widetilde{\mathcal{T}} \\ \downarrow & & \downarrow \\ \mathcal{T} \times \mathcal{T} \setminus F' & \longrightarrow & \mathcal{T} \setminus \{x^{-j}\}_{j\geq 0} \end{array}$$

where the bottom map is the restriction of the twisting map $\mathcal{T} \times \mathcal{T} \to \mathcal{T}$ sending (δ_1, δ_2) to $\delta_1\delta_2^{-1}$, the top map sends (δ_1, δ_2, L) to $(\delta_1\delta_2^{-1}, L)$ and the vertical maps are the blow-up maps at F and $\{x^i\chi\}_{i\geq 0}$, respectively, the sequence $(\delta_{1,n}, \delta_{2,n}, L_n)$ converges to (δ_1, δ_2, L) in $\widetilde{\mathcal{T}}_2$ if and only if the projections of this sequence under the left vertical map and the top horizontal map converge to the corresponding projections of (δ_1, δ_2, L). Indeed, the top map is obtained from the universal property of the blow-up map on the right (this can be checked on charts using the definition of the map g constructed after Lemma 2.3), so there is an induced map from $\widetilde{\mathcal{T}}_2$ to the fiber product of $\mathcal{T} \times \mathcal{T} \setminus F'$ and $\widetilde{\mathcal{T}}$ over $\mathcal{T} \setminus \{x^{-j}\}_{j\geq 0}$ which is a closed immersion [Schoutens 1995, Proposition 3.1.2], and hence induces an inclusion on points $\widetilde{\mathcal{T}}_2(E)$ to the fiber product of $(\mathcal{T} \times \mathcal{T} \setminus F')(E)$ and $\widetilde{\mathcal{T}}(E)$ over $(\mathcal{T} \setminus \{x^{-j}\}_{j\geq 0})(E)$ with closed image, and then this is the definition of convergence in the last fiber product.

An easy check shows that the characters μ_{y_n} and $\mu_{1/y_n}\chi^{1-k_n}$ converge to the characters $\mu_{p^{r/2}}$ and $\mu_{1/p^{r/2}}\chi^{1-k}$ respectively. The ratio of these characters is the exceptional character $\mu_{p^r}\chi^{k-1} = x^r\chi$. Thus, if the sequence $(\mu_{y_n}, \mu_{1/y_n}\chi^{1-k_n}, \infty)$ converges in $\widetilde{\mathcal{T}}_2$, then it converges to a point in the fiber over F_r. Thus, it remains to check that $(\mu_{y_n^2}\chi^{k_n-1}, \infty)$ converges in $\widetilde{\mathcal{T}}$ to

$$(x^r\chi, -\mathcal{L}) \quad \text{if } \mathcal{L} \neq \infty,$$
$$(x^r\chi, \infty) \quad \text{if } \mathcal{L} = \infty. \tag{4}$$

In order to prove (4), we set up local coordinates U_r around the point $x^r\chi$, describe the blow-up $\widetilde{U}_r$ of this coordinate patch with center at the exceptional character $x^r\chi$ explicitly (see Section 2 for details) and compute the limit in $\widetilde{U}_r$. Let ζ_{p-1} be a fixed primitive $(p-1)$-th root of unity. Associating the tuple $(\delta(p), \delta(\zeta_{p-1}), \delta(1+p)-1)$ to a character $\delta \in \mathcal{T}(\mathbb{Q}_p)$ identifies $\mathcal{T}(\mathbb{Q}_p)$ with $\mathbb{Q}_p^* \times \mu_{p-1} \times p\mathbb{Z}_p$. Under this identification, the exceptional character $\mu_{p^r}\chi^{k-1}$ goes to the tuple $(p^r, \zeta_{p-1}^{k-1}, (1+p)^{k-1}-1)$. The set $p^r\mathbb{Z}_p^* \times \{\zeta_{p-1}^{k-1}\} \times p\mathbb{Z}_p$ is a neighborhood of $\mu_{p^r}\chi^{k-1}$ in $\mathcal{T}(\mathbb{Q}_p)$. This leads us to consider the affinoid algebra

$$U_r = \operatorname{Sp} \mathbb{Q}_p\langle S_1, S_2, T_1, T_2, T_3\rangle/(p^r T_1 - S_1, 1 - T_1T_2, pT_3 - S_2)$$

as a neighborhood of $\mu_{p^r}\chi^{k-1}$ in $\mathcal{T}$ because clearly $U_r(\mathbb{Q}_p) = p^r\mathbb{Z}_p^* \times p\mathbb{Z}_p$. The variable S_1 corresponds to the first factor and S_2 to the second factor. From now on, by fixing the tame part of the characters under consideration, we identify $U_r(E)$ with the subset $p^r\mathcal{O}_E^* \times \{\zeta_{p-1}^{k-1}\} \times p\mathcal{O}_E$ of $\mathcal{T}(E)$, where $\mathcal{O}_E$ is the ring of integers of E.

The character $\mu_{p^r}\chi^{k-1} = x^r\chi$ corresponds to the maximal ideal

$$m = (S_1 - p^r, S_2 - ((1+p)^{k-1}-1))$$

of $\mathcal{O}(U_r)$. The blow-up $\widetilde{U}_r$ of U_r at the maximal ideal m turns out to have the following standard description (see (5)):

$$\widetilde{U}_r(E) = \{(s_1, s_2, \xi_1 : \xi_2) \in U_r(E) \times \mathbb{P}^1(E) \mid (s_1 - p^r)\xi_2 = (s_2 - ((1+p)^{k-1}-1))\xi_1\}.$$

For large n, the points $(\mu_{y_n^2}\chi^{k_n-1}, \infty)$ in $\widetilde{\mathcal{T}}$ lie in $\widetilde{U}_r$. We prove that the sequence converges in the blow-up $\widetilde{U}_r$ to the point (see Section 6.1)

$$\begin{cases} \left(p^r, (1+p)^{k-1}-1, \mathcal{L}\dfrac{p^r}{(1+p)^{k-1}\log(1+p)} : 1\right) & \text{if } \mathcal{L} \neq \infty, \\ (p^r, (1+p)^{k-1}-1, 1:0) & \text{if } \mathcal{L} = \infty, \end{cases}$$

where log is normalized so that $\log(p) = 0$. The proof of (4) then follows immediately from Theorem 5.2, a technical but important formula for the $\mathcal{L}$-invariant of a point in the exceptional fiber, noting that the fudge factor there cancels with the extra factor appearing in the third coordinate of the limit point above when $\mathcal{L} \neq \infty$ and flips the sign.

Theorem 5.2 is proved as follows. Given a point in the exceptional fiber, we convert it to a tangent direction in U_r at the point $(p^r, (1+p)^{k-1}-1)$, i.e., an element of $\mathbb{P}(\operatorname{Hom}(m/m^2 \otimes_{\mathbb{Q}_p} E, E))$. This is done using the map in Proposition 2.5. We then explicitly describe the isomorphism $\mathbb{P}(\operatorname{Hom}(m/m^2 \otimes_{\mathbb{Q}_p} E, E)) \to \mathbb{P}(H^1(\mathcal{R}_E(x^r\chi)))$ stated in [Chenevier 2013, Theorem 2.33], using some preparatory material on the cohomology of "big" (φ, Γ)-modules in Section 4. The image of the given point in the exceptional fiber under the composition of these two maps yields a cohomology class in $H^1(\mathcal{R}_E(x^r\chi))$ (up to scalars). The corresponding (φ, Γ)-module (up to isomorphism) is referred to as "the (φ, Γ)-module" associated to the given point (Definition 5.1). We then represent this cohomology class as an explicit linear combination of the basis elements of $H^1(\mathcal{R}_E(x^r\chi))$ studied by Benois [2011, Proposition 1.5.4]. The original formula

for the $\mathcal{L}$-invariant due to Colmez is, however, in terms of a different basis of $H^1(\mathcal{R}_E(x^r\chi))$, namely, the one constructed in [Colmez 2008, Proposition 2.19] (see Section 3). Restating the formula for the $\mathcal{L}$-invariant in terms of Benois' basis (Definition 3.6) allows us to give a formula for the $\mathcal{L}$-invariant of the given point in the exceptional fiber.

1.4. ***Reductions of semistable representations.*** Chenevier [2013, Proposition 3.9] proved that points of the space $\widetilde{\mathcal{T}}_2$ lie in families of (φ, Γ)-modules. By [Kedlaya and Liu 2010, Theorem 0.2], such a family comes from a family of Galois representations at least affinoid locally around an étale point. Furthermore (see, e.g., the discussion on [Chenevier 2013, p. 1513], which uses results of [Chenevier 2014] on pseudorepresentations), the semisimplification of the reduction of Galois representations living in connected families are isomorphic. This means that if the points in $\widetilde{\mathcal{T}}_2$ corresponding to two Galois representations are close, then the semisimplifications of the reductions of the two Galois representations are the same. Moreover, the dual of the reduction of a lattice in a p-adic representation is the same as the reduction of the dual lattice in the dual representation. Using these facts, along with Theorem 1.1, we see that if one knows the reductions of the crystalline representations appearing in Theorem 1.1, then one can compute the reduction of $V_{k,\mathcal{L}}$ for any $k \geq 3$ and $\mathcal{L} \neq \infty$ (the case $k = 2$ is not as interesting since the reduction is always reducible).

More generally, the above method allows one to compute the reduction of any irreducible two-dimensional noncrystalline semistable representation with distinct Hodge–Tate weights. Indeed, suppose V is such a semistable representation. Twisting by χ^a for some integer a, we may assume that the Hodge–Tate weights of the semistable representation $V \otimes \chi^a$ are $(0, k-1)$ for an integer $k \geq 2$. By, for instance, [Guerberoff and Park 2019, Lemma 3.1.2(3)], the filtered (φ, N)-module $D_{\mathrm{st}}((V \otimes \chi^a)^*)$ is the module $D(\lambda, \mathcal{L})$ described in [Guerberoff and Park 2019, Example 3.1.1] with $\lambda = up^{(k-2)/2}$ for some unit u (since r there is equal to $k-1$). Now $V \otimes \chi^a \otimes \mu_u$ is isomorphic to $V_{k,\mathcal{L}}$, as can be seen by comparing the corresponding filtered (φ, N)-modules. By [loc. cit., Lemma 3.1.2(4)], we must have $k \geq 3$.

This approach to computing the reduction of semistable representations using crystalline representations is of some importance because the reductions of these two classes of representations are nowadays largely studied by completely different methods: the crystalline case uses the compatibility of reduction between the p-adic and mod p local Langlands correspondences, or computes the reduction of the corresponding Wach module, whereas the reductions in the semistable case are determined by studying the reductions of the corresponding strongly divisible modules. In our experience, the former methods, while quite intricate, are not as complicated as the latter method. Thus, in view of the remarks above, the techniques used in the crystalline case may be brought to bear on the study of the reductions of semistable representations. We note, however, that the former method is only available for two-dimensional representations of $\mathrm{Gal}(\overline{\mathbb{Q}}_p/\mathbb{Q}_p)$, whereas the latter method is available in principle for representations of $\mathrm{Gal}(\overline{\mathbb{Q}}_p/\mathbb{Q}_p)$ of any dimension (though in practical terms only for those of small Hodge–Tate weights).

Let us illustrate this with some examples. The reductions of semistable representations have been computed completely for even weights in the range $[2, p-1]$ by Breuil and Mézard [2002], and for

odd weights in the same range by Guerberoff and Park [2019] at least on inertia. In [Ghate 2021], the second author made the following conjecture called the zig-zag conjecture[1] describing the reductions of crystalline representations of exceptional weights and half-integral slopes in terms of an alternating sequence of reducible and irreducible mod p representations.

Conjecture 1.2 (Zig-zag conjecture). *Say that $k \equiv k_0 = 2v(a_p) + 2 \bmod (p-1)$ is an exceptional congruence class of weights for a particular half-integral slope $\frac{1}{2} \le v_p(a_p) \in \frac{1}{2}\mathbb{Z} \le \frac{1}{2}(p-1)$. Let $r = k-2$ and $r_0 = k_0 - 2$. Define two parameters*

$$\tau = v_p\left(\frac{a_p^2 - \binom{r-v_-}{v_+}\binom{r-v_+}{v_-}p^{r_0}}{pa_p}\right) \quad \text{and} \quad t = v_p(k-k_0),$$

where v_- and v_+ are the largest and smallest integers such that $v(a_p)$ lies in (v_-, v_+). Then, for all weights $k > k_0$ with t sufficiently large, the (semisimplification of the) reduction $\overline{V}_{k,a_p}$ of the crystalline representation V_{k,a_p} on the inertia subgroup $I_{\mathbb{Q}_p}$ is given by

$$\overline{V}_{k,a_p}|_{I_{\mathbb{Q}_p}} \sim \begin{cases} \operatorname{ind}(\omega_2^{r_0+1}) & \text{if } \tau < t, \\ \omega^{r_0} \oplus \omega & \text{if } \tau = t, \\ \operatorname{ind}(\omega_2^{r_0+p}) & \text{if } t < \tau < t+1, \\ \omega^{r_0-1} \oplus \omega^2 & \text{if } \tau = t+1, \\ \operatorname{ind}(\omega_2^{r_0+2p-1}) & \text{if } t+1 < \tau < t+2, \\ \omega^{r_0-2} \oplus \omega^3 & \text{if } \tau = t+2, \\ \quad\vdots & \quad\vdots \\ \left.\begin{array}{ll} \operatorname{ind}(\omega_2^{r_0+1+\frac{1}{2}(r_0-2)(p-1)}) & \text{if } t + \frac{1}{2}(r_0-4) < \tau < t + \frac{1}{2}(r_0-2), \\ \omega^{\frac{1}{2}(r_0+2)} \oplus \omega^{\frac{1}{2}r_0} & \text{if } \tau = t + \frac{1}{2}(r_0-2), \\ \operatorname{ind}(\omega_2^{r_0+1+\frac{1}{2}r_0(p-1)}) & \text{if } \tau > t + \frac{1}{2}(r_0-2), \end{array}\right\} & \text{and } r_0 \text{ is even} \\ \qquad\qquad \text{or} & \\ \left.\begin{array}{ll} \operatorname{ind}(\omega_2^{r_0+1+\frac{1}{2}(r_0-1)(p-1)}) & \text{if } t + \frac{1}{2}(r_0-3) < \tau < t + \frac{1}{2}(r_0-1), \\ \omega^{\frac{1}{2}(r_0+1)} \oplus \omega^{\frac{1}{2}(r_0+1)} & \text{if } \tau \ge t + \frac{1}{2}(r_0-1) \end{array}\right\} & \text{and } r_0 \text{ is odd.} \end{cases}$$

This conjecture has been verified for some small slopes (see [Buzzard and Gee 2013] for slope $\frac{1}{2}$, [Bhattacharya, Ghate, and Rozensztajn 2018] for slope 1 and [Ghate and Rai 2025] for slope $\frac{3}{2}$) even with $t = 0$.[2]

This "crystalline" conjecture is intimately connected to the "semistable" results in [Breuil and Mézard 2002, Theorem 1.2; Guerberoff and Park 2019, Theorem 5.0.5]. More precisely, the zig-zag conjecture and Theorem 1.1 can be used to completely recover the description of the reductions of semistable representations in these theorems when $k \neq 2$, and even (conjecturally) extend it to the cases $k = p+1$ and $k = p$, respectively, at least on the inertia subgroup (see Theorem 1.3 below). Moreover, in the

[1] This version is mildly different from [Ghate 2021, Conjecture 1.1] in that there we require k to be sufficiently far away from some weights which are strictly larger than $p+1$, whereas here k is required to be sufficiently close to the weights $3 \le k_0 \le p+1$.

[2] The condition that t is sufficiently large is required for larger slopes due to some numerical observations made by Rozensztajn.

odd-weight cases for which the zig-zag conjecture has been proved, we obtain new information about the reductions of semistable representations on the full Galois group $\mathrm{Gal}(\overline{\mathbb{Q}}_p/\mathbb{Q}_p)$.

To elaborate further, let us set up some notation. For an integer k in the interval $[3, p+1]$, define v_- and v_+ to be the largest and smallest integers, respectively, such that $v_- < \frac{1}{2}(k-2) < v_+$. For $l \geq 1$, let

$$H_l = \sum_{i=1}^{l} \frac{1}{i}$$

be the l-th partial harmonic sum and set $H_0 = 0$. For convenience, write $H_- = H_{v_-}$ and $H_+ = H_{v_+}$. For any $\mathcal{L}$ in a finite extension of $\mathbb{Q}_p$, let

$$\nu = v_p(\mathcal{L} - H_- - H_+)$$

be the p-adic valuation of the $\mathcal{L}$-invariant shifted by the partial harmonic sums H_- and H_+. Let ω and ω_2 denote the mod p fundamental characters of levels 1 and 2, respectively, on the inertia group $I_{\mathbb{Q}_p}$. The character ω may be thought of as a character of $\mathrm{Gal}(\overline{\mathbb{Q}}_p/\mathbb{Q}_p)$; we choose an extension of ω_2 to $\mathrm{Gal}(\overline{\mathbb{Q}}_p/\mathbb{Q}_{p^2})$ such that for any integer c with $p+1 \nmid c$, the representation $\mathrm{ind}(\omega_2^c)$ obtained by inducing ω_2^c from $\mathrm{Gal}(\overline{\mathbb{Q}}_p/\mathbb{Q}_{p^2})$ to $\mathrm{Gal}(\overline{\mathbb{Q}}_p/\mathbb{Q}_p)$ has determinant ω^c. Let μ_λ be the unramified character of $\mathrm{Gal}(\overline{\mathbb{Q}}_p/\mathbb{Q}_p)$ mapping a geometric Frobenius element at p to $\lambda \in \overline{\mathbb{F}}_p^*$ or $\overline{\mathbb{Q}}_p^*$. Normalizing local class field theory by sending p to a geometric Frobenius element at p, we obtain the character μ_λ of $\mathbb{Q}_p^*$ taking p to λ and $\mathbb{Z}_p^*$ to 1 used in Section 1.3. Our second main theorem is the following:

Theorem 1.3. *If the zig-zag conjecture is true, then for any weight k satisfying $3 \leq k \leq p+1$, we have the following $(k-1)$-fold description of the (semisimplification of the) reduction of the semistable representation $V_{k,\mathcal{L}}$ for $\mathcal{L} \in \mathbb{P}^1(\overline{\mathbb{Q}}_p)$ on the inertia subgroup*:

$$\overline{V}_{k,\mathcal{L}}|_{I_{\mathbb{Q}_p}} \sim \begin{cases} \omega_2^{k-1} \oplus \omega_2^{p(k-1)} & \text{if } \nu < 1 - \frac{1}{2}(k-2), \\ \omega^{k-2} \oplus \omega & \text{if } \nu = 1 - \frac{1}{2}(k-2), \\ \omega_2^{k-2+p} \oplus \omega_2^{p(k-2)+1} & \text{if } 1 - \frac{1}{2}(k-2) < \nu < 2 - \frac{1}{2}(k-2), \\ \omega^{k-3} \oplus \omega^2 & \text{if } \nu = 2 - \frac{1}{2}(k-2), \\ \quad \vdots & \quad \vdots \\ \left.\begin{array}{ll} \omega_2^{\frac{1}{2}k+1+p\left(\frac{1}{2}k-2\right)} \oplus \omega_2^{\frac{1}{2}k-2+p\left(\frac{1}{2}k+1\right)} & \text{if } -1 < \nu < 0, \\ \omega^{\frac{1}{2}k} \oplus \omega^{\frac{1}{2}(k-2)} & \text{if } \nu = 0, \\ \omega_2^{\frac{1}{2}k+p\left(\frac{1}{2}k-1\right)} \oplus \omega_2^{\frac{1}{2}k-1+\frac{1}{2}pk} & \text{if } \nu > 0, \end{array}\right\} \text{and } k \text{ is even} \\ \quad \text{or} & \\ \left.\begin{array}{ll} \omega_2^{\frac{1}{2}(k+1)+p\left(\frac{1}{2}(k-3)\right)} \oplus \omega_2^{\frac{1}{2}(k-3)+p\left(\frac{1}{2}(k+1)\right)} & \text{if } -\frac{1}{2} < \nu < \frac{1}{2}, \\ \omega^{\frac{1}{2}(k-1)} \oplus \omega^{\frac{1}{2}(k-1)} & \text{if } \nu \geq \frac{1}{2} \end{array}\right\} \text{and } k \text{ is odd.} \end{cases}$$

Remark. The theorem holds when $\mathcal{L} = \infty$ even without assuming the zig-zag conjecture by the work of Fontaine and Edixhoven [1992], since, by convention, $v_p(\infty) = -\infty$.

A generalization of the $v < 1 - \frac{1}{2}(k-2)$ case of Theorem 1.3 has recently been proved for all weights $k \geq 4$ (and odd primes p) by [Bergdall, Levin and Liu 2023, Theorem 1.1]. Note that the term $v_p((k-2)!)$ in their result vanishes in our setting.

Also Theorem 1.3 above indeed matches with the results in [Breuil and Mézard 2002; Guerberoff and Park 2019] for $\mathcal{L} \neq \infty$ and for $k \in [3, p-1]$:

Even $k \in [3, p-1]$: Breuil and Mézard [2002, Theorem 1.2] have computed the reduction of $V_{k,\mathcal{L}}$ in terms of the valuations $v_p(a)$ and $v_p(\mathcal{L})$, where

$$a = (-1)^{\frac{1}{2}k}\Big(-1 + \frac{k}{2}\Big(\frac{k}{2} - 1\Big)(-\mathcal{L} + 2H_{\frac{1}{2}k-1})\Big).$$

Since $k \in [3, p-1]$, we see that $v_p\big(\frac{1}{2}k\big(\frac{1}{2}k-1\big)\big) = 0$, and therefore

$$\begin{aligned} v_p(a) &= v_p\bigg(\frac{-1}{\big(\frac{1}{2}k\big)\big(\frac{1}{2}k-1\big)} + (-\mathcal{L} + 2H_{\frac{1}{2}k-1})\bigg) = v_p\bigg(\frac{1}{\frac{1}{2}k} - \frac{1}{\frac{1}{2}k-1} - \mathcal{L} + 2\sum_{i=1}^{\frac{1}{2}k-1}\frac{1}{i}\bigg) \\ &= v_p\bigg(-\mathcal{L} + \sum_{i=1}^{\frac{1}{2}k}\frac{1}{i} + \sum_{i=1}^{\frac{1}{2}k-2}\frac{1}{i}\bigg) = v_p(\mathcal{L} - H_{\frac{1}{2}k} - H_{\frac{1}{2}k-2}) = v, \end{aligned}$$

where we have used k is even in the last equality. Now:

- If $v > 0$, then Theorem 1.3 yields
$$\overline{V}_{k,\mathcal{L}}|_{I_{\mathbb{Q}_p}} \sim \omega_2^{\frac{1}{2}k + p\left(\frac{1}{2}k-1\right)} \oplus \omega_2^{\frac{1}{2}k - 1 + \frac{1}{2}pk},$$
which agrees with [Breuil and Mézard 2002, Theorem 1.2(ii)].
- If $v = 0$, then Theorem 1.3 yields
$$\overline{V}_{k,\mathcal{L}}|_{I_{\mathbb{Q}_p}} \sim \omega^{\frac{1}{2}k} \oplus \omega^{\frac{1}{2}(k-2)},$$
which agrees with [loc. cit., Theorem 1.2(i)].
- So assume $v < 0$. Then $v = v_p(\mathcal{L})$. If $v < 2 - \frac{1}{2}k$, then Theorem 1.3 yields $\overline{V}_{k,\mathcal{L}}|_{I_{\mathbb{Q}_p}} \sim \omega_2^{k-1} \oplus \omega_2^{p(k-1)}$, which agrees with [loc. cit., Theorem 1.2(iii)]. Now if $2 - \frac{1}{2}k \leq v < 0$ and $v \in \mathbb{Z}$, then [loc. cit., Theorem 1.2(iii)] yields
$$\overline{V}_{k,\mathcal{L}}|_{I_{\mathbb{Q}_p}} \sim \omega^{\frac{1}{2}k - v} \oplus \omega^{\frac{1}{2}k + v - 1},$$
which is the same as the reduction computed in Theorem 1.3. Finally, if $2 - \frac{1}{2}k \leq v < 0$ and $v \notin \mathbb{Z}$, then [loc. cit., Theorem 1.2(iii)] yields
$$\overline{V}_{k,\mathcal{L}}|_{I_{\mathbb{Q}_p}} \sim \omega_2^{\frac{1}{2}k - \lfloor v \rfloor + p\left(\frac{1}{2}k + \lfloor v \rfloor - 1\right)} \oplus \omega_2^{\frac{1}{2}k + \lfloor v \rfloor - 1 + p\left(\frac{1}{2}k - \lfloor v \rfloor\right)},$$
which matches with the reduction computed in Theorem 1.3.

Remark. Since the zig-zag conjecture has already been proved for $p \geq 5$ and slope 1 in [Bhattacharya, Ghate, and Rozensztajn 2018, Theorem 1.1], even on $\mathrm{Gal}(\overline{\mathbb{Q}}_p/\mathbb{Q}_p)$, we recover the following result of

Breuil and Mézard when $k=4$ (see [Breuil and Mézard 2002, Theorem 1.2]):[3] if $p \geq 5$ and $k=4$, then

$$\overline{V}_{k,\mathcal{L}} \sim \begin{cases} \mathrm{ind}(\omega_2^3) & \text{if } \nu < 0, \\ \mu_\lambda \omega^2 \oplus \mu_{\lambda^{-1}}\omega & \text{if } \nu = 0, \\ \mathrm{ind}(\omega_2^{2+p}) & \text{if } \nu > 0, \end{cases}$$

where

$$\lambda = \overline{-2\big(\mathcal{L} - \tfrac{3}{2}\big)}.$$

If all nearby étale points in $\widetilde{\mathcal{T}}_2$ close to a given étale point, which has a lattice with nonsplit reduction, also have lattices with isomorphic reduction (an assumption which is stronger than the reductions being isomorphic up to semisimplification, and which might follow by extending results of [Chenevier 2014] from pseudorepresentations to Galois representations), then one can read off more subtle information about whether the representation $\overline{V}_{k,\mathcal{L}}$ is peu or très ramifiée (when it is reducible and $\lambda = \pm 1$) from the corresponding information of a sufficiently close crystalline representation V_{k_n,a_n}, which, in turn, is controlled by the size of $v_p(u_n - \varepsilon_n)$ in the notation of [Bhattacharya, Ghate, and Rozensztajn 2018, Theorem 1.3]. Indeed, in the middle case of the trichotomy above, we have $v_p(3-2\mathcal{L}) = 0$, putting us in part (i) of [Breuil and Mézard 2002, Theorem 1.2]. Note that $\overline{a_n/p} = 1$ and $\overline{k_n - 2} = 2$. There are now two cases to consider depending on whether $\varepsilon_n = \pm 1$. If $v_p(\mathcal{L}-2) = 0$, then a small check shows that $\varepsilon_n = 1$, so by [Bhattacharya, Ghate, and Rozensztajn 2018, Theorem 1.3(1a)], the reduction $\overline{V}_{k,\mathcal{L}}$ (without semisimplification) is peu ramifiée. If $v_p(\mathcal{L}-2) > 0$, then another small check shows that $\varepsilon_n = -1$ and that $v_p(u_n - \varepsilon_n) = v_p(\mathcal{L}-2)$. Thus, by [loc. cit., Theorem 1.3(1b)], the reduction $\overline{V}_{k,\mathcal{L}}$ (without semisimplification) is peu ramifiée if and only if $v_p(\mathcal{L}-2) < 1$, at least if $\mathcal{L}$ lies in an unramified extension of $\mathbb{Q}_p$. Both these conclusions are consistent with the corresponding conclusions in [Breuil and Mézard 2002, Theorem 1.2(i)]!

Odd $k \in [3, p-1]$: Guerberoff and Park [2019, Theorem 5.0.5] computed the reduction of $V_{k,\mathcal{L}}$ in terms of the valuation $v_p(\mathcal{L} - a(k-1))$, where

$$a(j) = H_{j/2} + H_{j/2-1}$$

for $j \geq 1$. We clearly have $a(k-1) = H_- + H_+$. Therefore, the regions used to classify the reductions in Theorem 1.3 match with those used in [loc. cit., Theorem 5.0.5]. Now:

- If $\nu < 1 - (k-2)/2$, then Theorem 1.3 yields

$$\overline{V}_{k,\mathcal{L}}|_{I_{\mathbb{Q}_p}} \sim \omega_2^{k-1} \oplus \omega_2^{p(k-1)},$$

which agrees with the reduction computed in [loc. cit., Theorem 5.0.5(2)].

- Assume $-\frac{1}{2} - l < \nu < \frac{1}{2} - l$ for some $l \in \big\{0, 1, \ldots, \frac{1}{2}(k-5)\big\}$. This region can be written as

$$\big(-l + \tfrac{1}{2}(k-3)\big) - \tfrac{1}{2}(k-2) < \nu < \big(-l + \tfrac{1}{2}(k-1)\big) - \tfrac{1}{2}(k-2).$$

[3] The computation $\bar{a}\overline{(a_p/p)} = \overline{(-1+2(-\mathcal{L}+2))}\overline{(p/p)} = \overline{-2\big(\mathcal{L}-\frac{3}{2}\big)}$ shows that the reduction agrees with the reduction computed in [Breuil and Mézard 2002, Theorem 1.2].

Theorem 1.3 yields

$$\overline{V}_{k,\mathcal{L}}|_{I_{\mathbb{Q}_p}} \sim \omega_2^{(k-1)+(p-1)\left(-l+\frac{1}{2}(k-3)\right)} \oplus \omega_2^{p\left((k-1)+(p-1)\left(-l+\frac{1}{2}(k-3)\right)\right)},$$

which agrees with the reduction computed in [Guerberoff and Park 2019, Theorem 5.0.5(1)].

- Assume $\nu = -\frac{1}{2} - l$ for some $l \in \left\{0, 1, \ldots, \frac{1}{2}(k-5)\right\}$. This can be written as

$$\nu = \left(-l + \tfrac{1}{2}(k-3)\right) - \tfrac{1}{2}(k-2).$$

Theorem 1.3 yields

$$\overline{V}_{k,\mathcal{L}}|_{I_{\mathbb{Q}_p}} \sim \omega^{k-1-\left(-l+\frac{1}{2}(k-3)\right)} \oplus \omega^{\left(-l+\frac{1}{2}(k-3)\right)},$$

which agrees with the reduction computed in [loc. cit., Theorem 5.0.5].

- Finally, if $\nu \geq \frac{1}{2}$, then Theorem 1.3 yields

$$\overline{V}_{k,\mathcal{L}}|_{I_{\mathbb{Q}_p}} \sim \omega^{\frac{1}{2}(k-1)} \oplus \omega^{\frac{1}{2}(k-1)},$$

which agrees with the reduction computed in [loc. cit., Theorem 5.0.5].

For the small weights $k = 3$ and 5, we may in fact improve on [loc. cit., Theorem 5.0.5] by computing the reductions $\overline{V}_{k,\mathcal{L}}$ on the full Galois group $\mathrm{Gal}(\overline{\mathbb{Q}}_p/\mathbb{Q}_p)$ (by giving formulas for λ). Indeed, since zig-zag has been proved for slopes $\frac{1}{2}$ and $\frac{3}{2}$ in [Buzzard and Gee 2013, Theorem A; Ghate and Rai 2025, Theorem 1.1], respectively, we obtain the following theorems.

Theorem 1.4. *Let $p \geq 3$ and $k = 3$. We have the following dichotomy for the shape of the semisimplification of the reduction of the semistable representation $V_{k,\mathcal{L}}$:*

$$\overline{V}_{k,\mathcal{L}} \sim \begin{cases} \mathrm{ind}(\omega_2^2) & \text{if } \nu < \frac{1}{2}, \\ \mu_\lambda \omega \oplus \mu_{\lambda^{-1}}\omega & \text{if } \nu \geq \frac{1}{2}, \end{cases}$$

where

$$\lambda + \frac{1}{\lambda} = \overline{-p^{-\frac{1}{2}}(\mathcal{L}-1)}.$$

Theorem 1.5. *Let $p \geq 5$ and $k = 5$. We have the following tetrachotomy for the shape of the semisimplification of the reduction of the semistable representation $V_{k,\mathcal{L}}$:*

$$\overline{V}_{k,\mathcal{L}} \sim \begin{cases} \mathrm{ind}(\omega_2^4) & \text{if } \nu < -\frac{1}{2} \\ \mu_{\lambda_1}\omega^3 \oplus \mu_{\lambda_1^{-1}}\omega & \text{if } \nu = -\frac{1}{2}, \\ \mathrm{ind}(\omega_2^{3+p}) & \text{if } -\frac{1}{2} < \nu < \frac{1}{2}, \\ \mu_{\lambda_2}\omega^2 \oplus \mu_{\lambda_2^{-1}}\omega^2 & \text{if } \nu \geq \frac{1}{2}, \end{cases}$$

where the constants λ_i are given by

$$\lambda_1 = \overline{-3p^{\frac{1}{2}}\left(\mathcal{L} - \tfrac{5}{2}\right)}, \quad \lambda_2 + \frac{1}{\lambda_2} = \overline{2p^{-\frac{1}{2}}\left(\mathcal{L} - \tfrac{5}{2}\right)}.$$

Remark. We remark that recently the second author noticed that it is possible to reverse the arguments used to prove Theorem 1.3 to give a *proof* of the zig-zag conjecture, at least on inertia and for *all* half-integral slopes $0 < v_p(a_p) \leq \frac{1}{2}(p-3)$, [Ghate 2022, v1] (these restrictions can be removed [Ghate 2022, v2] using very recent work of the first two authors [Chitrao and Ghate 2023] which computes $\overline{V}_{k,\mathcal{L}}$ directly on $\mathrm{Gal}(\overline{\mathbb{Q}}_p/\mathbb{Q}_p)$ using the Iwahori mod p local Langlands correspondence [Chitrao 2025]).

1.5. ***Relation with local constancy.*** Berger [2012] proved the following theorem on the local constancy in the weight of the semisimplification of the reduction of crystalline representations.

Theorem 1.6 [Berger 2012, Theorem B]. *Let $a_p \neq 0$ and $k > 3v_p(a_p) + \alpha(k-1) + 1$, where*

$$\alpha(j) = \sum_{n \geq 1} \left\lfloor \frac{j}{p^{n-1}(p-1)} \right\rfloor.$$

Then there exists $m = m(k, a_p)$ such that $\overline{V}_{k',a_p} = \overline{V}_{k,a_p}$ if $k' \geq k$ and $k' - k \in p^{m-1}(p-1)\mathbb{Z}$.

This local constancy result does not extend to small weights k in the sense that the bound in the theorem is sharp. This was noticed in [Ghate 2021] using the following examples.

• Let $k = 4$ and $a_p = p \geq 5$. Then $4 \not> 3(1) + 1$ and the crystalline representation V_{k,a_p} does not satisfy Berger's bound on the weight. Now let $k' = 4 + p^n(p-1)$ for large n. If the above local constancy result were to hold for V_{k,a_p}, then we would have $\overline{V}_{k',a_p} \sim \overline{V}_{k,a_p}$. However, $\overline{V}_{k,a_p}$ is irreducible as k belongs to the range $[2, p+1]$ treated by Fontaine and Edixhoven [1992], but $\overline{V}_{k',a_p} \sim \mu_3\omega^2 \oplus \mu_{3^{-1}}\omega$ is reducible by [Bhattacharya, Ghate, and Rozensztajn 2018, Theorem 1.1].

• Let $k = 5$ and $a_p = p^{3/2}$ for $p \geq 7$. Then $5 \not> 3(1.5) + 1$ and the crystalline representation V_{k,a_p} again does not satisfy the bound on the weight. Now let $k' = 5 + p^n(p-1)$, for large n. If the above local constancy result were to hold for V_{k,a_p}, then we would have $\overline{V}_{k',a_p} \sim \overline{V}_{k,a_p}$. However, $\overline{V}_{k,a_p} \sim \mathrm{ind}(\omega_2^4)$ since k belongs to the Fontaine–Edixhoven range, but $\overline{V}_{k',a_p} \sim \mathrm{ind}(\omega_2^{3+p})$ by [Ghate and Rai 2025, Theorem 1.1].

However, we have the following corollary to Theorem 1.1.

Corollary 1.7. *For $k \geq 3$, the sequence of crystalline representations $V^*_{k+p^n(p-1),\,p^{r/2}}$ converges to the semistable representation $V^*_{k,\mathcal{L}}$ for $\mathcal{L} = 0$.*

Therefore, for large n the reductions of the crystalline representations $V_{k+p^n(p-1),\,p^{r/2}}$ in the examples above should be isomorphic to the reduction of the semistable representation $V_{k,\mathcal{L}}$ with $\mathcal{L} = 0$, and not necessarily to the reduction of the crystalline representation $V_{k,\,p^{r/2}}$ (though all may be the same, as is the case when $k = 3$). Indeed, one checks that

- if $k = 4$ and $p \geq 5$, then by [Breuil and Mézard 2002, Theorem 1.2], the reduction of the semistable representation $V_{4,\mathcal{L}}$ for $\mathcal{L} = 0$ is $\mu_3\omega^2 \oplus \mu_{3^{-1}}\omega$, and,
- if $k = 5$ and $p \geq 7$, then by [Guerberoff and Park 2019, Theorem 5.0.5] (or, better, by Theorem 1.5), the reduction of the semistable representation $V_{5,\mathcal{L}}$ for $\mathcal{L} = 0$ is $\mathrm{ind}(\omega_2^{3+p})$.

Thus, there is no apparent contradiction to local constancy in the weight for the reductions of the crystalline representations above when k is small if one works in the Colmez–Chenevier space which includes semistable representations of weight k.

2. The blow-up of U_r

In this section, we provide details about blow-ups in the rigid-analytic setting which may be of independent interest. This is inspired by [Schoutens 1995], but we carefully work out the details and also do not work over algebraically closed fields. For background on rigid-analytic geometry, see [Bosch, Güntzer, and Remmert 1984]. We also recall the important Proposition 2.5, which for each finite extension E of $\mathbb{Q}_p$ establishes a bijection between the E-valued points of $\widetilde{U}_r$ that lie above $(p^r, (1+p)^{k-1}-1)$ and the tangent directions in U_r at the exceptional point.

2.1. ***The blow-up as a rigid-analytic variety.*** The blow-up of U_r consists of a rigid-analytic variety $\widetilde{U}_r$ and a map $\pi : \widetilde{U}_r \to U_r$ satisfying the properties given in [Schoutens 1995, Definition 1.2.1]. In this subsection, we will construct $\widetilde{U}_r$.

Recall that $U_r = \operatorname{Sp} \mathcal{O}(U_r)$, where

$$\mathcal{O}(U_r) = \mathbb{Q}_p\langle S_1, S_2, T_1, T_2, T_3\rangle/(p^r T_1 - S_1, 1 - T_1T_2, pT_3 - S_2).$$

The $\mathbb{Q}_p$-valued point $(p^r, (1+p)^{k-1}-1)$ corresponds to the maximal ideal $m = (f_1, f_2)$ of $\mathcal{O}(U_r)$, where $f_1 = S_1 - p^r$ and $f_2 = S_2 - ((1+p)^{k-1}-1)$. By the blow-up of U_r at $(p^r, (1+p)^{k-1}-1)$, we mean the blow-up at the maximal ideal m.

We will construct $\widetilde{U}_r$ by a patching argument. Let Q_1, Q_2, Q_1' and Q_2' be indeterminates. For $i = 0, 1, \ldots$, consider the affinoid algebra

$$A_i = \mathcal{O}(U_r)\langle p^i Q_1\rangle/(f_2 - Q_1 f_1),$$

which describes the subset of U_r where $|f_2| \le p^i |f_1|$. Sending $p^{i+1}Q_1$ to $p \cdot p^i Q_1$ and X to $p^i Q_1$ induces an isomorphism $A_{i+1}\langle X\rangle/(pX - p^{i+1}Q_1) \cong A_i$, showing that $\operatorname{Sp} A_i$ is the affinoid subdomain (hence an open subvariety) of $\operatorname{Sp} A_{i+1}$ consisting of maximal ideals $\mathfrak{m}$ such that $|Q_1 \bmod \mathfrak{m}| \le p^i$. We therefore get a sequence of inclusions $\operatorname{Sp} A_0 \to \operatorname{Sp} A_1 \to \operatorname{Sp} A_2 \to \cdots$ associated to the sequence of affinoid algebra homomorphisms $\cdots \to A_2 \to A_1 \to A_0$. Using [Bosch, Güntzer, and Remmert 1984, Proposition 9.3.2/1] we paste together the $\operatorname{Sp} A_i$ for $i \ge 0$ to get a rigid-analytic variety $\widetilde{V}_1$. The same proposition also states that $\{\operatorname{Sp} A_i\}_{i\ge 0}$ is an admissible cover of $\widetilde{V}_1$. Similarly, for $i \ge 0$, we define

$$B_i = \mathcal{O}(U_r)\langle p^i Q_2\rangle/(f_1 - Q_2 f_2)$$

and glue the corresponding affinoid spaces to get a rigid-analytic variety $\widetilde{V}_2$. The blow-up $\widetilde{U}_r$ is the rigid-analytic variety obtained by gluing $\widetilde{V}_1$ and $\widetilde{V}_2$ along certain open subvarieties.

We describe these open subvarieties now. For each $i \ge 0$, consider the affinoid algebra

$$A_i' = \mathcal{O}(U_r)\langle p^i Q_1, p^i Q_1'\rangle/(f_2 - Q_1 f_1, 1 - Q_1 Q_1'),$$

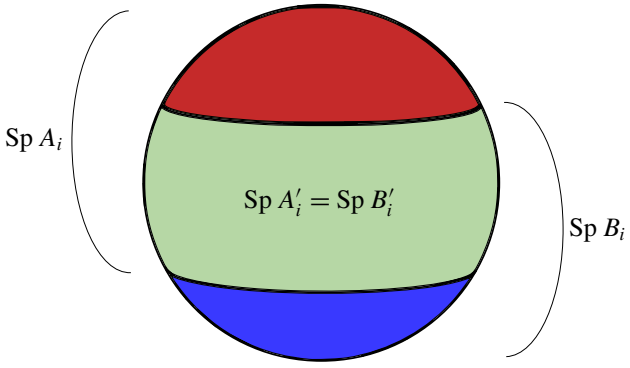


Figure 2. Exceptional fiber of the blow-up in the rigid setting.

which describes a Laurent subdomain $\mathrm{Sp}\, A'_i$ of $\mathrm{Sp}\, A_i$ given by the condition $|Q_1| \geq p^{-i}$. Since we have an isomorphism $A'_{i+1}\langle X, Y\rangle/(pX - p^{i+1}Q_1, pY - p^{i+1}Q'_1) \cong A'_i$ (sending X to $p^i Q_1$ and Y to $p^i Q'_1$), we see that $\mathrm{Sp}\, A'_i$ is an affinoid subdomain (and hence an open subvariety) of $\mathrm{Sp}\, A'_{i+1}$. Using [Bosch, Güntzer, and Remmert 1984, Proposition 9.3.2/1], we paste together $\mathrm{Sp}\, A'_i$ for $i \geq 0$ to get a rigid-analytic variety $\widetilde{V}'_1$. Moreover, using [loc. cit., Proposition 9.3.3/1], we see that the canonical inclusions $\mathrm{Sp}\, A'_i \to \mathrm{Sp}\, A_i$ induce an inclusion $\widetilde{V}'_1 \hookrightarrow \widetilde{V}_1$, identifying $\widetilde{V}'_1$ as a subvariety of $\widetilde{V}_1$. Similarly, we construct a subvariety $\widetilde{V}'_2$ of $\widetilde{V}_2$ by gluing all the $\mathrm{Sp}\, B'_i$ together, where

$$B'_i = \mathcal{O}(U_r)\langle p^i Q_2, p^i Q'_2\rangle/(f_1 - Q_2 f_2, 1 - Q_2 Q'_2).$$

It turns out that $\mathrm{Sp}\, A'_i$ and $\mathrm{Sp}\, B'_i$ are the intersections of $\mathrm{Sp}\, A_i$ and $\mathrm{Sp}\, B_i$ in the blow-up $\widetilde{U}_r$ that we will soon construct. The spaces above are summarized by Figure 2.

The space $\widetilde{V}_1$ corresponds to the surface of the sphere except the south pole in Figure 2, whereas $\widetilde{V}_2$ corresponds to the surface of the sphere except the north pole.

We claim that $\widetilde{V}'_j$ is an open subvariety of $\widetilde{V}_j$ for $j = 1, 2$. Without loss of generality, assume $j = 1$. To prove this claim, we need to show that $\widetilde{V}'_1 \cap \mathrm{Sp}\, A_i$ is an admissible open subset of $\mathrm{Sp}\, A_i$ for each $i \geq 0$. Write $\widetilde{V}'_1 \cap \mathrm{Sp}\, A_i = \bigcup_{k=i}^{\infty}(\mathrm{Sp}\, A'_k \cap \mathrm{Sp}\, A_i)$. If $k \geq i$, then $\mathrm{Sp}\, A'_k \cap \mathrm{Sp}\, A_i$ is the set of maximal ideals $\mathfrak{m}$ of A_i satisfying $|Q_1 \bmod \mathfrak{m}| \geq p^{-k}$. Therefore $\widetilde{V}'_1 \cap \mathrm{Sp}\, A_i$ is the set of maximal ideals $\mathfrak{m}$ of A_i such that $|Q_1 \bmod \mathfrak{m}| > 0$. In other words, it is the complement of the vanishing set of Q_1 in $\mathrm{Sp}\, A_i$; i.e., it is a Zariski open subset. Since Zariski open subsets are admissible open, we have proved the claim.

To glue $\widetilde{V}_1$ and $\widetilde{V}_2$, we define an isomorphism $\phi : \widetilde{V}'_1 \to \widetilde{V}'_2$ as follows. Recall that, for $i \geq 0$,

$$A'_i = \mathcal{O}(U_r)\langle p^i Q_1, p^i Q'_1\rangle/(f_2 - Q_1 f_1, 1 - Q_1 Q'_1),$$
$$B'_i = \mathcal{O}(U_r)\langle p^i Q_2, p^i Q'_2\rangle/(f_1 - Q_2 f_2, 1 - Q_2 Q'_2).$$

Consider an isomorphism of affinoid algebras $B'_i \to A'_i$ given by

$$p^i Q_2 \to p^i Q'_1,$$
$$p^i Q'_2 \to p^i Q_1.$$

This isomorphism gives rise to an isomorphism $\phi_i : \operatorname{Sp} A'_i \to \operatorname{Sp} B'_i$ of the corresponding affinoid spaces. For $i, j \geq 0$, these maps clearly restrict to the same map on $\operatorname{Sp} A'_i \cap \operatorname{Sp} A'_j$. Therefore applying [Bosch, Güntzer, and Remmert 1984, Proposition 9.3.3/1], we get an isomorphism of rigid-analytic varieties $\phi : \widetilde{V}'_1 \to \widetilde{V}'_2$.

Define $\widetilde{U}_r$ to be the rigid-analytic variety obtained by gluing $\widetilde{V}_1$ and $\widetilde{V}_2$ along $\widetilde{V}'_1$ and $\widetilde{V}'_2$, respectively, using the isomorphism ϕ. This rigid-analytic variety is analogous to the blow-up in the classical algebraic geometry setting.

We see this by computing its E-valued points for any finite extension E of $\mathbb{Q}_p$. Since $\widetilde{U}_r$ is covered by $\widetilde{V}_1$ and $\widetilde{V}_2$, we compute $\widetilde{V}_1(E)$ and $\widetilde{V}_2(E)$ first. Recall, for $i \geq 0$,

$$A_i = \mathcal{O}(U_r)\langle p^i Q_1 \rangle / (f_2 - Q_1 f_1).$$

We identify the set of E-valued points of $\operatorname{Sp} A_i$ with

$$\{(s_1, s_2, q_1) \in U_r(E) \times E \mid (1 + s_2 - (1+p)^{k-1}) = q_1(s_1 - p^r),\ |q_1| \leq p^i\}.$$

The first condition is a consequence of the relation $f_2 = Q_1 f_1$. For $i \geq j$, the gluing map $A_i \to A_j$ induces the canonical inclusion map on E-valued points $\operatorname{Sp} A_j(E) \to \operatorname{Sp} A_i(E)$. Therefore, the set of E-valued points of $\widetilde{V}_1$ is the union of all the $\operatorname{Sp} A_i(E)$, i.e.,

$$\widetilde{V}_1(E) = \{(s_1, s_2, q_1) \in U_r(E) \times E \mid (1 + s_2 - (1+p)^{k-1}) = q_1(s_1 - p^r)\}.$$

We similarly have

$$\widetilde{V}_2(E) = \{(s_1, s_2, q_2) \in U_r(E) \times E \mid (s_1 - p^r) = q_2(1 + s_2 - (1+p)^{k-1})\}.$$

We can now describe $\widetilde{U}_r(E)$. If P is an E-valued point of $\widetilde{V}_1 \cap \widetilde{V}_2$, it is an E-valued point of $\operatorname{Sp} A_i \cap \operatorname{Sp} B_i = \operatorname{Sp} A'_i$ for some $i \geq 0$. Being a point of $\operatorname{Sp} A_i$, it is of the form (s_1, s_2, q_1) for some $s_1, s_2, q_1 \in E$. Since the gluing map between B'_i and A'_i takes S_1 to S_1, S_2 to S_2 and Q_2 to the inverse of Q_1 in A'_i, we see that (s_1, s_2, q_1^{-1}) represents P as an E-valued point of $\operatorname{Sp} B_i$ (the condition $|q_1^{-1}| \leq p^i$ is satisfied because $P \in \operatorname{Sp} A'_i$ implies $|q_1| \geq p^{-i}$). We can therefore identify $(s_1, s_2, q_1) \in \widetilde{V}_1(E)$ with the point $(s_1, s_2, 1 : q_1) \in U_r(E) \times \mathbb{P}^1(E)$ and $(s_1, s_2, q_2) \in \widetilde{V}_2(E)$ with the point $(s_1, s_2, q_2 : 1) \in U_r(E) \times \mathbb{P}^1(E)$, so that a point in the intersection goes to the same point in $U_r(E) \times \mathbb{P}^1(E)$ under both of these identifications. As a result of this discussion, we see that

$$\widetilde{U}_r(E) = \{(s_1, s_2, \xi_1 : \xi_2) \in U_r(E) \times \mathbb{P}^1(E) \mid (s_1 - p^r)\xi_2 = (1 + s_2 - (1+p)^{k-1})\xi_1\} \tag{5}$$

exactly as in the classical algebraic geometry setting.

2.2. ***The blow-up map $\pi : \widetilde{U}_r \to U_r$.*** In this subsection, we define a candidate for the blow-up map $\pi : \widetilde{U}_r \to U_r$ using [Bosch, Güntzer, and Remmert 1984, Proposition 9.3.3/1].

To define π, we first define its restrictions π_i to $\widetilde{V}_i$ for $i = 1, 2$. We first define the map $\pi_1 : \widetilde{V}_1 \to U_r$. For each $i \geq 0$, consider the map $\operatorname{Sp} A_i \to U_r$ associated to the canonical homomorphism

$$\mathcal{O}(U_r) \to \mathcal{O}(U_r)\langle p^i Q_1 \rangle / (f_2 - Q_1 f_1) = A_i.$$

These maps clearly restrict to the same map on $\operatorname{Sp} A_i \cap \operatorname{Sp} A_j$ for $i, j \geq 0$. So using [Bosch, Güntzer, and Remmert 1984, Proposition 9.3.3/1], we glue these maps together to get a map $\pi_1 : \widetilde{V}_1 \to U_r$ of rigid-analytic varieties. The other map $\pi_2 : \widetilde{V}_2 \to U_r$ is constructed similarly.

Now, to get a map $\pi : \widetilde{U}_r \to U_r$, we have to check that π_1 and π_2 agree on the intersection of $\widetilde{V}_1$ and $\widetilde{V}_2$ in $\widetilde{U}_r$. Since this intersection is equal to $\widetilde{V}_1'$ ($\cong \widetilde{V}_2'$), we have to check that the following diagram commutes:

$$\begin{array}{ccc} \widetilde{V}_1' & \xrightarrow{\pi_1|_{\widetilde{V}_1'}} & \\ \downarrow{\scriptstyle \phi} & & U_r \\ \widetilde{V}_2' & \xrightarrow{\pi_2|_{\widetilde{V}_2'}} & \end{array}$$

Since ϕ is obtained by pasting the maps $\phi_i : \operatorname{Sp} A_i' \to \operatorname{Sp} B_i'$, the commutativity of the diagram above is equivalent to the commutativity of the following diagram for all $i \geq 0$:

$$\begin{array}{ccc} \operatorname{Sp} A_i' & \xrightarrow{\pi_1|_{\operatorname{Sp} A_i'}} & \\ \downarrow{\scriptstyle \phi_i} & & U_r \\ \operatorname{Sp} B_i' & \xrightarrow{\pi_2|_{\operatorname{Sp} B_i'}} & \end{array}$$

The commutativity of the diagram above can be checked by reversing all the arrows and noting that the resulting maps are $\mathcal{O}(U_r)$-algebra homomorphisms. This shows that π_1 and π_2 agree on the intersection of $\widetilde{V}_1$ and $\widetilde{V}_2$ in $\widetilde{U}_r$. Therefore, we glue π_1 and π_2 using [loc. cit., Proposition 9.3.3/1] to get a map $\pi : \widetilde{U}_r \to U_r$.

2.3. ***Proof that $\pi : \widetilde{U}_r \to U_r$ is the blow-up.*** In this subsection, we prove that the map $\pi : \widetilde{U}_r \to U_r$ defined in the previous subsection is the blow-up of U_r at the maximal ideal $m = (f_1, f_2)$. To prove this, we need to check that π satisfies the two properties stated in [Schoutens 1995, Definition 1.2.1], namely the invertibility of a certain sheaf, and a corresponding universal property.

We first show that the ideal sheaf on U_r corresponding to the ideal m of $\mathcal{O}(U_r)$ extends to an invertible ideal sheaf on $\widetilde{U}_r$ in the sense of [loc. cit., Definition 1.1.1]. Since this is a local criterion (this means that the criterion can be checked over an admissible cover), we need to check that, for all $i \geq 0$, the maps $\pi|_{\operatorname{Sp} A_i} : \operatorname{Sp} A_i \to U_r$ and $\pi|_{\operatorname{Sp} B_i} : \operatorname{Sp} B_i \to U_r$ satisfy the same property. Let us prove this statement for $\pi|_{\operatorname{Sp} A_i} : \operatorname{Sp} A_i \to U_r$. In this case, the extended ideal sheaf corresponds to the ideal mA_i on $\operatorname{Sp} A_i$. Moreover, [loc. cit., Proposition 1.1.4] states that the invertibility of this sheaf is equivalent to the invertibility of the ideal $mA_{i,\mathfrak{m}}$ of $A_{i,\mathfrak{m}}$ for all maximal ideals $\mathfrak{m}$ of A_i, where $A_{i,\mathfrak{m}}$ is the localization of A_i at $\mathfrak{m}$. To show this, it is enough to prove that mA_i is generated by a regular element of A_i because regular elements go to regular elements under flat base change.

Since mA_i is generated by f_1, we prove that f_1 is not a zero divisor in A_i for $i \geq 0$. Note that we only need to prove this statement for $i = 0$ because of the injections $A_i \hookrightarrow A_0$.

Lemma 2.1. *f_1 is not a zero divisor in $A_0 = \mathcal{O}(U_r)\langle Q_1\rangle/(f_2 - Q_1 f_1)$.*

Proof. Recall that

$$\mathcal{O}(U_r) = \mathbb{Q}_p\langle S_1, S_2, T_1, T_2, T_3\rangle/(p^r T_1 - S_1, 1 - T_1 T_2, pT_3 - S_2).$$

To prove that f_1 is not a zero divisor in A_0, it is enough to prove that f_1 is not a zero divisor in $\mathbb{Q}_p\langle S_1, S_2, Q_1\rangle/(f_2 - Q_1 f_1)$. Indeed, since $\operatorname{Sp} A_0$ is an affinoid subdomain (a Laurent subdomain) of $\operatorname{Sp}\mathbb{Q}_p\langle S_1, S_2, Q_1\rangle/(f_2 - Q_1 f_1)$, the canonical map $\mathbb{Q}_p\langle S_1, S_2, Q_1\rangle/(f_2 - Q_1 f_1) \to A_0$ is flat by [Bosch, Güntzer, and Remmert 1984, Corollary 7.3.2/6]. Therefore, if f_1 is not a zero divisor in the former then, by flatness, it is not a zero divisor in the latter.

So let us now prove that f_1 is not a zero divisor in $\mathbb{Q}_p\langle S_1, S_2, Q_1\rangle/(f_2 - Q_1 f_1)$. Assume that there exists $h \in \mathbb{Q}_p\langle S_1, S_2, Q_1\rangle$ such that $f_2 - Q_1 f_1$ divides $f_1 h$. We know that

$$f_2 - Q_1 f_1 = 1 + S_2 - (1+p)^{k-1} - Q_1(S_1 - p^r)$$

is S_2-distinguished of degree 1 (see [loc. cit., Definition 5.2.1/1]). Applying the Weierstrass division theorem (see [loc. cit., Theorem 5.2.1/2]) to h and $f_2 - Q_1 f_1$, we see that there exist unique power series $q \in \mathbb{Q}_p\langle S_1, S_2, Q_1\rangle$ and $r \in \mathbb{Q}_p\langle S_1, Q_1\rangle$ (note that S_2 does not appear in r because $f_2 - Q_1 f_1$ is of degree 1 in S_2) such that $h = q(f_2 - Q_1 f_1) + r$. The fact that $f_2 - Q_1 f_1$ divides $f_1 h$ implies that $f_2 - Q_1 f_1$ divides $f_1 r$. This is possible only if $r = 0$, i.e., only if $f_2 - Q_1 f_1$ divides h. Therefore $h = 0$ in $\mathbb{Q}_p\langle S_1, S_2, Q_1\rangle/(f_2 - Q_1 f_1)$. □

Having checked the first property of blow-ups for π, we now check the second property. Using [Schoutens 1995, Lemma 1.2.4], we note that it is enough to check it for affinoid spaces. In other words, given any affinoid space $Y = \operatorname{Sp} R$ and a map of rigid-analytic varieties $f : Y \to U_r$ such that the ideal sheaf on Y associated to the ideal mR is invertible, we prove that there exists a unique map $g : Y \to \widetilde{U}_r$ such that the following diagram commutes:

$$\begin{array}{ccc} & & \widetilde{U}_r \\ & \overset{g}{\nearrow} & \big\downarrow \pi \\ Y & \xrightarrow{\ f\ } & U_r \end{array}$$

For $i = 1, 2$, define $\mathfrak{a}_i = (f^*(f_i)R : mR)$ and let $Y_i = Y \setminus V(\mathfrak{a}_i)$. We prove two lemmas about the Y_i.

Lemma 2.2. *For $i = 1, 2$, a point $y \in Y$ belongs to Y_i if and only if $m\mathcal{O}_{Y,y} = f^*(f_i)\mathcal{O}_{Y,y}$.*

Proof. We thank one of the referees for pointing out the useful [Atiyah and Macdonald 1969, Corollary 3.15], which shortens the proof. Let y be a point in Y and $\mathfrak{m}_y$ be the corresponding maximal ideal of R. Fix $i = 1, 2$. Using [loc. cit., Corollary 3.15], we see that $\mathfrak{a}_i R_{\mathfrak{m}_y} = (f^*(f_i)R_{\mathfrak{m}_y} : mR_{\mathfrak{m}_y})$. Now,

$$y \in Y_i \iff \mathfrak{a}_i R_{\mathfrak{m}_y} = R_{\mathfrak{m}_y} \iff mR_{\mathfrak{m}_y} = f^*(f_i)R_{\mathfrak{m}_y}.$$

Using [Bosch, Güntzer, and Remmert 1984, Proposition 7.3.2/3], we have a map $R_{\mathfrak{m}_y} \to \mathcal{O}_{Y,y}$, which induces an isomorphism $\widehat{R}_{\mathfrak{m}_y} \simeq \widehat{\mathcal{O}}_{Y,y}$. We conclude that

$$mR_{\mathfrak{m}_y} = f^*(f_i)R_{\mathfrak{m}_y} \quad \Longleftrightarrow \quad m\mathcal{O}_{Y,y} = f^*(f_i)\mathcal{O}_{Y,y}.$$

Indeed, the forward implication is obtained by extending scalars and the reverse implication follows by extending scalars to the completion $\widehat{\mathcal{O}}_{Y,y}$ and using the fact that $R_{\mathfrak{m}_y} \to \widehat{R}_{\mathfrak{m}_y}$ is faithfully flat. □

Using [Schoutens 1995, Lemma 1.1.2] and the lemma above one easily checks that $Y = Y_1 \cup Y_2$. The sets Y_1 and Y_2, being Zariski open subsets of Y, are admissible open and form an admissible cover of Y.

Before proving the next lemma, we remark that, by [loc. cit., Lemma 0.4], we have $\mathfrak{a}_i\mathcal{O}(Y_i) = \mathcal{O}(Y_i)$, which implies that $m\mathcal{O}(Y_i) = f^*(f_i)\mathcal{O}(Y_i)$ for $i = 1, 2$.

Lemma 2.3. *Let $i = 1, 2$. Then, $f^*(f_i)$ is not a zero divisor in $\mathcal{O}(Y')$ for any admissible open subset Y' of Y contained in Y_i.*

Proof. The proof has three steps. In the first step, we prove that $f^*(f_i)$ is not a zero divisor in $\mathcal{O}_{Y,y}$ for any $y \in Y'$. In the second step, we prove that the canonical map $\mathcal{O}(Y') \to \prod_{y\in Y'} \mathcal{O}_{Y,y}$ is injective. In the third step, we use these two facts to conclude that $f^*(f_i)$ is not a zero divisor in $\mathcal{O}(Y')$. Let $i = 1$ or 2.

• Using the remark preceding this lemma, we get $m\mathcal{O}(Y') = f^*(f_i)\mathcal{O}(Y')$. Extending to the stalks, we get $m\mathcal{O}_{Y,y} = f^*(f_i)\mathcal{O}_{Y,y}$ for each $y \in Y'$. Since the sheaf associated to the ideal $m\mathcal{O}(Y)$ is invertible, we see that $m\mathcal{O}_{Y,y}$ is generated by a regular element of $\mathcal{O}_{Y,y}$ for each $y \in Y$. Therefore $f^*(f_i)$ is not a zero divisor in $\mathcal{O}_{Y,y}$ for each $y \in Y'$.

• Consider the map $\mathcal{O}(Y') \to \prod_{y\in Y'} \mathcal{O}_{Y,y}$. Let a be an element of $\mathcal{O}(Y')$ that maps to 0 in $\mathcal{O}_{Y,y}$ for each $y \in Y'$. Choose an admissible cover $\{U_j\}_{j\in J}$ of Y' by affinoid subdomains of Y. For any $j \in J$, the image of a in $\mathcal{O}_{Y,y}$ is 0 for each $y \in U_j$. Using [Bosch, Güntzer, and Remmert 1984, Corollary 7.3.2/4], we see that the restriction of a to $\mathcal{O}(U_j)$ is 0 for each $j \in J$. Since the cover $\{U_j\}_{j\in J}$ is admissible, we see that $a = 0$ in $\mathcal{O}(Y')$.

• Suppose there exists a $b \in \mathcal{O}(Y')$ such that $f^*(f_i)b = 0$. This means that $f^*(f_i)b = 0$ in $\mathcal{O}_{Y,y}$ for each $y \in Y'$. Using the fact that $f^*(f_i)$ is not a zero divisor in $\mathcal{O}_{Y,y}$ for $y \in Y'$, we see that the image of b in $\mathcal{O}_{Y,y}$ is 0. The injectivity of the map $\mathcal{O}(Y') \to \prod_{j\in J} \mathcal{O}_{Y,y}$ implies that $b = 0$ in $\mathcal{O}(Y')$. Therefore $f^*(f_i)$ is not a zero divisor in $\mathcal{O}(Y')$. □

We now construct $g : Y \to \widetilde{U}_r$. In order to do this, we use the above lemmas to define the restrictions of g to the elements of a refinement of the cover $Y = Y_1 \cup Y_2$ and then apply the usual patching argument.

As $f^*(f_2) \in m\mathcal{O}(Y_1) = f^*(f_1)\mathcal{O}(Y_1)$, there exists a $q_1 \in \mathcal{O}(Y_1)$ such that $f^*(f_2) = q_1 f^*(f_1)$. Similarly, there exists a $q_2 \in \mathcal{O}(Y_2)$ such that $f^*(f_1) = q_2 f^*(f_2)$. Furthermore, q_1 and q_2 are unique because of Lemma 2.3.

Since Y_1 is an admissible open subset of Y, there exists an admissible cover $\{Y_{1,m}\}_{m\in I}$ of Y_1 by affinoid subdomains $Y_{1,m}$ of Y. Here and just below m is an element of the indexing set I and should not confused with the maximal ideal m of $\mathcal{O}(U_r)$ used throughout this paper. For each $m \in I$, there

is a restriction map $\mathcal{O}(Y_1) \to \mathcal{O}(Y_{1,m})$ and so we can think of q_1 as an element of $\mathcal{O}(Y_{1,m})$. For each $m \in I$, fix a natural number $\underline{m}$ such that $|p^{\underline{m}} q_1| \leq 1$. Consider the affinoid algebra map $f^* : \mathcal{O}(U_r) \to \mathcal{O}(Y)$ and compose it with the restriction map $\mathcal{O}(Y) \to \mathcal{O}(Y_{1,m})$. Extend this composition to a map $\mathcal{O}(U_r)\langle p^{\underline{m}} Q_1 \rangle \to \mathcal{O}(Y_{1,m})$ by sending $p^{\underline{m}} Q_1$ to $p^{\underline{m}} q_1$. This is possible because of [Bosch, Güntzer, and Remmert 1984, Corollary 1.4.3/2]. Since we have the relation $f^*(f_1) = q_1 f^*(f_2)$ in $\mathcal{O}(Y_{1,m})$, we see that this extension factors through an affinoid algebra map $A_{\underline{m}} \to \mathcal{O}(Y_{1,m})$. Let $g_{1,m} : Y_{1,m} \to \operatorname{Sp} A_{\underline{m}}$ be the corresponding map of affinoid spaces. Similarly, there is an admissible cover $\{Y_{2,n}\}_{n \in J}$ of Y_2 by affinoid subdomains $Y_{2,n}$ of Y and, for each $n \in J$, a fixed natural number $\underline{n}$ such that $|p^{\underline{n}} q_2| \leq 1$ and a map of affinoid spaces $g_{2,n} : Y_{2,n} \to \operatorname{Sp} B_{\underline{n}}$ associated to the map of affinoid algebras $B_{\underline{n}} \to \mathcal{O}(Y_{2,n})$ sending $p^{\underline{n}} Q_2$ to $p^{\underline{n}} q_2$. Since $g_{1,m}^*$ and $g_{2,n}^*$ are $\mathcal{O}(U_r)$-algebra homomorphisms, we see that $\pi \circ g_{1,m} = f$ on $Y_{1,m}$ and $\pi \circ g_{2,n} = f$ on $Y_{2,n}$ for $m \in I,\ n \in J$.

Note that $\{Y_{1,m}, Y_{2,n}\}_{m \in I, n \in J}$ is an admissible cover of Y. To obtain $g : Y \to \widetilde{U}_r$, we glue all the maps $g_{1,m}$ and $g_{2,n}$. Without loss of generality, assume $\underline{m} \geq \underline{m}'$. Then, $g_{1,m}$ and $g_{1,m'}$ agree on $Y_{1,m} \cap Y_{1,m'}$ because the following diagram commutes:

$$\begin{array}{ccc} A_{\underline{m}'} & \xrightarrow{\ g_{1,m'}^*\ } & \mathcal{O}(Y_{1,m} \cap Y_{1,m'}) \\ \uparrow & & \\ A_{\underline{m}} & \xrightarrow{\ g_{1,m}^*\ } & \end{array}$$

A similar argument shows that $g_{2,n}$ and $g_{2,n'}$ agree on $Y_{2,n} \cap Y_{2,n'}$.

We check that the maps $g_{1,m}$ and $g_{2,n}$ agree on $Y_{1,m} \cap Y_{2,n}$. Let $N \geq \underline{m}, \underline{n}$. Think of $g_{1,m}$ as a map from $Y_{1,m}$ to $\operatorname{Sp} A_N$ (by composing with the map $\operatorname{Sp} A_{\underline{m}} \to \operatorname{Sp} A_N$) and $g_{2,n}$ as a map from $Y_{2,n}$ to $\operatorname{Sp} B_N$ (by composing with the map $\operatorname{Sp} B_{\underline{n}} \to \operatorname{Sp} B_N$). We claim that $g_{1,m}$ maps $Y_{1,m} \cap Y_{2,n}$ into $\operatorname{Sp} A'_N$ and that $g_{2,n}$ maps $Y_{1,m} \cap Y_{2,n}$ into $\operatorname{Sp} B'_N$. To show this, we first prove that q_1 is the inverse of q_2 in $\mathcal{O}(Y_{1,m} \cap Y_{2,n})$. Transfer the relations $f^*(f_2) = q_1 f^*(f_1)$ and $f^*(f_1) = q_2 f^*(f_2)$ from $\mathcal{O}(Y_1)$ and $\mathcal{O}(Y_2)$, respectively, to $\mathcal{O}(Y_{1,m} \cap Y_{2,n})$ by the restriction maps. Substituting the first relation into the second relation, we get $f^*(f_1) = q_2 q_1 f^*(f_1)$. Using Lemma 2.3, we cancel $f^*(f_1)$ to see that q_1 is indeed the inverse of q_2 in $\mathcal{O}(Y_{1,m} \cap Y_{2,n})$.

Compose $g_{1,m}^* : A_N \to \mathcal{O}(Y_{1,m})$ with the restriction map $\mathcal{O}(Y_{1,m}) \to \mathcal{O}(Y_{1,m} \cap Y_{2,n})$. Extend this composition to a map $A_N\langle p^N Q'_1 \rangle \to \mathcal{O}(Y_{1,m} \cap Y_{2,n})$ by sending $p^N Q'_1$ to $p^N q_2$. This is possible because $|p^N q_2| \leq 1$ since the intersection of two affinoid subdomains of an affinoid space is again an affinoid space and because under the homomorphism $\mathcal{O}(Y_{2,n}) \to \mathcal{O}(Y_{1,m} \cap Y_{2,n})$ of affinoid algebras, the image of an element has norm at most the norm of the element itself. Using the fact that q_2 is the inverse of q_1 in $\mathcal{O}(Y_{1,m} \cap Y_{2,n})$, we may further extend the above map to A'_N to obtain the following commutative diagram:

$$\begin{array}{ccc} A_N = \mathcal{O}(U_r)\langle p^N Q_1 \rangle / (f_2 - Q_1 f_1) & \xrightarrow{\ g_{1,m}^*\ } & \mathcal{O}(Y_{1,m}) \\ \downarrow & & \downarrow \\ A'_N = \mathcal{O}(U_r)\langle p^N Q_1,\, p^N Q'_1 \rangle / (f_2 - Q_1 f_1,\, 1 - Q_1 Q'_1) & \longrightarrow & \mathcal{O}(Y_{1,m} \cap Y_{2,n}) \end{array}$$

This shows that the affinoid algebra map $A_N \to \mathcal{O}(Y_{1,m} \cap Y_{2,n})$ factors through the map $A_N \to A'_N$. Reversing the arrows, we see that the restriction of $g_{1,m}$ to $Y_{1,m} \cap Y_{2,n}$ factors through the inclusion $\operatorname{Sp} A'_N \to \operatorname{Sp} A_N$. In other words, $g_{1,m}$ maps $Y_{1,m} \cap Y_{2,n}$ into $\operatorname{Sp} A'_N$. Similarly, we can prove that $g_{2,n}$ maps $Y_{1,m} \cap Y_{2,n}$ into $\operatorname{Sp} B'_N$.

Now we can prove that $g_{1,m}$ and $g_{2,n}$ agree on $Y_{1,m} \cap Y_{2,n}$. Indeed, this statement is equivalent to the commutativity of the following diagram:

$$\begin{array}{ccc} A'_N & \xrightarrow{g^*_{1,m}} & \\ {\scriptstyle \phi^*_N}\uparrow & & \mathcal{O}(Y_{1,m} \cap Y_{2,n}) \\ B'_N & \xrightarrow[g^*_{2,n}]{} & \end{array}$$

where the vertical map is the gluing map. This diagram commutes because under the top two maps $Q_2 \mapsto Q'_1 \mapsto q_2$ and $Q'_2 \mapsto Q_1 \mapsto q_1$, and these are exactly the images under the lower map.

By [Bosch, Güntzer, and Remmert 1984, Proposition 9.3.3/1], there exists a map of rigid-analytic spaces $g : Y \to \widetilde{U}_r$ such that the following diagram commutes:

$$\begin{array}{ccc} & & \widetilde{U}_r \\ & {\scriptstyle g}\nearrow & \downarrow{\scriptstyle \pi} \\ Y & \xrightarrow{f} & U_r \end{array}$$

To prove that $\pi : \widetilde{U}_r \to U_r$ is the blow-up of U_r at the ideal $m = (f_1, f_2)$, it remains to check that g is unique in making the above diagram commute. Suppose $g' : Y \to \widetilde{U}_r$ is another candidate. We prove that $g = g'$. We need:

Lemma 2.4. *For $i = 1, 2$, we have $g'^{-1}(\widetilde{V}_i) = Y_i$.*

Proof. Without loss of generality, assume that $i = 1$.

• We prove that $g'^{-1}(\widetilde{V}_1) \subseteq Y_1$. Recall that $\widetilde{V}_1 = \bigcup_{i \geq 0} \operatorname{Sp} A_i$. Therefore it is enough to prove that $Y_{1,i} := g'^{-1}(\operatorname{Sp} A_i) \subseteq Y_1$ for each $i \geq 0$. (The notation $Y_{1,i}$ here and $Y_{2,i}$ below are local to this proof.) Fix $i \geq 0$. Consider the following commutative diagram:

$$\begin{array}{ccc} & & \operatorname{Sp} A_i \\ & {\scriptstyle g'}\nearrow & \downarrow{\scriptstyle \pi} \\ Y_{1,i} = g'^{-1}(\operatorname{Sp} A_i) & \xrightarrow{f} & U_r \end{array}$$

We know that $mA_i = f_1 A_i$. Extending this ideal from A_i to $\mathcal{O}(Y_{1,i})$ using g'^*, we see that $m\mathcal{O}(Y_{1,i}) = g'^*(f_1)\mathcal{O}(Y_{1,i}) = f^*(f_1)\mathcal{O}(Y_{1,i})$. Let $y \in Y_{1,i}$. Noting that $Y_{1,i}$ is an admissible open subset of Y, we have $m\mathcal{O}_{Y,y} = f^*(f_1)\mathcal{O}_{Y,y}$. By Lemma 2.2, we see that $y \in Y_1$.

• We prove that $Y_1 \subseteq g'^{-1}(\widetilde{V}_1)$. This is equivalent to proving $g'^{-1}(\widetilde{U}_r \setminus \widetilde{V}_1) \subseteq Y \setminus Y_1$. Recall that $\widetilde{U}_r$ is covered by $\widetilde{V}_1$ and $\widetilde{V}_2$. Since $\widetilde{V}_2 = \bigcup_{i \geq 0} \operatorname{Sp} B_i$, we have

$$\widetilde{U}_r \setminus \widetilde{V}_1 = \bigcup_{i \geq 0} (\operatorname{Sp} B_i \setminus \widetilde{V}_1).$$

Thus, it is enough to prove that $Y_{2,i} := g'^{-1}(\operatorname{Sp} B_i \setminus \widetilde{V}_1) \subseteq Y \setminus Y_1$ for each $i \geq 0$. Fix $i \geq 0$. We prove that if $y \in Y_{2,i}$, then $y \notin Y_1$. We do so by proving that the ideal $m\mathcal{O}_{Y,y}$ of $\mathcal{O}_{Y,y}$ is not generated by $f^*(f_1)$ (see Lemma 2.2). So, let $y \in Y_{2,i}$ map to $\operatorname{Sp} B_i \setminus \widetilde{V}_1$ under g'. Assume towards a contradiction that $y \in Y_1$. By Lemma 2.2, we see that $m\mathcal{O}_{Y,y}$ is generated by $f^*(f_1)$. By the analog of the first bullet point, $Y_{2,i} \subseteq Y_2$, so we also have $m\mathcal{O}_{Y,y} = f^*(f_2)\mathcal{O}_{Y,y}$. Therefore there exists a unit $u \in \mathcal{O}_{Y,y}$ such that $f^*(f_1) = uf^*(f_2)$. Consider the following commutative diagram:

$$\begin{array}{ccc} & & \operatorname{Sp} B_i \\ & \overset{g'}{\nearrow} & \downarrow \pi \\ Y_{2,i} = g'^{-1}(\operatorname{Sp} B_i \setminus \widetilde{V}_1) & \xrightarrow{f} & U_r \end{array}$$

Note that $Y_{2,i}$ is an admissible open subset of Y. We obtain the relation $f^*(f_1) = g'^*(Q_2)f^*(f_2)$ in $\mathcal{O}_{Y,y}$ by pushing the relation $f_1 = Q_2 f_2$ from B_i to $\mathcal{O}(Y_{2,i})$ using g'^* and then from $\mathcal{O}(Y_{2,i})$ to $\mathcal{O}_{Y,y}$. Since $f^*(f_2)$ is not a zero divisor in $\mathcal{O}_{Y,y}$, we see that $g'^*(Q_2) = u$, which is a unit of $\mathcal{O}_{Y,y}$. We prove that this is a contradiction. We know that $\operatorname{Sp} B_i \cap \widetilde{V}_1$ is the set of maximal ideals $\mathfrak{m}$ of B_i satisfying the extra condition $|Q_2 \bmod \mathfrak{m}| > 0$. Therefore $\operatorname{Sp} B_i \setminus \widetilde{V}_1$ is the set of maximal ideals $\mathfrak{m}$ of B_i satisfying $|Q_2 \bmod \mathfrak{m}| = 0$. Since $g'(y) \in \operatorname{Sp} B_i \setminus \widetilde{V}_1$, we see that Q_2 vanishes at $g'(y)$. In other words, if we denote the maximal ideal of $\mathcal{O}_{\widetilde{U}_r, g'(y)}$ by $\mathfrak{m}_{g'(y)}$, then we have $Q_2 \bmod \mathfrak{m}_{g'(y)} = 0$. If $\mathfrak{n}_y$ is the maximal ideal of $\mathcal{O}_{Y,y}$, then using the commutative diagram

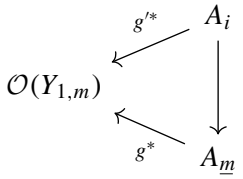
$$\begin{array}{ccc} \mathcal{O}_{\widetilde{U}_r, g'(y)} & \xrightarrow{g'^*} & \mathcal{O}_{Y,y} \\ \downarrow & & \downarrow \\ \mathcal{O}_{\widetilde{U}_r, g'(y)}/\mathfrak{m}_{g'(y)} & \longrightarrow & \mathcal{O}_{Y,y}/\mathfrak{n}_y \end{array}$$

we get $g'^*(Q_2) \bmod \mathfrak{n}_y = 0$. This means that $g'^*(Q_2)$ is contained in the maximal ideal of $\mathcal{O}_{Y,y}$. In other words, $g'^*(Q_2)$ is not a unit. This is a contradiction. Therefore $y \notin Y_1$. □

To show that $g = g'$, we prove that their restrictions to $Y_{1,m}$ and $Y_{2,n}$ are equal for all $m \in I$ and $n \in J$. We prove that g and g' agree on $Y_{1,m}$. Using the definition of g, we see that the restriction of g to $Y_{1,m}$ factors as $Y_{1,m} \to \operatorname{Sp} A_{\underline{m}} \to \widetilde{U}_r$. Since g' maps the affinoid space $Y_{1,m}$ to $\widetilde{U}_r$ and $\{\operatorname{Sp} A_i, \operatorname{Sp} B_j\}_{i,j\geq 0}$ is an admissible cover of $\widetilde{U}_r$, we deduce that g' maps $Y_{1,m}$ into the union of finitely many $\operatorname{Sp} A_i$ and $\operatorname{Sp} B_j$. By Lemma 2.4, we see that $g' : Y_{1,m} \to \widetilde{U}_r$ factors as $Y_{1,m} \to \operatorname{Sp} A_i \to \widetilde{U}_r$ for some $i \gg 0$. We may assume that $i \geq \underline{m}$. To check that g and g' agree on $Y_{1,m}$, it is therefore enough to check that the following diagram commutes:

$$\begin{array}{ccc} & \overset{g'}{\nearrow} & \operatorname{Sp} A_i \\ Y_{1,m} & & \uparrow \\ & \underset{g}{\searrow} & \operatorname{Sp} A_{\underline{m}}. \end{array}$$

In other words, we have to check that the following diagram commutes:



Since both g^* and g'^* are equal to f^* on $\mathcal{O}(U_r)$, we only need to check that $g^*(Q_1) = g'^*(Q_1)$. Pushing the relation $f_2 = Q_1 f_1$ from A_i to $\mathcal{O}(Y_{1,m})$ under g'^*, we get $f^*(f_2) = g'^*(Q_1) f^*(f_1)$. Similarly, pushing the same relation from $A_{\underline{m}}$ to $\mathcal{O}(Y_{1,m})$ under g^*, we get $f^*(f_2) = g^*(Q_1) f^*(f_1)$. Using Lemma 2.3, we get $g^*(Q_1) = g'^*(Q_1)$. This proves the commutativity of the diagram above. In other words, g and g' agree on $Y_{1,m}$. A similar argument shows that g and g' agree on $Y_{2,n}$. Therefore $g = g'$ on Y. This finally proves that $\pi : \widetilde{U}_r \to U_r$ is the blow-up of U_r at the maximal ideal (f_1, f_2).

2.4. ***Points in the exceptional fiber as tangent directions.*** The following standard fact allows us to realize E-valued points of the fiber over the exceptional point $(p^r, (1+p)^{r+1} - 1)$ as tangent directions in U_r at this point. Recall that the maximal ideal m of $\mathcal{O}(U_r)$ corresponding to this exceptional point is the ideal generated by $f_1 = S_1 - p^r$ and $f_2 = S_2 - ((1+p)^{k-1} - 1)$.

Proposition 2.5. *There is a bijection*

$$\pi^{-1}(p^r, (1+p)^{r+1} - 1)(E) \to \mathbb{P}(\mathrm{Hom}(m/m^2 \otimes_{\mathbb{Q}_p} E, E))$$

between the E-valued points of the fiber over the point $(p^r, (1+p)^{k-1} - 1)$ and the elements of the projectivization of the tangent space $\mathrm{Hom}(m/m^2 \otimes_{\mathbb{Q}_p} E, E)$ *over E, which sends a point $(p^r, (1+p)^{k-1} - 1, a:b) \in \widetilde{U}_r(E)$ to the class in $\mathbb{P}(\mathrm{Hom}(m/m^2 \otimes_{\mathbb{Q}_p} E, E))$ represented by v, where*

$$v(\bar{f}_1 \otimes 1) = a \quad \textit{and} \quad v(\bar{f}_2 \otimes 1) = b,$$

with bar denoting image modulo m^2.

Proof. Let $P = (p^r, (1+p)^{k-1} - 1, a : b)$ be an E-valued point in the fiber above the exceptional point. Assume $b \neq 0$. Write $P = (p^r, (1+p)^{k-1} - 1, a/b : 1)$. From the construction of $\widetilde{U}_r$, we see that $P \in \widetilde{V}_2(E)$. Therefore P is an E-valued point of $\mathrm{Sp}\, B_i$ for some $i \geq 0$. Let g_P be the homomorphism $B_i \to E$ associated with P. The discussion at the end of Section 2.1 implies that g_P is defined by $g_P(S_1) = p^r$, $g_P(S_2) = (1+p)^{k-1} - 1$ and $g_P(Q_2) = a/b$. Let m_P be the kernel of g_P. The map $\mathcal{O}(U_r) \to B_i$ induces an E-linear surjection

$$(m/m^2) \otimes_{\mathbb{Q}_p} E \to f_2 B_i / f_2 B_i m_P \otimes_{B_i, g_P} E,$$

given by

$$\bar{f}_1 \otimes 1 \mapsto \bar{f}_2 \otimes (a/b),$$
$$\bar{f}_2 \otimes 1 \mapsto \bar{f}_2 \otimes 1,$$

where the bars over the elements on the left denote their images modulo m^2 and the bars over the elements on the right denote their images modulo $f_2 B_i m_P$, and where we have used $\bar{f}_1 \otimes 1 = \bar{f}_2 \overline{Q}_2 \otimes 1 = \bar{f}_2 \otimes (a/b)$ in the codomain. The codomain of this map can be identified with E up to multiplication by a nonzero scalar. The class of the above map in $\mathbb{P}(\mathrm{Hom}(m/m^2 \otimes_{\mathbb{Q}_p} E, E))$ is the same as that of the map v, where $v(\bar{f}_1 \otimes 1) = a$ and $v(\bar{f}_2 \otimes 1) = b$. A similar construction works if $a \neq 0$. One checks immediately that the resulting assignment $P \mapsto v$ is well-defined and a bijection. □

3. Explicit bases of $H^1(\mathcal{R}_{\mathbb{Q}_p}(x^r\chi))$ and $\mathcal{L}$-invariants

Let $\Gamma = \mathrm{Gal}(\mathbb{Q}_p(\mu_{p^\infty})/\mathbb{Q}_p)$ and think of the cyclotomic character χ as a character $\chi : \Gamma \to \mathbb{Z}_p^*$. In this section, we study two explicit bases of the first Fontaine–Herr cohomology group $H^1(\mathcal{R}_{\mathbb{Q}_p}(x^r\chi))$ and the corresponding formulas for the $\mathcal{L}$-invariant. Here $\mathcal{R}_{\mathbb{Q}_p}(x^r\chi)$ is the Robba ring over $\mathbb{Q}_p$ in the variable T, with the standard actions of φ and Γ twisted by $x^r\chi$.

Colmez [2008, Proposition 2.19] has constructed two power series $G(|x|, r+1)$ and $G'(|x|, r+1)$ in $\mathcal{R}_{\mathbb{Q}_p}$, which he uses to define a basis of $H^1(\mathcal{R}_{\mathbb{Q}_p}(x^r\chi))$. In the first subsection, we explicitly compute $G(|x|, r+1)$. In the second subsection, we give a partial description of $G'(|x|, r+1)$. In the third subsection, we state Colmez's formula for the $\mathcal{L}$-invariant of a nonzero element of $H^1(\mathcal{R}_{\mathbb{Q}_p}(x^r\chi))$ expressed as a linear combination in this basis. In the same subsection, we describe another basis of $H^1(\mathcal{R}_{\mathbb{Q}_p}(x^r\chi))$ studied by Benois [2011, Proposition 1.5.4]. Finally, we find the change of basis matrix between these two bases and restate the formula for the $\mathcal{L}$-invariant in terms of Benois' basis (Definition 3.6).

We first recall that the Fontaine–Herr cohomology groups $H^i(D)$ of a (φ, Γ)-module D over $\mathcal{R}_{\mathbb{Q}_p}$ for $i = 0, 1, 2$ are defined to be the cohomology groups of the complex

$$0 \to D \to D \oplus D \to D \to 0, \tag{6}$$

where the second map is $x \mapsto ((\varphi - 1)x, (\gamma - 1)x)$ and the third map is $(y, z) \mapsto (\gamma - 1)y - (\varphi - 1)z$ for γ a fixed topological generator of Γ. We note that this complex is as in [Chenevier 2013, Section 2.1] and is a bit different from the one in [Colmez 2008, Section 2.1].

We now recall some standard facts about Robba rings. For $M > 0$, let $\mathcal{E}_E^{]0,M]}$ be the set of bidirectional power series in the variable T with coefficients in E that converge on the elements of $\overline{\mathbb{Q}}_p$ with valuation belonging to $(0, M]$. More precisely, it is the set of the series $\sum_{n=-\infty}^{\infty} a_n T^n$ satisfying

$$\lim_{n \to \pm\infty} v_p(a_n) + ns \to \infty \quad \text{for all } 0 < s \leq M.$$

For each $0 < s \leq M$, there is a valuation $v_p(\cdot, s)$ on $\mathcal{E}_E^{]0,M]}$ defined by

$$v_p\left(\sum_{n=-\infty}^{\infty} a_n T^n, s\right) = \inf\{v_p(a_n) + ns \mid n \in \mathbb{Z}\}.$$

Any sequence in $\mathcal{E}_E^{]0,M]}$ that is Cauchy with respect to $v_p(\cdot, s)$ for each $0 < s \leq M$ is convergent in $\mathcal{E}_E^{]0,M]}$ (see [Kedlaya 2006, Definition 2.5.1]). Moreover, choosing a sequence $0 < r_l \leq M$ converging to 0 with $r_1 = M$, we get a countable family of valuations $v_p(\cdot, r_l)$ on $\mathcal{E}_E^{]0,M]}$. Note that if $r_l \leq s \leq M$

for some $l \geq 1$, then $v_p(f(T), s) \geq \inf\{v_p(f(T), r_l), v_p(f(T), M)\}$. Therefore to check if a sequence in $\mathcal{E}_E^{]0,M]}$ is convergent, we only need to check that it converges to a common limit with respect to the valuations $v_p(\cdot, r_l)$ for $l \geq 1$. In particular, $\mathcal{E}_E^{]0,M]}$ is a Fréchet space with respect to the valuations $v_p(\cdot, r_l)$. For the space $\mathcal{E}_E^{]0,1/(p-1)]}$, we set $r_l = 1/\phi(p^l)$, where ϕ is Euler's totient function.

The Robba ring over E is defined to be $\mathcal{R}_E = \bigcup_{M>0} \mathcal{E}_E^{]0,M]}$. It is endowed with commuting actions of an operator φ and the group Γ via $\varphi T = (1+T)^p - 1$ and $\gamma T = (1+T)^{\chi(\gamma)} - 1$ for $\gamma \in \Gamma$, respectively.

We recall some facts from [Lazard 1962]. Let $\phi_n(T)$ for $n \geq 1$ be the p^n-th cyclotomic polynomial defined by

$$\phi_n(T) = \frac{(1+T)^{p^n} - 1}{(1+T)^{p^{n-1}} - 1} = 1 + (1+T)^{p^{n-1}} + \cdots + (1+T)^{(p-1)p^{n-1}}. \tag{7}$$

The polynomials $\phi_n(T)$ are $(1/\phi(p^n))$-extremal in the sense of [loc. cit., Definition 2.7]. This means that $v_p(\phi_n(T), 1/\phi(p^n))$ is attained at the constant and leading terms:

$$v_p(a_0) = v_p\left(\phi_n(T), \frac{1}{\phi(p^n)}\right) = v_p(a_{p^{n-1}(p-1)}) + \frac{p^{n-1}(p-1)}{\phi(p^n)} = 1, \tag{8}$$

where $\phi_n(T) = a_0 + a_1 T + \cdots + a_{p^{n-1}(p-1)} T^{p^{n-1}(p-1)}$, with $a_0 = p$ and $a_{p^{n-1}(p-1)} = 1$. For $n \geq 1$, we also have

$$\varphi(\phi_n(T)) = \phi_{n+1}(T). \tag{9}$$

Let $t = \log(1+T) \in \mathcal{E}^{]0,1/(p-1)]}$. The following formula relates t and the cyclotomic polynomials:

$$t = T \prod_{n \geq 1} \frac{\phi_n(T)}{p}. \tag{10}$$

We also have

$$\varphi(t) = pt \tag{11}$$

and

$$\gamma(t) = \chi(\gamma) t \tag{12}$$

for all $\gamma \in \Gamma$.

Let $r \geq 0$. By [Colmez 2008, Proposition 2.16(i)], there is an isomorphism

$$\mathcal{E}^{]0,\frac{1}{p-1}]}/t^{r+1} \to \prod_{n \geq 1} \mathcal{E}^{]0,\frac{1}{p-1}]}/\phi_n^{r+1}. \tag{13}$$

By [loc. cit., Proposition 2.16(ii)], for $n \geq 1$, there is a Γ-equivariant isomorphism

$$\iota_n : \mathcal{E}^{]0,\frac{1}{p-1}]}/\phi_n^{r+1} \to \mathbb{Q}_p(\zeta_{p^n})[t]/t^{r+1}$$

obtained by sending T to $\zeta_{p^n} e^{t/p^n} - 1$. Thus the Γ-equivariant homomorphism

$$\mathcal{E}^{]0,\frac{1}{p-1}]}/t^{r+1} \to \prod_{n \geq 1} \mathbb{Q}_p(\zeta_{p^n})[t]/t^{r+1} \tag{14}$$

induced by the maps ι_n for $n \geq 1$ is an isomorphism.

For the rest of the paper, we fix a topological generator γ of Γ such that $\chi(\gamma) = \zeta_{p-1}^{a}(1+p)$ for some fixed integer a.

3.1. *An explicit description of $G(|x|, r+1)$.* Let $r \geq 1$. The first basis vector of $H^1(\mathcal{R}_{\mathbb{Q}_p}(x^r\chi))$ constructed by Colmez is represented by the element

$$c_1 = \big(t^{-(r+1)}(p^{-1}\varphi - 1)G(|x|, r+1), t^{-(r+1)}(\gamma - 1)G(|x|, r+1)\big)$$

in $\mathcal{R}_{\mathbb{Q}_p}(x^r\chi) \oplus \mathcal{R}_{\mathbb{Q}_p}(x^r\chi)$ (for the untwisted action of φ and γ); see [Colmez 2008, Proposition 2.19]. Here $G(|x|, r+1)$ is any power series $f(T)$ in $\mathcal{E}^{]0,1/(p-1)]}$ mapping to $\prod_{n\geq 1} 1/p^n$ under the map (13) (see [loc. cit., Section 2.6]). In other words, $G(|x|, r+1)$ satisfies the system of congruences

$$\begin{aligned} f(T) &\equiv \frac{1}{p} \bmod \phi_1(T)^{r+1}, \\ f(T) &\equiv \frac{1}{p^2} \bmod \phi_2(T)^{r+1}, \\ &\vdots \\ f(T) &\equiv \frac{1}{p^n} \bmod \phi_n(T)^{r+1}, \\ &\vdots \end{aligned}$$

The goal of this section is to write down $G(|x|, r+1)$ explicitly.

Definition 3.1. For each $n \geq 1$ and $r \geq 1$, define

$$G_{n,r+1}(T) := \left[1 - \left(\frac{\phi_n(T)}{p}\right)^{r+1}\right]^{r+1} \frac{1}{p^n} \prod_{i>n} \left[1 - \left(1 - \frac{\phi_i(T)}{p}\right)^{r+1}\right]^{r+1}.$$

To check $G_{n,r+1}(T)$ is well-defined, we need to show that the infinite product converges. We prove that for any $n, r \geq 1$ the product $\prod_{i>n}(1 - (1 - (\phi_i(T)/p))^{r+1})$ converges.

Lemma 3.2. *For any $c = 1, 2, \ldots, p-1$ and any $j = 1, 2, \ldots, cp^{i-1}$, we have*

$$v_p\left(\binom{cp^{i-1}}{j}\right) = i - 1 - v_p(j).$$

Proof. We leave this as an exercise for the reader. □

Lemma 3.3. *For any $i \geq 1$, we have*

$$v_p\left(1 - \frac{\phi_i(T)}{p}, \frac{1}{\phi(p^i)}\right) = 0$$

and, for $l \geq 1$, we have

$$v_p\left(1 - \frac{\phi_i(T)}{p}, \frac{1}{\phi(p^l)}\right) > i - l - 1.$$

Proof. Since $\phi_i(T)$ is $(1/\phi(p^i))$-extremal, we see that by (8), $v_p(\phi_i(T), 1/\phi(p^i)) = 1$. Since all the terms except the constant terms of the polynomials $\phi_i(T)$ and $\phi_i(T) - p$ are the same, we get $v_p(\phi_i(T) - p, 1/\phi(p^i)) = 1$. Therefore $v_p(1 - \phi_i(T)/p, 1/\phi(p^i)) = 0$.

Write $\phi_i(T) - p = a_1 T + a_2 T^2 + \cdots + a_{\phi(p^i)} T^{\phi(p^i)}$. By expanding the powers of $(1+T)$ in $\phi_i(T)$, we see that

$$a_j = \binom{p^{i-1}}{j} + \binom{2p^{i-1}}{j} + \cdots + \binom{(p-1)p^{i-1}}{j}.$$

By Lemma 3.2,

$$v_p(a_j) + \frac{j}{\phi(p^l)} \geq i - 1 - v_p(j) + \frac{p^{v_p(j)}}{\phi(p^l)} = i - 1 - v_p(j) + \frac{p^{v_p(j)-(l-1)}}{p-1} > i - 1 - v_p(j) + v_p(j) - (l-1) = i - l.$$

Therefore

$$v_p\left(1 - \frac{\phi_i(T)}{p}, \frac{1}{\phi(p^l)}\right) = v_p\left(\phi_i(T) - p, \frac{1}{\phi(p^l)}\right) - 1 > i - l - 1. \qquad \square$$

We now prove that the product $\prod_{i>n}(1-(1-(\phi_i(T)/p))^{r+1})$ converges in $\mathcal{E}^{]0,1/(p-1)]}$ for $n, r \geq 1$. Consider the polynomials $1 - (1 - (\phi_i(T)/p))^{r+1}$ as elements of $\mathcal{E}^{]0,\infty]}$. By [Lazard 1962, Proposition 4.11], the product $\prod_{i>n}(1 - (1 - \phi_i(T)/p)^{r+1})$ converges in $\mathcal{E}^{]0,\infty]}$ if and only if

$$v_p\left(\left(1 - \frac{\phi_i(T)}{p}\right)^{r+1}, s\right) \to \infty \quad \text{as } i \to \infty \tag{15}$$

for each $s > 0$. Fix $s > 0$ and pick $l_s \geq 1$ such that $1/\phi(p^{l_s}) < s$. As $(1 - \phi_i(T)/p)^{r+1}$ contains no negative power of T, we have

$$v_p\left(\left(1 - \frac{\phi_i(T)}{p}\right)^{r+1}, s\right) \geq (r+1) v_p\left(1 - \frac{\phi_i(T)}{p}, \frac{1}{\phi(p^{l_s})}\right) > (r+1)(i - l_s - 1),$$

by Lemma 3.3, whence (15) holds. Therefore the product $\prod_{i>n}(1 - (1 - (\phi_i(T)/p))^{r+1})$ converges in $\mathcal{E}^{]0,\infty]} \subseteq \mathcal{E}^{]0,1/(p-1)]}$.

Our candidate for $G(|x|, r+1)$ is $\sum_{n=1}^{\infty} G_{n,r+1}(T)$. To see that the infinite sum is well-defined, we prove the following theorem.

Theorem 3.4. *For any positive integer l, the sequence $G_{n,r+1}(T)$ converges to 0 with respect to the valuation $v_p(\cdot, 1/\phi(p^l))$.*

Proof. Fix $l \geq 1$. We need to show

$$v_p\left(G_{n,r+1}(T), \frac{1}{\phi(p^l)}\right) \to \infty \quad \text{as } n \to \infty.$$

Consider

$$v_p\left(G_{n,r+1}(T), \frac{1}{\phi(p^l)}\right)$$
$$= (r+1) v_p\left(1 - \left(\frac{\phi_n(T)}{p}\right)^{r+1}, \frac{1}{\phi(p^l)}\right) - n + (r+1) \sum_{i>n} v_p\left(1 - \left(1 - \frac{\phi_i(T)}{p}\right)^{r+1}, \frac{1}{\phi(p^l)}\right).$$

Note that

$$1-\left(\frac{\phi_n(T)}{p}\right)^{r+1}=\left(1-\frac{\phi_n(T)}{p}\right)\left(1+\frac{\phi_n(T)}{p}+\cdots+\left(\frac{\phi_n(T)}{p}\right)^{r}\right).$$

Now if $n \geq l$, then

$$v_p\left(\phi_n(T), \frac{1}{\phi(p^l)}\right)=1.$$

Indeed, since $\phi_n(T)=a_0+a_1T+\cdots+a_{\phi(p^n)}T^{\phi(p^n)}$ is $(1/\phi(p^n))$-extremal, we have

$$v_p(a_i)+\frac{i}{\phi(p^l)} \geq v_p(a_i)+\frac{i}{\phi(p^n)} \geq v_p(a_0)=1 \quad \text{for any } i.$$

Hence

$$v_p\left(\phi_n(T), \frac{1}{\phi(p^l)}\right)=v_p(a_0)=1.$$

Thus

$$v_p\left((\phi_n(T)/p)^j, \frac{1}{\phi(p^l)}\right)=0 \quad \text{for any } j \geq 0.$$

Therefore

$$v_p\left(1-\left(\frac{\phi_n(T)}{p}\right)^{r+1}, \frac{1}{\phi(p^l)}\right) \geq v_p\left(1-\frac{\phi_n(T)}{p}, \frac{1}{\phi(p^l)}\right) > n-l-1,$$

where the last inequality is a consequence of Lemma 3.3. On the other hand, writing

$$1-\left(1-\left(\frac{\phi_i(T)}{p}\right)\right)^{r+1}=a_0'+a_1'T+\cdots+a'_{(r+1)\phi(p^i)}T^{(r+1)\phi(p^i)},$$

we observe that $a_0'=1$ and $a_1'T+\cdots+a'_{(r+1)\phi(p^i)}T^{(r+1)\phi(p^i)}=-(1-(\phi_i(T)/p))^{r+1}$. By Lemma 3.3,

$$v_p\left(-\left(1-\frac{\phi_i(T)}{p}\right)^{r+1}, \frac{1}{\phi(p^l)}\right) > 0 \quad \text{if } i > l.$$

We also know that $v_p(a_0')=0$. Therefore

$$v_p\left(1-\left(1-\frac{\phi_i(T)}{p}\right)^{r+1}, \frac{1}{\phi(p^l)}\right)=0.$$

Using these estimates together, we see that, for $n \geq l$,

$$v_p\left(G_{n,r+1}(T), \frac{1}{\phi(p^l)}\right) > (r+1)(n-l-1)-n=nr-(r+1)l-(r+1).$$

Letting $n \to \infty$, we see that $v_p(G_{n,r+1}(T), 1/\phi(p^l)) \to \infty$, as desired. Therefore the sequence $G_{n,r+1}(T)$ converges to 0 with respect to the valuation $v_p(\cdot\,, 1/\phi(p^l))$ for each $l > 0$. □

Using the theorem above, we see that the series $G(|x|, r+1)=\sum_{n=1}^{\infty} G_{n,r+1}(T)$ converges. It is easily checked, using $\phi_i(T) \equiv p \bmod \phi_n(T)$ for $i > n$, that it is a solution to the system of congruences that we started with:

$$G(|x|, r+1) \equiv \frac{1}{p^n} \bmod \phi_n(T)^{r+1} \quad \text{for all } n \in \mathbb{N}.$$

Therefore, we have written down $G(|x|, r+1)$ explicitly.

3.2. ***A partial description of $G'(|x|, r+1)$.*** Let $r \geq 1$. The second basis vector of $H^1(\mathcal{R}_{\mathbb{Q}_p}(x^r\chi))$ constructed by Colmez is represented by the element

$$c_2 = \big(t^{-(r+1)}(p^{-1}\varphi - 1)(\log T - G'(|x|, r+1)),\, t^{-(r+1)}(\gamma - 1)(\log T - G'(|x|, r+1))\big)$$

in $\mathcal{R}_{\mathbb{Q}_p}(x^r\chi) \oplus \mathcal{R}_{\mathbb{Q}_p}(x^r\chi)$ (for the untwisted action of φ and γ); see [Colmez 2008, Proposition 2.19]. Here $G'(|x|, r+1)$ is an element of $\mathcal{E}^{]0,1/(p-1)]}$ mapping to $\prod_{n\geq 1} \log(\zeta_{p^n} e^{t/p^n} - 1)$ under the map (14) (see [loc. cit., Section 2.7]).

We wish to compute $G'(|x|, r+1)$ explicitly, just as we did for $G(|x|, r+1)$. We are only able to determine $G'(|x|, r+1)$ modulo t but this is sufficient for our purposes.

Consider the element

$$g'(T) = \frac{\log \gamma}{\gamma - 1} \log T - \log T - \frac{\log \chi(\gamma) t}{\chi(\gamma) - 1} \in \mathcal{E}^{]0, \frac{1}{p-1}]}. \tag{16}$$

In the proof of his Theorem 1.5.7, Benois [2011] relates $g'(T)$ (which he calls y) to the element $d_1 = -t^{-1}\nabla_0(\log T) + (\chi(\gamma) - 1)^{-1}$ of $\mathcal{R}_{\mathbb{Q}_p}[\log T, 1/t]$ using the equation

$$d_1 = -\log(\chi(\gamma))^{-1} t^{-1}(\log T + g'(T)), \tag{17}$$

where

$$\nabla_0 = \frac{1}{\log \chi(\gamma)} \frac{\log \gamma}{\gamma - 1}.$$

We claim that $\iota_n(-g'(T)) \equiv \log(\zeta_{p^n} e^{t/p^n} - 1) \bmod t$. Indeed,

$$\begin{aligned}
\iota_n(g'(T)) &= \iota_n\left(\frac{\log \gamma}{\gamma - 1} \log T - \log T - \frac{\log \chi(\gamma) t}{\chi(\gamma) - 1}\right) \\
&\equiv \frac{\log \gamma}{\gamma - 1} \log(\zeta_{p^n} - 1) - \log(\zeta_{p^n} - 1) \bmod t \\
&\equiv \frac{1 + \gamma + \cdots + \gamma^{p^{n-1}(p-1)-1}}{p^{n-1}(p-1)} \frac{\log \gamma^{p^{n-1}(p-1)}}{\gamma^{p^{n-1}(p-1)} - 1} \log(\zeta_{p^n} - 1) - \log(\zeta_{p^n} - 1) \bmod t.
\end{aligned}$$

Since $\gamma^{p^{n-1}(p-1)} \log(\zeta_{p^n} - 1) = \log(\zeta_{p^n} - 1)$, we get

$$\begin{aligned}
\iota_n(g'(T)) &\equiv \frac{1 + \gamma + \cdots + \gamma^{p^{n-1}(p-1)-1}}{p^{n-1}(p-1)} \log(\zeta_{p^n} - 1) - \log(\zeta_{p^n} - 1) \bmod t \\
&\equiv \frac{1}{p^{n-1}(p-1)} \log(N_{\mathbb{Q}_p(\zeta_{p^n})/\mathbb{Q}_p}(\zeta_{p^n} - 1)) - \log(\zeta_{p^n} - 1) \bmod t \\
&\equiv -\log(\zeta_{p^n} - 1) \bmod t \\
&\equiv -\log(\zeta_{p^n} e^{t/p^n} - 1) \bmod t.
\end{aligned}$$

Since the images of $-g'(T)$ and $G'(|x|, r+1)$ under the isomorphism (14) are equal (with r there equal to zero!), there exists $g''(T) \in \mathcal{E}^{]0,1/(p-1)]}$ such that

$$G'(|x|, r+1) = -g'(T) + t g''(T). \tag{18}$$

This determines $G'(|x|, r+1)$ explicitly modulo t.

3.3. ***Another basis of $H^1(\mathcal{R}_{\mathbb{Q}_p}(x^r\chi))$.*** Let $r \geq 1$. In this section, we consider a different basis $\{\bar{\alpha}_{r+1}, \bar{\beta}_{r+1}\}$ of $H^1(\mathcal{R}_{\mathbb{Q}_p}(x^r\chi))$ described in [Benois 2011, Proposition 1.5.4] which is more suitable for computations. Here bar denotes class in $H^1(\mathcal{R}_{\mathbb{Q}_p}(x^r\chi))$. By [loc. cit., Corollary 1.5.5], for any class in $H^1(\mathcal{R}_{\mathbb{Q}_p}(x^r\chi))$ represented by $(a, b) \in \mathcal{R}_{\mathbb{Q}_p}(x^r\chi) \oplus \mathcal{R}_{\mathbb{Q}_p}(x^r\chi)$, we have

$$\overline{(a, b)} = \lambda' \cdot \bar{\alpha}_{r+1} + \mu' \cdot \bar{\beta}_{r+1}, \quad \text{where } \lambda' = \operatorname{res}(at^r\,dt) \text{ and } \mu' = \operatorname{res}(bt^r\,dt). \tag{19}$$

We want to compute the change of basis relation between Colmez's basis and Benois' basis. This allows us to state the formula for the $\mathcal{L}$-invariant of a nonzero cohomology class in terms of certain residues.

Proposition 3.5. *Let*

$$c_1 = \big(t^{-(r+1)}(p^{-1}\varphi - 1)G(|x|, r+1), t^{-(r+1)}(\gamma - 1)G(|x|, r+1)\big),$$
$$c_2 = \big(t^{-(r+1)}(p^{-1}\varphi - 1)(\log T - G'(|x|, r+1)), t^{-(r+1)}(\gamma - 1)(\log T - G'(|x|, r+1))\big)$$

be elements of $\mathcal{R}_{\mathbb{Q}_p}(x^r\chi) \oplus \mathcal{R}_{\mathbb{Q}_p}(x^r\chi)$ representing Colmez's basis of $H^1(\mathcal{R}_{\mathbb{Q}_p}(x^r\chi))$. Then

$$\bar{c}_1 = -\frac{p-1}{p} \cdot \bar{\alpha}_{r+1}, \quad \bar{c}_2 = \log(\chi(\gamma)) \cdot \bar{\beta}_{r+1}.$$

Proof. We prove the proposition using (19). We therefore need to compute the four residues

$$\operatorname{res}\big(t^{-1}(p^{-1}\varphi - 1)G(|x|, r+1)\,dt\big), \qquad \operatorname{res}\big(t^{-1}(\gamma - 1)G(|x|, r+1)\,dt\big),$$
$$\operatorname{res}\big(t^{-1}(p^{-1}\varphi - 1)(\log T - G'(|x|, r+1))\,dt\big), \quad \operatorname{res}\big(t^{-1}(\gamma - 1)(\log T - G'(|x|, r+1))\,dt\big).$$

We shall use the following formulas: for any $f(T) \in \mathcal{R}_{\mathbb{Q}_p}$, we have

$$\operatorname{res}(\varphi(f(T))\,dt) = \operatorname{res}(f(T)\,dt), \tag{20}$$
$$\operatorname{res}(\gamma(f(T))\,dt) = \chi(\gamma)^{-1}\operatorname{res}(f(T)\,dt). \tag{21}$$

- Recall that $G(|x|, r+1) = \sum_{n=1}^{\infty} G_{n,r+1}(T)$, where

$$G_{n,r+1}(T) = \left[1 - \left(\frac{\phi_n(T)}{p}\right)^{r+1}\right]^{r+1} \frac{1}{p^n} \prod_{i>n}\left[1 - \left(1 - \frac{\phi_i(T)}{p}\right)^{r+1}\right]^{r+1}.$$

Using equation (9), we see that $(p^{-1}\varphi)G_{n,r+1}(T) = G_{n+1,r+1}(T)$. Therefore

$$(p^{-1}\varphi - 1)G(|x|, r+1) = -G_{1,r+1}(T).$$

Hence $\operatorname{res}\big(t^{-1}(p^{-1}\varphi - 1)G(|x|, r+1)\,dt\big) = -\operatorname{res}(t^{-1}G_{1,r+1}(T)\,dt)$. Now

$$\operatorname{res}(t^{-1}G_{1,r+1}(T)\,dt) = \operatorname{res}\left(t^{-1}\left[1 - \left(\frac{\phi_1(T)}{p}\right)^{r+1}\right]^{r+1} \frac{1}{p}\prod_{i>1}\left[1 - \left(1 - \frac{\phi_i(T)}{p}\right)^{r+1}\right]^{r+1} dt\right)$$
$$\overset{(10)}{=} \operatorname{res}\left(T^{-1}\left(\frac{\phi_1(T)}{p}\right)^{-1}\left[1 - \left(\frac{\phi_1(T)}{p}\right)^{r+1}\right]^{r+1} \frac{1}{p} f(T)\,dt\right),$$

where

$$f(T)=\prod_{i>1}\left(\frac{\phi_i(T)}{p}\right)^{-1}\left[1-\left(1-\frac{\phi_i(T)}{p}\right)^{r+1}\right]^{r+1}.$$

An argument similar to those above shows that $f(T)\in\mathcal{E}^{]0,1/(p-1)]}$. Moreover, since the value of $\phi(T)/p$ at $T=0$ is 1, we see that $f(T)\in 1+T\mathbb{Q}_p[[T]]$. Using (9), we see that there exists $g(T)\in 1+T\mathbb{Q}_p[[T]]\cap\mathcal{E}^{]0,1/(p-1)]}$ such that $f(T)=\varphi(g(T))$. We therefore have

$$\operatorname{res}(t^{-1}G_{1,r+1}(T)\,dt)=\operatorname{res}\left(T^{-1}\left(\frac{\phi_1(T)}{p}\right)^{-1}\left[1-\left(\frac{\phi_1(T)}{p}\right)^{r+1}\right]^{r+1}\frac{1}{p}\varphi(g(T))\,dt\right).$$

Adding and subtracting 1 from the term $[1-(\phi_1(T)/p)^{r+1}]^{r+1}$, we get

$$\begin{aligned}&\operatorname{res}(t^{-1}G_{1,r+1}(T)\,dt)\\&\overset{(7)}{=}\operatorname{res}\left(\frac{1}{(1+T)^p-1}\varphi(g(T))\,dt\right)+\operatorname{res}\left(T^{-1}\left(\frac{\phi_1(T)}{p}\right)^{-1}\left(\left[1-\left(\frac{\phi_1(T)}{p}\right)^{r+1}\right]^{r+1}-1\right)\frac{1}{p}\varphi(g(T))\,dt\right).\end{aligned}$$

The first term is equal to $\operatorname{res}(\varphi(g(T)/T)\,dt)$, which by (20), equals $\operatorname{res}(g(T)/T\,dt)=1$, since the constant term of $g(T)$ is 1. For the second term, we see that $\phi_1(T)/p$ divides $[1-(\phi_1(T)/p)^{r+1}]^{r+1}-1$ as polynomials. Also, the constant term of

$$\left(\frac{\phi_1(T)}{p}\right)^{-1}\left(\left[1-\left(\frac{\phi_1(T)}{p}\right)^{r+1}\right]^{r+1}-1\right)\varphi(g(T))$$

is -1. Hence, the second residue is $-1/p$. Therefore,

$$\operatorname{res}\big(t^{-1}(p^{-1}\varphi-1)G(|x|,r+1)\,dt\big)=\frac{1}{p}-1.$$

- We prove $\operatorname{res}\big(t^{-1}(\gamma-1)G(|x|,r+1)\,dt\big)=0$, by showing that, for all $n\geq 1$,

$$\operatorname{res}(t^{-1}(\gamma-1)G_{n,r+1}(T)\,dt)=0.$$

Again by (9), we see that $(p^{-1}\varphi)G_{n,r+1}(T)=G_{n+1,r+1}(T)$. Therefore

$$\begin{aligned}\operatorname{res}(t^{-1}(\gamma-1)G_{n,r+1}(T)\,dt)&=\operatorname{res}(t^{-1}(\gamma-1)(p^{-1}\varphi)^{n-1}G_{1,r+1}(T)\,dt)\\&\overset{(11)}{=}\operatorname{res}(\varphi^{n-1}t^{-1}(\gamma-1)G_{1,r+1}(T)\,dt)\\&\overset{(12)}{=}\operatorname{res}(\varphi^{n-1}(\chi(\gamma)\gamma-1)(t^{-1}G_{1,r+1}(T))\,dt),\end{aligned}$$

which vanishes by (20) and (21).

- Using (18), we get $\log T-G'(|x|,r+1)=\log T+g'(T)-tg''(T)$, whereas rearranging the terms in (17), we get $\log T+g'(T)=-\log(\chi(\gamma))td_1$. These two equations imply

$$\log T-G'(|x|,r+1)=-\log(\chi(\gamma))td_1-tg''(T).$$

Therefore we have

$$t^{-1}(p^{-1}\varphi-1)(\log T-G'(|x|,r+1))\overset{(11)}{=}-\log(\chi(\gamma))(\varphi-1)d_1-(\varphi-1)g''(T).$$

By the discussion on [Benois 2011, p. 1604, p. 1601], $\operatorname{res}(((\varphi-1)d_1)\,dt)=0$. By (20), we get

$$\operatorname{res}(t^{-1}(p^{-1}\varphi-1)(\log T - G'(|x|, r+1))\,dt)=0.$$

- Similarly, we have

$$t^{-1}(\gamma-1)(\log T - G'(|x|, r+1)) \overset{(12)}{=} -\log(\chi(\gamma))(\chi(\gamma)\gamma-1)d_1 - (\chi(\gamma)\gamma-1)g''(T).$$

Using $(\chi(\gamma)\gamma-1)d_1=-1/T$ (see the proof of [loc. cit., Theorem 1.5.7(iiib)]) and using (21), we get

$$\operatorname{res}(t^{-1}(\gamma-1)(\log T - G'(|x|, r+1))\,dt)=\log\chi(\gamma).$$

Using (19), we see that $\bar{c}_1$ is $-((p-1)/p)\cdot\bar{\alpha}_{r+1}$ and $\bar{c}_2$ is $\log(\chi(\gamma))\cdot\bar{\beta}_{r+1}$. □

Let us state the formula for the $\mathcal{L}$-invariant using [Colmez 2008, Definition 2.20]. Let (a, b) represent a nonzero element of $H^1(\mathcal{R}_{\mathbb{Q}_p}(x^r\chi))$. Writing $\overline{(a,b)}=\lambda\cdot\bar{c}_1+\mu\cdot\bar{c}_2$, where the bar over an element denotes its class in $H^1(\mathcal{R}_{\mathbb{Q}_p}(x^r\chi))$, Colmez defines the $\mathcal{L}$-invariant associated to the class of (a, b) to be $((p-1)/p)(\lambda/\mu)\in\mathbb{P}^1(\mathbb{Q}_p)$.[4] Using Proposition 3.5, we can restate the formula for the $\mathcal{L}$-invariant in terms of Benois' basis.

Definition 3.6. If $\overline{(a,b)}=\lambda'\cdot\bar{\alpha}_{r+1}+\mu'\cdot\bar{\beta}_{r+1}\neq 0$, where $\overline{(a,b)}$ is the class of (a, b) in $H^1(\mathcal{R}_{\mathbb{Q}_p}(x^r\chi))$, then the $\mathcal{L}$-invariant associated to this class is

$$\mathcal{L}=-\log(\chi(\gamma))\cdot\frac{\lambda'}{\mu'}.$$

4. Cohomology of $m\mathcal{R}_{\mathcal{O}(U_r)}(\delta_{U_r})$

The aim of this section is to find a generator of $H^1(m\mathcal{R}_{\mathcal{O}(U_r)}(\delta_{U_r}))$ for the tautological character $\delta_{U_r}:\mathbb{Q}_p^*\to\mathcal{O}(U_r)^*$, which will be defined in the first subsection. This generator is an essential ingredient in the proof of Theorem 5.2.

4.1. *The tautological character $\delta_{U_r}:\mathbb{Q}_p^*\to\mathcal{O}(U_r)^*$.*

Define δ_{U_r} by the equations

$$\begin{aligned}\delta_{U_r}(p)&=S_1,\\ \delta_{U_r}(\zeta_{p-1})&=\zeta_{p-1}^{r+1},\\ \delta_{U_r}(1+p)&=1+S_2.\end{aligned}\tag{22}$$

This character is tautological in the following sense. If δ is an E-valued character of $\mathbb{Q}_p^*$ and if $(\delta(p), \delta(\zeta_{p-1}), \delta(1+p)-1)\in U_r(E)$, then we have the following commutative diagram:

[4]The sign is the opposite of the one in [Colmez 2008, Definition 2.20]. We believe that this choice of sign is correct. Indeed, consider the image of $p(1+p)$ under the Kummer map $\kappa:\mathbb{Q}_p^*\to H^1(\mathbb{Q}_p,\mathbb{Q}_p(1))$. By the discussion in [Benois 2011, Section 1.5.6], we see that with the original choice of sign, the $\mathcal{L}$-invariant associated to the image of $\kappa(p(1+p))$ under the canonical isomorphism $H^1(\mathbb{Q}_p,\mathbb{Q}_p(1))\simeq H^1(\mathcal{R}_{\mathbb{Q}_p}(x|x|))$ is $-\log(1+p)$. However, using Tate's formula, the $\mathcal{L}$-invariant associated to $\kappa(p(1+p))$ is equal to $\log(p(1+p))/(v_p(p(1+p)))=\log(1+p)$. The original sign is also incompatible with our results in the Introduction.

$$\begin{array}{ccc} & & \mathcal{O}(U_r)^* \\ & \overset{\delta_{U_r}}{\nearrow} & \downarrow \\ \mathbb{Q}_p^* & \xrightarrow{\delta} & E^* \end{array}$$

where the vertical map is induced by the affinoid algebra map $\mathcal{O}(U_r) \to E$ associated to the E-valued point $(\delta(p), \delta(\zeta_{p-1}), \delta(1+p)-1)$ of U_r.

We prove that δ_{U_r} is bien plaçé in the sense of [Chenevier 2013, Definition 2.31].

Lemma 4.1. *The character δ_{U_r} is bien plaçé.*

Proof. • Given any $i \in \mathbb{Z}$, we have to check that $1 - \delta_{U_r}(p)p^i = 1 - S_1 p^i$ is not a zero divisor in $\mathcal{O}(U_r)$. But this follows immediately from the fact that $\mathbb{Q}_p\langle S_1, S_2\rangle$ is a domain and $\mathbb{Q}_p\langle S_1, S_2\rangle \to \mathcal{O}(U_r)$ is flat [Bosch, Güntzer, and Remmert 1984, Corollary 7.3.2/6].

• Given any $i \geq 0$, we have to check that the image of

$$1 - \delta_{U_r}(\chi(\gamma))\chi(\gamma)^{1-i} = 1 - \zeta_{p-1}^{a(r+2-i)}(1+S_2)(1+p)^{1-i}$$

in the quotient $\mathcal{O}(U_r)/(1-\delta_{U_r}(p)p^{-i})$ is not a zero divisor. This again follows from the facts that

$$\mathbb{Q}_p\langle S_1, S_2\rangle/(1-\delta_{U_r}(p)p^{-i}) \to \mathcal{O}(U_r)/(1-\delta_{U_r}(p)p^{-i})$$

is flat [loc. cit., Corollary 7.3.2/6], that the above element is nonzero on the left (easy to check) and that the ring on the left is a domain (indeed, if $f(S_1, S_2)g(S_1, S_2) = (1 - S_1 p^{-i})h(S_1, S_2)$, then applying the Weierstrass division theorem [loc. cit., Corollary 5.2.1/2], we have $f = q(1 - S_1 p^{-i}) + r$ and $g = q'(1 - S_1 p^{-i}) + r'$ for some $q, q' \in \mathbb{Q}_p\langle S_1, S_2\rangle$ and $r, r' \in \mathbb{Q}_p\langle S_2\rangle$, so that $(1 - S_1 p^{-i}) \mid rr'$ and hence one of r or r' must be zero). □

4.2. *Computation of $H^1(m\mathcal{R}_{\mathcal{O}(U_r)}(\delta_{U_r}))$.* In this section, we make explicit some facts in [Chenevier 2013] about the first Fontaine–Herr cohomology groups of certain "big" (φ, Γ)-modules.

Let $\mathcal{R}_{\mathcal{O}(U_r)}(\delta_{U_r})$ be the Robba ring over $\mathcal{O}(U_r)$ in the variable T, with the standard actions of φ and Γ twisted by δ_{U_r}: $\varphi T = \delta_{U_r}(p)((1+T)^p - 1)$ and $\gamma T = \delta_{U_r}(\chi(\gamma))((1+T)^{\chi(\gamma)} - 1)$. Let ψ be the usual left inverse of φ, twisted by $\delta_{U_r}(p^{-1})$. Let $\mathcal{R}^+_{\mathcal{O}(U_r)}(\delta_{U_r})$ be the power series in $\mathcal{R}_{\mathcal{O}(U_r)}(\delta_{U_r})$ consisting of nonnegative powers of T.

We work with two different versions of the Fontaine–Herr cohomology groups for (φ, Γ)-modules D over the Robba ring $\mathcal{R}_{\mathcal{O}(U_r)}$, namely, the (φ, Γ)-version $H^1_{(\varphi,\Gamma)}(D)$ and the (ψ, Γ)-version $H^1_{(\psi,\Gamma)}(D)$. The former groups are defined as usual as the cohomology groups of the complex (6). For the definition of the latter groups, one replaces φ by ψ in the maps in the complex (6). There is a map η from the (φ, Γ)-complex to the (ψ, Γ)-complex

$$\begin{array}{ccccccccc} 0 & \longrightarrow & D & \longrightarrow & D \oplus D & \longrightarrow & D & \longrightarrow & 0 \\ & & \downarrow{\scriptstyle \eta_0} & & \downarrow{\scriptstyle \eta_1} & & \downarrow{\scriptstyle \eta_2} & & \\ 0 & \longrightarrow & D & \longrightarrow & D \oplus D & \longrightarrow & D & \longrightarrow & 0 \end{array}$$

where

$$\eta_0(x) = x, \quad \eta_1(x, y) = (-\psi(x), y) \quad \text{and} \quad \eta_2(x) = -\psi(x).$$

If $\gamma - 1$ is bijective on $D^{\psi=0}$, then η induces an isomorphism on cohomology: $H^1_{(\varphi,\Gamma)}(D) \simeq H^1_{(\psi,\Gamma)}(D)$ (see [Chenevier 2013, Section 2]).

By [loc. cit., Theorem 2.33], the cohomology group $H^1_{(\varphi,\Gamma)}(\mathcal{R}_{\mathcal{O}(U_r)}(\delta_{U_r}))$ is a free module of rank 1 over $\mathcal{O}(U_r)$. First, we compute an $\mathcal{O}(U_r)$-generator of $H^1_{(\varphi,\Gamma)}(\mathcal{R}_{\mathcal{O}(U_r)}(\delta_{U_r}))$. By [loc. cit., Theorem 2.33], the inclusion $m\mathcal{R}_{\mathcal{O}(U_r)}(\delta_{U_r}) \to \mathcal{R}_{\mathcal{O}(U_r)}(\delta_{U_r})$ induces an isomorphism

$$H^1_{(\varphi,\Gamma)}(m\mathcal{R}_{\mathcal{O}(U_r)}(\delta_{U_r})) \to H^1_{(\varphi,\Gamma)}(\mathcal{R}_{\mathcal{O}(U_r)}(\delta_{U_r})).$$

Second, we use this isomorphism to lift the $\mathcal{O}(U_r)$-generator of $H^1_{(\varphi,\Gamma)}(\mathcal{R}_{\mathcal{O}(U_r)}(\delta_{U_r}))$ to an $\mathcal{O}(U_r)$-generator of $H^1_{(\varphi,\Gamma)}(m\mathcal{R}_{\mathcal{O}(U_r)}(\delta_{U_r}))$ (see Definition 4.5).

Let $D = \mathcal{R}_{\mathcal{O}(U_r)}(\delta_{U_r})$. In Section 4.1, we showed that δ_{U_r} is bien plaçé. By [loc. cit., Proposition 2.32] we have the isomorphisms

$$C(D^+)/(\gamma - 1) \longrightarrow C(D)/(\gamma - 1) \longleftarrow D^{\psi=1}/(\gamma - 1) \longrightarrow H^1_{(\psi,\Gamma)}(D),$$

where $C(D) = (1 - \varphi)D^{\psi=1}$, $D^+ = \mathcal{R}^+_{\mathcal{O}(U_r)}(\delta_{U_r})$ and $C(D^+) = (1 - \varphi)(D^+)^{\psi=1}$, and where we recall that the actions of φ and ψ are the usual ones twisted by δ_{U_r}. The first map is induced by the canonical injection of D^+ into D. The second map is induced by $f(T) \mapsto (1 - \varphi) f(T)$. The third map is induced by $f(T) \mapsto (0, f(T))$. We emphasize that the cohomology group appearing in the display above is in the sense of (ψ, Γ)-modules.

We wish to find an $\mathcal{O}(U_r)$-generator of $H^1_{(\psi,\Gamma)}(D)$. We do so by finding an $\mathcal{O}(U_r)$-generator of $C(D^+)/(\gamma - 1)$ and then using the isomorphisms above to get an $\mathcal{O}(U_r)$-generator of $H^1_{(\psi,\Gamma)}(D)$. To find an $\mathcal{O}(U_r)$-generator of $C(D^+)/(\gamma - 1)$, we prove a couple of lemmas.

Lemma 4.2. $C(D^+) = (D^+)^{\psi=0}$.

Proof. By definition, $C(D^+)$ is the image of the map $1 - \varphi : (D^+)^{\psi=1} \to (D^+)^{\psi=0}$. Applying [Chenevier 2013, Lemma 2.9(vi)] with $\lambda = S_1$ and $N = 0$ ($r \geq 1 \Rightarrow |S_1| < 1$), we see that this map is a bijection. The lemma follows. □

Lemma 4.3. $S_1(1 + T)$ *generates* $(D^+)^{\psi=0}/(\gamma - 1)$ *as a free* $\mathcal{O}(U_r)$*-module of rank* 1.

Proof. In order to prove this, we apply [loc. cit., Proposition 2.14]. We know that D^+ is a free $\mathcal{R}^+_{\mathcal{O}(U_r)}$-module of rank 1, say with basis w. To check that D^+ is Γ-bounded in the sense of [loc. cit., Section 1.8], we choose the following model of $\mathcal{O}(U_r)$:

$$\mathcal{A} = \mathbb{Z}_p\langle S_1, S_2, T_1, T_2, T_3\rangle/(p^r T_1 - S_1, 1 - T_1T_2, pT_3 - S_2).$$

Then for $\gamma' \in \Gamma$ we have $\mathrm{Mat}(\gamma') = [\delta_{U_r}(\chi(\gamma'))] \in M_1(\mathcal{A}[[T]])$, so D^+ is Γ-bounded. Using [loc. cit., Proposition 2.14], we see that $\{S_1(1 + T)w\}$ is a basis of $(D^+)^{\psi=0}$ over $\mathcal{R}^+_{\mathcal{O}(U_r)}(\Gamma)$ (see [loc. cit., Section 2.12] for the definition of the last ring). Now $\mathcal{R}^+_{\mathcal{O}(U_r)}(\Gamma)/(\gamma - 1) \cong \mathcal{O}(U_r)$. Therefore $(D^+)^{\psi=0}/(\gamma - 1)$ is a free $\mathcal{O}(U_r)$-module of rank 1 generated by $S_1(1 + T)w$. □

For the remainder of section, we drop w when we talk about the $\mathcal{O}(U_r)$-generator $S_1(1+T)w$ of $(D^+)^{\psi=0}/(\gamma-1)$. Using these lemmas, we see that the quotient $C(D^+)/(\gamma-1)$ is generated by $S_1(1+T)$ as a free $\mathcal{O}(U_r)$-module. Pushing this generator to $C(D)/(\gamma-1)$ under the first isomorphism $C(D^+)/(\gamma-1) \to C(D)/(\gamma-1)$, we see that $S_1(1+T)$ is an $\mathcal{O}(U_r)$-generator of $C(D)/(\gamma-1)$. Now we want a generator of $D^{\psi=1}/(\gamma-1)$.

Lemma 4.4. *The element $y = S_1 \sum_{n=0}^{\infty} S_1^n (1+T)^{p^n}$ is the preimage of $S_1(1+T)$ under the isomorphism $1-\varphi : D^{\psi=1}/(\gamma-1) \to C(D)/(\gamma-1)$.*

Proof. Note $y \in D^{\psi=1}$ (for the twisted ψ). To see this, write (for the twisted φ)

$$S_1 \sum_{n=0}^{\infty} S_1^n (1+T)^{p^n} = \varphi\left(S_1 \sum_{n=0}^{\infty} S_1^n (1+T)^{p^n}\right) + (1+T)\varphi(1).$$

From the definition of ψ (see [Chenevier 2013, Section 2.1]), we see that $y \in D^{\psi=1}$. Clearly $(1-\varphi)y = S_1(1+T)$. □

Finally, consider the third isomorphism $D^{\psi=1}/(\gamma-1) \to H^1_{(\psi,\Gamma)}(D)$ induced by sending $f(T)$ to $(0, f(T))$. Using this isomorphism, we see that the class of $(0, y)$ is a generator of $H^1_{(\psi,\Gamma)}(D)$ as a free $\mathcal{O}(U_r)$-module. We repeat that this cohomology group is in the sense of (ψ, Γ)-modules. In order to get a generator of the (φ, Γ)-cohomology group, we note that D is a rank-1 module so it is tautologically trianguline, so $\gamma - 1$ is bijective on $D^{\psi=0}$ by [loc. cit., Corollary 2.5], so

$$H^1_{(\varphi,\Gamma)}(D) \xrightarrow{(-\psi,\mathrm{id})} H^1_{(\psi,\Gamma)}(D)$$

is an isomorphism of cohomology groups, by the discussion at the beginning of this section. Since $y \in D^{\psi=1}$, we have $(\varphi-1)y \in D^{\psi=0}$. Since $\gamma-1$ is bijective on $D^{\psi=0}$, there exists a unique $x \in D^{\psi=0}$ such that $(\gamma-1)x = (\varphi-1)y$. Therefore (x, y) represents a class in $H^1_{(\varphi,\Gamma)}(D)$. Moreover, the class of (x, y) maps to the class of $(0, y)$ under the isomorphism above. Therefore, the class of (x, y) generates $H^1_{(\varphi,\Gamma)}(D)$ as a free $\mathcal{O}(U_r)$-module.

For the rest of this paper, every cohomology group that we write will be the (φ, Γ)-version, and we drop the subscript "(φ, Γ)" from the notation.

We conclude this section by describing an $\mathcal{O}(U_r)$-generator of $H^1(m\mathcal{R}_{\mathcal{O}(U_r)}(\delta_{U_r}))$.

Using the isomorphism $H^1(mD) \to H^1(D)$, we see that the class of (x, y) is represented by elements of mD: there exists a $d \in D$ such that $x - (\varphi-1)d,\ y - (\gamma-1)d \in mD$.

Definition 4.5. We denote by e the class of $(x-(\varphi-1)d,\ y-(\gamma-1)d)$ in $H^1(mD)$.

Thus e is an explicit generator of the free $\mathcal{O}(U_r)$-module $H^1(m\mathcal{R}_{\mathcal{O}(U_r)}(\delta_{U_r}))$ of rank 1.

5. The $\mathcal{L}$-invariant of the limit point

Recall that the fiber in the blow-up $\widetilde{U}_r$ of U_r over the exceptional point $(p^r, (1+p)^{r+1}-1)$ corresponding to the maximal ideal $m = (f_1, f_2)$ is parametrized by $\mathbb{P}^1(\overline{\mathbb{Q}}_p)$. The main theorem in this section gives us

a formula for the $\mathcal{L}$-invariant (in the sense of Definition 3.6) of the (φ, Γ)-module associated to a general E-valued point $(p^r, (1+p)^{r+1}-1, a:b)$ in the exceptional fiber. Fix such a point $(p^r, (1+p)^{r+1}-1, a:b)$. Let $v \in \mathrm{Hom}(m/m^2 \otimes_{\mathbb{Q}_p} E, E)$ be a tangent vector representing the tangent direction associated with this point by Proposition 2.5. Recall that $D = \mathcal{R}_{\mathcal{O}(U_r)}(\delta_{U_r})$. Consider the specialization map $v^* : E \otimes_{\mathbb{Q}_p} mD \to \mathcal{R}_E(x^r\chi)$ induced by the composition of maps

$$mD \to m \otimes D \to m/m^2 \otimes_{\mathbb{Q}_p} D/mD \xrightarrow{v} E \otimes_{\mathbb{Q}_p} \mathcal{R}_{\mathbb{Q}_p}(x^r\chi) = \mathcal{R}_E(x^r\chi)$$

given by

$$fg \mapsto f \otimes g \mapsto \bar{f} \otimes \bar{g} \mapsto v(\bar{f})\bar{g}$$

for all $f \in m$, $g \in D$, where the first map is the inverse of the multiplication map $m \otimes D \to mD$ (which is an isomorphism since $\mathcal{R}_{\mathcal{O}(U_r)}$ is flat over $\mathcal{O}(U_r)$ by [Chenevier 2013, Lemma 1.3(v)]), and the second map is the map obtained by going mod m on each factor. In particular, we have

$$v^*(f_1g_1 + f_2g_2) = v(\bar{f_1})\bar{g}_1 + v(\bar{f_2})\bar{g}_2,$$

where $g_1, g_2 \in D$ are arbitrary, $\bar{g}_1, \bar{g}_2$ are their images in D/mD, and $\bar{f_1}$, $\bar{f_2}$ are the images of $f_1 = S_1 - p^r$ and $f_2 = 1 + S_2 - (1+p)^{r+1}$ in m/m^2, respectively.

The map v^* yields a specialization map $H^1(v^*) : E \otimes_{\mathbb{Q}_p} H^1(mD) \to H^1(\mathcal{R}_E(x^r\chi))$ induced by

$$H^1(v^*)(\eta_1, \eta_2) = (v^*(\eta_1), v^*(\eta_2))$$

for $(\eta_1, \eta_2) \in mD \oplus mD$ representing a cohomology class in $H^1(mD)$. Recall that e is represented by

$$(x - (\delta_{U_r}(p)\varphi - 1)d,\ y - (\delta_{U_r}(\chi(\gamma))\gamma - 1)d) \in mD \oplus mD$$

and is an $\mathcal{O}(U_r)$-generator of $H^1(mD)$, where now φ and γ are the usual untwisted operators (see Definition 4.5). Here we have rewritten e in terms of these untwisted operators since this explicit formula for e will be needed below. We make the following definition.

Definition 5.1. The (φ, Γ)-module associated[5] to the point $(p^r, (1+p)^{r+1}-1, a:b)$ is the image of e in $H^1(\mathcal{R}_E(x^r\chi))$ under $H^1(v^*)$.

Note that the generator e and the tangent vector v are only well-defined up to scalars, but the isomorphism class of this (φ, Γ)-module is unchanged under multiplication by scalars. The theorem below gives us a formula for the $\mathcal{L}$-invariant of the (φ, Γ)-module associated to $(p^r, (1+p)^{r+1}-1, a:b)$ in the exceptional fiber in terms of a and b. Note that the $\mathcal{L}$-invariant only depends on the projective image of the cohomology class $H^1(v^*)(e)$.

Theorem 5.2. *The $\mathcal{L}$-invariant of the (φ, Γ)-module associated to the E-valued point in the exceptional fiber with coordinates $(p^r, (1+p)^{r+1}-1, a:b)$ is*

$$\mathcal{L} = -\frac{(1+p)^{r+1}\log(1+p)}{p^r} \cdot \frac{a}{b} \in \mathbb{P}^1(E).$$

[5]There is a more conceptual definition of this (φ, Γ)-module in terms of a cover U_i of $\widetilde{U}_r$ as the module $(D_i)_z$, with notation as in the proof of [Chenevier 2013, Proposition 3.9], but this definition coincides with the more computational one given here by [loc. cit., Theorem 2.33] (see the remarks at the end of that proof).

Proof. We compute the $\mathcal{L}$-invariant of the (φ, Γ)-module given by $H^1(v^*)(e)$ using Definition 3.6. For convenience of notation, let $\mathrm{Res}(f(T)) = \mathrm{res}(f(T)\,dt)$ for $f(T) \in \mathcal{R}_E, \mathcal{R}_{\mathcal{O}(U_r)}$. Write

$$H^1(v^*)(e) = \lambda_v \cdot \bar{\alpha}_{r+1} + \mu_v \cdot \bar{\beta}_{r+1},$$

where, by (19),

$$\lambda_v = \mathrm{Res}(t^r v^*(x - (\delta_{U_r}(p)\varphi - 1)d)), \quad \mu_v = \mathrm{Res}(t^r v^*(y - (\delta_{U_r}(\chi(\gamma))\gamma - 1)d)).$$

To compute these residues, we first note that the following square commutes:

$$\begin{array}{ccc} mD \otimes_{\mathbb{Q}_p} E & \xrightarrow{v^*} & \mathcal{R}_E(x^r\chi) \\ \downarrow{\scriptstyle \mathrm{Res}(t^r\_)} & & \downarrow{\scriptstyle \mathrm{Res}(t^r\_)} \\ m \otimes_{\mathbb{Q}_p} E & \xrightarrow{v} & E \end{array}$$

where we write v again for the map $m \otimes E \xrightarrow{-} m/m^2 \otimes E \xrightarrow{v} E$. Indeed, given any element $f_1 g_1 + f_2 g_2 \in mD$, we have

$$\begin{aligned} \mathrm{Res}(t^r v^*(f_1 g_1 + f_2 g_2)) &= \mathrm{Res}(t^r (v(\bar{f}_1)\bar{g}_1 + v(\bar{f}_2)\bar{g}_2)) \\ &= v(\bar{f}_1)\,\mathrm{Res}(t^r \bar{g}_1) + v(\bar{f}_2)\,\mathrm{Res}(t^r \bar{g}_2) \\ &= v(\mathrm{Res}(t^r g_1) f_1 + \mathrm{Res}(t^r g_2) f_2) \\ &= v(\mathrm{Res}(t^r (f_1 g_1 + f_2 g_2))). \end{aligned}$$

Now we prove that $\mathrm{Res}(t^r x) = 0$ and $\mathrm{Res}(t^r y) = 0$. We use the following formulas, which are similar to (20) and (21). For any $f(T) \in \mathcal{R}_{\mathcal{O}(U_r)}$, we have

$$\mathrm{Res}(\varphi(f(T))) = \mathrm{Res}(f(T)), \tag{23}$$

$$\mathrm{Res}(\gamma(f(T))) = \chi(\gamma)^{-1}\,\mathrm{Res}(f(T)). \tag{24}$$

Using the definition of x and y (see the discussion before Definition 4.5), we get

$$(\delta_{U_r}(\chi(\gamma))\gamma - 1)x = (\delta_{U_r}(p)\varphi - 1)y = -S_1(1+T),$$

by Lemma 4.4. Multiplying by t^r and taking residues, by (12) and (11), we get

$$\begin{aligned} \mathrm{Res}((\delta_{U_r}(\chi(\gamma))\chi(\gamma)^{-r}\gamma - 1)t^r x) &= \mathrm{Res}(-t^r S_1(1+T)) = 0, \\ \mathrm{Res}((\delta_{U_r}(p)p^{-r}\varphi - 1)t^r y) &= \mathrm{Res}(-t^r S_1(1+T)) = 0. \end{aligned}$$

Applying (24) and (23) to the terms on the left, we get $\mathrm{Res}(t^r x) = 0 = \mathrm{Res}(t^r y)$.

Thus, using the commutativity of the diagram above, we get

$$\begin{aligned} \lambda_v &= v(\mathrm{Res}(t^r (x - (\delta_{U_r}(p)\varphi - 1)d))) \\ &= -v(\mathrm{Res}(t^r (\delta_{U_r}(p)\varphi - 1)d)) \\ &\overset{(11)}{=} -v(\mathrm{Res}((\delta_{U_r}(p)p^{-r}\varphi - 1)t^r d)). \end{aligned}$$

Using (23), we get

$$\begin{aligned}\lambda_v &= -v((\delta_{U_r}(p)p^{-r}-1)\operatorname{Res}(t^r d))\\ &\overset{(22)}{=} -v((S_1-p^r)p^{-r}\operatorname{Res}(t^r d))\\ &= -v(f_1 p^{-r}\operatorname{Res}(t^r d)) = -p^{-r}v(f_1\operatorname{Res}(t^r d)).\end{aligned}$$

The expression in the brackets belongs to m. Recall that $v : m\otimes E\to E$ is the composition of the maps $^- : m\to m/m^2$ and $v : m/m^2\otimes E\to E$. We therefore get

$$\lambda_v = -p^{-r}\overline{\operatorname{Res}(t^r d)}\cdot v(\bar{f}_1) = -p^{-r}\overline{\operatorname{Res}(t^r d)}\cdot a.$$

Similarly,

$$\begin{aligned}\mu_v &= v\big(\operatorname{Res}(t^r(y-(\delta_{U_r}(\chi(\gamma))\gamma-1)d))\big)\\ &= -v\big(\operatorname{Res}(t^r(\delta_{U_r}(\chi(\gamma))\gamma-1)d)\big)\\ &\overset{(12)}{=} -v\big(\operatorname{Res}((\delta_{U_r}(\chi(\gamma))\chi(\gamma)^{-r}\gamma-1)t^r d)\big).\end{aligned}$$

Using (24), we get

$$\mu_v = -v\big((\delta_{U_r}(\chi(\gamma))\chi(\gamma)^{-(r+1)}-1)\operatorname{Res}(t^r d)\big).$$

Since $\chi(\gamma)=\zeta_{p-1}^a(1+p)$, for some a, we get

$$\begin{aligned}\mu_v &\overset{(22)}{=} -v\big((\zeta_{p-1}^{a(r+1)}(1+S_2)\zeta_{p-1}^{-a(r+1)}(1+p)^{-(r+1)}-1)\operatorname{Res}(t^r d)\big)\\ &= -v(f_2(1+p)^{-(r+1)}\operatorname{Res}(t^r d))\\ &= -(1+p)^{-(r+1)}v(f_2\operatorname{Res}(t^r d)).\end{aligned}$$

The expression inside the brackets is an element of m. As above, we get

$$\mu_v = -(1+p)^{-(r+1)}\overline{\operatorname{Res}(t^r d)}\cdot v(\bar{f}_2) = -(1+p)^{-(r+1)}\overline{\operatorname{Res}(t^r d)}\cdot b.$$

Both λ_v and μ_v cannot be equal to 0 simultaneously because $H^1(v^*)(e)\neq 0$. So $\overline{\operatorname{Res}(t^r d)}\neq 0$. By Definition 3.6 (or directly by [Benois 2011, Proposition 2.3.7]), the $\mathcal{L}$-invariant of $H^1(v^*)(e)$ is

$$-\log(\chi(\gamma))\cdot\frac{\lambda_v}{\mu_v} = -\frac{(1+p)^{r+1}\log(1+p)}{p^r}\cdot\frac{a}{b}.$$ □

6. Proof of Theorem 1.1 and generalizations

6.1. ***Proof of Theorem 1.1.*** Let $k\geq 3$ and let $r=k-2$. Let (k_n, a_n) for $n\geq 1$ be as in (1). These quantities depend on $\mathcal{L}\in\mathbb{P}^1(\overline{\mathbb{Q}}_p)$. We prove that the sequence of crystalline representations $V^*_{k_n,a_n}$ converges to the semistable representation $V^*_{k,\mathcal{L}}$.

We recall some facts from Section 1.3. Recall that the ordered pair of characters $(\delta_{1,n},\delta_{2,n})$ is associated to the representation $V^*_{k_n,a_n}$ and the sequence of characters $\delta_{1,n}\delta_{2,n}^{-1}=\mu_{y_n^2}\chi^{k_n-1}$, where y_n is as in (3), converges to the exceptional character $x^r\chi$. Moreover, the character $\mu_{y_n^2}\chi^{k_n-1}$ corresponds to the following point of $\widetilde{U}_r$:

$$(y_n^2,(1+p)^{k_n-1}-1,y_n^2-p^r : (1+p)^{k_n-1}-(1+p)^{k-1}). \tag{25}$$

We compute the limit of the above points in $\widetilde{U}_r$ as $n \to \infty$ using the following lemmas.

Lemma 6.1. *We have*

$$\lim_{n\to\infty} \frac{y_n^2 - p^r}{p^n(p-1)} = \begin{cases} \mathcal{L}p^r & \text{if } \mathcal{L} \neq \infty, \\ 2p^r/(p-1) & \text{if } \mathcal{L} = \infty. \end{cases}$$

Proof. Assume $\mathcal{L} \neq \infty$. We have

$$y_n^2 - p^r = \left(\frac{a_n + \sqrt{a_n^2 - 4p^{k_n-1}}}{2}\right)^2 - p^r = a_n^2\left(\frac{1 + \sqrt{1 - 4p^{k_n-1}a_n^{-2}}}{2}\right)^2 - p^r.$$

For large n, the valuation of $a_n = p^{r/2} + \frac{1}{2}\mathcal{L}p^{n+r/2}(p-1)$ is equal to $\frac{1}{2}r$. Therefore for large n, the expression inside the second radical sign is an element of $1 + p^{1+p^n(p-1)}\mathcal{O}_E$, where $\mathcal{O}_E$ is the ring of integers of E. Since taking square roots is an automorphism of this group, we see that $\sqrt{1 - 4p^{k_n-1}a_n^{-2}} \in 1 + p^{1+p^n(p-1)}\mathcal{O}_E$. Therefore we can write

$$\frac{1 + \sqrt{1 - 4p^{k_n-1}a_n^{-2}}}{2} = 1 + up^{1+p^n(p-1)}$$

for some $u \in \mathcal{O}_E$. Substituting this in the previous equation, we get

$$\begin{aligned} y_n^2 - p^r &= \left(p^r + \mathcal{L}p^{n+r}(p-1) + \tfrac{1}{4}\mathcal{L}^2p^{2n+r}(p-1)^2\right)(1 + 2up^{1+p^n(p-1)} + u^2p^{2+2p^n(p-1)}) - p^r \\ &= p^r + \mathcal{L}p^{n+r}(p-1) + p^{2n}(u') - p^r \end{aligned}$$

for some u' whose valuation is bounded below. So

$$\lim_{n\to\infty} \frac{y_n^2 - p^r}{p^n(p-1)} = \mathcal{L}p^r.$$

The limit in the case $\mathcal{L} = \infty$ is computed similarly. □

Lemma 6.2. *We have*

$$\lim_{n\to\infty} \frac{(1+p)^{k_n-1} - (1+p)^{k-1}}{p^n(p-1)} = \begin{cases} (1+p)^{k-1}\log(1+p) & \text{if } \mathcal{L} \neq \infty, \\ 0 & \text{if } \mathcal{L} = \infty. \end{cases}$$

Proof. Assume $\mathcal{L} \neq \infty$. Since

$$(1+p)^{k_n-1} - (1+p)^{k-1} = (1+p)^{k-1}[(1+p)^{p^n(p-1)} - 1],$$

we see that

$$\begin{aligned} \lim_{n\to\infty} \frac{(1+p)^{k_n-1} - (1+p)^{k-1}}{p^n} &= (1+p)^{k-1} \lim_{n\to\infty} \frac{(1+p)^{p^n(p-1)} - 1}{p^n} \\ &= (1+p)^{k-1} \lim_{n\to\infty} \frac{\{1 + [(1+p)^{(p-1)} - 1]\}^{p^n} - 1}{p^n} \\ &\overset{(10)}{=} (1+p)^{k-1}\log(1+p)^{p-1}. \end{aligned}$$

The limit in the case $\mathcal{L} = \infty$ is proved similarly. □

Write the sequence (25) as

$$\left(y_n^2,\,(1+p)^{k_n-1}-1,\,\frac{y_n^2-p^r}{p^n(p-1)}:\frac{(1+p)^{k_n-1}-(1+p)^{k-1}}{p^n(p-1)}\right).$$

Using Lemmas 6.1 and 6.2, we see that the sequence above converges to the point

$$\begin{cases}\big(p^r,\,(1+p)^{k-1}-1,\,\mathcal{L}p^r:(1+p)^{k-1}\log(1+p)\big) & \text{if } \mathcal{L}\neq\infty,\\ (p^r,\,(1+p)^{k-1}-1,\,1:0) & \text{if } \mathcal{L}=\infty.\end{cases}$$

Assume $\mathcal{L}\neq\infty$. By Theorem 5.2, the $\mathcal{L}$-invariant of the (φ,Γ)-module associated to this limit point is $-\mathcal{L}$. By [Colmez 2008, Théorème 0.5(i)], this (φ,Γ)-module is also étale since $(\mu_{p^{r/2}},\mu_{1/p^{r/2}}\chi^{1-k},-\mathcal{L})\in\mathscr{S}_*\setminus\mathscr{S}_*^{\mathrm{ncl}}$ in the notation of [loc. cit.] since $\frac{1}{2}r-\frac{1}{2}r=0$, $\frac{1}{2}r>0$ and $\frac{1}{2}r\neq k-1$. The corresponding Galois representation is $V(\mu_{p^{r/2}},\mu_{1/p^{r/2}}\chi^{1-k},-\mathcal{L})$ in the notation of [loc. cit.]. Comparing the filtered (φ,N)-module associated to $V^*_{k,\mathcal{L}}$ given in the Introduction with the one associated to $V(\mu_{p^{r/2}},\mu_{1/p^{r/2}}\chi^{1-k},-\mathcal{L})$ using [loc. cit., Section 4.6] (more specifically, [loc. cit., Proposition 4.18] with $a=1-k$, $b=0$, $\alpha=\mu_{p^{k/2}}$, and $\mathcal{L}$ replaced by $-\mathcal{L}$), we see that $V(\mu_{p^{r/2}},\mu_{1/p^{r/2}}\chi^{1-k},-\mathcal{L})\simeq V^*_{k,\mathcal{L}}$. Thus the sequence of crystalline representations $V^*_{k_n,a_n}$ converges to the semistable representation $V^*_{k,\mathcal{L}}$ for $\mathcal{L}\in E$.

Now assume we are in the $\mathcal{L}=\infty$ case. By Theorem 5.2, the $\mathcal{L}$-invariant associated to the above limit point is ∞. The corresponding Galois representation is $V(\mu_{p^{r/2}},\mu_{1/p^{r/2}}\chi^{1-k},\infty)$ in the notation of [loc. cit.]. Comparing the filtered φ-module associated to $V^*_{k,\infty}$ in the Introduction with the one associated to $V(\mu_{p^{r/2}},\mu_{1/p^{r/2}}\chi^{1-k},\infty)$ in [loc. cit., Section 4.5] (more precisely, take $a=1-k$, $b=0$, $\beta=\mu_{p^{r/2}}$ and $\alpha=\mu_{p^{k/2}}$ in the second isomorphism in [loc. cit., Proposition 4.13]), we see that $V(\mu_{p^{r/2}},\mu_{1/p^{r/2}}\chi^{1-k},\infty)\simeq V^*_{k,\infty}$. (Alternatively, the corresponding Galois representation is $V(\mu_{p^{r/2}},\mu_{1/p^{r/2}}\chi^{1-k})$ in the notation of [Berger 2012] and, by Proposition 3.1 of that work, this last representation is isomorphic to the crystalline representation V^*_{k,a_p} with $a_p=p^{k/2}+p^{k/2-1}$, which as mentioned in the Introduction, is isomorphic to $V^*_{k,\infty}$.) Thus, the sequence of crystalline representations $V^*_{k_n,a_n}$ again converges to the (crystalline) representation $V^*_{k,\infty}$.

6.2. ***$\mathcal{L}$-invariants as logarithmic derivatives.*** In this subsection, all $\mathcal{L}$-invariants are finite.

We prove formula (2) from the Introduction showing that $\mathcal{L}$ is twice the logarithmic derivative of a_p. More precisely, we prove that if $a_p:\mathbb{Z}_p\to E$ is a differentiable function of l with $a_p(k)=p^{r/2}$ for $r=k-2$ and $k\geq 3$, then the crystalline representations $V^*_{l,a_p(l)}$ converge in $\widetilde{\mathcal{T}}_2$ to the semistable representation $V^*_{k,\mathcal{L}}$ with

$$\mathcal{L}=2a_p(k)^{-1}a_p'(k)$$

as l tends to k in the p-adic topology through integers $l\equiv k \bmod (p-1)$ with $l\neq k$. The condition $l\equiv k \bmod (p-1)$ is necessary because $V^*_{l,a_p(l)}$ converges to $V^*_{k,\mathcal{L}}$ implies that (the tame part of) $\det V^*_{l,a_p(l)}=\chi^{1-l}$ converges to (the tame part of) $\det V^*_{k,\mathcal{L}}=\chi^{1-k}$.

This formula is a variant of a classical formula due to [Greenberg and Stevens 1993, Theorem 3.18; Stevens 2010, Theorem B; Bertolini, Darmon, and Iovita 2010, Theorem 4; Colmez 2010, Théorème 0.5, Corollaire 0.7; Benois 2010, Theorem 2] and others (see Remark 6.6). Our proof of the formula seems

new. It uses some elementary p-adic analysis (see the two lemmas below) and Theorem 5.2, which in turn uses some geometry (the blow-up space $\widetilde{\mathcal{T}}_2$) and some algebra (the interpretation of this space in terms of trianguline (φ, Γ)-modules over the Robba ring).

Recall that for integer $l \geq 2$, the (φ, Γ)-module $\mathbf{D}_{\mathrm{rig}}(V^*_{l,a_p(l)})$ is an extension of $\mathcal{R}_E(\mu_{1/y(l)}\chi^{1-l})$ by $\mathcal{R}_E(\mu_{y(l)})$, where

$$y(l) = \frac{a_p(l) + \sqrt{a_p(l)^2 - 4p^{l-1}}}{2}.$$

Let $\delta_1(l) = \mu_{y(l)}$ and $\delta_2(l) = \mu_{1/y(l)}\chi^{1-l}$. Since $l \equiv k \bmod (p-1)$, the characters $\delta_1(l)\delta_2(l)^{-1}$ converge to the exceptional character $x^r\chi$ as $l \to k$. Therefore the characters $\delta_1(l)\delta_2(l)^{-1}$ eventually belong to $\widetilde{U}_r$. Moreover, $\delta_1(l)\delta_2(l)^{-1}$ corresponds to the following point of $\widetilde{U}_r$:

$$\big(y(l)^2,\, (1+p)^{l-1} - 1,\, y(l)^2 - p^r : (1+p)^{l-1} - (1+p)^{k-1}\big). \tag{26}$$

We compute the limit of these points in $\widetilde{U}_r$ as $l \to k$ using the following two lemmas.

Lemma 6.3. *We have*

$$\lim_{l \to k} \frac{y(l)^2 - p^r}{l-k} = 2a_p(k)a'_p(k).$$

Proof. The statement generalizes that of Lemma 6.1 in the case $\mathcal{L} \neq \infty$ and we give a slightly different proof. Note that

$$\begin{aligned} y(l)^2 - p^r &= \frac{2a_p(l)^2 + 2a_p(l)\sqrt{a_p(l)^2 - 4p^{l-1}} - 4p^{l-1}}{4} - p^r \\ &= \frac{a_p(l)^2 - p^r}{2} + \frac{a_p(l)^2\sqrt{1 - 4a_p(l)^{-2}p^{l-1}} - p^r}{2} - p^{l-1}. \end{aligned}$$

Therefore

$$\frac{y(l)^2 - p^r}{l-k} = \frac{a_p(l)^2 - p^r}{2(l-k)} + \frac{a_p(l)^2 - p^r}{2(l-k)}\sqrt{1 - 4a_p(l)^{-2}p^{l-1}} + p^r\,\frac{\sqrt{1 - 4a_p(l)^{-2}p^{l-1}} - 1}{2(l-k)} - \frac{p^{l-1}}{l-k}.$$

By the definition of the derivative, the first and the second summands in the equation above converge to $a_p(k)a'_p(k)$ as $l \to k$. The last summand clearly converges to 0. The third summand also converges to 0. Indeed, as $l \to k$, the valuation of $a_p(l)$ becomes $\frac{1}{2}r$. Therefore for such $l \geq k$ we can write

$$\sqrt{1 - 4a_p(l)^{-2}p^{l-1}} = \sum_{i=0}^{\infty} \binom{\frac{1}{2}}{i} (-4a_p(l)^{-2}p^{l-1})^i.$$

Using the fact that $p^{l-1}/l - k$ converges to 0 as $l \to k$, we see that

$$\lim_{l \to k} \frac{\sqrt{1 - 4a_p(l)^{-2}p^{l-1}} - 1}{2(l-k)} = 0.$$

Putting everything together, we see that

$$\lim_{l \to k} \frac{y(l)^2 - p^r}{l-k} = 2a_p(k)a'_p(k). \qquad \square$$

Lemma 6.4. *We have*

$$\lim_{l\to k}\frac{(1+p)^{l-1}-(1+p)^{k-1}}{l-k}=(1+p)^{k-1}\log(1+p).$$

Proof. This is the same as Lemma 6.2 in the case $\mathcal{L}\neq\infty$, so we give a slightly different proof. Write

$$\frac{(1+p)^{l-1}-(1+p)^{k-1}}{l-k}=(1+p)^{k-1}\frac{[(1+p)^{l-k}-1]}{l-k}=(1+p)^{k-1}\Bigg[\sum_{i=1}^{\infty}\frac{1}{i}\binom{l-k-1}{i-1}p^i\Bigg].$$

Since $\binom{-1}{i-1}=(-1)^{i-1}$, taking the limit as $l\to k$ we obtain

$$\lim_{l\to k}\frac{(1+p)^{l-1}-(1+p)^{k-1}}{l-k}=(1+p)^{k-1}\log(1+p).$$ □

Now rewrite the point (26) in the blow-up as

$$\left(y(l)^2,\,(1+p)^{l-1}-1,\,\frac{y(l)^2-p^r}{l-k}:\frac{(1+p)^{l-1}-(1+p)^{k-1}}{l-k}\right).$$

Using Lemmas 6.3 and 6.4, we see that as $l\to k$, the points above converge to

$$\big(p^r,\,(1+p)^{k-1}-1,\,2a_p(k)a_p'(k):(1+p)^{k-1}\log(1+p)\big).$$

By Theorem 5.2, we see that the $\mathcal{L}$-invariant of the (φ,Γ)-module associated to this limit point is $-2a_p(k)^{-1}a_p'(k)$. Working as at the end of Section 6.1, we see that the corresponding Galois representation is $V(\mu_{p^{r/2}},\mu_{1/p^{r/2}}\chi^{1-k},-\mathcal{L})$ with $\mathcal{L}=2a_p(k)^{-1}a_p'(k)$, so the crystalline representations $V^*_{l,a_p(l)}$ converge to the semistable representation $V^*_{k,\mathcal{L}}$ with $\mathcal{L}=2a_p(k)^{-1}a_p'(k)$.

Remark 6.5. We thank one of the referees for pointing out the following simplification to the computation of the limits in Lemmas 6.1 and 6.3. Assume as above that $a_p:\mathbb{Z}_p\to E$ is a differentiable function of l with $a_p(k)=p^{r/2}$ with $r=k-2$ and $k\geq 3$. If we replace the crystalline representations $V^*_{l,a_p(l)}$ by $V^*_{l,a_p(l)+p^{l-1}/a_p(l)}$, then the limit computation in Lemma 6.3 becomes easier. Indeed, the $y(l)$ for the crystalline representation $V^*_{l,a_p(l)+p^{l-1}/a_p(l)}$ is equal to $a_p(l)$. Therefore, for l close to k and $l\equiv k \bmod (p-1)$, the point in $\widetilde{\mathcal{T}}$ associated to $V^*_{l,a_p(l)+p^{l-1}/a_p(l)}$ belongs to $\widetilde{U}_r$ and is given by

$$\big(a_p(l)^2,\,(1+p)^{l-1}-1,\,a_p(l)^2-p^r:(1+p)^{l-1}-(1+p)^{k-1}\big).$$

To evaluate the limit of these points in the third coordinate, we have to evaluate the limit

$$\lim_{\substack{l\to k\\ l\equiv k \bmod p-1}}\frac{a_p(l)^2-p^r}{l-k}.$$

But, this limit is just the derivative of $a_p(l)^2$ at $l=k$ and so is immediately equal to $2a_p(k)a_p'(k)$. Similarly, for the limits in Lemma 6.1. In principle, since the limit of the above sequence is the same as that of the original sequence, the crystalline representations $V_{l,a_p(l)+p^{l-1}/a_p(l)}$ can also be used to compute the reduction of the semistable representation $V_{k,\mathcal{L}}$. Note that $\overline{V}_{l,a_p(l)+p^{l-1}/a_p(l)}\simeq\overline{V}_{l,a_p(l)}$ for l close to k by [Berger 2012, Theorem A].

Also, Lemmas 6.2 and 6.4 follow immediately by differentiating the function $f : \mathbb{Z}_p \to \mathbb{Q}_p$ defined by $f(x) = (1+p)^{x-1}$ at $x = k$.

Remark 6.6 (relation to work of Greenberg and Stevens, etc.). Our formula (2) is slightly different from the classical formula due to [Greenberg and Stevens 1993; Stevens 2010; Colmez 2010; Bertolini, Darmon, and Iovita 2010; Benois 2010] and others. We thank D. Benois for pointing this out.

Let us explain the difference. In the classical setting, one starts with a newform f of weight $k \geq 2$ for the subgroup $\Gamma_0(Np)$, where $(N, p) = 1$ (for simplicity, we assume that the nebentypus at N is trivial) with U_p-eigenvalue $\alpha_p(f) = p^{(k-2)/2}$ (as opposed to $-p^{(k-2)/2}$ for simplicity). Then one takes the Hida family F if f is ordinary (equivalently k = 2), or the Coleman family F if f has positive slope (equivalently $k > 2$), passing through f with U_p-eigenvalue α_p, which is an analytic function of the weight. Note that $\alpha_p(k) = \alpha_p(f)$. The classical formula states that the $\mathcal{L}$-invariant of the form f is

$$\mathcal{L}(f) = -2\frac{\alpha_p'(k)}{\alpha_p(k)}.$$

Incidentally, the above authors use various definitions of the $\mathcal{L}$-invariant of the form f but these are all equal (see [Colmez 2005] for a survey comparing these alternative definitions).

For $l \equiv k \bmod (p-1)$, the nebentypus of the weight l member F_l of the family F is trivial. Since forms living in a Hida or Coleman family have the same slope and since the slope of the U_p-eigenvalue of a $\Gamma_0(Np)$-newform of weight l is equal to $\frac{1}{2}(l-2)$, we see that $f = F_k$ is the unique classical member in the family F of weight $l \equiv k \bmod (p-1)$ that is p-new. In fact, any classical member F_l with $l \equiv k \bmod (p-1)$ and $l \neq k$ arises as a p-stabilization of a form $\widetilde{F}_l$ that is only N-new. The U_p-eigenvalue $\alpha_p(l)$ of F_l is a root of

$$x^2 - a_p(l)x + p^{l-1},$$

where $a_p(l)$ is the T_p-eigenvalue of the form $\widetilde{F}_l$. The local Galois representation

$$\rho_{F_l}|_{\mathrm{Gal}(\bar{\mathbb{Q}}_p/\mathbb{Q}_p)} \simeq \rho_{\widetilde{F}_l}|_{\mathrm{Gal}(\bar{\mathbb{Q}}_p/\mathbb{Q}_p)}$$

is isomorphic to the crystalline representation $V_{l,a_p(l)}$. Moreover, $\rho_f|_{\mathrm{Gal}(\bar{\mathbb{Q}}_p/\mathbb{Q}_p)}$ is isomorphic to the semistable representation $V_{k,\mathcal{L}}$ with $\mathcal{L} = -\mathcal{L}(f)$. The change in sign is because the $\mathcal{L}$-invariant in the filtration on $D_{k,\mathcal{L}}$ given in the Introduction is the negative of the one in the filtration on the filtered module in [Mazur 1994]. Therefore in the classical setup,

$$\mathcal{L} = -\mathcal{L}(f) = 2\frac{\alpha_p'(k)}{\alpha_p(k)}.$$

In our setup, we have a smooth function $a_p : \mathbb{Z}_p \to E$ such that $a_p(k) = p^{(k-2)/2}$. We prove that the sequence of crystalline representations $V^*_{l,a_p(l)}$ converge to the semistable representation $V^*_{k,\mathcal{L}}$, with

$$\mathcal{L} = 2\frac{a_p'(k)}{a_p(k)}.$$

Note that there does not seem to be a common framework within which one may compare these two formulas. Since the functions $a_p(l)$ and $\alpha_p(l)$ associated to the Hida or Coleman families above are related by the formula

$$a_p(l) = \alpha_p(l) + \frac{p^{l-1}}{\alpha_p(l)},$$

the $a_p(l)$ above cannot be interpolated by a smooth function $a_p : \mathbb{Z}_p \to E$. Indeed, p^{l-1} is not even defined on the whole of $\mathbb{Z}_p$. However, we can obtain the classical formula above using the trick in Remark 6.5. Indeed, suppose $\alpha_p : \mathbb{Z}_p \to E$ is a differentiable function of l with $\alpha_p(k) = p^{(k-2)/2}$ with $k \geq 3$. Consider the crystalline representations $V^*_{l,\alpha_p(l)+p^{l-1}/\alpha_p(l)}$ for integer $l \equiv k \bmod (p-1)$ with $l \neq k$. The $y(l)$ for these representations is $\alpha_p(l)$. Therefore, for l close to k and $l \equiv k \bmod (p-1)$, the point in $\widetilde{\mathcal{T}}$ associated to $V^*_{l,\alpha_p(l)+p^{l-1}/\alpha_p(l)}$ belongs to $\widetilde{U}_r$ and is given by

$$\big(\alpha_p(l)^2, (1+p)^{l-1} - 1, \alpha_p(l)^2 - p^r : (1+p)^{l-1} - (1+p)^{k-1}\big).$$

Arguing as in Remark 6.5 (with $a_p(l)$ there replaced by $\alpha_p(l)$), we see that as l tends to k the limit of the above sequence of points is

$$\big(p^r, (1+p)^{k-1} - 1, 2\alpha_p(k)\alpha_p'(k) : (1+p)^{k-1}\log(1+p)\big).$$

Using Theorem 5.2, we see that the $\mathcal{L}$-invariant of the (φ, Γ)-module associated to the limit point is $-\mathcal{L}$ with

$$\mathcal{L} = 2\frac{\alpha_p'(k)}{\alpha_p(k)}.$$

In the notation of [Colmez 2008], the corresponding Galois representation is $V(\mu_{p^{r/2}}, \mu_{1/p^{r/2}}, -\mathcal{L})$, which is isomorphic to $V^*_{k,\mathcal{L}}$ as explained at the end of Section 6.1.

7. Proof of Theorem 1.3

In this section, we use Theorem 1.1 along with local constancy for the reduction and the zig-zag conjecture to compute the reductions of semistable representations $V_{k,\mathcal{L}}$ with weights k in the range $[3, p+1]$ for p odd.

Consider the crystalline representation V_{k_n,a_n}, with (k_n, a_n) as in (1). Since $v_p(a_n) = \frac{1}{2}r$ for $r = k-2$, for n sufficiently large, the weight k_n is an *exceptional* weight with respect to the half-integral slope $0 < v = \frac{1}{2}r \leq \frac{1}{2}(p-1)$ in the sense of [Ghate 2021, Section 1], namely $k_n \equiv 2v+2 \bmod (p-1)$ for n large. Therefore, we can apply Conjecture 1.2 to compute the reduction of V_{k_n,a_n} for large n. The zig-zag conjecture specifies the reduction in terms of an alternating sequence of irreducible and reducible mod p representations depending on the relative size of a rational parameter τ with respect to (certain integer shifts of) another integer parameter t.

For V_{k_n,a_n}, the formula for τ (see [Ghate 2021, (1.1)]), which we call τ_n, is

$$\tau_n = v_p\left(\frac{a_n^2 - \binom{k_n-2-v_-}{v_+}\binom{k_n-2-v_+}{v_-}p^r}{pa_n}\right),$$

where v_- and v_+ are the largest and the smallest integers, respectively, such that $v_- < v < v_+$. Let us simplify this. We first treat the case $\mathcal{L} \neq \infty$. We have

$$\tau_n = v_p\left(\frac{p^r\left(1+\frac{1}{2}p^n(p-1)\mathcal{L}\right)^2 - \binom{k+p^n(p-1)-2-v_-}{v_+}\binom{k+p^n(p-1)-2-v_+}{v_-}p^r}{p^{1+\frac{1}{2}r}\left(1+\frac{1}{2}p^n(p-1)\mathcal{L}\right)}\right).$$

For large n, we have $v_- + v_+ = k-2$, so

$$\begin{aligned}\binom{k+p^n(p-1)-2-v_-}{v_+} &= \frac{(1+p^n(p-1))(2+p^n(p-1))\cdots(v_+ + p^n(p-1))}{v_+!}\\ &= 1 + p^n(p-1)H_+ + \text{terms involving } p^{2n}\end{aligned}$$

for the harmonic sum[6] $H_+ = \sum_{i=1}^{v_+} 1/i$. Similarly,

$$\binom{k+p^n(p-1)-2-v_+}{v_-} = 1 + p^n(p-1)H_- + \text{terms involving } p^{2n}$$

for $H_- = \sum_{i=1}^{v_-}(1/i)$ (if $v_- \geq 1$; $H_- = 0$ if $v_- = 0$). Substituting, we get

$$\begin{aligned}\tau_n = r + v_p\Big(\big(1+\tfrac{1}{2}\mathcal{L}p^n(p-1)\big)^2 - [1+p^n(p-1)(H_-+H_+)+\text{terms involving } p^{2n}]\Big)\\ - v_p\big(p^{1+\frac{1}{2}r}\big(1+\tfrac{1}{2}\mathcal{L}p^n(p-1)\big)\big)\\ = r + v_p(\mathcal{L}p^n(p-1) - (H_-+H_+)p^n(p-1) + \text{terms involving } p^{2n}) - \big(1+\tfrac{1}{2}r\big) - v_p\big(1+\tfrac{1}{2}\mathcal{L}p^n(p-1)\big).\end{aligned}$$

Thus, for large n,

$$\tau_n = \tfrac{1}{2}r - 1 + n + v_p(\mathcal{L} - H_- - H_+)$$

if $\mathcal{L} \neq H_- + H_+$ (and $\tau_n \geq \frac{1}{2}r - 1 + 2n + c$ for some $c \in \mathbb{Q}$ independent of n if $\mathcal{L} = H_- + H_+$).

As mentioned above, the zig-zag conjecture involves another parameter t which for V_{k_n,a_n} we call t_n. We have

$$t_n = v_p(k_n - 2 - r) = n.$$

These formulas for τ_n and t_n show that for large n the parameter $\tau_n - t_n$ lies, independently of n, in one of the intervals (which may possibly be a point) that appear in the statement of [Ghate 2021, Conjecture 1.1]. This interval is determined by the size of $\nu = v_p(\mathcal{L} - H_- - H_+)$ (and is the rightmost one when $\nu = \infty$). The conjecture accordingly specifies the exact shape of the reductions of the crystalline representations V_{k_n,a_n} for large n in terms of the size of ν. Since taking duals commutes with taking reduction, we obtain the reductions of the $V^*_{k_n,a_n}$ for large n in terms of the size of ν. Using Theorem 1.1 and local constancy for the reduction, we get the reduction of the limiting semistable representation $V^*_{k,\mathcal{L}}$ in terms of the size of ν. Finally, taking duals again, we obtain Theorem 1.3 for $\mathcal{L}$ finite.

Now assume we are in the case $\mathcal{L} = \infty$. Then a computation similar to the one above shows that $\tau_n = \frac{1}{2}r - 1 + n$ and $t_n = n^2$, so that for large n we have $\tau_n < t_n$ (and so $\tau_n - t_n$ is in the leftmost interval

[6]Thus, the provenance of the harmonic sums that are prevalent in all the computation of the reductions of semistable representations in the literature can now be traced back to the p-adic expansions of the binomial coefficients appearing in the zig-zag conjecture.

appearing in the conjecture). Using [Ghate 2021, Conjecture 1.1] and Theorem 1.1 again, we see that $\overline{V}_{k,\infty} \sim \mathrm{ind}(\omega_2^{k-1})$, at least on the inertia subgroup $I_{\mathbb{Q}_p}$. Thus Theorem 1.3 also holds for $\mathcal{L} = \infty$ as $\nu = -\infty$ (but as remarked after the theorem, the result in this case is classical by [Edixhoven 1992] and even holds without assuming zig-zag!).

8. Bounded Hodge–Tate weights

In this section, we use the techniques of this paper to give a proof of the fact that the limit of a sequence of two-dimensional (irreducible) crystalline representations V_n for $n \geq 1$ with Hodge–Tate weights in an interval $[a, b]$ such that the difference of the Hodge–Tate weights is at least 2 infinitely often is also (irreducible) crystalline with Hodge–Tate weights in the interval $[a, b]$. However, note that Berger [2004, Théorème 1] has proved more generally that the limit of subquotients of crystalline representations with bounded Hodge–Tate weights belonging to $[a, b]$ is also crystalline with Hodge–Tate weights in $[a, b]$.

Twisting by a fixed power of the cyclotomic character, we may assume that infinitely often the Hodge–Tate weights of V_n are $(0, k_n - 1)$ with $k_n \geq 3$. This sequence of integers k_n is not to be confused with the one defined in (1). Then there exists a sequence of unramified characters μ_n for $n \geq 1$ such that $V_n \simeq V_{k_n,a_n} \otimes \mu_n$ for some a_n with $v_p(a_n) > 0$. For each $n \geq 1$, consider the ordered pair of characters $(\delta_{1,n}, \delta_{2,n})$ associated with V_n^* under triangulation. We have $\delta_{1,n} = \mu_{y_n} \cdot \mu_n^{-1}$ and $\delta_{2,n} = \mu_{1/y_n} \chi^{1-k_n} \cdot \mu_n^{-1}$, where y_n is as in (3). The assumption that the sequence V_n converges means that the associated sequence of points in $\widetilde{\mathcal{T}}_2$ converge to a point in $\widetilde{\mathcal{T}}_2$. Say that the sequence $(\delta_{1,n}, \delta_{2,n})$ converges to (δ_1, δ_2) in $(\mathcal{T} \times \mathcal{T}) \setminus F'$.

First assume that $\delta_1 \delta_2^{-1}$ is not of the form $x^r \chi$ for $r \geq 0$. Then, by convention, the $\mathcal{L}$-invariant associated to the limit point is ∞. Using the convergence of (the unramified parts of) $\delta_{1,n}\delta_{2,n}$ and $\delta_{1,n}\delta_{2,n}^{-1}$, we see that μ_n^{-2} converges to μ_{α^2} for some $\alpha \in \overline{\mathbb{Q}}_p^*$ and $\mu_{y_n^2}$ converges to μ_{y^2} for some $y \in \overline{\mathbb{Q}}_p^*$. Therefore $\delta_{1,n} = \mu_{y_n} \cdot \mu_n^{-1}$ converges to $\delta_1 = \mu_{\pm\alpha y}$. Similarly, $\delta_{2,n}$ converges to $\delta_2 = \mu_{\pm\alpha/y} \chi^{1-k}$ for some k. So the sequence V_n^* converges to $V(\mu_{\pm\alpha y}, \mu_{\pm\alpha/y} \chi^{1-k}, \infty)$, for some k, in the notation of [Colmez 2008]. Now k_n is a bounded sequence of integers converging to k. Therefore $k_n = k$ for large n. In particular, $k \geq 3$. But $V(\mu_{\pm\alpha y}, \mu_{\pm\alpha/y} \chi^{1-k}, \infty)$ is equal to the crystalline representation $V^*_{k,y+p^{k-1}/y} \otimes \mu_{\pm\alpha}$. Taking duals, we see that the sequence V_n converges to the crystalline representation $V_{k,y+p^{k-1}/y} \otimes \mu_{\pm\alpha}^{-1}$. Untwisting by the fixed power of the cyclotomic character, the original sequence V_n also converges to a crystalline representation with Hodge–Tate weights in $[a, b]$.

Now assume that $\delta_1 \delta_2^{-1} = x^r \chi$ for some $r \geq 0$. This means that $\delta_{1,n}\delta_{2,n}^{-1} = \mu_{y_n^2} \chi^{k_n - 1}$ converges to $x^r \chi$ in the neighborhood U_r of $x^r \chi$ in $\mathcal{T}$. Let $k = r + 2$. Since k_n is a bounded sequence of integers, the convergence implies that $k_n = k$ for all but finitely many n. In particular, we have $k \geq 3$. The point corresponding to the character $\mu_{y_n^2} \chi^{k-1}$ in the blow-up $\widetilde{U}_r$ of U_r is

$$(y_n^2, (1+p)^{k-1} - 1, y_n^2 - p^r : 0) = (y_n^2, (1+p)^{k-1} - 1, 1 : 0),$$

at least if $y_n^2 \neq p^r$ (if $y_n^2 = p^r$, then it is also the last point since this is the only crystalline point in the fiber above $x^r \chi$). Since y_n^2 converges to p^r, this sequence converges to the point $(p^r, (1+p)^{k-1} - 1, 1 : 0)$

in $\widetilde{U}_r$. By Theorem 5.2, we see that the $\mathcal{L}$-invariant associated with this limit point is ∞. So the limit of the sequence V_n^* is $V(\delta_1, \delta_2, \infty)$ in the notation of [Colmez 2008].

Now $\delta_1\delta_2^{-1} = x^r\chi = \mu_{p^{r/2}} \cdot (\mu_{1/p^{r/2}}\chi^{1-k})^{-1}$ implies that there is a character δ such that $\delta_1 = \delta \cdot \mu_{p^{r/2}}$ and $\delta_2 = \delta \cdot \mu_{1/p^{r/2}}\chi^{1-k}$. Using [loc. cit., Proposition 4.4(i)], we see that $(\delta_1\delta_2)(p)$ is a unit. Hence $\delta(p)$ is a unit. Consequently, $V(\delta_1, \delta_2, \infty) \simeq V(\mu_{p^{r/2}}, \mu_{1/p^{r/2}}\chi^{1-k}, \infty) \otimes \delta$. Therefore the limit of the crystalline representations V_n^* is the crystalline representation $V_{k,\infty}^* \otimes \delta$. Taking duals, we see that the limit of the V_n is $V_{k,\infty} \otimes \delta^{-1}$. Note that the $\delta_{1,n} = \mu_{y_n} \cdot \mu_n^{-1}$ are unramified characters converging to $\delta_1 = \delta \cdot \mu_{p^{r/2}}$. This forces δ to be unramified and hence crystalline. Therefore $V_{k,\infty} \otimes \delta^{-1}$ is crystalline. The last representation has Hodge–Tate weights $(0, k-1)$. Untwisting by the fixed power of the cyclotomic character, we see that the Hodge–Tate weights of the limit representation of the original sequence again lie in $[a, b]$.

Acknowledgements

We would like to thank D. Benois, J. Bergdall, L. Berger, C. Breuil, P. Colmez, G. Chenevier, H. Schoutens and C. Park for useful conversations and the anonymous referees for numerous remarks which greatly helped improve the paper. We would also like to thank the participants of the rigid-analytic geometry seminar held at TIFR in 2019. Chitrao and Ghate acknowledge support of the Department of Atomic Energy under project number 12-R&D-TFR-5.01-0500. During this work, Yasuda was partially supported by JSPS KAKENHI grant numbers JP15H03610, JP21H00969, JP23K20782.

References

[Atiyah and Macdonald 1969] M. F. Atiyah and I. G. Macdonald, *Introduction to commutative algebra*, Addison-Wesley, Reading, MA, 1969. MR Zbl

[Benois 2010] D. Benois, "Infinitesimal deformations and the ℓ-invariant", *Doc. Math.* (2010), 5–31. MR Zbl

[Benois 2011] D. Benois, "A generalization of Greenberg's $\mathscr{L}$-invariant", *Amer. J. Math.* **133**:6 (2011), 1573–1632. MR Zbl

[Bergdall, Levin and Liu 2023] J. Bergdall, B. Levin, and T. Liu, "Reductions of 2-dimensional semistable representations with large $\mathcal{L}$-invariant", *J. Inst. Math. Jussieu* **22**:6 (2023), 2619–2644. MR Zbl

[Berger 2004] L. Berger, "Limites de représentations cristallines", *Compos. Math.* **140**:6 (2004), 1473–1498. MR Zbl

[Berger 2012] L. Berger, "Local constancy for the reduction mod p of 2-dimensional crystalline representations", *Bull. Lond. Math. Soc.* **44**:3 (2012), 451–459. MR Zbl

[Bertolini, Darmon, and Iovita 2010] M. Bertolini, H. Darmon, and A. Iovita, "Families of automorphic forms on definite quaternion algebras and Teitelbaum's conjecture", pp. 29–64 in *Représentations p-adiques de groupes p-adiques*, *III*: *méthodes globales et géométriques*, Astérisque **331**, Société Mathématique de France, Paris, 2010. MR Zbl

[Bhattacharya, Ghate, and Rozensztajn 2018] S. Bhattacharya, E. Ghate, and S. Rozensztajn, "Reductions of Galois representations of slope 1", *J. Algebra* **508** (2018), 98–156. MR Zbl

[Bosch, Güntzer, and Remmert 1984] S. Bosch, U. Güntzer, and R. Remmert, *Non-Archimedean analysis*: *a systematic approach to rigid analytic geometry*, Grundl. Math. Wissen. **261**, Springer, 1984. MR Zbl

[Breuil and Mézard 2002] C. Breuil and A. Mézard, "Multiplicités modulaires et représentations de $\mathrm{GL}_2(\mathbb{Z}_p)$ et de $\mathrm{Gal}(\overline{\mathbb{Q}}_p/\mathbb{Q}_p)$ en $l = p$", *Duke Math. J.* **115**:2 (2002), 205–310. MR Zbl

[Buzzard and Gee 2013] K. Buzzard and T. Gee, "Explicit reduction modulo p of certain 2-dimensional crystalline representations, II", *Bull. Lond. Math. Soc.* **45**:4 (2013), 779–788. MR Zbl

[Chenevier 2013] G. Chenevier, "Sur la densité des représentations cristallines de $\mathrm{Gal}(\overline{\mathbb{Q}}_p/\mathbb{Q}_p)$", *Math. Ann.* **355**:4 (2013), 1469–1525. MR Zbl

[Chenevier 2014] G. Chenevier, "The p-adic analytic space of pseudocharacters of a profinite group and pseudorepresentations over arbitrary rings", pp. 221–285 in *Automorphic forms and Galois representations*, *Vol.* 1, edited by F. Diamond et al., London Math. Soc. Lecture Note Ser. **414**, Cambridge Univ. Press, 2014. MR Zbl

[Chitrao 2025] A. Chitrao, "An Iwahori theoretic mod p local Langlands correspondence", *Canad. Math. Bull.* (online publication February 2025).

[Chitrao and Ghate 2023] A. Chitrao and E. Ghate, "Reductions of semi-stable representations using the Iwahori mod p local Langlands correspondence", preprint, 2023. arXiv 2311.03740

[Colmez 2005] P. Colmez, "Zéros supplémentaires de fonctions L p-adiques de formes modulaires", pp. 193–210 in *Algebra and number theory*, edited by R. Tandon, Hindustan Book Agency, Delhi, 2005. MR Zbl

[Colmez 2008] P. Colmez, "Représentations triangulines de dimension 2", pp. 213–258 in *Représentations p-adiques de groupes p-adiques*, *I*: *Représentations galoisiennes et* (ϕ, Γ)*-modules*, edited by L. Berger et al., Astérisque **319**, Société Mathématique de France, Paris, 2008. MR Zbl

[Colmez 2010] P. Colmez, "Invariants $\mathscr{L}$ et dérivées de valeurs propres de Frobenius", pp. 13–28 Astérisque **331**, Société Mathématique de France, Paris, 2010. MR Zbl

[Edixhoven 1992] B. Edixhoven, "The weight in Serre's conjectures on modular forms", *Invent. Math.* **109**:3 (1992), 563–594. MR Zbl

[Ghate 2021] E. Ghate, "A zig-zag conjecture and local constancy for Galois representations", pp. 249–268 in *Algebraic Number Theory and Related Topics* 2018, edited by T. Yamazaki and S. Yamamoto, RIMS Kôkyûroku Bessatsu **B86**, Res. Inst. Math. Sci. (RIMS), Kyoto, 2021. MR Zbl

[Ghate 2022] E. Ghate, "Zig-zag for Galois representations", preprint, 2022. arXiv 2211.12114

[Ghate and Rai 2025] E. Ghate and V. Rai, "Reductions of Galois representations of slope $\frac{3}{2}$", *Kyoto J. Math.* (online publication May 2025).

[Greenberg and Stevens 1993] R. Greenberg and G. Stevens, "p-adic L-functions and p-adic periods of modular forms", *Invent. Math.* **111**:2 (1993), 407–447. MR Zbl

[Guerberoff and Park 2019] L. Guerberoff and C. Park, "Semistable deformation rings in even Hodge–Tate weights", *Pacific J. Math.* **298**:2 (2019), 299–374. MR Zbl

[Kedlaya 2006] K. S. Kedlaya, "Finiteness of rigid cohomology with coefficients", *Duke Math. J.* **134**:1 (2006), 15–97. MR Zbl

[Kedlaya and Liu 2010] K. Kedlaya and R. Liu, "On families of ϕ, Γ-modules", *Algebra Number Theory* **4**:7 (2010), 943–967. MR Zbl

[Lazard 1962] M. Lazard, "Les zéros des fonctions analytiques d'une variable sur un corps valué complet", *Inst. Hautes Études Sci. Publ. Math.* 14 (1962), 47–75. MR Zbl

[Mazur 1994] B. Mazur, "On monodromy invariants occurring in global arithmetic, and Fontaine's theory", pp. 1–20 in *p-adic monodromy and the Birch and Swinnerton–Dyer conjecture* (Boston, MA, 1991), edited by B. Mazur and G. Stevens, Contemp. Math. **165**, Amer. Math. Soc., Providence, RI, 1994. MR Zbl

[Schoutens 1995] H. Schoutens, "Blowing up in rigid analytic geometry", *Bull. Belg. Math. Soc. Simon Stevin* **2**:4 (1995), 399–417. MR Zbl

[Stevens 2010] G. Stevens, "Coleman's $\mathscr{L}$-invariant and families of modular forms", pp. 1–12 in *Représentations p-adiques de groupes p-adiques*, *III*: *méthodes globales et géométriques*, Astérisque **331**, Société Mathématique de France, Paris, 2010. MR Zbl

Communicated by Samit Dasgupta
Received 2022-07-19 Revised 2024-02-29 Accepted 2024-04-29

anandchitrao@gmail.com *School of Mathematics, Tata Institute of Fundamental Research, Mumbai, India*

eghate@math.tifr.res.in *School of Mathematics, Tata Institute of Fundamental Research, Mumbai, India*

sese@math.sci.hokudai.ac.jp *Department of Mathematics, Hokkaido University, Hokkaido, Japan*

ALGEBRA AND NUMBER THEORY 19:6 (2025)
https://doi.org/10.2140/ant.2025.19.1099

msp

Geometry-of-numbers methods in the cusp

Arul Shankar, Artane Siad, Ashvin A. Swaminathan and Ila Varma

We develop new methods for counting integral orbits having bounded invariants that lie inside the cusps of fundamental domains for coregular representations. We illustrate these methods for a representation of cardinal interest in number theory, namely that of the split orthogonal group acting on the space of quadratic forms.

1. Introduction

A *coregular representation* (G, W) consists of a reductive algebraic group G, defined over $\mathbb{Z}$, and a finite-dimensional representation W of G, also defined over $\mathbb{Z}$, such that the ring of polynomial invariants for the action of the semisimplification of G on W is freely generated (say by the elements $p_1, \dots, p_k$). We then say that $U = \mathrm{Aff}^k$ is the *space of relative invariants*, and note that for any ring R we have the natural map $W(R) \to U(R)$ given by $w \mapsto (p_1(w), \dots, p_k(w))$[1] is freely and finitely generated over $\mathbb{Z}$. Many objects of interest in number theory and arithmetic geometry admit natural parametrizations in terms of integral orbits of coregular representations. Typically, these parametrizations impose the following three pieces of structure on the pair (G, W) and the associated invariant space U:

(1) an algebraic notion of *nondegeneracy* for the orbits of G on W;

(2) a natural notion of *height* on $U(\mathbb{R})$, which then lifts to a $G(\mathbb{R})$-invariant notion of height on $W(\mathbb{R})$;

(3) an arithmetic notion of *irreducibility* for the orbits of $G(\mathbb{Z})$ on $W(\mathbb{Z})$.

We note that *prehomogeneous* representations, i.e., representations whose rings of invariants are generated by a single element, are all coregular. In that case, the single generating invariant (usually termed the discriminant) is the height that is used.

MSC2020: primary 11R29, 11R45; secondary 11H55, 11E76.
Keywords: geometry-of-numbers, coregular representations, fundamental domain, cusps, irreducibility.

[1] More algebrogeometrically, the ring of relative invariants is $\mathbb{Z}[W]^G$, while the space of relative invariants is $U := W /\!\!/ G = \mathrm{Spec}(\mathbb{Z}[W]^G)$.

A landmark result of Borel and Harish-Chandra [19] implies that the number of nondegenerate $G(\mathbb{Z})$-orbits on $W(\mathbb{Z})$ having bounded height is finite. In light of this result, it is natural to ask the following fundamental question: *what are the asymptotics for the number of nondegenerate $G(\mathbb{Z})$-orbits on $W(\mathbb{Z})$ having bounded height*? Answering this question for just those integral orbits that are *irreducible* has proven to be a problem with significant applications in arithmetic statistics, the study of distributions of arithmetic objects. Indeed, counting integral irreducible orbits of coregular representations played a key role in the proofs of many breakthrough results, from determining asymptotics for S_n-number fields of small fixed degree ordered by discriminant (see [2; 4; 24; 29]), to computing average sizes of p-class groups in certain families of number fields for small primes p (see [2; 14; 15; 24; 31; 46; 47; 49]), to calculating average n-Selmer ranks of elliptic curves and hyperelliptic Jacobians for small n (see [6; 7; 8; 9; 10; 15; 32; 33; 40; 43]), and many more.

More recently, counting *reducible* integral orbits of coregular representations has emerged as a problem of significant interest in its own right, yielding applications toward determining the sizes of the 3-torsion in the class groups of quadratic orders [11; 45], counting octic D_4-number fields ordered by Artin conductor [1] and by discriminant [42], studying families of elliptic curves ordered by conductor [44], and carrying out squarefree sieves on families of polynomials [16] and binary n-ic forms [17]. However, in all of these applications, the counts of reducible orbits were obtained using ad hoc methods. Furthermore, the latter three results cited above do not actually prove asymptotics, and merely obtain upper bounds on the number of reducible orbits of bounded height. This is in stark contrast with the case of irreducible orbits, for which systematic methods have been developed to obtain precise asymptotics with power-saving error terms.

The purpose of this article is to devise new systematic techniques for counting reducible orbits. We illustrate our techniques for a representation that features prominently in the literature on arithmetic statistics, namely the action of the orthogonal group of the split n-ary quadratic form on the space of quadratic forms in n variables, where $n \geq 3$ is an arbitrary fixed integer. In particular, we provide a complete answer to the fundamental question stated above for each representation in this infinite family indexed by n; see Section 1.3 for precise statements of our main theorems. Before proving these general theorems, we provide the reader with a gentle introduction to our new method by illustrating how it applies in the context of a low-dimensional example, namely the action of PGL_2 on the space $\mathrm{Sym}_4(2)$ of binary quartic forms; see Section 2.

1.1. ***Background on orbit-counting, and relation to earlier work.*** In this section, we summarize the orbit-counting methods that feature in the literature on arithmetic statistics leading up to the present article, and we describe the context for our new methods of counting reducible orbits.

The critical dichotomy. Let (G, W) be a coregular representation with space of relative invariants U, and suppose that a family of arithmetic objects can be parametrized in terms of certain $G(\mathbb{Z})$-orbits on $W(\mathbb{Z})$. We note that a natural source of examples of such coregular representations are *prehomogeneous* representations, i.e., the ring of relative invariants is generated by a single element, usually called the *discriminant*.

Further suppose that the space $U(\mathbb{R})$ admits a natural notion of height, and define the height of (the $G(\mathbb{R})$-orbit of) an element $w \in W(\mathbb{R})$ to be the height of the image of w under the natural map $W(\mathbb{R}) \to U(\mathbb{R})$. In the prehomogeneous case, this height is usually taken to be the absolute value of the discriminant. Under the above assumptions, the problem of counting arithmetic objects in the family with bounded height can be translated into the problem of counting lattice points of bounded height in a fundamental domain $\mathcal{D}$ for the action of $G(\mathbb{Z})$ on $W(\mathbb{R})$. This latter problem is complicated by the fact that the fundamental domain $\mathcal{D}$ is usually not compact, even for the subset $\mathcal{D}_X$ of points in $\mathcal{D}$ with height less than X. Indeed, the height-bounded fundamental domain $\mathcal{D}_X$ typically consists of a bounded region known as the *main body*, along with one or more *cusps*, which are long tentacle-like regions going off to infinity.

Surprisingly, in most of the coregular representations considered in the literature thus far, the following dichotomy holds:

(1) A proportion of 100% of irreducible orbits lie in the main body; and a proportion of 100% of points in the main body are irreducible.

(2) A proportion of 100% of reducible orbits lie in the cusp; and a proportion of 100% of points in the cusp are reducible.

If properties (1) and (2) above hold, then the count of points in the main body of $\mathcal{D}_X$ gives an asymptotic for the number of irreducible $G(\mathbb{Z})$-orbits on $W(\mathbb{Z})$ having height less than X. This main body count can be determined in a systematic way: using geometry-of-numbers methods, the count can be expressed in terms of the volume of the main body, and this volume can then be computed by performing a suitable change-of-variables. The difficulty in counting the reducible orbits, which predominate in the cusp(s) of $\mathcal{D}_X$, is that geometry-of-numbers methods do not directly apply in such regions. In fact, the asymptotic count of reducible orbits is *not* given in terms of the volume of the corresponding cuspidal regions. In the examples treated in this paper, we find that the volume of the cuspidal region is an underestimate, and that the actual answer is given in terms of a certain weighted volume.

Historical context. The development and use of the orbit-counting strategy summarized above goes back centuries. Mertens [34] and Siegel [48] studied the action of GL_2 on the space $\mathrm{Sym}_2(2)$ of binary quadratic forms, which has a unique polynomial invariant, namely the discriminant. They developed geometry-of-numbers methods to count the number of $\mathrm{GL}_2(\mathbb{Z})$-orbits on $\mathrm{Sym}_2 \mathbb{Z}^2$ with bounded discriminant, resolving conjectures of Gauss on the average sizes of class numbers of quadratic orders. In the series of papers [22; 23], Davenport considered the action of GL_2 on the space $(\mathrm{Sym}^3(2))^*$ of binary cubic forms, which also has the discriminant as its unique polynomial invariant. Using the strategy described above, he obtained asymptotics for the number of irreducible $\mathrm{GL}_2(\mathbb{Z})$-orbits on $(\mathrm{Sym}^3 \mathbb{Z}^2)^*$ having bounded discriminant, and in collaboration with Heilbronn, he combined these asymptotics with results from class field theory and sieve methods to determine the density of discriminants of cubic fields [24].

The main obstruction to generalizing the work of Mertens, Siegel, and Davenport–Heilbronn to other representations was proving that the number of irreducible points in the cusps is negligible. A pioneering advance was made by Bhargava (see [4]), who introduced an "averaging" method that has the effect

of thickening the cusps, making them a bit more amenable to doing geometry-of-numbers. Bhargava's averaging method, which applies to general coregular representations, gives a systematic way to bound the number of points in cuspidal regions and thus opened the door to study the distributions of a wide variety of arithmetic objects. However, the averaging method has, until the present paper, only been used to obtain upper bounds in the cusp and new methods are needed to obtain precise asymptotics.

1.2. ***Notation and setup.*** We first study a concrete representation in Section 2, before proceeding to our main results on families of representations. The concrete representation we consider is that of PGL_2 acting on binary quartic forms. The notation in this case is quite manageable, and largely follows that of [9]. In this section, we introduce the other (family of) representations studied in this paper, and we set up the notation necessary to state our main results. Let $\mathcal{A}$ be the antidiagonal $n \times n$ matrix with all antidiagonal entries equal to 1; if viewed as a quadratic form, we note that $\mathcal{A}$ is split (i.e., it has a maximal isotropic space defined over $\mathbb{Q}$) and unimodular.

The representation. We first define the group G, which is slightly different for n odd and n even:

- When n is odd, we take $G := \mathrm{SO}_{\mathcal{A}}$ to be the split special orthogonal group scheme over $\mathbb{Z}$ corresponding to $\mathcal{A}$. That is, we have $G(R) = \{g \in \mathrm{SL}_n(R) : g^t \mathcal{A}\, g = \mathcal{A}\}$ for any $\mathbb{Z}$-algebra R.
- When n is even, we take $G := \mathrm{O}_{\mathcal{A}}/\mu_2$ to be the split projective orthogonal group scheme over $\mathbb{Z}$ corresponding to $\mathcal{A}$. That is, we take G to be the cokernel of the inclusion $\mu_2 \hookrightarrow \mathrm{O}_{\mathcal{A}}$ of group schemes over $\mathbb{Z}$, where $\mathrm{O}_{\mathcal{A}} := \{g \in \mathrm{GL}_n(R) : g^t \mathcal{A}\, g = \mathcal{A}\}$ for any $\mathbb{Z}$-algebra R.

We now define the representation W of G. Let W denote the affine $\mathbb{Z}$-scheme whose R-points consist of the set of $n \times n$ symmetric matrices with entries in R (i.e., classically integral quadratic forms over R). Then W has a natural structure as a G-representation, where the (left) action is given by $g \cdot B = (g^{-1})^t B g^{-1}$ for $g \in G$ and $B \in W$. We note that W can also be interpreted as the space of self-adjoint operators for $\mathcal{A}$ over R by identifying a self-adjoint operator T with $B = -\mathcal{A}T \in W$.

The orbits of the representation of G on W have been studied extensively in the literature. For example, Bhargava and Gross [6] (resp. Shankar and Wang [43]) obtained asymptotics for the number of *irreducible* orbits of $G(\mathbb{Z})$ on $W(\mathbb{Z})$ having bounded height when n is odd (resp., when n is even). These asymptotics have yielded a striking array of applications: they were utilized by the aforementioned authors to bound the average sizes of the 2-Selmer groups of monic hyperelliptic Jacobians of any given dimension, by Swaminathan to prove that most odd-degree binary forms fail to primitively represent a square [50], and by Siad [46; 47] to bound the average size of the 2-torsion subgroup of the class groups of monogenic fields of any given degree. Furthermore, by proving an upper bound on the *reducible* orbits of this representation, Bhargava, Shankar, and Wang [16] determined the probability that a monic integral polynomial has squarefree discriminant.

The invariants. Let U denote the affine $\mathbb{Z}$-scheme whose R-points consist of monic degree-n polynomials with coefficients in R. For any element $B \in W(R)$, the monic degree-n polynomial

$$\mathrm{inv}(B) := (-1)^{\lfloor \frac{n}{2} \rfloor} \det(x\mathcal{A} + B) \in U(R) \tag{1}$$

is invariant under the action of $G(R)$; in fact, by [20, Section 8.3, part (VI) of Section 13.2], its coefficients freely generate the ring of polynomial invariants for the action of G on W. Given $f \in U(R)$, we write

$$\mathrm{inv}^{-1}(f) := \{B \in W(R) : \mathrm{inv}(B) = f\}.$$

The notion of nondegeneracy. Let R be an integral domain. Then we say that a monic polynomial with coefficients in R is *nondegenerate* if it has nonzero discriminant. We say that an element $B \in W(R)$ is *nondegenerate* if the monic polynomial $\mathrm{inv}(B)$ has nonzero discriminant.

When $R = \mathbb{Z}$ or $\mathbb{R}$, it is convenient to partition the set of nondegenerate elements in $U(R)$ according to the number of real roots, and to lift this partition to $W(R)$. To this end, let $0 < r \le n$ be odd if n is odd, and let $0 \le r \le n$ be even if n is even. For $R = \mathbb{Z}$ or $\mathbb{R}$, we define

$$U(R)^{(r)} := \{f \in U(R) : f \text{ is nondegenerate and exactly } r \text{ real roots}\},$$
$$W(R)^{(r)} := \{B \in W(R) : \mathrm{inv}(B) \in U(R)^{(r)}\}.$$

The height. We order elements of $U(\mathbb{R})$ and $W(\mathbb{R})$ by *height*. Given a polynomial

$$f(x) = x^n + \sum_{i=1}^{n} f_i x^{n-i} \in U(\mathbb{R}),$$

we define its *height* $\mathrm{H}(f)$ by

$$\mathrm{H}(f) := \max_{1 \le i \le n} \{|f_i|^{1/i}\},$$

and given $B \in W(\mathbb{R})$, we define its height by $\mathrm{H}(B) := \mathrm{H}(\mathrm{inv}(B))$. For any $X > 0$, we write

$$N^{(r)}(X) := \#\{f \in U(\mathbb{Z})^{(r)} : \mathrm{H}(f) < X\}.$$

Notice that we have

$$N^{(r)}(X) \sim \mathcal{V}^{(r)}(X) := \mathrm{Vol}(\{f \in U(\mathbb{R})^{(r)} : \mathrm{H}(f) < X\}) \asymp X^{\dim W} = X^{\frac{n^2+n}{2}},$$

where the volume is computed with respect to the Euclidean measure on $U(\mathbb{R})$, normalized so that $U(\mathbb{Z})$ has covolume 1.

The notion of reducibility. When R is a principal ideal domain with field of fractions K, we can partition the nondegenerate elements in $W(R)$ into two natural $G(R)$-invariant subsets. We call (the $G(R)$-orbit of) a nondegenerate element $B \in W(R)$ *reducible* if:

- $\mathcal{A}$ and B, viewed as quadratic forms, share a maximal isotropic subspace over K if n is odd, or
- B has an isotropic subspace of dimension $\frac{n-2}{2}$ over K contained within a maximal isotropic subspace for $\mathcal{A}$ over K if n is even.

We say that $B \in W(R)$ is *irreducible* if it is nondegenerate and not reducible. There are a few good arithmetic reasons to think of such elements as reducible. For example, pairs $(\mathcal{A}, B)$ correspond to triples (R', I, δ), where R' is a rank n ring over R, I is an ideal in R', and $\delta \in L^\times / L^{\times 2}$, where $L = R' \otimes K$, such that $I^2 \subset (\delta)$ and $N(I)^2 = N(\delta)$. Then it was proven in [46; 47] that $(\mathcal{A}, B)$ is reducible if and only if $\delta \equiv 1$.

As another example, pairs $(\mathcal{A}, B)$ also correspond to classes $\sigma \in H^1(K, J[2])$, where J is the Jacobian of the monic hyperelliptic curve $y^2 = \pm\det(\mathcal{A} + B)$, and $(\mathcal{A}, B)$ is reducible if and only if σ is trivial.

Write the coordinates of W as $[b_{ij}]_{1\le i\le j\le n}$. Let W_0 be the subscheme of W obtained by setting $b_{ij} = 0$ for all (i, j) with $i + j < n$. Observe that, if R is a principal ideal domain, then every element of $W_0(R)$ is reducible; for this reason, we call W_0 the *reducible hyperplane*. The reducible hyperplane is sent to itself under the action of the lower-triangular subgroup $\mathcal{P} \subset G$, and so we obtain a well-defined representation of $\mathcal{P}$ on W_0. In what follows, for a $\mathbb{Z}$-algebra R and any subset $S \subset W(R)$, we sometimes write $S_0 \subset S$ to denote the subset $S \cap W_0(R)$.

1.3. ***Statements of main theorems.*** Having set up the notation, we are now in position to state our first main result. Define the constants C_n^{fin} and $C_{n,r}^{\mathrm{inf}}$ as

$$C_n^{\mathrm{fin}} := \begin{cases} \prod_{i=1}^{\lfloor \frac{n}{2} \rfloor} \zeta(2i) & \text{if } 2 \nmid n, \\ \zeta\left(\frac{n}{2}\right) \prod_{i=1}^{\frac{n-2}{2}} \zeta(2i) & \text{if } 2 \mid n, \end{cases} \qquad C_{n,r}^{\mathrm{inf}} := \begin{cases} \mathcal{V}^{(r)}(1) & \text{if } 2 \nmid n, \\ 2^{-\frac{n}{2}} \mathcal{V}^{(r)}(1) & \text{if } 2 \mid n. \end{cases}$$

Then we have the following result, which gives the total count of reducible $G(\mathbb{Z})$-orbits on $W(\mathbb{Z})^{(r)}$:

Theorem 1. *The number of reducible $G(\mathbb{Z})$-orbits on $W(\mathbb{Z})^{(r)}$ having height up to X is given by*

$$C_n^{\mathrm{fin}} \cdot (C_{n,r}^{\mathrm{inf}} \cdot X^{\frac{n^2+n}{2}}) + o(X^{\frac{n^2+n}{2}}).$$

Remark. The factor $C_{n,r}^{\mathrm{inf}} \cdot X^{(n^2+n)/2}$ occurring in Theorem 1 is an asymptotic for the number of invariant polynomials that arise from orbits of height up to X. Notice that $C_{n,r}^{\mathrm{inf}} \cdot X^{(n^2+n)/2} \sim N^{(r)}(X)$ when n is odd and that $C_{n,r}^{\mathrm{inf}} \cdot X^{(n^2+n)/2} \sim 2^{-n/2} N^{(r)}(X)$ when n is even. The extra factor of $2^{-n/2}$ occurs when n is even because, for any $B \in W(\mathbb{Z})$, the x^i-coefficient of $\mathrm{inv}(B)$ is divisible by 2 for every odd number $i \in \{1, 3, \ldots, n-1\}$ (see [47, Theorem 80]).

As mentioned above, the corresponding asymptotics for the irreducible $G(\mathbb{Z})$-orbits on $W(\mathbb{Z})^{(r)}$ have been obtained previously in [6, Theorem 10.1 and equation (10.27)] when n is odd and in [43, Theorem 20 and equation (38)] when n is even. The main term exponents for the irreducible and reducible cases are the same, and remarkably, the asymptotics for both cases have the same leading constant C_n^{fin}, up to a rational factor! Even more surprisingly, the manner in which C_n^{fin} arises is completely different for the irreducible and reducible counts. In the irreducible case, this constant arises from the fundamental volume of the group G—i.e., the volume with respect to a suitably normalized Haar measure of $G(\mathbb{Z})\backslash G(\mathbb{R})$. However, in the reducible case considered in this paper, it arises from a summation of volumes of lower-dimensional slices of the cuspidal regions of a fundamental domain for $G(\mathbb{Z})$ on $W(\mathbb{R})$. These slices have negligible volume unless they are very high up in the cusp. Thus, they do not detect most of the main body volume, and this appears to be a genuinely different way of obtaining the constant C_n^{fin}.

In light of the above discussion, we are led to pose the following natural question, which we have answered in the affirmative for the action of G on W:

Question 2. *Let (G, W) be a coregular representation with natural notions of nondegeneracy, height, and reducibility. Let $N^{\mathrm{irr}}(X)$ (resp., $N^{\mathrm{red}}(X)$) denote the number of irreducible (resp., reducible) $G(\mathbb{Z})$-orbits on $W(\mathbb{Z})$ having height less than X.*

(a) *Is it true that as X tends to infinity we have $N^{\mathrm{irr}}(X) \asymp N^{\mathrm{red}}(X)$?*

(b) *If so, then is it true that $\lim_{X\to\infty}(N^{\mathrm{irr}}(X)/N^{\mathrm{red}}(X))$ is a rational constant?*

We note that the answers to both parts of Question 2 are also "yes" for the action of GL_2 on $\mathrm{Sym}_3(2)$ by work of Shintani [45, Chapter 2, Section 7, Remark 2] and Bhargava and Varma [11, Section 4.1.2]. It follows by combining our results in Section 2 with the irreducible count obtained by Bhargava and Shankar in their paper on binary quartic forms [9] that the answers are also "yes" for the action of GL_2 on $\mathrm{Sym}_4(2)$. On the other hand, the answer to the first part of Question 2 is "no" for the action of SL_n on $2 \otimes \mathrm{Sym}^2(n)$, where $n \geq 4$ is an arbitrary even integer; indeed, it was shown in work of Bhargava [5] that $N^{\mathrm{red}}(X) = o(N^{\mathrm{irr}}(X))$ for this representation. The situation is more complicated for the action of $\mathrm{GL}_2 \times \mathrm{SL}_3$ on $2 \otimes \mathrm{Sym}_2(3)$. Bhargava proved in [2, Theorem 7] that $N^{\mathrm{irr}}(X) \asymp X$, and it is well known that $N^{\mathrm{red}}(X) \asymp X \log X$. However, the reducible orbits of this representation can be partitioned into several natural subsets, and as was shown by Swaminathan [51] using the methods developed in the present paper, the answers to both parts of Question 2 are "yes" if we restrict to one of these subsets.

Our next main results concern families defined by certain infinite sets of congruence conditions. We call a subset $S \subset W(\mathbb{Z})$ a *big family* if $S = W(\mathbb{Z})^{(r)} \cap \bigcap_p S_p$, where the sets $S_p \subset W(\mathbb{Z}_p)$ satisfy the following properties:

(1) S_p is $G(\mathbb{Z}_p)$-invariant and is the preimage under reduction modulo p^j of a nonempty subset of $W(\mathbb{Z}/p^j\mathbb{Z})$ for some $j > 0$ for each p; and

(2) S_p contains all ($G(\mathbb{Z}_p)$-orbits of) elements $B \in W_0(\mathbb{Z}_p)$ such that, for all $p \gg 1$, we have that $b_{i(n-i)}(B)$ is a p-adic unit for some i.

Note that a big family S is necessarily $G(\mathbb{Z})$-invariant.

Our first result on reducible orbits in big families is stated in terms of a polynomial function λ on the reducible hyperplane W_0, defined explicitly by

$$\lambda(B) := \begin{cases} \prod_{i=1}^{\lfloor \frac{n}{2} \rfloor} b_{i(n-i)}(B)^{2i-1} & \text{if } 2 \nmid n, \\ b_{\frac{n}{2}\frac{n}{2}}(B)^{\frac{n-2}{2}} \prod_{i=1}^{\frac{n-2}{2}} b_{i(n-i)}(B)^{2i-1} & \text{if } 2 \mid n. \end{cases} \tag{2}$$

Given this definition of λ, we have the following asymptotic formula:

Theorem 3. *Let S be a big family. Then the number of reducible $G(\mathbb{Z})$-orbits on S of height up to X is given by*

$$C_{n,r}^{\mathrm{inf}} \cdot \left(\prod_p \left(1 - \frac{1}{p}\right)^{-\lfloor \frac{n}{2} \rfloor} \int_{B \in (S_p)_0} |\lambda(B)|_p \, dB \right) \cdot X^{\frac{n^2+n}{2}} + o(X^{\frac{n^2+n}{2}}),$$

where dB denotes the Euclidean measure on $W_0(\mathbb{Z}_p)$, normalized so that $W_0(\mathbb{Z}_p)$ has volume 1, and where $|-|_p$ denotes the usual p-adic absolute value.

Our second result on reducible orbits in big families, which is equivalent to Theorem 3, is stated in terms of a product of local orbit counts for the action of $\mathcal{P}$ on W_0:

Theorem 4. *Let S be a big family. Then the number of reducible $G(\mathbb{Z})$-orbits on S of height up to X is given by*

$$\left(\prod_p \int_{f \in U(\mathbb{Z}_p)} \#\left(\frac{\mathrm{inv}^{-1}(f) \cap (S_p)_0}{\mathcal{P}(\mathbb{Z}_p)}\right) df\right) \cdot N^{(r)}(X) + o(X^{\frac{n^2+n}{2}}), \tag{3}$$

where df denotes the Euclidean measure on $U(\mathbb{Z}_p)$, normalized so that $U(\mathbb{Z}_p)$ has volume 1.

The local product in (3) looks similar to many other mass formulas in related works with one major difference: the group in question, $\mathcal{P}$, is not reductive. To prove Theorem 4, it is therefore necessary for us to get some control over orbits of this nonreductive group, which we do in, e.g., Section 6.3.

1.4. ***Methods of proof.*** To prove our main results, we introduce two new methods of determining asymptotics for reducible orbits, and we describe them both as follows.

Method I. Our first method proceeds by directly counting points in the cusp(s) of an "averaged" fundamental domain $\mathcal{D}$ for the action of $G(\mathbb{Z})$ on $W(\mathbb{R})$. This suffices to get the count of reducible orbits because, by the results in [6; 43], properties (1) and (2) in Section 1.1 are satisfied for the action of G on W. This method requires us to construct fundamental domains $\mathcal{D}$ for the action of $G(\mathbb{Z})$ on $W(\mathbb{R})$, which in turn requires us to construct a fundamental domain $\mathcal{F}$ for the action of $G(\mathbb{Z})$ on $G(\mathbb{R})$. Unlike in most previous situations, it is simply not enough for us to invoke the work of Borel and Harish-Chandra [18; 19], who constructed fundamental domains for general semisimple groups. Indeed, our argument relies on $\mathcal{F}$ being *box-shaped at infinity*, meaning that $\mathcal{F}$ looks like a Siegel domain in a neighborhood of the cusp. We prove the existence of such fundamental domains for our groups G.

The region $\mathcal{D}$ is too skewed for a direct geometry-of-numbers argument to give anything better than an upper bound for the number of points it contains. To resolve this issue, we cut up the region $\mathcal{D}$ into a countable collection of nicer-looking slices. Within each slice, we prove that the count of the integral points is asymptotic to the volume of the slice. Summing up over all slices yields the desired total asymptotic. Our slicing method constitutes the first higher-dimensional generalization of an argument developed in [13], which treated the simpler case of the cusps arising from the group GL_2.

Summing up the volumes of the slices gives us the desired asymptotic in terms of weighted volumes of certain sets in the reducible hyperplane $W_0(\mathbb{R}) \subset W(\mathbb{R})$, where the volumes are computed with respect to the weight λ defined in (2). We evaluate these weighted volumes by proving a Jacobian change-of-variables formula that transforms the measure λ on W_0 into the product of the Euclidean measure on $U(\mathbb{R})$ with the Haar measure on the lower-triangular subgroup $\mathcal{P}(\mathbb{R}) \subset G(\mathbb{R})$. Such change-of-variable results have previously been proven when the group under consideration is unimodular (see, e.g., [9, Section 3.4]), but the fact that the group $\mathcal{P}$ fails to be unimodular presents significant new challenges.

Method I, as applied to the representation of G on W, requires very complicated indexing and notation. To make Method I more readily comprehensible for the reader, we begin in Section 2 by illustrating the

method in the case of PGL_2 acting on $\mathrm{Sym}_4(2)$, before proceeding with the parallel, but much more complicated, case of G acting on W starting in Section 3. We note that Section 2 can be read more or less independently of the rest of the paper.

Method II. Our second method proceeds by means of the following four steps. First, we claim that the asymptotics for the number of reducible $G(\mathbb{Z})$-orbits on $W(\mathbb{Z})$ are the same as the asymptotics for the number of $\mathcal{P}(\mathbb{Z})$-orbits on $W_0(\mathbb{Z})$. This claim is an immediate corollary of the following two facts:

(a) If $B_1, B_2 \in W_0(\mathbb{Z})$ are equivalent under $G(\mathbb{Z})$ but not under $\mathcal{P}(\mathbb{Z})$, then B_1 and B_2 have nontrivial stabilizer in $G(\mathbb{Q})$, but as we establish in Proposition 32, all but negligibly many $G(\mathbb{Z})$-equivalence classes on $W_0(\mathbb{Z})$ have trivial stabilizer in $G(\mathbb{Q})$.

(b) As shown in [6, Proposition 10.7] and [43, Proposition 23], all but negligibly many reducible $G(\mathbb{Z})$-orbits on $W(\mathbb{Z})$ have a representative lying on the reducible hyperplane $W_0(\mathbb{Z})$.

The second step is to develop the reduction theory for the action of $\mathcal{P}$ on W_0. Specifically, we prove that, for any field K, the group $\mathcal{P}(K)$ acts *simply transitively* on the set of elements in $W_0(K)$ having any fixed nondegenerate invariant polynomial. Using the fact that the group $\mathcal{P}$ has class number 1 over $\mathbb{Q}$, we then deduce that the orbits of $\mathcal{P}(\mathbb{Z})$ on $W_0(\mathbb{Z})$ satisfy the following strong local-to-global principle:

Theorem 5. *Let $f \in U(\mathbb{Z})$ be a nondegenerate monic degree-n polynomial, and when n is even, suppose that the x^i-coefficient of f is divisible by 2 for each odd i. For each prime p, choose $B_p \in W_0(\mathbb{Z}_p)$ such that* $\mathrm{inv}(B) = f$. *Then there exists $B \in W_0(\mathbb{Z})$, unique up to the action of $\mathcal{P}(\mathbb{Z})$, such that B is $\mathcal{P}(\mathbb{Z}_p)$-equivalent to B_p for each prime p.*

In fact, we prove a more general version of Theorem 5, namely Theorem 34, applying to representations of groups having class number 1 over $\mathbb{Q}$.

By Theorem 5, the number of $\mathcal{P}(\mathbb{Z})$-orbits on $W_0(\mathbb{Z})$ with nondegenerate invariant polynomial f is simply the product over all primes p of the number of $\mathcal{P}(\mathbb{Z}_p)$-orbits on $W_0(\mathbb{Z}_p)$ lying above f. The third step is to sum this local product formula over all invariant polynomials of bounded height. A key ingredient for evaluating the sum is to verify that there are not too many orbits with large value of λ. This was verified in [16], where an upper bound of roughly the correct order of magnitude was obtained for the number of $\mathcal{P}(\mathbb{Z})$-orbits on $W_0(\mathbb{Z})$ with large value of λ.[2] We thus arrive at the surprising conclusion that, at least for the representation of G on W, an upper bound on the number of reducible orbits can be indirectly used to deduce a precise asymptotic!

The resulting asymptotic for the total count of reducible orbits is now expressed in terms of a product of local orbit counts, as in Theorem 4. The fourth and final step is to evaluate this product. We do this by using the previously mentioned Jacobian change-of-variables formula in reverse! This expresses the local orbit count at a prime p as a certain p-adic integral, thus yielding Theorem 3; evaluating each of these integrals and multiplying them all together yields the total asymptotic in Theorem 1.

[2] An even stronger upper bound for the number of orbits with large value of λ can be obtained using Method I, by simply summing over those slices with large value of λ; see Theorem 33 (to follow).

Remark. For the coregular representations considered in this paper, both of the above methods are sufficient to obtain asymptotics for the number of reducible elements. Method I gives a direct proof of Theorem 3, while Method II gives a direct proof of Theorem 4. Moreover, the main terms in these two theorems can be related to each other using a Jacobian change-of-variables. However, for certain representations such as those considered in [51; 36], Method II can be directly applied while applying Method I seems more complicated. Moreover, for certain representations such as those considered in [1], Method II is inapplicable, while Method I can be used.

1.5. ***Other applications of our methods.*** We expect that both of the methods that we introduce in this paper can be used to count reducible orbits for other representations.

Our methods have already been used to derive arithmetic applications beyond the results of this paper. Swaminathan [51] counted reducible orbits for the action of SL_n on $2 \otimes \mathrm{Sym}^2(n)$, where n is odd, and used this to prove asymptotics for counts of 2-torsion in the ideal class groups of cubic orders in the case $n = 3$. (A similar application could be pursued for counting quintic rings, which also correspond to integral orbits of a coregular representation via a parametrization of Bhargava [3].) Also, Oller [36] counted reducible orbits in all the Vinberg representations in Thorne's thesis [52], and also carried out squarefree sieves in all these cases.

One possible line of inquiry is to study the action of GL_2 on the space of binary n-ic forms for $n \geq 5$. We carry out the case $n = 4$ in Section 2; the $n = 3$ case is similar and would give a simpler proof of the results of Shintani and Bhargava–Varma mentioned previously. The spaces are not coregular for $n \geq 5$, but when integral orbits are ordered by Julia invariant, Bhargava and Yang [12] determined asymptotics for the number of irreducible orbits. We expect the methods introduced in this paper to have applications towards counting the reducible orbits, ordered by Julia invariant, for these spaces. Other arithmetically interesting families of coregular representations for which the question of counting reducible orbits remains open may be found in the thesis of Ho [30].

2. Counting reducible integral orbits on binary quartic forms

In this section, we determine asymptotics for the number of reducible orbits of bounded height for the action of PGL_2 on the space of integral binary quartic forms. Our purpose is to illustrate Method I (see Section 1.4) in the context of a low-dimensional example, with the view of making the higher-dimensional application treated in Sections 3–6 more readily comprehensible.

This section can be read more or less independently of the rest of the paper. We remark that some of the notation used within this section is recycled in subsequent sections to denote different but analogous objects. As none of the objects introduced in this section are used in subsequent sections, we do not expect this to cause ambiguity.

Setup. Let V denote the affine $\mathbb{Z}$-scheme whose R-points consist of binary quartic forms with coefficients in R; i.e., we have

$$V(R) := \{f(x, y) = ax^4 + bx^3y + cx^2y^2 + dxy^3 + ey^4 : a, b, c, d, e \in R\}.$$

The group PGL_2 acts on V via $(g \cdot f)(x, y) := (\det g)^{-2} \times f((x, y) \cdot g)$. The ring of invariants for the action of PGL_2 on V is freely generated by two invariants, denoted by I and J. For the form $f(x, y)$ written as above, these invariants are given explicitly by

$$I(f) = 12ae - 3bd + c^2, \quad J(f) = 72ace + 9bcd - 27ad^2 - 27eb^2 - 2c^3.$$

For convenience, we define $\mathrm{inv} : V \to \mathbb{A}^2$ to be the map that sends f to $(I(f), J(f))$. The image of this map over $\mathbb{Z}$ is not all of $\mathbb{A}^2(\mathbb{Z})$, but is defined by congruence conditions modulo 27; see [9, Theorem 1.7].

We define a $\mathrm{PGL}_2(\mathbb{R})$-invariant *height* H on $V(\mathbb{R})$ via $H(f) := \max\{|I(f)|^3, J(f)^2/4\}$. We say that a binary quartic form f is *nondegenerate* if its *discriminant* $\Delta(f) := (4I(f)^3 - J(f)^2)/27$ is nonzero. More generally, we say that a pair (I, J) is nondegenerate if the quantity $\Delta(I, J) := (4I^3 - J^2)/27$ is nonzero (so a binary quartic form is nondegenerate if and only if its invariants are nondegenerate). For $i \in \{0, 1, 2\}$ and $R = \mathbb{Z}$ or $\mathbb{R}$, we let $V(R)^{(i)}$ denote the set of nondegenerate elements in $V(R)$ having i pairs of complex conjugate roots and $4 - 2i$ real roots in $\mathbb{P}^1(\mathbb{C})$.

A nondegenerate binary quartic form $f \in V(R)$ is said to be *reducible* if it factors over R. It follows from [9, Lemma 2.3] that the number of reducible orbits on $V(\mathbb{Z})$ having height bounded by X and factoring into the product of two irreducible quadratic forms is $O(X^{2/3+\epsilon})$, and thus negligible. That is, 100% of reducible orbits have at least one rational (and thus at least one real) linear factor. Therefore, for the purposes of this section, it suffices to restrict our attention to counting orbits of binary quartic forms that possess a rational linear factor, which amounts to counting reducible integral orbits in the sets $V(\mathbb{R})^{(0)}$ and $V(\mathbb{R})^{(1)}$.

Let $V_0(R) \subset V(R)$ be the "reducible hyperplane" consisting of those forms f with $a = 0$ (i.e., those forms f that are divisible by y). The reducible hyperplane is sent to itself under the action of the lower-triangular subgroup $P \subset \mathrm{PGL}_2$,[3] and so we obtain a well-defined representation of P on V_0. In what follows, for a $\mathbb{Z}$-algebra R and any subset $S \subset V(R)$, we sometimes write $S_0 \subset S$ to denote the subset $S \cap V_0(R)$.

Main results. Given the above setup, the main result of this section is as follows, by analogy with Theorem 1:

Theorem 6. *The number of reducible* $\mathrm{PGL}_2(\mathbb{Z})$*-orbits on* $V(\mathbb{Z})^{(i)}$ *having height less than* X *is*

$$\zeta(2) \cdot \left(\tfrac{8+24i}{135} \cdot X^{\frac{5}{6}}\right) + O_\epsilon(X^{\frac{3}{4}+\epsilon}).$$

Remark. The factor $\frac{8+24i}{135} \cdot X^{5/6}$ occurring in Theorem 6 is an asymptotic for the number of pairs (I, J) that arise as invariants of orbits of height up to X (see [9, Proposition 2.10]). The factor of $\frac{8+24i}{135}$ comprises two parts: the first is a factor of $\frac{8+24i}{5}$, which is the volume of the space of invariants of height at most 1 in $V(\mathbb{R})^{(i)}$, and the second is a factor of $\frac{1}{27}$, which occurs because not every pair $(I, J) \in R^2$ arises as the set of invariants of a binary quartic form in $V(R)$.

[3] More precisely, for a ring R, we define $P(R)$ to be the image of the subgroup of lower triangular matrices in $\mathrm{GL}_2(R)$ under the map $\mathrm{GL}_2(R) \to \mathrm{PGL}_2(R)$.

Next, we consider subsets of $V(\mathbb{Z})$ cut out by certain (possibly) infinite sets of congruence conditions. We call $S \subset V(\mathbb{Z})$ a *big family* if $S = V(\mathbb{Z})^{(i)} \cap \bigcap_p S_p$, for $i \in \{0, 1\}$, where the sets $S_p \subset V(\mathbb{Z}_p)$ satisfy the following properties:

(1) S_p is $\mathrm{PGL}_2(\mathbb{Z}_p)$-invariant and is the preimage under reduction modulo p^j of a nonempty subset of $V(\mathbb{Z}/p^j\mathbb{Z})$ for some $j > 0$ for each p.

(2) For all $p \gg 1$, the set S_p contains (all $\mathrm{PGL}_2(\mathbb{Z}_p)$-orbits of) all elements $f(x, y) \in V(\mathbb{Z}_p)$ such that $a(f) = 0$ and $b(f)$ is a p-adic unit, where $a(f)$ and $b(f)$ denote the x^4- and x^3y-coefficients of $f(x, y)$, respectively.

We then have the following result, by analogy with Theorem 3:

Theorem 7. *Let $S \subset V(\mathbb{Z})^{(i)}$ be a big family. Then the number of reducible $\mathrm{PGL}_2(\mathbb{Z})$-orbits on S of height less than X is*

$$\left(\tfrac{8+24i}{135} \cdot X^{\frac{5}{6}}\right) \cdot \prod_p (1 - p^{-1})^{-1} \int_{f \in (S_p)_0} |b(f)|_p \, df + O_\epsilon(X^{\frac{3}{4}+\epsilon}).$$

Finally, by analogy with Theorem 4, we have the following result, which is equivalent to Theorem 7:

Theorem 8. *Let $S \subset V(\mathbb{Z})^{(i)}$ be a big family. Then the number of reducible $\mathrm{PGL}_2(\mathbb{Z})$-orbits on S of height less than X is*

$$\left(\prod_p \int_{(I,J) \in \mathbb{Z}_p^2} \#\left(\frac{\mathrm{inv}^{-1}(I, J) \cap (S_p)_0}{P(\mathbb{Z}_p)}\right) dI \, dJ\right) \cdot \left(\tfrac{8+24i}{5} \cdot X^{\frac{5}{6}}\right) + O_\epsilon(X^{\frac{3}{4}+\epsilon}).$$

2.1. ***Reduction theory for the action of $\mathbf{PGL_2}(\mathbb{Z})$ on $V(\mathbb{R})$.*** To count orbits of $\mathrm{PGL}_2(\mathbb{Z})$ on $V(\mathbb{Z})$, we realize these orbits as lattice points in fundamental sets for the action of $\mathrm{PGL}_2(\mathbb{Z})$ on $V(\mathbb{R})$. In this section, we construct such fundamental sets by means of a two-step process: first, in Section 2.1.1, we describe a fundamental domain $\mathcal{F}$ for the action of $\mathrm{PGL}_2(\mathbb{Z})$ on $\mathrm{PGL}_2(\mathbb{R})$; subsequently, in Section 2.1.2, we combine $\mathcal{F}$ with fundamental sets for the action of $\mathrm{PGL}_2(\mathbb{R})$ on $V(\mathbb{R})$.

2.1.1. *A box-shaped fundamental domain for $\mathrm{PGL}_2(\mathbb{Z}) \curvearrowright \mathrm{PGL}_2(\mathbb{R})$.* We start by recalling Gauss' fundamental domain for the action of $\mathrm{PGL}_2(\mathbb{Z})$ on $\mathrm{PGL}_2(\mathbb{R})$, rephrased in terms of the Iwasawa decomposition of $\mathrm{PGL}_2(\mathbb{R})$, which we now recall. Let N be the (algebraic) subgroup of PGL_2 consisting of lower triangular unipotent matrices $\left[\begin{smallmatrix} 1 & 0 \\ u & 1 \end{smallmatrix}\right]$, let T be the maximal torus defined by $T = \left\{\left[\begin{smallmatrix} t^{-1} & 0 \\ 0 & t \end{smallmatrix}\right] : t \in \mathbb{R}_{>0}\right\}$, and let $K = \mathrm{SO}_2(\mathbb{R})/\{\pm \mathrm{id}\}$. We often abuse notation by writing u and t for the corresponding elements of $N(\mathbb{R})$ and A. If we let $T' = \{t \in T : t \geq \sqrt[4]{3}/\sqrt{2}\}$, then for each $t \in T'$, there exists a compact subset $N'(t) \subset \left[-\frac{1}{2}, \frac{1}{2}\right]$ such that the set

$$\{ut : u \in N'(t), \, t \in T'\} \cdot K \tag{4}$$

is a fundamental domain $\mathcal{F}$ for the action of $\mathrm{PGL}_2(\mathbb{Z})$ on $\mathrm{PGL}_2(\mathbb{R})$. It is well known that $N'(t) = \left[-\frac{1}{2}, \frac{1}{2}\right]$ for all $t \geq 1$; consequently, we say that this fundamental domain is *box-shaped at infinity*. This property of the fundamental domain is essential for our proof.

We denote elements of K by θ. With respect to the coordinates u on $N(\mathbb{R})$, t on T, and θ on K, the Haar measure dg on $\mathrm{PGL}_2(\mathbb{R})$ is given by

$$dg = d\theta\, du(t^{-2}\, d^\times t), \tag{5}$$

where $d^\times t = dt/t$. Above, $d\theta$ is normalized so that $\int_{\theta\in K} d\theta = 1$, and du is normalized so that $N(\mathbb{Z})$ has covolume 1 in $N(\mathbb{R})$.

2.1.2. *Fundamental sets for* $\mathrm{PGL}_2(\mathbb{R}) \curvearrowright V(\mathbb{R})$ *and* $\mathrm{PGL}_2(\mathbb{Z}) \curvearrowright V(\mathbb{R})$. The action of $\mathrm{PGL}_2(\mathbb{R})$ on $V(\mathbb{R})^{(0)} \sqcup V(\mathbb{R})^{(1)}$ has one orbit per set of *nondegenerate invariants* — i.e., a pair of invariants $(I, J) \in \mathbb{R}^2$ such that $\Delta(I, J) := 4I^3 - J^2 \neq 0$. This orbit belongs to $V(\mathbb{R})^{(0)}$ when $\Delta(I, J) > 0$ and to $V(\mathbb{R})^{(1)}$ when $\Delta(I, J) < 0$. Consider the function σ_0 given by

$$\sigma_0 : \mathbb{R}^2 \smallsetminus \{\Delta = 0\} \to V(\mathbb{R}), \quad (I, J) \mapsto x^3 y - \tfrac{1}{3} I x y^3 - \tfrac{1}{27} J y^4. \tag{6}$$

It is easy to check that σ_0 is a section of the map inv, meaning that the invariants of $\sigma_0(I, J)$ are I and J. For $i \in \{0, 1\}$, we take our fundamental sets for the action of $\mathrm{PGL}_2(\mathbb{R})$ on $V(\mathbb{R})$ to be

$$\mathcal{R}^{(i)} := \mathbb{R}_{>0} \cdot \{\sigma_0(I, J) : (-1)^i \Delta(I, J) > 0,\ H(I, J) = 1\}.$$

Finally, we note that the stabilizer $\mathrm{Stab}_{\mathrm{PGL}_2(\mathbb{R})}(F)$ is independent of the choice of $f \in V(\mathbb{R})^{(i)}$; letting $n_i := \#\,\mathrm{Stab}_{\mathrm{PGL}_2(\mathbb{R})}(f)$ for any $f \in V(\mathbb{R})^{(i)}$, one readily verifies that $n_0 = 4$ and $n_1 = 2$.

We conclude that the multiset $\mathcal{F}\cdot\mathcal{R}^{(i)}$ is a cover for a fundamental domain for the action of $\mathrm{PGL}_2(\mathbb{Z})$ on $V(\mathbb{R})^{(i)}$. More precisely, every $\mathrm{PGL}_2(\mathbb{Z})$-orbit of $f \in V(\mathbb{R})^{(i)}$ is represented exactly $n_i/\#\,\mathrm{Stab}_{\mathrm{PGL}_2(\mathbb{Z})}(f)$ times in $\mathcal{F}\cdot\mathcal{R}^{(i)}$.

2.2. ***The action of the subgroup P on the reducible hyperplane V_0.*** In this section, we examine the action of the lower-triangular (parabolic) subgroup P on the reducible hyperplane V_0. Specifically, in Section 2.2.1, we show that over many interesting base rings R, the action of $P(R)$ on the set of forms in $V_0(R)$ lying over a given nondegenerate pair of invariants is simply transitive. Then, in Section 2.2.2, we prove a Jacobian change-of-variables formula relating the Euclidean measure on $V_0(R)$ to the product of the Haar measure on $P(R)$ with the Euclidean measure on R^2. This formula will be applied in Sections 2.3.3–2.3.4.

2.2.1. *Reduction theory for the action of P on V_0.* Let R be a field or $\mathbb{Z}_p$ for some prime p. Then we have the following result, which classifies the orbits and stabilizers of $P(R)$ on $V_0(R)$:

Proposition 9. *Let R be as above, and let $(I, J) \in R^2$ be such that $\Delta(I, J)$ is a unit. Then the set of binary quartic forms in $V_0(R)$ with invariants I and J is either empty or consists of a single $P(R)$-orbit, and the stabilizer of any element in this orbit is trivial.*

Proof. Given a form $f(x, y) = bx^3 y + cx^2 y^2 + dxy^3 + ey^4 \in V_0(R)$ having invariants I and J, we first note that $b^2 \mid \Delta(f) = \Delta(I, J)$. Thus, if $\Delta(I, J)$ is a unit, then so is b. As a consequence, by replacing $f(x, y)$ with a $P(R)$-translate, we can arrange that $b = 1$. When $R \neq \mathbb{Z}_3$, we have that $3 \in R^\times$, and

hence by replacing f with the $P(R)$-translate $f(x-c/3y, y)$, we may assume that $c = 0$. When $R = \mathbb{Z}_3$, we may similarly replace f with a $P(R)$-translate to arrange that $c \in \{0, 1, 2\}$ (depending on the residue classes of I modulo 9 and J modulo 27). Once this has been done, the values of d and e are respectively determined by I and J (since $a = 0$ implies linear relations between (I, J) and (d, e)). This proves that the set of elements in $V_0(R)$ with invariants I and J (if nonempty) form a single $P(R)$-orbit.

Next, we prove the claim regarding the stabilizer in $P(R)$ of $f \in V_0(R)$. Suppose that an element $g \in P(R)$, represented by a matrix with coefficients 1 and u on the diagonal and n in the lower left coordinate, fixes f. First note that since b is a unit, the x^2y^2-coefficient of $g \cdot f(x, y)$ will change unless $n = 0$. Assume thus that $n = 0$. Next note that the x^3y-coefficient of $g \cdot f(x, y)$ is $u^{-1}b$, implying that $u = 1$, as needed. □

The above result has the following immediate consequence by specializing to the case $R = \mathbb{R}$.

Lemma 10. *Let $(I, J) \in \mathbb{R}^2$ be nondegenerate. Then the set $\{f \in \mathrm{inv}^{-1}(f)_0 : b(f) > 0\}$ consists of a single $N(\mathbb{R})T$-orbit.*

2.2.2. *A Jacobian change-of-variables formula.* Proposition 9 implies that when $R = \mathbb{R}$ or $\mathbb{Z}_p$ for a prime p, the space $V_0(R)$ is a fibration over R^2, where the generic fiber can be identified with $P(R)$, so long as it is nonempty. Thus, the Euclidean measure on $V_0(R)$ should be related to the product of the Haar measure on $P(R)$ with the Euclidean measure on R^2. The following proposition gives a Jacobian change-of-variables formula relating these measures:

Proposition 11. *Let $R = \mathbb{R}$ or $\mathbb{Z}_p$ for some prime p, and let $\phi : V_0(R) \to \mathbb{R}$ be a measurable function. Then we have*

$$\int_{f \in V_0(R)} \phi(f)|b(f)|\, df = \frac{2}{27} \int_{\substack{(I,J)\in R^2 \\ \Delta(I,J)\neq 0}} \Bigg(\sum_{f \in \frac{\mathrm{inv}^{-1}(I,J)_0}{P(R)}} \int_{h \in P(R)} \phi(h \cdot f)\, dh \Bigg) dI\, dJ,$$

where df, dI, and dJ are Euclidean measures, where dh is the right Haar measure on P given by $dh = t\, du\, dt$, and where $|-|$ denotes the usual absolute value on R.

Proof. First note that the result for $R = \mathbb{R}$ implies the result for $R = \mathbb{Z}_p$ by the principle of permanence of identities, so it suffices to treat the case $R = \mathbb{R}$. Recall the construction of the section σ_0 in (6). Proposition 9 implies that we have

$$V_0(\mathbb{R}) = P(\mathbb{R}) \cdot \sigma_0(\{(I, J) \in \mathbb{R}^2 : \Delta(I, J) \neq 0\}).$$

Thus, the theorem follows from the equality

$$\int_{f \in P(\mathbb{R})\cdot\sigma_0(\mathbb{R}^2 \smallsetminus \{\Delta=0\})} \phi(f)|b(f)|\, df = \frac{2}{27} \int_{(I,J)\in\mathbb{R}^2} \int_{h \in P(\mathbb{R})} \phi(h \cdot \sigma_0(I, J))\, dh\, dI\, dJ, \tag{7}$$

which in turn is a consequence of the following Jacobian change-of-variables computation. First note that a typical element of the region of integration on the left-hand side of (7) is given by

$$(tu) \cdot \sigma_0(I, J) = \frac{1}{t^2}x^3y + 3ux^2y^2 + \Big(3u^2t^2 - \frac{It^2}{3}\Big)xy^3 + \Big(u^3t^4 - \frac{Iut^4}{3} - \frac{Jt^4}{27}\Big)y^4.$$

By taking partial derivatives with respect to t, u, I, and J of the coefficients of the binary quartic form on the right-hand side above and arranging these partials into matrix form, we find that the Jacobian determinant relating the measures df and $dt\,du\,dI\,dJ$ is given by

$$\begin{vmatrix} -2t^{-3} & 0 & 6u^2t - \frac{2}{3}It & 4u^3t^3 - \frac{4}{3}Iut^3 - \frac{4}{27}Jt^3 \\ & 3 & 6ut^2 & 3u^2t^4 - \frac{1}{3}It^4 \\ & & -\frac{1}{3}t^2 & -\frac{1}{3}ut^4 \\ & & & -\frac{1}{27}t^4 \end{vmatrix} = -\tfrac{2}{27}t^3.$$

Therefore, we have

$$df = \tfrac{2}{27}t^3\,dt\,du\,dI\,dJ.$$

Equation (7) then follows, since $b((tu)\cdot\sigma_0(I,J)) = t^{-2}$. □

2.3. ***Counting reducible*** $\mathbf{PGL_2}(\mathbb{Z})$***-orbits on*** $V(\mathbb{Z})$**.** Let $i \in \{0,1\}$. In this section, we obtain asymptotics for the number of reducible orbits of $\mathrm{PGL}_2(\mathbb{Z})$ on $V_0(\mathbb{Z})^{(i)}$ of bounded height, thus proving Theorems 6–8. To simplify the exposition in the rest of this section, we introduce the following notation:

• For any set $S \subset V(\mathbb{Z})$, let $S_{\mathrm{red}} \subset S$ be the subset of forms in S having a rational linear factor; for $X > 0$, let

$$S_X := \{B \in S : H(B) < X\},$$

and as before, let $S_0 := S \cap V_0(\mathbb{Z})$ be the set of elements of S that lie on the reducible hyperplane.

• Let $G_0 \subset \mathrm{PGL}_2(\mathbb{R})$ be a fixed nonempty open bounded set such that $G_0^{-1} = G_0$ and G_0 is left- and right-invariant under the group K' generated by K together with the diagonal matrix having diagonal entries 1 and -1. Such a set can be constructed by starting with a nonempty open bounded set G_0' and taking $G_0 = K'(G_0' \cup G_0'^{-1})K'$.

• Define the multiset $\mathcal{B}_\infty$ by

$$\mathcal{B}_\infty := G_0 \cdot \mathcal{R}^{(i)} \cap V_0(\mathbb{R}).$$

Set $\mathcal{B} := (\mathcal{B}_\infty)_1$, and note that by the construction of $\mathcal{R}$, we have $(\mathcal{B}_\infty)_X = X^{1/6}\mathcal{B}$.

• We define the quantity $C(\mathcal{B})$ by

$$C(\mathcal{B}) := \frac{1}{\tilde{n}_i \operatorname{Vol}(G_0)} \cdot \int_{f\in\mathcal{B}} |b(f)|\,df, \tag{8}$$

where the volume of G_0 is computed using the Haar measure dg, and where $\tilde{n}_i = 2n_i$, with n_i as defined in Section 2.1.2.

• For a finite set Σ of $\mathrm{PGL}_2(\mathbb{Z})$-orbits on $V(\mathbb{Z})$, let $\#'\,\Sigma$ be the number of elements of Σ, where each $f \in \Sigma$ is counted with weight $1/\#\operatorname{Stab}_{\mathrm{PGL}_2(\mathbb{Z})}(f)$.

2.3.1. *Averaging over fundamental domains.* We begin by applying Bhargava's averaging technique, developed in [2; 4]. By an argument identical to [9, Theorem 2.5], which involves averaging over translates

of the fundamental domain $\mathcal{F}$ by elements of G_0 and performing a suitable change-of-variables, we have

$$\begin{aligned}\#'\left(\frac{(V^{(i)}(\mathbb{Z})_{\mathrm{red}})_X}{\mathrm{PGL}_2(\mathbb{Z})}\right) &= \frac{1}{n_i}\cdot\#(\mathcal{F}h\cdot\mathcal{R}^{(i)}\cap(V(\mathbb{Z})_{\mathrm{red}})_X)\\ &= \frac{1}{n_i\,\mathrm{Vol}(G_0)}\int_{g\in\mathcal{F}}\#(gG_0\cdot\mathcal{R}_X^{(i)}\cap V(\mathbb{Z})_{\mathrm{red}})\,dg,\end{aligned}\tag{9}$$

Then we have the following result:

Proposition 12. *We have*

$$\#\left(\frac{(V^{(i)}(\mathbb{Z})_{\mathrm{red}})_X}{\mathrm{PGL}_2(\mathbb{Z})}\right) = \frac{1}{n_i\,\mathrm{Vol}(G_0)}\int_{t=1}^{\infty}\int_{u=-\frac{1}{2}}^{\frac{1}{2}}\#(utX^{\frac{1}{6}}\mathcal{B}\cap V_0(\mathbb{Z}))t^{-2}\,du\,d^{\times}t + O_\epsilon(X^{\frac{3}{4}+\epsilon}),\tag{10}$$

where $\mathcal{B}$ is the multiset

$$\mathcal{B} := (\mathcal{B}_\infty)_1 := G_0\cdot\mathcal{R}_1^{(i)}\cap V_0(\mathbb{R}).$$

Proof. The number of reducible (and irreducible) $\mathrm{PGL}_2(\mathbb{Z})$-orbits on $V(\mathbb{Z})$ that have nontrivial integral stabilizer was proven in [9, Lemma 2.4] to be bounded $O(X^{3/4+\epsilon})$. Thus, we may replace the $\#'$ in (9) with $\#$ at the cost of an error of $O(X^{3/4+\epsilon})$. We split the fundamental domain $\mathcal{F}$ as

$$\mathcal{F} = \mathcal{F}'\sqcup\left\{utk : u\in\left[-\tfrac{1}{2},\tfrac{1}{2}\right],\ t\geq 1,\ k\in K\right\},$$

where $\mathcal{F}'$ is absolutely bounded. When the integral over $\mathcal{F}$ is restricted to the compact region $\mathcal{F}'$, it is clear that the number of forms with vanishing x^4-coefficient is bounded by $O(X^{2/3})$. By [9, Lemma 2.3], the number of reducible forms with nonzero x^4-coefficient is bounded by $O_\epsilon(X^{2/3+\epsilon})$. These error terms are sufficiently small. The proposition then follows upon noting that, by the left K-invariance of G_0, the set $gG_0\mathcal{R}^{(i)}$ is independent of θ when g is written as $g = ut\theta$ in Iwasawa coordinates. □

2.3.2. *Slicing.* Throughout this subsection we set $Y := X^{1/6}$. The integrand of the right-hand side of (10) is the number of integral points in the region $utY\mathcal{B}\cap V_0(\mathbb{R})$. This region is typically quite skewed: indeed, whenever t is high up in the cusp, the x^3y-coefficient is small, so the volumes of the projections of the set $utY\mathcal{B}$ away from this coefficient have the same order of magnitude as the volume of $utY\mathcal{B}$ itself. Furthermore, the region where t is high up in the cusp contributes most of the lattice points that we are interested in counting! We resolve this issue in this section by fibering the region $utY\mathcal{B}\cap V_0(\mathbb{R})$ by the x^3y-coefficient and using a result of Davenport to estimate the number of lattice points on each fiber.

We now partition the region $utY\mathcal{B}$ into slices, one for each possible value of the x^3y-coefficient. For any $b\in\mathbb{R}\smallsetminus\{0\}$, and any $\mathcal{S}\subset V(\mathbb{R})$, let $\mathcal{S}|_b$ denote the *slice of $\mathcal{S}$ at b*, i.e., the subset of forms in $\mathcal{S}$ with x^4-coefficient equal to 0 and x^3y-coefficient equal to b. Then we can express the integrand of the right-hand side of (10) as

$$\#\big(utY\mathcal{B}\cap V_0(\mathbb{Z})\big) = \sum_{b\in\mathbb{Z},\,b\neq 0}\#\big((utY\mathcal{B})|_b\cap V(\mathbb{Z})\big) = \sum_{b\in\mathbb{Z},\,b\neq 0}\mathrm{Vol}((utY\mathcal{B})|_b)(1+O(Y^{-1})),\tag{11}$$

where the final estimate is a consequence of the following proposition, due to Davenport:

Proposition 13 [21]. *Let $\mathcal{R}$ be a bounded, semialgebraic multiset in $\mathbb{R}^n$ having maximum multiplicity m that is defined by at most k polynomial inequalities, each having degree at most ℓ. Let $\mathcal{R}'$ denote the image of $\mathcal{R}$ under any (upper or lower) triangular, unipotent transformation of $\mathbb{R}^n$. Then the number of integer lattice points (counted with multiplicity) contained in the region $\mathcal{R}'$ is given by*

$$\mathrm{Vol}(\mathcal{R}) + O(\max\{\mathrm{Vol}(\overline{\mathcal{R}}), 1\}),$$

where $\mathrm{Vol}(\overline{\mathcal{R}})$ denotes the greatest d-dimensional volume of any projection of $\mathcal{R}$ onto a coordinate subspace obtained by equating $n-d$ coordinates to zero, where d ranges over all values in $\{1,\dots,n-1\}$. The implied constant in the second summand depends only on n, m, k, and ℓ.

Now, since unipotent transformations preserve both the value of b and the volume, we have

$$\mathrm{Vol}((utY\mathcal{B})|_b) = \mathrm{Vol}((tY\mathcal{B})|_b).$$

Combining (11) with Proposition 12 yields

$$\#\left(\frac{(V^{(i)}(\mathbb{Z})_{\mathrm{red}})_X}{\mathrm{PGL}_2(\mathbb{Z})}\right) = \frac{1}{n_i\,\mathrm{Vol}(G_0)} \sum_{b\in\mathbb{Z}\smallsetminus\{0\}} \int_{1\le t\ll Y^{1/2}/|b|^{1/2}} \mathrm{Vol}((tY\mathcal{B})|_b)(1+O(Y^{-1}))t^{-2}\,du\,d^\times t + O_\epsilon(X^{\frac{3}{4}+\epsilon}). \quad (12)$$

The bound in the region of integration above is obtained by noting that the value of $|b|$ is bounded above by $O(Y)$, and, for each fixed $b\neq 0$, the range of t goes up to $Y^{1/2}/\sqrt{|b|}$, since the x^3y-coefficients of elements in $utY\mathcal{B}$ are $\ll t^{-2}Y$. We bound the error term in the right-hand side of (12) by

$$\ll Y^{-1}\,\mathrm{Vol}((tY\mathcal{B})|_b) \ll \sum_{b=1}^{O(Y)} \int_{t=1}^{O(Y^{1/2}/\sqrt{b})} Y^2t^6\frac{d^\times t}{t^2} \ll Y^4 \sum_{b=1}^{O(Y)} b^{-2} \ll X^{\frac{2}{3}}. \quad (13)$$

Substituting the estimate (13) into (12) yields

$$\#\left(\frac{(V^{(i)}(\mathbb{Z})_{\mathrm{red}})_X}{\mathrm{PGL}_2(\mathbb{Z})}\right) = \frac{1}{n_i\,\mathrm{Vol}(G_0)} \sum_{b\in\mathbb{Z}\smallsetminus\{0\}} \int_{1\le t\ll Y^{1/2}/|b|^{1/2}} \mathrm{Vol}((tY\mathcal{B})|_b)\frac{d^\times t}{t^2} + O_\epsilon(X^{\frac{3}{4}+\epsilon}). \quad (14)$$

We now manipulate the integrand in (14) to extract its dependence on the slicing index b. Because unipotent transformations leave volumes unchanged, and the action of t on $V(\mathbb{R})$ scales the x^iy^j-coefficient by t^{i-j}, we have

$$\mathrm{Vol}((tY\mathcal{B})|_b) = Y^3\,\mathrm{Vol}((t\mathcal{B})|_{b/Y}) = t^6X^{1/2}\,\mathrm{Vol}(\mathcal{B}|_{t^2b/Y}). \quad (15)$$

Since G_0 is left-K'-invariant and since the diagonal matrix with entries 1 and -1 belongs to K', it follows that $\mathrm{Vol}(\mathcal{B}|_\beta) = \mathrm{Vol}(\mathcal{B}|_{-\beta})$ for any $\beta\in\mathbb{R}\smallsetminus\{0\}$. Hence, setting $\beta = t^2b/Y$ (which gives $d^\times\beta = 2d^\times t$),

we obtain

$$\#\left(\frac{(V^{(i)}(\mathbb{Z})_{\mathrm{red}})_X}{\mathrm{PGL}_2(\mathbb{Z})}\right) = \frac{X^{1/2}}{2n_i \operatorname{Vol}(G_0)} \sum_{b\in\mathbb{Z}\smallsetminus\{0\}} b^{-2} \int_{|b|/Y\le\beta\ll 1} Y^2\beta^2 \operatorname{Vol}(\mathcal{B}|_\beta)\, d^\times\beta + O_\epsilon(X^{\frac{3}{4}+\epsilon})$$
$$= \frac{X^{5/6}}{n_i \operatorname{Vol}(G_0)} \sum_{b=1}^{\infty} b^{-2} \int_{\beta\ge 0} \beta \operatorname{Vol}(\mathcal{B}|_\beta)\, d\beta + O_\epsilon(X^{\frac{3}{4}+\epsilon}), \tag{16}$$

where the second line above follows since $\mathcal{B}|_\beta$ is a bounded set and $\operatorname{Vol}(\mathcal{B}|_\beta) \ll 1$. It therefore follows that

$$\#\left(\frac{(V^{(i)}(\mathbb{Z})_{\mathrm{red}})_X}{\mathrm{PGL}_2(\mathbb{Z})}\right) = \zeta(2)\cdot C(\mathcal{B})\cdot X^{\frac{5}{6}} + O_\epsilon(X^{\frac{3}{4}+\epsilon}), \tag{17}$$

where $C(\mathcal{B})$ was defined in (8). Note that while going from (14) to (17), we pick up a factor of $\frac{1}{2}$ from $d^\times\beta = 2d^\times t$, a factor of 2 from restricting the sum over $b\in\mathbb{Z}$ to the sum over $b>0$, and another factor of $\frac{1}{2}$ by replacing the integral over $\beta\ge 0$ to the integral over all β in the definition of $C(\mathcal{B})$ (which is equivalent to the integral over all $f\in\mathcal{B}$).

2.3.3. *Computing the constant.* In this section, we compute the value of $C(\mathcal{B})$. Recall that we defined $\mathcal{B}$ to be the multiset $\mathcal{B} := G_0\cdot\mathcal{R}_1^{(i)}\cap V_0(\mathbb{R})$. Since G_0 is right K'-invariant, we may write $G_0 = \mathcal{S}K'$ for some $\mathcal{S}\subset N(\mathbb{R})T$. Hence, we have

$$\mathcal{B} = \mathcal{S}K'\cdot\mathcal{R}_1^{(i)}\cap V_0(\mathbb{R}) = \mathcal{S}\cdot(K'\mathcal{R}_1^{(i)})_0.$$

We then have the following lemma concerning the multiplicity of the fiber of $(K'\mathcal{R}^{(i)})_0$ over $\mathrm{inv}(V(\mathbb{R})^{(i)})\subset\mathbb{R}^2$:

Lemma 14. *The map* $\mathrm{inv}:(K'\mathcal{R}^{(i)})_0\to\{\mathrm{inv}(V(\mathbb{R})^{(i)})$ *is* $\tilde{n}_i$ *to* 1.

Proof. We begin by noting the map is certainly surjective since the invariants I and J of $x^3y+dxy^2+ey^3$ are linear in d and e, respectively. It thus suffices to prove that the map

$$\mathrm{inv}:\{f\in(K'\mathcal{R}^{(i)})_0 : b(f)>0\}\to\mathrm{inv}(V(\mathbb{R})^{(i)})$$

is n_i to 1. This is a consequence of the following two facts: first, the stabilizer in $\mathrm{PGL}_2(\mathbb{R})$ of any element in the image has size n_i, and second, the group $N(\mathbb{R})T$ acts simply transitively on $\mathrm{inv}^{-1}(I,J)\cap V_0(\mathbb{R})$ for any $(I,J)\in\mathrm{inv}(V(\mathbb{R})^{(i)})$ by Lemma 10. Indeed, given $f\in\mathcal{R}^{(i)}$ having invariants (I,J), and $p\theta\in\mathrm{Stab}_{\mathrm{PGL}_2(\mathbb{R})}(f)$ with $p\in N(\mathbb{R})T$ and $\theta\in K'$, the element $\theta f = p^{-1}f$ belongs to $(K'\mathcal{R}^{(i)})_0$ and has invariants (I,J). This association yields the result. □

We are now in position to compute the constant $C(\mathcal{B})$:

Proposition 15. *We have*

$$C(\mathcal{B}) = \tfrac{1}{27}\operatorname{Vol}\{(I,J)\in\mathbb{R}^2 : (-1)^i\Delta(I,J)>0,\ H(I,J)<1\}.$$

Proof. We have

$$\begin{aligned} C(\mathcal{B}) &= \frac{1}{\tilde{n}_i \operatorname{Vol}(G_0)} \int_{f\in\mathcal{B}} |b(f)|\, df = \frac{1}{\tilde{n}_i \operatorname{Vol}(G_0)} \int_{f\in\mathcal{S}\cdot(K'\mathcal{R}_1^{(i)})_0} |b(f)|\, df \\ &= \frac{2}{27\operatorname{Vol}(G_0)} \int_{\substack{(I,J)\in\mathbb{R}^2 \\ (-1)^i\Delta(I,J)>0 \\ H(I,J)<1}} \int_{h\in\mathcal{S}} dh\, dI\, dJ \\ &= \frac{2\operatorname{Vol}_{\text{right}}(\mathcal{S})\operatorname{Vol}\{(I,J)\in\mathbb{R}^2 : (-1)^i\Delta(I,J)>0,\ H(I,J)<1\}}{27\operatorname{Vol}(\mathcal{S}K')}, \end{aligned} \tag{18}$$

where the second line follows by applying the Jacobian change-of-variables established in Proposition 11 along with Lemma 14. In the third line, $\operatorname{Vol}_{\text{right}}(\mathcal{S})$ denotes the volume of $\mathcal{S}$ with respect to the right Haar measure on $P(\mathbb{R})$. But since $\operatorname{Vol}(K)$ is normalized to be equal to 1, we have that $\operatorname{Vol}_{\text{right}}(\mathcal{S}) = \frac{1}{2}\operatorname{Vol}(K'\mathcal{S})$, where the volume is computed with respect to the Haar measure on $G(\mathbb{R})$. The next lemma demonstrates that $\operatorname{Vol}(K'\mathcal{S}) = \operatorname{Vol}(\mathcal{S}K')$:

Lemma 16. *We have $K'\mathcal{S} = \mathcal{S}K'$, and in particular* $\operatorname{Vol}(K'\mathcal{S}) = \operatorname{Vol}(\mathcal{S}K')$.

Proof of Lemma 16. Since $G_0 = \mathcal{S}K'$ is left-K'-invariant and inversion-invariant, it follows that

$$K'\mathcal{S} \subset K'\mathcal{S}K' = \mathcal{S}K' = G_0 = G_0^{-1} = K'\mathcal{S}^{-1}. \tag{19}$$

Since the Iwasawa decomposition of $\mathrm{PGL}_2(\mathbb{R})$ is unique, (19) implies that $\mathcal{S} \subset \mathcal{S}^{-1}$, and hence also that $\mathcal{S} = \mathcal{S}^{-1}$. Thus, $\mathcal{S}K' = K'\mathcal{S}^{-1} = K'\mathcal{S}$, as desired. □

Combining (18) with the result of Lemma 16 completes the proof of Proposition 15. □

2.3.4. *Congruence conditions.* We now prove Theorem 7. Let $S \subset V(\mathbb{Z})^{(i)}$ be a big family, and suppose for now that S is defined by congruence conditions at finitely many places (i.e., suppose that $S_p = V(\mathbb{Z}_p)$ for all primes $p \gg 1$). For each $b \in \mathbb{Z}_p \smallsetminus \{0\}$, let

$$(S_p)_0|_b := \{f \in (S_p)_0 : b(f) = b\},$$

and for each $b \in \mathbb{Z} \smallsetminus \{0\}$, let $\nu(S_0|_b) := \prod_p \operatorname{Vol}((S_p)_0|_b)$ denote the density of the slice $S_0|_b$ in $V_0(\mathbb{Z})|_b$; here, each p-adic volume $\operatorname{Vol}((S_p)_0|_b)$ is computed with respect to the Euclidean measure on $V_0(\mathbb{Z}_p)|_b$, normalized so that $V_0(\mathbb{Z}_p)|_b$ has volume 1. Then an argument identical to the one used to obtain (14) yields the asymptotic formula

$$\#\left(\frac{(S_{\text{red}})_X}{\mathrm{PGL}_2(\mathbb{Z})}\right) = \frac{1}{n_i \operatorname{Vol}(G_0)} \sum_{b\in\mathbb{Z}\smallsetminus\{0\}} \nu(S_0|_b) \int_{t\geq 1} \operatorname{Vol}((tY\mathcal{B})|_b) t^{-2}\, d^\times t + O_\epsilon(X^{\frac{3}{4}+\epsilon}). \tag{20}$$

Note that the upper bound on t in (14) can be omitted, since if $t > cY^{1/2}/b^{1/2}$ for some sufficiently large constant c, then $(tY\mathcal{B})|_b$ is empty. The $\mathrm{PGL}_2(\mathbb{Z})$-invariance of S implies that $\nu(S_0|_b) = \nu(S_0|_{|b|})$ for all b. Hence, the argument used to deduce (16) and (17) yields the estimate

$$\#\left(\frac{(S_{\text{red}})_X}{\mathrm{PGL}_2(\mathbb{Z})}\right) = C(\mathcal{B}) \cdot \left(\sum_{b=1}^{\infty} \frac{\nu(S_0|_b)}{b^2}\right) \cdot X^{\frac{5}{6}} + O_\epsilon(X^{\frac{3}{4}+\epsilon}). \tag{21}$$

To evaluate the sum over b on the right-hand side of (21), we use the following property, which is a consequence of the fact that S is a big family: if p is a prime and $b, b' \in \mathbb{Z}_p \smallsetminus \{0\}$ are elements such that $|b_i|_p = |b_i'|_p$ for each i, then $\mathrm{Vol}((S_p)_0|_b) = \mathrm{Vol}((S_p)_0|_{b'})$. By repeatedly using this property, we obtain the chain of equalities

$$\begin{aligned}\sum_{b=1}^{\infty} \frac{\nu(S_0|_b)}{b^2} &= \prod_p \sum_{i=0}^{\infty} \frac{\mathrm{Vol}((S_p)_0|_{p^i})}{p^{2i}} \\ &= \prod_p \Big(1-\frac{1}{p}\Big)^{-1} \int_{\substack{b\in\mathbb{Z}_p \\ b\neq 0}} |b|_p \,\mathrm{Vol}((S_p)_0|_b)\, db \\ &= \prod_p \Big(1-\frac{1}{p}\Big)^{-1} \int_{f\in(S_p)_0} |b(f)|_p \, df,\end{aligned}$$

where the second line above follows by partitioning the region of integration $\mathbb{Z}_p \smallsetminus \{0\}$ into level sets for the integrand and summing over all such level sets, and where the last line above follows just as in (17).

It remains to handle the case where S is a big family defined by congruence conditions at infinitely many places. This case follows by using an inclusion-exclusion sieve[4] in conjunction with the following bound on the number of $\mathrm{PGL}_2(\mathbb{Z})$-equivalence classes of forms f with large $b(f)$-value:

Theorem 17. *Fix a real number $M > 0$. Then the number of $\mathrm{PGL}_2(\mathbb{Z})$-equivalence classes of (or equivalently, $P(\mathbb{Z})$-orbits of) elements of the set $\{f \in V_0(\mathbb{Z}) : H(f) < X,\ |b(f)| \geq M\}$ is bounded by $O(X^{5/6}/M) + O_\epsilon(X^{3/4+\epsilon})$, where the implied constant is independent of M.*

Proof. The required bound follows immediately from the proof of Theorem 6 by simply summing (16) over only those b such that $|b| \geq M$. □

This concludes the proof of Theorem 7. We finish by noting that Theorem 8 follows from Theorem 7 by applying the Jacobian change-of-variables result in Proposition 11 to each p-adic integral.

3. Reduction theory for the action of $G(\mathbb{Z})$ on $W(\mathbb{R})$

Fix an integer $n \geq 3$. In this section, we construct finite covers of a fundamental set for the action of $G(\mathbb{Z})$ on $W(\mathbb{R})$. As in Section 2.1, we achieve this in two steps: first, in Sections 3.1–3.2, we choose a certain fundamental domain $\mathcal{F}$ for the action of $G(\mathbb{Z})$ on $G(\mathbb{R})$, and then in Section 3.3, we combine $\mathcal{F}$ with fundamental sets for the action of $G(\mathbb{R})$ on $W(\mathbb{R})$ to construct our required covers.

Note that it suffices to construct a fundamental domain for $\mathrm{O}_{\mathcal{A}}(\mathbb{Z})$ on $\mathrm{O}_{\mathcal{A}}(\mathbb{R})$, as such a domain would also be a fundamental domain for $G(\mathbb{Z})$ on $G(\mathbb{R})$. Indeed, when n is odd, this follows because $\mathrm{O}_{\mathcal{A}}(\mathbb{Z})$ contains an element of determinant -1, namely, the negative of the identity matrix. When n is even, this follows because the 2-torsion elements of $\mathrm{O}_{\mathcal{A}}(\mathbb{R})$ are contained in $\mathrm{O}_{\mathcal{A}}(\mathbb{Z})$. Thus, in Sections 3.1–3.2, we work with the group $\mathrm{O}_{\mathcal{A}}$. Furthermore, for our counting purposes, we cannot simply use any fundamental

[4]We do not flesh out the sieving argument here to avoid being repetitive, because in Section 6.3 (to follow), we use the same sort of argument to prove Theorem 4.

domain $\mathcal{F}$ for the action of $\mathrm{O}_{\mathcal{A}}(\mathbb{Z})$ on $\mathrm{O}_{\mathcal{A}}(\mathbb{R})$; rather, we require that $\mathcal{F}$ be *box-shaped at infinity*, and we prove that such a choice of fundamental domain exists in Section 3.2.

3.1. ***A coordinate system for*** $\mathbf{O}_{\mathcal{A}}$**.** We begin by recalling the Iwasawa decomposition

$$\mathrm{O}_{\mathcal{A}}(\mathbb{R}) = N(\mathbb{R})TK,$$

where N denotes the (algebraic) group of lower triangular unipotent matrices in $\mathrm{O}_{\mathcal{A}}$, T denotes the set of diagonal matrices with positive entries contained in $\mathrm{O}_{\mathcal{A}}(\mathbb{R})$, and K denotes a maximal compact subgroup of $\mathrm{O}_{\mathcal{A}}(\mathbb{R})$. We note that T is a maximal torus of $\mathrm{O}_{\mathcal{A}}(\mathbb{R})$, and that the elements of T normalize $N(\mathbb{R})$.

A calculation shows that the elements of $N(\mathbb{R})$ are parametrized as

$$\begin{bmatrix}
1 & & & & & & & & & & \\
u_{21} & 1 & & & & & & & & & \\
u_{31} & u_{32} & 1 & & & & & & & & \\
\vdots & \vdots & \vdots & \ddots & & & & & & & \\
u_{(\lceil \frac{n}{2}\rceil -1)1} & u_{(\lceil \frac{n}{2}\rceil -1)2} & u_{(\lceil \frac{n}{2}\rceil -1)3} & \cdots & 1 & & & & & & \\
u_{\lceil \frac{n}{2}\rceil 1} & u_{\lceil \frac{n}{2}\rceil 2} & u_{\lceil \frac{n}{2}\rceil 3} & \cdots & u_{\lceil \frac{n}{2}\rceil \lfloor \frac{n}{2}\rfloor} & 1 & & & & & \\
u_{(\lceil \frac{n}{2}\rceil +1)1} & u_{(\lceil \frac{n}{2}\rceil +1)2} & u_{(\lceil \frac{n}{2}\rceil +1)3} & \cdots & -\frac{1}{2}u^2_{\lceil \frac{n}{2}\rceil \lfloor \frac{n}{2}\rfloor} + * & -u_{\lceil \frac{n}{2}\rceil \lfloor \frac{n}{2}\rfloor} & 1 & & & & \\
\vdots & \vdots & \vdots & \iddots & \vdots & \vdots & \vdots & \ddots & & & \\
u_{(n-2)1} & u_{(n-2)2} & -\frac{1}{2}u^2_{\lceil \frac{n}{2}\rceil 3} + * & \cdots & * & * & * & \cdots & 1 & & \\
u_{(n-1)1} & -\frac{1}{2}u^2_{\lceil \frac{n}{2}\rceil 2} + * & * & \cdots & * & * & * & \cdots & -u_{32} & 1 & \\
-\frac{1}{2}u^2_{\lceil \frac{n}{2}\rceil 1} + * & * & * & \cdots & * & * & * & \cdots & * & -u_{21} & 1
\end{bmatrix}, \tag{22}$$

$$\begin{bmatrix}
1 & & & & & & & & & & \\
u_{21} & 1 & & & & & & & & & \\
u_{31} & u_{32} & 1 & & & & & & & & \\
\vdots & \vdots & \vdots & \ddots & & & & & & & \\
u_{(\frac{n-2}{2})1} & u_{(\frac{n-2}{2})2} & u_{(\frac{n-2}{2})3} & \cdots & 1 & & & & & & \\
u_{\frac{n}{2}1} & u_{\frac{n}{2}2} & u_{\frac{n}{2}3} & \cdots & u_{\frac{n}{2}(\frac{n-2}{2})} & 1 & & & & & \\
u_{(\frac{n+2}{2})1} & u_{(\frac{n+2}{2})2} & u_{(\frac{n+2}{2})3} & \cdots & u_{(\frac{n+2}{2})(\frac{n-2}{2})} & 0 & 1 & & & & \\
u_{(\frac{n+4}{2})1} & u_{(\frac{n+4}{2})2} & u_{(\frac{n+4}{2})3} & \cdots & * & * & -u_{\frac{n}{2}(\frac{n-2}{2})} & 1 & & & \\
\vdots & \vdots & \vdots & \iddots & \vdots & \vdots & \vdots & \vdots & \ddots & & \\
u_{(n-2)1} & u_{(n-2)2} & * & \cdots & * & * & * & * & \cdots & 1 & \\
u_{(n-1)1} & * & * & \cdots & * & * & * & * & \cdots & -u_{32} & 1 \\
* & * & * & \cdots & * & * & * & * & \cdots & * & -u_{21} \quad 1
\end{bmatrix}, \tag{23}$$

where the $u_{ij} \in \mathbb{R}$ for $i \in \{2, \dots, n-1\}$ and $j \in \{1, \dots, \min\{i-1, n-i\}\}$ are free parameters, and where the symbol "$*$" is shorthand and is read as follows: if the "$*$" occurs in the row-i, column-j entry, then it denotes "some polynomial of positive degree in the variables $\{u_{i'j'} : i' - j' \le i - j\}$ with integer coefficients and no constant term" (the polynomial being abbreviated depends on the matrix entry in

which it occurs). We often abbreviate the tuple

$$\big(u_{ij} : i \in \{2, \dots, n-1\},\ j \in \{1, \dots, \min\{i-1, n-i\}\}\big)$$

by u and abuse notation by writing u for the corresponding element of $N(\mathbb{R})$.

Elements of T have the form $s = \mathrm{diag}(t_1, \dots, t_n)$, with $t_i t_{n-i+1} = 1$ for $i \in \big\{1, \dots, \lceil \frac{n}{2} \rceil\big\}$. (Note in particular that when n is odd, we have $t_{(n+1)/2} = 1$.) In the sequel, it will be convenient to use the following alternative coordinates for T. Define the coordinates $(s_1, \dots, s_{\lfloor n/2 \rfloor})$ to be such that $(t_1, \dots, t_{\lfloor n/2 \rfloor})$ is equal to

$$\left(\prod_{i=1}^{\lfloor \frac{n}{2} \rfloor} s_i^{-1}, \prod_{i=2}^{\lfloor \frac{n}{2} \rfloor} s_i^{-1}, \dots, s_{\lfloor \frac{n}{2} \rfloor}^{-1}\right) \qquad \text{if } 2 \nmid n, \tag{24}$$

$$\left(\mathfrak{s}_n^{-1} \cdot \prod_{i=1}^{\frac{n-4}{2}} s_i^{-1},\ \mathfrak{s}_n^{-1} \cdot \prod_{i=2}^{\frac{n-4}{2}} s_i^{-1}, \dots, \mathfrak{s}_n^{-1} \cdot s_{\frac{n-4}{2}}^{-1},\ \mathfrak{s}_n^{-1},\ \mathfrak{s}_n^{-1} \cdot s_{\frac{n-2}{2}}\right) \qquad \text{if } 2 \mid n, \tag{25}$$

where $s_i > 0$ for each i and where $\mathfrak{s}_n := \sqrt{s_{(n-2)/2} s_{(n)/2}}$ when n is even.

We denote the elements of K by θ. With respect to the coordinates u on $N(\mathbb{R})$, s on T, and θ on K, the Haar measure dg on $\mathrm{O}_{\mathcal{A}}(\mathbb{R})$ is given by

$$dg = d\theta\, du(\delta(s)\, d^{\times} s), \tag{26}$$

where

$$du := \prod_{i=1}^{\lceil \frac{n}{2} \rceil} \prod_{j=1}^{i-1} du_{ij}, \quad d^{\times} s := \prod_{i=1}^{\lfloor \frac{n}{2} \rfloor} \frac{ds_i}{s_i}, \quad \delta(s) := \begin{cases} \prod_{i=1}^{\lfloor \frac{n}{2} \rfloor} s_i^{\,i^2 - 2i\lfloor \frac{n}{2} \rfloor} & \text{if } 2 \nmid n, \\ (s_{\frac{n-2}{2}} s_{\frac{n}{2}})^{-\frac{n^2-2n}{8}} \prod_{i=1}^{\frac{n-4}{2}} s_i^{\,i^2 - i(n-1)} & \text{if } 2 \mid n. \end{cases}$$

Above, $d\theta$ is normalized so that $\int_{\theta \in \{\pm \mathrm{id}\} \backslash K} d\theta = 1$, and du is normalized so that $N(\mathbb{Z})$ has covolume 1 in $N(\mathbb{R})$.

3.2. ***A box-shaped fundamental domain for*** $\mathbf{O}_{\mathcal{A}}(\mathbb{Z}) \curvearrowright \mathbf{O}_{\mathcal{A}}(\mathbb{R})$**.** A fundamental domain $\mathcal{F}$ is said to be *box-shaped at infinity* if it can be sandwiched as $\mathcal{S}_1 \subset \mathcal{F} \subset \mathcal{S}_2$, where $\mathcal{S}_1 \subset \mathcal{S}_2$ are nested generalized Siegel sets[5] satisfying the following conditions:

(a) There exists an open subset $\mathcal{U}_1 \subset \mathfrak{S}_1$ of full measure such that every $\mathrm{O}_{\mathcal{A}}(\mathbb{Z})$-orbit on $\mathrm{O}_{\mathcal{A}}(\mathbb{R})$ meets $\mathcal{U}_1$ at most once.

(b) Every $\mathrm{O}_{\mathcal{A}}(\mathbb{Z})$-orbit on $\mathrm{O}_{\mathcal{A}}(\mathbb{R})$ meets $\mathcal{S}_2$ at least once.

(c) The set $\mathcal{S}_2 \smallsetminus \mathcal{S}_1$ is empty "sufficiently high in the cusp," in the sense that, for some $c > 0$, the set $T_c := \{s = (s_1, \dots, s_{\lfloor n/2 \rfloor}) \in T : s_i > c \text{ for all } i\}$ has the property that $\mathcal{S}_1 \cap N T_c K = \mathcal{S}_2 \cap N T_c K$.

[5]Here, a fundamental domain $\mathcal{F} \subset \mathrm{O}_{\mathcal{A}}(\mathbb{R})$ is defined to be a measurable subset such that there exists a subset $\mathcal{M} \subset \mathrm{O}_{\mathcal{A}}(\mathbb{R})$ of full measure with the property that every $g \in \mathcal{M}$ is $\mathrm{O}_{\mathcal{A}}(\mathbb{Z})$-equivalent to a unique element of $\mathcal{F}$. By a generalized Siegel set, we mean a finite union of Siegel sets.

Because they are defined by simple equations in the cusp, box-shaped fundamental domains are particularly amenable to explicit computations. For instance, we use the box-shaped property to evaluate a certain integral that arises in the proof of Theorem 1 (see (35)). As another example of the utility of box-shaped fundamental domains, see [27; 28], where Grenier proved the analogue of Theorem 18 for the group SL_n and remarked that his result could be used to compute certain integrals of Eisenstein series that arise when generalizing Selberg's trace formula to the group $\mathrm{SL}_n(\mathbb{Z})$. Grenier's work has had a number of applications in the literature (see, e.g., [26; 38; 53]), and as explained in Section 3.2 (to follow), it plays a central role in our proof of Theorem 18.

In this subsection, we construct a box-shaped fundamental domain for the action of $\mathrm{O}_{\mathcal{A}}(\mathbb{Z})$ on $\mathrm{O}_{\mathcal{A}}(\mathbb{R})$. Specifically, we prove the following result:

Theorem 18. *There exists a fundamental domain for the action of* $\mathrm{O}_{\mathcal{A}}(\mathbb{Z})$ *on* $\mathrm{O}_{\mathcal{A}}(\mathbb{R})$ *that is box-shaped at infinity.*

Our proof of Theorem 18 occurs over the next five subsubsections and is structured as follows:

- First, in Section 3.2.1, we show that it suffices to construct the nested generalized Siegel sets $\mathcal{S}_1 \subset \mathcal{S}_2$ satisfying the properties (a)–(c) enumerated above.
- Next, in Section 3.2.2, we construct $\mathcal{S}_2$ in terms of a certain compact subset $\mathcal{N} \subset N(\mathbb{R})$, and in Section 3.2.3, we make a convenient explicit choice for the set $\mathcal{N}$.
- It then remains to construct $\mathcal{S}_1$, which we do using the aforementioned work of Grenier. Specifically, in Section 3.2.4, we recall the construction of Grenier's domain, and in Section 3.2.5, we use his result to construct $\mathcal{S}_1$.

3.2.1. *Reduction to constructing* $\mathcal{S}_1$ *and* $\mathcal{S}_2$. The following lemma reduces the problem of constructing the desired fundamental domain $\mathcal{F}$ into the simpler problem of constructing $\mathcal{S}_1$ and $\mathcal{S}_2$:

Lemma 19. *Let* Λ *be a discrete subgroup of a Lie group* $\mathcal{G}$ *and denote by* $\mathcal{B}(\mathcal{G})$ *the Borel* σ*-algebra of* $\mathcal{G}$*. Suppose that* $\mathcal{S}$ *and* $\mathcal{S}'$ *are sets in* $\mathcal{B}(\mathcal{G})$ *with the property that the maps* $\mathcal{S} \to \mathcal{G}/\Lambda$ *and* $\mathcal{S}' \to \mathcal{G}/\Lambda$ *induced by* $s \mapsto s\Lambda$ *are, respectively, injective and surjective. Then there is a fundamental domain* $\mathcal{F}$ *in* $\mathcal{B}(\mathcal{G})$ *for the action of* Λ *on* $\mathcal{G}$ *such that* $\mathcal{S} \subset \mathcal{F} \subset \mathcal{S}'$.

Proof. The argument is analogous to the proof of [35, Lemma 4.1.1]. Since Λ is discrete, we can find a nonempty open subset $U \subset \mathcal{G}$ such that $U^{-1}U \cap \Lambda = \{\mathrm{id}\}$. Since $\mathcal{G}$ is second countable, we can find a sequence of elements $\{g_n\} \subset \mathcal{G}$ such that $\mathcal{G} = \bigcup_{n=1}^{\infty} g_n U$. Let $\mathcal{S}'' = \mathcal{S}' \smallsetminus \mathcal{S}\Lambda$ and set

$$\mathcal{F}' = \bigcup_{n=1}^{\infty} \Big(g_n U \cap \mathcal{S}'' \smallsetminus \bigcup_{i<n} (g_i U \cap \mathcal{S}'')\Lambda \Big).$$

Lastly, define $\mathcal{F} = \mathcal{S} \cup \mathcal{F}'$ and note that this union is disjoint. Then $\mathcal{F} \in \mathcal{B}(\mathcal{G})$ since all the operations used in its construction keep us in the σ-algebra. It is simple to check that the induced map $\mathcal{F} \to \mathcal{G}/\Lambda$ sending x to $x\Lambda$ is bijective. We conclude that $\mathcal{F}$ is a fundamental domain. □

3.2.2. *Constructing $\mathcal{S}_2$.* We now construct $\mathcal{S}_2$ in terms of a certain compact subset $\mathcal{N} \subset N(\mathbb{R})$, to be chosen explicitly in the next subsubsection. This construction is in essence due to Borel and Harish-Chandra (see [6, Section 9.2] and [43, Section 4.1], which specialize the results of [18; 19] for semisimple groups to the case of the group $\mathrm{SO}_{\mathcal{A}}$). By [18, Théorème 2.4 and Exemple 2.5], there exist a constant $c_2 > 0$ and a compact set $\mathcal{N} \subset N(\mathbb{R})$ such that if we take

$$T_2 := \{s = (s_1, \dots, s_{\lfloor n/2 \rfloor}) \in T : s_i > c_2 \text{ for all } i\},$$

then the set

$$\mathcal{S}_2 := \mathcal{N}\, T_2\, (\{\pm \mathrm{id}\} \backslash K) \tag{27}$$

meets every orbit of $\mathrm{O}_{\mathcal{A}}(\mathbb{Z})$ on $\mathrm{O}_{\mathcal{A}}(\mathbb{R})$ at least once (hence satisfying property (b) above), where $\{\pm \mathrm{id}\}\backslash K$ denotes some strict fundamental domain for the action of the group $\{\pm \mathrm{id}\}$ by left-multiplication on K (where "strict" means that every coset of $\{\pm \mathrm{id}\}$ has a unique representative).[6]

3.2.3. *Choosing $\mathcal{N}$.* Having constructed $\mathcal{S}_2$ in terms of $\mathcal{N}$, we now make an explicit choice of $\mathcal{N}$ that will be convenient in what follows. Let $\overline{\mathcal{N}} \subset N(\mathbb{R})$ be the subset defined as

$$\overline{\mathcal{N}} := \begin{cases} \{u \in N(\mathbb{R}) : |u_{ij}| \le 1 \text{ for } i = \lceil \frac{n}{2} \rceil,\ |u_{ij}| \le \frac{1}{2} \text{ for } i \ne \lceil \frac{n}{2} \rceil\} & \text{if } n \text{ is odd,} \\ \{u \in N(\mathbb{R}) : |u_{ij}| \le \frac{1}{2} \text{ for all } i, j\} & \text{if } n \text{ is even.} \end{cases}$$

The following lemma implies that by chopping $\mathcal{N}$ into pieces and translating them via elements of $N(\mathbb{R}) \cap \mathrm{O}_{\mathcal{A}}(\mathbb{Z})$, we can replace $\mathcal{N}$ with a subset of $\overline{\mathcal{N}}$:

Lemma 20. *Let $u \in N(\mathbb{R})$. Then there exists $\bar{u} \in N(\mathbb{R}) \cap \mathrm{O}_{\mathcal{A}}(\mathbb{Z})$ such that $\bar{u}u \in \overline{\mathcal{N}}$. Moreover, there exists an open subset $\mathcal{U}_3 \subset N(\mathbb{R})$ of full measure such that for any $u \in \mathcal{U}_3$ there is precisely one element $\bar{u} \in N(\mathbb{R}) \cap \mathrm{O}_{\mathcal{A}}(\mathbb{Z})$ such that $\bar{u}u \in \overline{\mathcal{N}}$.*

Proof. We construct $\bar{u}$ inductively. Upon inspecting the coordinate system on N provided in Section 3.1, we arrive at the following observation: if $k \in \{0, \dots, n-3\}$ is an integer and $u' \in N(\mathbb{R})$ is an element such that $u'_{ij} = 0$ for all i, j such that $i - j \le k$, then $(u'u)_{ij} = u'_{ij} + u_{ij}$ for all i, j such that $i - j = k + 1$. By the observation, we may choose $u_1 \in N(\mathbb{R}) \cap \mathrm{O}_{\mathcal{A}}(\mathbb{Z})$ such that $|(u_1)_{i(i-1)} + u_{i(i-1)}| \le \frac{1}{2}$ for each $i \in \{2, \dots, \lfloor \frac{n}{2} \rfloor\}$ and such that, for n odd, we have $|(u_1)_{\lceil n/2 \rceil \lfloor n/2 \rfloor} + u_{\lceil n/2 \rceil \lfloor n/2 \rfloor}| \le 1$. Suppose for some $k \in \{0, \dots, n-4\}$ we have chosen $u_\ell \in N(\mathbb{R}) \cap \mathrm{O}_{\mathcal{A}}(\mathbb{Z})$ for each $\ell \in \{1, \dots, k+1\}$; then, by the observation, we may choose $u_{k+2} \in N(\mathbb{R}) \cap \mathrm{O}_{\mathcal{A}}(\mathbb{Z})$ such that $|(u_1)_{i(i-k-2)} + u_{i(i-k-2)}| \le \frac{1}{2}$ for each $i \in \{k+3, \dots, \lfloor \frac{n+k+2}{2} \rfloor\}$, unless n is odd and $i = \lceil \frac{n}{2} \rceil$, in which case we can only arrange for $|(u_1)_{\lceil n/2 \rceil (\lceil n/2 \rceil - k - 2)} + u_{\lceil n/2 \rceil (\lceil n/2 \rceil - k - 2)}| \le 1$. Having constructed u_ℓ for each $\ell \in \{1, \dots, n-2\}$, we then take $\bar{u} = \prod_{\ell=1}^{n-2} u_{n-2-\ell}$.

As for uniqueness on an open subset of full measure, it suffices to show that $\bar{u}$ is unique when u lies in the interior of $\overline{\mathcal{N}}$, for then we can take $\mathcal{U}_3$ to be the union of all $(N(\mathbb{R}) \cap \mathrm{O}_{\mathcal{A}}(\mathbb{Z}))$-translates of the interior

[6] A priori, [18, Théorème 2.4] states that a finite number σ of translates of Siegel sets can be found such that their union meets every orbit of $\mathrm{O}_{\mathcal{A}}(\mathbb{Z})$ on $\mathrm{O}_{\mathcal{A}}(\mathbb{R})/K$ at least once; here, $\sigma = \#(\mathrm{O}_{\mathcal{A}}(\mathbb{Z}) \backslash \mathrm{O}_{\mathcal{A}}(\mathbb{Q}) / \mathcal{P}(\mathbb{Q}))$. But since the algebraic group $\mathrm{O}_{\mathcal{A}}$ has class number 1, it follows from [37, Propositions 5.4 and 5.10] that $\sigma = 1$.

of $\overline{\mathcal{N}}$. So take $u \in \overline{\mathcal{N}}$. If $\bar{u}u \in \overline{\mathcal{N}}$ for some $\bar{u} \in N(\mathbb{R}) \cap \mathrm{O}_{\mathcal{A}}(\mathbb{Z})$, then the aforementioned observation, together with the fact that u lies in the interior of $\overline{\mathcal{N}}$, implies that $\bar{u}_{i(i-1)} = 0$ for each i. Proceeding inductively as we did to prove existence, we find that $\bar{u} = \mathrm{id}$. □

By Lemma 20, we may assume that $\mathcal{N} \subset \overline{\mathcal{N}}$. We next show that $\mathcal{N}$ can be chosen to lie within an even smaller subset of $\overline{\mathcal{N}}$. Let $\Gamma \subset \mathrm{O}_{\mathcal{A}}(\mathbb{Z})$ denote the subgroup of diagonal matrices with integer entries. One readily verifies that Γ satisfies the following properties:

- Γ is a subgroup of K of order $2^{\lceil n/2 \rceil}$ centralizing T;
- Conjugation by elements of Γ defines a group action on $\overline{\mathcal{N}}$ with the property that for any $\rho \in \Gamma$ and $u \in \overline{\mathcal{N}}$, we have $|(\rho \cdot u)_{ij}| = |u_{ij}|$ for all i, j.
- The orbit of every element of $\overline{\mathcal{N}}$ under the action of Γ has a representative u such that for every n we have $u_{i(i-1)} \in \left[0, \frac{1}{2}\right]$ for each $i \in \left\{2, \dots, \left\lfloor \frac{n}{2} \right\rfloor\right\}$ and such that for odd n we have $u_{\lceil n/2 \rceil \lfloor n/2 \rfloor} \in \left[-1, -\frac{1}{2}\right] \cup \left[0, \frac{1}{2}\right]$. The representative u is unique if each u_{ij} lies in the interior of the corresponding interval or union of intervals.

It follows that we can take $\mathcal{N}$ to lie within the subset

$$\widetilde{\mathcal{N}} := \left\{u \in \overline{\mathcal{N}} : u_{i(i-1)} \in \left[0, \tfrac{1}{2}\right] \text{ for each } i \in \left\{2, \dots, \left\lfloor \tfrac{n}{2} \right\rfloor\right\},\ u_{\lceil \frac{n}{2} \rceil \lfloor \frac{n}{2} \rfloor} \in \left[-1, -\tfrac{1}{2}\right] \cup \left[0, \tfrac{1}{2}\right] \text{ for odd } n\right\}.$$

By possibly expanding $\mathcal{N}$, we may in fact choose $\mathcal{N}$ to be equal to $\widetilde{\mathcal{N}}$.

3.2.4. *Grenier's domain.* In the following subsubsection, we shall deduce the construction of the set $\mathcal{S}_1$ from a slight reformulation of Grenier's explicit box-shaped fundamental domain $\mathscr{F}$ for the action of $\mathrm{SL}_n^{\pm}(\mathbb{Z})$ on $\mathrm{SL}_n^{\pm}(\mathbb{R})$ (see [28]). To state this reformulation, we must introduce some notation. Let $\mathscr{N} \subset \mathrm{SL}_n^{\pm}(\mathbb{R})$ denote the subgroup of lower-triangular unipotent matrices; for an element $u \in \mathscr{N}$, we denote by u_{ij} the row-i, column-j entry of u. Let $\overline{\mathscr{N}}$ be the set defined by

$$\overline{\mathscr{N}} := \left\{u \in \mathscr{N} : |u_{ij}| \le \tfrac{1}{2} \text{ for all } i, j\right\}.$$

Let $\mathscr{T} \subset \mathrm{SL}_n^{\pm}(\mathbb{R})$ denote the subgroup of diagonal matrices with positive entries; for an element $s \in \mathscr{T}$, we denote by s_i the quotient of the row-$(i+1)$, column-$(i+1)$ entry of s by the row-i, column-i entry. Let $\mathscr{K} \subset \mathrm{SL}_n^{\pm}(\mathbb{R})$ denote a maximal compact subgroup containing K.

Consider the subset $\mathscr{N}' \subset \overline{\mathscr{N}}$ defined as

$$\begin{aligned}\mathscr{N}' := \Big\{u \in \overline{\mathscr{N}} : u_{i(i-1)} \in \left[0, \tfrac{1}{2}\right] \text{ for each } i \in \left\{2, \dots, \left\lceil \tfrac{n}{2} \right\rceil\right\},\\ u_{i(i-1)} \in \left[-\tfrac{1}{2}, 0\right] \text{ for each } i \in \left\{\left\lceil \tfrac{n}{2} \right\rceil + 2, \dots, n\right\},\\ u_{(\lceil \frac{n}{2} \rceil + 1)\lceil \frac{n}{2} \rceil} \in \left[-\tfrac{1}{2}, 0\right] \text{ if } 2 \nmid n,\ u_{(\frac{n+2}{2})\frac{n}{2}} \in \left[-\tfrac{1}{2}, \tfrac{1}{2}\right] \text{ if } 2 \mid n\Big\}.\end{aligned}$$

Let $\mathscr{T}' := \{s \in \mathscr{T} : s_i > c_1 \text{ for all } i\}$ for a sufficiently large constant $c_1 > c_2$. Let $\varepsilon_n \in \mathscr{K}$ denote the identity matrix when n is odd, and the diagonal matrix whose diagonal entries are given by $\frac{n}{2}$ copies of -1 followed by $\frac{n}{2}$ copies of 1 when n is even, and let $\{\pm \mathrm{id}, \pm\varepsilon_n\}\backslash\mathscr{K}$ denote some strict fundamental domain for the action of the group $\{\pm \mathrm{id}, \pm\varepsilon_n\}$ by left-multiplication on K.

We are now in position to state our reformulation of Grenier's result:

Lemma 21. *For every sufficiently large $c_1 > 0$, the following property holds*: *If*

$$us\theta, u's'\theta' \in \mathscr{N}'\,\mathscr{T}'\,(\{\pm \mathrm{id}, \pm\varepsilon_n\}\backslash\mathscr{K})$$

are $\mathrm{SL}_n^{\pm}(\mathbb{Z})$-equivalent elements such that $s_i, s_i' > c_1$ for all i and such that u, u' lie in the interior of $\mathscr{N}'$, then $us\theta = u's'\theta'$.

Proof. In [27], Grenier constructed a fundamental domain $\overline{\mathscr{F}} \subset \mathscr{N}\mathscr{T}$ for the action of $\mathrm{SL}_n^{\pm}(\mathbb{Z})$ on $\mathrm{SL}_n^{\pm}(\mathbb{R})/\mathscr{K}$. The domain $\overline{\mathscr{F}}$ has the property that no two points in its interior are $\mathrm{SL}_n^{\pm}(\mathbb{Z})$-equivalent. In [28, Theorem 1], Grenier established that, for every sufficiently large $c_1 > 0$, we have $\mathscr{N}''\mathscr{T}' \subset \overline{\mathscr{F}}$, where $\mathscr{N}'' = \left\{u \in \mathscr{N}' : u_{((n+2)/2)(n/2)} \in \left[0, \frac{1}{2}\right]\right\}$. Consequently, the set $\overline{\mathscr{F}}(\{\pm \mathrm{id}\}\backslash\mathscr{K})$ is a fundamental domain for the action of $\mathrm{SL}_n^{\pm}(\mathbb{Z})$ on $\mathrm{SL}_n^{\pm}(\mathbb{R})$ containing $\mathscr{N}''\mathscr{T}'(\{\pm \mathrm{id}\}\backslash\mathscr{K})$. Since $\varepsilon_n\mathscr{N}''\varepsilon_n = \mathscr{N}'$, it follows that there is a fundamental domain for the action of $\mathrm{SL}_n^{\pm}(\mathbb{Z})$ on $\mathrm{SL}_n^{\pm}(\mathbb{R})$ containing $\mathscr{N}'\mathscr{T}'(\{\pm \mathrm{id}, \pm\varepsilon_n\}\backslash\mathscr{K})$. If we have two distinct $\mathrm{SL}_n^{\pm}(\mathbb{Z})$-equivalent elements $us\theta, u's'\theta' \in \mathscr{N}'\mathscr{T}'(\{\pm \mathrm{id}, \pm\varepsilon_n\}\backslash\mathscr{K})$ such that u, u' lie in the interior of $\mathscr{N}'$, then one can find two distinct elements of the interior of $\overline{\mathscr{F}}$ that are $\mathrm{SL}_n^{\pm}(\mathbb{Z})$-equivalent, which is a contradiction. □

3.2.5. *Constructing $\mathcal{S}_1$.* Let $T_1 := \{s \in T : s_i > c_1 \text{ for all } i\}$, where the constant $c_1 > 0$ is to be chosen shortly. We then take

$$\mathcal{S}_1 := \mathcal{N}\,T_1\,(\{\pm \mathrm{id}\}\backslash K).$$

It is evident that $\mathcal{S}_1$ and $\mathcal{S}_2$ satisfy property (c) above. The following lemma states that we can choose c_1 so that $\mathcal{S}_1$ lies within $\mathcal{S}_2$ and satisfies property (a) above:

Lemma 22. *There exists a constant $c_1 > 0$ for which* (1) *$\mathcal{S}_1 \subset \mathcal{S}_2$ and* (2) *there exists an open subset $\mathcal{U}_1 \subset \mathcal{S}_1$ of full measure such that every orbit of $\mathrm{O}_{\mathcal{A}}(\mathbb{Z})$ on $\mathrm{O}_{\mathcal{A}}(\mathbb{R})$ meets $\mathcal{U}_1$ at most once.*

Proof. Let $\mathcal{U}_1$ be an open subset of full measure contained in the interior of $\mathcal{S}_1$ consisting of elements $us\theta$ satisfying the following two properties:

- The stabilizer under left-multiplication by $\mathrm{O}_{\mathcal{A}}(\mathbb{Z})$ of $us\theta \in \mathrm{O}_{\mathcal{A}}(\mathbb{R})/K$ is given by $\{\pm \mathrm{id}\}$.
- There exists a unique element $u_0 \in \mathscr{N} \cap \mathrm{SL}_n^{\pm}(\mathbb{Z})$ such that $u_0 u$ lies in the interior of $\mathscr{N}'$.

To see why we can arrange for the first property above to hold on a open subset of full measure, we use the following lemma:

Lemma 23. *There exists an open subset $\mathcal{U} \subset \mathrm{O}_{\mathcal{A}}(\mathbb{R})/K$ of full measure such that the stabilizer in $\mathrm{O}_{\mathcal{A}}(\mathbb{Z})$ of any $g \in \mathcal{U}$ is given by $K \cap \{\pm \mathrm{id}\}$.*

Proof of Lemma 23. Let $\overline{\mathcal{F}}$ be any fundamental domain for $\mathrm{O}_{\mathcal{A}}(\mathbb{Z})$ on $\mathrm{O}_{\mathcal{A}}(\mathbb{R})/K$, and let g be an element of the interior of $\overline{\mathcal{F}}$. If $\gamma \in \mathrm{O}_{\mathcal{A}}(\mathbb{Z})$ stabilizes g, then there is an open neighborhood $U \ni g$ contained in the interior of $\overline{\mathcal{F}}$ such that $\gamma \cdot U$ is contained in the interior of $\overline{\mathcal{F}}$, implying in fact that γ stabilizes every element of U. Since left-multiplication by γ defines a real-analytic function on $\mathrm{O}_{\mathcal{A}}(\mathbb{R})/K$, and since

$\mathrm{O}_{\mathcal{A}}(\mathbb{R})/K$ is connected (as K meets both of the two connected components of $\mathrm{O}_{\mathcal{A}}(\mathbb{R})$), it follows that γ stabilizes all of $\mathrm{O}_{\mathcal{A}}(\mathbb{R})/K$.

Now, the stabilizer in $\mathrm{O}_{\mathcal{A}}(\mathbb{R})$ of any element $h \in \mathrm{O}_{\mathcal{A}}(\mathbb{R})/K$ is given by hKh^{-1}. Since γ stabilizes all of $\mathrm{O}_{\mathcal{A}}(\mathbb{R})/K$, it follows that $\gamma \in \mathcal{K} := \bigcap_{h\in \mathrm{O}_{\mathcal{A}}(\mathbb{R})} hKh^{-1}$. But because $\mathcal{K}$ is a compact normal subgroup of $\mathrm{O}_{\mathcal{A}}(\mathbb{R})$, it follows that $\mathcal{K}$ is discrete. Let $\mathcal{K}^+$ denote the intersection of $\mathcal{K}$ with the identity component $\mathrm{O}_{\mathcal{A}}(\mathbb{R})^+$ of $\mathrm{O}_{\mathcal{A}}(\mathbb{R})$. Then $\mathcal{K}^+$ is a discrete normal subgroup of the connected group $\mathrm{O}_{\mathcal{A}}(\mathbb{R})^+$, and so $\mathcal{K}^+$ is central in $\mathrm{O}_{\mathcal{A}}(\mathbb{R})^+$. It follows that $\mathcal{K}^+ \subset \{\pm \mathrm{id}\}$, and hence that $\mathcal{K}^+$ is central in $\mathrm{O}_{\mathcal{A}}(\mathbb{R})$. If $\gamma' \in \mathcal{K} \smallsetminus \mathcal{K}^+$, then γ' commutes with elements of both connected components of $\mathrm{O}_{\mathcal{A}}(\mathbb{R})$, so γ' is central in $\mathrm{O}_{\mathcal{A}}(\mathbb{R})$ since $\mathcal{K}$ is normal. Thus, $\mathcal{K} \subset \{\pm \mathrm{id}\}$, and so $\gamma = \pm \mathrm{id}$.

Finally, let $\mathcal{U}$ be the union of all $\mathrm{O}_{\mathcal{A}}(\mathbb{Z})$-translates of the interior of $\overline{\mathcal{F}}$. Then $\mathcal{U}$ is an open subset of full measure, and we have shown above that the stabilizer in $\mathrm{O}_{\mathcal{A}}(\mathbb{Z})$ of any $g \in \mathcal{U}$ is contained in $\{\pm \mathrm{id}\}$, as desired. This completes the proof of Lemma 23. □

As for the second property, observe that by the definitions of $\mathcal{N}$ and $\mathcal{N}'$, the desired element u_0 must be such that $(u_0)_{i(i-1)} = 0$ for each i. Then, by proving the analogue of Lemma 20 for the group $\mathrm{SL}_n^{\pm}$, one finds that there exists at least one $u_0 \in \mathcal{N} \cap \mathrm{SL}_n^{\pm}(\mathbb{Z})$ such that $u_0 u \in \mathcal{N}'$. By our explicit characterization of the elements of $\mathcal{N}$ (see (22) and (23)), the row-i, column-j entry of u is a nonconstant polynomial in the unipotent coordinates for every pair (i, j) with $i > j + 1$. Thus, there exists an open subset $\mathcal{U}_2 \subset \mathcal{N}$ of full measure such that for any $u \in \mathcal{U}_2$ and $u_0 \in \mathcal{N} \cap \mathrm{SL}_n^{\pm}(\mathbb{Z})$, the row-$i$, column-$j$ entry of $u_0 u$ is not an integer multiple of $\frac{1}{2}$ for every pair (i, j) with $i > j$, unless $i + 1 = j = \frac{n}{2}$, in which case $u_{ij} = 0$. In particular, if for $u \in \mathcal{U}_2$ and $u_0 \in \mathcal{N} \cap \mathrm{SL}_n^{\pm}(\mathbb{Z})$ we have $u_0 u \in \mathcal{N}'$, then $u_0 u$ must in fact lie in the interior of $\mathcal{N}'$, and imitating the proof of uniqueness in Lemma 20 yields that u_0 must be unique.

Having defined $\mathcal{U}_1$, take any $us\theta, u's'\theta' \in \mathcal{U}_1$ such that $g \cdot us\theta = u's'\theta'$ for some $g \in \mathrm{O}_{\mathcal{A}}(\mathbb{Z})$. Let $\varepsilon, \varepsilon' \in \{\pm \mathrm{id}, \pm\varepsilon_n\}$ be such that $\varepsilon\theta, \varepsilon'\theta' \in \{\pm \mathrm{id}, \pm\varepsilon_n\} \backslash \mathscr{K}$. Let $u_0, u_0' \in \mathcal{N} \cap \mathrm{SL}_n^{\pm}(\mathbb{Z})$ be the unique elements such that $u_0(\varepsilon u \varepsilon), u_0'(\varepsilon' u' \varepsilon')$ lie in the interior of $\mathcal{N}'$. Since $s_i, s_i' > c_1$ for all i, it follows from Lemma 21 that $u_0(\varepsilon u \varepsilon) s(\varepsilon\theta) = u_0'(\varepsilon' u' \varepsilon') s'(\varepsilon'\theta')$. By the uniqueness of the Iwasawa decomposition, we must have

$$u_0(\varepsilon u \varepsilon) = u_0'(\varepsilon' u' \varepsilon'), \quad s = s', \quad \varepsilon\theta = \varepsilon'\theta'. \tag{28}$$

The third equality in (28) implies that $\varepsilon\varepsilon' \in K$, so since $\varepsilon_n \notin K$ when n is even, it follows that $\varepsilon\varepsilon' = \pm \mathrm{id}$. But $(\{\pm \mathrm{id}\}) \backslash K$ contains either θ or $-\theta$ and not both, meaning that $\varepsilon\varepsilon' = \mathrm{id}$. Combining this with the first equality in (28) yields that u is a translate of u' by an element of $\mathcal{N} \cap \mathrm{SL}_n^{\pm}(\mathbb{Z})$, and hence by an element of $\mathcal{N} \cap \mathrm{O}_{\mathcal{A}}(\mathbb{Z})$. The uniqueness statement in Lemma 20 then implies that $u = u'$. We conclude that $us\theta = us'\theta'$.

This completes the proof of Lemma 22, and hence also that of Theorem 18. □

3.3. ***Fundamental sets for $G(\mathbb{R}) \curvearrowright W(\mathbb{R})$ and $G(\mathbb{Z}) \curvearrowright W(\mathbb{R})$.*** Let r and s be nonnegative integers with $r + 2s = n$. We define $U(\mathbb{R})^{(r)}$ to be the set of monic polynomial $f(x)$ of degree n with r distinct real roots and s distinct pairs of complex conjugate roots. We let $W(\mathbb{R})^{(r)}$ denote $\mathrm{inv}^{-1}(U(\mathbb{R})^{(r)})$, and let $W(\mathbb{R})^{(r),\mathrm{red}}$ be the set of elements in $W(\mathbb{R})^{(r)}$ that are reducible over $\mathbb{R}$. Let f be an element in $U(\mathbb{R})^{(r)}$. From [6, Section 9] and [43, Section 4], we know that the set $\{B \in W(\mathbb{R})^{(r),\mathrm{red}} : \mathrm{inv}(B) = f\}$ consists of

a single $G(\mathbb{R})$-orbit, and also that the quantity

$$\theta_r := \#\,\mathrm{Stab}_{G(\mathbb{R})}(B) \tag{29}$$

is independent of the choice of $B \in W(\mathbb{R})^{(r)}$.

We next exhibit an explicit representative of the reducible $G(\mathbb{R})$-orbit of $\mathrm{inv}^{-1}(f)$ for polynomials $f \in U(\mathbb{R})$. Denote the coefficients of f by $f(x) = x^n + f_1 x^{n-1} + \cdots + f_n$, and define $\sigma_0 : U(\mathbb{R}) \to W_0(\mathbb{R})$ by

$$\sigma_0(f) := \begin{bmatrix}
 & & & & & & 1 & 0 \\
 & & & & & \iddots & 0 & \\
 & & & & 1 & & & \\
 & & & 1 & 0 & & & \\
 & & 1 & f_1 & -\frac{f_2}{2} & & & \\
 & 1 & 0 & -\frac{f_2}{2} & -f_3 & \ddots & & \\
 & \iddots & & & \ddots & \ddots & -\frac{f_{n-3}}{2} & \\
1 & 0 & & & & -\frac{f_{n-3}}{2} & -f_{n-2} & -\frac{f_{n-1}}{2} \\
0 & & & & & & -\frac{f_{n-1}}{2} & -f_n
\end{bmatrix}$$

if n is odd, and by

$$\sigma_0(f) := \begin{bmatrix}
 & & & & & & & 1 & 0 \\
 & & & & & & \iddots & 0 & \\
 & & & & & 1 & & & \\
 & & & & 1 & 0 & & & \\
 & & & 1 & -\frac{f_1}{2} & & & & \\
 & & 1 & -\frac{f_1}{2} & \frac{f_1^2}{4} - f_2 & -\frac{f_3}{2} & & & \\
 & 1 & 0 & & -\frac{f_3}{2} & -f_4 & \ddots & & \\
 & \iddots & & & & \ddots & \ddots & -\frac{f_{n-3}}{2} & \\
1 & 0 & & & & & -\frac{f_{n-3}}{2} & -f_{n-2} & -\frac{f_{n-1}}{2} \\
0 & & & & & & & -\frac{f_{n-1}}{2} & -f_n
\end{bmatrix}$$

if n is even. These sections are constructed and used in [16, equations (11) and (27)]. It is easy to check that we have $\mathrm{inv}(\sigma_0(f)) = f$, and so σ_0 is indeed a section. This explicit section σ_0 is useful in the proof of Proposition 27 (to follow), for its image consists of matrices whose entries are polynomials in the f_i. However, in Section 5.2 (also to follow), we shall require a section σ with image consisting of elements B whose entries are $O(\mathrm{H}(B))$. To this end, we rescale σ_0: given $f \in U(\mathbb{R})$ having height Y and nonzero discriminant, set $f_1(x) := f(x/Y)$. Then $\mathrm{H}(f_1) = 1$, and we define the section σ by $\sigma(f) := \mathrm{H}(f)\sigma(f_1)$. It is easy to check that σ is also a section, and that the coefficients of B in the image of σ are bounded by $O(\mathrm{H}(B))$. We now define $\mathcal{R}^{(r)}$ to be $\sigma(U(\mathbb{R})^{(r)}) \subset W(\mathbb{R})^{(r),\mathrm{red}}$.

Finally, we have the following result, which follows immediately from an argument parallel to [9, Section 2.1]:

Proposition 24. *The multiset $\mathcal{F}\cdot\mathcal{R}^{(r)}$ is a cover of a fundamental domain for the action of $G(\mathbb{Z})$ on $W(\mathbb{R})^{(r),\mathrm{red}}$. More precisely, every $G(\mathbb{Z})$-orbit of $B\in W(\mathbb{R})^{(r),\mathrm{red}}$ is represented exactly $\theta_r/\#\operatorname{Stab}_{G(\mathbb{R})}(B)$ times in $\mathcal{F}\cdot\mathcal{R}^{(r)}$.*

4. The action of the subgroup $\mathcal{P}$ on the reducible hyperplane W_0

In this section, we examine the action of the lower-triangular (parabolic) subgroup $\mathcal{P}$ on the reducible hyperplane W_0. Specifically, in Section 4.1, we show that over many interesting base rings R, the action of $\mathcal{P}(R)$ on the set of elements in $W_0(R)$ lying over a given nondegenerate invariant polynomial is simply transitive. Then, in Section 4.2, we prove a Jacobian change-of-variables formula relating the Euclidean measure on $W_0(R)$ to the product of the Haar measure on $\mathcal{P}(R)$ with the Euclidean measure on $U(R)$. This formula will be applied in Sections 5.3–5.4 and 6.3.

4.1. ***Reduction theory for the action of $\mathcal{P}$ on W_0.*** Let R be either a field or $\mathbb{Z}_p$ for some prime p. Then, by analogy with Proposition 9, we have the following result, which classifies the orbits and stabilizers of $\mathcal{P}(R)$ on $W_0(R)$:

Proposition 25. *Let R be as above, and let $f\in U(R)$ be an element with unit discriminant. Then $\mathrm{inv}^{-1}(f)\cap W_0(R)$ consists of a single $\mathcal{P}(R)$-orbit, and the stabilizer of any element of this unique orbit is trivial.*

Proof. We first prove the transitivity claim for *any* integral domain R in which 2 is invertible (this includes every field of characteristic not 2, as well as $\mathbb{Z}_p$ for odd primes p). Let $B=[b_{ij}]$ be any element in $\mathrm{inv}^{-1}(f)\cap W_0(R)$. The idea of the proof is to show that B can be transformed under the group $\mathcal{P}(R)$ into an element of the image of the section σ_0, which is defined over R because we assumed that $2\in R^\times$. Since distinct elements in the image of σ_0 have distinct invariants, it will follow that B is $\mathcal{P}(R)$-equivalent to $\sigma_0(f)$, as necessary.

We now translate B into the image of σ_0. First, notice that each $b_{i(n-i)}$ is a unit in R: indeed, the discriminant of f is a unit in R, and each $b_{i(n-i)}$ divides the discriminant of f. Thus, we may use the action of the diagonal matrices in $\mathcal{P}(R)$ to transform each $b_{i(n-i)}$ into 1, and we may assume in what follows that $b_{i(n-i)}=1$ for each i.

Next, using the coefficients of N given in (22) and (23), we abuse notation by denoting, for any $v_{ij}\in R$, an element in $N(R)$ called $\tilde{u}_{ij}$, whose u_{ij}-coefficient is v_{ij}, and whose other coefficients are 0. We successively replace B by $N(R)$-translates of itself by means of the following steps:

(1) Let $v_{21}\in R$ be such that $b_{1n}+\tilde{u}_{21}=0$. In other words, $v_{21}=-b_{1n}$. We redefine B to be $\tilde{u}_{21}\cdot B$. We now have $b_{1n}=0$.

(2) Set $v_{32}:=-b_{2(n-1)}$, and redefine B to be $\tilde{u}_{32}\cdot B$. Next, set $v_{31}:=-b_{2n}$ and redefine B to be $\tilde{u}_{31}\cdot B$. We now have $b_{2(n-1)}=b_{2n}=0$.

(k) Let $k\in\{3,\dots,n-2\}$, and suppose that we have transformed the first $k-1$ rows (and hence the first $k-1$ columns) of B into the required form. We now explain how to transform the k-th row (and column)

of B into what is needed; i.e., we explain how to clear out the entries $b_{k(n-j)}$ for j in the appropriate range. First set

$$v_{(k+1)\min\{k,n-k-1\}} := -b_{k\max\{k+2,n-k+1\}}$$

and redefine B to be $\tilde{u}_{(k+1)\min\{k,n-k-1\}} \cdot B$. Next set

$$v_{(k+1)\min\{k-1,n-k-2\}} := -b_{k\max\{k+3,n-k+2\}}$$

and redefine $B$to be $\tilde{u}_{(k+1)\min\{k-1,n-k-2\}} \cdot B$. These two steps clear out $b_{k\max\{k+2,n-k+1\}}$ and $b_{k\max\{k+3,n-k+2\}}$. Continuing in this manner, let $k' \in \{3,\dots,\min\{k-1,n-k-2\}\}$, and suppose that we have cleared out b_{kj} for $j \in \max\{k+2,n-k+1\},\dots,\max\{k+k'+1,n-k+k'\}$. Then set

$$v_{(k+1)\min\{k-k',n-k-k'-1\}} := b_{k\max\{k+k'+2,n-k+k'+1\}},$$

and redefine B to be $\tilde{u}_{(k+1)\min\{k-k',n-k-k'-1\}} \cdot B$. This process has now cleared out the all of the required entries in the k-th row of B.

Having transformed the first $n-2$ rows (and therefore columns) of B into the required form, note that the last two rows are already as required. We have thus replaced B by an $\mathcal{P}(R)$-translate to ensure that B lies in the image of σ_0, as desired.

We now handle the case where $R = \mathbb{Z}_2$. In this case, when n is even, the argument given above works without change. But it does not quite work when n is odd, because every element $u \in N(\mathbb{Z}_2)$ has the property that $u_{\lceil n/2\rceil j}$ is divisible by 2 for each $j \in \left\{1,\dots,\left\lfloor\frac{n}{2}\right\rfloor\right\}$ (see (22)). Thus, in step $\left(\left\lfloor\frac{n}{2}\right\rfloor\right)$ of the process outlined above, the action of $N(\mathbb{Z}_2)$ can only be used to make the entries $b_{\lfloor n/2\rfloor j}$ equal to either 0 or 1. Consequently, once all the steps have been completed, the resulting matrix B may fail to lie in the image of $\sigma_0(f)$, but only because some of the entries $b_{\lfloor n/2\rfloor j}$ might be equal to 1. Now, it is easy to verify that such a matrix B has $b_{\lfloor n/2\rfloor j} = 1$ if and only if the $x^{2n-2j+1}$-coefficient of f is odd. Thus, the arrangement of 0's and 1's in the $\left\lfloor\frac{n}{2}\right\rfloor$-th row of B is uniquely determined by the invariants, and the claim follows.

We finally handle the case where R is a field of characteristic 2. In this case, when n is even, every element of $W_0(R)$ has discriminant zero, and so the claim is moot. When n is odd, the argument used in the case $R = \mathbb{Z}_2$ works without change.

The claim regarding the stabilizers of elements in $W_0(R)$ follows by inspection. Indeed, any element of $\mathcal{P}(R)$ stabilizing B must fix each $b_{i(n-i)}$ and hence must lie in the subgroup $N(R) \subset \mathcal{P}(R)$; moreover, any element of $u \in N(R)$ can be factored as $u = \prod \tilde{u}_{ij}$, and it follows from the points enumerated above that u stabilizes B if and only if each $\tilde{u}_{ij}$ stabilizes B, which happens if and only if each $\tilde{u}_{ij}$ is the identity matrix. (Note: this stabilizer computation works over any integral domain R.) □

By analogy with Lemma 10, the above result has the following immediate consequence by specializing to the case $R = \mathbb{R}$:

Lemma 26. *Let $f \in U(\mathbb{R})$ be nondegenerate. Then the set $\{B \in \mathrm{inv}^{-1}(f)_0 : b_{k(n-k)}(B) > 0$ for each $k\}$ consists of a single $N(\mathbb{R})T$-orbit.*

4.2. ***A Jacobian change-of-variables formula.*** Proposition 25 implies that when $R = \mathbb{R}$ or $\mathbb{Z}_p$ for a prime p, the space $W_0(R)$ is a fibration over $U(R)$, where the generic fiber can be identified with $\mathcal{P}(R)$, so long as it is nonempty. Thus, the Euclidean measure on $W_0(R)$ should be related to the product of the Haar measure on $\mathcal{P}(R)$ with the Euclidean measure on $U(R)$. By analogy with Proposition 11, the following result shows that the relationship between these measures is governed by the polynomial function λ on W_0 defined as in (2):

Proposition 27. *Let $R = \mathbb{R}$ or $\mathbb{Z}_p$ for some prime p. Let $\phi : W_0(R) \to \mathbb{R}$ be a measurable function. Then there exists a nonzero rational number $\mathcal{J} \in \mathbb{Q}^\times$ such that*

$$\int_{B \in W_0(R)} \phi(B)|\lambda(B)|\, dB = |\mathcal{J}| \int_{\substack{f \in U(R) \\ \operatorname{disc}(f) \neq 0}} \Bigg(\sum_{B \in \frac{\operatorname{inv}^{-1}(f)_0}{\mathcal{P}(R)}} \int_{h \in \mathcal{P}(R)} \phi(h \cdot B)\, dh \Bigg)\, df,$$

where dB and df are Euclidean measures, where dh is the right Haar measure on $\mathcal{P}$ given by $dh = \delta(s)^{-1}\, du\, d^\times s$, and where $|-|$ denotes the usual absolute value on R.

Remark. Unlike in Proposition 11, which concerns a representation of small dimension, we cannot prove Proposition 27 by means of a direct computation. Instead, we follow the general four-step strategy used to prove [9, Propositions 3.7, 3.10], which establish an analogous change-of-variables formula in a simpler setting, where the group $\mathcal{P}$ is replaced by a semisimple group, and the factor of $|\lambda(B)|$ is not present. The first step is to note that a similar equation holds, where the Jacobian constant $\mathcal{J}$ is replaced by a Jacobian function that a priori depends on the section σ, the group element h, and the invariant f. Second, the Jacobian function is shown to be independent of the group element h using left-invariance of the Haar measure (which does not hold in our case). Third, the Jacobian function is shown to be independent of the section σ using right-invariance of Haar measure (which does hold in our case, and the independence on the section follows in exactly the same way). Finally, once independence of σ has been established, the section can be chosen to be a polynomial map, from which independence on the invariant can be easily deduced by comparing degrees and dimensions of the invariants and spaces involved. This final step also goes through without change for us.

Thus, the main difference in our case is that the Haar measure dh is not left-invariant. As we demonstrate below, the extra factor of $|\lambda(B)|$ captures how the volumes of sets in $W_0(R)$ transform under left-translation by group elements and therefore compensates for the failure of the measure dh to be left-invariant.

Proof of Proposition 27. Let $\mathcal{U} \subset U(\mathbb{R})$ be an open set, and let $\sigma : \mathcal{U} \to W_0(\mathbb{R})$ be a continuous section (such as the section σ_0 constructed in Section 3.3) with respect to inv. We first claim that we have

$$\int_{B \in \mathcal{P}(\mathbb{R}) \cdot \sigma(\mathcal{U})} \phi(B)|\lambda(B)|\, dB = |\mathcal{J}| \int_{f \in \mathcal{U}} \int_{h \in \mathcal{P}(\mathbb{R})} \phi(h \cdot \sigma(f))\, dh\, df \tag{30}$$

for some nonzero rational constant $\mathcal{J} \in \mathbb{Q}^\times$. We prove (30) in a series of steps. First, by the Stone–Weierstrass theorem, we may assume that σ is piecewise analytic, in which case we have

$$\int_{B \in \mathcal{P}(\mathbb{R}) \cdot \sigma(\mathcal{U})} \phi(B)|\lambda(B)|\, dB = \int_{f \in \mathcal{U}} \int_{h \in \mathcal{P}(\mathbb{R})} |\mathcal{J}_\supset(h, f)| \phi(h \cdot \sigma(f))\, dh\, df,$$

where $\mathcal{J}_\sigma(h, f)$ denotes the determinant of the Jacobian matrix arising from the change-of-variables taking the measure $\lambda(B)\,dB$ on to the product measure $dh\,df$. Note in particular that the function $\mathcal{J}_\sigma$ is piecewise continuous in both arguments h and f.

Second, we show that $\mathcal{J}_\sigma(h, f)$ is independent of h. To do this, fix $\gamma \in \mathcal{P}(\mathbb{R})$, and consider the transformation on $W_0(\mathbb{R})$ sending B to $\gamma \cdot B$. Then there is a character $\chi_\lambda : \mathcal{P}(\mathbb{R}) \to \mathbb{R}_{>0}$ such that $\lambda(\gamma \cdot B)d(\gamma \cdot B) = \chi_\lambda(\gamma)\lambda(B)dB$; note that χ_λ exists because $\lambda(B)$ is a product of coefficients that are unchanged by the action of $N(\mathbb{R})$. In fact, writing $\gamma = tu = su$, the elementary computation

$$\begin{aligned}\lambda(\gamma \cdot B)\, d(\gamma \cdot B) &= \prod_{i=1}^{\lfloor \frac{n}{2} \rfloor} (t_i/t_{i+1})^{1-2i} \lambda(B) t_1^{n-1} \prod_{i=2}^{\lfloor \frac{n}{2} \rfloor} t_i^{n-2i+2}\, dB \\ &= t_1^{n-2} \prod_{i=2}^{\lfloor \frac{n}{2} \rfloor} t_i^{n-2i} \lambda(B)\, dB = \delta(s)^{-1} \lambda(B)\, dB\end{aligned}$$

for odd n (and a similar one for even n) reveals that $\chi_\lambda(\gamma) = \delta(s)^{-1}$ for $\gamma = su$. On the other hand, the transformation $B \mapsto \gamma \cdot B$ acts on $\mathcal{P}(\mathbb{R}) \times \mathcal{U}$ by sending (h, f) to $(\gamma h, f)$. Letting $\rho : \mathcal{P}(\mathbb{R}) \to \mathbb{R}_{>0}$ denote the modulus for the left action of $\mathcal{P}$ on the right Haar measure dh, we see that

$$\mathcal{J}_\sigma(\gamma h, f)\, d(\gamma h)\, df = \rho(\gamma) \mathcal{J}_\sigma(\gamma h, f)\, dh\, df. \tag{31}$$

By definition, $\rho(\gamma) = \delta(s)^{-1}$ for $\gamma = su$, and so we have $\rho(\gamma) = \chi_\lambda(\gamma)$. Therefore, we have

$$\mathcal{J}_\sigma(\gamma h, f)\, d(\gamma h)\, df = \lambda(\gamma B)\, dB = \chi_\lambda(\gamma)\lambda(B)\, dB = \chi_\lambda(\gamma) \mathcal{J}_\sigma(h, f)\, dh\, df. \tag{32}$$

Comparing (31) and (32), we see that $\mathcal{J}_\sigma(h, f) = \mathcal{J}(f)$ is independent of h.

Third, that $\mathcal{J}_\sigma(h, f)$ is independent of σ follows from an argument identical to Step 2 in the proof of [9, Proposition 3.10]; this step relies crucially on the fact that the measure dh is right-invariant. Thus, we can take σ to be the polynomial section σ_0 defined in Section 3.3. Having made this choice of section, that $\mathcal{J}_{\sigma_0}(h, f)$ is independent of f and given by a nonzero rational constant follows from an argument identical to Steps 3 and 4 in the proof of [9, Proposition 3.10].

We have therefore proven (30). Proposition 27 now follows from (30) and the principle of permanence of identities in a manner identical to how [9, Proposition 3.7] is deduced from [9, Proposition 3.10]. □

We conclude this section by computing the value of the Jacobian constant $|\mathcal{J}| \in \mathbb{Q}^\times$ that arises in Proposition 27:

Proposition 28. *The value of $|\mathcal{J}|$ is 1 when n is odd and $2^{-n/2}$ when n is even.*

Proof. To compute $|\mathcal{J}|$, it suffices to compute $|\mathcal{J}|_p$ for each p since $\mathcal{J} \in \mathbb{Q}^\times$. To do this, we construct convenient sets in $W_0(\mathbb{Z}_p)$ whose volumes are computed in two different ways: first, using Proposition 27, and second, via an $\mathbb{F}_p$-point count. Equating the two answers yields the value of $|\mathcal{J}|_p$.

Case 1: $p > 2$. Fix a nondegenerate polynomial $f \in U(\mathbb{F}_p)$, and let $\phi_p : W_0(\mathbb{Z}_p) \to \mathbb{R}$ be the indicator function of the set

$$\Sigma := \{B \in W_0(\mathbb{Z}_p) : \mathrm{inv}(B) \equiv f \pmod p\}.$$

By Proposition 25, the group $\mathcal{P}(\mathbb{Z}_p)$ acts simply transitively on the set of elements in Σ having any fixed invariant polynomial. Hence, from Proposition 27, we obtain

$$\mathrm{Vol}(\Sigma) = |\mathcal{J}|_p \cdot \mathrm{Vol}(\mathcal{P}(\mathbb{Z}_p)) \int_{\substack{g \in U(\mathbb{Z}_p) \\ g \equiv f \pmod p}} dg = |\mathcal{J}|_p \cdot \mathrm{Vol}(\mathcal{P}(\mathbb{Z}_p)) \cdot p^{-\dim U}. \tag{33}$$

On the other hand, Proposition 25 also implies that the group $\mathcal{P}(\mathbb{F}_p)$ acts simply transitively on the mod-p reduction $\overline{\Sigma}$ of Σ. Thus, we have

$$\#\overline{\Sigma} = \#\mathcal{P}(\mathbb{F}_p). \tag{34}$$

Since $\mathrm{Vol}(\Sigma) = p^{-\dim W_0} \cdot \#\overline{\Sigma}$, $\mathrm{Vol}(\mathcal{P}(\mathbb{Z}_p)) = p^{-\dim \mathcal{P}} \cdot \#\mathcal{P}(\mathbb{F}_p)$, and $\dim \mathcal{P} + \dim U = \dim W_0$, we obtain from (33) and (34) that $|\mathcal{J}|_p = 1$ for all odd primes p.

Case 2: $p = 2$. The proof here is similar to Case 1, so we highlight the differences. Pick an integer $m \gg 1$, and set $q = 2^m$. This time, we pick the polynomial $f(x) = x^n \in U(\mathbb{Z}/q\mathbb{Z})$, and we define the set

$$\Sigma := \left\{B \in W_0(\mathbb{Z}_2) : |b_{i(n-i)}(B)|_2 = 1 \text{ for all } i \in \left\{1, \dots, \left\lfloor \tfrac{n}{2} \right\rfloor\right\},\ \mathrm{inv}(B) \equiv f \pmod q\right\}.$$

As before, we obtain

$$\mathrm{Vol}(\Sigma) = |\mathcal{J}|_2 \cdot \mathrm{Vol}(\mathcal{P}(\mathbb{Z}_2)) \cdot q^{-\dim U}.$$

However, the situation over $\mathbb{Z}/q\mathbb{Z}$ is more complicated. Here, the mod-q reduction $\overline{\Sigma}$ of Σ breaks up into $2^{\lfloor n/2 \rfloor}$ different $\mathcal{P}(\mathbb{Z}/q\mathbb{Z})$-orbits. Indeed, the $\left\lfloor \frac{n}{2} \right\rfloor$ different coefficients labeled $-\frac{f_i}{2}$ in the image of $\sigma_0(f)$ in Section 3.3 can be taken to be either 0 or $\frac{q}{2}$, and this gives exactly $2^{\lfloor n/2 \rfloor}$ different elements that are inequivalent under the action of $\mathcal{P}(\mathbb{Z}/q\mathbb{Z})$. Therefore, this time we have

$$\#\overline{\Sigma} = 2^{\lfloor \frac{n}{2} \rfloor} \cdot \#\mathcal{P}(\mathbb{Z}/q\mathbb{Z}).$$

As before, we have $\mathrm{Vol}(\Sigma) = q^{-\dim W_0} \cdot \#\overline{\Sigma}$, and it is easy to check that we have $\mathrm{Vol}(\mathcal{P}(\mathbb{Z}_2)) = 2^{\lfloor n/2 \rfloor} q^{-\dim \mathcal{P}} \cdot \#\mathcal{P}(\mathbb{Z}/q\mathbb{Z})$ when n is odd and $\mathrm{Vol}(\mathcal{P}(\mathbb{Z}_2)) = q^{-\dim \mathcal{P}} \cdot \#\mathcal{P}(\mathbb{Z}/q\mathbb{Z})$ when n is even. It follows that $|\mathcal{J}|_2 = 1$ when n is odd and $|\mathcal{J}|_2 = 2^{n/2}$ when n is even. □

5. Counting reducible $G(\mathbb{Z})$-orbits on $W(\mathbb{Z})$

Let $n \geq 3$, r, and s be nonnegative integers with $r + 2s = n$.[7] In this section, we obtain asymptotics for the number of reducible orbits of $G(\mathbb{Z})$ on $W(\mathbb{Z})^{(r)}$ of bounded height, thereby proving Theorems 1, 3, and 4 using Method I. The proofs using Method II are given in the next section.

To simplify the exposition in the rest of this section, we introduce the following notation:

• For any set $S \subset W(\mathbb{Z})$, let $S_{\mathrm{red}} \subset S$ be the subset of reducible elements of S; for $X > 0$, let $S_X := \{B \in S : \mathrm{H}(B) < X\}$; and as before, let $S_0 := S \cap W_0(\mathbb{Z})$ be the set of elements of S that lie on the reducible hyperplane.

[7] Note that definitions of quantities introduced in what follows may implicitly depend on r.

• Let $G_0 \subset G(\mathbb{R})$ be a fixed nonempty open bounded set such that $G_0^{-1} = G_0$ and G_0 is left- and right-K-invariant. As explained in Section 2.3, such a set can be constructed by starting with a nonempty open bounded set G_0' and taking $G_0 = K(G_0' \cup G_0'^{-1})K$.

• Define the multiset $\mathcal{B}_\infty$ by

$$\mathcal{B}_\infty := G_0 \cdot \mathcal{R}^{(r)} \cap W_0(\mathbb{R}).$$

Set $\mathcal{B} := (\mathcal{B}_\infty)_1$, and note that by the construction of $\mathcal{R}$, we have $(\mathcal{B}_\infty)_X = X\mathcal{B}$.

• We define the quantity $C(\mathcal{B})$ by

$$C(\mathcal{B}) := \frac{1}{\tilde{\theta}_r \operatorname{Vol}(G_0)} \cdot \int_{B \in \mathcal{B}} |\lambda(B)|\, dB,$$

where the volume of G_0 is computed using the Haar measure dg, and where $\tilde{\theta}_r := 2^{\lceil n/2 \rceil}\theta_r$, with θ_r as defined in (29).

• For a finite set Σ of $G(\mathbb{Z})$-orbits on $W(\mathbb{Z})$, let $\#'\,\Sigma$ be the number of elements of Σ, where each $B \in \Sigma$ is counted with weight $1/\#\operatorname{Stab}_{G(\mathbb{Z})}(B)$.

This section is organized as follows. After setting up Bhargava's averaging method in Section 5.1, we reduce the problem of counting reducible $G(\mathbb{Z})$-orbits on $W(\mathbb{Z})$ to a question of counting integer points in various regions of $W_0(\mathbb{Z})$. This is accomplished in Proposition 29 by combining previously obtained "main ball" counting estimates. The advantage of this result over simply applying the averaging method over the nonreductive group $\mathcal{P}$ is that the integral in the right-hand side of (36) goes over T_1 instead of T. This comes at a cost of a fairly large (but sufficient for our purposes) error term. The regions in $W_0(\mathbb{R})$ that we need to count in are very skewed. In Section 5.2, we use a slicing method to express the point count in terms of certain constants that are expressed as products of local integrals. In Section 5.3, we use our Jacobian change-of-variables from the previous section to evaluate the contribution to these constants from the infinite place. Finally, in Section 5.4, we evaluate the contribution from finite places, and also carry out a squarefree sieve proving Theorems 1, 3, and 4.

5.1. ***Averaging over fundamental domains.*** As in Section 2.3.1, we begin by applying Bhargava's averaging technique, developed in [2; 4]. Let $\mathcal{F}$ be a fundamental domain for the action of $G(\mathbb{Z})$ on $G(\mathbb{R})$ that is box-shaped at infinity (recall that such an $\mathcal{F}$ exists by Theorem 18). Then, by analogy with (9), we obtain

$$\#'\left(\frac{(W(\mathbb{Z})^{(r)}_{\mathrm{red}})_X}{G(\mathbb{Z})}\right) = \frac{1}{\theta_r \operatorname{Vol}(G_0)} \int_{g \in \mathcal{F}} \#(gG_0 \cdot \mathcal{R}_X^{(r)} \cap W(\mathbb{Z})_{\mathrm{red}})\, dg. \tag{35}$$

Since $\mathcal{F}$ is a box-shaped fundamental domain, it follows that we can write, up to a measure-0 set, $\mathcal{F}$ as the disjoint union $\mathcal{F}' \cup \mathcal{N}T_1(\{\pm \mathrm{id}\}\backslash K)$, where $\mathcal{N}$ is the compact subset of $N(\mathbb{R})$ determined in Section 3.2.3,

$$T_1 := \{s = (s_1, \ldots, s_{\lfloor n/2 \rfloor}) \in T : s_i > c_1 \text{ for all } i\}$$

is a subset of T, and $\mathcal{F}'$ is a subset of

$$\mathcal{N}\{s = (s_1, \ldots, s_{\lfloor n/2 \rfloor}) \in T : s_i \le c_1 \text{ for some } i\}(\{\pm \mathrm{id}\}\backslash K).$$

By combining counting results from [6; 16; 43], we prove the following result, in partial analogy with Proposition 12:

Proposition 29. *We have*

$$\#'\left(\frac{(W(\mathbb{Z})^{(r)}_{\mathrm{red}})_X}{G(\mathbb{Z})}\right) = \frac{1}{\tilde{\theta}_r \operatorname{Vol}(G_0)} \int_{us\in\overline{\mathcal{N}}T_1} \#(usX\mathcal{B}\cap W_0(\mathbb{Z}))\delta(s)\,du\,d^\times s + O_\epsilon(X^{\frac{n^2+n-0.4}{2}+\epsilon}), \quad (36)$$

where $\mathcal{B}$ is the multiset $\mathcal{B} := (\mathcal{B}_\infty)_1 := G_0\cdot\mathcal{R}^{(r)}_1\cap W_0(\mathbb{R})$, and $\overline{\mathcal{N}}$ is defined in Section 3.2.3.

Proof. Recall that $B=[b_{ij}]\in W_0(\mathbb{R})$ if and only if $b_{ij}=0$ for all $i+j<n$. Borrowing terminology from [17], we break up the integral on the right-hand side of (35) into three regions, the main body, the shallow cusp, and the deep cusp, and we estimate the contributions from each region separately.

The main body is the region of the integral over $\mathcal{F}$ consisting of $g\in\mathcal{F}$ such that $gG_0\cdot\mathcal{R}^{(r)}_X$ contains integer points $B\in W(\mathbb{Z})$ with $b_{11}\neq 0$. The number of *reducible* elements in the main ball has been shown to be negligible in [6, Proposition 10.7] (for odd n) and [43, Proposition 23] (for even n). These estimates have been improved to a power-saving bound of $O_\epsilon(X^{(n^2+n-0.4)/2+\epsilon})$ in Propositions 2.6 and 3.5 of [16], respectively.

Next, the shallow cusp is the region of the integral over $\mathcal{F}$ consisting of $g\in\mathcal{F}$ such that every $B\in gG_0\cdot\mathcal{R}^{(r)}_X$ satisfies $|b_{11}|<1$, and such that $gG_0\cdot\mathcal{R}^{(r)}_X$ contains integer points $B\in W(\mathbb{Z})$ with $b_{\lfloor(n-1)/2\rfloor\lfloor(n-1)/2\rfloor}\neq 0$. The proofs of [6, Proposition 10.5] (for odd n) and [43, Proposition 21] (for even n) prove that the number of elements in the shallow cusp is bounded by $O(X^{(n^2+n-2)/2})$. Together, these bounds imply

$$\#'\left(\frac{(W(\mathbb{Z})^{(r)}_{\mathrm{red}})_X}{G(\mathbb{Z})}\right) = \frac{1}{\theta_r \operatorname{Vol}(G_0)} \int_{g\in\mathcal{F}} \#(gG_0\cdot\mathcal{R}^{(r)}_X\cap W_0(\mathbb{Z}))\,dg + O_\epsilon(X^{\frac{n^2+n-0.4}{2}+\epsilon}).$$

We now claim that the above estimate also holds when the region of integration $\mathcal{F}$ is replaced by the region $\mathcal{N}T_1$. To prove this claim, we show that the above integral is negligible when $\mathcal{F}$ is replaced by $\mathcal{F}'$. However, this also follows from the previous bounds since $\mathcal{F}'$ lies within the main body and the shallow cusp. Indeed, note that for an element $ntk\in\mathcal{F}$ to lie within the deep cusp (i.e., in the cusp but not the shallow cusp), we must have $\log t_{\lfloor(n-1)/2\rfloor}\gg\log X$. Moreover, since $ntk\in\mathcal{F}$, the condition $\log t_{\lfloor(n-1)/2\rfloor}\gg\log X$ automatically implies that $\log s_i\gg\log X$ for every i, and thus $ntk\notin\mathcal{F}'$. Finally, we replace $\mathcal{N}$ with $\overline{\mathcal{N}}$ (see Section 3.2.3 for the definition), and to compensate, we divide by the order of the subgroup $\Gamma\subset K$, which is $2^{\lceil n/2\rceil}$. The result now follows from the definitions of θ_r and $\tilde{\theta}_r$. □

5.2. ***Slicing.*** Just like in Section 2.3.2, Proposition 13 is not by itself sufficient to estimate the number of integral points in the region $usX\mathcal{B}$, which is typically quite skewed. Instead, we fiber the region $usX\mathcal{B}$ by the coefficients $b_{1(n-1)},\dots,b_{\lfloor n/2\rfloor\lceil n/2\rceil}$. For any $b=(b_1,\dots,b_{\lfloor n/2\rfloor})\in(\mathbb{R}\smallsetminus\{0\})^{\lfloor n/2\rfloor}$, and any $\mathcal{S}\subset W(\mathbb{R})$, let $\mathcal{S}|_b$ denote the *slice of $\mathcal{S}$ at b*, i.e.,

$$\mathcal{S}|_b := \left\{B\in\mathcal{S}\cap W_0(\mathbb{R}) : b_{k(n-k)}(B)=b_k \text{ for all } k\in\left\{1,\dots,\left\lfloor\tfrac{n}{2}\right\rfloor\right\}\right\}.$$

We can express the integrand of the right-hand side of (36) as

$$\#(usX\mathcal{B}\cap W(\mathbb{Z})) = \sum_{\substack{b\in\mathbb{Z}^{\lfloor n/2\rfloor}\\ b_i\neq 0\ \forall i}} \#\big((usX\mathcal{B})|_b\cap W(\mathbb{Z})\big). \tag{37}$$

By examining the action of s on an element $B=[b_{ij}]\in W(\mathbb{R})$, we define the *weight* $w_{ij}=w(b_{ij})$ to be the quantity by which s scales the matrix entry b_{ij} for each pair $(i,j)\in\{1,\dots,n\}^2$. For any subset S of coefficients b_{ij} of $W(\mathbb{R})$, we let $w(S)$ denote the product of the weights of all the elements in S. From Proposition 13, it follows that

$$\#\big((usX\mathcal{B})|_b\cap W(\mathbb{Z})\big) = \mathrm{Vol}((usX\mathcal{B})|_b)(1+O(X^{-1})), \tag{38}$$

where the error term is seen to be X^{-1} times the main term as follows. The weight of every coefficient in $W_0(\mathbb{R})$ not being sliced over is $\gg 1$: indeed, note that $w(b_{k(n+1-k)})=1$ for each k, and that the remaining weights are all at least as big. As a consequence, the range of each coefficient varying in $(usX\mathcal{B})_b$ is $\gg X$, and the volume of $(usX\mathcal{B})|_b$ is asymptotic to the product of the ranges of these varying coefficients. Proposition 13 then yields a saving of size X, as necessary.

Now, since unipotent transformations preserve both the value of b and the volume, we have

$$\mathrm{Vol}((usX\mathcal{B})|_b) = \mathrm{Vol}((sX\mathcal{B})|_b).$$

Recall that we have normalized measures to ensure $\mathrm{Vol}(\overline{\mathcal{N}})=1$. Therefore, (36), (37), and (38) yield

$$\#'\left(\frac{(W(\mathbb{Z})^{(r)}_{\mathrm{red}})_X}{G(\mathbb{Z})}\right) = \frac{1}{\tilde{\theta}_r\,\mathrm{Vol}(G_0)}\sum_{\substack{b\in\mathbb{Z}^{\lfloor n/2\rfloor}\\ b_i\neq 0\ \forall i}}\int_{s\in T_1}\mathrm{Vol}((sX\mathcal{B})|_b)\delta(s)\,d^\times s + O_\epsilon(X^{\frac{n^2+n-0.4}{2}+\epsilon}). \tag{39}$$

Define an action of T on $(\mathbb{R}\smallsetminus\{0\})^{\lfloor n/2\rfloor}$ by setting $s((b_i)_i):=(w(b_{i(n-i)})b_i)_i$. For fixed X and $b=(b_i)_i\in(\mathbb{R}\smallsetminus\{0\})^{\lfloor n/2\rfloor}$, we write $X^{-1}s^{-1}(b)=:\beta=:(\beta_i)_i$. Let S denote the set of coefficients of W_0, and write $S=S_0\sqcup S^\flat$, where S_0 is the set of all $b_{i(n-i)}$, and $S^\flat:=S\smallsetminus S_0$. Since $(sX\mathcal{B})|_b=sX(\mathcal{B}|_\beta)$, it follows that

$$\mathrm{Vol}\big((sX\mathcal{B})|_b\big)=\mathrm{Vol}\big(sX(\mathcal{B}|_\beta)\big)=X^{\dim(S^\flat)}w(S^\flat)\,\mathrm{Vol}(\mathcal{B}|_\beta). \tag{40}$$

Consider the change of variables $s_i\mapsto\beta_i$, and note that $d^\times s=d^\times\beta:=\prod_i(d\beta_i/\beta_i)$. Write $n=2g+1$ when n is odd, and $n=2g+2$ when n is even. We have $s_k=X\beta_k b_k^{-1}$ for all $1\le k\le g$. When n is even, we further have $s_{g+1}=X^2\beta_g\beta_{g+1}b_g^{-1}b_{g+1}^{-1}$. A direct computation now yields that

$$X^{\dim S^\flat}w(S^\flat)\delta(s)=\begin{cases}X^{\dim S^\flat}\prod_{k=1}^g s_k^{2k} & \text{for } n \text{ odd},\\ X^{\dim S^\flat}(s_g^{g-1}s_{g+1}^{g+1})\prod_{k=1}^{g-1}s_k^{2k} & \text{for } n \text{ even}.\end{cases}$$

Therefore, defining $\mathcal{Z}(a_1,\dots,a_g):=\prod_{k=1}^g a_k^{2k}$ when $n=2g+1$ is odd and $\mathcal{Z}(a_1,\dots,a_{g+1}):=a_{g+1}^{g+1}\prod_{k=1}^g a_k^{2k}$ when $n=2g+2$ is even, we have

$$X^{\dim S^\flat}w(S^\flat)\delta(s)=X^{\frac{n^2+n}{2}}\frac{\mathcal{Z}(\beta)}{\mathcal{Z}(b)}. \tag{41}$$

We now examine each individual summand on the right-hand side of (39). For $b = (b_i)_i \in (\mathbb{R} \smallsetminus \{0\})^{\lfloor n/2 \rfloor}$, let $|b|$ denote $(|b_i|)_i$. Recall that G_0 is K-invariant, and recall from Section 3.2.3 that K contains every diagonal matrix in $G(\mathbb{Z})$ with each entry ± 1. It follows that we have $\mathrm{Vol}(\mathcal{B}|_b) = \mathrm{Vol}(\mathcal{B}|_{|b|})$. Therefore, from (40) and (41), we deduce that

$$\int_{s \in T_1} \mathrm{Vol}((sX\mathcal{B})|_b)\delta(s)\, d^\times s = \frac{X^{\frac{n^2+n}{2}}}{\mathcal{Z}(|b|)} \int_{\beta \in \mathbb{R}^{\lfloor n/2 \rfloor}_{\geq 0} \smallsetminus T_1'} \mathcal{Z}(\beta)\, \mathrm{Vol}(\mathcal{B}|_\beta)\, d^\times \beta$$

for each $b = (b_i)_i \in (\mathbb{Z} \smallsetminus \{0\})^{\lfloor n/2 \rfloor}$, where T_1' is a region contained in the set of those β for which $\beta_i \ll X^{-1}$ for at least one i. Since $\mathcal{B}$ is a bounded set and the integral of $\mathcal{Z}(\beta) d^\times \beta$ over T_1' is clearly bounded by $O(X^{-1})$, we find that

$$\begin{aligned}\int_{s \in T_1} \mathrm{Vol}((sX\mathcal{B})|_b)\delta(s)\, d^\times s &= \frac{X^{\frac{n^2+n}{2}}}{\mathcal{Z}(|b|)} \int_{\beta \in \mathbb{R}^{\lfloor n/2 \rfloor}_{>0}} \mathcal{Z}(\beta)\, \mathrm{Vol}(\mathcal{B}|_\beta) d^\times \beta + O(X^{\frac{n^2+n-2}{2}}) \\ &= \frac{X^{\frac{n^2+n}{2}}}{\mathcal{Z}(|b|)} \int_{B \in \mathcal{B}^+} \lambda(B)\, dB + O(X^{\frac{n^2+n-2}{2}}), \end{aligned} \tag{42}$$

where $\mathcal{B}^+ := \{B \in \mathcal{B} : b_{i(n-i)}(B) > 0 \text{ for all } i \in \{1, \dots, \lfloor \frac{n}{2} \rfloor\}\}$. Substituting (42) into (39) and summing over b, we immediately obtain

$$\#'\left(\frac{(W(\mathbb{Z})^{(r)}_{\mathrm{red}})_X}{G(\mathbb{Z})}\right) = C(\mathcal{B}) \cdot C_n^{\mathrm{fin}} \cdot X^{\frac{n^2+n}{2}} + O_\epsilon(X^{\frac{n^2+n-0.4}{2}+\epsilon}). \tag{43}$$

To recover Theorem 1 from (43), it remains to prove two facts: first, that the constant $C(\mathcal{B})$ is equal to $C_{n,r}^{\mathrm{inf}}$, and second, that the $\#'$-count and the $\#$-count differ by only a negligible amount. We verify these facts in the next two subsections.

5.3. ***Computing the constant.*** In this subsection, we compute the value of $C(\mathcal{B})$. Recall that we defined $\mathcal{B}$ to be the multiset $\mathcal{B} := G_0 \cdot \mathcal{R}_1^{(r)} \cap W_0(\mathbb{R})$. Since G_0 is right-K-invariant, we may write $G_0 = \mathcal{S}K$ for some $\mathcal{S} \subset N(\mathbb{R})T$. Hence, we have that $\mathcal{B} = \mathcal{S}K \cdot \mathcal{R}_1^{(r)} \cap W_0(\mathbb{R}) = \mathcal{S} \cdot (K\mathcal{R}_1^{(r)})_0$. Then, by analogy with Lemma 14, we have the following lemma concerning the multiplicity of the fiber of $(K\mathcal{R}^{(r)})_0$ over $U(\mathbb{R})^{(r)}$:

Lemma 30. *The map* $\mathrm{inv} : (K\mathcal{R}^{(r)})_0 \to U(\mathbb{R})^{(r)}$ *is* $\tilde{\theta}_r$*-to-*1.

Proof. It suffices to prove the map $\mathrm{inv} : \{B \in (K\mathcal{R}^{(r)})_0 : b_{i(n-i)}(B) > 0 \text{ for all } i \in \{1, \dots, \lfloor \frac{n}{2} \rfloor\}\} \to U(\mathbb{R})^{(r)}$ is θ_r-to-1. This is a consequence of the following two facts: first, the stabilizer in $G(\mathbb{R})$ of any element of $W(\mathbb{R})^{(r)}$ has size θ_r, and second, the group $N(\mathbb{R})T$ acts simply transitively on $\mathrm{inv}^{-1}(f) \cap W_0(\mathbb{R})$ for any $f \in U(\mathbb{R})^{(r)}$ by Lemma 26. Indeed, given $B \in \mathcal{R}^{(r)}$ having invariant polynomial f, and $pk \in \mathrm{Stab}_{G(\mathbb{R})}(B)$ with $p \in N(\mathbb{R})T$ and $\theta \in K$, the element $\theta B = p^{-1}B$ belongs to $(K\mathcal{R}^{(r)})_0$ and has invariant polynomial f. This association yields the result. □

Now, by analogy with Proposition 15, we are in position to compute the constant $C(\mathcal{B})$:

Proposition 31. *We have that* $C(\mathcal{B}) = C_{n,r}^{\mathrm{inf}}$.

Proof. We have

$$C(\mathcal{B}) = \frac{1}{\theta_r \operatorname{Vol}(G_0)} \int_{B \in \mathcal{S}\cdot(K\mathcal{R}_1^{(r)})_0} |\lambda(B)|\, dB = \frac{|\mathcal{J}| \operatorname{Vol}_{\text{right}}(\mathcal{S}) \operatorname{Vol}(\{f \in U(\mathbb{R})^{(r)} : \mathrm{H}(f) < 1\})}{\operatorname{Vol}(\mathcal{S}K)}, \quad (44)$$

where the second equality follows from applying the Jacobian change-of-variables established in Proposition 27 along with Lemma 30, and where $\operatorname{Vol}_{\text{right}}(\mathcal{S})$ denotes the volume of $\mathcal{S}$ with respect to the right Haar measure on $\mathcal{P}(\mathbb{R})$. But since $\operatorname{Vol}(K)$ is normalized to be equal to 1, we have $\operatorname{Vol}_{\text{right}}(\mathcal{S}) = \operatorname{Vol}(K\mathcal{S})$. In complete analogy with Lemma 16, we have $\operatorname{Vol}(K\mathcal{S}) = \operatorname{Vol}(\mathcal{S}K)$. Combining this with (44) and the computation of $|\mathcal{J}|$ performed in Proposition 28 completes the proof of Proposition 31. □

For the last missing ingredient in the proof of Theorem 1, we prove that there is no asymptotic difference between the #′-count and the #-count. To this end, let $W(\mathbb{Z})_{\text{bs}} := \{B \in W(\mathbb{Z})_{\text{red}}^{(r)} : \operatorname{Stab}_{G(\mathbb{Z})}(B) \neq 1\}$. Then we have the following result:

Proposition 32. *We have*

$$\#\left(\frac{(W(\mathbb{Z})_{\text{bs}})_X}{G(\mathbb{Z})}\right) = O_\epsilon(X^{\frac{n^2+n-0.4}{2}+\epsilon}).$$

Proof. Following (36) and (39), we have

$$\begin{aligned}\#\left(\frac{(W(\mathbb{Z})_{\text{bs}})_X}{G(\mathbb{Z})}\right) &\ll \int_{us\in\mathcal{N}T_1} \#(usX\mathcal{B} \cap W(\mathbb{Z})_{\text{bs}})\delta(s)\, du\, d^\times s \\ &\ll \sum_{\substack{b\in\mathbb{Z}^{\lfloor n/2\rfloor} \\ b_i\neq 0\ \forall i}} \int_{s\in T_1} \#(sX\mathcal{B}|_b \cap (W(\mathbb{Z})_{\text{bs}}))\delta(s)\, du\, d^\times s,\end{aligned}$$

up to an error of $X^{(n^2+n-0.4)/2+\epsilon}$.

Next, note that if $B \in W(\mathbb{Z})_{\text{bs}}$, then the reduction of B mod p has a nontrivial stabilizer for every prime p. Fix $b = (b_i)_i \in (\mathbb{Z}\smallsetminus\{0\})^{\lfloor n/2\rfloor}$, and let p be a prime such that $p \nmid \prod b_i$. We claim that a positive proportion of elements in $W_0(\mathbb{F}_p)|_b := \left\{B \in W_0(\mathbb{F}_p) : b_{k(n-k)}(B) = b_k \text{ for all } k \in \left\{1, \dots, \left\lfloor \frac{n}{2} \right\rfloor\right\}\right\}$ have trivial stabilizer in $G(\mathbb{F}_p)$. Indeed, this is true for every $B \in W(\mathbb{F}_p)$ whose invariant polynomial is irreducible over $\mathbb{F}_p$ when n is odd (see [6, Section 10.7]), and for every $B \in W(\mathbb{F}_p)$ whose invariant polynomial is a linear polynomial times an irreducible polynomial over $\mathbb{F}_p$ when n is even (see [43, proof of Proposition 23]). A positive proportion of invariant polynomials satisfy these splitting criteria, and the fiber in $W_0(\mathbb{F}_p)|_b$ over each polynomial has the same size (in fact, this size is $\#N(\mathbb{F}_p)$). A power-saving estimate for each summand in the above equation now follows by using the Selberg sieve analogously to the argument in [41]. (Indeed, a power-saving bound from the Selberg sieve only requires the ability to count these integer points with a power-saving error term, and requires a positive proportion of mod p residue classes to be avoided. We omit the details of the computation of the precise power-saving exponent since the argument closely follows that in [41].) The proposition now follows since the sum over b converges absolutely. □

Theorem 1 now follows from (43) and Propositions 31 and 32.

5.4. ***Congruence conditions.*** We now prove Theorem 3. Let $S \subset W(\mathbb{Z})^{(r)}$ be a big family, and suppose for now that S is defined by congruence conditions at finitely many places (i.e., suppose that $S_p = W(\mathbb{Z}_p)$ for all primes $p \gg 1$). For each $b \in (\mathbb{Z}_p \smallsetminus \{0\})^{\lfloor n/2 \rfloor}$, let

$$(S_p)_0|_b := \left\{B \in (S_p)_0 : b_{i(n-i)}(B) = b_i \text{ for all } i \in \left\{1, \dots, \left\lfloor \tfrac{n}{2} \right\rfloor\right\}\right\},$$

and for each $b \in (\mathbb{Z} \smallsetminus \{0\})^{\lfloor n/2 \rfloor}$, let $\nu(S_0|_b) := \prod_p \mathrm{Vol}((S_p)_0|_b)$ denote the density of the slice $S_0|_b$ in $W_0(\mathbb{Z})|_b$; here, each p-adic volume $\mathrm{Vol}((S_p)_0|_b)$ is computed with respect to the Euclidean measure on $W_0(\mathbb{Z}_p)|_b$, normalized so that $W_0(\mathbb{Z}_p)|_b$ has volume 1. Then an argument identical to the one used to obtain (39) yields the asymptotic formula

$$\#\left(\frac{(S_{\mathrm{red}})_X}{G(\mathbb{Z})}\right) = \frac{1}{\theta_r \,\mathrm{Vol}(G_0)} \sum_{\substack{b \in \mathbb{Z}^{\lfloor n/2 \rfloor} \\ b_i \neq 0 \ \forall i}} \nu(S_0|_b) \int_{s \in T_1} \mathrm{Vol}((sX\mathcal{B})|_b)\, \delta(s)\, d^\times s + O_\epsilon(X^{\frac{n^2+n-0.4}{2}+\epsilon}). \quad (45)$$

The $G(\mathbb{Z})$-invariance of S implies that $\nu(S_0|_b) = \nu(S_0|_{|b|})$ for all b. Hence, (42) and (45) yield the estimate

$$\#\left(\frac{(S_{\mathrm{red}})_X}{G(\mathbb{Z})}\right) = C(\mathcal{B}) \cdot \left(\sum_{\substack{b \in \mathbb{Z}^{\lfloor n/2 \rfloor} \\ b_i > 0 \ \forall i}} \frac{\nu(S_0|_b)}{\mathcal{Z}(b)}\right) \cdot X^{\frac{n^2+n}{2}} + O_\epsilon(X^{\frac{n^2+n-0.4}{2}+\epsilon}). \quad (46)$$

To evaluate the sum over b on the right-hand side of (46), we use the following property, which is a consequence of the fact that S is a big family: if p is a prime and $b, b' \in (\mathbb{Z}_p \smallsetminus \{0\})^{\lfloor n/2 \rfloor}$ are elements such that $|b_i|_p = |b'_i|_p$ for each i, then $\mathrm{Vol}((S_p)_0|_b) = \mathrm{Vol}((S_p)_0|_{b'})$. By repeatedly using this property, we obtain the chain of equalities

$$\begin{aligned}
\sum_{\substack{b \in \mathbb{Z}^{\lfloor n/2 \rfloor} \\ b_i > 0 \ \forall i}} \frac{\nu(S_0|_b)}{\mathcal{Z}(b)} &= \prod_p \sum_{\substack{(i) \in \mathbb{Z}^{\lfloor n/2 \rfloor} \\ i_j \geq 0 \ \forall j}} \frac{\mathrm{Vol}((S_p)_0|_{(p^{i_1}, \dots, p^{i_{\lfloor n/2 \rfloor}})})}{\mathcal{Z}(p^{i_1}, \dots, p^{i_{\lfloor n/2 \rfloor}})} \\
&= \prod_p \left(1 - \frac{1}{p}\right)^{-\lfloor \frac{n}{2} \rfloor} \int_{\substack{b \in \mathbb{Z}_p^{\lfloor n/2 \rfloor} \\ b_i \neq 0 \ \forall i}} \frac{\mathrm{Vol}((S_p)_0|_{(|b_1|_p^{-1}, \dots, |b_{\lfloor \frac{n}{2} \rfloor}|_p^{-1})})}{\mathcal{Z}(|b_1|_p^{-1}, \dots, |b_{\lfloor \frac{n}{2} \rfloor}|_p^{-1}) \prod_i |b_i|_p}\, db \\
&= \prod_p \left(1 - \frac{1}{p}\right)^{-\lfloor \frac{n}{2} \rfloor} \int_{\substack{b \in \mathbb{Z}_p^{\lfloor n/2 \rfloor} \\ b_i \neq 0 \ \forall i}} \left|\frac{\mathcal{Z}(b)}{\prod_i b_i}\right|_p \mathrm{Vol}((S_p)_0|_b)\, db \\
&= \prod_p \left(1 - \frac{1}{p}\right)^{-\lfloor \frac{n}{2} \rfloor} \int_{B \in (S_p)_0} |\lambda(B)|_p\, dB,
\end{aligned}$$

where the second line above follows by partitioning the region of integration $(\mathbb{Z}_p \smallsetminus \{0\})^{\lfloor n/2 \rfloor}$ into level sets for the integrand and summing over all such level sets, and where the last line above follows just as in (42).

It remains to handle the case where S is a big family defined by congruence conditions at infinitely many places. By abuse of notation, let $\mathcal{Z}$ be the polynomial on W_0 defined by $\mathcal{Z}(B) := \mathcal{Z}(b_{1(n-1)}(B), \dots, b_{\lfloor n/2 \rfloor \lceil n/2 \rceil}(B))$. Then the case of infinitely many places follows from the case

of finitely many places by using the following bound on the number of $G(\mathbb{Z})$-equivalence classes of elements with large $\mathcal{Z}$-value, in conjunction with an inclusion-exclusion sieve:[8]

Theorem 33. *Fix a real number $M > 0$. Then the number of $G(\mathbb{Z})$-equivalence classes of (or equivalently, $\mathcal{P}(\mathbb{Z})$-orbits of) elements of the set $\{B \in W_0(\mathbb{Z}) : \mathrm{H}(B) < X, |\mathcal{Z}(B)| \geq M^2\}$ is bounded by $O(X^{(n^2+n)/2}/M) + O_\epsilon(X^{(n^2+n-0.4)/2+\epsilon})$, where the implied constant is independent of M.*

Proof. The required bound follows immediately from the proof of Theorem 1 by simply summing (42) over only those b such that $|\mathcal{Z}(b)| \geq M^2$. □

Remark. Theorem 33 constitutes a slight strengthening of a result previously proven in [16, Sections 2.3–2.4 and 3.3–3.4], where Bhargava, Shankar, and Wang used the averaging method to obtain an upper bound of $O_\epsilon(X^{(n^2+n)/2+\epsilon}/M) + O_\epsilon(X^{(n^2+n-0.4)/2+\epsilon})$.

This concludes the proof of Theorem 3. We finish by noting that Theorem 4 follows from Theorem 3 by applying the Jacobian change-of-variables result in Proposition 27 to each p-adic integral, with ϕ taken to be the characteristic function of $(S_p)_0$. Indeed, Propositions 27 and 28 together imply the equality of each factor at the odd primes; at $p = 2$, there is an extra factor of $2^{n/2}$ when n is even, which perfectly accounts for the corresponding factor at infinity.

6. A local-to-global principle for the action of $\mathcal{P}(\mathbb{Z})$ on $W_0(\mathbb{Z})$

In this section, we develop an alternative method for counting reducible $G(\mathbb{Z})$-orbits on $W(\mathbb{Z})$. This method, which we called "Method II" in Section 1.4, consists of the following steps:

- First, in Section 6.1, we consider a general representation with trivial generic stabilizer, and we prove that the integral orbits of such a representation satisfy a strong local-to-global principle, which is stated precisely in Theorem 34.
- As shown in Proposition 25, the action of the group $\mathcal{P}$ on the reducible hyperplane W_0 is a representation with trivial generic stabilizer. In Section 6.2, we deduce Theorem 5 by applying Theorem 34 to the action of $\mathcal{P}$ on W_0.
- Finally, in Section 6.3, we explain how to use Theorem 5 to deduce asymptotics for the number of $\mathcal{P}(\mathbb{Z})$-orbits on $W_0(\mathbb{Z})$—and hence also for the number of reducible $G(\mathbb{Z})$-orbits on $W(\mathbb{Z})$. The main analytic ingredient is an upper bound obtained by Bhargava, Shankar, and Wang on the number of $\mathcal{P}(\mathbb{Z})$-orbits of elements $B \in W_0(\mathbb{Z})$ with the property that $\lambda(B)$ is large.

6.1. ***Group actions with trivial stabilizers.*** In this section, we work with the following data: an algebraic group H, finite-dimensional H-representations V and I, and an H-equivariant morphism $\phi : V \to I$, all defined over $\mathbb{Z}$. Note that we do not require ϕ to be a linear map. Any polynomial map will suffice. Suppose that the group H has class number 1 over $\mathbb{Q}$, meaning that $H(\mathbb{A}_\mathbb{Q})$ is the product of its subgroups

[8] Just as in Section 2.3.4, we do not flesh out the sieving argument here to avoid being repetitive, because in Section 6.3 (to follow), we use the same sort of argument to prove Theorem 4.

$H(\mathbb{A}_{\mathbb{Z}})$ and $H(\mathbb{Q})$ (i.e., for every $h \in H(\mathbb{A}_{\mathbb{Q}})$, there exist $h' \in H(\mathbb{A}_{\mathbb{Z}})$ and $h'' \in H(\mathbb{Q})$ such that $h = h'h''$). Here $\mathbb{A}_{\mathbb{Q}}$ denotes the ring of adeles, and $\mathbb{A}_{\mathbb{Z}}$ is the ring of everywhere integral adeles. Suppose further that, for some nonempty open subscheme $\mathcal{I} \subset I$, defined over $\mathbb{Z}$, the following two assumptions hold:

(1) For every $i \in \mathcal{I}(\mathbb{C})$, the set $\{v \in V(\mathbb{C}) : \phi(v) = i\}$ forms a single nonempty $H(\mathbb{C})$-orbit.

(2) For every $i \in \mathcal{I}(\mathbb{C})$ and every (or equivalently, any) $v \in V(\mathbb{C})$ with $\phi(v) = i$, we have $\mathrm{Stab}_{H(\mathbb{C})}(v) = 1$.

In this setting, we prove the following strong local-to-global principle for the action of H on V:

Theorem 34. *Suppose that $i \in \mathcal{I}(\mathbb{Z})$ is an element contained in the image $\phi(V(\mathbb{Z}))$. For each prime p, let $v_p \in V(\mathbb{Z}_p)$ be such that $\phi(v_p) = i$. Then there exists $v \in V(\mathbb{Z})$, unique up to the action of $H(\mathbb{Z})$, such that v is $H(\mathbb{Z}_p)$-equivalent to v_p for each prime p.*

Before we give the proof of Theorem 34, we first use assumptions (1) and (2) enumerated above to deduce that an analogue of assumption (1) holds for *any* subfield $K \subset \mathbb{C}$:

Lemma 35. *Let $K \subset \mathbb{C}$ be a subfield. For every $i \in \mathcal{I}(K)$, the set $\{v \in V(K) : \phi(v) = i\}$ consists of single $H(K)$-orbit.*

Proof. Let $i \in \mathcal{I}(K)$, and let $v, v' \in V(K)$ with $\phi(v) = \phi(v') = i$. By assumption (1), there exists $h \in H(\mathbb{C})$ such that $v' = h \cdot v$. Then for any $\sigma \in \mathrm{Gal}(\mathbb{C}/K)$, we have $v' = {}^{\sigma}h \cdot v$, so $h^{-1\sigma}h \in \mathrm{Stab}_{H(\mathbb{C})}(v) = 1$ by assumption (2). It follows that h is fixed by $\mathrm{Gal}(\mathbb{C}/K)$, and so $h \in H(K)$, as necessary. □

We now use Lemma 35, assumption (2), and the fact that H has class number 1 over $\mathbb{Q}$ to complete the proof of the theorem:

Proof of Theorem 34. Let us temporarily drop assumption (2). For a principal ideal domain R with fraction field K and an element $v \in V(R)$, define $H(K)_v := \{h \in H(K) : h \cdot v \in H(R)\}$. Then it is clear that the set of $H(R)$-orbits contained in the $H(K)$-orbit of an element $v \in V(R)$ is in natural bijection with the double coset space $H(R)\backslash H(K)_v / \mathrm{Stab}_{H(K)}(v)$. With assumption (2) reinstated, this double coset space is simply given by $H(R)\backslash H(K)_v$ as long as the H-invariant of v lies in the subset $\mathcal{I}(R) \subset I(R)$.

Now, using the fact that H has class number 1 over $\mathbb{Q}$, it is proven in [9, proof of Proposition 3.6][9] that the diagonal embedding $H(\mathbb{Q}) \hookrightarrow \prod_p H(\mathbb{Q}_p)$ induces a bijection

$$H(\mathbb{Z})\backslash H(\mathbb{Q})_{v_0} \longrightarrow \prod_p H(\mathbb{Z}_p)\backslash H(\mathbb{Q}_p)_{v_0}. \tag{47}$$

Let $v_0 \in V(\mathbb{Z})$ be an element such that $\phi(v_0) = i$. By the result of the previous paragraph, the bijection in (47) may be regarded as identifying the set of $H(\mathbb{Z})$-orbits contained in the $H(\mathbb{Q})$-orbit of v_0 with the product over all primes p of the set of $H(\mathbb{Z}_p)$-orbits contained in the $H(\mathbb{Q}_p)$-orbit of v_0.

By Lemma 35, which implies that the $H(\mathbb{Q}_p)$-orbit of v_0 is equal to that of v_p for each prime p, we may view the tuple $(v_p)_p$ as an element of the right-hand side of (47). Then, under the bijection, the tuple $(v_p)_p$ corresponds to the $H(\mathbb{Z})$-orbit of the desired element $v \in V(\mathbb{Z})$. □

[9] To be clear, [9, Proposition 3.6] concerns the case $H = \mathrm{PGL}_2$, but it is evident that the same argument goes through for any group H of class number 1 over $\mathbb{Q}$.

6.2. ***Proof of Theorem 5.*** We now deduce Theorem 5 from Theorem 34. To do so, we take $H = \mathcal{P}$, $V = W_0$, $I = U$, $\phi = \mathrm{inv}$, and $\mathcal{I} \subset U$ to be the open subscheme consisting of nondegenerate polynomials. Note that the group $\mathcal{P}$ clearly has class number 1 over $\mathbb{Q}$, and note that assumptions (1) and (2) follow immediately from Proposition 25. Applying Theorem 34 then yields the following: for any $f \in \mathcal{I}(\mathbb{Z})$, *if* there exists $B_0 \in W_0(\mathbb{Z})$ such that $\mathrm{inv}(B_0) = f$ and *if* there exists $B_p \in W_0(\mathbb{Z}_p)$ such that $\mathrm{inv}(B_p) = f$ for each prime p, then there exists $B \in W_0(\mathbb{Z})$, unique up to the action of $\mathcal{P}(\mathbb{Z})$, such that B is $\mathcal{P}(\mathbb{Z}_p)$-equivalent to B_p for each prime p. To prove Theorem 5, it now remains to verify the existence of the elements B_0 and B_p for each prime p.

When n is even, the existence of the elements B_0 and B_p is implied by the existence of the integral section σ_0 given in Section 3.3—indeed, when n is even, we have restricted our consideration to those monic polynomials whose x^i-coefficients are divisible by 2 for each odd i, so the section σ_0 is defined over $\mathbb{Z}$ in this case. On the other hand, when n is odd, the section σ_0 is not defined over $\mathbb{Z}$. Instead, it is shown in [31, Section 4.1] that for any principal ideal domain R and for each $f \in U(R)$ there exists a pair (A, B) of $n \times n$ symmetric matrices with entries in R such that: A is split over $K = \mathrm{Frac}(R)$; A and B share a maximal isotropic space over K; and $\det(xA + yB) = (-1)^{\lfloor n/2 \rfloor} f(x, y)$. By the classification of unimodular symmetric bilinear forms (see [39, Chapter V]), A is $\mathrm{GL}_n(R)$-equivalent to $\mathcal{A}$. By translating B with the same element of $\mathrm{GL}_n(R)$, we obtain $B' \in W_0(R)$ satisfying $\mathrm{inv}(B') = f$.

6.3. ***Proof of Theorem 4.*** We now use Theorem 5 to give a second proof of Theorem 4. We call a subset $\mathfrak{S} \subset W_0(\mathbb{Z})$ a *big family* if $\mathfrak{S} = W(\mathbb{Z})_0^{(r)} \cap \bigcap_p \mathfrak{S}_p$, where the sets $\mathfrak{S}_p \subset W_0(\mathbb{Z}_p)$ satisfy the following properties:

(1) $\mathfrak{S}_p$ is $\mathcal{P}(\mathbb{Z}_p)$-invariant and is the preimage under reduction modulo p^j of a nonempty subset of $W_0(\mathbb{Z}/p^j\mathbb{Z})$ for some $j > 0$ for each p.

(2) $\mathfrak{S}_p$ contains all elements $B \in W_0(\mathbb{Z}_p)$ such that, for all $p \gg 1$, we have that $b_{i(n-i)}(B)$ is a p-adic unit for some i.

We then have the following variant of Theorem 4, from which Theorem 4 readily follows by taking $\mathfrak{S}_p = (S_p)_0$ for each p. Indeed, given this choice of $\mathfrak{S}$, the asymptotics for reducible $G(\mathbb{Z})$-orbits on S are the same as those for $\mathcal{P}(\mathbb{Z})$-orbits on $\mathfrak{S}$, as explained in Section 1.4.

Theorem 36. *Let $\mathfrak{S} \subset W_0(\mathbb{Z})$ be a big family. Then the number of $\mathcal{P}(\mathbb{Z})$-orbits on $\mathfrak{S}$ of height up to X is given by*

$$\left(\prod_p \int_{f \in U(\mathbb{Z}_p)} \#\left(\frac{\mathrm{inv}^{-1}(f) \cap \mathfrak{S}_p}{\mathcal{P}(\mathbb{Z}_p)}\right) df\right) \cdot N^{(r)}(X) + o(X^{\frac{n^2+n}{2}}), \tag{48}$$

where df denotes the Euclidean measure on $U(\mathbb{Z}_p)$, normalized so that $U(\mathbb{Z}_p)$ has volume 1.

Proof. To start, fix an integer $\mathfrak{b} \geq 1$, and suppose its prime factorization is given by $\mathfrak{b} = \prod_p p^{e_p}$. We first prove an analogue of Theorem 4 for the subfamily $\mathfrak{S}(\mathfrak{b}) := \{B \in \mathfrak{S} : |\mathcal{Z}(B)| = \mathfrak{b}\}$. Note that the subfamily $\mathfrak{S}(\mathfrak{b})$ is itself a big family, where $\mathfrak{S}(\mathfrak{b})_p = \{B \in \mathfrak{S}_p : |\mathcal{Z}(B)|_p = |\mathfrak{b}|_p\}$.

If $\mathfrak{S}(\mathfrak{b}) = \varnothing$, then the result is tautologically true, so we may assume that $\mathfrak{S}(\mathfrak{b}) \neq \varnothing$. For each prime $p \mid \mathfrak{b}$, we partition $U(\mathbb{Z}_p)$ as $U(\mathbb{Z}_p) = \bigsqcup_{j=1}^{m_p} U_{p,j}$, where each $U_{p,j}$ is a level set for the function that sends $f \in U(\mathbb{Z}_p)$ to $\#(\mathcal{P}(\mathbb{Z}_p)\backslash(\mathrm{inv}^{-1}(f) \cap \mathfrak{S}(\mathfrak{b})_p))$. Write "$E(m)$" to mean "a power of m that depends only on n and $\mathfrak{S}$." Then set $U_{p,j}$ is defined by congruence conditions modulo $E(p^{e_p})$. The quantity $\#(\mathcal{P}(\mathbb{Z}_p)\backslash(\mathrm{inv}^{-1}(f) \cap \mathfrak{S}(\mathfrak{b})_p))$ is independent of the choice of $f \in U_{p,j}$ (by the definition of a level set) and is evidently bounded by $E(p^{e_p})$.

Now for each prime $p \nmid \mathfrak{b}$, the proof of Proposition 25 implies that $\#(\mathcal{P}(\mathbb{Z}_p)\backslash(\mathrm{inv}^{-1}(f) \cap \mathfrak{S}(\mathfrak{b})_p)) = 1$ for each $f \in \mathrm{inv}(\mathfrak{S}(\mathfrak{b})_p)$. It then follows from Theorem 5 that the quantity

$$\#\left(\frac{\mathrm{inv}^{-1}(f) \cap \mathfrak{S}(\mathfrak{b})}{\mathcal{P}(\mathbb{Z})}\right) = \prod_p \#\left(\frac{\mathrm{inv}^{-1}(f) \cap \mathfrak{S}(\mathfrak{b})_p}{\mathcal{P}(\mathbb{Z}_p)}\right) = \prod_{p \mid \mathfrak{b}} \#\left(\frac{\mathrm{inv}^{-1}(f) \cap \mathfrak{S}(\mathfrak{b})_p}{\mathcal{P}(\mathbb{Z}_p)}\right) \tag{49}$$

is independent of the choice of $f \in \mathrm{inv}(\mathfrak{S}(\mathfrak{b})) \cap \bigcap_p U_{p,j_p}$. Therefore, for each tuple $(j_p)_{p \mid \mathfrak{b}} \in \prod_{p \mid \mathfrak{b}} \{1, \dots, m_p\}$, we have

$$\sum_{\substack{f \in U(\mathbb{Z}) \cap \bigcap_p U_{p,j_p} \\ \mathrm{H}(f) < X}} \#\left(\frac{\mathrm{inv}^{-1}(f) \cap \mathfrak{S}(\mathfrak{b})}{\mathcal{P}(\mathbb{Z})}\right) = \#\left(\frac{\mathrm{inv}^{-1}(f^*) \cap \mathfrak{S}(\mathfrak{b})}{\mathcal{P}(\mathbb{Z})}\right) \cdot \sum_{\substack{f \in \mathrm{inv}(\mathfrak{S}(\mathfrak{b})) \cap \bigcap_p U_{p,j_p} \\ \mathrm{H}(f) < X}} 1, \tag{50}$$

where $f^* \in \mathrm{inv}(\mathfrak{S}(\mathfrak{b})) \cap \bigcap_{p \mid \mathfrak{b}} U_{p,j_p}$ is any fixed element. Since $\mathfrak{S}$ is a big family, it follows that $\mathrm{inv}(\mathfrak{S}(\mathfrak{b})_p) = U(\mathbb{Z}_p)$ for every $p \gg 1$ that does not divide $\mathfrak{b}$. As the set $\mathrm{inv}(\mathfrak{S}(\mathfrak{b})) \cap \bigcap_p U_{p,j_p}$ is defined by congruence conditions modulo $E(\mathfrak{b})$, since $\mathrm{inv}(\mathfrak{S}(\mathfrak{b})_p) \cap U_{p,j_p}$ is defined by congruence conditions modulo $E(p^{e_p})$ for each p, we obtain the asymptotic

$$\begin{aligned}\sum_{\substack{f \in \mathrm{inv}(\mathfrak{S}(\mathfrak{b})) \cap \bigcap_p U_{p,j_p} \\ \mathrm{H}(f) < X}} 1 = N^{(r)}(X) \cdot \prod_{p \mid \mathfrak{b}} \int_{f \in \mathrm{inv}(\mathfrak{S}(\mathfrak{b})_p) \cap U_{p,j_p}} df \\ \cdot \prod_{p \nmid \mathfrak{b}} \int_{f \in \mathrm{inv}(\mathfrak{S}(\mathfrak{b})_p)} df + O_\epsilon(E(\mathfrak{b}) X^{\frac{n^2+n-0.4}{2}+\epsilon}).\end{aligned} \tag{51}$$

Substituting the asymptotic (51) into the right-hand side of (50), applying (49) to the resulting expression, and summing that over tuples $(j_p)_{p \mid \mathfrak{b}} \in \prod_{p \mid \mathfrak{b}} \{1, \dots, m_p\}$ yields

$$\begin{aligned}\sum_{\substack{f \in U(\mathbb{Z}) \\ \mathrm{H}(f) < X}} \#\left(\frac{\mathrm{inv}^{-1}(f) \cap \mathfrak{S}(\mathfrak{b})}{\mathcal{P}(\mathbb{Z})}\right) = N^{(r)}(X) \cdot \prod_p \int_{f \in U(\mathbb{Z}_p)} \#\left(\frac{\mathrm{inv}^{-1}(f) \cap \mathfrak{S}(\mathfrak{b})_p}{\mathcal{P}(\mathbb{Z}_p)}\right) df \\ + O_\epsilon(E(\mathfrak{b}) X^{\frac{n^2+n-0.4}{2}+\epsilon}).\end{aligned} \tag{52}$$

Next, we prove that the theorem holds with "=" replaced by "≥". For any real number $M > 1$, let $\mathfrak{S}(M) := \{B \in \mathfrak{S} : |\mathcal{Z}(B)| < M^2\}$. Summing (52) over $\mathfrak{b} < M$ yields

$$\begin{aligned}\sum_{\substack{f \in U(\mathbb{Z}) \\ \mathrm{H}(f) < X}} \#\left(\frac{\mathrm{inv}^{-1}(f) \cap \mathfrak{S}(M)}{\mathcal{P}(\mathbb{Z})}\right) = N^{(r)}(X) \cdot \sum_{\mathfrak{b} < M} \prod_p \int_{f \in U(\mathbb{Z}_p)} \#\left(\frac{\mathrm{inv}^{-1}(f) \cap \mathfrak{S}(\mathfrak{b})_p}{\mathcal{P}(\mathbb{Z}_p)}\right) df \\ + O_\epsilon(E(M) X^{\frac{n^2+n-0.4}{2}+\epsilon}).\end{aligned} \tag{53}$$

Dividing (53) through by $N^{(r)}(X)$ and letting $X \to \infty$, we find that

$$\liminf_{X\to\infty} \frac{\sum_{\substack{f\in U(\mathbb{Z})\\ \mathrm{H}(f)<X}} \#\left(\frac{\mathrm{inv}^{-1}(f)\cap \mathfrak{S}}{\mathcal{P}(\mathbb{Z})}\right)}{N^{(r)}(X)} \geq \sum_{\mathfrak{b}<M} \prod_p \int_{f\in U(\mathbb{Z}_p)} \#\left(\frac{\mathrm{inv}^{-1}(f)\cap \mathfrak{S}(\mathfrak{b})_p}{\mathcal{P}(\mathbb{Z}_p)}\right) df. \tag{54}$$

Now, letting $M \to \infty$ on the right-hand side of (54) and factoring the sum into an Euler product, we obtain

$$\begin{aligned}\sum_{\mathfrak{b}=1}^{\infty} \prod_p \int_{f\in U(\mathbb{Z}_p)} \#\left(\frac{\mathrm{inv}^{-1}(f)\cap \mathfrak{S}(\mathfrak{b})_p}{\mathcal{P}(\mathbb{Z}_p)}\right) df &= \prod_p \sum_{e=0}^{\infty} \int_{f\in U(\mathbb{Z}_p)} \#\left(\frac{\mathrm{inv}^{-1}(f)\cap \mathfrak{S}(p^e)_p}{\mathcal{P}(\mathbb{Z}_p)}\right) df \\ &= \prod_p \int_{f\in U(\mathbb{Z}_p)} \#\left(\frac{\mathrm{inv}^{-1}(f)\cap \mathfrak{S}_p}{\mathcal{P}(\mathbb{Z}_p)}\right) df. \end{aligned} \tag{55}$$

Combining (54) with (55), we find that Theorem 4 holds with "=" replaced by "≥". Note that it is not a priori clear whether the infinite sum on the left-hand side of (55) converges, but even if it were to diverge, it would still be equal to the product on the right-hand side! To show that the sum does indeed converge, one can apply the Jacobian change-of-variables result in Proposition 27 to each p-adic integral; it is then clear that the summand at $\mathfrak{b}$ is $O\left(\mathfrak{b}^{-1}\prod_{p\mid \mathfrak{b}} p^{-1}\right)$, which is sufficient, as the sum of the reciprocals of the powerful numbers converges (see [25]).

It thus remains to prove the theorem with "=" replaced by "≤". Let $\mathfrak{S}(M)' := \mathfrak{S} \smallsetminus \mathfrak{S}(M)$. Then for each $B \in \mathfrak{S}(M)'$, we have that $|\mathcal{Z}(B)| \geq M^2$. From Theorem 33, it follows that

$$\sum_{\substack{f\in U(\mathbb{Z})\\ \mathrm{H}(f)<X}} \#\left(\frac{\mathrm{inv}^{-1}(f)\cap \mathfrak{S}(M)'}{\mathcal{P}(\mathbb{Z})}\right) = O(X^{\frac{n^2+n}{2}}/M) + O_\epsilon(X^{\frac{n^2+n-0.4}{2}+\epsilon}). \tag{56}$$

On the other hand, it follows from (53) that

$$\begin{aligned}\sum_{\substack{f\in U(\mathbb{Z})\\ \mathrm{H}(f)<X}} \#\left(\frac{\mathrm{inv}^{-1}(f)\cap \mathfrak{S}(M)}{\mathcal{P}(\mathbb{Z})}\right) &\\ \leq N^{(r)}(X)\cdot \prod_p \int_{f\in U(\mathbb{Z}_p)} \#\left(\frac{\mathrm{inv}^{-1}(f)\cap \mathcal{S}_p}{\mathcal{P}(\mathbb{Z}_p)}\right) df &+ O_\epsilon(E(M)X^{\frac{n^2+n-0.4}{2}+\epsilon}). \end{aligned} \tag{57}$$

Taking M to grow as a sufficiently small power of X and combining (56) with (57) yields Theorem 36, and hence also Theorem 4. □

We finish by noting that Theorems 1 and 3 follow from Theorem 4 by applying the Jacobian change-of-variables result in Proposition 27 to each p-adic integral.

Acknowledgments

We thank Manjul Bhargava for several enlightening conversations, for providing numerous helpful comments and suggestions, and for offering feedback on earlier drafts of this paper. We thank the

anonymous referee for giving us detailed feedback that helped us improve the exposition and readability of the paper. We are also grateful to Alex Bartel, Noam D. Elkies, Andrew Granville, Jef Laga, Aaron Landesman, Aurel Page, Peter Sarnak, and Melanie Matchett Wood for helpful discussions, and to Will Sawin for suggesting the proof of Lemma 23.

Shankar was supported by a National Sciences and Engineering Research Council of Canada discovery grant and a Sloan fellowship. Siad was supported by a Queen Elizabeth II/Steve Halperin Scholarship in Science and Technology at the University of Toronto during the initial phase of this project and is grateful to Princeton University and the Institute for Advanced Study for providing excellent working conditions during its final stage. Swaminathan was supported by the National Science Foundation, under the Graduate Research Fellowship, as well as award no. 2202839.

References

[1] S. A. Altuğ, A. Shankar, I. Varma, and K. H. Wilson, "The number of D_4-fields ordered by conductor", *J. Eur. Math. Soc.* **23**:8 (2021), 2733–2785. MR Zbl

[2] M. Bhargava, "The density of discriminants of quartic rings and fields", *Ann. of Math.* (2) **162**:2 (2005), 1031–1063. MR Zbl

[3] M. Bhargava, "Higher composition laws, IV: The parametrization of quintic rings", *Ann. of Math.* (2) **167**:1 (2008), 53–94. MR Zbl

[4] M. Bhargava, "The density of discriminants of quintic rings and fields", *Ann. of Math.* (2) **172**:3 (2010), 1559–1591. MR Zbl

[5] M. Bhargava, "Most hyperelliptic curves over $\mathbb{Q}$ have no rational points", preprint, 2013. arXiv 1308.0395

[6] M. Bhargava and B. H. Gross, "The average size of the 2-Selmer group of Jacobians of hyperelliptic curves having a rational Weierstrass point", pp. 23–91 in *Automorphic representations and L-functions* (Mumbai, 2012), edited by D. Prasad et al., Tata Inst. Fundam. Res. Stud. Math. **22**, Tata Inst. Fund. Res., Mumbai, 2013. MR Zbl

[7] M. Bhargava and A. Shankar, "The average number of elements in the 4-Selmer groups of elliptic curves is 7", preprint, 2013. arXiv 1312.7333

[8] M. Bhargava and A. Shankar, "The average size of the 5-Selmer group of elliptic curves is 6, and the average rank is less than 1", preprint, 2013. arXiv 1312.7859

[9] 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

[10] 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

[11] M. Bhargava and I. Varma, "The mean number of 3-torsion elements in the class groups and ideal groups of quadratic orders", *Proc. Lond. Math. Soc.* (3) **112**:2 (2016), 235–266. MR Zbl

[12] M. Bhargava and A. Yang, "On the number of integral binary n-ic forms having bounded Julia invariant", *Bull. Lond. Math. Soc.* **54**:4 (2022), 1232–1248. MR Zbl

[13] 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

[14] M. Bhargava, J. Hanke, and A. Shankar, "The mean number of 2-torsion elements in the class groups of n-monogenized cubic fields", preprint, 2020. arXiv 2010.15744

[15] M. Bhargava, A. Shankar, and A. Swaminathan, "The second moment of the size of the 2-Selmer group of elliptic curves", preprint, 2021. arXiv 2110.09063

[16] M. Bhargava, A. Shankar, and X. Wang, "Squarefree values of polynomial discriminants, I", *Invent. Math.* **228**:3 (2022), 1037–1073. MR Zbl

[17] M. Bhargava, A. Shankar, and X. Wang, "Squarefree values of polynomial discriminants, II", preprint, 2022. arXiv 2207.05592

[18] A. Borel, "Ensembles fondamentaux pour les groupes arithmétiques", pp. 23–40 in *Colloque sur la théorie des groupes algébriques* (Brussels, 1962), Librairie Univ., Louvain, Belgium, 1962. MR Zbl

[19] A. Borel and Harish-Chandra, "Arithmetic subgroups of algebraic groups", *Ann. of Math.* (2) **75** (1962), 485–535. MR Zbl

[20] N. Bourbaki, *Groupes et algèbres de Lie*, *Chapitres VII–VIII*, Actualités Scientifiques et Industrielles **1364**, Hermann, Paris, 1975. Translated as *Lie groups and Lie algebras*, Springer, 2005. MR Zbl

[21] H. Davenport, "On a principle of Lipschitz", *J. Lond. Math. Soc.* **26** (1951), 179–183. MR Zbl

[22] H. Davenport, "On the class-number of binary cubic forms, I", *J. Lond. Math. Soc.* **26** (1951), 183–192. Correction in **27**:4 (1952), 52. MR Zbl

[23] H. Davenport, "On the class-number of binary cubic forms, II", *J. Lond. Math. Soc.* **26** (1951), 192–198. MR Zbl

[24] H. Davenport and H. Heilbronn, "On the density of discriminants of cubic fields, II", *Proc. Roy. Soc. Lond. Ser. A* **322**:1551 (1971), 405–420. MR Zbl

[25] S. W. Golomb, "Powerful numbers", *Amer. Math. Monthly* **77** (1970), 848–855. MR Zbl

[26] D. Gordon, D. Grenier, and A. Terras, "Hecke operators and the fundamental domain for $\mathrm{SL}(3,\mathbb{Z})$", *Math. Comp.* **48**:177 (1987), 159–178. MR Zbl

[27] D. Grenier, "Fundamental domains for the general linear group", *Pacific J. Math.* **132**:2 (1988), 293–317. MR Zbl

[28] D. Grenier, "On the shape of fundamental domains in $\mathrm{GL}(n,\mathbb{R})/\mathrm{O}(n)$", *Pacific J. Math.* **160**:1 (1993), 53–66. MR Zbl

[29] F. Gundlach, *Parametrizing extensions with fixed Galois group*, Ph.D. thesis, Princeton University, 2019, available at https://www.proquest.com/docview/2302691956.

[30] W. Ho, *Orbit parametrizations of curves*, Ph.D. thesis, Princeton University, 2009, available at https://www.proquest.com/docview/304989584.

[31] W. Ho, A. Shankar, and I. Varma, "Odd degree number fields with odd class number", *Duke Math. J.* **167**:5 (2018), 995–1047. MR Zbl

[32] J. Laga, "The average size of the 2-Selmer group of a family of non-hyperelliptic curves of genus 3", *Algebra Number Theory* **16**:5 (2022), 1161–1212. MR Zbl

[33] J. Laga, "Graded Lie algebras, compactified Jacobians and arithmetic statistics", *J. Eur. Math. Soc.* (online publication October 2024).

[34] F. Mertens, "Ueber einige asymptotische Gesetze der Zahlentheorie", *J. Reine Angew. Math.* **77** (1874), 289–338. MR Zbl

[35] D. W. Morris, *Introduction to arithmetic groups*, Deductive Press, Lethbridge, AB, 2015. MR Zbl

[36] M. Oller, "The density of ADE families of curves having squarefree discriminant", preprint, 2023. arXiv 2306.05961

[37] V. Platonov and A. Rapinchuk, *Algebraic groups and number theory*, Pure Appl. Math. **139**, Academic Press, Boston, MA, 1994. MR Zbl

[38] P. Sarnak and A. Strömbergsson, "Minima of Epstein's zeta function and heights of flat tori", *Invent. Math.* **165**:1 (2006), 115–151. MR Zbl

[39] J.-P. Serre, *A course in arithmetic*, Grad. Texts in Math. **7**, Springer, 1973. MR Zbl

[40] A. N. Shankar, "2-Selmer groups of hyperelliptic curves with marked points", *Trans. Amer. Math. Soc.* **372**:1 (2019), 267–304. MR Zbl

[41] A. Shankar and J. Tsimerman, "Counting S_5-fields with a power saving error term", *Forum Math. Sigma* **2** (2014), art. id. e13. MR Zbl

[42] A. Shankar and I. Varma, "Malle's conjecture for octic d_4-fields", in preparation.

[43] A. Shankar and X. Wang, "Rational points on hyperelliptic curves having a marked non-Weierstrass point", *Compos. Math.* **154**:1 (2018), 188–222. MR Zbl

[44] A. N. Shankar, A. Shankar, and X. Wang, "Large families of elliptic curves ordered by conductor", *Compos. Math.* **157**:7 (2021), 1538–1583. MR Zbl

[45] T. Shintani, "On Dirichlet series whose coefficients are class numbers of integral binary cubic forms", *J. Math. Soc. Japan* **24** (1972), 132–188. MR Zbl

[46] A. Siad, "Effect of monogenicity on 2-torsion in the class group of number fields of odd degree", preprint, 2020. arXiv 2011.08834

[47] A. Siad, "Monogenic fields with odd class number, II: Even degree", preprint, 2020. arXiv 2011.08842

[48] C. L. Siegel, "The average measure of quadratic forms with given determinant and signature", *Ann. of Math.* (2) **45** (1944), 667–685. MR Zbl

[49] A. A. Swaminathan, "A new parametrization for ideal classes in rings defined by binary forms, and applications", *J. Reine Angew. Math.* **798** (2023), 143–191. MR Zbl

[50] A. A. Swaminathan, "Most odd-degree binary forms fail to primitively represent a square", *Compos. Math.* **160**:3 (2024), 481–517. MR Zbl

[51] A. A. Swaminathan, "The mean number of 2-torsion elements in the class groups of cubic orders", *Comment. Math. Helv.* **100**:2 (2025), 225–267. MR Zbl

[52] J. A. Thorne, *The arithmetic of simple singularities*, 2012.

[53] L. Y. Vulakh, "Units in some families of algebraic number fields", *Trans. Amer. Math. Soc.* **356**:6 (2004), 2325–2348. MR Zbl

Communicated by Andrew Granville
Received 2022-08-24 Revised 2024-07-30 Accepted 2024-09-03

arul.shnkr@gmail.com *Department of Mathematics, University of Toronto, Toronto, ON, Canada*

as4426@princeton.edu *Department of Mathematics, Princeton University, Princeton, NJ, United States*

School of Mathematics, Institute for Advanced Study, Princeton, NJ, United States

swaminathan@math.harvard.edu *Department of Mathematics, Harvard University, Cambridge, MA, United States*

ila@math.toronto.edu *Department of Mathematics, University of Toronto, Toronto, ON, Canada*

ALGEBRA AND NUMBER THEORY 19:6 (2025)
https://doi.org/10.2140/ant.2025.19.1147

msp

Explicit isogenies of prime degree over number fields

Barinder S. Banwait and Maarten Derickx

Dedicated to the memory of Sebastiaan Johan Edixhoven, 1962–2022

We provide an explicit and algorithmic version of a theorem of Momose classifying isogenies of prime degree of elliptic curves over number fields, which we implement in Sage and PARI/GP. Combining this algorithm with recent work of Box, Gajović and Goodman we obtain the first classifications of the possible prime degree isogenies of elliptic curves over cubic number fields, as well as for several quadratic fields not previously known. While the correctness of the general algorithm relies on the generalised Riemann hypothesis, the algorithm is unconditional for the restricted class of semistable elliptic curves.

1. Introduction

Let k be a number field, and consider the set $\mathsf{IsogPrimeDeg}(k)$ of primes p which arise as k-rational isogenies of degree p as one varies over all elliptic curves over k; we refer to such p as *isogeny primes for* k. The set $\mathsf{IsogPrimeDeg}(k)$ is necessarily infinite if k contains the Hilbert class field of an imaginary quadratic field, which follows from the basic theory of complex multiplication on elliptic curves, and it is a consequence of work of Momose [1995] with Merel's proof of uniform boundedness for torsion on elliptic curves [Merel 1996] that the converse holds assuming the generalised Riemann hypothesis (GRH). This consequence was written in the literature explicitly by Larson and Vaintrob.

Theorem 1.1 [Larson and Vaintrob 2014, Corollary 2]. *Assume the generalised Riemann hypothesis. For a number field* k, $\mathsf{IsogPrimeDeg}(k)$ *is finite if and only if* k *does not contain the Hilbert class field of an imaginary quadratic field.*

The question of exactly determining $\mathsf{IsogPrimeDeg}(k)$ when it is finite originated with the seminal work of Mazur who determined the base case of $k = \mathbb{Q}$.

Theorem 1.2 [Mazur 1978, Theorem 1]. $\mathsf{IsogPrimeDeg}(\mathbb{Q}) = \{2, 3, 5, 7, 11, 13, 17, 19, 37, 43, 67, 163\}$.

Recent work of Banwait [2023] found the first instances of the determination of $\mathsf{IsogPrimeDeg}(k)$ for some quadratic fields.

MSC2020: primary 11G05; secondary 11G15, 11Y60.
Keywords: elliptic curves, isogenies, prime degree.

Theorem 1.3 (Banwait). *Assuming GRH, we have*

$$\begin{aligned}\mathsf{IsogPrimeDeg}(\mathbb{Q}(\sqrt{-10})) &= \mathsf{IsogPrimeDeg}(\mathbb{Q}),\\ \mathsf{IsogPrimeDeg}(\mathbb{Q}(\sqrt{5})) &= \mathsf{IsogPrimeDeg}(\mathbb{Q}) \cup \{23, 47\},\\ \mathsf{IsogPrimeDeg}(\mathbb{Q}(\sqrt{7})) &= \mathsf{IsogPrimeDeg}(\mathbb{Q}).\end{aligned}$$

More generally, an algorithm was presented in [Banwait 2023] which took as input a quadratic field which is not imaginary quadratic of class number one, and output a small superset for $\mathsf{IsogPrimeDeg}(k)$. Results on quadratic points on low-genus modular curves [Bruin and Najman 2015; Ozman and Siksek 2019; Box 2021] together with work on everywhere local solubility of twists of modular curves [Ozman 2012] then allowed one to pass from the superset to the actual set itself.

This algorithm relied crucially on work of Momose [1995, Theorem A/Theorem 1], who proved that if there exists an elliptic curve admitting a k-rational p-isogeny, then for p larger than a constant C_k depending only on k, the associated isogeny character obtained from considering the Galois action on the kernel of the isogeny must be one of three defined "types", hereafter referred to as *Momose types* 1, 2 *and* 3. We refer to this result as *Momose's isogeny classification theorem*. We will recap the definition of these "types" in Section 3.4; for now we mention only that Momose type 3 requires that k contain the Hilbert class field of an imaginary quadratic field, and that, with this assumption, the infinitely many resulting isogeny primes from CM elliptic curves are of Momose type 3. In principle this reduces the task of bounding isogeny primes for k (when this set of isogeny primes is finite) to bounding isogeny primes arising from Momose types 1 and 2. The former appears as Theorem 3 of [Momose 1995], and the latter — which requires GRH — as Remark 8 of [loc. cit.]. Taken together we refer to these three results as *Momose's isogeny theorems*, and we note furthermore that Momose assumed in his work that p is unramified in k.

Momose did not make the constant C_k explicit; this was done subsequently by David [2011b, Section 2.4; 2008, Section 2.3] in the case that k is Galois over $\mathbb{Q}$. In the course of this work, David gave a more precise and careful treatment of the proof of Momose's Theorem 1 in the Galois case; in particular a reproof of Momose's Lemmas 1 and 2 [1995] (which originally assumed that k is Galois over $\mathbb{Q}$); see Section 2 of [Banwait 2023] for an overview of David's work and for precise references to it. We note that David also assumed that p is unramified in k.

While Momose's Lemmas 1 and 2 *do* assume that k is Galois over $\mathbb{Q}$, Momose gives an argument in the proof of Theorem 1 to cover the non-Galois case by passing to the Galois closure K over $\mathbb{Q}$. Unfortunately, the details given there contain some mistakes and gaps, which we indicate as the first three items of the following; the last item indicates a mistake in Momose's Theorem 3:

(1) Momose defines a certain group ring character ε at the outset of the proof, but it is unclear whether he is taking this to be over k or over K. Based on his exposition, and his definition of d as $[k : \mathbb{Q}]$, it appears that he is taking the base field to be k; however he is using Lemma 1 to define ε, and Lemma 1

assumes that the field is Galois over $\mathbb{Q}$, suggesting he is taking K to be the ground field. It is not *a priori* clear that such an ε exists over non-Galois number fields k.

(2) Later in the proof Momose takes a set of generators of the class group of k consisting of completely split primes in k. Such a generating set only necessarily exists in the Galois case.

(3) Immediately after the displayed equation (2), Momose writes "In the case of type 2, $\beta = \zeta\sqrt{-q}$ for a $12h$-th root ζ of unity". This would follow if one had the interpretation of "type 2" as $\varepsilon = 6\,\mathrm{Nm}_{k/\mathbb{Q}}$ as promised by Lemma 2. However, again, since Lemma 2 also assumes that the ground field is Galois over $\mathbb{Q}$, this move is not valid.

(4) In the proof of Theorem 3 of [Momose 1995], Momose concludes that $p - 1 \mid 12h_k$ in the potentially multiplicative reduction case (h_k being the class number of k); however this bound is too strict, and should rightly be

$$p \mid \mathrm{Nm}(\mathfrak{q})^{12h_k} - 1$$

for $\mathfrak{q}$ a prime of k coprime to p. See Remark 5.10 and its preceding discussion for more details.

One of the main contributions of this paper is to offer corrected proofs of Momose's isogeny theorems which furthermore strengthen them to deal with the case that p ramifies in k. In addition, since our chief motivation is to compute exact sets of isogeny primes, we take this opportunity to recast Momose's Theorem 1 into a more algorithmic framing, providing a generalisation of the previous algorithm of the first named author to arbitrary number fields. This algorithm may be found as Algorithm 3.26, and it yields the following explicit version of Momose's isogeny classification theorem.

Theorem 1.4. *Let k be a number field. Then Algorithm 3.26 computes a nonzero integer* $\mathsf{MMIB}(k)$ *such that, if p is an isogeny prime for k whose associated isogeny character is not of Momose type* 1, 2 *or* 3, *then p divides* $\mathsf{MMIB}(k)$.

We refer to $\mathsf{MMIB}(k)$ as the *Momose multiplicative isogeny bound* of k. Contrary to Momose's passing to the Galois closure of k over $\mathbb{Q}$, our approach will be to strengthen Momose's Lemmas 1 and 2 (or rather, David's versions of these lemmas) to remove the Galois assumption, which furthermore provides for a favourable improvement to the algorithm (see Remark 4.5).

Since isogenies arising from CM elliptic curves are necessarily of Momose type 3, we cannot hope to bound such "Momose type 3 isogeny primes". Thus, in our attempt to find a multiplicative bound on $\mathsf{IsogPrimeDeg}(k)$, we are reduced to bounding isogeny primes which arise from isogeny characters of Momose types 1 or 2.

By building on the explicit criteria — given by work of Derickx, Kamienny, Stein and Stoll [Derickx et al. 2023] — for when the natural map $X_0(p)^d \to J_e$ of the d-th symmetric power modular curve into the winding quotient of $J_0(p)$ is a formal immersion in positive characteristic — we are able to explicitly and algorithmically determine a multiplicative bound on Momose type 1 primes.

Theorem 1.5. *Let k be a number field. Then Algorithm 5.1 computes a nonzero integer* $\mathsf{TypeOneBound}(k)$ *such that, if p is an isogeny prime for k whose associated isogeny character is of Momose type* 1*, then p divides* $\mathsf{TypeOneBound}(k)$.

Algorithms 3.26 and 5.1 referenced respectively in Theorems 1.4 and 1.5 are of a similar nature, and are implemented in the computer algebra system Sage [2021]. Dealing with isogeny primes of Momose type 2, however, requires a different approach. Rather than identifying integers which such isogeny primes must divide, we instead, on the one hand, identify a certain necessary condition (Proposition 6.1) which such isogeny primes must satisfy; and then, on the other hand, determine an upper bound — conditional on GRH — for such isogeny primes. This is the only result in our work which requires GRH. The situation is summarised as follows, the details being given in Section 6.

Theorem 1.6. *Let k be a number field of degree d and discriminant Δ_k, and let E/k be an elliptic curve admitting a k-rational p-isogeny of Momose type* 2*. Then we have the following*:

(1) *E is not semistable.*

(2) *The pair (k, p) satisfies the necessary condition in Proposition 6.1.*

(3) *Assuming GRH, p satisfies*

$$p \leq (8d\log(12p) + 16\log(\Delta_k) + 10d + 6)^4 \tag{1-1}$$

and hence there are only finitely many isogeny primes of Momose type 2*.*

Determining whether a pair (k, p) satisfies the necessary condition of Proposition 6.1 is a matter of checking splitting conditions on primes in the imaginary quadratic field $\mathbb{Q}(\sqrt{-p})$, and is thus easily checked via Legendre symbol computations. However, the conditional bound implied by (1-1) becomes rather large, as illustrated in Table 1, which shows the conditional bounds on type 2 primes for the number fields of smallest absolute discriminant and class number one for each $2 \leq d \leq 10$ which do not contain the Hilbert class field of an imaginary quadratic field. Checking the necessary condition of Proposition 6.1 on all primes up to this bound is the content of Algorithm 6.4, and is implemented in PARI/GP [2021].

Combining Algorithms 3.26, 5.1 and 6.4 into Algorithm 8.1, we may summarise Theorems 1.4–1.6 as follows. By *semistable isogeny primes for k* we mean the isogeny primes for k obtained by varying over only semistable elliptic curves.

Theorem 1.7. *Let k be a number field. Then Algorithm 8.1 outputs a finite set of primes S_k such that, if p is an isogeny prime for k whose associated isogeny character is not of Momose type* 3*, then, conditional on GRH, $p \in S_k$. In particular*:

(1) *If k does not contain the Hilbert class field of an imaginary quadratic field, then S_k contains* $\mathsf{IsogPrimeDeg}(k)$.

(2) *The above results are unconditional for the restricted set of semistable isogeny primes for k.*

d	Δ_K	LMFDB label	type 2 bound
2	5	2.2.5.1	5.65×10^{10}
3	49	3.3.49.1	4.09×10^{11}
4	125	4.0.125.1	1.46×10^{12}
5	14641	5.5.14641.1	4.75×10^{12}
6	300125	6.6.300125.1	1.12×10^{13}
7	594823321	7.7.594823321.1	2.65×10^{13}
8	64000000	8.0.64000000.2	4.16×10^{13}
9	16983563041	9.9.16983563041.1	7.60×10^{13}
10	572981288913	10.10.572981288913.1	1.24×10^{14}

Table 1. The bound, conditional on GRH, on type 2 primes for the smallest Galois number fields of class number one not containing the Hilbert Class field of an imaginary quadratic field for each degree between 2 and 10.

d	Δ_k	LMFDB label	possible isogeny primes	time(s)
2	5	2.2.5.1	23, 47	0.92
3	49	3.3.49.1	23, 29, 31, 73	3.23
4	125	4.0.125.1	23, 29, 31, 41, 47, 53, 61, 73, 97, 103	3.77
5	14641	5.5.14641.1	23, 29, 31, 41, 47, 59, 71, 73, 97	35.42
6	300125	6.6.300125.1	23, 29, 31, 41, 47, 53, 59, 61, 71, 73, 79, 83, 97, 103, 107, 109, 113, 127, 131, 191, 211, 263, 311, 503	206.62

Table 2. The possible isogeny primes $p \notin \mathsf{IsogPrimeDeg}(\mathbb{Q})$ for the number fields from Table 1 for $2 \leq d \leq 6$, as well as the time taken to obtain this running the algorithm on an old laptop. The time only measures checking possible Momose type 2 primes up to 10^6; checking all up to the conditional bound adds several hours to the runtime in each case.

Here we have highlighted that our results are unconditional for semistable isogeny primes since the restricted class of semistable elliptic curves arises naturally in the context of Frey curves and solving Diophantine equations [Freitas and Siksek 2015]. Moreover, Algorithm 8.1 does more than merely combine the different subalgorithms dealing with each of the isogeny types in Momose's classification; it also includes several methods for ruling out possible isogeny primes, based on congruence conditions, class field theory, and explicit computations with Jacobians of modular curves. These methods are discussed in Section 7.

To give the reader a sense of the size of this superset S_k for $\mathsf{IsogPrimeDeg}(k)$ as well as the runtime, we show in Table 2 the output of the algorithm on the number fields from Table 1 for $2 \leq d \leq 6$, as well as the time taken for this to complete on an old laptop. Since the output necessarily contains $\mathsf{IsogPrimeDeg}(\mathbb{Q})$ in each case, we show only those possible primes *not* in $\mathsf{IsogPrimeDeg}(\mathbb{Q})$. In addition, in Table 2 the timings refer to only checking possible Momose type 2 primes up to 10^6; checking up to the bounds given in Table 1 is the main bottleneck of the algorithm, and doing so for each number field in the table takes several hours of parallel computation in PARI/GP, which results in no additional possible isogeny primes beyond those listed.

Determining which of the primes p in these supersets are actually isogeny primes for k requires one to determine whether $X_0(p)(k)$ contains any noncuspidal points, a subject which has hardly been explicitly studied for $\deg(k) \geq 3$. However, by building on recent work of Box, Gajović and Goodman [Box et al. 2023] in which the authors determine the finitely many cubic points on some modular curves $X_0(N)$, we are able to exactly determine $\mathsf{IsogPrimeDeg}(k)$ for some cubic fields, yielding the first instances of the determination of $\mathsf{IsogPrimeDeg}(k)$ for number fields of degree at least three.

Theorem 1.8. *Assuming GRH, we have*

$$\begin{aligned}\mathsf{IsogPrimeDeg}(\mathbb{Q}(\zeta_7)^+) &= \mathsf{IsogPrimeDeg}(\mathbb{Q}),\\ \mathsf{IsogPrimeDeg}(\mathbb{Q}(\alpha)) &= \mathsf{IsogPrimeDeg}(\mathbb{Q}) \cup \{29\},\\ \mathsf{IsogPrimeDeg}(\mathbb{Q}(\beta)) &= \mathsf{IsogPrimeDeg}(\mathbb{Q}).\end{aligned}$$

where $\alpha^3 - \alpha^2 - 2\alpha - 20 = 0$ *and* $\beta^3 - \beta^2 - 3\beta + 1 = 0$.

In addition to cubic fields, our algorithm provides for several key improvements which allow us to determine several more instances of quadratic isogeny primes. The following is obtained by combining our algorithm with standard techniques for working explicitly with modular curves in Magma [1997]; see [Banwait 2023, Section 7] for an overview of these techniques.

Theorem 1.9. *Let* $D \neq 1$ *be a squarefree integer such that* $|D| < 50$ *and* $\mathbb{Q}(\sqrt{D})$ *is not imaginary quadratic of class number* 1. *Then* $\mathsf{IsogPrimeDeg}(\mathbb{Q}(\sqrt{D})) = \mathsf{IsogPrimeDeg}(\mathbb{Q})$ *unless* D *is listed in Table 3, in which case the additional isogeny primes are listed in the column "new isogeny primes", and any primes not yet determined are listed in the column "undetermined"'.*

The implementation of the combined Algorithm 8.1 is available at https://github.com/isogeny-primes/isogeny-primes. As with its predecessor [Banwait 2021], it has been released as a command line tool under the GPLv3+ licence and the README.md contains detailed instructions on its use and an overview of the testing strategy. All filenames will refer to files in this repository [Banwait and Derickx 2021].

The outline of the paper is as follows: Section 2 sets the notation to be used throughout the paper as well as some basic results about roots of characteristic polynomials of Frobenius of elliptic curves over finite fields. The heart of the paper is Section 3 which strengthens and fixes Momose's isogeny classification theorem, and presents the algorithmic version of it. Section 4 discusses some optimisation aspects of the implementation of the algorithm dealing with the so-called *generic isogeny primes*, those which arise from isogeny characters which are not of type 1, 2 or 3.[1] Sections 5 and 6 deal respectively with isogeny primes arising from isogeny characters of Momose type 1 and 2, and Section 7 presents methods to further eliminate possible isogeny primes, based largely on methods of explicit class field theory. Combining all of the algorithms up to this point into the main Algorithm 8.1 is done in Section 8. Finally in Section 9 we present results related to cubic points on modular curves, and prove Theorem 1.8. (The verification required for Theorem 1.9 is given in magma_scripts/QuadraticVerifs.m in the above repository.)

[1] Here we mean "signature type" rather than "Momose type". This distinction will be made clear in Section 3.

D	new isogeny primes	undetermined
-47	31	61
-39	–	97
-37	–	59, 131
-31	73	–
-23	29, 31	–
-15	23	–
-5	23	–
5	23, 47	–
13	31	–
17	–	23
29	29	–
37	–	23
41	41	–
47	–	59

Table 3. Determination of $\mathsf{IsogPrimeDeg}(\mathbb{Q}(\sqrt{D}))$ for squarefree $|D| < 50$, excluding the nine imaginary quadratic fields of class number one. If D is not listed here, then $\mathsf{IsogPrimeDeg}(\mathbb{Q}(\sqrt{D})) = \mathsf{IsogPrimeDeg}(\mathbb{Q})$. The primes in "new isogeny primes" have been verified to be isogeny primes; it is currently not known whether the primes in "undetermined" are isogeny primes for $\mathbb{Q}(\sqrt{D})$ or not.

2. Notation and preliminaries

In this section we set the notation for relevant objects to be used throughout the paper. Certain sections below will introduce their own notation in addition to those set here. This is the default notation to be taken, unless a particular result overrides the notation set here. We will also prove some results of a preliminary nature to be used later:

d : an integer ≥ 1
k : a number field of degree d
Cl_k : the class group of k
h_k : the class number of k
G_k : $\mathrm{Gal}(\bar{k}/k)$, the absolute Galois group of k
K : the Galois closure of k over $\mathbb{Q}$
Σ : $\mathrm{Hom}(k, K)$, the embeddings of k in K
p : a rational prime (denoting the isogeny prime we seek to bound)
χ_p : the mod-p cyclotomic character of G_k
$\mathfrak{p}_0$: a chosen and fixed prime ideal of K lying above p
q : a rational prime different from p
$\mathfrak{q}$: a prime ideal of k lying above q
$h_{\mathfrak{q}}$: the order of $\mathfrak{q}$ in Cl_k

$\gamma_\mathfrak{q}$: a generator of $\mathfrak{q}^{h_\mathfrak{q}}$ in Cl_k
$\mathbb{F}_\mathfrak{q}$: the residue field of $\mathfrak{q}$
$\mathrm{Frob}_\mathfrak{q}$: a choice of lift of Frobenius in G_k
$\iota_\mathfrak{q}$: the embedding of k into its $\mathfrak{q}$-adic completion
E : an elliptic curve over k admitting a k-rational p-isogeny
λ : the isogeny character of E; that is, the Galois action on the isogeny's kernel
μ : λ^{12}
ε : the signature of λ (see Section 3.1).

Note that μ is unramified outside of p [David 2011b, Propositions 1.4 and 1.5; 2011a, Propositions 3.3 and 3.5] and its image is abelian, hence $\mu(\mathrm{Frob}_q)$ is well defined. Neither of these facts are true for λ itself, which is why we take the twelfth power; it is then expedient and minimises overload of notation to introduce a new symbol for λ^{12}.

Remark 2.1. By precomposing with the Artin map from class field theory, we may view each of λ and μ as character on the group of fractional ideals of k coprime to p. This is done in a minority of places in the paper (specifically, Proposition 3.4, the proof of Lemma 4.3, and Section 7.2), and is done largely so that one may employ the following convenient shorthand notation for $\mu((\alpha))$:

$$\mu((\alpha)) = \prod_\mathfrak{q} \mu(\mathrm{Frob}_\mathfrak{q})^{v_\mathfrak{q}(\alpha)}.$$

The following result will be used in Section 3.4.

Proposition 2.2. *Let L be an imaginary quadratic field, q an odd rational prime, $\mathfrak{q}$ a prime of $\mathcal{O}_L$ above q, and f a positive integer such that $\mathfrak{q}^f = \overline{\mathfrak{q}^f}$. Suppose that $\mathfrak{q}^f$ is principal, generated by $\alpha_\mathfrak{q} \in L$. Then $\alpha_\mathfrak{q}^{12} = \mathrm{Nm}_{L/\mathbb{Q}}(\mathfrak{q})^{6f}$.*

Proof. We split up the prove according to the splitting behaviour of q in L. Firstly, if q splits in L then $\mathfrak{q}^f \neq \overline{\mathfrak{q}^f}$ and there is nothing to prove. Secondly, if q is inert in L then $\mathfrak{q} = q\mathcal{O}_L$ and $\mathrm{Nm}(\mathfrak{q}) = q^2$. This implies $\alpha_\mathfrak{q}\mathcal{O}_L = q^f\mathcal{O}_L$ and hence that $\alpha_\mathfrak{q}/q^f$ is a unit. Since L is imaginary quadratic this means $\alpha_\mathfrak{q}^{12} = q^{12f} = \mathrm{Nm}(\mathfrak{q})^{6f}$.

Finally, suppose that q ramifies in L. Then $\mathfrak{q}^2 = q\mathcal{O}_L$ and $\mathrm{Nm}(\mathfrak{q}) = q$. This implies $\alpha_\mathfrak{q}^2\mathcal{O}_L = q^f\mathcal{O}_L$ and hence that $\alpha_\mathfrak{q}^2/q^f$ is a unit. Since $q > 2$ is ramified in L we know $L \neq \mathbb{Q}(i)$ and hence the order of this unit divides 6. This means $\alpha_\mathfrak{q}^{12} = q^{6f} = \mathrm{Nm}(\mathfrak{q})^{6f}$. □

Remark 2.3. The condition $q > 2$ is necessary, since $\mathfrak{q} := (i+1)\mathbb{Z}[i]$ is an ideal invariant under complex conjugation and $(i+1)^{12} = -64 = -\mathrm{Nm}(\mathfrak{q})^6 \neq \mathrm{Nm}(\mathfrak{q})^6$. However when $q = 2$ one can in general still conclude that $\alpha_\mathfrak{q}^{24} = \mathrm{Nm}(\mathfrak{q})^{12f}$.

Remark 2.4. The above proposition is in fact true if one replaces $\mathfrak{q}^f$ by an arbitrary principal ideal I coprime to 2 such that $I = \bar{I}$, but we won't need this.

The remainder of this section will collect consequences and explicit statements derived from Theorem 4.1 of [Waterhouse 1969], which is an explicit version of the Honda–Tate theorem that describes isogeny classes of abelian varieties in terms of characteristic polynomials of Frobenius. We will first state the theorem using notation which is compatible with the variables used in our work.

Theorem 2.5 (Waterhouse). *Let q be a prime and f an integer. The polynomials that occur as the characteristic polynomial of Frobenius of an elliptic curve over $\mathbb{F}_{q^f}$ are exactly the polynomials of the form $x^2 + tx + q^f$ with t an integer such that $|t| \leq 2\sqrt{q^f}$ that satisfy one of the following conditions*:

(1) $(t, q) = 1$.

(2) *f is even and $t = \pm 2q^{f/2}$.*

(3) *f is even, $q \not\equiv 1 \pmod 3$ and $t = \pm q^{f/2}$.*

(4) *f is odd, $q = 2, 3$ and $t = \pm q^{(f+1)/2}$.*

(5) (i) *f is odd and $t = 0$.*
(ii) *f is even, $p \not\equiv 1 \pmod 4$ and $t = 0$.*

The first case only occurs for ordinary elliptic curves, while the other cases only occur for supersingular elliptic curves.

The description of the roots of the characteristic polynomials of Frobenius for supersingular elliptic curves given in Theorem 2.5 is summarised in Table 4. In the sequel we refer to such roots as *supersingular Frobenius roots*, while by *ordinary Frobenius roots* we shall mean the roots of the characteristic polynomials of Frobenius for ordinary elliptic curves.

Corollary 2.6. *Let β be an ordinary Frobenius root of an elliptic curve over $\mathbb{F}_{q^f}$. Then for all $n \geq 1$, β^n is not rational.*

Proof. Let $E/\mathbb{F}_{q^f}$ be an ordinary elliptic curve with characteristic polynomial of Frobenius equal to $(x - \beta)(x - \bar{\beta})$. Then $(x - \beta^n)(x - \bar{\beta}^n)$ is the characteristic polynomial of Frobenius of E viewed over $\mathbb{F}_{q^{fn}}$. Since being ordinary is invariant under base change one has that β^n is a root of a polynomial $x^2 + t'x + q^{fn}$ of the form of Case 1 of Theorem 2.5; that is, $\gcd(t', q) = 1$, and therefore $|t'| < 2\sqrt{q^f}$, which implies that the discriminant of this polynomial is negative. In particular β^n is the root of an irreducible polynomial, and hence not rational. □

Corollary 2.7. *Let β be a supersingular Frobenius root of an elliptic curve over $\mathbb{F}_{q^f}$. Then β^{12} is rational.*

Proof. This is clear from the last column of Table 4. □

Corollary 2.8. *Let E/k be an elliptic curve over a number field, and $\mathfrak{q}$ a prime of k of odd residue characteristic q such that E admits potentially good supersingular reduction at $\mathfrak{q}$. Let β be a root of the characteristic polynomial of Frobenius at $\mathfrak{q}$ acting on the p-adic Tate module of E (for any $p \neq q$; this is independent of the choice of p). Then $\beta^{12} = \mathrm{Nm}(\mathfrak{q})^6$.*

case of Theorem 2.5	f	t	β	β^{12}
(2)	even	$\pm 2q^{f/2}$	$\mp q^{f/2}$	q^{6f}
(3)	even	$\pm q^{f/2}$	$\pm\zeta_3 q^{f/2}$ or $\pm\zeta_3^2 q^{f/2}$	q^{6f}
(4)	odd	$\pm 2^{(f+1)/2}$	$\pm(i-1)2^{(f-1)/2}$ or $\pm(-i-1)2^{(f-1)/2}$	$-q^{6f}$
		$\pm 3^{(f+1)/2}$	$\pm(\zeta_3-1)3^{(f-1)/2}$ or $\pm(\zeta_3^2-1)3^{(f-1)/2}$	q^{6f}
(5)(i)	odd	0	$\sqrt{-q^f}$	q^{6f}
(5)(ii)	even		$iq^{f/2}$	q^{6f}

Table 4. Description of the roots of the characteristic polynomials of Frobenius of supersingular elliptic curves.

Proof. By considering the different types of additive reduction from the Kodaira–Néron classification of the special fibre of the Néron model of E over $\mathcal{O}_k$, the elliptic curve E attains good reduction over a totally ramified extension L of k of degree 1, 2, 3, 4 or 6. Since L is totally ramified, the reduction of E/L at the unique prime of L above $\mathfrak{q}$ is an elliptic curve over the residue field $\mathbb{F}_\mathfrak{q}$ of k at $\mathfrak{q}$, and hence β is a root of the characteristic polynomial of Frobenius of a supersingular elliptic curve over $\mathbb{F}_\mathfrak{q}$. This elliptic curve over $\mathbb{F}_\mathfrak{q}$ is not unique, since it depends on the choice of L; however another choice of L will only lead to a curve which is a quadratic, quartic or sextic twist of E. Thus β^{12} is independent of this choice.

Writing f for the residue field degree of $\mathfrak{q}$, we see from the last column of Table 4 that $\beta^{12} = \mathrm{Nm}(\mathfrak{q})^6$. □

We conclude this background section with a recap on Serre's fundamental characters.

2.1. ***Fundamental characters of level n.*** The canonical reference for the material here is [Serre 1972, Section 1].

For our number field k and prime p, let $\mathfrak{p}$ be a prime of k over p, and consider the tame inertia group $I_\mathfrak{p} = \mathrm{Gal}(\overline{k_\mathfrak{p}}/k_\mathfrak{p}^{nr})$, where $k_\mathfrak{p}^{nr}$ is the maximal unramified extension of $k_\mathfrak{p}$ in $\overline{k_\mathfrak{p}}$. Both of these fields have the same residue field, which is an algebraic closure of $\mathbb{F}_p$ that we denote by $\overline{\mathbb{F}_p}$.

Inside $\overline{\mathbb{F}_p}$ one has all finite extensions $\mathbb{F}_{p^n}$ of $\mathbb{F}_p$ with norm maps $\mathbb{F}_{p^m}^\times \to \mathbb{F}_{p^n}^\times$ whenever $n \mid m$, and there is a natural identification θ of $I_\mathfrak{p}$ with the inverse limit of this system of norm maps [Serre 1972, Proposition 2].

We therefore obtain, for each power $q = p^n$, the natural projection

$$\theta_{q-1} : I_\mathfrak{p} \to \mathbb{F}_q^\times;$$

this may be considered *the* fundamental character of level n, and by composing it with an automorphism of $\mathbb{F}_q^\times$ (i.e., a power ϕ^i of the Frobenius automorphism $\phi : x \mapsto x^p$) we obtain all other fundamental characters of level n, viz. $\theta_{q-1}^{p^i} := \phi^i \circ \theta_{q-1}$, for $i = 0, \ldots, n-1$.

3. On Momose's isogeny classification theorem

In this section we generalise and make Theorem 1 of [Momose 1995] algorithmic, and address the gaps and mistakes in the original proof in [loc. cit.] mentioned in the introduction.

3.1. ***Strengthening Momose's Lemma 1.*** In this subsection we provide a version of Momose's Lemma 1 without the assumptions that p is unramified in k or that k is Galois over $\mathbb{Q}$.

The first step is to describe the possible integers a_σ which can occur in the statement of Momose's Lemma 1. A fuller description of these integers was given by David [2011a, Proposition 3.2] under the assumption that p is unramified in k. We thus begin by providing the more general version of this result of David.

Let $\mathfrak{p}$ be a prime of k lying above p, $I_\mathfrak{p} = \mathrm{Gal}(\overline{k_\mathfrak{p}}/k_\mathfrak{p}^{nr}) \subseteq \mathrm{Gal}(\overline{k_\mathfrak{p}}/k_\mathfrak{p})$ the inertia subgroup at $\mathfrak{p}$ and let μ be the twelfth power of a p-isogeny character over $k_\mathfrak{p}$. Then we would like to describe the possible actions of μ restricted to $I_\mathfrak{p}$.

This is provided by following result, which is a generalisation of Proposition 3.2 of [David 2011a], since for an unramified extension, the zeroth fundamental character θ_{p-1} is equal to the cyclotomic character χ_p restricted to $I_\mathfrak{p}$. See Section 2.1 for more on fundamental characters.

Proposition 3.1. *Let $\mu = \lambda^{12}$ be the twelfth power of a p-isogeny character, and $\mathfrak{p} \mid p$ a prime of k with ramification index $e_\mathfrak{p}$. Suppose that $p \geq 5$. Then there exists an integer $0 \leq a_\mathfrak{p} \leq 12e_\mathfrak{p}$ such that $\mu|_{I_\mathfrak{p}} = \theta_{p-1}^{a_\mathfrak{p}}$. We have $a_\mathfrak{p} \equiv 0, 4, 6$ or $8 \pmod{12}$, and if E is semistable, then $a_\mathfrak{p} \equiv 0 \pmod{12}$.*

Proof. This proof is essentially the same as the proof of Proposition 3.2 of [David 2011a], the main difference being that we here are keeping track of the ramification index $e_\mathfrak{p}$ (which for David was equal to 1). In the interest of being self-contained, we supply the details.

In the case that we have potentially multiplicative reduction we have $\lambda^2|_{I_\mathfrak{p}}$ is either trivial or χ_p^2. Now since $\theta_{p-1}^{e_\mathfrak{p}} = \chi_p$ it follows that $\mu|_{I_\mathfrak{p}} = \theta_{p-1}^{a_\mathfrak{p}}$, where $a_\mathfrak{p} = 0$ or $a_\mathfrak{p} = 12e_\mathfrak{p}$.

In the case that we have potentially good reduction there exists an integer $a'_\mathfrak{p}$ such that $\lambda|_{I_\mathfrak{p}} = \theta_{p-1}^{a'_\mathfrak{p}}$. Additionally there exists a purely ramified extension $k_\mathfrak{p} \subseteq k'_\mathfrak{p}$ over which the elliptic curve E obtains good reduction. Writing $\widetilde{E_\mathfrak{p}}$ for this reduced elliptic curve, we have that the degree $e'_\mathfrak{p} := [k'_\mathfrak{p} : k_\mathfrak{p}]$ of this extension divides the size of the geometric automorphism group $\mathrm{Aut}(\widetilde{E_\mathfrak{p}})$; see the proof of Theorem 2 in Section 2 of [Serre and Tate 1968]. Since $p \geq 5$, we obtain that $e'_\mathfrak{p}$ divides 2, 4 or 6. Let $I'_\mathfrak{p}$ denote the inertia subgroup of $\mathrm{Gal}(\overline{k'_\mathfrak{p}}/k'_\mathfrak{p})$. Then $e'_\mathfrak{p}e_\mathfrak{p}$ is the ramification degree of $k'_\mathfrak{p}/\mathbb{Q}_p$. Writing θ'_{p-1} for the fundamental character of level one of $k'_\mathfrak{p}$, we have on $I'_\mathfrak{p}$ the equality $\theta_{p-1} = (\theta'_{p-1})^{e'_\mathfrak{p}}$. Since E has good reduction over $k'_\mathfrak{p}$ we know by [Serre 1972, Section 1.13] that there exists an integer $0 \leq r_\mathfrak{p} \leq e'_\mathfrak{p}e_\mathfrak{p}$ such that on $I'_\mathfrak{p}$ we have $\lambda = (\theta'_{p-1})^{r_\mathfrak{p}}$. Putting all of these relations between characters on $I'_\mathfrak{p}$ together we obtain

$$(\theta'_{p-1})^{r_\mathfrak{p}} = (\theta_{p-1})^{a'_\mathfrak{p}} = (\theta'_{p-1})^{e'_\mathfrak{p}a'_\mathfrak{p}}.$$

Since the order of θ'_{p-1} is $p-1$ one obtains

$$r_\mathfrak{p} \equiv e'_\mathfrak{p}a'_\mathfrak{p} \bmod p-1.$$

Multiplying by the integer $12/e'_{\mathfrak{p}}$ gives $(12/e'_{\mathfrak{p}})r_{\mathfrak{p}} \equiv (12/e'_{\mathfrak{p}})e'_{\mathfrak{p}}a'_{\mathfrak{p}} \equiv 12a'_{\mathfrak{p}} \bmod p-1$, so that

$$\mu|_{I_{\mathfrak{p}}} = \lambda^{12}|_{I_{\mathfrak{p}}} = \theta_{p-1}^{a_{\mathfrak{p}}},$$

with $a_{\mathfrak{p}} = (12/e'_{\mathfrak{p}})r_{\mathfrak{p}}$. Observe that the possible range of $r_{\mathfrak{p}}$ gives that $0 \leq a_{\mathfrak{p}} \leq 12e_{\mathfrak{p}}$.

We now show that $a_{\mathfrak{p}} = (12/e'_{\mathfrak{p}})r_{\mathfrak{p}} \equiv 0, 4, 6$ or $8 \bmod 12$. This can be done by studying the possible values $1, 2, 3, 4, 6$ of $e'_{\mathfrak{p}}$ separately. In the cases $e'_{\mathfrak{p}} = 1, 2$ or 3 one has $12/e'_{\mathfrak{p}} \equiv 0, 4, 6$ or $8 \bmod 12$, hence $a_{\mathfrak{p}}$ satisfies that congruence as well. If $e'_{\mathfrak{p}} = 4$ respectively 6, then $12/e'_{\mathfrak{p}} = 3$ respectively 2. Furthermore, $r_p \equiv e'_{\mathfrak{p}}a'_{\mathfrak{p}} \bmod p-1$ implies $r_{\mathfrak{p}}$ is even, which gives the result in these two cases also.

Finally we establish that, if E is semistable, then we have $a_{\mathfrak{p}} \equiv 0 \pmod{12}$. If E has multiplicative reduction, then this was established at the very beginning of the proof. If E has good reduction, then this follows from observing that the integer $e'_{\mathfrak{p}}$ is equal to 1, whence $a_{\mathfrak{p}} = 12r_{\mathfrak{p}}$. □

Remark 3.2. The character θ_{p-1} is surjective, so in particular the residue class $a_{\mathfrak{p}} \bmod p-1$ is determined by μ. Thus if $p-1 > 12e_{\mathfrak{p}}$ then the integer $a_{\mathfrak{p}}$ is unique.

In order to get the results we need without the Galois assumption we will first make the following definition following Freitas and Siksek [2015]; see the discussion just before Proposition 2.2 in [loc. cit.].

Definition 3.3. Let k be a number field with Galois closure K over $\mathbb{Q}$. For each $\sigma \in \Sigma := \mathrm{Hom}(k, K)$ let $a_\sigma \in \{0, 4, 6, 8, 12\}$ be an integer and denote by $\varepsilon = \sum_\sigma a_\sigma \sigma$ a formal sum. Then ε is called a *k-isogeny signature*.

For $\alpha \in k^\times$, and $\varepsilon = \sum_\sigma a_\sigma \sigma$ a k-isogeny signature we prefer to use Momose's notation α^ε to denote what David calls $\mathcal{N}(\alpha)$:

$$(-)^\varepsilon : k \to K, \quad \alpha \mapsto \prod_{\sigma \in \Sigma} \sigma(\alpha)^{a_\sigma}.$$

We prefer this approach since it makes explicit the dependence on ε.

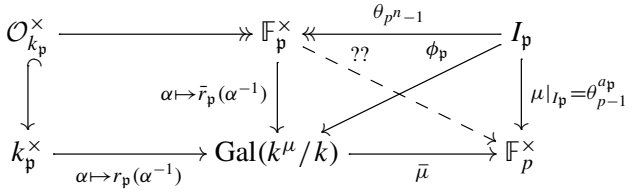
Proposition 3.1 describes μ on the local inertia groups at the different primes $\mathfrak{p}$ above p in terms of fundamental characters. However since μ is unramified outside of p these are the only inertia subgroups on which μ acts nontrivially.

Now let $k \subset k^\mu$ be the smallest abelian extension that trivializes μ. Then μ induces an injective morphism $\bar{\mu} : \mathrm{Gal}(k^\mu/k) \to \mathbb{F}_p^\times$. Let $\mathbb{I}_k$ denote the idéles of k. Then class field theory provides the global reciprocity map

$$r : \mathbb{I}_k \to \mathrm{Gal}(k^\mu/k).$$

The units $k^\times$ embed diagonally into $\mathbb{I}_k$ and one has $r(k^\times) = \{\mathrm{Id}_{k^\mu}\}$, and composing the inclusions $k_{\mathfrak{p}}^\times \hookrightarrow \mathbb{I}_k$ with r gives the local reciprocity map $r_{\mathfrak{p}}$ at $\mathfrak{p}$. Since the order of $\mathrm{Gal}(k^\mu/k)$ is coprime to p the local reciprocity map vanishes on $1 + \mathfrak{p}\mathcal{O}_{k_{\mathfrak{p}}}$ and hence we also get a map $\bar{r}_{\mathfrak{p}} : \mathbb{F}_{\mathfrak{p}}^\times \to \mathrm{Gal}(k^\mu/k)$. Let $\mathbb{F}_{\mathfrak{p}}$ denote the residue field at $\mathfrak{p}$ and write $p^n = \#\mathbb{F}_{\mathfrak{p}}$. Then the fundamental character of level n can be seen as a map $\theta_{p^n-1} : I_{\mathfrak{p}} \to \mathbb{F}_{\mathfrak{p}}^\times$. Finally by choosing an embedding $k^\mu \to \bar{k}_{\mathfrak{p}}$ we get a map

$\phi_{\mathfrak{p}} : I_{\mathfrak{p}} \subseteq \mathrm{Gal}(\overline{k_{\mathfrak{p}}}/k_{\mathfrak{p}}) \to \mathrm{Gal}(k^{\mu}/k)$. This can be summarised in the following commutative diagram:



where the equality $\phi_{\mathfrak{p}}(s) = \bar{r}_{\mathfrak{p}} \circ \theta_{p^n-1}(s^{-1})$ is Proposition 3 of [Serre 1972]. Since $\theta_{p-1} = \mathrm{Nm}_{\mathbb{F}_{\mathfrak{p}}/\mathbb{F}_p} \circ \theta_{p^n-1}$ we know that the map that is needed at the place of the question marks to make everything commute is $x \mapsto \mathrm{Nm}_{\mathbb{F}_{\mathfrak{p}}/\mathbb{F}_p}(x)^{a_{\mathfrak{p}}}$.

In particular this commutative diagram allows one to describe $\bar{\mu} \circ r_{\mathfrak{p}} : \mathcal{O}_{k_{\mathfrak{p}}}^{\times} \to \mathbb{F}_p^{\times}$ more directly by

$$\bar{\mu} \circ r_{\mathfrak{p}}(\alpha) = \mathrm{Nm}_{\mathbb{F}_{\mathfrak{p}}/\mathbb{F}_p}(\alpha \pmod{\mathfrak{p}})^{-a_{\mathfrak{p}}}.$$

Having this concise description of $\bar{\mu} \circ r_{\mathfrak{p}}$ on $\mathcal{O}_{k_{\mathfrak{p}}}^{\times}$ for all primes $\mathfrak{p} \mid p$ using local class field theory, as well as the fact that μ is unramified outside p, allows one to describe μ more generally as in the following proposition, in which we have used the ideal theoretic description of class field theory to see μ as a character on the fractional ideals coprime to p; see Remark 2.1.

Proposition 3.4 (generalisation of [David 2011b, Proposition 2.6]). *Let k be a number field, K its Galois closure over $\mathbb{Q}$ and μ the twelfth power of a p-isogeny character over k. Then for every prime ideal $\mathfrak{p}_0$ lying above p in K there exists a k-isogeny signature $\varepsilon = \varepsilon_{\mathfrak{p}_0} = \sum_{\sigma} a_{\sigma}\sigma$ such that for all $\alpha \in k^{\times}$ prime to p,*

$$\mu((\alpha)) \equiv \alpha^{\varepsilon} \pmod{\mathfrak{p}_0}.$$

Furthermore if $p > 13$ and p is unramified in k, then for every $\mathfrak{p}_0$ there is an unique such signature $\varepsilon_{\mathfrak{p}_0}$.

Proof. View $\bar{\mu} \circ r$ as a character on $\mathbb{I}_k$ that vanishes on $k^{\times}$. Since μ is the twelfth power of another character, $\bar{\mu} \circ r_v$ is trivial for all infinite places v of k. We can use this to compute

$$1 = \prod_{\mathfrak{q} \nmid p} \bar{\mu} \circ r_{\mathfrak{q}}(\alpha_{\mathfrak{q}}) \times \prod_{\mathfrak{p} \mid p} \bar{\mu} \circ r_{\mathfrak{p}}(\alpha_{\mathfrak{p}}) = \prod_{\mathfrak{q} \nmid p} \mu(\mathrm{Frob}_{\mathfrak{q}})^{v_{\mathfrak{q}}(\alpha)} \times \left(\prod_{\mathfrak{p} \mid p} \mathrm{Nm}_{\mathbb{F}_{\mathfrak{p}}/\mathbb{F}_p}(\alpha \bmod \mathfrak{p})^{a_{\mathfrak{p}}} \right)^{-1}.$$

If p were unramified then we could write

$$\mathrm{Nm}_{\mathbb{F}_{\mathfrak{p}}/\mathbb{F}_p}(\alpha \pmod{\mathfrak{p}}) \equiv \prod_{\sigma \in \mathrm{Hom}(k_{\mathfrak{p}}, \overline{\mathbb{Q}_{\mathfrak{p}}})} \sigma(\alpha) \pmod{\mathfrak{p}'},$$

where $\mathfrak{p}'$ is the prime of $\overline{\mathbb{Q}_{\mathfrak{p}}}$ lying over p, and the proposition would follow by setting $a_{\sigma} := a_{\sigma^{-1}(\mathfrak{p}_0)}$ and rewriting as follows:

$$\begin{aligned} \prod_{\mathfrak{p} \mid p} \mathrm{Nm}_{\mathbb{F}_{\mathfrak{p}}/\mathbb{F}_p}(\alpha \pmod{\mathfrak{p}})^{a_{\mathfrak{p}}} &\equiv \prod_{\mathfrak{p} \mid p} \prod_{\sigma \in \mathrm{Hom}(k_{\mathfrak{p}}, \overline{\mathbb{Q}_{\mathfrak{p}}})} \sigma(\alpha)^{a_{\mathfrak{p}}} \pmod{\mathfrak{p}'} \\ &\equiv \prod_{\sigma \in \mathrm{Hom}(k, K)} \sigma(\alpha)^{\alpha_{\sigma}} \pmod{\mathfrak{p}_0} = \alpha^{\varepsilon} \pmod{\mathfrak{p}_0}, \end{aligned}$$

where in the last step we choose an embedding $K \subseteq K_{\mathfrak{p}_0} \hookrightarrow \overline{\mathbb{Q}_p}$.

Suppose now that a prime $\mathfrak{p} \mid p$ ramifies in k. Let $e_\mathfrak{p}$ denote the ramification index at $\mathfrak{p}$. Then we can write

$$\mathrm{Hom}(k_\mathfrak{p}, \overline{\mathbb{Q}_p}) = \bigcup_{i=1}^{e_\mathfrak{p}} S_{\mathfrak{p},i}$$

as the union of $e_\mathfrak{p}$ different sets of size $[\mathbb{F}_\mathfrak{p} : \mathbb{F}_p]$, so that for each $S_{\mathfrak{p},i}$ we have

$$\mathrm{Nm}_{\mathbb{F}_\mathfrak{p}/\mathbb{F}_p}(\alpha \pmod{\mathfrak{p}}) \equiv \prod_{\sigma \in S_{\mathfrak{p},i}} \sigma(\alpha) \pmod{\mathfrak{p}'}.$$

Furthermore the conditions on $a_\mathfrak{p}$ from Proposition 3.1 allow us to write $a_\mathfrak{p} = \sum_{i=1}^{e_\mathfrak{p}} a_{\mathfrak{p},i}$ with each $0 \leq a_{\mathfrak{p},i} \leq 12$ and $a_{\mathfrak{p},i} \equiv 0, 4, 6$ or $8 \bmod 12$, whence

$$\mathrm{Nm}_{\mathbb{F}_\mathfrak{p}/\mathbb{F}_p}(\alpha \pmod{\mathfrak{p}})^{a_\mathfrak{p}} \equiv \prod_{i=1}^{e_\mathfrak{p}} \prod_{\sigma \in S_i} \sigma(\alpha)^{a_{\mathfrak{p},i}} \pmod{\mathfrak{p}'}.$$

Writing

$$\mathrm{Hom}(k, K) = \mathrm{Hom}(k, \overline{\mathbb{Q}_p}) = \bigcup_{\mathfrak{p} \mid p} \bigcup_{i=1}^{e_\mathfrak{p}} S_{\mathfrak{p},i}$$

and setting $a_\sigma = a_{\mathfrak{p},i}$ for $\sigma \in S_{\mathfrak{p},i}$ we can rewrite

$$\begin{aligned}\prod_{\mathfrak{p} \mid p} \mathrm{Nm}_{\mathbb{F}_\mathfrak{p}/\mathbb{F}_p}(\alpha \pmod{\mathfrak{p}})^{a_\mathfrak{p}} &\equiv \prod_{\mathfrak{p} \mid p} \prod_{i=1}^{e_\mathfrak{p}} \prod_{\sigma \in S_{\mathfrak{p},i}} \sigma(\alpha)^{a_{\mathfrak{p},i}} \pmod{\mathfrak{p}'} \\ &\equiv \prod_{\sigma \in \mathrm{Hom}(k,K)} \sigma(\alpha)^{\alpha_\sigma} \pmod{\mathfrak{p}_0} = \alpha^\varepsilon \pmod{\mathfrak{p}_0}\end{aligned}$$

as in the unramified case and we are done. □

We refer to $\varepsilon_{\mathfrak{p}_0}$ as *the isogeny signature of* λ *with respect to* $\mathfrak{p}_0$. Note that since $\mathrm{Gal}(K/\mathbb{Q})$ acts transitively on the primes above p, a different choice of prime $\mathfrak{p}_0$ merely permutes the integers a_σ. We will therefore often drop the subscript $\mathfrak{p}_0$ and speak of ε as *the isogeny signature of* λ. Note that fixing an ordering to the embeddings in $\Sigma = \mathrm{Hom}(k, K)$ allows one to think of the signature as a d-tuple valued in the set $\{0, 4, 6, 8, 12\}$, and hence one sees that there are precisely 5^d possible isogeny signatures for a degree d number field. If all the integers a_σ are the same integer a, then clearly the ordering on Σ does not matter, and in the sequel we denote this signature as the d-tuple $(a, \ldots, a)$.

From the construction of $\varepsilon_{\mathfrak{p}_0}$ in the above proof, we obtain the following condition which must be satisfied by the integers a_σ.

Corollary 3.5. *Let* $\varepsilon_{\mathfrak{p}_0} = \sum_{\sigma \in \Sigma} a_\sigma \sigma$ *be an isogeny signature with respect to* $\mathfrak{p}_0$. *Then, for* $\sigma, \tau \in \Sigma = \mathrm{Hom}(k, K)$,

$$\sigma^{-1}(\mathfrak{p}_0) = \tau^{-1}(\mathfrak{p}_0) \quad \Longrightarrow \quad a_\sigma \equiv a_\tau \pmod{p-1}.$$

In particular, if p *is inert in* k *and* $p \geq 17$, *then all the integers* a_σ *are the same.*

Isogeny signatures satisfying this condition will be referred to as *admissible*. Note furthermore that if one of the integers in the signature ε is 4 or 8 (respectively 6), then from the proof of Proposition 3.1

we have that $\mathrm{Aut}(\widetilde{E}_{\mathfrak{p}}) \cong \mu_6$ and $p \equiv 2 \pmod 3$ (respectively $\mathrm{Aut}(\widetilde{E}_{\mathfrak{p}}) \cong \mu_4$ and $p \equiv 3 \pmod 4$). We therefore refer to such signatures as *sextic* (respectively *quartic*). The overlap of these two designations (i.e., the signature contains both 4 and 6 or both 8 and 6) is an interesting source of isogenies, and will be referred to as *mixed*. Signatures, all of whose defining integers are the same, will be referred to as *constant*.

Example 3.6. In the case that $X_0(p)$ is elliptic or hyperelliptic, one may generate quadratic points on $X_0(p)$ by taking the pullback of rational points along the hyperelliptic map to $\mathbb{P}^1$. The corresponding isogeny ϕ may be constructed in Magma. After passing to a suitable ramified extension of $k_{\mathfrak{p}}$ over which E obtains semistable reduction, we let $\widehat{\phi}$ be the power series representation of ϕ on the formal groups of the elliptic curves. Then as in [Serre 1972, Section 1.10], we may directly compute the $a_{\mathfrak{p}}$ integers as the $\mathfrak{p}$-adic valuation of the coefficient of X in $\widehat{\phi}(X)$. In this way one may actually compute the signature of isogenies, as the following examples show; in both cases σ denotes the nontrivial automorphism of the corresponding quadratic field:

(1) Let $k = \mathbb{Q}(\sqrt{-1120581})$, and let E/k be an elliptic curve with j-invariant

$$\tfrac{1}{837568512}(-344992121\sqrt{-1120581} - 182301639894).$$

Then E admits a k-rational 11-isogeny of signature $6\,\mathrm{Id} + 8\sigma$, and in particular shows that isogenies of mixed signature exist.

(2) Let $k = \mathbb{Q}(\sqrt{38731793})$, and let E/k be an elliptic curve with j-invariant

$$432949901898808770597450 0\sqrt{38731793} + 269445817519329500833893356 25.$$

Then E admits a k-rational 23-isogeny of signature $(6, 6)$.

The Magma code in magma_scripts/EpsilonTypes.m was used to find and verify these examples (and others); this script has also been used to generate test cases for our software package.

Remark 3.7. Note that Larson and Vaintrob have similarly removed the Galois assumption in Momose's Lemma 1, using the notion of *algebraic characters*; see Definition 2.2 and Corollary 2.4 of [Larson and Vaintrob 2014]. The left-hand side of the equation in their Corollary 2.4 is the space of algebraic characters, whilst the right-hand side may be identified with the space of isogeny signatures.

3.2. ***A general divisibility criterion.*** Let $\mathfrak{q}$ be a prime ideal of k which is coprime to p, and consider the reduction of E at $\mathfrak{q}$, which is either potentially multiplicative, potentially good supersingular, or potentially good ordinary. In both of the potentially good cases, the characteristic polynomial of Frobenius $\mathrm{Frob}_{\mathfrak{q}}$ acting on the p-adic Tate module of E has coefficients in $\mathbb{Z}$ and is independent of p [Serre and Tate 1968, Theorem 3]; we may thus write $P_{\mathfrak{q}}(X)$ for this polynomial. This is a quadratic polynomial whose roots have absolute value $\sqrt{\mathrm{Nm}(\mathfrak{q})}$. We write $L^{\mathfrak{q}}$ for the splitting field of this polynomial, which is either $\mathbb{Q}$ or an imaginary quadratic field.

In all cases of reduction at $\mathfrak{q}$, one obtains congruence conditions modulo p on $\lambda(\mathrm{Frob}_\mathfrak{q})$. If E has potentially multiplicative reduction at $\mathfrak{q}$, then $\lambda(\mathrm{Frob}_\mathfrak{q})$ is either 1 or $\pm\mathrm{Nm}(\mathfrak{q}) \pmod p$ [David 2011a, Proposition 3.3; 2011b, Proposition 1.4]. In the potentially good reduction case, for $\mathcal{P}^\mathfrak{q}$ a prime of $L^\mathfrak{q}$ above p, the images of the roots of $P_\mathfrak{q}(X)$ in $\mathcal{O}_{L^\mathfrak{q}}/\mathcal{P}^\mathfrak{q}$ are in $\mathbb{F}_p^\times$, and there is a root $\beta_\mathfrak{q}$ of $P_\mathfrak{q}(X)$ such that $\lambda(\mathrm{Frob}_\mathfrak{q}) = \beta_\mathfrak{q} \pmod{\mathcal{P}^\mathfrak{q}}$ [David 2011a, Proposition 3.6; 2011b, Proposition 1.8]. For simplicity we will sometimes drop the $\mathfrak{q}$ subscript on $\beta_\mathfrak{q}$, particularly when "looping" over several such roots.

Write $h_\mathfrak{q}$ for the order of $\mathfrak{q}$ in the class group Cl_k of k, and let $\gamma_\mathfrak{q}$ be a generator of the principal ideal $\mathfrak{q}^{h_\mathfrak{q}}$. We apply Proposition 3.4 to $\gamma_\mathfrak{q}$ and obtain, with the ideal-theoretic interpretation for the domain of μ as Remark 2.1, the following expression:

$$\mu((\gamma_\mathfrak{q})) \equiv \gamma_\mathfrak{q}^\varepsilon \pmod{\mathfrak{p}_0}.$$

Replacing this with the Galois character interpretation for μ, and using $\mathfrak{q}^{h_\mathfrak{q}} = (\gamma_\mathfrak{q})$, we obtain

$$\mu^{h_\mathfrak{q}}(\mathrm{Frob}_\mathfrak{q}) \equiv \gamma_\mathfrak{q}^\varepsilon \pmod{\mathfrak{p}_0}.$$

This expression, when combined with the aforementioned congruence conditions modulo p on $\mu(\mathrm{Frob}_\mathfrak{q})$, yield divisibility conditions for p, namely that p must divide one of the integers defined as follows.

Definition 3.8. Let k be a number field with Galois closure K over $\mathbb{Q}$, $\mathfrak{q}$ a prime ideal of k of order $h_\mathfrak{q}$ in the class group of k, and $\gamma_\mathfrak{q}$ a generator of the principal ideal $\mathfrak{q}^{h_\mathfrak{q}}$. For a set S of integers, let $\mathrm{lcm}(S)$ denote the least common multiple of the integers in S, with the convention that this will be 0 if $0 \in S$. Then we define the integers

$$\begin{aligned}
A(\varepsilon,\mathfrak{q}) &:= \mathrm{Nm}_{K/\mathbb{Q}}(\gamma_\mathfrak{q}^\varepsilon - 1),\\
B(\varepsilon,\mathfrak{q}) &:= \mathrm{Nm}_{K/\mathbb{Q}}(\gamma_\mathfrak{q}^\varepsilon - \mathrm{Nm}(\mathfrak{q})^{12h_\mathfrak{q}}),\\
C_s(\varepsilon,\mathfrak{q}) &:= \mathrm{lcm}(\{\mathrm{Nm}_{K(\beta)/\mathbb{Q}}(\gamma_\mathfrak{q}^\varepsilon - \beta^{12h_\mathfrak{q}}) \mid \beta \text{ is a supersingular Frobenius root over } \mathbb{F}_\mathfrak{q}\}),\\
C_o(\varepsilon,\mathfrak{q}) &:= \mathrm{lcm}(\{\mathrm{Nm}_{K(\beta)/\mathbb{Q}}(\gamma_\mathfrak{q}^\varepsilon - \beta^{12h_\mathfrak{q}}) \mid \beta \text{ is an ordinary Frobenius root over } \mathbb{F}_\mathfrak{q}\}),\\
C(\varepsilon,\mathfrak{q}) &:= \mathrm{lcm}(C_o(\varepsilon,\mathfrak{q}), C_s(\varepsilon,\mathfrak{q})),
\end{aligned}$$

where in $C_s(\varepsilon,\mathfrak{q})$ (respectively $C_o(\varepsilon,\mathfrak{q})$) the lcm is taken over all roots β of characteristic polynomials of Frobenius of supersingular (respectively, ordinary) elliptic curves defined over the residue field $\mathbb{F}_\mathfrak{q}$ of k at $\mathfrak{q}$.

Remark 3.9. Using Waterhouse's Theorem 2.5, specifically Table 4 derived from it, one may show that

$$C_s(\varepsilon,\mathfrak{q}) := \begin{cases} B(2\varepsilon,\mathfrak{q}) & \text{if } |\mathbb{F}_\mathfrak{q}| = 2^f \text{ with } f \text{ odd},\\ \mathrm{Nm}_{k/\mathbb{Q}}(\gamma_\mathfrak{q}^\varepsilon - \mathrm{Nm}(\mathfrak{q})^{6h_\mathfrak{q}}) & \text{otherwise.}\end{cases}$$

We frame the above discussion as follows.

Corollary 3.10. (1) *If E has potentially multiplicative reduction at $\mathfrak{q}$, then p divides either $A(\varepsilon,\mathfrak{q})$ or $B(\varepsilon,\mathfrak{q})$.*

(2) *If E has potentially good ordinary reduction at $\mathfrak{q}$, then p divides $C_o(\varepsilon,\mathfrak{q})$.*

(3) *If E has potentially good supersingular reduction at $\mathfrak{q}$, then p divides $C_s(\varepsilon,\mathfrak{q})$.*

condition	$\gamma_{\mathfrak{q}}^{\varepsilon}$	$\mathbb{Q}(\beta)$	$\gamma_{\mathfrak{q}}^{\varepsilon} \in \mathbb{Q}$?	$(a_\tau)_{\tau\in\Sigma}$	everywhere unramified character	signature type
$A(\varepsilon, \mathfrak{q}) = 0$	1		yes	all 0	μ	type 1
$B(\varepsilon, \mathfrak{q}) = 0$	$q^{12h_{\mathfrak{q}}}$			all 12	μ/χ_p^{12}	
$C_s(\varepsilon, \mathfrak{q}) = 0$	$\pm q^{6h_{\mathfrak{q}}}$	$\mathbb{Q}(\sqrt{-q})$		all 6	μ/χ_p^{6}	type 2
$C_o(\varepsilon, \mathfrak{q}) = 0$	$\beta^{12h_{\mathfrak{q}}}$ for β an ordinary Frobenius root over $\mathbb{F}_{\mathfrak{q}}$	$\mathbb{Q}(\beta) \subseteq k$, p splits or ramifies in $\mathbb{Q}(\beta)$, $\mathrm{Nm}_{k/\mathbb{Q}(\beta)}(\mathfrak{q}) = (\beta)$ or $(\bar{\beta})$	no	$a_\tau = 12$ for $\tau \in \Sigma_{\mathbb{Q}(\beta)}$; 0 otherwise		type 3
				$a_\tau = 0$ for $\tau \in \Sigma_{\mathbb{Q}(\beta)}$; 12 otherwise		

Table 5. Summary of what happens if one of the integers A, B, C_s or C_o is zero.

3.3. ***Removing the Galois assumption in Momose's Lemma 2.*** Having identified integers which p must divide, the question of the nonzeroness of these integers becomes relevant. This is the motivation for Momose's Lemma 2, or [David 2011b, Proposition 2.15], both of which assume that k is Galois over $\mathbb{Q}$.

We thus provide the following non-Galois version of Momose's Lemma 2. The proof is modelled on David's proof, and many of the arguments carry over mutatis mutandis, though the details in type 3 require some additional ideas. For L a subfield of k, we denote by $\Sigma_L \subset \Sigma$ the subset of embeddings of k in K which act as the identity on L.

Proposition 3.11. *Let $p \geq 17$, let $q \neq p$ be a rational prime which splits completely in k, and let $\mathfrak{q}$ be a prime of k over q. If the condition shown in the left-most column of Table 5 is satisfied, then the corresponding assertions in the rest of the table hold.*

Proof. First observe, from the assumption that q splits completely in k, we have that q splits completely in K; see for example Chapter 1, Section 9, Exercise 4 in [Neukirch 1999]. This will be used throughout the proof.

Before considering the various cases, we set the following notation to be used throughout the proof. Let $\mathfrak{q}_1, \ldots, \mathfrak{q}_d$ be the distinct prime ideals of k lying over q, and we suppose $\mathfrak{q} = \mathfrak{q}_1$. We label the embeddings in Σ as $\sigma_1, \ldots, \sigma_d$, where σ_1 is the inclusion $k \subseteq K$. We let $M = \mathrm{Gal}(K/\sigma_1(k)) = \mathrm{Gal}(K/k)$, viewed as a subgroup of $G := \mathrm{Gal}(K/\mathbb{Q})$; let n denote the size of M. For each $1 \leq i \leq d$, we let $Q_1^{(i)}, \ldots, Q_n^{(i)}$ be the distinct prime ideals of K dividing $\sigma_i(\mathfrak{q})\mathcal{O}_K$ (recalling that if q splits completely in k then q splits completely in K). Because q is unramified, it follows that $\sigma_i(\mathfrak{q})\mathcal{O}_K = \prod_{j=1}^{n} Q_j^{(i)}$. Write $\varepsilon = \sum_{i=1}^{d} a_i\sigma_i$. We define the set

$$S_q := \{Q_j^{(i)} : 1 \leq i \leq d, 1 \leq j \leq n\},$$

which is a priori only a subset of the prime ideals of K lying over q. However, since S_q is stable under G, and G acts transitively on the set of primes in K lying over q, we have that S_q actually equals the set of

all primes lying over q. Finally we let $\tau_1, \ldots, \tau_d$ denote the (left) coset representatives of G/M, chosen such that $\tau_i \circ \sigma_1 = \sigma_i$. Fixing $\hat{Q} = Q_1^{(1)}$, this implies that

$$\sigma_i(\mathfrak{q})\mathcal{O}_K = \tau_i(\sigma_1(\mathfrak{q}))\mathcal{O}_K = \prod_{Q \in \tau_i M \hat{Q}} Q.$$

Since G acts faithfully on the primes lying over q, this means that $\sigma_i(\mathfrak{q})$ and $\sigma_j(\mathfrak{q})$ are coprime if $i \neq j$.

We now come to the three cases for the value of $\gamma_{\mathfrak{q}}^{\varepsilon}$.

Consider first that $\gamma_{\mathfrak{q}}^{\varepsilon} = 1$. By considering the ideal of $\mathcal{O}_K$ generated by $\gamma_{\mathfrak{q}}^{\varepsilon}$, we obtain

$$\mathcal{O}_K = \prod_{i=1}^{d} \sigma_i(\mathfrak{q})^{a_i h_{\mathfrak{q}}}.$$

Since the $\sigma_i(\mathfrak{q})$ are pairwise coprime we obtain that $a_i = 0$ for all i. Since $\mu|_{I_{\mathfrak{p}}} = \chi_p^{a_{\mathfrak{p}}}$ and μ is unramified away from p, we obtain that μ is unramified everywhere.

In the second case $\gamma_{\mathfrak{q}}^{\varepsilon} = q^{12h_{\mathfrak{q}}}$, one similarly obtains that $a_i = 12$ for all i, whence μ/χ_p^{12} is everywhere unramified.

In the third case, we have that $\gamma_{\mathfrak{q}}^{\varepsilon} = \beta^{12h_{\mathfrak{q}}}$, where β is a root of the characteristic polynomial of Frobenius of an elliptic curve over $\mathbb{F}_{\mathfrak{q}}$. We write $L = \mathbb{Q}(\beta)$, which is either $\mathbb{Q}$ or an imaginary quadratic field. By assumption, the element $\gamma_{\mathfrak{q}}^{\varepsilon}$ — a priori in K — is also in L; so either $\gamma_{\mathfrak{q}}^{\varepsilon}$ is rational, or it generates L and therefore L is contained in K.

In the first of these two subcases, $\beta^{12h_{\mathfrak{q}}}$ is rational; therefore it is equal to its complex conjugate $\bar{\beta}^{12h_{\mathfrak{q}}}$; in particular there is a $12h_{\mathfrak{q}}$-th root of unity $\zeta \in L$ such that $\beta = \zeta\bar{\beta}$. However, since L is imaginary quadratic, it only admits n-th roots of unity for $n = 2, 4$ or 6; thus β^{12} is rational. Moreover, since β is an algebraic integer, β^{12} is an integer. Since the absolute value of β is $\sqrt{\mathrm{Nm}(\mathfrak{q})} = \sqrt{q}$, we get that $\beta^{12} = \pm q^6$, and therefore

$$\prod_{\sigma \in \Sigma} \sigma(\mathfrak{q}^{h_{\mathfrak{q}}})^{a_\sigma}\mathcal{O}_K = \gamma_{\mathfrak{q}}^{\varepsilon}\mathcal{O}_K = \beta^{12h_{\mathfrak{q}}}\mathcal{O}_K = q^{6h_{\mathfrak{q}}}\mathcal{O}_K = \left(\prod_{\sigma \in \Sigma} \sigma(\mathfrak{q})\mathcal{O}_K\right)^{6h_{\mathfrak{q}}}.$$

Since $\sigma_i(\mathfrak{q})\mathcal{O}_K$ and $\sigma_j(\mathfrak{q})\mathcal{O}_K$ are coprime for $i \neq j$ we obtain that all a_σ are equal to 6. That μ/χ_p^6 is everywhere unramified follows as previously. To show that $\mathbb{Q}(\beta) = \mathbb{Q}(\sqrt{-q})$ one observes from Corollaries 2.6 and 2.7 that β is a supersingular Frobenius root, so one simply checks through the possibilities in Table 4 (observing that we have $f = 1$ because we are assuming $\mathfrak{q}$ is completely split).

In the second subcase, L is an imaginary quadratic field contained in K. We let $H := \mathrm{Gal}(K/L)$. As in the first subcase, we consider the K-ideal

$$(\gamma_{\mathfrak{q}}^{\varepsilon}) = \prod_{i=1}^{d}\left(\prod_{j=1}^{n} Q_j^{(i)}\right)^{a_i h_{\mathfrak{q}}};$$

the crucial observation is that, since this ideal is invariant under H (because $\gamma_{\mathfrak{q}}^{\varepsilon} \in L$), it must factor in the form

$$(\gamma_{\mathfrak{q}}^{\varepsilon}) = \left(\prod_{Q \in H\hat{Q}} Q\right)^{a_{\mathrm{Id}}}\left(\prod_{Q \in H\gamma\hat{Q}} Q\right)^{a_\gamma} \tag{3-1}$$

for $\gamma \in G$ an element which induces complex conjugation on L, and a_{Id} and a_γ are integers; in particular, there are only two choices (viz. a_{Id} or a_γ) for each a_i.

Suppose for a contradiction that $M \not\subseteq H$. Choose $\delta \in M \setminus H$, and consider the two primes $\hat{Q}$ and $\delta\hat{Q}$ in S_q. These are both ideals dividing $(\gamma_{\mathfrak{q}}^{\varepsilon})$ with the same exponent (viz. $a_1 h_{\mathfrak{q}}$), but are in different H-orbits under the H-action on S_q. This forces $a_{\mathrm{Id}} = a_\gamma$, whence $\gamma_{\mathfrak{q}}^{\varepsilon} = \mathrm{Nm}_{K/\mathbb{Q}}(\sigma_1(\gamma_{\mathfrak{q}}))^{a_\gamma}$; i.e., $\gamma_{\mathfrak{q}}^{\varepsilon} \in \mathbb{Q}$, contradicting the subcase that we are currently in. Therefore $M \subseteq H$, which establishes that L is contained in k.

We now have

$$(\gamma_{\mathfrak{q}}^{\varepsilon})\mathcal{O}_L = (\mathrm{Nm}_{k/L}(\gamma_{\mathfrak{q}})\mathcal{O}_L)^{a_{\mathrm{Id}}} \cdot (\overline{\mathrm{Nm}_{k/L}(\gamma_{\mathfrak{q}})}\mathcal{O}_L)^{a_\gamma} \text{ (from (3-1))}$$
$$\Longrightarrow \quad (\beta\mathcal{O}_L)^{12h_{\mathfrak{q}}} = ((\mathrm{Nm}_{k/L}(\mathfrak{q}))^{a_{\mathrm{Id}}}(\overline{\mathrm{Nm}_{k/L}(\mathfrak{q})})^{a_\gamma})^{h_{\mathfrak{q}}}.$$

Since $q = \beta\bar{\beta}$, we obtain that the L-ideal $\mathrm{Nm}_{k/L}(\mathfrak{q})$ is either $\beta\mathcal{O}_L$ or $\bar{\beta}\mathcal{O}_L$, which yields that the pair of integers $(a_{\mathrm{Id}}, a_\gamma)$ is either $(0, 12)$ or $(12, 0)$, yielding the two possible forms of ε as in the table.

The only remaining assertion to prove is that p splits or ramifies in $L = \mathbb{Q}(\beta)$ in the type 3 case. Suppose for a contradiction that p is inert in L, and let $\mathrm{Frob}_{\mathfrak{p}_0} \in \mathrm{Gal}(K/\mathbb{Q})$ be a choice of a lift of Frobenius at $\mathfrak{p}_0$. The automorphism $\mathrm{Frob}_{\mathfrak{p}_0}$ satisfies the following two properties:

(1) It fixes $\mathfrak{p}_0$.

(2) Its restriction to L is also a lift of Frobenius of the prime ideal $p\mathcal{O}$, of residue class degree 2, and hence its restriction to L is the nontrivial element of $\mathrm{Gal}(L/\mathbb{Q})$.

We now consider the two embeddings of k in K: σ_1 and $\mathrm{Frob}_{\mathfrak{p}_0} \circ \sigma_1$. By property (1) above, the preimage of $\mathfrak{p}_0$ in k under each of these embeddings is the same, and therefore, by Corollary 3.5, $a_1 \equiv a_{\mathrm{Frob}_{\mathfrak{p}_0} \circ \sigma_1} \pmod{p-1}$. On the other hand, property (2) forces one of these integers to be 0, and the other $12 \pmod{p-1}$. We therefore obtain a contradiction for primes p such that $p - 1 \nmid 12$ (and hence for all $p \geq 17$). □

Proposition 3.11 suggests that the signatures identified in Table 5 require particular attention. This motivates the following definition.

Definition 3.12. Let $\varepsilon = \sum_\sigma a_\sigma \sigma$ be an isogeny signature:

- If $\varepsilon = 0\sum_\sigma \sigma$ or $\varepsilon = 12\sum_\sigma \sigma$, then ε is of *type* 1.
- If $\varepsilon = 6\sum_\sigma \sigma$ (or equivalently $\varepsilon = 6\,\mathrm{Nm}(k/\mathbb{Q})$), then ε is of *type* 2.
- If there exists an index 2 subgroup $H \subseteq G := \mathrm{Gal}(K/\mathbb{Q})$ with $\mathrm{Gal}(K/k) \subseteq H$, $L := K^H$ is imaginary quadratic and either
$$\varepsilon = 12\sum_{\sigma \in \Sigma_L} \sigma \quad \text{or} \quad \varepsilon = 12\sum_{\sigma \in \Sigma \setminus \Sigma_L} \sigma,$$
then ε is of *type* 3 *with field* L.
- If ε is not of type 1, 2 or 3 then ε is *generic*.

An equivalent definition of a type 3 isogeny signature is that k contains an imaginary quadratic field L such that $\varepsilon = 12\,\mathrm{Nm}(K/L)$ or $\varepsilon = \rho \circ 12\,\mathrm{Nm}(K/L)$, where $\rho : L \to L$ is complex conjugation. Note that when given an explicit ε, the group H can be retrieved as $H = \{g \in \mathrm{Gal}(K/\mathbb{Q}) \mid g \circ \varepsilon = \varepsilon\}$, the stabiliser of ε. In particular, if ε is also type 3 with field L', then $L = L'$.

At this point it is instructive to recap what we are aiming to do in this section, and how far towards this goal we have come. This will clarify what will be happening in the rest of the section.

Our goal in this section (stated at the outset) is to generalise and make Momose's isogeny classification theorem algorithmic [Momose 1995, Theorem A/Theorem 1], as well as address the gaps and mistakes in his original proof. By generalise, we mean to remove the assumptions of k being Galois over $\mathbb{Q}$, and also address primes p that ramify in k.

Momose's isogeny classification theorem says that, for a given number field k, there is a constant C_k such that, if p is an isogeny prime for k that is larger than C_k, then the associated isogeny character must be one of three types, that in the introduction we called Momose types 1, 2 and 3.

In this section, we have identified (in Definition 3.12) three types for the signature for which certain integers (the A, B and C integers from Definition 3.8) could be zero. Outside of these special types of signature — that is, if the signature is generic — we know that these integers are nonzero; thus, by defining the integer

$$ABC(\varepsilon, \mathfrak{q}) := \mathrm{lcm}(\mathrm{Nm}(\mathfrak{q}), A(\varepsilon, \mathfrak{q}), B(\varepsilon, \mathfrak{q}), C(\varepsilon, \mathfrak{q})),$$

we may deal with the generic isogeny primes as a direct consequence of Corollary 3.10 and Proposition 3.11:

Corollary 3.13. *Let $p \geq 17$ be an isogeny prime whose associated isogeny signature ε is generic. Then for all completely split prime ideals $\mathfrak{q}$, p divides the nonzero integer $ABC(\varepsilon, \mathfrak{q})$.*

Therefore, taking a completely split prime ideal $\mathfrak{q}$ in k, and taking the lcm of $ABC(\varepsilon, \mathfrak{q})$ across all of the generic signatures ε, one would obtain a nonzero integer C' such that, if p is an isogeny prime for k that does not divide C', then the associated isogeny signature would be of type 1, 2 or 3.

If it were true that the signature types 1, 2 and 3 identified in Definition 3.12 coincided with Momose's types 1, 2 and 3, then we would be done with our goal. This is however not the case; while an isogeny character of Momose type n (to be defined in the next subsection) implies that the signature is of type n, the converse is not necessarily true (although it is true for $n = 1$).

What remains to be done, therefore, is to bound the prime degrees of isogenies whose signatures are of type n but that are not themselves of Momose type n. If we can do this for $n = 2$ and 3 (again this is not required for $n = 1$) then we have succeeded in making Momose's isogeny classification theorem algorithmic. We call this strategy *going from signature type n to Momose type n*. Specifically, this means that we will prove the following result.

Theorem 3.14. *Let k be a number field. Then for $n = 2$ or 3, there exist explicitly computable integers $B_n(K)$ such that if $p \nmid B_n(K)$ one has that if λ is a p-isogeny character of signature type n, then the isogeny character λ is of Momose type n.*

We refer to Corollary 3.18 and Proposition 3.24 that respectively establish this result for $n = 2$ and 3.

In the next section we will carry out this strategy, by studying the type 1, 2 and 3 signatures more closely.

3.4. ***From signature types to Momose types.*** We emphasise that the types identified in Definition 3.12 are properties of the *signature* of an isogeny, while the Momose types to be defined in the sequel are properties of the *isogeny character* λ. The three special signature types identified in the previous section do not in all cases correspond exactly to Momose's three types identified for the isogeny character λ. They do, however, for type 1 signatures, which Table 5 shows.

Definition 3.15. We say that an isogeny character λ is of *Momose type* 1 if either λ^{12} or $(\lambda/\chi_p)^{12}$ is everywhere unramified.

Momose type 1 primes will be handled in Section 5.

3.4.1. *Signature type 2.* We now consider isogenies of signature type 2; that is, $\varepsilon = 6\,\mathrm{Nm}(k/\mathbb{Q})$. Observe from Table 5 that this corresponds to μ and χ_p^6 being the same up to an everywhere unramified character. The notion of Momose type 2 goes further to say what this everywhere unramified character should be.

Definition 3.16. We say that an isogeny character λ is of *Momose type* 2 if $\lambda^{12} = \chi_p^6$.

That is, the everywhere unramified character should be trivial. Note in this case that necessarily $p \equiv 3 \pmod 4$, by the discussion appearing immediately after Corollary 3.5.

The following key result relates the notions of signature type 2 and Momose type 2.

Proposition 3.17. *Let E/k be an elliptic curve admitting a k-rational p-isogeny of isogeny character λ of signature ε. If the following two conditions hold*:

(1) *ε is of type* 2.

(2) *There exists a set of primes* Gen *generating* Cl_k, *such that for all* $\mathfrak{q} \in \mathrm{Gen}$

  (a) $\mathfrak{q}$ *is coprime to* p;

  (b) $\mathfrak{q}$ *does not lie over the rational prime* 2;

  (c) *E has potentially good supersingular reduction at* $\mathfrak{q}$.

Then λ is of Momose type 2.

Proof. From (1) we have that $\mu\chi_p^{-6}$ is an everywhere unramified character, and hence defines an abelian extension of k contained in the Hilbert class field of k, and thus is determined by its values at Frobenius automorphisms $\mathrm{Frob}_\mathfrak{q}$ for $\mathfrak{q}$ running through a set of generators of Cl_k. Let $\mathfrak{q} \in \mathrm{Gen}$. From (2c) and from the discussion at the beginning of Section 3.2, we obtain

$$\lambda(\mathrm{Frob}_\mathfrak{q}) \equiv \beta_\mathfrak{q} \pmod{\mathcal{P}^\mathfrak{q}}$$

for $\beta_\mathfrak{q}$ a supersingular Frobenius root over $\mathfrak{q}$ and for $\mathcal{P}^\mathfrak{q}$ a prime ideal above p inside a field that is either $\mathbb{Q}$ or an imaginary quadratic field (this is the field denoted as $L^\mathfrak{q}$ in Section 3.2). By the same reasoning as

in Remark 3.9 we have that, for such $\beta_\mathfrak{q}$, that $\beta_\mathfrak{q}^{12} = \mathrm{Nm}(\mathfrak{q})$ (this step requires $\mathfrak{q}$ to have odd characteristic), and hence

$$\mu(\mathrm{Frob}_\mathfrak{q}) = \mathrm{Nm}(\mathfrak{q})^6 \pmod{p} = \chi_p(\mathrm{Frob}_\mathfrak{q})^6.$$

Since this is true for all $\mathfrak{q} \in \mathrm{Gen}$, we obtain $\mu = \chi_p^6$ (i.e., not just equal up to an everywhere unramified character). □

Corollary 3.18. *Let E/k be an elliptic curve admitting a k-rational p-isogeny of isogeny character λ of signature ε. Suppose that ε is of type 2, but λ is not of Momose type 2. Then, for every set of prime ideal generators* Gen *of* Cl_k *of odd characteristic, p divides the nonzero integer*

$$ABC_o(\mathrm{Gen}) := \mathrm{lcm}_{\mathfrak{q}\in\mathrm{Gen}}(A(\varepsilon,\mathfrak{q}), B(\varepsilon,\mathfrak{q}), C_o(\varepsilon,\mathfrak{q}), \mathrm{Nm}(\mathfrak{q})).$$

Proof. Since λ is assumed not of Momose type 2, then by Proposition 3.17, there must exist a prime $\mathfrak{q}$ in Gen such that E does *not* have potentially good supersingular reduction at $\mathfrak{q}$; therefore p must divide the integer $ABC_o(\mathrm{Gen})$ by Corollary 3.10. The proof therefore consists in showing that this integer is nonzero.

Let $\mathfrak{q} \in \mathrm{Gen}$. Since ε is of type 2, we obtain that

$$\gamma_\mathfrak{q}^\varepsilon = \mathrm{Nm}(\mathfrak{q})^{6h_\mathfrak{q}}.$$

This is clearly not equal to 1, hence $A(\varepsilon,\mathfrak{q}) \neq 0$. It is also clearly not equal to $\mathrm{Nm}(\mathfrak{q})^{12h_\mathfrak{q}}$, whence $B(\varepsilon,\mathfrak{q}) \neq 0$. Finally suppose that $C_o(\varepsilon,\mathfrak{q}) = 0$. We would then obtain

$$\mathrm{Nm}(\mathfrak{q})^{6h_\mathfrak{q}} = \beta_\mathfrak{q}^{12h_\mathfrak{q}}$$

for $\beta_\mathfrak{q}$ an ordinary Frobenius root over $\mathbb{F}_\mathfrak{q}$, which contradicts Corollary 2.6. □

One then deals separately with isogenies of Momose type 2; this will be done in Section 6. Note that when $h_k = 1$ then we can take the empty set of generators in Corollary 3.18, and hence we obtain that an isogeny of signature type 2 is automatically of Momose type 2. The above Corollary 3.18 is the culmination of the strategy of going from signature type 2 to Momose type 2.

3.4.2. *Signature type 3.* Finally we consider isogenies of signature type 3 with field L. Note that Proposition 3.4 only implies that the signature corresponding to an isogeny is unique for unramified primes greater than 13. Thus, for ramified primes we have some freedom in the choice of signature that we can associate to an isogeny character in order to study it more precisely.

Proposition 3.19. *Let $L \subseteq k$ be an imaginary quadratic field in which p ramifies and let $\mathfrak{p}$ be the unique prime of L above p. Then a p-isogeny character λ defined over k has a signature of type 2 if and only if it has a signature of type 3.*

Proof. It suffices to prove that for all $\alpha \in k^\times$ coprime to p we have

$$\mathrm{Nm}_{k/L}(\alpha)^{12} \pmod{\mathfrak{p}} \equiv \mathrm{Nm}_{k/\mathbb{Q}}(\alpha)^6 \pmod{p} \equiv \overline{\mathrm{Nm}_{k/L}(\alpha)^{12}} \pmod{\mathfrak{p}}.$$

Since p ramifies in the quadratic field L we have $x \equiv \bar{x} \pmod{\mathfrak{p}}$ for all $x \in L$, and the proposition follows from the equality $\mathrm{Nm}_{k/\mathbb{Q}} = \mathrm{Nm}_{k/L} \cdot \overline{\mathrm{Nm}_{k/L}}$. □

The above proposition shows that when p ramifies in L, there is no difference between a character having a signature of type 2 and a signature of type 3. Since we already have a strategy for studying isogeny characters with a type 2 signature, we are reduced to studying isogenies with a type 3 signature where p is unramified in L, which we assume for the rest of this subsection.

We now define the notion of Momose type 3.

Definition 3.20. We say that a k-rational p-isogeny character λ is of *Momose type* 3 if k contains an imaginary quadratic field L as well as its Hilbert class field, p splits in L as $\mathfrak{p} \cdot \bar{\mathfrak{p}}$, and for any prime $\mathfrak{q}$ of k coprime to $\mathfrak{p}$,

$$\lambda^{12}(\mathrm{Frob}_{\mathfrak{q}}) = \alpha^{12} \pmod{\mathfrak{p}} \tag{3-2}$$

for any $\alpha \in L^{\times}$ a generator of $\mathrm{Nm}_{k/L}(\mathfrak{q})$.

Since this definition is rather more involved than Momose type 2, we describe how one may go from signature type 3 to Momose type 3 in three steps; these will be Corollary 3.22, and Propositions 3.23 and 3.24 below. We remark here that p splitting in L is automatic from ε being of signature type 3, since the proof given at the end of the proof of Proposition 3.11 is independent of $\mathfrak{q}$.

We take $\mathfrak{p}$ to be the prime of L lying below our choice of $\mathfrak{p}_0$ in K. Notice that by changing the prime $\mathfrak{p}_0$ if necessary we may assume $\varepsilon = 12 \sum_{\sigma \in \Sigma_L} \sigma$.

Lemma 3.21. *Let ε be an isogeny signature of type* 3 *with field $L \subseteq k$, and let $\mathfrak{q}$ be a prime ideal of k. Then the integers $A(\varepsilon, \mathfrak{q})$, $B(\varepsilon, \mathfrak{q})$ are nonzero. If in addition the ideal $\mathrm{Nm}_{k/L}(\mathfrak{q})$ is not principal, then $C(\varepsilon, \mathfrak{q})$ is nonzero.*

Proof. Write $\gamma_{\mathfrak{q}}$ for the generator of $\mathfrak{q}^{h_{\mathfrak{q}}}$. By definition of ε, we obtain that, up to complex conjugation in L,

$$\gamma_{\mathfrak{q}}^{\varepsilon} = \mathrm{Nm}_{k/L}(\gamma_{\mathfrak{q}})^{12}$$

and hence, by considering the ideals generated in L,

$$\gamma_{\mathfrak{q}}^{\varepsilon} \mathcal{O}_L = \mathrm{Nm}_{k/L}(\mathfrak{q})^{12h_{\mathfrak{q}}}.$$

If $A(\varepsilon, \mathfrak{q}) = 0$, then $\gamma_{\mathfrak{q}}^{\varepsilon} = 1$ and we would obtain $\mathrm{Nm}_{k/L}(\mathfrak{q})^{12h_{\mathfrak{q}}} = \mathcal{O}_L$ which is clearly not the case. If $B(\varepsilon, \mathfrak{q}) = 0$, then $\gamma_{\mathfrak{q}}^{\varepsilon} = \mathrm{Nm}(\mathfrak{q})^{12h_{\mathfrak{q}}}$ and we would obtain $\mathrm{Nm}_{k/L}(\mathfrak{q})^{12h_{\mathfrak{q}}} = (\mathrm{Nm}(\mathfrak{q})\mathcal{O}_L)^{12h_{\mathfrak{q}}}$, taking the norm form L to $\mathbb{Q}$ of this equation gives $\mathrm{Nm}(\mathfrak{q})^{12h_{\mathfrak{q}}} = (\mathrm{Nm}(\mathfrak{q})^2)^{12h_{\mathfrak{q}}}$ which cannot happen either. Finally, if $C(\varepsilon, \mathfrak{q}) = 0$, then $\gamma_{\mathfrak{q}}^{\varepsilon} = \beta_{\mathfrak{q}}^{12h_{\mathfrak{q}}}$ for a Frobenius root β over $\mathbb{F}_{\mathfrak{q}}$, which implies that $\mathrm{Nm}_{k/L}(\mathfrak{q}) = \beta_{\mathfrak{q}} \mathcal{O}_L$ is principal. □

We then obtain Step 1 of the three step process from signature type 3 to Momose type 3.

Corollary 3.22 (Step 1 of 3). *Let ε be an isogeny signature of type* 3 *with field $L \subseteq k$. Then either k contains the Hilbert class field of L, or for any prime ideal $\mathfrak{q}$ of k whose norm to L is nonprincipal, we*

have that the integer $ABC(\varepsilon, \mathfrak{q})$ *is nonzero*; *in particular*, *in this latter case*, *there are infinitely many such prime ideals* $\mathfrak{q}$.

Proof. The norm on ideals induces a norm map $\mathrm{Nm} : \mathrm{Cl}_k \to \mathrm{Cl}_L$, let $N := \ker \mathrm{Nm}$ denote its kernel. If $N = \mathrm{Cl}_k$ then by class field theory k contains the Hilbert class field of L and we are done. So from now on assume $N \neq \mathrm{Cl}_k$. Then by Lemma 3.21 $ABC(\mathfrak{q}, \varepsilon)$ will be nonzero for all primes $\mathfrak{q}$ in k that are in $\mathrm{Cl}_k \setminus N$, and since $\mathrm{Cl}_k \setminus N \neq \varnothing$ this set of primes is infinite. □

We are now reduced to the case of k containing the Hilbert class field of L, and would like to show that (3-2) is satisfied for all primes $\mathfrak{q}$ of k coprime to $\mathfrak{p}$. Step 2 reduces this task to showing (3-2) is satisfied for all prime ideals in a set of ideals generating Cl_k.

Proposition 3.23 (Step 2 of 3). *Let* λ *be a* k*-rational* p*-isogeny character of signature* ε. *If the following conditions hold*:

(1) $\varepsilon = 12\,\mathrm{Nm}_{k/L}$ *is of type* 3 *with field* L.

(2) k *contains the Hilbert class field of* L.

(3) *There is a generating set* Gen *of* Cl_k *such that* (3-2) *is satisfied for all primes* $\mathfrak{q} \in \mathrm{Gen}$.

Then λ *is of Momose type* 3.

Proof. Let $\mathfrak{q}$ be a prime of k. Then by (2) the ideal $\mathrm{Nm}_{k/L}(\mathfrak{q})$ is principal. Let $\alpha \in L$ be a generator of $\mathrm{Nm}_{k/L}(\mathfrak{q})$; then we need to show that $\lambda^{12}(\mathrm{Frob}_\mathfrak{q}) = \alpha^{12} \pmod{\mathfrak{p}}$.

Write $\mathfrak{q}_1, \ldots, \mathfrak{q}_n$ for the elements of Gen with $n = |\mathrm{Gen}|$. Since Gen generates the class group we can write

$$\mathfrak{q} = \xi \prod_{i=1}^{n} \mathfrak{q}_i^{e_i} \tag{3-3}$$

with $\xi \in k^\times$ and $e_1, \ldots, e_n$ positive integers. Again by (2) all the ideals $\mathrm{Nm}_{k/L}(\mathfrak{q}_i)$ are principal. Let $\alpha_i \in L^\times$ be a generator of $\mathrm{Nm}_{k/L}(\mathfrak{q}_i)$, then $\lambda^{12}(\mathrm{Frob}_{\mathfrak{q}_i}) = \alpha_i^{12} \pmod{\mathfrak{p}}$. This means

$$\lambda^{12}(\mathrm{Frob}_\mathfrak{q}) \equiv \mathrm{Nm}_{k/L}(\xi)^{12} \prod \lambda^{12}(\mathrm{Frob}_{\mathfrak{q}_i}) \pmod{\mathfrak{p}_0} \equiv \mathrm{Nm}_{k/L}(\xi)^{12} \prod \alpha_i^{12} \pmod{\mathfrak{p}}$$

where the first congruence follows from Proposition 3.4. Now applying $\mathrm{Nm}_{k/L}$ to (3-3) gives that $\alpha' := \mathrm{Nm}_{k/L}(\xi)^{12} \prod \alpha_i^{12}$ is a generator of $\mathrm{Nm}_{k/L}(\mathfrak{q})$ so in particular (3-2) holds for the generator α' of $\mathrm{Nm}_{k/L}(\mathfrak{q})$. Now since α and α' both generate $\mathrm{Nm}_{k/L}(\mathfrak{q})$ we have $\alpha/\alpha' \in L^\times$ is a unit. Since L is imaginary quadratic this unit has order dividing 12 and hence $(\alpha/\alpha')^{12} = 1$. In particular $\lambda^{12}(\mathrm{Frob}_\mathfrak{q}) \equiv \alpha'^{12} \equiv \alpha^{12} \pmod{\mathfrak{p}}$. □

The final step is then to show that we can arrange for Condition (3) above to be satisfied. We first define

$$C^*(\varepsilon, \mathfrak{q}) = \mathrm{lcm}(\{\mathrm{Nm}_{K(\beta)/\mathbb{Q}}(\gamma_\mathfrak{q}^\varepsilon - \beta^{12 h_\mathfrak{q}}) \mid \beta \in \bar{k} \text{ is a Frobenius root over } \mathbb{F}_\mathfrak{q}\}), \tag{3-4}$$

where in the lcm we only take nonzero terms. Note that this may be considered the necessarily nonzero part of the integer $C(\varepsilon, \mathfrak{q})$, and the notation has been chosen to reflect this.

Proposition 3.24 (Step 3 of 3). *Let λ be a k-rational p-isogeny character of signature ε of type* 3 *with field L. Assume moreover that $p \geq 17$ and is unramified in L, and that k contains the Hilbert class field of L. Let* Gen *be a set of prime ideal generators of* Cl_k *of odd residue characteristic. Then, either p divides the nonzero integer*

$$ABC^*(\varepsilon, \mathrm{Gen}) := \mathrm{lcm}_{\mathfrak{q}\in\mathrm{Gen}}(A(\varepsilon, \mathfrak{q}), B(\varepsilon, \mathfrak{q}), C^*(\varepsilon, \mathfrak{q}), \mathrm{Nm}(\mathfrak{q}));$$

or, for all $\mathfrak{q} \in$ Gen, *writing $\alpha_\mathfrak{q}$ for a generator of* $\mathrm{Nm}_{k/L}(\mathfrak{q})$, *we have that*

$$\lambda^{12}(\mathrm{Frob}_\mathfrak{q}) \equiv \alpha_\mathfrak{q}^{12} \pmod{\mathfrak{p}}$$

and thus λ is of Momose type 3.

Proof. Let $\mathfrak{q} \in$ Gen, and write q for the rational prime under $\mathfrak{q}$. Since k contains the Hilbert class field of L one has that $\mathrm{Nm}_{k/L}(\mathfrak{q})$ is principal, so write $\alpha_\mathfrak{q}\mathcal{O}_L = \mathrm{Nm}_{k/L}(\mathfrak{q})$. Let $h = h_\mathfrak{q}$ be the order of $\mathfrak{q}$ in Cl_k and $(\gamma) = \mathfrak{q}^h$ be a principal generator.

Since $p \geq 17$ and is assumed unramified in L, one shows in exactly the same way as in the final assertion of the proof of Proposition 3.11 that p splits in L. By applying Proposition 4.9 of [David 2011b] (which is just using her Proposition 2.4 together with the fact that p splits in L) to the principal ideal (γ) we obtain

$$\mu(\mathrm{Frob}_\mathfrak{q})^h \equiv \mathrm{Nm}_{k/L}(\gamma)^{12} \pmod{\mathfrak{p}}.$$

Since $\mathrm{Nm}_{k/L}(\gamma)/\alpha_\mathfrak{q}^h$ is a unit in $\mathcal{O}_L$ we get $(\mathrm{Nm}_{k/L}(\gamma)/\alpha_\mathfrak{q}^h)^{12} = 1$ and we may rewrite the above as

$$\mu(\mathrm{Frob}_\mathfrak{q})^h \equiv \alpha_\mathfrak{q}^{12h} \pmod{\mathfrak{p}};$$

but for the h-th power, this is what we want to obtain, and the rest of the proof is about how to gracefully take the h-th root.

If E has potentially multiplicative reduction at $\mathfrak{q}$, then p divides either $A(\varepsilon, \mathfrak{q})$ or $B(\varepsilon, \mathfrak{q})$, both of which are nonzero by Lemma 3.21. Thus, for p not dividing $\mathrm{lcm}(A(\varepsilon, \mathfrak{q}), B(\varepsilon, \mathfrak{q}))$, we have that E has potentially good reduction at $\mathfrak{q}$, and so there exists a Frobenius root $\beta_\mathfrak{q} \in L^\mathfrak{q}$ of an elliptic curve over $\mathbb{F}_\mathfrak{q}$ and a prime ideal $\mathfrak{p}'$ of $L^\mathfrak{q}$ lying over p such that

$$\lambda(\mathrm{Frob}_\mathfrak{q}) \equiv \beta_\mathfrak{q} \pmod{\mathfrak{p}'} \quad \Longrightarrow \quad \mu(\mathrm{Frob}_\mathfrak{q}) \equiv \beta_\mathfrak{q}^{12} \pmod{\mathfrak{p}'}.$$

We therefore obtain an equality in $\mathbb{F}_p^\times$:

$$\beta_\mathfrak{q}^{12h} \pmod{\mathfrak{p}'} = \alpha_\mathfrak{q}^{12h} \pmod{\mathfrak{p}}.$$

Suppose we had actual equality in characteristic 0:

$$\beta_\mathfrak{q}^{12h} = \alpha_\mathfrak{q}^{12h}.$$

There are two ways this equation could be satisfied: either $L = L^\mathfrak{q}$, or $L \neq L^\mathfrak{q}$ and this equality is between rational numbers. By Corollaries 2.6 and 2.7, the second case can only occur for supersingular values of $\beta_\mathfrak{q}$.

In the case $L = L^{\mathfrak{q}}$ we take the $12h$-th root to obtain $\beta_{\mathfrak{q}} = \zeta \cdot \alpha_{\mathfrak{q}}$ for $\zeta \in L$ a $12h$-th root of unity. However, since L is imaginary quadratic, all roots of unity are either second, fourth, or sixth roots of unity, whence we obtain $\beta_{\mathfrak{q}}^{12} = \alpha_{\mathfrak{q}}^{12}$ and conclude

$$\mu(\mathrm{Frob}_{\mathfrak{q}}) \equiv \alpha_{\mathfrak{q}}^{12} \pmod{\mathfrak{p}}$$

as required.

In the case $L \neq L^{\mathfrak{q}}$ we have that $\beta_{\mathfrak{q}}$ is supersingular and we argue as follows. The definition of $\alpha_{\mathfrak{q}}$ yields the equality of L-ideals

$$\mathrm{Nm}_{k/L}(\mathfrak{q})^{12h} = \alpha_{\mathfrak{q}}^{12h}\mathcal{O}_L.$$

Since $\alpha_{\mathfrak{q}}^{12h} \in \mathbb{Q}$, the $\mathrm{Gal}(L/\mathbb{Q})$-action acts trivially on it and hence also on $\mathrm{Nm}_{k/L}(\mathfrak{q})^{12h}$ and therefore also on $\mathrm{Nm}_{k/L}(\mathfrak{q})$. Let $\mathfrak{q}'$ be the prime of L lying below $\mathfrak{q}$ and f the integer such that $\mathrm{Nm}_{k/L}(\mathfrak{q}) = (\mathfrak{q}')^f$, then applying Proposition 2.2 to $(\mathfrak{q}')^f$ gives

$$\alpha_{\mathfrak{q}}^{12} = \mathrm{Nm}_{L/\mathbb{Q}}(\mathfrak{q}')^{6f} = \mathrm{Nm}_{L/\mathbb{Q}}(\mathrm{Nm}_{k/L}(\mathfrak{q})^6) = \mathrm{Nm}_{k/\mathbb{Q}}(\mathfrak{q})^6.$$

On the other hand, since $\beta_{\mathfrak{q}}$ is a supersingular Frobenius root, we have from Corollary 2.8 that $\beta_{\mathfrak{q}}^{12} = \mathrm{Nm}(\mathfrak{q})^6$ (using that $\mathfrak{q}$ has odd residue characteristic). Combining these we get $\alpha_{\mathfrak{q}}^{12} = \beta_{\mathfrak{q}}^{12}$ and again conclude

$$\mu(\mathrm{Frob}_{\mathfrak{q}}) \equiv \alpha_{\mathfrak{q}}^{12} \pmod{\mathfrak{p}}$$

as required. The final claim that λ is of Momose type 3 now follows from Proposition 3.23.

To summarise, if we had actual equality in characteristic zero between $\alpha_{\mathfrak{q}}^{12h}$ and $\beta_{\mathfrak{q}}^{12h}$, then we are finished. Of course, we cannot in general jump from an equality in characteristic p to one in characteristic 0; but we may do so for p outside the support of an explicitly computable integer. Namely, if $\beta_{\mathfrak{q}}^{12h} \neq \alpha_{\mathfrak{q}}^{12h}$, then p would divide the nonzero integer $\mathrm{Nm}_{L(\beta)/\mathbb{Q}}(\beta_{\mathfrak{q}}^{12h} - \alpha_{\mathfrak{q}}^{12h})$. Since we need to cover all possible roots $\beta_{\mathfrak{q}}$, and observing that we may take $\alpha_{\mathfrak{q}}^{12h}$ to be $\gamma_{\mathfrak{q}}^{\varepsilon}$, we obtain equality in characteristic zero outside of the integer $C^*(\varepsilon, \mathfrak{q})$ defined in (3-4). Doing this for all $\mathfrak{q} \in \mathrm{Gen}$ yields the integer defined in the statement of the proposition. □

In conclusion, we recap this treatment of going from signature type 3 to Momose type 3. Signature type 3 requires only that k contain an imaginary quadratic field, whereas Momose type 3 requires further that k contain the Hilbert class field of L. Step 1 of the above process Corollary 3.22 identifies a nonzero integer, outside of which an isogeny character of signature type 3 must have k containing the Hilbert Class field of L. However, a Momose type 3 isogeny character requires even more than just k containing the Hilbert Class field of L, so Step 3 (Proposition 3.24) identifies a nonzero integer outside of which the isogeny character is indeed of Momose type 3. Taking both of these nonzero integers gives us a multiplicative bound, outside of which an isogeny character of signature type 3 must be of Momose type 3. (This is the integer $D(\varepsilon_3)$ in Algorithm 3.26 in the next subsection.)

3.5. ***The algorithmic version of Momose's isogeny theorem.*** We summarise the previous subsections and present the algorithmic version of Momose's isogeny classification Algorithm 3.26 below takes as input a number field k, and outputs an integer $\mathsf{MMIB}(k)$, referred to as the *Momose multiplicative isogeny*

bound of k. Before showing the details of the algorithm, we present the main result concerning it, which is the main result of the paper.

Theorem 3.25. *Let k be a number field. Then the integer* $\mathsf{MMIB}(k)$ *output by Algorithm 3.26 is nonzero. Furthermore, if there exists an elliptic curve over k admitting a k-rational p-isogeny of isogeny character λ, with $p \nmid \mathsf{MMIB}(k)$, then λ is of Momose type* 1, 2 *or* 3.

For the reader's convenience we will collect the relevant results from this section into a unified proof of the above theorem. Before we can do that, however, we need to present the algorithm that computes $\mathsf{MMIB}(k)$. This depends on the choice of two "auxiliary sets" which are declared at the outset of the following.

Algorithm 3.26. *Given a number field k, compute an integer* $\mathsf{MMIB}(k)$ *as follows*:

(1) *Choose a finite set* Aux *of prime ideals of k which contains at least one totally split prime*; *if the class number h_k of k is greater than* 1, *and k contains an imaginary quadratic field L but not its Hilbert class field, then* Aux *is also required to contain at least one prime ideal of k whose norm to L is nonprincipal.*

(2) *Choose a finite set* AuxGen *of sets* Gen *of prime ideal generators for the class group* Cl_k *of k of odd characteristic.*

(3) *Make the following definitions*:

$$
\begin{aligned}
K &= \textit{Galois closure of } k \textit{ over } \mathbb{Q},\\
\mathfrak{q} &= \textit{a prime ideal of } k,\\
\mathbb{F}_{\mathfrak{q}} &= \textit{residue field of } \mathfrak{q} \textit{ of degree } f,\\
S &= \textit{the generic isogeny signatures for } k \textit{ (see Definition 3.12)},\\
\varepsilon_6 &= \textit{the type 2 signature } (6,\dots,6),\\
A(\varepsilon,\mathfrak{q}) &= \mathrm{Nm}_{K/\mathbb{Q}}(\gamma_{\mathfrak{q}}^{\varepsilon}-1),\\
B(\varepsilon,\mathfrak{q}) &= \mathrm{Nm}_{K/\mathbb{Q}}(\gamma_{\mathfrak{q}}^{\varepsilon}-\mathrm{Nm}(\mathfrak{q})^{12h_{\mathfrak{q}}}),\\
C_s(\varepsilon,\mathfrak{q}) &= \begin{cases} B(2\varepsilon,\mathfrak{q}) & \textit{if } |\mathbb{F}_{\mathfrak{q}}| = 2^f \textit{ with } f \textit{ odd},\\ \mathrm{Nm}_{K/\mathbb{Q}}(\gamma_{\mathfrak{q}}^{\varepsilon}-\mathrm{Nm}(\mathfrak{q})^{6h_{\mathfrak{q}}}) & \textit{otherwise}, \end{cases}\\
C_o(\varepsilon,\mathfrak{q}) &= \mathrm{lcm}\left(\left\{\mathrm{Nm}_{K(\beta)/\mathbb{Q}}(\gamma_{\mathfrak{q}}^{\varepsilon}-\beta^{12h_{\mathfrak{q}}}) \;\middle|\; \begin{array}{l}\beta \textit{ is an ordinary}\\ \textit{Frobenius root over } \mathbb{F}_{\mathfrak{q}}\end{array}\right\}\right),\\
C(\varepsilon,\mathfrak{q}) &= \mathrm{lcm}(C_o(\varepsilon,\mathfrak{q}), C_s(\varepsilon,\mathfrak{q})),\\
ABC(\varepsilon,\mathfrak{q}) &= \mathrm{lcm}(A(\varepsilon,\mathfrak{q}), B(\varepsilon,\mathfrak{q}), C(\varepsilon,\mathfrak{q}), \mathrm{Nm}(\mathfrak{q})),\\
ABC_o(\mathrm{Gen}) &= \mathrm{lcm}_{\mathfrak{q}\in\mathrm{Gen}}(A(\varepsilon_6,\mathfrak{q}), B(\varepsilon_6,\mathfrak{q}), C_o(\varepsilon_6,\mathfrak{q}), \mathrm{Nm}(\mathfrak{q})),\\
\mathrm{lcm}^* &= \textit{an } \mathrm{lcm} \textit{ over only nonzero terms in a given set},\\
C^*(\varepsilon,\mathfrak{q}) &= \mathrm{lcm}^*(\{\mathrm{Nm}_{K(\beta)/\mathbb{Q}}(\gamma_{\mathfrak{q}}^{\varepsilon}-\beta^{12h_{\mathfrak{q}}}) \mid \beta \textit{ is a Frobenius root over } \mathbb{F}_{\mathfrak{q}}\}),\\
ABC^*(\varepsilon,\mathrm{Gen}) &= \mathrm{lcm}_{\mathfrak{q}\in\mathrm{Gen}}(A(\varepsilon,\mathfrak{q}), B(\varepsilon,\mathfrak{q}), C^*(\varepsilon,\mathfrak{q}), \mathrm{Nm}(\mathfrak{q})),\\
\mathsf{GenericBound}(k) &= \mathrm{lcm}_{\varepsilon\in S}\Big(\gcd_{\mathfrak{q}\in\mathrm{Aux}}(ABC(\varepsilon,\mathfrak{q}))\Big),
\end{aligned}
$$

$$\begin{aligned}
\mathsf{TypeTwoNotMomoseBound}(k) &= \gcd_{\mathsf{Gen}\in\mathsf{AuxGen}} (ABC_o(\mathrm{Gen})),\\
S_3 &= \textit{the type 3 signatures for } k,\\
L(\varepsilon_3) &= \textit{the imaginary quadratic field corresponding to } \varepsilon_3 \in S_3,\\
\Delta_{L(\varepsilon_3)} &= \textit{the discriminant of } L(\varepsilon_3),\\
\mathsf{HCF}(L(\varepsilon_3)) &= \textit{the Hilbert class field of } L(\varepsilon_3),\\
D(\varepsilon_3) &= \begin{cases} \gcd_{\mathfrak{q}\in\mathsf{Aux}}(ABC(\varepsilon_3,\mathfrak{q})) & \textit{if } \mathsf{HCF}(L(\varepsilon_3)) \not\subseteq k,\\ \gcd_{\mathrm{Gen}\in\mathsf{AuxGen}}(ABC^*(\varepsilon_3,\mathrm{Gen})) & \textit{otherwise}, \end{cases}\\
\mathsf{TypeThreeNotMomoseBound}(k) &= \begin{cases} \mathrm{lcm}_{\varepsilon\in S_3}(D(\varepsilon)) & \textit{if } \mathsf{HCF}(L(\varepsilon_3)) \not\subseteq k,\\ \mathrm{lcm}_{\varepsilon\in S_3}(D(\varepsilon),\Delta_{L(\varepsilon)}) & \textit{otherwise}, \end{cases}\\
\mathsf{MMIB}(k) &= \mathrm{lcm}\begin{pmatrix} \mathsf{GenericBound}(k),\\ \mathsf{TypeTwoNotMomoseBound}(k),\\ \mathsf{TypeThreeNotMomoseBound}(k) \end{pmatrix}.
\end{aligned}$$

(4) *Return* $\mathsf{MMIB}(k)$.

Note that in the above, contrary to the previous sections, we are taking several prime ideals $\mathfrak{q}$ and several generating sets Gen in items (1) and (2); this allows us to take multiplicative sieves and thereby get a smaller $\mathsf{MMIB}(k)$; this explains the gcds appearing in item (3) of the above algorithm.

We now prove Theorem 3.25.

Proof of Theorem 3.25. Let E/k be an elliptic curve admitting a k-rational p-isogeny of isogeny character λ. We suppose that λ is not of Momose type 1, 2 or 3. We wish to show that $p \mid \mathsf{MMIB}(k)$.

Write ε for the signature of λ. If ε is generic, then by Corollary 3.13 we have that p divides the nonzero integer $ABC(\varepsilon,\mathfrak{q})$ for all completely split prime ideals $\mathfrak{q}$; and hence, by definition of Aux, p divides the nonzero integer $\mathsf{GenericBound}(k)$.

If ε is of type 1, then λ is of Momose type 1, which we have precluded. If ε is of type 2, then we are in the setup of Corollary 3.18, and we obtain that p must divide the nonzero integer $\mathsf{TypeTwoNotMomoseBound}(k)$.

Finally, if ε is of type 3 with field L, then we condition on whether or not $\mathsf{HCF}(L) \subseteq k$. If this is not the case, then we argue as follows. From Corollary 3.10 we have that p divides $\gcd_{\mathfrak{q}\in\mathsf{Aux}}(ABC(\varepsilon_3,\mathfrak{q}))$; to show that this integer is nonzero we use Corollary 3.22, observing that this is precisely why we included in Aux at least one prime ideal of k whose norm to L is nonprincipal. On the other hand, if $\mathsf{HCF}(L) \subseteq k$, then from Proposition 3.24 we have that p divides $D(\varepsilon)$ or p divides $\Delta_{L(\varepsilon)}$. In both cases we obtain that p divides the nonzero integer $\mathsf{TypeThreeNotMomoseBound}(k)$.

Thus, in all cases, we obtain that p divides the nonzero integer $\mathsf{MMIB}(k)$ as required. □

4. GenericBound

This section deals with some aspects of the implementation of $\mathsf{GenericBound}(k)$, particularly regarding how it may be optimised for speed and for obtaining smaller multiplicative bounds. Names of functions refer to functions defined in sage_code/generic.py unless otherwise specified.

4.1. ***Unit precomputation.*** We apply Proposition 3.4 in the case where α is a unit in k, from which we get the following immediate corollary.

Corollary 4.1. *Let λ be a p-isogeny character over k of signature ε and $\alpha \in \mathcal{O}_k^\times$ a unit in k. Then p divides* $\mathrm{Nm}_{K/\mathbb{Q}}(\alpha^\varepsilon - 1)$.

The following result shows that there are many situations in which $\mathrm{Nm}_{K/\mathbb{Q}}(\alpha^\varepsilon - 1)$ is guaranteed to be nonzero for some unit α and hence the divisibility bound on p from the above corollary is nontrivial.

Proposition 4.2. *Let k be a totally real number field of degree d and $\varepsilon = \sum_\sigma a_\sigma \sigma$ a k-isogeny signature such that $\alpha^\varepsilon = 1$ for all $a \in \mathcal{O}_k^\times$. Then all the integers a_σ are the same.*

Proof. This is essentially the same proof as Lemma 3.3 of [Freitas and Siksek 2015]. In this proof (and only in this proof) we override the λ and μ notation from the rest of the paper to accord with the notation in [Neukirch 1999]. Write $\sigma_1, \ldots, \sigma_d$ for the embeddings of k in K, and write a_i for a_{σ_i}. We consider the basic map from the setup of Dirichlet's unit theorem:

$$\lambda : \mathcal{O}_k^\times \to \mathbb{R}^d, \quad u \mapsto (\log|\sigma_1(u)|, \ldots, \log|\sigma_d(u)|),$$

and we have the following facts about λ; see for example Chapter 1, Section 7 of [Neukirch 1999]:

- The image of λ is contained in the *trace-zero hypersurface*
$$H := \{x \in \mathbb{R}^d : \mathrm{Tr}(x) = 0\},$$
and moreover is a complete lattice inside H.
- The kernel of λ is isomorphic to the roots of unity $\mu(k)$ inside k, which in our case of k totally real is just $\{\pm 1\}$.

We therefore obtain that $\mathcal{O}_k^\times/\{\pm 1\}$ maps isomorphically to a complete lattice of dimension $d-1$ under λ.

By our assumption, we also have that $\mathcal{O}_k^\times$ is contained in the hypersurface

$$a_1 x_1 + \cdots + a_d x_d = 0.$$

If not all of the a_i are the same, then this is a different hypersurface to H, and therefore $\mathcal{O}_k^\times$ would be contained in two distinct hypersurfaces; it would thus be contained in their intersection, which would be of dimension $d-2$. It is then clear that $\mathcal{O}_k^\times/\{\pm 1\}$ could not be a complete lattice of dimension $d-1$, yielding the desired contradiction. □

This observation is useful in practice to quickly obtain a bound on isogeny primes of signature ε, since it only requires the computation of a few norm values related to generators of the unit group $\mathcal{O}_k^\times$ which are typically very fast. As such, the resulting divisibilities are computed as an initial step in the main routine, in the function `get_U_integers`.

4.2. ***Choosing auxiliary primes.*** The set `Aux` of auxiliary primes is of central importance in the algorithm to compute `GenericBound`(k), and there are two strategies for constructing `Aux` implemented in *isogeny primes*.

The first is to take all prime ideals up to a bound on the norm which the user may specify (default: 50). The resulting divisibility conditions are not guaranteed to be nontrivial, as this set might not contain a completely split prime of k (which ensures that the integer $ABC(\varepsilon, \mathfrak{q})$ is nonzero); if it does not contain a completely split prime, then we simply add one in. This possibly newly added prime ideal is referred to in the code as an *emergency auxiliary prime*.

The default strategy, however, is an *auto-stop strategy* which successively takes only completely split prime ideals as auxiliary primes, computes the successive integers $\mathsf{GenericBound}(k)$, and terminates when a certain number of them (default: 4) are the same. This is the default behaviour since it was observed during testing that this results in faster runtime whilst not compromising on the tightness of the multiplicative bound.

4.3. ***Optimising signatures.*** For a number field k of degree d, the number of possible k-isogeny signatures ε to be considered (including the type 1 and 2 signatures) is 5^d, which becomes prohibitively large very quickly. Corollary 4.4 below allows one to reduce the number of signatures to be considered by a factor of $2d$ in the Galois case, and by a factor of 2 otherwise. We first establish the following lemma.

Lemma 4.3. *Let E/k be an elliptic curve over a number field admitting a k-rational p-isogeny with isogeny character of signature ε. Then, for $\sigma \in \mathrm{Aut}(k/\mathbb{Q})$, the conjugate elliptic curve $\sigma(E)$ admits a k-rational p-isogeny of signature $\varepsilon \circ \sigma$.*

Proof. Write $H = \mathrm{Aut}(k/\mathbb{Q})$ and $\varepsilon = \sum_{\tau \in \mathrm{Hom}(k,K)} a_\tau \tau$, and let λ be the isogeny character with signature ε. Write as before $\mu = \lambda^{12}$, and identify (as in Proposition 3.4; see also Remark 2.1) μ as an $\mathbb{F}_p^\times$-valued character on the group $I_k(p)$ of ideals of k coprime to p. Fix a prime ideal $\mathfrak{p}$ of K above p. From the natural $\mathrm{Aut}(k/\mathbb{Q})$-action on $I_k(p)$ we obtain, for $\alpha \in k^\times$ coprime to p that

$$\begin{aligned}\mu^\sigma((\alpha)) &= \mu((\sigma(\alpha))) \equiv (\sigma(\alpha))^\varepsilon \pmod{\mathfrak{p}} \\ &\equiv \prod_{\tau \in G} \tau(\sigma(\alpha))^{a_\tau} \pmod{\mathfrak{p}} \equiv \prod_{\tau \in G} \tau(\alpha)^{a_{\tau\sigma^{-1}}} \pmod{\mathfrak{p}} \equiv \alpha^{\varepsilon \circ \sigma} \pmod{\mathfrak{p}},\end{aligned}$$

whence the result follows. □

For $\varepsilon = \sum_\sigma a_\sigma \sigma$, we write $12 - \varepsilon$ for the isogeny signature $\sum_\sigma (12 - a_\sigma)\sigma$.

Corollary 4.4. *Let λ be a p-isogeny character over k of signature ε and $\mathfrak{q}$ a prime of k coprime to p:*

(1) *If p divides $ABC(\varepsilon, \mathfrak{q})$, then p also divides $ABC(12 - \varepsilon, \mathfrak{q})$.*

(2) *For $\sigma \in \mathrm{Aut}(k/\mathbb{Q})$, if p divides $ABC(\varepsilon, \mathfrak{q})$, then p also divides $ABC(\varepsilon \circ \sigma, \mathfrak{q})$.*

Proof. For λ a p-isogeny character of signature ε, we have from [Momose 1995, Remark 2] that the dual isogeny has character $\chi_p \lambda^{-1}$ and is of signature $12 - \varepsilon$; since the dual isogeny also has degree p, this proves (1), and (2) similarly follows from Lemma 4.3. □

type	description	examples
type 1	all $a_i = 0$ or all $a_i = 12$	(0, 0, 0), (12, 12, 12)
quadratic nonconstant	all $a_i \in \{0, 12\}$ not all identical	(0, 12, 0) (12, 12, 0)
sextic constant	All $a_i = 4$ or all $a_i = 8$	(4, 4, 4), (8, 8, 8)
sextic nonconstant	all $a_i \in \{0, 4, 8, 12\}$ not all identical at least one a_i is 4 or 8	(4, 0, 4) (4, 8, 8) (0, 0, 4)
type 2	all $a_i = 6$	(6, 6, 6)
quartic nonconstant	All $a_i \in \{0, 6, 12\}$ not all identical at least one a_i is 6	(0, 6, 0) (12, 6, 0) (6, 6, 0)
mixed	tuple contains 6 and either 4 or 8	(4, 6, 0), (8, 0, 6)

Table 6. The seven signature types. a_i refers to the integers in the signature. The examples show possible signatures for each type arising from a non-Galois cubic number field.

The code which generates the minimum set of ε tuples to be considered (i.e., modding out by the $\varepsilon \mapsto 12 - \varepsilon$ and $\mathrm{Aut}(k/\mathbb{Q})$ actions) is implemented in the function `generic_signatures`. This returns a dictionary with keys the minimum set of signatures, and values the *type* of that signature, as defined (for all signatures, not just the generic ones) in Table 6.

The reason for grouping all signatures into these particular groups is explained in Section 7.1.

Remark 4.5. Since the number of signatures to be considered is exponential in the degree of k, passing to the Galois closure of k over $\mathbb{Q}$ — which was Momose's method of proof of his isogeny theorem — incurs a significant cost to the algorithm. This was one of our main motivations for seeking a method for bounding k-isogenies of prime degree which did not pass to the Galois closure, which led to removing the Galois assumptions in Momose's Lemmas 1 and 2 (which are Propositions 3.4 and 3.11 in Section 3).

5. TypeOneBound

In this section we treat isogenies of signature type 1; that is, $\varepsilon = (0, \ldots, 0)$ or $(12, \ldots, 12)$. As mentioned in the proof of Corollary 4.4, if an elliptic curve admits an isogeny with character of signature $(12, \ldots, 12)$, then the dual isogeny will be of signature $\varepsilon = (0, \ldots, 0)$, so for our purposes of bounding the possible isogeny primes of this signature, we may assume without loss of generality throughout this section that $\varepsilon = \varepsilon_0 = (0, \ldots, 0)$.

The main goal of this section is to explain how the following algorithm produces a multiplicative bound on the isogeny primes of Momose type 1.

Algorithm 5.1. *Given a number field k of degree d, compute an integer* $\mathsf{TypeOneBound}(k)$ *as follows*:

(1) *Choose a finite set* Aux *of odd rational primes q.*

(2) *Make the following definitions.*

$$\begin{aligned}
\mathsf{BadFormalImmersion}(d) &= \textit{product of the bad formal immersion primes in degree } d \\
&\qquad (\textit{see Definition 5.4 and Theorem 5.5});\\
\mathsf{AGFI}_d(q) &= \textit{product of the almost good formal immersion primes in degree } d \textit{ which} \\
&\qquad \textit{are bad formal immersion primes in characteristic } q \\
&\qquad (\textit{see Definition 5.4 and Theorem 5.5});\\
\varepsilon_0 &= \textit{the type 1 signature } (0,\dots,0);\\
D(\mathsf{Aux}) &= \gcd_{\substack{\mathfrak{q}\mid q\\ q\in\mathsf{Aux}}} (\operatorname{lcm}(B(\varepsilon_0,\mathfrak{q}), C(\varepsilon_0,\mathfrak{q}), \operatorname{Nm}(\mathfrak{q}), \mathsf{AGFI}_d(q)))\\
&\qquad (\textit{see Definition 3.8 for } B \textit{ and } C);\\
\mathsf{TypeOneBound}(k) &= \operatorname{lcm}(\mathsf{BadFormalImmersion}(d), D(\mathsf{Aux})).
\end{aligned}$$

(3) *Return* $\mathsf{TypeOneBound}(k)$.

Specifically, the main result of this section is Theorem 1.5 from the introduction.

Theorem 5.2. *Let k be a number field. Then* $\mathsf{TypeOneBound}(k)$ *is nonzero. Moreover, if p is an isogeny prime for k whose associated isogeny character is of Momose type* 1, *then p divides* $\mathsf{TypeOneBound}(k)$.

As one may see from the definitions made in Algorithm 5.1, Theorem 5.2 is concerned with the notion of *good and bad formal immersion primes* of a given degree. We therefore explain these notions before commencing with the proof of Theorem 5.2.

For a scheme X, let $\mathcal{O}_X$ denote the structure sheaf, $\mathcal{O}_{X,x}$ the local ring at a point x of X, and $\widehat{\mathcal{O}}_{X,x}$ the completion of the local ring with respect to its maximal ideal $\mathfrak{m}_{X,x}$. Following [Mazur 1978, Section 3], we make the following definition.

Definition 5.3. If $f : X \to Y$ is a morphism of finite type between noetherian schemes, we say that f is a *formal immersion* at a point x if the induced map on the completion of local rings

$$\widehat{\mathcal{O}}_{Y,f(x)} \to \widehat{\mathcal{O}}_{X,x}$$

is surjective.

We will take X to be the d-th symmetric power (for $d \geq 1$ an integer) of the modular curve $X_0(p)$ (for p prime), which we write as $X_0(p)^{(d)}$, and which we regard as a smooth scheme over $S := \operatorname{Spec}(\mathbb{Z}[1/p])$. We then have a map

$$\begin{aligned}
f_p^{(d)} : X_0(p)^{(d)}_{/S} \to \quad J_0(p)_{/S} \quad &\to J_{e,/S}\\
D \mapsto [D - d(\infty)] &\mapsto [D - d(\infty)] \pmod{\gamma_{\mathfrak{J}} J_0(p)};
\end{aligned}$$

here $\mathfrak{J}$ is the winding ideal of (see [Mazur 1977, Chapter II, Section 18]) and J_e is the winding quotient of $J_0(p)$, the largest rank zero quotient of $J_0(p)$ assuming the Birch–Swinnerton-Dyer conjecture.

The point $x \in X$ at which we consider the formal immersion notion will be the S-section $\infty^d := (\infty, \ldots, \infty)$; however, we will only ever be concerned with a "local" notion of formal immersion; that is, we say that $f_p^{(d)}$ is a formal immersion at ∞^d in characteristic q (for $q \neq p$) if it is a formal immersion when the schemes are considered over the base $\mathrm{Spec}(\mathbb{Z}_{(q)})$.

Merel [1996, Proposition 3] proved, for p sufficiently large with respect to d, that the map $f_p^{(d)}$ *is* in fact a formal immersion along ∞^d in characteristic 3, and even gave an explicit bound on p, which was subsequently improved by Parent [1999, Theorem 1.8 and Proposition 1.9] and Oesterlé,[2] the latter of whom never published his bound, but which may be found explained in [Derickx et al. 2023, Section 6].

We however take a more algorithmic and general approach which works for all odd primes q (not just equal to 3; we will in fact be able to gain some information from considering $q = 2$ as well; see Remark 5.11). To explain our strategy, we make the following definitions.

Definition 5.4. Let $d \geq 1$ be an integer:

(1) For p and q distinct primes, we say that p is a *bad formal immersion prime in degree d at characteristic q* if the map $f_p^{(d)}$ is not a formal immersion in characteristic q.

(2) We say that p is a *bad formal immersion prime in degree d* if the map $f_p^{(d)}$ is not a formal immersion in characteristic q for infinitely many primes q.

(3) We say that p is an *almost good formal immersion prime in degree d* if the map $f_p^{(d)}$ is not a formal immersion in characteristic q for at least one but at most finitely many primes $q \neq 2, p$.

(4) We say that p is a *good formal immersion prime in degree d* if the map $f_p^{(d)}$ *is* a formal immersion in characteristic q for all primes $q \neq 2, p$.

The last definition here might be more accurately described as being *good outside* 2 *and* p, but for simplicity we prefer to have this implicit. Also, since d is usually fixed, we often drop the "in degree d" qualifier.

Our strategy is based on the following result.

Theorem 5.5. *Let $d \geq 1$ be an integer*:

(1) *The set of bad formal immersion primes in degree d is finite and effectively computable.*

(2) *For $q \neq 2$, the set of almost good formal primes in degree d which are bad formal immersion primes in characteristic q is finite and effectively computable.*

This result allows us to define the integers $\mathsf{BadFormalImmersion}(d)$ and $\mathsf{AGFI}_d(q)$ that were already shown in Algorithm 5.1.

To prove Theorem 5.5 we first establish an upper bound on the bad formal immersion primes. Parent [1999, Theorem 1.8 and Proposition 1.9] gives the general bound of $65(2d)^6$; however the following

[2]Pierre Parent was also a student of Bas Edixhoven.

result allows us to reduce this considerably. To state it, we let $M \geq 3$ denote an odd integer, we represent cosets of $(\mathbb{Z}/M\mathbb{Z})^\times$ by $a + M\mathbb{Z}$ with a chosen to satisfy $0 < a < M$, and we define the map

$$\varepsilon_M : (\mathbb{Z}/M\mathbb{Z})^\times \to \{0, 1\}, \quad a + M\mathbb{Z} \mapsto \begin{cases} 0 & \text{if } 1 \leq a < M/2, \\ 1 & \text{otherwise.} \end{cases}$$

For a given $u \in (\mathbb{Z}/M\mathbb{Z})^\times$ we may then define the matrix

$$R_{d,u} := (\varepsilon_M(na) - \varepsilon_M(nu/a))_{1 \leq n \leq d,\ a \in (\mathbb{Z}/M\mathbb{Z})^\times},$$

with entries taken in $\mathbb{Z}$. This is a $d \times \phi(M)$ matrix, where ϕ denotes the Euler totient function.

Proposition 5.6. *Let d be a positive integer, $M \geq 3$ an odd integer, and $u \in (\mathbb{Z}/M\mathbb{Z})^\times$. If the matrix $R_{d,u}$ has rank d over $\mathbb{Z}$, then for all primes $p > 2Md$ such that $pu \equiv 1 \pmod{M}$, p is a good formal immersion prime in degree d.*

Proof. This is in fact a reformulation of Corollary 6.8 in [Derickx et al. 2023]. Here one replaces the auxiliary prime 3 with q throughout Section 6 of [loc. cit.]. The integer M is chosen such that the matrix in the statement of Corollary 6.8 of [loc. cit.] with entries taken in $\mathbb{F}_q$ has rank d for all $u \in (\mathbb{Z}/M\mathbb{Z})^\times$, so the images of $L_1\mathbf{e}, \ldots, L_d\mathbf{e}$ in $H_1(X_0(p)(\mathbb{C}), \mathbb{F}_q)$ are linearly independent over $\mathbb{F}_q$, which by Corollary 6.6 in [loc. cit.] means that we have a formal immersion. □

We may therefore replace Parent's bound of $65(2d)^6$ with $2\hat{M}d$, where $\hat{M}$ is the smallest value of M such that the matrix $R_{d,u}$ has rank d *for all* $u \in (\mathbb{Z}/M\mathbb{Z})^\times$; this may be found algorithmically by trying successively larger odd values of M until one is found which satisfies the rank conditions; the search is terminated at Parent's bound. This is implemented in the `construct_M` function in `sage_code/type_one_primes.py`.

Having now found a bound depending only on d beyond which every prime p is a good formal immersion prime, it follows that, for a given d, there are only finitely many primes for which we need to determine whether they are good, almost good, or bad. The next theorem of Parent — often referred to as *Kamienny's criterion* — is the main ingredient for doing this; to state it, some notation is required:

$\mathbb{T}$: the Hecke algebra [Mazur 1977, Chapter II, Section 6].
T_n : the n-th Hecke operator [Mazur 1977, Chapter II, Section 6].
e : the winding element [Mazur 1977, Chapter II, Section 18].

Theorem 5.7 [Parent 1999, Theorem 4.18]. *Let p and q be distinct primes with $q \neq 2$. Then $f_p^{(d)}$ is a formal immersion at q if and only if the modular symbols $T_1e, \ldots, T_de$ are linearly independent in $\mathbb{T}e/q\mathbb{T}e$.*

Note that the condition $q \neq 2$, p is not mentioned explicitly in [Parent 1999, Theorem 4.18], but it is a running assumption in that is mentioned earlier in the text.

We may now finish the proof of Theorem 5.5.

Proof of Theorem 5.5. From Proposition 5.6 and the definition of $\hat{M}$, we have that for $p > 2\hat{M}d$, p is a good formal immersion prime; thus both of the sets in the statement of the theorem are finite and bounded

d	SGFIP	sporadic bad FI primes
2	23	37
3	41	43, 73
4	47	53, 61, 67, 73, 97
5	59	61, 67, 73, 97
6	71	73, 79, 83, 97, 103, 109, 113
7	101	103, 107, 109, 113, 127, 137, 157
8	131	137, 149, 157, 163, 193
9	131	137, 139, 149, 151, 157, 163, 181, 193
10	167	181, 193, 197, 211, 241

Table 7. The good and bad formal immersion primes in degrees $d \leq 10$. The prime is good if it is larger than or equal to "SGFIP" (smallest good formal immersion prime) and is not in "sporadic bad FI primes". For these degrees no almost bad formal immersion primes were found.

by $2\hat{M}d$, and it remains only to argue that they are effectively computable. We do this by showing that for a given prime p one can exactly compute the primes $q \neq 2$, p at which $f_p^{(d)}$ is not a formal immersion.

By computing Hecke operators up to the Sturm bound one may compute a basis of $\mathbb{T}$ and hence a basis of $\mathbb{T}e$. One then defines a linear map $A : \mathbb{Z}^d \to \mathbb{T}e$ which sends the i-th basis element to $T_i e$.

If A is not injective, then for all primes q the modular symbols $T_1 e, \ldots, T_d e$ are linearly dependent in $\mathbb{T}e/q\mathbb{T}e$, and hence p is bad formal immersion prime by Theorem 5.7.

If A is injective, then $T_1 e, \ldots, T_d e$ are linearly independent over $\mathbb{Z}$, and the finitely many primes q for which they become linearly dependent in $\mathbb{T}e/q\mathbb{T}e$ can be obtained from the Smith normal form of a matrix for A. In particular, p is a good formal immersion prime if all nonzero coefficients of the Smith normal form are coprime to $2p$, and p is an almost good formal immersion prime otherwise. □

Remark 5.8. Kamienny [1992b, Proposition 3.1] originally formulated his criterion in terms of the linear independence of Hecke operators in $\mathbb{T} \otimes \mathbb{F}_q$. This was subsequently developed by Merel, Oesterlé and Parent, and most recently generalised and made algorithmic by work of Derickx, Kamienny, Stein and Stoll; the state-of-the-art for general modular curves X_H is Proposition 5.3 of [Derickx et al. 2023] which also works when $q = 2$, but has the drawback of not giving an if and only if statement. Our algorithm only works for $X_0(p)$ and $q > 2$, however it gives more precise information.

The algorithm given by the proof of Theorem 5.5 is implemented in our package as the functions R_dp, is_formal_immersion and bad_formal_immersion_data in sage_code/type_one_primes.py. The results of the computation are stored in bad_formal_immersion_data.json as a method of caching these results to save time on subsequent runs of the algorithm for a given degree. (That is, this computation is only run once for each new d encountered; thereafter the values are looked up in the JSON file.) Table 7 shows the results of running this algorithm for degrees $d \leq 10$, where it was found that there were no almost good formal immersion primes (i.e., every prime $p \neq 2$ was either a good or a bad formal immersion prime in degree d).

Remark 5.9. We in fact ran the computation for all degrees up to 20, and still found no almost good formal immersion primes. We can find no reason to explain this.

We are now ready to present the proof of the main theorem of this section.

Proof of Theorem 5.2. We take $q \neq p$ a rational prime, and $\mathfrak{q} \mid q$ an auxiliary prime. If E has potentially good reduction at $\mathfrak{q}$, then by Corollary 3.10 p will divide $C(\varepsilon_0, \mathfrak{q})$. If this integer were zero, then by definition of $C(\varepsilon_0, \mathfrak{q})$ there would exist a Frobenius root β satisfying $\beta^{12h_\mathfrak{q}} = \gamma_\mathfrak{q}^{\varepsilon_0}$, where $\gamma_\mathfrak{q}$ is a generator of the principal ideal $\mathfrak{q}^{h_\mathfrak{q}}$; but since $\varepsilon_0 = (0, \ldots, 0)$, we deduce that $\beta^{12h_\mathfrak{q}} = 1$, which cannot happen (since β has absolute value an integer multiple of $\sqrt{q}$); thus $C(\varepsilon, \mathfrak{q}) \neq 0$, which gives a multiplicative bound on the primes p for which E has potentially good reduction at $\mathfrak{q}$, and explains the presence of the integer $C(\varepsilon_0, \mathfrak{q})$ in the integer $D(\mathsf{Aux})$ in item (2) of Algorithm 5.1.

We are thus reduced to considering E having potentially multiplicative reduction at $\mathfrak{q}$. That is, writing x for the noncuspidal K-point on $X_0(p)$ corresponding to E, we have that x reduces modulo $\mathfrak{q}$ to one of the two cusps 0 or ∞.

Suppose first that x reduces modulo $\mathfrak{q}$ to the zero cusp. This means that the kernel W of the isogeny specialises to the identity component $(E_{/\mathbb{F}_\mathfrak{q}})^0$ of the reduction of E at $\mathfrak{q}$. Since this coincides with the group scheme μ_p of p-th roots of unity over an (at most) quadratic extension of $\mathbb{F}_\mathfrak{q}$, we obtain an equality of groups between $W(\overline{K_\mathfrak{q}})$ and $\mu_p(\overline{K_\mathfrak{q}})$, where $K_\mathfrak{q}$ denotes the completion of K at $\mathfrak{q}$. This precise case is considered by David in the proof of Proposition 3.3 of [David 2011a], and corresponds to the isogeny character satisfying $\lambda^2(\mathrm{Frob}_\mathfrak{q}) \equiv \mathrm{Nm}(\mathfrak{q})^2 \pmod{p}$. However, since $\varepsilon = (0, \ldots, 0)$ we obtain that $\lambda^{12h_\mathfrak{q}}$ acts trivially on $\mathrm{Frob}_\mathfrak{q}$, yielding the divisibility

$$p \mid \mathrm{Nm}(\mathfrak{q})^{12h_\mathfrak{q}} - 1, \tag{5-1}$$

i.e., that $p \mid B(\varepsilon_0, \mathfrak{q})$.

Suppose next that x reduces modulo $\mathfrak{q}$ to the infinite cusp. We may then consider the point x^σ, for $\sigma \in \Sigma$ an embedding of k in K, and observe that, if there is an x^σ which specialises to the zero cusp modulo $\mathfrak{q}$, then we may apply the previous argument to the conjugate curve E^σ (which also has isogeny signature $(0, \ldots, 0)$) and conclude as before that $p \mid B(\varepsilon_0, \mathfrak{q})$.

We are thus reduced to considering the case that the S-section $(x^\sigma)_{\sigma \in \Sigma}$ (for $S := \mathrm{Spec}(\mathbb{Z}[1/p])$ as before) of $X_0(p)^{(d)}$ "meets" the section ∞^d at $\mathfrak{q}$ (to use Kamienny's terminology in the discussion immediately preceding his Theorem 3.4 in [Kamienny 1992a]; Mazur's language for this in the proof of Corollary 4.3 of [Mazur 1978] is that these two sections "cross" at $\mathfrak{q}$). By applying Kamienny's extension of Mazur's formal immersion argument [Kamienny 1992b, Theorem 3.3], augmented with Merel's use of the winding quotient instead of the Eisenstein quotient, this implies that the map $f_p^{(d)}$ is *not* a formal immersion along ∞^d in characteristic q; therefore, either p divides $\mathsf{BadFormalImmersion}(d)$, or p divides $\mathsf{AGFI}_d(q)$. Since these integers are defined as products of primes, they are nonzero. □

Remark 5.10. Note that Momose obtains the erroneous bound $p - 1 \mid 12h_k$ at the point where we obtain the multiplicative bound in (5-1). This occurs in the proof of [Momose 1995, Theorem 3], and is arrived

at through the claim that "the restriction of λ to the inertial subgroup $I_{\mathfrak{q}}$ of $\mathfrak{q}$ is $\pm\theta_p$" (θ_p being Momose's notation for the mod-p cyclotomic character). Momose cites the whole of [Deligne and Rapoport 1973] for this claim, which alas we were unable to find in those 174 pages. If this claim were true, then, since we also know that λ^2 is unramified at $\mathfrak{q}$, we would obtain that $p-1 \mid 2$, i.e., that $p = 2$ or 3.

Remark 5.11. While Theorem 5.5 is enough to obtain a finite superset for the type 1 primes, it is not making use of information that may be obtained from considering the auxiliary prime $q = 2$. The difficulty here is that there is no currently known lower bound (analogous to Proposition 5.6) on the primes p beyond which $f_p^{(d)}$ is a formal immersion at ∞^d in characteristic 2.

Nevertheless, one may still employ Kamienny's criterion for various small values of d and primes up to a reasonable bound and record the primes for which we *do* have formal immersion in characteristic 2. This was in fact already done for $3 \le d \le 7$ and for primes $11 \le p \le 2281$ in Lemma 5.4 of [Derickx et al. 2023], using Magma code which may be found on Stoll's webpage [2017]. Running this code also for $d = 2$ and slightly extending the bound on p allows us to make use of the auxiliary prime 2 to rule out possible type one primes p which are less than 2371, unless they are in an explicitly computed set which is found in the file formal_immersion_at_2.json, which contains the hardcoded information, for each $2 \le d \le 7$ and each prime $13 \le p \le 2371$, whether or not we have formal immersion at 2.

6. Momose type 2 primes

In this section we consider isogenies whose signature is not only of type 2, but whose isogeny character is furthermore of Momose type 2. We start with a necessary condition on p which must by satisfied by such isogenies. This generalises the condition ("Condition CC") given for quadratic fields by the first named author in [Banwait 2023].

Proposition 6.1. *Let k be a number field, and E/k an elliptic curve admitting a k-rational p-isogeny of Momose type 2. Let q be a rational prime, and $\mathfrak{q}$ a prime ideal of k dividing q of residue degree f satisfying the following conditions*:

(1) $q^f < p/4$.

(2) *f is odd.*

(3) $q^{2f} + q^f + 1 \not\equiv 0 \pmod{p}$.

Then q does not split in $\mathbb{Q}(\sqrt{-p})$.

Proof. If E has potentially good reduction at $\mathfrak{q}$, then by the first and second assumption, we get from [Momose 1995, Lemma 5] that q is inert in $\mathbb{Q}(\sqrt{-p})$.

We claim that if E has potentially multiplicative reduction at $\mathfrak{q}$, and q splits in $\mathbb{Q}(\sqrt{-p})$, then $q^{2f} + q^f + 1 \equiv 0 \pmod{p}$, contradicting assumption 3.

To establish the claim, denote by λ the isogeny character of E, assumed to be of Momose type 2. From [Momose 1995, Lemma 3], λ is of the form $\psi\chi_p^{(p+1)/4}$ for a character ψ of order dividing 6.

This claim is then established as follows:

$$\begin{aligned}
&\mathrm{Nm}(\mathfrak{q}) \equiv \psi(\mathrm{Frob}_{\mathfrak{q}})^{\pm 2} \pmod{p} \text{(by [Momose 1995, Lemma 4])} \\
&\implies \mathrm{Nm}(\mathfrak{q}) \text{ is a third root of unity in } \mathbb{F}_p^\times \text{ (since } \psi \text{ has order dividing 6 by [Momose 1995, Lemma 3])} \\
&\iff q^f \text{ is a third root of unity in } \mathbb{F}_p^\times \text{ (by definition of } f) \\
&\iff q^{2f} + q^f + 1 \equiv 0 \pmod{p} \text{ (since (1) implies } q^f \not\equiv 1 \pmod{p}).
\end{aligned}$$

□

Remark 6.2. See Lemma 5.1 of [Banwait 2023] for more details on the proofs of Lemmas 3, 4 and 5 in [Momose 1995].

Next we provide a bound on the isogeny primes of Momose type 2. This is the only result in our work which requires the generalised Riemann hypothesis.

Proposition 6.3. *Assume GRH. Let k be a number field of degree d, and E/k an elliptic curve possessing a k-rational p-isogeny, for p a type 2 prime. Then p satisfies*

$$p \leq (8d\log(12p) + 16\log(\Delta_k) + 10d + 6)^4. \tag{6-1}$$

In particular, there are only finitely many primes p as above.

Proof. In the proof of Theorem 6.4 of [Larson and Vaintrob 2014], the authors prove that a Momose type 2 prime p satisfies

$$p \leq (1 + \sqrt{\mathrm{Nm}_{\mathbb{Q}}^k(v)})^4, \tag{6-2}$$

where v is a prime ideal of k such that v is split in $k(\sqrt{-p})$, is of degree 1, does not lie over 3, and satisfies the inequality

$$\mathrm{Nm}_{k/\mathbb{Q}}(v) \leq c_7 \cdot (\log \Delta_{k(\sqrt{\pm p})} + n_{k(\sqrt{\pm p})} \log 3)^2$$

for an effectively computable absolute constant c_7 (note that they use ℓ instead of p, and n_k denotes the degree of k).

The existence of such a v follows from their Corollary 6.3, which requires GRH. Stepping into the proof of this Corollary, we arrive at a point where they apply the effective Chebotarev density theorem to $\mathrm{Gal}(E'/k)$ — for a certain Galois extension E' which fits into a tower of successive extensions $E'/E/k/\mathbb{Q}$ — to bound the norm of v as

$$\mathrm{Nm}_{k/\mathbb{Q}}(v) \leq c_5(\log \Delta_{E'})^2$$

for an effectively computable absolute constant c_5.

However, we may at this breakpoint instead use Theorem 5.1 of [Bach and Sorenson 1996] on the Galois extension E'/k of absolute degree $4d$ to obtain

$$\mathrm{Nm}_{k/\mathbb{Q}}(v) \leq (4\log \Delta_{E'} + 10d + 5)^2.$$

This then subsequently (stepping back out into the proof of Theorem 6.4) yields the bound

$$\mathrm{Nm}_{k/\mathbb{Q}}(v) \leq (16\log \Delta_k + 8d\log p + 8d\log 6 + 8d\log 2 + 10d + 5)^2.$$

Inserting this into (6-2) yields the result. □

Coupling Proposition 6.1 with the explicit bound on type 2 primes given in Proposition 6.3, we are able to algorithmically determine a superset for the type 2 primes.

Algorithm 6.4. *Given a number field k of degree d, compute a set of primes* TypeTwoPrimes(k) *as follows*:

(1) *Initialise* TypeTwoPrimes(k) *to the empty set.*

(2) *Compute the bound B_k on p implied by* (6-1).

(3) *For $p \leq B_k$*:

  (a) *For each of the finitely many q satisfying the conditions* (1)-(3) *of Proposition 6.1*:

    (i) *If q splits in $\mathbb{Q}(\sqrt{-p})$, then break this for-loop over q, and continue to the next p.*

    (ii) *else continue to the next q.*

    (iii) *If one has not continued to the next p by this point, then append p to* TypeTwoPrimes(k) *before continuing to the next p.*

(4) *Return* TypeTwoPrimes(k).

From the above discussion it is clear that the returned set is a superset for the primes p for which there exists an elliptic curve over k admitting a k-rational p-isogeny of Momose type 2. This algorithm is implemented in the function type_2_primes in sage_code/type_two_primes.py, though as in [Banwait 2023] the bounds become rather large, so an optimised and parallelised implementation in PARI/GP may be found in gp_scripts/partype2primes.gp. Note that, by Proposition 3.1, if E is semistable, then signature type 2 does not arise, so one does not have Momose type 2 primes, and hence the algorithm to compute a superset for the *semistable isogeny primes* is unconditional.

7. Automatic weeding

With the ideas presented in the paper up to this point, we obtain, for each signature ε, a nonzero integer $M(\varepsilon)$ such that, if there is an elliptic curve over k admitting a k-rational p-isogeny, then $p \mid M(\varepsilon)$. We refer to $M(\varepsilon)$ as the multiplicative upper bound on the isogeny primes of signature ε, since primes dividing this integer are not necessarily actual isogeny primes for k.

This gives rise to the notion of *weeding*, by which we mean techniques and conditions for ruling out possible isogeny primes for k. As part of our new software package *Isogeny Primes*, we have automated many of these methods, which this section gives an overview of. These methods are all applied after a call to prime_divisors has been made, so one has in hand candidate isogeny primes p at this point.

The top-level function which executes these automated weeding methods is apply_weeding in sage_code/weeding.py, called at the end of the routine.

7.1. ***Filtering isogeny primes of signature ε.*** By combining the admissibility criterion Corollary 3.5 with the necessary congruence conditions for the arising of the integers 4, 6 and 8 in an isogeny signature, we obtain the following necessary conditions.

type	p must split or ramify?	congruence conditions
type 1	no	none
quadratic nonconstant	yes	none
sextic constant	no	$p \equiv 2 \pmod 3$
sextic nonconstant	yes	$p \equiv 2 \pmod 3$
type 2	no	$p \equiv 3 \pmod 4$
quartic nonconstant	yes	$p \equiv 3 \pmod 4$
mixed	yes	$p \equiv 1 \pmod{12}$

Table 8. Necessary conditions for the existence of a k-rational p-isogeny for each broad ε-type. "NO" is to be interpreted as "not necessarily".

Corollary 7.1. *For each signature type, Table 8 gives necessary conditions that must be satisfied for the existence of a k-rational p-isogeny.*

These necessary conditions are implemented in filter_ABC_primes in sage_code/common_utils.py, and it explains the splitting of epsilons into the seven broad types expressed in Table 6. We take the LCM of the multiplicative upper-bound integers for each ε of a given type, and call prime_divisors on each of the resulting 7 integers, allowing one to significantly reduce the number of calls to this potentially expensive function. This filtering always occurs by default in the algorithm.

7.2. ***Isogeny character enumeration.*** The grouping of all signatures into seven broad types, while possibly allowing for faster runtime, does come at the expense of possibly worse filtering. Knowing that a p-isogeny must arise from a given isogeny signature ε allows for finer necessary conditions, which we refer to as *isogeny character enumeration*.

This approach is based on the interpretation of a p-isogeny character λ chiefly as a character of the group of ideals of k coprime to p (see Remark 2.1), which is achieved via class field theory. Since we need to be quite precise about this, we introduce the following notation to be used in addition to that set in Section 2:

$$\begin{aligned}
\mathcal{P} &: \text{the set of primes of } k \text{ lying over } p\\
\mathfrak{m}_p &: \textstyle\prod_{\mathfrak{p}\in\mathcal{P}} \mathfrak{p}\\
\mathcal{I}_k &: \text{the fractional ideals of } k\\
\mathcal{I}_k^{\mathfrak{m}_p} &: \text{the fractional ideals of } k \text{ coprime to } \mathfrak{m}_p\\
k^{\mathfrak{m}_p} &: \text{the units in } k^\times \text{ coprime to } \mathfrak{m}_p\\
k^{\mathfrak{m}_p,1} &: \text{the subgroup of } a \in k^{\mathfrak{m}_p} \text{ satisfying } v_{\mathfrak{p}}(a-1) \geq v_{\mathfrak{p}}(\mathfrak{m}_p) = 1 \text{ for } \mathfrak{p} \in \mathcal{P}\\
\mathcal{R}_k^{\mathfrak{m}_p} \subseteq \mathcal{I}_k^{\mathfrak{m}_p} &: \text{the principal ideals generated by } k^{\mathfrak{m}_p}\\
\mathcal{R}_k^{\mathfrak{m}_p,1} \subseteq \mathcal{I}_k^{\mathfrak{m}_p} &: \text{the principal ideals generated by } k^{\mathfrak{m}_p,1}\\
\mathrm{Cl}_k^{\mathfrak{m}_p} &: \mathcal{I}_k^{\mathfrak{m}_p}/R_k^{\mathfrak{m}_p,1}, \text{ the ray class group of modulus } \mathfrak{m}_p.
\end{aligned}$$

Since the codomain of λ^{12} is $\mathbb{F}_p^\times$ it follows that λ^{12} is tamely ramified at all primes in $\mathcal{P}$, and we have seen from earlier that λ^{12} is unramified at all primes outside of $\mathcal{P}$. Thus it is clear that $\mathfrak{m}_p$ is a modulus

for λ^{12}, meaning that $\mathcal{R}_k^{\mathfrak{m}_p,1} \subseteq \ker \lambda^{12}$ and hence that $\lambda^{12} : \mathcal{I}_k^{\mathfrak{m}_p} \to \mathbb{F}_p^\times$ factors through $\mathrm{Cl}_k^{\mathfrak{m}_p}$. From the exact sequence

$$1 \to \mathcal{R}_k^{\mathfrak{m}_p}/\mathcal{R}_k^{\mathfrak{m}_p,1} \to \mathrm{Cl}_k^{\mathfrak{m}_p} \to \mathrm{Cl}_k \to 1,$$

together with the isomorphism $\mathcal{R}_k^{\mathfrak{m}_p}/\mathcal{R}_k^{\mathfrak{m}_p,1} \cong (\mathcal{O}_k/\mathfrak{m}_p)^\times/\mathcal{O}_k^\times$, Proposition 3.4 can be interpreted as saying that ε classifies the possibilities of $\lambda^{12}|_{\mathcal{R}_k^{\mathfrak{m}_p}}$. In particular ε and $\mathfrak{p}_0$ together determine $\lambda^{12}|_{\mathcal{R}_k^{\mathfrak{m}_p}}$.

Now, given a prime $\mathfrak{p}_0$ in $\mathcal{O}_K$, and a signature ε one can define

$$\chi_{\varepsilon,\mathfrak{p}_0} : k^{\mathfrak{m}_p} \to (\mathcal{O}_K/\mathfrak{p}_0)^\times, \quad \alpha \mapsto \alpha^\varepsilon \pmod{\mathfrak{p}_0}.$$

There are two necessary conditions which the character $\chi_{\varepsilon,\mathfrak{p}_0}$ should satisfy in order for it to come from the twelfth power of an isogeny character:

(1) $\chi_{\varepsilon,\mathfrak{p}_0}(\mathcal{O}_k^\times) = 1$.

(2) $\mathrm{im}\, \chi_{\varepsilon,\mathfrak{p}_0} \subseteq (\mathbb{F}_p^\times)^{12} \subseteq (\mathcal{O}_K/\mathfrak{p}_0)^\times$.

The first condition comes from the fact that $\chi_{\varepsilon,\mathfrak{p}_0}(\alpha)$ should only depend on the ideal that α generates, and the second one is clear since isogeny characters are $\mathbb{F}_p^\times$ valued.

Condition (1) is easy to computationally verify since one can just compute $\chi_{\varepsilon,\mathfrak{p}_0}$ on a set of generators for $\mathcal{O}_k^\times$. Condition (2) can also be verified by observing that $\chi_{\varepsilon,\mathfrak{p}_0}$ factors through the natural map

$$k^{\mathfrak{m}_p} \to (\mathcal{O}_k/\mathfrak{m}_p)^\times \cong \prod_{\mathfrak{p}\in\mathcal{P}} (\mathcal{O}_{k_\mathfrak{p}}/\mathfrak{p})^\times,$$

so it suffices to verify condition (2) on lifts of generators of each finite field $(\mathcal{O}_{k_\mathfrak{p}}/\mathfrak{p})^\times$.

Once conditions (1) and (2) are satisfied by $\chi_{\varepsilon,\mathfrak{p}_0}$ we can define $\chi'_{\varepsilon,\mathfrak{p}_0}$ as the character $\chi'_{\varepsilon,\mathfrak{p}_0} : \mathcal{R}_k^{\mathfrak{m}_p} \to (\mathbb{F}_p^\times)^{12}$ making the following diagram commute:

$$\begin{array}{ccccccccc}
1 & \longrightarrow & \mathcal{O}_k^\times & \longrightarrow & k^{\mathfrak{m}_p} & \longrightarrow & \mathcal{R}_k^{\mathfrak{m}_p} & \longrightarrow & 1 \\
 & & & & \downarrow{\scriptstyle \chi_{\varepsilon,\mathfrak{p}_0}} & & \downarrow{\scriptstyle \chi'_{\varepsilon,\mathfrak{p}_0}} & & \\
 & & & & (\mathcal{O}_K/\mathfrak{m}_p)^\times & \hookleftarrow & (\mathbb{F}_p^\times)^{12} & &
\end{array}$$

Thus, the question of whether $\chi_{\varepsilon,\mathfrak{p}_0}$ arises as the twelfth power of an isogeny character is equivalent to whether $\chi'_{\varepsilon,\mathfrak{p}_0}$ can be extended to a character $\mu : \mathcal{I}_k^{\mathfrak{m}_p} \to (\mathbb{F}_p^\times)^{12}$:

$$\begin{array}{ccccccccc}
1 & \longrightarrow & \mathcal{R}_k^{\mathfrak{m}_p} & \longrightarrow & \mathcal{I}_k^{\mathfrak{m}_p} & \longrightarrow & \mathrm{Cl}_k & \longrightarrow & 1 \\
 & & \downarrow{\scriptstyle \chi'_{\varepsilon,\mathfrak{p}_0}} & \swarrow & \nearrow{\scriptstyle \exists \mu} & & & & \\
 & & (\mathbb{F}_p^\times)^{12} & & & & & &
\end{array}$$

Since Cl_k is a finitely generated group this can be determined algorithmically as follows. Starting with a set of ideals $I_1, \ldots, I_j \in \mathcal{I}_k^{\mathfrak{m}_p}$ which together generate Cl_k we iteratively compute a list L_i that contains

the different sequences of possible values $\mu(I_1), \ldots, \mu(I_i)$ for $i = 1, \ldots, j$. Precisely, we implement the following algorithm.

Algorithm 7.2. *Given a number field k, a prime p, and isogeny signature ε, return a set L as follows*:

(1) *Compute ideals $I_1, \ldots, I_j \in \mathcal{I}_k^{\mathfrak{m}_p}$ that together generate* Cl_k.

(2) *Let* $\mathrm{Cl}_0 := 1 \subset \mathrm{Cl}_k$ *be the trivial subgroup, and for* $i = 1, \ldots, j$ *let* $\mathrm{Cl}_i \subset \mathrm{Cl}_k$ *be the subgroup generated by* $I_1, \ldots, I_i$.

(3) *Let* $I_0 = (1)$ *and* $L_0 = [[1]]$.

(4) *For* $i = 1, \ldots, j$ *do*:

  (a) *Let* $L_i := []$ *be the empty list.*

  (b) *Compute* h_i *as the order of* I_i *in* $\mathrm{Cl}_k / \mathrm{Cl}_{i-1}$.

  (c) *Write* $I_i^{h_i} \sim \prod_{j=1}^{i-1} I_j^{e_{i,j}}$ *in* Cl_{i-1}.

  (d) *Compute a generator* α_i *of the principal ideal* $I_i^{h_i} \prod_{j=1}^{i-1} I_j^{-e_{i,j}}$.

  (e) *For each sequence of possible values* $c_0, c_1, \ldots, c_{i-1} \in (\mathbb{F}_p^\times)^{12}$ *of* L_{i-1} *do*:

    (i) *Compute* $c' := \chi_{\varepsilon,\mathfrak{p}_0}(\alpha_i) \prod_{j=1}^{i-1} c_j^{e_{i,j}}$.

    (ii) *Compute the set* $\{r_1, \ldots, r_{d_i}\}$ *of* h_i*-th roots of* c' *in* $(\mathbb{F}_p^\times)^{12}$.

    (iii) *Append* $[c_0, c_1, \ldots, c_{i-1}, r_1]$ *up to* $[c_0, c_1, \ldots, c_{i-1}, r_{d_i}]$ *to* L_i.

(5) *Return* $L = L_j$.

At the end of this algorithm L will be a list of $(j+1)$-tuples corresponding to all possible extensions μ of $\chi'_{\varepsilon,\mathfrak{p}_0}(\alpha_i)$ to $\mathcal{I}_k^{\mathfrak{m}_p}$ (for $i = 1, \ldots, j$); specifically, for $c = c_0, c_1, \ldots, c_j$, the extension μ_c corresponding to it is the character such that $\mu_c(I_i) = c_i$ for $i = 1, \ldots, j$.

Remark 7.3. At every single iteration of each loop, we have that $c_0 = 1$, so it can be left out. However the algorithm seems easier to understand with this initial condition.

Given one such j-tuple $c_1, \ldots, c_j$ corresponding to an extension μ of $\chi'_{\varepsilon,\mathfrak{p}_0}$, one may readily compute the value $\mu(I)$ for I any ideal of $I_k^{\mathfrak{m}_p}$ as follows.

Algorithm 7.4. *Given a j-tuple $c_1, \ldots, c_j$ of elements of $(\mathbb{F}_p^\times)^{12}$ corresponding to an extension μ of $\chi'_{\varepsilon,\mathfrak{p}_0}$, and an ideal I of $I_k^{\mathfrak{m}_p}$, compute $\mu(I) \in (\mathbb{F}_p^\times)^{12}$ as follows*:

(1) *Write* $I \sim \prod_{i=1}^{j} I_i^{e_i}$ *in* Cl_k.

(2) *Compute a generator* α *of the principal ideal* $I \prod_{i=1}^{j} I_i^{-e_i}$.

(3) *Return* $\mu(I) = \chi_{\varepsilon,\mathfrak{p}_0}(\alpha) \prod_{i=1}^{j} c_i^{e_i}$.

Having computed all characters μ extending $\chi'_{\varepsilon,\mathfrak{p}_0}$ we can try to prove that $\chi'_{\varepsilon,\mathfrak{p}_0}$ does not come from the twelfth power of an isogeny character by proving that any μ extending it does not come from an isogeny character. This is achieved by the following algorithm, which for an input prime ideal $\mathfrak{q}$ computes whether $\mu(\mathfrak{q})$ is a twelfth power of a Frobenius root mod p.

Algorithm 7.5. *Given a j-tuple $c_1, \ldots, c_j$ of elements of $(\mathbb{F}_p^\times)^{12}$ corresponding to an extension μ of $\chi'_{\varepsilon,\mathfrak{p}_0}$, and a prime ideal $\mathfrak{q}$ of $\mathcal{O}_k$ that is prime to p, return* True *or* False *as follows*:

(1) *Compute $c = \mu(\mathfrak{q})$ as in Algorithm 7.4 .*

(2) *For all twelfth roots $r \in \mathbb{F}_p^\times$ of c do*:

  (a) *Compute $\bar{a}_\mathfrak{q} := r + \mathrm{Nm}(\mathfrak{q})/r$.*

  (b) *For all $a_\mathfrak{q} \in \mathbb{Z}$ with $a_\mathfrak{q} \equiv \bar{a}_\mathfrak{q} \pmod{p}$ and $|a_\mathfrak{q}| \leq 2\sqrt{\mathrm{Nm}(\mathfrak{q})}$ do*:

    (i) *If $x^2 - a_\mathfrak{q} x + \mathrm{Nm}(\mathfrak{q})$ is a Frobenius polynomial, return* True *and terminate.*

(3) *Return* False.

Step (2.b.i) can be verified using Theorem 2.5. Note that it makes sense to run this algorithm for all $\mathfrak{q}$ with $4\sqrt{\mathrm{Nm}(\mathfrak{q})} < p$, since for these $\mathfrak{q}$ there is a chance that an $a_\mathfrak{q}$ as in step (b) does not exist, and hence that the algorithm returns False.

In the type 1 case things work differently, since here we have to consider that μ is either trivial or χ_p^{12}. Specifically, we apply the following.

Algorithm 7.6. *Given a j-tuple $c_1, \ldots, c_j$ of elements of $(\mathbb{F}_p^\times)^{12}$ corresponding to an extension μ of $\chi'_{\varepsilon,\mathfrak{p}_0}$, return* True *or* False *as follows*:

(1) *Set values* $= []$.

(2) *For all primes $\mathfrak{q}_1, \ldots, \mathfrak{q}_d \mid q\mathcal{O}_k$ do*:

  (a) *Compute $\mu(\mathfrak{q}_i)$ as in Algorithm 7.4.*

  (b) *If $\mu(\mathfrak{q}_i)$ is an $\mathfrak{q}_i$-Frobenius root mod p, return* False.

  (c) *Append $\mu(\mathfrak{q}_i)$ to values.*

(3) *If values* $= [1, 1, \ldots, 1]$ *or* $[\mathrm{Nm}(\mathfrak{q}_1)^{12}, \mathrm{Nm}(\mathfrak{q}_2)^{12}, \ldots, \mathrm{Nm}(\mathfrak{q}_d)^{12}]$, *return* True; *else return* False.

Note that if $\mu(\mathfrak{q}_i)$ at step (2.b) above is not 1 or $\mathrm{Nm}(\mathfrak{q}_i) \pmod{p}$ then in fact μ cannot be an isogeny character.

Putting all of these parts together, we arrive at the main algorithm based on isogeny character enumeration, which is implemented in the function character_enumeration_filter located in sage_code/character_enumeration.py, and which requires one to call to prime_divisors for each ε.

Algorithm 7.7. *Given a number field k, a prime p, and isogeny signature ε, return* True *or* False *as follows*:

(1) *Run Algorithm 7.2, and write L for the output of it.*

(2) *If ε is not of signature type* 1:

  (a) *For each j-tuple in L, apply Algorithm 7.4. If any tuple returns* True, *return* True; *else return* False.

(3) *If ε is of signature type* 1:

  (a) *For each j-tuple in L, apply Algorithm 7.6. If any tuple returns* True, *return* True; *else return* False.

Remark 7.8. The filtering arising from Isogeny character enumeration occurs as default in the implementation, but may be switched off via a command line argument by users who are experiencing performance issues, or who would like an answer more quickly. In practice we have found that calling prime_divisors on every $M(\varepsilon)$ integer is not a significant bottleneck for number fields of degree at most 12.

7.3. ***The "no growth in minus part" method.*** Lemma A.2 in the Appendix of [Banwait 2023] gives a general method to conclude that $X_0(p)(K) = X_0(p)(\mathbb{Q})$ provided certain conditions are satisfied, the most crucial of which is that the Mordell–Weil group of the minus part $J_0(p)_-$ of the Jacobian $J_0(p)$ does not grow when base extending from $\mathbb{Q}$ to K. Indeed — while it is not needed in the current paper — the method there works mutatis mutandis replacing p with an arbitrary integer N.

Algorithmically checking this nongrowth condition for a general K is at present not entirely straightforward; however in this section we illustrate how it may be checked in Sage in the simpler case that $K/\mathbb{Q}$ is an abelian extension of prime degree, using modular symbols computations.

After constructing $J_0(p)_-$ as the kernel of $w_p + 1$ on the cuspidal subspace of modular symbols of level $\Gamma_0(p)$, one may count the size of the reduction $\widetilde{J_0(p)}_-(\mathbb{F}_\mathfrak{p})$ at various primes $\mathfrak{p}$ of good reduction. One takes the GCD of the resulting point counts, and compares it to the size of the $\mathbb{Q}$-torsion subgroup of $J_0(p)_-$. Since all of the $\mathbb{Q}$-torsion of $J_0(p)$ is contained in $J_0(p)_-$ by Chapter 3, Corollary 1.5 of [Mazur 1977], this latter size is simply the numerator of $((p-1)/12)$ by Theorem 1 of [loc. cit.]. If the two quantities agree, then one can conclude that the torsion does not grow in the extension, and proceed to the rank check; although if they do not agree, then it is not necessarily the case that the torsion gets larger. In this case, since we are not sure, we abandon the routine.

The rank of $J_0(p)_-(K)$ is equal to the rank of $J_0(p)_-(\mathbb{Q})$ plus the rank of the twist $J_0(p)^\chi_-(\mathbb{Q})$, where χ is the Dirichlet character corresponding to K; this is where the assumption that $K/\mathbb{Q}$ is abelian of prime degree is used. One is therefore reduced to checking whether or not the rank of the χ-twist is zero, which can be checked with a modular symbols computation involving the so-called *χ-twisted winding element*; see Section 2.2.2 of Bosman's PhD thesis [2008]; this can also be found as Section 6.3.3 of [Bosman 2011].[3]

We therefore obtain the following algorithm which, if it returns False, then we may remove the possible isogeny prime p from the final superset.

Algorithm 7.9. *Given an abelian extension $K/\mathbb{Q}$ of prime degree d, and a prime p, return* True *or* False *as follows*:

(1) *If the class number of $\mathbb{Q}(\sqrt{-p})$ is either 1 or d, return* True *and terminate.*

(2) *If the gonality of $X_0(p)$ is ≤ 2, return* True *and terminate.*

(3) *Compute χ, the Dirichlet character associated to K.*

(4) *If $d = 2$ and $\chi(p) = -1$, return* True *and terminate.*

[3] Johan Bosman was also a PhD student of Bas Edixhoven.

(5) *For rational primes $q \neq p$ up to a fixed bound B, compute the cardinality $|\widetilde{J_0(p)}_-(\mathbb{F}_\mathfrak{q})|$ for $\mathfrak{q} \mid q$ (which must have residue field degree either 1 or d), and take the GCD. If this does not equal the numerator of $((p-1)/12)$, return* True *and terminate.*

(6) *Compute the χ-twisted winding element $e(\chi)$ associated to the modular symbols space of weight 2 and level p.*

(7) *Compute the image of $e(\chi)$ under the rational period mapping associated to each of the factors in the Hecke decomposition of the cuspidal modular symbol space associated to $J_0(p)_-$. If any of these images are zero, return* True *and terminate.*

(8) *If not returned by this point, return* False.

The implementation of this algorithm may be found in the method works_method_of_appendix in sage_code/weeding.py.

7.4. ***Quadratic weeding.*** In the case that K is a quadratic field, another method that was used previously by the first named author in determining IsogPrimeDeg(K) was the *Özman sieve*. This relied on the work of Bruin and Najman [2015], Özman and Siksek [2019], and Box [2021] who determined the so-called exceptional quadratic points on small genus modular curves, together with earlier work of Özman in deciding on everywhere local solubility of twisted modular curves $X_0^d(N)$.

This method has now been fully automated in the current package. The relevant information on quadratic points on modular curves has been encoded into the file quadratic_points_catalogue.json. For each candidate isogeny prime p, if the exceptional quadratic points on $X_0(p)$ have been determined, and none are rational over K, then the function oezman_sieve(q,p) is called for all ramified primes q in K which furthermore are unramified in $\mathbb{Q}(\sqrt{-p})$; if any of these return False, then one has a local obstruction at $\mathbb{Q}_q$ to $\mathbb{Q}$-rational points on $X_0^d(p)$, where $K = \mathbb{Q}(\sqrt{d})$.

As the literature on the cataloguing of quadratic points on modular curves grows, these will be added to quadratic_points_catalogue.json, so over time our algorithm will improve. Having catalogues of all degree d exceptional points on modular curves will likewise be of great benefit.

One further weeding method employed in the case of a quadratic field K is similar to the Özman sieve, and is based on [Najman and Trbović 2022, Theorem 2.13]. This applies in the case that $X_0(p)$ is hyperelliptic, and lists the primes which must be unramified in any quadratic field K such that $X_0(p)$ admits a K-point which is not a $\mathbb{Q}$-point. These unramified primes have also been encoded into quadratic_points_catalogue.json. We refer to this method as the *Najman–Trbović filter*.

Putting these together we obtain the following algorithm which, if it returns False, then we may remove the possible isogeny prime p from the returned superset.

Algorithm 7.10. *Given a quadratic field $K = \mathbb{Q}(\sqrt{d})$ and a prime p, return* True *or* False *as follows*:

(1) *If the exceptional quadratic points on $X_0(p)$ have all been determined*:

  (a) *If $X_0(p)$ admits an exceptional K-point, return* True *and terminate.*

(b) *Compute the primes S which are ramified in K but unramified in $\mathbb{Q}(\sqrt{-p})$.*
(c) *If* oezman_sieve(q,p) *returns* False *for any $q \in S$, return* False *and terminate.*

(2) *If $X_0(p)$ is hyperelliptic of genus ≥ 2:*
(a) *If any prime divisor p of the discriminant of K is listed as being an unramified prime by* [Najman and Trbović 2022, Theorem 2.13], *return* False *and terminate.*

(3) *If not returned by this point, return* True.

The function apply_quadratic_weeding in sage_code/weeding.py implements this algorithm.

8. The combined algorithm

In this section we collect all of the previously mentioned algorithms into one; this is the main algorithm which the package *Isogeny Primes* implements, and serves as an overview of all of the other algorithms in the paper.

Algorithm 8.1. *Given a number field k of degree d, compute a finite set of primes S_k as follows*:

(1) *Initialise S_k to the empty set.*
(2) *Run Algorithm 3.26 to produce* MMIB(k). *Append the prime divisors of* MMIB(k) *to S_k.*
(3) *Run Algorithm 5.1 to produce* TypeOneBound(k). *Append the prime divisors of* TypeOneBound(k) *to S_k.*
(4) *Run Algorithm 6.4 to produce* TypeTwoPrimes(k). *Add to S_k the primes in* TypeTwoPrimes(k).
(5) *For each p in S_k, apply the following filters. If any return False, then remove p from S_k, and go to the next p*:
(a) *Apply the congruence conditions expressed in Table 8.*
(b) *Apply Isogeny character enumeration* (*Algorithm 7.7*).
(c) *If k is an abelian extension of $\mathbb{Q}$ of prime degree, then apply the "no growth in minus part" method* (*Algorithm 7.9*).
(d) *If k is a quadratic extension of $\mathbb{Q}$, then apply quadratic weeding* (*Algorithm 7.10*).
(6) *Return S_k.*

We recall the main result (Theorem 1.7 from the introduction) concerning the significance of the above algorithm for the determination of IsogPrimeDeg(k), and provide a summary proof.

Theorem 8.2. *Let k be a number field. Then Algorithm 8.1 outputs a finite set of primes S_k such that, if p is an isogeny prime for k whose associated isogeny character is not of Momose type* 3, *then, conditional on GRH, $p \in S_k$. In particular,*

(1) *if k does not contain the Hilbert class field of an imaginary quadratic field, then S_k contains* IsogPrimeDeg(k);
(2) *the above results are unconditional for the restricted set of semistable isogeny primes for k.*

Proof. That S_k is finite follows from the nonzeroness of the integers $\mathsf{MMIB}(k)$ (Theorem 1.4) and $\mathsf{TypeOneBound}(k)$ (Theorem 1.5), and the finiteness of the set $\mathsf{TypeTwoPrimes}(k)$. That it is (conditional upon GRH) a superset for the isogeny primes for k which are not of Momose type 3 follows from the following facts:

(1) $\mathsf{MMIB}(k)$ is a multiplicative bound on isogeny primes which are not of Momose type 1, 2 or 3 (Theorem 1.4).

(2) $\mathsf{TypeOneBound}(k)$ is a multiplicative bound on isogeny primes which are of Momose type 1 (Theorem 1.5).

(3) $\mathsf{TypeTwoPrimes}(k)$ is, conditional upon GRH, a superset for the isogeny primes which are of Momose type 2 (Theorem 1.6).

Item (1) follows from the definition of Momose type 3 isogenies (Definition 3.20), which requires k to contain the Hilbert class field of an imaginary quadratic field, and Item (2) follows from the observation Proposition 3.1 that semistable elliptic curves cannot possess isogenies of signature type 2, and hence a fortiori not of Momose type 2; we therefore unconditionally have that $\mathsf{TypeTwoPrimes}(k)$ is empty. □

9. Degree three points on $X_0(p)$

In this next section we will determine $\mathsf{IsogPrimeDeg}(K)$ for various cubic fields K, and prove Theorem 1.8. This requires a study of cubic points on the modular curves $X_0(p)$ for small primes p. Such a study has recently been carried out by Box, Gajović and Goodman [2023], who explicitly determine all cubic points on $X_0(p)$ for the primes $p = 53, 61, 67$ and 73. These primes are such that $J_0(p)$ has positive rank, and the authors of [loc. cit.] skip the relatively easy cases of where $J_0(p)$ is of rank 0. We therefore augment their work with the following two results: Theorem 9.1 which deals with the genus 2 cases, and Theorem 9.4, which deals with the higher genus cases. Note that for all p considered below, $X_0(p)$ is hyperelliptic.

Theorem 9.1. *Suppose $p = 23, 29$ or 31. Let K be a cubic field such that $X_0(p)$ admits a noncuspidal K-rational point. Then we have the following*:

(1) *The discriminant Δ_K of K is negative.*

(2) *There is a finite set of explicitly computable hyperelliptic curves over $\mathbb{Q}$ of genera 2 or 3 such that the quadratic twist at Δ_K of at least one of them admits a $\mathbb{Q}$-rational point.*

Proof. We use the usual strategy for studying degree 3 points by studying $X_0(p)^{(3)}(\mathbb{Q})$ and it's canonical map to $\mathrm{Pic}^3 X_0(p)(\mathbb{Q})$. For each of these 3 values of p the curve $X_0(p)$ is hyperelliptic with the Atkin–Lehner involution being the hyperelliptic involution. Additionally $J_0(p)$ is of rank 0. This implies that

$$J_0(p)(\mathbb{Q}) = J_0(p)(\mathbb{Q})_{\mathrm{tors}} = \mathbb{Z}/N\mathbb{Z}(0-\infty)$$

with $N = \mathrm{Numerator}((p-1)/12)$. In particular if a runs over all residue classes in $\mathbb{Z}/N\mathbb{Z}$ then $3\infty + a(0-\infty)$ runs over a complete set of representatives of $\mathrm{Pic}^3 X_0(p)(\mathbb{Q})$, meaning that every divisor of degree 3 is linearly equivalent to one of the form $3\infty + a(0-\infty)$.

Consider the associated Riemann–Roch spaces $\mathcal{L}(3\infty+a(0-\infty))$ for all $a\in\mathbb{Z}/N\mathbb{Z}$. Note that these are all 2-dimensional, since the genus of these curves is 2, so the Riemann–Roch theorem for a degree 3 divisor D gives $\dim H_0(X, D)=\dim H_0(X, D)-\dim H_0(X, K-D)=\deg D+1-g=2$. In particular, this shows that every effective divisor of degree 3 is contained in exactly one 1-dimensional family parametrised by $\mathbb{P}^1$.

By following the techniques in work of Derickx with Najman [2019, see particularly the proofs of Lemmas 4.9 and 4.11] one may explicitly compute defining equations for each of these finitely many families. Briefly, one constructs models $f(x,t)=0$ for $X_0(p)$ whose projections to the t-coordinate coincides with the degree 3 map to $\mathbb{P}^1$ induced by the global sections of $\mathcal{L}a(3\infty+a(0-\infty))$ for each $a\in\mathbb{Z}/N\mathbb{Z}$. The cubic field corresponding to a value of $t=b\in\mathbb{Q}$ is $K_b=\mathbb{Q}[x]/f(x,b)$. Hence the discriminant of K differs from the discriminant of f by a square in $\mathbb{Q}$. Write $g(t)=\Delta_x f(x,t)$, where Δ_x means taking the discriminant with respect to the variable x. Then in particular we have

$$\Delta(K_b)y^2=g(b) \tag{9-1}$$

for some value of y in $\mathbb{Q}$. We checked that $g(x)\leq 0$ proving that $\Delta(K_b)\leq 0$ and hence (1). For (2), we computed the genera of the curves $\mathcal{C}: y^2=g(x)$ and checked that they were all of genus 2 or 3. Since (9-1) gives the quadratic twist of $\mathcal{C}$, this completes the proof of (2). The computations described in this proof are carried out in magma_scripts/Genus2Cubic.m. □

Since totally real cubic fields must have positive discriminant, and cyclic cubic fields must always be totally real, we immediately obtain the following result.

Corollary 9.2. *Suppose $p=23$, 29 or 31, and let K be a totally real cubic field. Then $X_0(p)(K)$ consists only of the two $\mathbb{Q}$-rational cusps.*

Remark 9.3. The hyperelliptic curves constructed in the proof of Theorem 9.1 are computed in the main function of magma_scripts/Genus2Cubic.m. For $p=23$, 29 and 31, there are respectively 9, 5 and 3 such curves. For $p=31$, since there are only three of them, we list them here:

$$\mathcal{C}_1: y^2=-3x^8-8x^7+392x^6+992x^5-32976x^4-286336x^3-980352x^2-1560064x-963328.$$
$$\mathcal{C}_2: y^2=-3x^6+104x^5-1400x^4+9040x^3-27696x^2+34880x-22208.$$
$$\mathcal{C}_3: y^2=-108x^6+2552x^5-12276x^4-64944x^3-124436x^2-134728x-66188.$$

Theorem 9.4. *Suppose $p=41$, 47, 59 or 71; then $X_0(p)$ has 2, 2, 1 and 0 degree 3 points respectively. Furthermore the degree 3 number fields over which these points are defined are as follows*:

$p=41$: $\mathbb{Q}[x]/(x^3-x^2+x+2)$ *of discriminant* -139.

$p=47$: $\mathbb{Q}[x]/(x^3+x^2+2x+12)$ *of discriminant* -883.

$p=59$: $\mathbb{Q}[x]/(x^3-x^2-x+2)$ *of discriminant* -59.

Proof. The proof proceeds as in the proof of Theorem 9.1, although now it is no longer true that $\mathcal{L}(3\infty+a(0-\infty))$ has dimension 2 for all $a\in\mathbb{Z}/N\mathbb{Z}$.

Δ_K	f_K	LMFDB label	new isogeny primes
49	$x^3 - x^2 - 2x + 1$	3.3.49.1	–
148	$x^3 - x^2 - 3x + 1$	3.3.148.1	–
-2891	$x^3 - x^2 - 2x - 20$	3.1.2891.3	29

Table 9. Determination of $\mathsf{IsogPrimeDeg}(K)$ for some cubic number fields K.

We proceed via explicit Magma computation. For each value of p we found exactly two values of a where $\mathcal{L}(3\infty + a(0-\infty))$ was 2-dimensional; these two cases correspond to the 1-dimensional families of degree 3 divisor of the form $\infty + D$ and $0 + D$ where D is a divisor in the hyperelliptic class. Hence these two families don't correspond to degree 3 points.

In each of the remaining cases $\mathcal{L}(3\infty + a(0-\infty))$ was either 0- or 1-dimensional. In the 0-dimensional cases there is nothing to do. For the 1-dimensional cases we explicitly found the unique effective divisor of degree 3 linearly equivalent to $3\infty + a(0-\infty)$. In the cases this divisor was irreducible we explicitly computed a defining polynomial for the residue field of this divisor, this gives the classification. The code carrying out these explicit computations is found in magma_scripts/HigherGenusCubic.m. □

Corollary 9.5. *Let K be a cubic field of positive discriminant. If there exists an elliptic curve over K admitting a K-rational p-isogeny, for p not in* $\mathsf{IsogPrimeDeg}(\mathbb{Q})$, *then $p \geq 79$.*

Proof. We combine Theorems 9.1 and 9.4 with Theorem 1.1 of [Box et al. 2023], and check that the finitely many cubic fields given in Section 5.2 of [loc. cit.] all have negative discriminant. □

By combining the results of this section with the program *Isogeny Primes* explained in this paper, we determine $\mathsf{IsogPrimeDeg}(K)$ for some cubic fields K, thereby proving Theorem 1.8.

Theorem 9.6. *Assume GRH. For each number field in Table 9, the column "new isogeny primes" lists the isogeny primes for that number field which are not in* $\mathsf{IsogPrimeDeg}(\mathbb{Q})$.

Proof. Running *Isogeny primes* on the first two number fields in Table 9 shows that the largest possible isogeny prime is at most 73, so we conclude with Corollary 9.5. For the last cubic field K in Table 9, we run *isogeny primes* to obtain a superset as follows:

$$\mathsf{IsogPrimeDeg}(K) \subseteq \mathsf{IsogPrimeDeg}(\mathbb{Q}) \cup \{23, 29, 31, 73\}.$$

From [Box et al. 2023] we know that 73 is not an isogeny prime for this cubic field. A search for K-rational points on the modular curve $X_0(29)$ in Magma reveals that there are elliptic curves admitting a K-rational 29-isogeny. The primes 23 and 31 are ruled out with part (2) of Theorem 9.1. We computed the finite set of hyperelliptic curves, and showed that the quadratic twist by the squarefree part of -2891 (i.e., -59) of each of these curves was not everywhere locally soluble, thereby showing that none of them admit a $\mathbb{Q}$-point. □

Acknowledgments

We are grateful to Edgar Costa, David Roe and Andrew Sutherland for providing us access to computational resources of the Simons Collaboration in Arithmetic Geometry, Number Theory, and Computation, with which many of the computations have been performed. Our access to the Magma algebra system was made available by a generous initiative of the Simons Foundation. We would also like to thank Josha Box for useful conversations and providing some Magma code, as well as Nicolas Billerey, Pip Goodman, Filip Najman and Andrew Sutherland for comments and corrections to an earlier version of the manuscript. Banwait is grateful to Jennifer Balakrishnan and Boston University for hosting a research visit where this version of the manuscript was finalised. We are also extremely grateful to the anonymous referee for their careful and meticulous review of an earlier version of the manuscript.

During the final stages of this project we were deeply saddened to learn of the untimely passing of Bas Edixhoven, who was Derickx's PhD advisor. We were both touched by his support and care for more junior mathematicians, were inspired by his generous sharing of his time and ideas, and we wish to dedicate this paper in his memory.

References

[Bach and Sorenson 1996] E. Bach and J. Sorenson, "Explicit bounds for primes in residue classes", *Math. Comp.* **65**:216 (1996), 1717–1735. MR Zbl

[Banwait 2021] B. S. Banwait, "Quadratic isogeny primes, an algorithm to compute isogeny primes over given quadratic fields", 2021, available at https://github.com/barinderbanwait/quadratic_isogeny_primes.

[Banwait 2023] B. S. Banwait, "Explicit isogenies of prime degree over quadratic fields", *Int. Math. Res. Not.* **2023**:14 (2023), 11829–11876. MR Zbl

[Banwait and Derickx 2021] B. S. Banwait and M. Derickx, "Isogeny primes, an algorithm to compute isogeny primes over given number field", 2021, available at https://github.com/isogeny-primes/isogeny-primes.

[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

[Bosman 2008] J. G. Bosman, *Explicit computations with modular Galois representations*, Ph.D. thesis, Universiteit Lieden, 2008, available at https://hdl.handle.net/1887/13364.

[Bosman 2011] J. Bosman, "Computations with modular forms and Galois representations", pp. 129–157 in *Computational aspects of modular forms and Galois representations*, edited by B. Edixhoven and J.-M. Couveignes, Ann. of Math. Stud. **176**, Princeton Univ. Press, 2011. MR

[Box 2021] J. Box, "Quadratic points on modular curves with infinite Mordell–Weil group", *Math. Comp.* **90**:327 (2021), 321–343. MR Zbl

[Box et al. 2023] J. Box, S. Gajović, and P. Goodman, "Cubic and quartic points on modular curves using generalised symmetric Chabauty", *Int. Math. Res. Not.* **2023**:7 (2023), 5604–5659. MR Zbl

[Bruin and Najman 2015] P. Bruin and F. Najman, "Hyperelliptic modular curves $X_0(n)$ and isogenies of elliptic curves over quadratic fields", *LMS J. Comput. Math.* **18**:1 (2015), 578–602. MR Zbl

[David 2008] A. David, *Caractère d'isogénie et borne uniforme pour les homothéties*, Ph.D. thesis, Université Louis Pasteur, 2008, available at https://theses.hal.science/tel-00343355.

[David 2011a] A. David, "Borne uniforme pour les homothéties dans l'image de Galois associée aux courbes elliptiques", *J. Number Theory* **131**:11 (2011), 2175–2191. MR Zbl

[David 2011b] A. David, "Caractère d'isogénie et critères d'irréductibilité", preprint, 2011. arXiv 1103.3892

[Deligne and Rapoport 1973] P. Deligne and M. Rapoport, "Les schémas de modules de courbes elliptiques", pp. 143–316 in *Modular functions of one variable*, *II* (Antwerp, Belgium, 1972), edited by P. Deligne and W. Kuyk, Lecture Notes in Math. **349**, Springer, 1973. MR Zbl

[Derickx and Najman 2019] M. Derickx and F. Najman, "Torsion of elliptic curves over cyclic cubic fields", *Math. Comp.* **88**:319 (2019), 2443–2459. MR Zbl

[Derickx et al. 2023] M. Derickx, S. Kamienny, W. Stein, and M. Stoll, "Torsion points on elliptic curves over number fields of small degree", *Algebra Number Theory* **17**:2 (2023), 267–308. MR Zbl

[Freitas and Siksek 2015] N. Freitas and S. Siksek, "Criteria for irreducibility of mod p representations of Frey curves", *J. Théor. Nombres Bordeaux* **27**:1 (2015), 67–76. MR Zbl

[Kamienny 1992a] S. Kamienny, "Torsion points on elliptic curves and q-coefficients of modular forms", *Invent. Math.* **109**:2 (1992), 221–229. MR Zbl

[Kamienny 1992b] S. Kamienny, "Torsion points on elliptic curves over fields of higher degree", *Int. Math. Res. Not.* **1992**:6 (1992), 129–133. MR Zbl

[Larson and Vaintrob 2014] E. Larson and D. Vaintrob, "Determinants of subquotients of Galois representations associated with abelian varieties", *J. Inst. Math. Jussieu* **13**:3 (2014), 517–559. MR Zbl

[Mazur 1977] B. Mazur, "Modular curves and the Eisenstein ideal", *Inst. Hautes Études Sci. Publ. Math.* **47** (1977), 33–186. MR Zbl

[Mazur 1978] B. Mazur, "Rational isogenies of prime degree", *Invent. Math.* **44**:2 (1978), 129–162. MR Zbl

[Merel 1996] L. Merel, "Bornes pour la torsion des courbes elliptiques sur les corps de nombres", *Invent. Math.* **124**:1-3 (1996), 437–449. MR Zbl

[Momose 1995] F. Momose, "Isogenies of prime degree over number fields", *Compos. Math.* **97**:3 (1995), 329–348. MR Zbl

[Najman and Trbović 2022] F. Najman and A. Trbović, "Splitting of primes in number fields generated by points on some modular curves", *Res. Number Theory* **8**:2 (2022), art. id. 28. MR Zbl

[Neukirch 1999] J. Neukirch, *Algebraic number theory*, Grundl. Math. Wissen. **322**, Springer, 1999. MR Zbl

[Ozman 2012] E. Ozman, "Points on quadratic twists of $X_0(N)$", *Acta Arith.* **152**:4 (2012), 323–348. MR Zbl

[Ozman and Siksek 2019] E. Ozman and S. Siksek, "Quadratic points on modular curves", *Math. Comp.* **88**:319 (2019), 2461–2484. MR Zbl

[Parent 1999] P. Parent, "Bornes effectives pour la torsion des courbes elliptiques sur les corps de nombres", *J. Reine Angew. Math.* **506** (1999), 85–116. MR Zbl

[PARI/GP 2021] "PARI/GP", 2021, available at http://pari.math.u-bordeaux.fr. Version 2.14.0.

[SageMath 2021] *SageMath, the Sage Mathematics Software System*, 2021, available at http://www.sagemath.org. Version 9.4.

[Serre 1972] J.-P. Serre, "Propriétés galoisiennes des points d'ordre fini des courbes elliptiques", *Invent. Math.* **15**:4 (1972), 259–331. MR Zbl

[Serre and Tate 1968] J.-P. Serre and J. Tate, "Good reduction of abelian varieties", *Ann. of Math.* (2) **88** (1968), 492–517. MR Zbl

[Stoll 2017] M. Stoll, Magma programs related to [Derickx et al. 2023], 2017, available at http://www.mathe2.uni-bayreuth.de/stoll/magma/#DKSS.

[Waterhouse 1969] W. C. Waterhouse, "Abelian varieties over finite fields", *Ann. Sci. École Norm. Sup.* (4) **2** (1969), 521–560. MR Zbl

Communicated by Bjorn Poonen
Received 2022-10-19 Revised 2024-03-28 Accepted 2024-09-03

barinder.s.banwait@gmail.com *Department of Mathematics and Statistics, Boston University, Boston, MA, United States*

maarten@mderickx.nl *Department of Mathematics, University of Zagreb, Zagreb, Croatia*

ALGEBRA AND NUMBER THEORY 19:6 (2025)
https://doi.org/10.2140/ant.2025.19.1199

msp

Ideals in enveloping algebras of affine Kac–Moody algebras

Rekha Biswal and Susan J. Sierra

Let L be an affine Kac–Moody algebra, with central element c, and let $\lambda \in \mathbb{C}$. We study two-sided ideals in the central quotient $U_\lambda(L) := U(L)/(c-\lambda)$ of the universal enveloping algebra of L and prove:

(1) If $\lambda \neq 0$ then $U_\lambda(L)$ is simple.
(2) The algebra $U_0(L)$ has *just-infinite growth*, in the sense that any proper quotient has polynomial growth.

As an immediate corollary, we show that the annihilator of any nontrivial integrable highest-weight representation of L is centrally generated, extending a result of Chari for Verma modules.

We also show that universal enveloping algebras of loop algebras and current algebras of finite-dimensional simple Lie algebras have just-infinite growth, and prove similar results to the two results above for quotients of symmetric algebras of these Lie algebras by Poisson ideals.

1. Introduction

Fix an algebraically closed field $\Bbbk$ of characteristic 0. Let L be an affine Kac–Moody algebra over $\Bbbk$. This paper is concerned with the structure, and particularly the *size*, of two-sided ideals of the universal enveloping algebra $U(L)$; our main theorem shows that such ideals are extremely large, in a sense that we make precise below.

In the introduction, to simplify discussion, we assume that L is untwisted; that is, that there is a finite-dimensional simple Lie algebra $\mathfrak{g}$ so that, as a vector space,

$$L = \mathfrak{g}[t, t^{-1}] \oplus \Bbbk c \oplus \Bbbk d,$$

where c is central, d is the derivation measuring degree, and $\mathfrak{g}[t, t^{-1}]$ is the loop algebra of $\mathfrak{g}$. (We consider general affine algebras in the body of the paper.) The derived subalgebra L' of L is

$$L' = \mathfrak{g}[t, t^{-1}] \oplus \Bbbk c$$

and is the unique (up to isomorphism) nontrivial central extension of the loop algebra $\mathfrak{g}[t, t^{-1}]$. To emphasize the relationship between L and $\mathfrak{g}$, we sometimes write $L = \widehat{\mathfrak{g}}$.

MSC2020: primary 16P90, 16S30, 17B10, 17B67; secondary 16D30, 17B65.
Keywords: Kac–Moody algebra, affine algebra, highest-weight representation, Gelfand–Kirillov dimension, simple ring.

If $\lambda \in \Bbbk$, define

$$U_\lambda(L) = U(L)/(c-\lambda) \quad \text{and} \quad U_\lambda(L') = U(L')/(c-\lambda),$$

so

$$U(\mathfrak{g}[t,t^{-1}]) \cong U_0(L').$$

We study two-sided ideals in $U(L)$ and in the central quotients $U_\lambda(L)$. We will see that they are very big, in a precise sense. Our main result is:

Theorem 1.1. *Let $\lambda \in \Bbbk$:*

(0) *The algebras $U_\lambda(L)$ and $U_\lambda(L')$ have just-infinite growth. That is, let J be a nonzero ideal of $U_\lambda(L)$, and let J' be a nonzero ideal of $U_\lambda(L')$. Then $U_\lambda(L)/J$ and $U_\lambda(L')/J'$ have polynomial growth.*

(1) *In fact, if $\lambda \neq 0$ then $U_\lambda(L)$ and $U_\lambda(L')$ are simple rings.*

(2) *Any nonzero ideal of $U(L)$ or of $U(L')$ contains a nonzero element of $\Bbbk[c]$; equivalently, the algebras $U(L') \otimes_{\Bbbk[c]} \Bbbk(c)$ and $U(L) \otimes_{\Bbbk[c]} \Bbbk(c)$ are simple.*

Here recall that a $\Bbbk$-algebra R has *polynomial growth* if there is a polynomial $p(t) \in \mathbb{R}[t]$ so that, for every finite-dimensional subspace V of R, $\dim V^n \leq p(n)$ for all sufficiently large n. (For example, enveloping algebras of finite-dimensional Lie algebras have polynomial growth.) Further, R has *exponential growth* if there exists $V \subset R$ so that $\lim_{n\to\infty}(\dim V^n)^{1/n} > 1$ and (strictly) *subexponential growth* if R has neither exponential nor polynomial growth. It is well-known that $U_\lambda(L)$ and $U_\lambda(L')$ have subexponential growth. Thus Theorem 1.1(0) tells us that two-sided ideals in these algebras are extremely large: large enough to cut these very big algebras down to a reasonable size.

We also prove a similar theorem for Poisson ideals in the symmetric algebras of L and L'. Recall that the symmetric algebra $S(\mathfrak{k})$ of a Lie algebra $\mathfrak{k}$ is a Poisson algebra under the Kostant–Kirillov Poisson bracket $\{-,-\}$ induced by defining $\{x,y\} = [x,y]$ for $x, y \in \mathfrak{k}$. A *Poisson ideal* of a Poisson algebra R is an ideal of R which is also a Lie ideal for the Poisson bracket of R. We prove:

Theorem 1.2. *Let $\lambda \in \Bbbk$ and consider the Poisson algebras $S_\lambda(L) = S(L)/(c-\lambda)$ and $S_\lambda(L') = S(L')/(c-\lambda)$:*

(0) *Let I be a nonzero Poisson ideal of $S_\lambda(L)$ and let I' be a nonzero Poisson ideal of $S_\lambda(L')$. Then $S_\lambda(L)/I$ and $S_\lambda(L')/I'$ have polynomial growth.*

(1) *In fact, if $\lambda \neq 0$ then $S_\lambda(L)$ and $S_\lambda(L')$ are **Poisson simple** in the sense that they have no nontrivial Poisson ideals.*

(2) *Any nonzero Poisson ideal of $S(L)$ or of $S(L')$ contains a nonzero element of $\Bbbk[c]$; equivalently, $S(L') \otimes_{\Bbbk[c]} \Bbbk(c)$ and $S(L) \otimes_{\Bbbk[c]} \Bbbk(c)$ are Poisson simple.*

As an immediate consequence of Theorem 1.1 we compute the annihilators of a large class of representations of L, and in particular for *all* nontrivial integrable highest-weight representations of L; we show these annihilators are all centrally generated.

Theorem 1.3. *Let N be a nonzero representation of L so that $(c-\lambda)N = 0$ for some $\lambda \neq 0$. Then $\mathrm{Ann}_{U(L)}(N) = (c-\lambda)$. In particular, if M is a nontrivial integrable highest-weight representation of L with central character λ, then*

$$\mathrm{Ann}_{U(L)}(M) = (c-\lambda).$$

Proof. Theorem 1.1(1) shows that $(c-\lambda)$ is a maximal ideal of $U(L)$. The second statement is an immediate consequence, since nontrivial highest-weight representations of L have nontrivial central character. □

Theorem 1.3 extends Vyjayanthi Chari's well-known calculation of the annihilators of Verma modules for infinite-dimensional symmetrizable Kac–Moody algebras [Chari 1985]. To our knowledge, no annihilator of a (nontrivial) integrable highest-weight representation of an affine algebra has been known until now.

Our investigation was partially motivated by work of Natalia Iyudu and Sierra [2020], showing that the analogue of Theorem 1.1(0) holds for the Virasoro Lie algebra: central quotients of the enveloping algebra of the Virasoro algebra have just-infinite growth; see [Iyudu and Sierra 2020, Theorem 1.2]. Affine Kac–Moody algebras (more precisely, their derived subalgebras) and the Virasoro algebra are both central extensions of a graded simple Lie algebra of linear growth, and their representation theory is related through the Sugawara construction. It is thus natural to ask whether results for the Virasoro algebra can be extended to cover affine algebras. However, affine algebras are much more commutative than the Virasoro algebra, where the centralizers of elements are in general finite-dimensional (in fact two-dimensional); thus, naively, two-sided ideals in their enveloping algebras are smaller. It is surprising to find that the two-sided structure of their enveloping algebras is similarly constrained.

The phrasing of Theorem 1.1 appears redundant; of course, since $U_\lambda(L)$ is simple for $\lambda \neq 0$, it has just-infinite growth. However, it reflects the structure of the paper. We focus first on proving Theorems 1.1(0) and 1.2(0). Our proof strategy here is broadly similar to the methods of [Iyudu and Sierra 2020], but significantly more delicate because of the commutativity problem. Centralizers in L are large in general and so the adjoint action of L is more difficult to control. Note that to prove just-infinite growth for a (left and right) nonnoetherian algebra like $U_\lambda(L)$, it is helpful for as many commutators as possible to be nonzero in order to ensure that two-sided ideals are big.

To prove Theorems 1.1(0) and 1.2(0), we construct an ordered PBW basis for $U_\lambda(L)$ and then show, through analyzing the adjoint action of L on $U_\lambda(L)$, that if B is a sufficiently large element of this basis, then there is an element of J with leading term B. (See Lemma 5.4.) This allows us to reduce almost all basis elements, modulo J, to smaller elements and thus to bound the growth of $U_\lambda(L)/J$.

We work first with the symmetric algebra of the positive current algebra $t\mathfrak{g}[t]$ of $\mathfrak{g}$. This is the associated graded ring of $U(t\mathfrak{g}[t])$ under the natural (length) filtration and so has a Poisson algebra structure; further it is finitely graded under the grading induced by giving elements of $\mathfrak{g}$ degree 0 and t degree 1. We show that if I is a nontrivial Poisson ideal of $S(t\mathfrak{g}[t])$, then we can reduce almost all monomials in an ordered PBW basis of $S(t\mathfrak{g}[t])$ to smaller monomials, modulo I. This reduction result allows us to

prove (Theorem 3.9) that $U(\mathfrak{g}[t])$ has just-infinite growth. We then extend our analysis to $U(\mathfrak{g}[t,t^{-1}])$, to general $U_\lambda(L')$, and to $U_\lambda(L)$.

A corollary of Theorem 1.1(0) is the following somewhat counterintuitive result:

Proposition 1.4 (Proposition 4.10, Corollary 7.2). *Let $\mathfrak{g}$ be a finite-dimensional simple Lie algebra, let $U(\mathfrak{g}[t,t^{-1}])$ be the enveloping algebra of the loop algebra of $\mathfrak{g}$, and let J be a nontrivial ideal of $U(\mathfrak{g}[t,t^{-1}])$. If $X \subseteq \mathbb{Z}$ is any infinite set (for example, X consists of all the primes), and $g \in \mathfrak{g} \setminus \{0\}$, then J contains a (nonzero) element involving only Lie algebra elements of the form gt^x for $x \in X$.*

In particular, if L is the (untwisted) affine algebra associated to $\mathfrak{g}$ and J is a nonzero ideal of $U(L)$, then $J \cap U(\mathfrak{g}[t]) \neq (0)$.

Proposition 1.4 allows us to conclude that if h is a semisimple element of $\mathfrak{g}$, then any nontrivial ideal of $U_\lambda(L)$ meets the subalgebra $\mathbb{A}_\lambda$ generated by $\{ht^i : i \in \mathbb{Z} \setminus \{0\}\}$ nontrivially. If $\lambda \neq 0$ then $\mathbb{A}_\lambda$ is isomorphic to an infinite Weyl algebra and is thus simple, and simplicity of $U_\lambda(L)$ follows immediately. Similar techniques give Theorem 1.1(2) and parts (1), (2) of Theorem 1.2.

To conclude the Introduction, we briefly describe the structure of the paper. Section 2 establishes notation and basic facts about affine Lie algebras and loop algebras, as well as the basics of growth and GK-dimension. Key results on Poisson ideals of symmetric algebras of positive current algebras are proved in Section 3. In Section 4 we prove preparatory results needed to extend our methods from $U(t\mathfrak{g}[t])$ to $U_\lambda(L)$ before proving Theorems 1.1(0) and 1.2(0) in Section 5. In Section 6 we prove the remaining parts of Theorems 1.1 and 1.2. Section 7 gives several other applications of Theorem 1.1, including Proposition 1.4.

2. Bases and basic facts

If L is an affine Kac–Moody algebra with central element c then there is a finite-dimensional simple Lie algebra $\mathfrak{g}$ and $\sigma \in \mathrm{Aut}(\mathfrak{g})$ so that $L'/(c)$ is isomorphic to the (possibly twisted) loop algebra $\mathfrak{g}[t,t^{-1}]^\sigma$. In this section we establish notation and basic facts about loop algebras, and further establish ordered bases for loop algebras and for their enveloping/symmetric algebras, which we will use in the next section. We also recall some background on growth and Gelfand–Kirillov dimension of algebras and modules.

We adopt the following notation: Let $\mathfrak{g}$ be a finite dimensional simple Lie algebra. Let $\sigma \in \mathrm{Aut}(\mathfrak{g})$ which (without loss of generality) we assume comes from an automorphism of the Dynkin diagram of $\mathfrak{g}$. We also let σ denote the diagram automorphism and the associated automorphism of the root system of $\mathfrak{g}$. Let r be the order of σ. Let η be a primitive r-th root of unity. We extend the action of σ to $\mathfrak{g}[t,t^{-1}]$ by defining $\sigma(gt^s) = \sigma(g)\eta^{-s}t^s$, and denote the corresponding twisted current algebra by $\mathfrak{g}[t]^\sigma$ and the twisted loop algebra by $\mathfrak{g}[t,t^{-1}]^\sigma$.

2.1. ***Basic facts about $\mathfrak{g}$.*** In this subsection we establish notation and basic facts about $\mathfrak{g}$. Fix a Cartan subalgebra $\mathfrak{h}$ of $\mathfrak{g}$. If $\alpha \in \mathfrak{h}^*$, we denote the α-eigenspace of $\mathfrak{g}$ by $\mathfrak{g}_\alpha$.

If $s \in \mathbb{Z}$, let $\bar{s}$ denote the congruence class of s in $\mathbb{Z}/r\mathbb{Z}$. The automorphism σ induces a $\mathbb{Z}/r\mathbb{Z}$-grading $\mathfrak{g} = \bigoplus_{s=0}^{r-1} \mathfrak{g}_{\bar{s}}$ of $\mathfrak{g}$, where $\mathfrak{g}_{\bar{s}} = \{g \in \mathfrak{g} : \sigma(g) = \eta^s g\}$. If $g \in \mathfrak{g}_{\bar{s}}$, we write $|g| = \bar{s}$. The $\mathfrak{g}_{\bar{s}}$ for $0 \le s \le r-1$ are irreducible $\mathfrak{g}_{\bar{0}}$-modules under the adjoint action.

Let $\mathfrak{h}_{\bar{0}} = \mathfrak{g}_{\bar{0}} \cap \mathfrak{h}$ be the Cartan subalgebra of $\mathfrak{g}_{\bar{0}}$. As in [Kac 1990, page 130], for $\bar{s} \in \mathbb{Z}/r\mathbb{Z}$ let $\Delta_{\bar{s}}$ be the set of nonzero weights of $\mathfrak{h}_{\bar{0}}$ on $\mathfrak{g}_{\bar{s}}$ and define the weight space decomposition

$$\mathfrak{g}_{\bar{s}} = \bigoplus_{\alpha \in \Delta_{\bar{s}} \cup \{0\}} \mathfrak{g}_{\bar{s},\alpha}.$$

We say an element $g \in \mathfrak{g}$ is *weight-homogeneous* if g is in some $\mathfrak{g}_{\bar{s},\alpha}$. Any weight-homogeneous element is by definition σ-equivariant. It is clear that $[\mathfrak{g}_{\bar{s},\alpha}, \mathfrak{g}_{\bar{p},\beta}] \subseteq \mathfrak{g}_{\bar{s}+\bar{p},\alpha+\beta}$.

We observe the following basic facts. As these are standard, most are stated without proof:

(0) We have $\mathfrak{g}[t, t^{-1}]^\sigma = \bigoplus_{s \in \mathbb{Z}} \mathfrak{g}_{\bar{s}} t^s$.

(1) The automorphism σ preserves the set of simple roots of $\mathfrak{g}$ and so the height of elements of $\mathfrak{g}$. In particular, the positive Borel subalgebra $\mathfrak{b}^+$ of $\mathfrak{g}$ and the nilpotent radical $\mathfrak{n}^+$ of $\mathfrak{b}^+$ are σ-invariant and thus decompose as sums

$$\mathfrak{n}^+ = \bigoplus_{s=0}^{r-1} \mathfrak{n}^+_{\bar{s}} \quad \text{and} \quad \mathfrak{b}^+ = \bigoplus_{s=0}^{r-1} \mathfrak{b}^+_{\bar{s}}$$

of σ-weight spaces.

(2) For each $\bar{s} \in \mathbb{Z}/r\mathbb{Z}$ and each $\alpha \in \Delta_{\bar{s}}$ we either have $\mathfrak{g}_{\bar{s},\alpha} \subseteq \mathfrak{n}^+$ or $\mathfrak{g}_{\bar{s},\alpha} \subseteq \mathfrak{n}^-$. Thus we can define

$$\Delta^+_{\bar{s}} = \{\alpha \in \Delta_{\bar{s}} : \mathfrak{g}_{\bar{s},\alpha} \subseteq \mathfrak{n}^+\}.$$

Let $\Delta^+ = \{(\bar{s}, \alpha) : \alpha \in \Delta^+_{\bar{s}}\}$. Similarly, define $\Delta^-_{\bar{s}}$ and Δ^-. Let $\Delta = \Delta^+ \cup \Delta^- = \{(\bar{s}, \alpha) : \alpha \in \Delta_{\bar{s}}\}$.

(3) Let θ be the longest root of $\mathfrak{g}$. There are $(\bar{s}, \alpha) \in \Delta^+$ so that $\mathfrak{g}_\theta = \mathfrak{g}_{\bar{s},\alpha}$.

(4) For $\alpha \neq 0$ the spaces $\mathfrak{g}_{\bar{s},\alpha}$ are one-dimensional, so we may choose a generator $g_{\bar{s},\alpha}$. Also, if γ is a positive root of $\mathfrak{g}$, let e_γ, f_γ be the Chevalley generators, respectively, of $\mathfrak{g}_\gamma$ and $\mathfrak{g}_{-\gamma}$. Let $h_\gamma = [e_\gamma, f_\gamma]$.

(5) If $(\bar{p}, \beta) \in \Delta^\pm$, then there are $(\bar{s}_1, \gamma_1), \ldots, (\bar{s}_k, \gamma_k) \in \Delta^\mp$ so that

$$[\mathfrak{g}_{\bar{s}_k,\gamma_k}, [\ldots, [\mathfrak{g}_{\bar{s}_1,\gamma_1}, \mathfrak{g}_{\pm\theta}], \ldots]] = \mathfrak{g}_{\bar{p},\beta}.$$

(6) The centralizer of $\mathfrak{h}_{\bar{0}}$ in $\mathfrak{g}$ is a Cartan subalgebra of $\mathfrak{g}$ and is thus equal to $\mathfrak{h}$, and so

$$\mathfrak{h} = \mathfrak{g}_{\bar{0},0} \oplus \cdots \oplus \mathfrak{g}_{\overline{r-1},0},$$

and $\mathfrak{h}_{\bar{0}} = \mathfrak{g}_{\bar{0},0}$.

(7) If $(\bar{s}, \alpha) \in \Delta^+$, then $(\overline{-s}, -\alpha) \in \Delta^-$ and $[\mathfrak{g}_{\bar{s},\alpha}, \mathfrak{g}_{\overline{-s},-\alpha}] \subseteq \mathfrak{h}_{\bar{0}}$. Let κ be the Killing form of $\mathfrak{g}$, which restricts to a nondegenerate bilinear form on $\mathfrak{h}_{\bar{0}}$, and let $\nu : \mathfrak{h}_{\bar{0}} \to \mathfrak{h}_{\bar{0}}^*$ be the induced isomorphism. Then $[\mathfrak{g}_{\bar{s},\alpha}, \mathfrak{g}_{-\bar{s},-\alpha}] = \Bbbk \nu^{-1}(\alpha)$. Further, there is a positive root γ of $\mathfrak{g}$ so that

$$\mathfrak{g}_{\bar{s},\alpha} \subseteq \sum_{i=0}^{r-1} \mathfrak{g}_{\sigma^i(\gamma)}.$$

r	$\mathfrak{g}$	$\mathfrak{h}_{\bar{0}}$	$\mathfrak{h}_{\bar{1}}$
2	$A_{2\ell}, \ell \geq 1$	$\{h_i + h_{2\ell-i+1} : 1 \leq i \leq \ell\}$	$\{h_i - h_{2\ell-i+1} : 1 \leq i \leq \ell\}$
2	$A_{2\ell-1}, \ell \geq 1$	$\{h_i + h_{2\ell-i} : 1 \leq i \leq \ell-1, h_\ell\}$	$\{h_i - h_{2\ell-i} : 1 \leq i \leq \ell-1\}$
2	$D_{\ell+1}, \ell \geq 3$	$\{h_i : 1 \leq i \leq \ell-1, h_\ell + h_{\ell+1}\}$	$h_\ell - h_{\ell+1}$
2	E_6	$h_1 + h_5, h_2 + h_4, h_3, h_6$	$h_1 - h_5, h_2 - h_4$
3	D_4	$h_1 + h_3 + h_4, h_2$	$h_1 + \eta h_3 + \eta^2 h_4$

r	$\mathfrak{h}_{\bar{2}}$
2	0
3	$h_1 + \eta^2 h_3 + \eta h_4$

Table 1. Basis for Cartan subalgebra.

Let r' be the σ-order of γ. Then $(g_{\bar{s},\alpha}, g_{\overline{-s},-\alpha})$ is, up to nonzero scalar multiple,

$$\begin{cases} (e_\gamma + \eta e_{\sigma(\gamma)} + \eta^2 e_{\sigma^2(\gamma)},\ f_\gamma + \eta^2 f_{\sigma(\gamma)} + \eta f_{\sigma^2(\gamma)}) & \text{if } r' = 3 \text{ and } \bar{s} = \bar{1}, \\ (e_\gamma + \eta^2 e_{\sigma(\gamma)} + \eta e_{\sigma^2(\gamma)},\ f_\gamma + \eta f_{\sigma(\gamma)} + \eta^2 f_{\sigma^2(\gamma)}) & \text{if } r' = 3 \text{ and } \bar{s} = \bar{2}, \\ (e_\gamma + e_{\sigma(\gamma)} + e_{\sigma^2(\gamma)},\ f_\gamma + f_{\sigma(\gamma)} + f_{\sigma^2(\gamma)}) & \text{if } r' = 3 \text{ and } \bar{s} = \bar{0}, \\ (e_\gamma + e_{\sigma(\gamma)},\ f_\gamma + f_{\sigma(\gamma)}) & \text{if } r' = 2 \text{ and } \bar{s} = \bar{0}, \\ (e_\gamma - e_{\sigma(\gamma)},\ f_\gamma - f_{\sigma(\gamma)}) & \text{if } r' = 2 \text{ and } \bar{s} = \bar{1}, \\ (e_\gamma,\ f_\gamma) & \text{if } r' = 1. \end{cases}$$

Also,

$$[g_{\bar{s},\alpha}, g_{\overline{-s},-\alpha}] = \begin{cases} h_\gamma + h_{\sigma(\gamma)} + h_{\sigma^2(\gamma)} & \text{if } r' = 3, \\ h_\gamma + h_{\sigma(\gamma)} & \text{if } r' = 2, \\ h_\gamma & \text{if } r' = 1. \end{cases}$$

Finally, up to scalars $\{g_{\bar{s},\alpha}, g_{\overline{-s},-\alpha}, [g_{\bar{s},\alpha}, g_{\overline{-s},-\alpha}]\}$ forms an $\mathfrak{sl}_2$-triple.

As σ preserves the height of elements, each $g_{\bar{s},\alpha}$ has a well-defined height in terms of the root system of $\mathfrak{g}$.

(8) When $\sigma \neq \mathrm{id}$ (that is, when $r = 2, 3$) we give notation for a σ-equivariant basis of $\mathfrak{h}$ in Table 1. In this case $\mathfrak{g}$ must be type A, D, or E. Let the simple roots of $\mathfrak{g}$ be $\alpha_1, \alpha_2, \ldots$ and let $h_i = [e_{\alpha_i}, f_{\alpha_i}]$ for $1 \leq i \leq \dim \mathfrak{h}$. We number the simple roots of $\mathfrak{g}$ as in [Kac 1990, Table Fin, page 53].

(9) Let $\mathcal{B}_\Delta = \{g_{\bar{s},\alpha} : (\bar{s}, \alpha) \in \Delta\}$ and let

$$h_{\bar{s}}^i = \sum_{j=0}^{r'} \eta^{sj} h_{\sigma^j(i)}.$$

(Here r' is the σ-order of the simple root α_i.) Then

$$\mathcal{B} = \mathcal{B}_\Delta \cup \{h_{\bar{s}}^i\}$$

is a σ-equivariant basis of $\mathfrak{g}$. If $0 \leq s \leq r-1$, let $\mathcal{B}_{\bar{s}} = \{x \in \mathcal{B} : |x| = s\}$.

2.2. ***Two monomial orders on $S(\mathfrak{g}[t, t^{-1}]^\sigma)$.*** In this section, we define PBW bases of $S(\mathfrak{g}[t, t^{-1}]^\sigma)$ and $U(\mathfrak{g}[t, t^{-1}]^\sigma)$. We will refer to elements of these bases as *monomials*, and we will define two different orders on monomials. The interplay between these monomial orders is key to proving our reduction results.

We first extend the σ-equivariant basis $\mathcal{B}$ of $\mathfrak{g}$ to a basis of $\mathfrak{g}[t, t^{-1}]^\sigma$. Let

$$\mathcal{B}[t, t^{-1}]^\sigma = \{gt^{rn+|g|} : g \in \mathcal{B}, n \in \mathbb{Z}\},$$

which is a basis of $\mathfrak{g}[t, t^{-1}]^\sigma$. Let

$$\mathcal{B}[t]^\sigma = \mathcal{B}[t, t^{-1}]^\sigma \cap \mathfrak{g}[t]^\sigma.$$

We will need an order on $\mathcal{B}[t, t^{-1}]^\sigma$, which we define as follows.

We first give an order on the roots of $\mathfrak{g}$. Let R be the set of roots of $\mathfrak{g}$ and let R^+, R^- be respectively the sets of positive and negative roots. Let $\{\alpha_i : i \in \{1, \dots, n\}\}$ be the set of simple roots of $\mathfrak{g}$. If $\alpha = \sum b_i\alpha_i$ is a positive root, then we define the height of α to be $\operatorname{ht}\alpha = \sum b_i$. Recall that we denote the highest root of $\mathfrak{g}$ by θ.

We will define a total order $<$ on R by:

- $\alpha < \beta$ if $\operatorname{ht}\alpha < \operatorname{ht}\beta$.
- If $\alpha = \sum b_i\alpha_i$, $\beta = \sum c_i\alpha_i$ and $\operatorname{ht}\alpha = \operatorname{ht}\beta$, then $\alpha < \beta$ if the tuple $(b_1, \dots, b_n)$ is less than the tuple $(c_1, \dots, c_n)$ in lexicographic order.

For each $\beta \in R$ let $g_\beta = e_\beta$ if $\beta \in R^+$ and $g_\beta = f_{-\beta}$ if $\beta \in R^-$. Then

$$\mathcal{C} = \{g_\beta : \beta \in R\} \cup \{h_{\alpha_i} : 1 \le i \le n\}$$

is a basis for $\mathfrak{g}$. Let $(\beta_1, \dots, \beta_\ell)$ be an enumeration of R^+ written in increasing order w.r.t. the above total order (so $\beta_\ell = \theta$). We define the following total order on $\mathcal{C}$:

$$g_{-\beta_\ell} < \cdots < g_{-\beta_1} < h_{\alpha_1} < \cdots < h_{\alpha_n} < g_{\beta_1} < \cdots < g_{\beta_\ell}.$$

Given $g \in \mathfrak{g}$, define the *$\mathcal{C}$-leading term* of g, which we denote by $\operatorname{LT}_\mathcal{C}(g)$, to be the largest element of $\mathcal{C}$ occurring in g with nonzero coefficient. (The reason for the subscript $\mathcal{C}$ in the notation is to distinguish this from other uses of the terminology "leading term" in this paper.)

The ordering defined above on $\mathcal{C}$ has the property that if $x, y, z \in \mathcal{C}$ with $[x, y] \ne 0 \ne [x, z]$, then

$$y < z \quad \Longleftrightarrow \quad \operatorname{LT}_\mathcal{C}[x, y] < \operatorname{LT}_\mathcal{C}[x, z]. \tag{2.1}$$

We now induce orderings on each basis $\mathcal{B}_{\bar{s}}$ of $\mathfrak{g}_{\bar{s}}$ from our order on $\mathcal{C}$. Let $0 \le s \le r-1$. If $x, y \in \mathcal{B}_{\bar{s}}$, define

$$x < y \quad \Longleftrightarrow \quad \operatorname{LT}_\mathcal{C}(x) < \operatorname{LT}_\mathcal{C}(y).$$

(We note here that it can be seen by inspection that, since every element of $\mathcal{B}_{\bar{s}}$ is associated to a unique σ-orbit in $\mathcal{C}$, if $\operatorname{LT}_\mathcal{C}(x) = \operatorname{LT}_\mathcal{C}(y)$ then $x = y$.) For $g \in \mathfrak{g}_{\bar{s}}$, we define $\operatorname{LT}_\mathcal{B}(g)$ to be the largest element of $\mathcal{B}$ occurring in g with nonzero coefficient.

It follows from (2.1) that if $x \in \mathcal{B}_{\bar{s}}$, $y \in \mathfrak{g}_{\bar{s}}$ with $\operatorname{LT}_\mathcal{B}(y) < x$, and z is weight-homogeneous with $[z, x] \ne 0 \ne [z, y]$, then

$$\operatorname{LT}_\mathcal{B}[z, y] < \operatorname{LT}_\mathcal{B}[z, x]. \tag{2.2}$$

As an example, we give the orderings on the $\mathcal{B}_{\bar{s}}$ in the case $\mathfrak{g} = A_2$, $|\sigma| = 2$. We have

$$\mathcal{B}_{\bar{0}} : g_{-\alpha_1} < g_{-\alpha_2} < h_1 + h_2 < g_{\alpha_2} < g_{\alpha_1} \quad \text{and} \quad \mathcal{B}_{\bar{1}} : g_{-\theta} < h_1 - h_2 < g_{\theta}.$$

Using the total order $<$ on the $\mathcal{B}_{\bar{s}}$, we define a total order on the basis $\mathcal{B}[t, t^{-1}]^{\sigma}$ of $\mathfrak{g}[t, t^{-1}]^{\sigma}$ as follows: If $xt^a, yt^b \in \mathcal{B}[t, t^{-1}]^{\sigma}$, then $xt^a < yt^b$ if and only if

- $a < b$; or
- $a = b$, and $x < y$ in the total order defined on $\mathcal{B}_{\bar{a}}$.

We now define a basis of normal words or standard monomials for $S(\mathfrak{g}[t, t^{-1}]^{\sigma})$ and $U(\mathfrak{g}[t, t^{-1}]^{\sigma})$. We will refer to elements of $\mathcal{B}[t, t^{-1}]^{\sigma}$ as *letters*, as "words" (or monomials) in these letters will span the algebras of interest. A *standard monomial* or *normally ordered monomial* in the letters $\mathcal{B}[t, t^{-1}]^{\sigma}$ is an expression of the form

$$M = g_1 t^{rm_1 + |g_1|} \cdots g_k t^{rm_k + |g_k|}, \tag{2.3}$$

where the $g_i \in \mathcal{B}$, and $g_1 t^{rm_1+|g_1|} \leq \cdots \leq g_k t^{rm_k+|g_k|}$. We sometimes write a standard monomial as

$$M = g_1 t^{n_1} \cdots g_k t^{n_k},$$

and when we do so we assume that $\overline{n_a} = |g_a|$ for all a: in other words, that each $g_i t^{n_i} \in \mathfrak{g}[t, t^{-1}]^{\sigma}$, so $M \in S(\mathfrak{g}[t, t^{-1}]^{\sigma})$. We say that k is the *length* of M, which we denote by $\operatorname{len} M$, and that the total t-power $\sum_a n_a$ is the *degree* of M, which we denote by $\deg(M)$.

We introduce two total orderings on the set of standard monomials. For two standard monomials M_1 and M_2, we write $M_1 < M_2$ if

- $\operatorname{len} M_1 < \operatorname{len} M_2$ or
- $\operatorname{len} M_1 = \operatorname{len} M_2$ and $\deg M_1 < \deg M_2$ or
- $\operatorname{len} M_1 = \operatorname{len} M_2$, $\deg M_1 = \deg M_2$, and M_1 is less than M_2 with respect to the *left to right* lexicographic order when both M_1 and M_2 are written in increasing (that is, normal) order.

Similarly, we write $M_1 \prec M_2$ if

- $\operatorname{len} M_1 < \operatorname{len} M_2$ or
- $\operatorname{len} M_1 = \operatorname{len} M_2$ and $\deg M_1 < \deg M_2$ or
- $\operatorname{len} M_1 = \operatorname{len} M_2$, $\deg M_1 = \deg M_2$, and M_1 is less than M_2 with respect to the *right to left* lexicographic order when both M_1 and M_2 are written in increasing order.

By the PBW theorem, both $U(\mathfrak{g}[t, t^{-1}]^{\sigma})$ and $S(\mathfrak{g}[t, t^{-1}]^{\sigma})$ have a basis of standard monomials. Given nonzero $F \in U(\mathfrak{g}[t, t^{-1}]^{\sigma})$ or $F \in S(\mathfrak{g}[t, t^{-1}]^{\sigma})$, we define $\operatorname{LT}_{<} F$, respectively $\operatorname{LT}_{\prec} F$, to be the $<$-largest, respectively $\prec$-largest, standard monomial occurring in F with nonzero coefficient when F is written as a linear combination of standard monomials.

2.3. ***Growth.*** In this short subsection we recall some definitions and basic facts about the growth of algebras and modules. Let R be a finitely generated associative $\Bbbk$-algebra and let M be a finitely generated representation of R. Then R has *polynomial growth* if there is a polynomial $p(t) \in \mathbb{R}[t]$ so that, for every finite-dimensional subspace V of R, $\dim V^n \leq p(n)$ for all sufficiently large n. Further, R has *exponential growth* if there exists $V \subset R$ so that $\lim_{n\to\infty}(\dim V^n)^{1/n} > 1$ (note that this limit always exists) and (strictly) *subexponential growth* if R has neither exponential or polynomial growth. The growth of M is defined similarly; here we let $V \subset R$ and $W \subset M$ be finite-dimensional vector spaces and consider $\dim V^n W$ as $n \to \infty$. If $\mathfrak{k}$ is an infinite-dimensional Lie algebra of polynomial growth (for example, an affine or loop algebra) then $U(\mathfrak{k})$ has subexponential growth by [Smith 1976].

The idea of growth may be refined through considering the *Gelfand–Kirillov dimension* or GKdim of R and of M. Here we define

$$\operatorname{GKdim} R = \sup_V \overline{\lim} \log_n \dim V^n,$$

where the supremum is taken over all finite-dimensional subspaces V of R. Likewise,

$$\operatorname{GKdim} M = \sup_{V,W} \overline{\lim} \log_n \dim V^n W,$$

where the supremum is taken over all finite-dimensional subspaces V of R and W of M. If R is a finitely generated $\Bbbk$-algebra and M is a finitely generated R-module, let V be any finite-dimensional generating subspace of R that contains 1, and let W be any finite-dimensional generating subspace of M. Then

$$\operatorname{GKdim} R = \overline{\lim} \log_n \dim V^n, \quad \operatorname{GKdim} M = \overline{\lim} \log_n \dim V^n W,$$

and does not depend on the choice of V or W.

Note that R (respectively, M) has polynomial growth if and only if $\operatorname{GKdim} R < \infty$ (respectively, $\operatorname{GKdim} M < \infty$). We say that R has *just-infinite growth* (equivalently, *just-infinite GK-dimension*) if $\operatorname{GKdim} R = \infty$ but for any nontrivial ideal J of R, then $\operatorname{GKdim} R/J < \infty$. For more background about growth and GK-dimension, we refer the reader to [Krause and Lenagan 1985].

If $\mathfrak{k}$ is a Lie algebra, then $\operatorname{GKdim} U(\mathfrak{k}) = \dim \mathfrak{k}$. Further, by [Smith 1976], if L is an affine Kac–Moody algebra, then the central quotients $U_\lambda(L)$ and $U_\lambda(L')$ have intermediate growth.

We will repeatedly use the following basic fact about growth of algebras.

Scholium 2.4. *Let K be a field and let $A \subseteq B$ be K-algebras. Suppose that* $\operatorname{GKdim} A = \infty$ *and that J is an ideal of B such that B/J has polynomial growth. Then $J \cap A \neq (0)$.*

Proof. This follows directly from the fact that GK-dimension does not increase on subalgebras [Krause and Lenagan 1985, Lemma 3.1], so we cannot have $A \hookrightarrow B/J$. □

3. A reduction result for symmetric algebras of current algebras

Let $\mathfrak{g}$ be a finite-dimensional simple Lie algebra and let $\sigma \in \operatorname{Aut}(\mathfrak{g})$. We adopt the notation of Section 2; in particular, let r be the order of σ. The *positive twisted current algebra* of $\mathfrak{g}$ is the subalgebra $(t\mathfrak{g}[t])^\sigma$

of the (twisted) loop algebra $\mathfrak{g}[t, t^{-1}]^\sigma$. In this section, we consider the symmetric algebra of the positive twisted current algebra of $\mathfrak{g}$ and show that it has just-infinite growth as a Poisson algebra: any factor by a proper Poisson ideal has polynomial growth. We will generalize this result in later sections to prove Theorems 1.1 and 1.2.

The symmetric algebra $S(\mathfrak{g}[t, t^{-1}]^\sigma)$ is a Lie algebra under the Poisson bracket $\{-, -\}$ induced from the Lie bracket on $\mathfrak{g}[t, t^{-1}]^\sigma$. That is, for $x, y \in \mathfrak{g}[t, t^{-1}]^\sigma$ we have $\{x, y\} = [x, y]$, and $\{-, -\}$ is anticommutative and a derivation in each input.

A *Poisson ideal* of $S(\mathfrak{g}[t, t^{-1}]^\sigma)$ is an ideal of the underlying commutative algebra which is also a Lie ideal for the Poisson bracket. Note that a Lie ideal of $S(\mathfrak{g}[t, t^{-1}]^\sigma)$ is also a $\mathfrak{g}[t, t^{-1}]^\sigma$-subrepresentation, where $\mathfrak{g}[t, t^{-1}]^\sigma$ acts on the symmetric algebra by the adjoint action $\operatorname{ad} gt^a = \{gt^a, -\}$. Each $S^m(\mathfrak{g}[t, t^{-1}]^\sigma)$ is a subrepresentation of $S(\mathfrak{g}[t, t^{-1}]^\sigma)$ under the action of $\mathfrak{g}[t, t^{-1}]^\sigma$, although of course not a Poisson ideal.

Thus if I is a Poisson ideal of $S(\mathfrak{g}[t, t^{-1}]^\sigma)$, each $I \cap S^m(\mathfrak{g}[t, t^{-1}]^\sigma)$ is a $\mathfrak{g}[t, t^{-1}]^\sigma$-subrepresentation of $S(\mathfrak{g}[t, t^{-1}]^\sigma)$. We will spend the bulk of this section analyzing the structure of such subrepresentations. In fact, because we want our results to apply to Poisson ideals of current algebras, we will consider the structure of $(t\mathfrak{g}[t])^\sigma$-subrepresentations of the m-th symmetric power $S^m(\mathfrak{g}[t, t^{-1}]^\sigma)$.

Our key technical result is the following:

Proposition 3.1. *Fix $m \in \mathbb{N}$, and let I be any nonzero $(t\mathfrak{g}[t])^\sigma$ subrepresentation of $S^m(\mathfrak{g}[t, t^{-1}]^\sigma)$. There are $n, \ell \in \mathbb{Z}$ such that for any standard monomial $M = g_1 t^{i_1} \cdots g_m t^{i_m}$ with $i_1 \geq n$, there is $H_M \in I$ such that $\mathrm{LT}_< H_M = M$ and all t-powers in H_M are $\geq \ell$.*

This result allows us to reduce "big" monomials modulo I to a linear combination of smaller monomials. The proof is based on the interplay between the $<$ and $\prec$-orders on standard monomials, generalizing the key arguments of [Iyudu and Sierra 2020] to work in our substantially more commutative context.

3.1. ***Preparatory results.*** Before proving the main reduction result Proposition 3.1, we will need several preparatory lemmata which will help us control the $\prec$-leading terms of elements of I. If $M = g_1 t^{i_1} \cdots g_m t^{i_m}$ is a standard monomial, we say that $(g_1, \ldots, g_m)$ is the *congruence class* of M and write $M \equiv (g_1, \ldots, g_m)$. We will construct, for any congruence class $\underline{g} \in (\mathcal{B}_\Delta)^m$, an element $H_{\underline{g}} \in I$ with $\mathrm{LT}_\prec H_{\underline{g}} \equiv \underline{g}$. We will see that the existence of these elements is sufficient to prove Proposition 3.1.

Constructing the $H_{\underline{g}}$ will take several steps. We first show that $H_{\underline{g}}$ exists for $\underline{g} = (g_\theta, \ldots, g_\theta)$ or $(g_{-\theta}, \ldots, g_{-\theta})$; and further, we can assume, in this case, that all monomials of $H_{\underline{g}}$ have the same congruence class.

Lemma 3.2. *Let $0 \neq F \in S^m(\mathfrak{g}[t, t^{-1}]^\sigma)$, and consider the adjoint action of $(t\mathfrak{g}[t])^\sigma$ on $S^m(\mathfrak{g}[t, t^{-1}]^\sigma)$. Let I denote the subrepresentation generated by F:*

(1) *If $\{g_{\bar{s},\alpha} t^{rp+s}, F\} = 0$ for all $p \in \mathbb{Z}_{\geq 1}$ and all $(\bar{s}, \alpha) \in \Delta^+$ then $F \in S^m(g_\theta[t, t^{-1}])$: that is, F is a linear combination of monomials of the form $g_\theta t^{k_1} \cdots g_\theta t^{k_m}$ for some $k_1 \leq \cdots \leq k_m$.*

(2) *If* $\{g_{\bar{s},\alpha}t^{rp+s}, F\} = 0$ *for all* $p \in \mathbb{Z}_{\geq 1}$ *and all* $(\bar{s}, \alpha) \in \Delta^-$ *then* $F \in S^m(g_{-\theta}[t, t^{-1}])$: *that is, F is a linear combination of monomials of the form* $g_{-\theta}t^{k_1} \cdots g_{-\theta}t^{k_m}$ *for some* $k_1 \leq \cdots \leq k_m$.

(3) *There are nonzero* $G \in I \cap S^m(g_\theta[t, t^{-1}])$ *and* $G' \in I \cap S^m(g_{-\theta}[t, t^{-1}])$.

Proof. (1) Let $(\bar{s}, \alpha) \in \Delta^+$. Let us take $q \in \mathbb{Z}_{\geq 1}$ such that $\bar{q} = \bar{s}$ and so that if $g_1 t^{j_1} \cdots g_m t^{j_m}$ is a monomial in F. Then $q > |j_1| + \cdots + |j_m|$. Thus, if $M = g_1 t^{j_1} \cdots g_m t^{j_m}$ is a monomial appearing in F, then

$$\{g_{\bar{s},\alpha}t^{rq+s}, M\} = g_2 t^{j_2} \cdots g_m t^{j_m}[g_{\bar{s},\alpha}, g_1]t^{j_1+q} + g_1 t^{j_1} g_3 t^{j_3} \cdots g_m t^{j_m}[g_{\bar{s},\alpha}, g_2]t^{j_2+q} + \cdots + g_1 t^{j_1} \cdots g_{v-1} t^{j_{v-1}}[g_{\bar{s},\alpha}, g_m]t^{j_m+q}.$$

Note that each of the monomials above are written in normal order. Since we have taken q to be large enough, none of those monomials will appear in expressions coming from the action of $g_{\bar{s},\alpha}t^q$ on any monomial $M' \neq M$ occurring in F. Hence by hypothesis the Lie brackets $[g_{\bar{s},\alpha}, g_i]$ are all zero. This is true for all $(\bar{s}, \alpha) \in \Delta^+$ which implies that $[\mathfrak{n}^+, g_i] = 0$ for $1 \leq i \leq m$. Hence $g_1 = \cdots = g_m = g_\theta$, which spans the center of $\mathfrak{n}^+$. The proof of (2) is similar.

To prove (3), we first construct $0 \neq G \in I$ so that

$$\{g_{\bar{s},\alpha}t^{rp+s}, G\} = 0 \quad \text{for all } p \in \mathbb{Z}_{\geq 1} \text{ and } (\bar{s}, \alpha) \in \Delta^+. \tag{3.3}$$

If (3.3) holds for $G = F$, we are done. Otherwise there exist $a \in \mathbb{Z}_{\geq 1}$ and $(\bar{u}, \beta) \in \Delta^+$ such that $F_1 = \{g_{\bar{u},\beta}t^{ra+u}, F\} \neq 0$. If (3.3) holds for $G = F_1$, we are done; else there exist $b \in \mathbb{Z}_{\geq 1}$ and $(\bar{v}, \gamma) \in \Delta^+$ such that $F_2 = \{g_{\bar{v},\gamma}t^{rb+v}, F_1\} \neq 0$. As $\{g_{\bar{s},\alpha}, -\}$ increases the height of monomials and the height of a monomial in $S^m(\mathfrak{g}[t, t^{-1}]^\sigma)$ is at most $m \operatorname{ht}\theta$, this process must terminate. Thus, continuing this procedure repeatedly, we will reach a nonzero element G of I such that (3.3) holds.

By symmetry there is $0 \neq G' \in I$ such that $\{g_{\bar{s},\alpha}t^{rp+s}, G'\} = 0$ for all $p \in \mathbb{Z}_{\geq 1}$ and $(\bar{s}, \alpha) \in \Delta^-$. The two parts of (3) then follow by applying (1) and (2). □

Enumerate $\Delta^+ = \{(\bar{\xi}_1, \beta_1), \ldots, (\bar{\xi}_\ell, \beta_\ell)\}$ where $0 \leq \xi_i \leq r-1$. Then the set $\{g_{\bar{\xi}_1,\beta_1}, \ldots, g_{\bar{\xi}_\ell,\beta_\ell}\}$ forms a basis of $\mathfrak{n}^+$ and the set $\{g_{\overline{-\xi_1},-\beta_1}, \ldots, g_{\overline{-\xi_\ell},-\beta_\ell}\}$ forms a basis of $\mathfrak{n}^-$. Recall that if $g \in \mathfrak{g}_{\bar{s}}$, then we write $\bar{s} = |g|$.

We now prove the existence of $H_{\underline{g}}$ with $\mathrm{LT}_\prec H_{\underline{g}} \equiv \underline{g}$ for any $\underline{g} \in (\mathcal{B}_\Delta)^m$.

Lemma 3.4. *Let I be the* $(t\mathfrak{g}[t])^\sigma$*-subrepresentation of* $S^m(\mathfrak{g}[t, t^{-1}]^\sigma)$ *generated by an element* $F \neq 0$. *Let* $(g_1, \ldots, g_m) \in (\mathcal{B}_\Delta)^m$. *Then there exists G in I such that* $\mathrm{LT}_\prec(G) \equiv (g_1, \ldots, g_m)$.

Proof. Let $d = \operatorname{ht}(\theta) > 0$. Write $\mathfrak{g}_\theta = \mathfrak{g}_{\bar{s},\alpha}$, which implies that $\mathfrak{g}_{-\theta} = \mathfrak{g}_{\overline{-s},-\alpha}$ i.e., $|\mathfrak{g}_\theta| = -|\mathfrak{g}_{-\theta}| \pmod r$. We prove by induction that for all $n \in \{0, \ldots, m\}$ there is $G(n) \in I$ with the following properties:

(1) $\mathrm{LT}_\prec(G(n)) = g_1 t^{k_1} \cdots g_n t^{k_n} g_\theta t^{k_{n+1}} \cdots g_\theta t^{k_m}$ for some $k_1 \leq \cdots \leq k_m \in \mathbb{Z}$.

(2) $k_i + 2rd \leq k_{i+1}$ for all $i > n$.

(3) $k_i < k_{i+1}$ for all $i \leq n$.

Then $G = G(m)$ is the element we seek.

To construct $G(0)$ we use Lemma 3.2. By that result, there is a nonzero $H \in I \cap S^m(\mathfrak{g}_{-\theta}[t, t^{-1}])$. Write $\mathrm{LT}_\prec H = g_{-\theta}t^{j_1} \dots g_{-\theta}t^{j_m}$ where $\bar{j}_1 = \cdots = \bar{j}_m = |\mathfrak{g}_{-\theta}|$. Let

$$G(0) = (\operatorname{ad} g_\theta t^{rd+|\mathfrak{g}_\theta|})^2(\operatorname{ad} g_\theta t^{2rd+|\mathfrak{g}_\theta|})^2 \cdots (\operatorname{ad} g_\theta t^{mrd+|\mathfrak{g}_\theta|})^2(H).$$

The monomials in $G(0)$ are obtained from monomials in H by applying exactly two $g_\theta t^i$ to each $g_{-\theta}t^{j_k}$. (This is because $(\operatorname{ad} g_\theta)^3(g_{-\theta}) = 0$.) The $\prec$-leading term of $G(0)$ is obtained from $\mathrm{LT}_\prec(H)$ by applying the highest powers of t to the rightmost letters. Thus $\mathrm{LT}_\prec(G(0)) = g_\theta t^{k_1} \cdots g_\theta t^{k_m}$ where $k_i = j_i + 2(ird + |\mathfrak{g}_\theta|)$. It is immediate that this has the claimed properties.

Now assume we have constructed $G(n)$ as claimed for some $1 \le n \le m-1$; we construct $G(n+1)$. There are two cases, depending on whether g_{n+1} lies in a positive or negative root space:

Negative case: $g_{n+1} \in \mathfrak{n}^-$.

Step 1$^-$. Let $H_1^- = (\operatorname{ad} g_{-\theta}t^{r-|\mathfrak{g}_\theta|})^{2(m-n)}(G(n))$ and let

$$N_1^- = g_1 t^{k_1} \cdots g_n t^{k_n} g_{-\theta} t^{k_{n+1}+2(r-|\mathfrak{g}_\theta|)} \cdots g_{-\theta} t^{k_m+2(r-|\mathfrak{g}_\theta|)}.$$

We claim that $N_1^- = \mathrm{LT}_\prec(H_1^-)$. This is because the $\prec$-leading term of H_1^- is obtained from $\mathrm{LT}_\prec(G(n))$ by increasing the right-most powers of t as much as possible, and this clearly gives N_1^-.

Step 2$^-$. Let $H_2^- = (\operatorname{ad} g_\theta t^{r+|\mathfrak{g}_\theta|})^{2(m-n-1)}(H_1^-)$ and let

$$N_2^- = g_1 t^{k_1} \cdots g_n t^{k_n} g_{-\theta} t^{k_{n+1}+2(r-|\mathfrak{g}_\theta|)} g_\theta t^{k_{n+2}+4r} \cdots g_\theta t^{k_m+4r}.$$

We claim that $N_2^- = \mathrm{LT}_\prec(H_2^-)$.

First, we show that $N_2^- = \mathrm{LT}_\prec(\operatorname{ad} g_{-\theta}t^{r-|g_\theta|})^{2(m-n)}(N_1^-)$. This is similar to the previous paragraph; again, the $\prec$-leading term is obtained by increasing the right-most powers of t as much as possible, giving N_2^-.

We now show that $N_2^- = \mathrm{LT}_\prec(H_2^-)$. Let $N_2' \ne N_2^-$ be any other monomial of H_2^-. Suppose that $N_2' \in \mathfrak{g}t^{k_1} \cdots \mathfrak{g}t^{k_n} \mathfrak{g}t^{k_{n+1}+2(r-|\mathfrak{g}_\theta|)} \mathfrak{g}t^{k_{n+2}+4r} \cdots \mathfrak{g}t^{k_m+4r}$ — that is, N_2' involves the same t-powers as N_2^-. (If this is not the case, it is clear that $N_2' \prec N_2^-$.) By construction, there must be a monomial $M = h_1 t^{k_1} \cdots h_m t^{k_m}$ of $G(n)$ so that N_2' is a monomial in $(\operatorname{ad} g_\theta t^{r-|\mathfrak{g}_{-\theta}|})^{2(m-n-1)}(\operatorname{ad} g_{-\theta}t^{r-|\mathfrak{g}_\theta|})^{2(m-n)}(M)$; that is, N_2' is obtained from a monomial in $G(n)$ with the same t-powers as $\mathrm{LT}_\prec G(n)$. Further, N_2' must have been obtained from M by applying $g_{-\theta}t^{r-|\mathfrak{g}_\theta|}$ twice to each letter of M in the $(n+1)$-st place or further right, and $g_\theta t^{r+|\mathfrak{g}_\theta|}$ twice to each of the letters in the $(n+2)$-nd place and further right; any other combination of applying $g_{-\theta}t^{r-|\mathfrak{g}_\theta|}$ and $g_\theta t^{r+|\mathfrak{g}_\theta|}$ would give either the wrong t-powers or a zero result. The only possibility to obtain a nonzero result is if $M = h_1 t^{k_1} \dots h_n t^{k_n} g_\theta t^{k_{n+1}} \dots g_\theta t^{k_m}$, as we needed $(\operatorname{ad} g_{-\theta}t^{r-|\mathfrak{g}_\theta|})^2$ to be nonzero on the $(n+1)$-st and greater letters of M. Thus we have $N_2' = h_1 t^{k_1} \dots h_n t^{k_n} g_{-\theta} t^{k_{n+1}+2(r-|\mathfrak{g}_\theta|)} g_\theta t^{k_{n+2}+4r} \dots g_\theta t^{k_m+4r}$, where necessarily $h_1 t^{k_1} \dots h_n t^{k_n} \prec g_1 t^{k_1} \dots g_n t^{k_n}$. So $N_2' \prec N_2^-$.

Step 3$^-$. As θ is the longest root of $\mathfrak{g}$, by Fact (5) there are $p \in \mathbb{Z}_{\geq 0}$ and $(\overline{s_1}, \eta_1), \ldots, (\overline{s_p}, \eta_p) \in \Delta^+$ so that g_{n+1} is a nonzero multiple of

$$[g_{\bar{s}_1,\eta_1}, [\ldots, [g_{\bar{s}_p,\eta_p}, g_{-\theta}]\ldots]].$$

Let us define $H_3^- = \{g_{\bar{s}_1,\eta_1}t^{s_1}, \{\ldots, \{g_{\bar{s}_p,\eta_p}t^{s_p}, H_2^-\}\ldots\}\}$ and $S = \sum_{i=1}^{p} s_i$ and let

$$N_3^- = g_1 t^{k_1} \cdots g_n t^{k_n} g_{n+1} t^{k_{n+1}+2(r-|\mathfrak{g}_\theta|)+S} g_\theta t^{k_{n+2}+4r} \cdots g_\theta t^{k_m+4r}.$$

We claim that $N_3^- = \mathrm{LT}_\prec(H_3^-)$. First, $\mathrm{LT}_\prec(\{g_{\bar{s}_1,\eta_1}t^{s_1}, \{\ldots, \{g_{\bar{s}_p,\eta_p}t^{s_p}, N_2^-\}\ldots\}\})$ is obtained from N_2^- by applying the elements $g_{\bar{s}_i,\eta_i}t^{s_i}$ as far to the right as possible. As $[g_{\bar{s}_i,\eta_i}, g_\theta] = 0$ for all i (this is because θ is the highest root of $\mathfrak{g}$), applying the $g_{\bar{s}_i,\eta_i}t^{s_i}$ as far to the right as possible means applying them to the $(n+1)$-st letter $g_{-\theta}t^{k_{n+1}+2(r-|\mathfrak{g}_\theta|)}$ to obtain N_3^-. In other words,

$$N_3^- = \mathrm{LT}_\prec(\{g_{\bar{s}_1,\eta_1}t^{s_1}, \{\ldots, \{g_{\bar{s}_p,\eta_p}t^{s_p}, N_2^-\}\ldots\}\}).$$

Now let $N_2' \neq N_2^-$ be any other monomial of H_2^- and let

$$H_3' = \{g_{\bar{s}_1,\eta_1}t^{s_1}, \{\ldots, \{g_{\bar{s}_p,\eta_p}t^{s_p}, N_2'\}\ldots\}\}.$$

If $N_2' \notin \mathfrak{g}t^{k_1} \cdots \mathfrak{g}t^{k_n}\mathfrak{g}t^{k_{n+1}+2(r-|\mathfrak{g}_\theta|)}\mathfrak{g}t^{k_{n+2}+4r} \cdots \mathfrak{g}t^{k_m+4r}$, then by comparing t-powers we must have $\mathrm{LT}_\prec H_3' \prec N_3^-$. Suppose now that N_2' involves the same t-powers as N_2^-; that is,

$$N_2' \in \mathfrak{g}t^{k_1} \cdots \mathfrak{g}t^{k_n}\mathfrak{g}t^{k_{n+1}+2(r-|\mathfrak{g}_\theta|)}\mathfrak{g}t^{k_{n+2}+4r} \cdots \mathfrak{g}t^{k_m+4r}.$$

We have seen in the proof of Step 2$^-$ that we must have

$$N_2' = h_1 t^{k_1} \cdots h_n t^{k_n} g_{-\theta} t^{k_{n+1}+2(r-|\mathfrak{g}_\theta|)} g_\theta t^{k_{n+2}+4r} \cdots g_\theta t^{k_m+4r}$$

for some $h_1, \ldots, h_n \in \mathcal{B}$. Then

$$\mathrm{LT}_\prec(H_3') = h_1 t^{k_1} \cdots h_n t^{k_n} g_{n+1} t^{k_{n+1}+2(r-|\mathfrak{g}_\theta|)+S} g_\theta t^{k_{n+2}+4r} \cdots g_\theta t^{k_m+4r}.$$

But as $N_2' \prec N_2^-$, thus $h_1 t^{k_1} \cdots h_n t^{k_n} \prec g_1 t^{k_1} \cdots g_n t^{k_n}$: that is, $\mathrm{LT}_\prec(H_3') \prec N_3^-$. Thus $N_3^- = \mathrm{LT}_\prec H_3^-$, as claimed.

Let $G(n+1) = H_3^-$. We have shown that (1) is satisfied; but the t-powers in N_3^- also satisfy (2) and (3).

Positive case: $g_{n+1} \in \mathfrak{n}^+$.

Step 1$^+$. Let $H_1^+ = (\mathrm{ad}\, g_{-\theta}t^{r-|\mathfrak{g}_\theta|})^{2(m-n-1)}(G(n))$ and let

$$N_1^+ = g_1 t^{k_1} \cdots g_n t^{k_n} g_\theta t^{k_{n+1}} g_{-\theta} t^{k_{n+2}+2(r-|\mathfrak{g}_\theta|)} \cdots g_{-\theta} t^{k_m+2(r-|\mathfrak{g}_\theta|)}.$$

We claim that $N_1^+ = \mathrm{LT}_\prec H_1^+$. This is because the $\prec$-leading term of H_1^+ must come from the $\prec$-leading term of $G(n)$ by increasing the rightmost t-powers as much as possible, which gives N_1^+.

Step 2$^+$. As θ is the longest root of $\mathfrak{g}$, by Fact (5) there are $q \in \mathbb{Z}_{\geq 0}$ and $(\overline{s_1}, \gamma_1), \ldots, (\overline{s_q}, \gamma_q) \in \Delta^-$ so that g_{n+1} is a nonzero multiple of

$$[g_{\bar{s}_1,\gamma_1}, [\ldots, [g_{\bar{s}_q,\gamma_q}, g_\theta] \ldots]].$$

Let $H_2^+ = \{g_{\bar{s}_1,\gamma_1} t^{r+s_1}, \{\ldots, \{g_{\bar{s}_q,\gamma_q} t^{r+s_q}, H_1^+\} \ldots\}\}$. Let $S = qr + \sum_{i=1}^q s_i$ and let

$$N_2^+ = g_1 t^{k_1} \cdots g_n t^{k_n} g_{n+1} t^{k_{n+1}+S} g_{-\theta} t^{k_{n+2}+2(r-|\mathfrak{g}_\theta|)} \cdots g_{-\theta} t^{k_m+2(r-|\mathfrak{g}_\theta|)}.$$

We claim that $N_2^+ = \mathrm{LT}_\prec(H_2^+)$. This is because, again, the $\prec$-leading term of H_2^+ comes by increasing the rightmost t-powers as much as possible. As $[g_{\bar{s}_i,\gamma_i}, g_{-\theta}] = 0$ for all i, this is done by acting on $g_\theta t^{k_{n+1}}$ with all of the $g_{\bar{s}_i,\gamma_i} t^{r+s_i}$ to obtain N_2^+.

Note that $q \leq d$ and each $s_i \leq r$. Thus $k_{n+1} + S < k_{n+1} + 2rd \leq k_{n+2} < k_{n+2} + 2(r - |\mathfrak{g}_\theta|)$, using the induction hypothesis. Therefore N_2^+ is a standard monomial.

Step 3$^+$. Let $H_3^+ = (\operatorname{ad} g_\theta t^{r+|\mathfrak{g}_\theta|})^{2(m-n-1)}(H_2^+)$ and let

$$N_3^+ = g_1 t^{k_1} \cdots g_n t^{k_n} g_{n+1} t^{k_{n+1}+S} g_\theta t^{k_{n+2}+4r} \cdots g_\theta t^{k_m+4r}.$$

We claim that $N_3^+ = \mathrm{LT}_\prec(H_3^+)$. This is because the $\prec$-leading term of H_3^+ comes from applying $\operatorname{ad} g_\theta t^{r+|g_\theta|}$ as far to the right as possible: that is, applying $g_\theta t^{r+|\mathfrak{g}_\theta|}$ twice to each $g_{-\theta} t^{k_i+2(r-|\mathfrak{g}_\theta|)}$, where $i \in \{n+2, \ldots, m\}$.

Let $G(n+1) = H_3^+$. We have verified that $\mathrm{LT}_\prec G(n+1)$ has the claimed congruence class, so (1) is satisfied. The observation at the end of Step 2$^+$ ensures that (3) holds; and (2) follows as $k_{i+1} + 4r - (k_i + 4r) = k_{i+1} - k_i$. □

Let I be a nontrivial $\mathfrak{g}[t]^\sigma$-subrepresentation of $S^m(\mathfrak{g}[t, t^{-1}]^\sigma)$. We have seen that for any $\underline{g} \in (\mathcal{B}_\Delta)^m$ there is $H_{\underline{g}} \in I$ with $\mathrm{LT}_\prec H_{\underline{g}} \equiv \underline{g}$. We will use this to construct elements of I with arbitrary $\prec$-leading term, as long as all t-powers involved are sufficiently large.

The next result considers which $\prec$-leading terms may be obtained from a particular $H_{\underline{g}}$. It is modeled on the methods of [Iyudu and Sierra 2020].

Lemma 3.5. *Let $(g_1, \ldots, g_m) \in (\mathcal{B}_\Delta)^m$. Let $G \in S^m(\mathfrak{g}[t, t^{-1}]^\sigma)$ with $\mathrm{LT}_\prec G = g_1 t^{a_1} \cdots g_m t^{a_m}$. Let I be the $(t\mathfrak{g}[t])^\sigma$-subrepresentation of $S^m(\mathfrak{g}[t, t^{-1}]^\sigma)$ generated by G. Let s be the smallest power of t occurring in G.*

Suppose that $(g'_1, \ldots, g'_m) \in \mathcal{B}^m$ so that

for all j there is weight-homogeneous $g''_j \in \mathfrak{g}$

$$\text{so that } [g''_j, g_j] \text{ is a nonzero scalar multiple of } g'_{m+1-j}. \tag{3.6}$$

Then for all standard monomials $M = g'_1 t^{i_1} \cdots g'_m t^{i_m}$ with $i_1 > \max(0, 2a_m - s)$, there is $H_M \in I$ with $\mathrm{LT}_\prec H_M = M$.

Proof. Let

$$H_M = \{g_1'' t^{i_m - a_1}, \{g_2'' t^{i_{m-1} - a_2} \dots, \{g_m'' t^{i_1 - a_m}, G\} \dots\}\}.$$

Then $H_M \in I$ because each $g_j'' t^{i_{m+1-j} - a_j} \in \mathfrak{g}[t]^\sigma$ we have $\bar{a}_j = |g_j|$ and $\bar{i}_j = |g_j'|$ implies that $\overline{i_{m+1-j} - a_j} = |g_j''|$. We claim that $\mathrm{LT}_< H_M = M$.

To prove our claim, let us consider the process by which monomials of H_M are produced from monomials of G. Let

$$A = h_1 t^{b_1} \cdots h_m t^{b_m}$$

be a monomial of G. Then A induces two kinds of monomials of H_M: either we act on each letter of A with exactly one $g_j'' t^{i_{m+1-j} - a_j}$, or we act on some letter of A more than once and in consequence do not act at all on at least one letter of A. We call the first of these actions *permutational* and the second *nonpermutational*. The $<$-leading term in H_M always comes from a permutational action. This is because in a nonpermutational action, the smallest power of t is always bounded above by a_m since there will be a missing $h_i t^{b_i}$ which is not acted upon and $b_i \le a_m$ by the definition of the monomial order $\prec$; whereas in a permutational action, the lowest power of t is bounded below by $i_1 + b_1 - a_m$ which is strictly bigger than a_m because of our assumption that $i_1 > 2a_m - s$.

Now, if $\sum b_i < \sum a_i$, then all monomials of

$$\{g_1'' t^{i_m - a_1}, \{g_2'' t^{i_{m-1} - a_2} \dots, \{g_m'' t^{i_1 - a_m}, A\} \dots\}\}$$

have degree $< \deg M$ and so are strictly $< M$. We may thus assume that $\sum b_i = \sum a_i$.

A monomial coming from a permutational action upon A looks like

$$[g_m'', h_{\tau(1)}] t^{i_1 + b_{\tau(1)} - a_m} [g_{m-1}'', h_{\tau(2)}] t^{i_2 + b_{\tau(2)} - a_{m-1}} \cdots [g_1'', h_{\tau(m)}] t^{i_m + b_{\tau(m)} - a_1} \tag{3.7}$$

for some $\tau \in \mathfrak{S}_m$, where all the Lie brackets are nonzero. We claim that

$$(3.7) \le M,$$

with equality only when $b_{\tau(i)} = m + 1 - i$ and $h_{\tau(i)} = g_{m+1-i}$ for all i.

Note that $b_{\tau(1)} \le a_m$ thanks to the definition of $\prec$. We will have several different cases:

- If $b_{\tau(1)} < a_m$ then our claim is obvious because $i_1 + b_{\tau(1)} - a_m < i_1$.
- If $b_{\tau(1)} = a_m$, then we must have $[g_m'', h_{\tau(1)}] \le g_1'$, with equality only if $h_{\tau(1)} = g_m$. This is because, by the definition of $\prec$, $h_{\tau(1)} \le g_m$ and $<$ has property (2.2).
- If $b_{\tau(1)} = a_m$ and $h_{\tau(1)} = g_m$, then we want to show that

  $$g_1' t^{i_1} [g_{m-1}'', h_{\tau(2)}] t^{i_2 + b_{\tau(2)} - a_{m-1}} \cdots [g_1'', h_{\tau(m)}] t^{i_m + b_{\tau(m)} - a_1} \le g_1' t^{i_1} \cdots g_m' t^{i_m},$$

  with equality only when $b_{\tau(i)} = a_{m+1-i}$ and $h_{\tau(i)} = g_{m+1-i}$ for all $2 \le i \le m$. This can be proved by repeating the same procedure and moving towards the right.

This analysis show that $\mathrm{LT}_{\prec} G$ is the only monomial which contributes to the occurrence of M in H_M. To finish, we note that by (3.6), then

$$[g''_m, g_m] \cdots [g''_1, g_1] = \lambda g'_1 \cdots g'_m$$

for some $\lambda \neq 0$. Thus M occurs in H_M with nonzero coefficient $C\lambda$, where

$$C = \#\{\tau \in \mathfrak{S}_m : a_{\tau(i)} = a_{m+1-i} \text{ and } g_{\tau(i)} = g_{m+1-i} \text{ for all } i \} \neq 0.$$

Thus $\mathrm{LT}_{<} H_M = M$. □

To use Lemma 3.5 we must prove that we can always find a situation where (3.6) holds. This is given by the next elementary lemma.

Lemma 3.8. *For any $g \in \mathcal{B}$ there are $g' \in \mathcal{B}_\Delta$ and weight-homogeneous $g'' \in \mathfrak{g}$ so that $[g'', g']$ is a nonzero scalar multiple of g.*

Proof. If $g = g_{\pm\xi_i, \pm\beta_i} \in \mathcal{B}_\Delta$, set $g^- = g_{\mp\xi_i, \mp\beta_i}$. Let $g' = g$ and $g'' = [g, g^-]$. As $\{g, g^-, [g, g^-]\}$ form an $\mathfrak{sl}_2$-triple, $[g'', g']$ is a nonzero scalar multiple of g.

If $g \in \mathfrak{h}$ then we consider the action of σ on g. We give the proof for $r = 3$. If $|g| = \bar{0}$ then $g = h_{\alpha_i} + h_{\sigma(\alpha_i)} + h_{\sigma^2(\alpha_i)}$ for some simple root α_i and we can take $g' = g_{|\alpha_i|, \alpha_i}$ and $g'' = g_{|-\alpha_i|, -\alpha_i}$. If $|g| = \bar{1}$ then $g = h_{\alpha_i} + \eta h_{\sigma(\alpha_i)} + \eta^2 h_{\sigma(\alpha_i)}$ for some α_i. Set $g' = e_{\alpha_i} + \eta e_{\sigma(\alpha_i)} + \eta^2 e_{\sigma^2(\alpha_i)}$ and $g'' = f_{\alpha_i} + f_{\sigma(\alpha_i)} + f_{\sigma^2(\alpha_i)}$. Likewise, if $|g| = \bar{2}$ then $g = h_{\alpha_i} + \eta^2 h_{\sigma(\alpha_i)} + \eta h_{\sigma(\alpha_i)}$ for some α_i. Set $g' = e_{\alpha_i} + \eta^2 e_{\sigma(\alpha_i)} + \eta e_{\sigma^2(\alpha_i)}$ and $g'' = f_{\alpha_i} + f_{\sigma(\alpha_i)} + f_{\sigma^2(\alpha_i)}$. □

3.2. ***The main reduction result.*** We now prove our main reduction result, Proposition 3.1.

Proof of Proposition 3.1. We may assume that I is generated by a single element $F \neq 0$. For every $\underline{g}' = (g'_1, \ldots, g'_m) \in (\mathcal{B}_\Delta)^m$, using Lemma 3.4 choose $G_{\underline{g}'} \in I$ with $\mathrm{LT}_{\prec} G_{\underline{g}'} \equiv \underline{g}'$. Let $s_{\underline{g}'}$ be the smallest power of t occurring in $G_{\underline{g}'}$, and let $S_{\underline{g}'}$ be the largest power of t. Let n be an integer so that

$$n \geq \max(0, \{2S_{\underline{g}'} - s_{\underline{g}'} : \underline{g}' \in (\mathcal{B}_\Delta)^m\}).$$

Let ℓ be the smallest power of t occurring in F.

Fix $M = g_1 t^{i_1} \cdots g_m t^{i_m}$ with $i_1 > n$. We construct H_M. Using Lemma 3.8, choose $g'_1, \ldots, g'_m \in \mathcal{B}_\Delta$ and weight-homogeneous $g''_1, \ldots, g''_m \in \mathfrak{g}$ so that $[g''_j, g'_j]$ is a nonzero multiple of g_{m+1-j} for $1 \leq j \leq m$. Let $G = G_{(g'_1, \ldots, g'_m)} \in I$, and write $\mathrm{LT}_{\prec} G = g'_1 t^{a_1} \cdots g'_m t^{a_m}$. Let

$$H_M = \{g''_1 t^{i_m - a_1}, \{\ldots, \{g''_m t^{i_1 - a_m}, G\}, \ldots\}\}.$$

The proof of Lemma 3.5 shows that

$$M = \mathrm{LT}_{<} H_M,$$

which is in I as our definition of n ensures that $0 \leq i_1 - a_m \leq i_j - a_{m+1-j}$ for any j.

Throughout the proofs of Lemmata 3.2, 3.4, and 3.5 we acted on F only with positive powers of t. Thus the smallest t-power in H_M is $\geq \ell$. □

From Proposition 3.1 we deduce our first just-infinite growth result.

Theorem 3.9. *Let $\mathfrak{g}$ be a finite-dimensional simple Lie algebra and let $\sigma \in \operatorname{Aut}\mathfrak{g}$:*

(1) *The enveloping algebra $U(\mathfrak{g}[t]^\sigma)$ has just-infinite growth. That is, if J is a nonzero ideal of $U(\mathfrak{g}[t]^\sigma)$, then $U(\mathfrak{g}[t]^\sigma)/J$ has polynomial growth.*

(2) *Let I be a nonzero Poisson ideal of $S(\mathfrak{g}[t]^\sigma)$. Then $S(\mathfrak{g}[t]^\sigma)/I$ has polynomial growth.*

Proof. If M is a standard monomial in the letters $\mathcal{B}[t]^\sigma$, we define the *modified degree* of M to be

$$\operatorname{md} M = \deg M + \operatorname{len} M.$$

Let $\mathcal{U}^j$ be the subspace of $U(\mathfrak{g}[t]^\sigma)$ spanned by standard monomials of modified degree $\leq j$, and let $\mathcal{S}^j$ be the corresponding subspace of $S(\mathfrak{g}[t]^\sigma)$. As $U(\mathfrak{g}[t]^\sigma)$ is degree-graded, and, by the PBW theorem, is filtered by length, thus $\mathcal{U}^\bullet$ defines a filtration on $U(\mathfrak{g}[t]^\sigma)$: that is, $\mathcal{U}^i\mathcal{U}^j \subseteq \mathcal{U}^{i+j}$ for all $i, j \in \mathbb{N}$. Likewise, $\mathcal{S}^\bullet$ defines a filtration on $S(\mathfrak{g}[t]^\sigma)$. It suffices to prove that

$$\dim \frac{\mathcal{U}^j + J}{J} \quad \text{and} \quad \dim \frac{\mathcal{S}^j + I}{I}$$

have polynomial growth.

Let

$$\operatorname{gr}_{\mathrm{md}}(U(\mathfrak{g}[t]^\sigma)) = \bigoplus_j \mathcal{U}^j/\mathcal{U}^{j-1}.$$

This is a (commutative) Poisson algebra: if $F \in \mathcal{U}^i$ and $G \in \mathcal{U}^j$ then $FG - GF \in \mathcal{U}^{i+j-1}$ so $\operatorname{gr}_{\mathrm{md}} U(\mathfrak{g}[t]^\sigma)$ is commutative.

Define

$$\{\operatorname{gr}_{\mathrm{md}} F, \operatorname{gr}_{\mathrm{md}} G\} = \frac{FG - GF + \mathcal{U}^{i+j-2}}{\mathcal{U}^{i+j-2}} \in (\operatorname{gr}_{\mathrm{md}} U(\mathfrak{g}[t]^\sigma))_{i+j-1}.$$

In fact, as degree alone defines a grading on $U(\mathfrak{g}[t]^\sigma)$, there is a canonical identification $\operatorname{gr}_{\mathrm{md}} U(\mathfrak{g}[t]^\sigma) \cong \operatorname{gr}_{\mathrm{len}}(U(\mathfrak{g}[t]^\sigma)) = S(\mathfrak{g}[t]^\sigma)$ as Poisson algebras. Further, $\operatorname{gr}_{\mathrm{md}} J$ is a nontrivial md-homogeneous Poisson ideal of $S(\mathfrak{g}[t]^\sigma)$. Likewise, $\operatorname{gr}_{\mathrm{md}} I$ is a nontrivial md-homogeneous Poisson ideal of $S(\mathfrak{g}[t]^\sigma)$. As

$$\dim \frac{\mathcal{U}^j + J}{J} = \dim \frac{\mathcal{S}^j + \operatorname{gr}_{\mathrm{md}} J}{\operatorname{gr}_{\mathrm{md}} J}$$

and similarly for I, it suffices to prove:

Claim: Let K be a nontrivial md-homogeneous Poisson ideal of $S(\mathfrak{g}[t]^\sigma)$. Then $\dim((\mathcal{S}^j + K)/K)$ is bounded by a polynomial in j.

Let $K' = \operatorname{gr}_{\mathrm{len}}(K)$, which is a nontrivial len-graded Poisson ideal of $S(\mathfrak{g}[t]^\sigma)$, and thus meets some $S^m(\mathfrak{g}[t]^\sigma)$ nontrivially. Note that $K' \cap S^m(\mathfrak{g}[t]^\sigma)$ is a $(t\mathfrak{g}[t])^\sigma$-subrepresentation of $S^m(\mathfrak{g}[t]^\sigma)$. By Proposition 3.1, there is $n \in \mathbb{Z}_{\geq 1}$ so that for any standard monomial $M = g_1 t^{i_1} \cdots g_m t^{i_m}$ with $i_1 \geq n$, there is $H'_M \in K'$ such that $\operatorname{LT}_< H'_M = M$. Now, for each standard monomial M as above, there is $H_M \in K$

with $H'_M = \mathrm{gr}_{\mathrm{len}} H_M$. As the monomial ordering $<$ compares length first, $M = \mathrm{LT}_< H_M$ as well. Further, as K is md-graded we may take H_M to be md-homogeneous.

Let $F \in \mathcal{S}^j$. Repeatedly using the H_M to reduce monomials involving m or more powers of t which are bigger than n, we may rewrite F (modulo K) without increasing $\mathrm{md}\, F$ so that no monomial in F contains more than $m-1$ t-powers bigger than n. That is, $(\mathcal{S}^j + K)/K$ is spanned by the image of the set of standard monomials M with $\mathrm{md}(M) \leq j$ which admit a factorization $M = M_1 M_2$, where M_1 is a standard monomial involving t-powers $\leq n-1$, and M_2 is a standard monomial of length $< m$ involving t-powers $\geq n$. Let us call such monomials *normal words*, and let

$$r(j) = \#\{\text{normal words } M : \mathrm{md}(M) \leq j\} \geq \dim \frac{\mathcal{S}^j + K}{K}.$$

It is clear that $r(j) \leq b(j)c(j)$, where

$$b(j) = \#\{ \text{ standard monomials } M_1 \text{ involving only } t\text{-powers} \leq n-1 \text{ with } \mathrm{md}\, M_1 \leq j \ \}$$

and

$$c(j) = \#\{ \text{ standard monomials } M_2 \text{ of length } < m \text{ involving only } t\text{-powers } \geq n \text{ and with } \mathrm{md}\, M_2 \leq j\}.$$

Now, modified degree exceeds length, so $b(j)$ does not exceed the number of standard monomials of length $\leq j$ involving t-powers between 0 and $n-1$, which is $\binom{(\dim \mathfrak{g})n+j}{j}$ and is bounded above by a polynomial in j of degree $n(\dim \mathfrak{g})$. And in a monomial M_2 of degree at most j and length at most $m-1$ in t-powers $\geq n$ there are no more than $j \dim \mathfrak{g}$ choices for each letter of M_2, so $c(j) \leq (j \dim \mathfrak{g})^{m-1}$. Thus $r(j)$ is bounded above by a polynomial in j of degree $n(\dim \mathfrak{g}) + m - 1$. □

Remark 3.10. A small modification to the proof of Theorem 3.9 shows that if J is a nonzero ideal of $U((t\mathfrak{g}[t])^\sigma)$ and I is a nonzero Poisson ideal of $S((t\mathfrak{g}[t])^\sigma)$ then $U((t\mathfrak{g}[t])^\sigma)/J$ and $S((t\mathfrak{g}[t])^\sigma)/I$ have polynomial growth. We leave the details to the reader.

4. Useful technical results

We will prove the just-infinite growth results Theorems 1.1(0) and 1.2(0) in the next section. In this section we establish a number of useful preparatory results which will allow similar counting arguments to those in the proof of Theorem 3.9 to apply to affine Kac–Moody algebras.

We establish notation, which will apply to the next two sections. Let L be an affine Kac–Moody algebra, with central element c and derivation d. Then there are a finite-dimensional simple Lie algebra $\mathfrak{g}$ and an automorphism σ of $\mathfrak{g}$ so that

$$L'/(c) \cong \mathfrak{g}[t, t^{-1}]^\sigma.$$

(Here L' is the derived subalgebra of L.) We may assume that σ induces an automorphism of the Dynkin diagram and the root system of $\mathfrak{g}$; we denote this diagram automorphism by σ as well. Let r be the order of σ. Throughout the rest of the paper we fix the meanings of L, $\mathfrak{g}$, σ, r, d, c as in this paragraph. We

further fix a primitive r-th root of unity, η, and induce a $\mathbb{Z}/r\mathbb{Z}$-grading on $\mathfrak{g}$ as in Section 2. We will use other notation from Section 2 without comment.

Let $\lambda \in \Bbbk$. Define $U_\lambda(L) = U(L)/(c-\lambda)$ and $S_\lambda(L) = S(L)/(c-\lambda)$. As $c-\lambda$ is Poisson central in $S(L)$, the factor $S_\lambda(L)$ is a Poisson algebra. Note that $S_0(L') \cong S(\mathfrak{g}[t,t^{-1}]^\sigma)$ as Poisson algebras.

We first show that every nonzero Poisson ideal of $S_\lambda(L)$ meets $S_\lambda(L')$. This requires a technical lemma.

Lemma 4.1. *Let $\lambda \in \Bbbk$. Let $G \in S_\lambda(L)$, and suppose that there exists a σ-equivariant $x \in \mathfrak{g}$ so that $\{xt^{rj+|x|}, G\} = 0$ for all $j \in \mathbb{Z}$. Then $G \in S_\lambda(L')$.*

Proof. Write $G = \sum_{i=0}^n d^i G_i$, where $G_i \in S_\lambda(L')$. Then for any σ-equivariant $x \in \mathfrak{g}$ and $m \in r\mathbb{Z} + |x|$,

$$0 = \{xt^m, G\} = \sum_{i=0}^n d^i(-(i+1)mxt^m G_{i+1} + \{xt^m, G_i\}). \tag{4.2}$$

Thus for all m, i, the coefficient of d^i in (4.2) must vanish, and so we have

$$\{xt^m, G_i\} = (i+1)mxt^m G_{i+1}. \tag{4.3}$$

Fix i and let $m \gg 0$ be large enough so that t^{-m} does not occur in G_i. As $m \gg 0$ varies, from the definition of the Poisson bracket on $S_\lambda(L)$ we see that the LHS of (4.3) changes only in the powers of t which occur, and not in the coefficients. On the other hand, the expressions for the coefficients of the right-hand side involve a factor of m and so vary with m. Thus (4.3) cannot hold for all values of $m \in r\mathbb{Z} + |x|$ unless both sides are identically 0. We conclude that $G_i = 0$ for all $i \geq 1$, and $G = G_0 \in S_\lambda(L')$. □

Corollary 4.4. *Let $\lambda \in \Bbbk$ and let I be a nonzero Poisson ideal of $S_\lambda(L)$. Then $I \cap S_\lambda(L') \neq (0)$.*

Proof. Let $0 \neq G \in I$, and choose $(\bar{s}, \alpha) \in \Delta^+$. There is $k \geq 1$ so that

$$\{g_\alpha t^{ri_k+s}, \ldots, \{g_\alpha t^{ri_1+s}, G\} \ldots\} = 0$$

for all $i_1, \ldots, i_k \in \mathbb{Z}$; without loss of generality let k be minimal. Thus there is some

$$H = \{g_\alpha t^{ri_{k-1}+s}, \ldots, \{g_\alpha t^{ri_1+s}, G\} \ldots\} \neq 0.$$

By Lemma 4.1, $H \in I \cap S_\lambda(L')$. □

We need to extend the orderings $<$ and $\prec$ defined on loop algebras to $U_\lambda(L)$ and $S_\lambda(L)$. To this end, define a σ-equivariant basis $\mathcal{B}$ of $\mathfrak{g}$ as in Basic Fact (9), and note that $\{d\} \cup \{gt^{rj+|g|} : g \in \mathcal{B}, j \in \mathbb{Z}\}$ is a basis of $L/(c)$. We denote this basis by $\mathcal{B}^d[t,t^{-1}]^\sigma$. We extend the ordering $<$ on $\mathcal{B}[t,t^{-1}]^\sigma$ to an ordering on $\mathcal{B}^d[t,t^{-1}]^\sigma$ by saying that $d > gt^n$ if and only if $n < 0$ and $d < gt^n$ if and only if $n \geq 0$.

We must modify (2.3) to give a basis of $U_\lambda(L)$ and $S_\lambda(L)$. We will say that a *standard monomial* in the elements of $\mathcal{B}^d[t,t^{-1}]^\sigma$ is an expression of the form

$$M = g_1 t^{rm_1+|g_1|} \cdots g_i t^{rm_i+|g_i|} d^j g_1' t^{rn_1+|g_1'|} \cdots g_k' t^{rn_k+|g_k'|}, \tag{4.5}$$

where $g_a, g_b' \in \mathcal{B}$ and

$$g_1 t^{rm_1+|g_1|} \leq \cdots \leq g_i t^{rm_i+|g_i|} < d < g_1' t^{rn_1+|g_1'|} \leq \cdots \leq g_k' t^{rn_k+|g_k'|}.$$

We sometimes write a standard monomial as

$$M = g_1 t^{m_1} \cdots g_i t^{m_i} d^j g_1' t^{n_1} \cdots g_k' t^{n_k},$$

and when we do so we assume that $\overline{m_a} = |g_a|$ and $\overline{n_a} = |g_a'|$ for all a; in other words, that this monomial is an element of $S(L)$.

By the PBW theorem, both $U_\lambda(L)$ and $S_\lambda(L)$ have a basis of standard monomials in the elements of $\mathcal{B}^d[t, t^{-1}]^\sigma$. We say that $i + j + k$ is the *length* of M, which we denote by len M, and that the total t-power $\sum_a rm_a + |g_a| + \sum_b rn_b + |g_b'|$ is the *degree* of M, which we denote by $\deg(M)$. Note that $U_\lambda(L')$ and $S_\lambda(L')$ have a basis of standard monomials in the elements of $\mathcal{B}[t, t^{-1}]^\sigma$.

We extend the monomial orderings $<$ and $\prec$ from Section 2.2 to define two orderings $<$ and $\prec$ on standard monomials in the elements of $\mathcal{B}^d[t, t^{-1}]^\sigma$: the ordering $<$ compares length first, then degree, and then compares monomials lexicographically from left to right, whereas $\prec$ compares length first, then degree, then compares monomials lexicographically from right to left.

We use the following reduction lemma, which applies Proposition 3.1 to ideals of $U_\lambda(L)$, $U_\lambda(L')$ or Poisson ideals of $S_\lambda(L)$, $S_\lambda(L')$.

Lemma 4.6. *Let $\lambda \in \Bbbk$:*

(1) *Let J be a nonzero ideal of $U_\lambda(L)$ or of $U_\lambda(L')$. There are $m, \ell, n \in \mathbb{Z}$, with $m, n > 0$, so that if*

$$M = g_1 t^{i_1} \cdots g_m t^{i_m}$$

is a standard monomial in the elements of $\mathcal{B}[t, t^{-1}]^\sigma$ with $i_1 \geq n$, then there is $H_M \in J$ so that

$$\mathrm{LT}_< H_M = M$$

and so that all t-powers occurring in H_M are $\geq \ell$.

(2) *Let I be a nonzero Poisson ideal of $S_\lambda(L)$ or of $S_\lambda(L')$. There are $m, \ell, n \in \mathbb{Z}$, with $m, n > 0$, so that if*

$$M = g_1 t^{i_1} \cdots g_m t^{i_m}$$

is a standard monomial in the elements of $\mathcal{B}[t, t^{-1}]^\sigma$ with $i_1 \geq n$, then there is $G_M \in I$ so that

$$\mathrm{LT}_< G_M = M$$

and so that all t-powers occurring in G_M are $\geq \ell$.

Proof. (1) We give the proof for $J \lhd U_\lambda(L)$; the proof for $J \lhd U_\lambda(L')$ is similar but easier. We filter $U_\lambda(L)$ by length of monomials, and define a Poisson bracket on $\mathrm{gr}_{\mathrm{len}}\, U_\lambda(L)$ as in the proof of Theorem 3.9. The relation

$$xt^i yt^j - yt^j xt^i = [x, y]t^{i+j} + i\delta_{i+j,0}\kappa(x, y)\lambda \tag{4.7}$$

on $U_\lambda(L)$ induces the Poisson bracket

$$\{xt^i, yt^j\} = [x, y]t^{i+j}$$

in $\mathrm{gr}_{\mathrm{len}} U_\lambda(L)$ and so $\mathrm{gr}_{\mathrm{len}} U_\lambda(L)$ may be identified with $S_0(L)$ as Poisson algebras. By Corollary 4.4 applied to $\mathrm{gr}_{\mathrm{len}} J$, which is a nonzero Poisson ideal of $S_0(L)$, there is $0 \neq H \in J$ so that $\mathrm{gr}_{\mathrm{len}} H \in S_0(L')$.

Let $m = \mathrm{len}(\mathrm{gr}_{\mathrm{len}} H) = \mathrm{len}\, H$ and let ℓ be the smallest t-power occurring in H. Let I be the Poisson ideal of $S_0(L)$ generated by $\mathrm{gr}_{\mathrm{len}} H$. By Proposition 3.1 there is $n \in \mathbb{Z}$ so that if $M = g_1 t^{i_1} \cdots g_m t^{i_m}$ is a standard monomial with $i_1 \geq n$, there is $G_M \in I \cap S_0(L')$ with $\mathrm{LT}_< G_M = M$. The procedure in the proof of Proposition 3.1 that produces G_M (that is, the procedure in Lemmata 3.2, 3.4, and 3.5) produces $B_1, \ldots, B_s \in \mathcal{B}[t, t^{-1}]^\sigma$, involving only nonnegative t-powers, so that

$$\{B_1, \{\ldots, \{B_s, \mathrm{gr}_{\mathrm{len}} H\}, \ldots\} = G_M.$$

Now, if $B \in \mathcal{B}[t, t^{-1}]^\sigma$ and $P \in U_\lambda(L)$, then

$$\mathrm{gr}_{\mathrm{len}}[B, P] = \{B, \mathrm{gr}_{\mathrm{len}} P\}, \quad \text{if } \{B, \mathrm{gr}_{\mathrm{len}} P\} \neq 0. \tag{4.8}$$

Let

$$H_M = [B_1, [\ldots, [B_s, H]\ldots].$$

Applying (4.8) we see that $G_M = \mathrm{gr}_{\mathrm{len}} H_M$, and as the monomial ordering $<$ compares length first, $\mathrm{LT}_< H_M = \mathrm{LT}_< G_M = M$. As the t-powers in the B_i are nonnegative, the t-powers in H_M are no smaller than ℓ.

(2) This proof is similar to the proof of (1). By Corollary 4.4 it suffices to give the proof for a Poisson ideal I of $S_\lambda(L')$. Let $0 \neq G \in I$. Let $m = \mathrm{len}\, G$ and let ℓ be the smallest t-power occurring in G.

Filter $S_\lambda(L')$ by length, as in the proof of (1), so $\mathrm{gr}_{\mathrm{len}} S_\lambda(L') = S_0(L')$. We again apply Proposition 3.1 to obtain n so that if M is a standard monomial of length m involving only t-powers $\geq n$, then there is G_M in the Poisson ideal of $S_0(L')$ generated by $\mathrm{gr}_{\mathrm{len}} G$ with $\mathrm{LT}_< G_M = M$.

Temporarily, let $\{-,-\}_0$ denote the Poisson bracket in $S_0(L')$ and let $\{-,-\}_\lambda$ denote the Poisson bracket in $S_\lambda(L')$. Then, similarly to (4.8), if $B \in \mathcal{B}[t, t^{-1}]^\sigma$ and $P \in S_\lambda(L')$, then

$$\mathrm{gr}_{\mathrm{len}}(\{B, P\}_\lambda) = \{B, \mathrm{gr}_{\mathrm{len}} P\}_0, \quad \text{if } \{B, \mathrm{gr}_{\mathrm{len}} P\}_0 \neq 0. \tag{4.9}$$

As in the proof of (1), there are $B_1, \ldots, B_s \in \mathcal{B}[t, t^{-1}]^\sigma$, involving only nonnegative powers of t, so that

$$G_M = \{B_1, \{\ldots, \{B_s, \mathrm{gr}_{\mathrm{len}} G\}_0, \ldots, \}_0.$$

Let

$$H_M = \{B_1, \{\ldots, \{B_s, G\}_\lambda, \ldots, \}_\lambda.$$

As I is a Poisson ideal, $H_M \in I$. As before, applying (4.9) this time, $\mathrm{LT}_< H_M = M$, and H_M does not involve any t-powers smaller than ℓ. □

Let L_+ be the sub-Lie algebra of L generated by all gt^n with $n > 0$, and similarly define L_- to be generated by all gt^n with $n < 0$. For all λ, then $U(L_+)$, $U(L_-)$ are subalgebras of $U_\lambda(L)$, and similarly $S(L_+)$, $S(L_-) \subseteq S_\lambda(L)$, where these second inclusions are inclusions of Poisson algebras.

We next apply Lemma 4.6 to show that a nontrivial ideal of $U_\lambda(L)$ must meet $U(L_+)$, and, symmetrically, $U(L_-)$. We do not know of a way to show this without using growth.

Proposition 4.10. *Let $\lambda \in \Bbbk$:*

(1) *Let J be a nonzero ideal of $U_\lambda(L)$ or of $U_\lambda(L')$. There are nonzero elements $H^+ \in J \cap U(L_+)$ and $H^- \in J \cap U(L_-)$.*

(2) *Let I be a nonzero Poisson ideal of $S_\lambda(L)$ or of $S_\lambda(L')$. There are nonzero elements $G^+ \in I \cap S(L_+)$ and $G^- \in I \cap S(L_-)$.*

Proof. We give the proof if $J \lhd U_\lambda(L)$.

By symmetry, it suffices to prove the result for H_+. Let $A = U_\lambda(L)/J$. We claim that A has finite GK-dimension as a right or left $U(L_+)$-module; by symmetry it suffices to consider the GK-dimension as a right module.

The proof of Lemma 4.6(1) produces $H \in J$ and, from H, integers m, ℓ, n so that if

$$M = g_1 t^{i_1} \cdots g_m t^{i_m}$$

is a standard monomial with $i_1 \geq n$, then there is $H_M \in U(L'_{\geq 0}) H U(L'_{\geq 0})$ so that

$$\mathrm{LT}_< H_M = M$$

and so that all t-powers occurring in H_M are $\geq \ell$. Without loss of generality, $\ell \leq 0$. Let

$$D = (\dim \mathfrak{g})(n-1) + m - 1.$$

We claim that the GK-dimension of A as a $U(L_+)$-module is at most D.

Let $V \subset U_\lambda(L)$ be a finite-dimensional subspace which includes 1; we will show that the $U(L_+)$-module $X = (U(L_+)VU(L_+) + J)/J$ has GKdim $\leq D$. (As $\mathrm{GKdim}_{U(L_+)} A$ is by the definition the supremum over all finitely generated $U(L_+)$ submodules $A' \subseteq A$ of $\mathrm{GKdim}_{U(L_+)} A'$, this is sufficient to prove the claim.) Since $U(L_+)$ is finitely graded, for any $j \in \mathbb{Z}_{\geq 0}$ the subspace $\{x \in X : \deg x \leq j\}$ is finite dimensional, so it suffices to show that the dimension, considered as a function of j, grows as a polynomial of degree $\leq D$.

We may enlarge V without damage, so assume that $H \in VU(L_+)$. Let s be the minimum t-power occurring in any element of V and let m' be the maximum length of any element of V. We may enlarge V again so that $s \leq \ell$ (recall that $\ell \leq 0$) and $m' \geq m$ and so that $V = S^{\leq m'}(\mathfrak{g}_{\bar{s}} t^s \oplus \mathfrak{g}_{\overline{s+1}} t^{s+1} \oplus \cdots \oplus \mathfrak{g}_{\bar{0}} t^0 \oplus \Bbbk d)$. The minimal degree of an element of V is $sm' \leq 0$.

Note that if $xt^k \in L_+$ and M is a monomial in V, then

$$xt^k M = Mxt^k + \text{ a sum of monomials of length } \leq \mathrm{len}(M) \text{ involving } t\text{-powers } \geq s.$$

Thus $xt^k M \in VU(L_+)$. By induction, $U(L_+)VU(L_+) = VU(L_+)$. Thus X is spanned by standard monomials M which admit a factorization $M = M^0 M^1$, where M^0 is a standard monomial of length $\leq m'$ involving d and t-powers between s and 0, and M^1 is a standard monomial involving t-powers ≥ 1. For all such M, our assumption that $s \leq \ell$ and our choice of V mean that $H_M \in VU(L_+) = U(L_+)VU(L_+)$. Working modulo J to rewrite the monomials M^1 and repeatedly applying Lemma 4.6, we see that X is

spanned by the image of standard monomials M in $VU(L_+)$ which admit a factorization $M = M^0 M_1 M_2$, where M^0 is a standard monomial of length $\leq m'$ involving d and t-powers between s and 0, M_1 is a standard monomial involving t-powers between 1 and $n-1$, and M_2 is a standard monomial of length $< m$ involving only t-powers $\geq n$.

Thus, the growth of X is thus bounded by the growth of

$$Y = \{M^0 M_1 M_2\},$$

where M^0 is a standard monomial of length $\leq m$ involving d and t-powers between s and 0, M_1 is a standard monomial involving t-powers between 1 and $n-1$, and M_2 is a standard monomial involving only t-powers $\geq n$ and with $\operatorname{len} M_2 < m$. It thus suffices to show that

$$q(j) = \#\{M \in Y : \deg(M) \leq j\}$$

is bounded by a polynomial in j of degree D.

Clearly $q(j)$ is bounded by $ab(j)c(j)$, where

$$\begin{aligned} a &= \dim S^{\leq m'}(\mathfrak{g}t^s \oplus \cdots \oplus \mathfrak{g}t^0 \oplus \Bbbk d), \\ b(j) &= \#\{\text{standard monomials } M_1 \text{ of degree } \leq j - sm' \text{ involving only } t\text{-powers between 1 and } n-1\}, \\ c(j) &= \#\{\text{standard monomials } M_2 \text{ of length } < m \text{ and degree } \leq j - sm' \text{ involving only } t\text{-powers } \geq n\}. \end{aligned}$$

As M_1 involves only positive t-powers, $\deg M_1 \geq \operatorname{len} M_1$, and so $b(j)$ does not exceed the number of standard monomials of length $\leq j - sm'$ involving t-powers between 1 and $n-1$, which is $\leq \binom{(\dim \mathfrak{g})(n-1)+j-sm'}{j-sm'}$ and is bounded above by a polynomial in j of degree $(\dim \mathfrak{g})(n-1)$. And in a monomial M_2 of degree at most $j - sm'$ and length at most $m-1$ in t-powers $\geq n$ there are no more than $(j - sm')\dim \mathfrak{g}$ choices for each letter of M_2, so $c(j) \leq ((j - sm')\dim \mathfrak{g})^{m-1}$. Thus there is $\lambda \in \mathbb{R}$ so that $q(j) \leq \lambda j^D$ for all but finitely many j, proving the claim.

As $\operatorname{GKdim} U(L_+) = \infty$, the natural map $U(L_+) \to A$ cannot be injective, and thus $U(L_+) \cap J \neq (0)$. (This argument is essentially Scholium 2.4, but for modules, not algebras.)

The proofs of other statements are similar, using other parts of Lemma 4.6. □

5. Just-infinite growth

We now prove our first main results, Theorems 1.1(0) and 1.2(0). Throughout, fix $\lambda \in \Bbbk$. Let J be a nonzero ideal of $U_\lambda(L)$ and let $A = U_\lambda(L)/J$. Let J' be a nonzero ideal of $U_\lambda(L')$ and let $A' = U_\lambda(L')/J'$. Let I be a nonzero Poisson ideal of $S_\lambda(L)$ and let $C = S_\lambda(L)/I$, and let I' be a nonzero Poisson ideal of $S_\lambda(L')$ and let $C' = S_\lambda(L')/I'$.

We first prove Theorem 1.1(0). As $U_\lambda(L)$ is not finitely graded, there are some technical issues in the proof. Our solution is to extend the definition of modified degree from the proof of Theorem 3.9 so that the associated graded ring $B = \operatorname{gr}_{\mathrm{md}} U_\lambda / \operatorname{gr}_{\mathrm{md}} J$ of A will be (finitely) connected graded. Here recall that an

$\mathbb{N}$-graded $\Bbbk$-algebra $R = \bigoplus_{n\geq 0} R_n$ is *finitely connected graded* if each $\dim R_n < \infty$ and $R_0 = \Bbbk$. We then use the Poisson GK-dimension of [Petukhov and Sierra 2020] to bound the GK-dimension of A and of A'.

We extend the earlier definition of modified degree to define the *modified degree* of an element of $\mathcal{B}^d[t, t^{-1}]^\sigma$ to be

$$\operatorname{md} gt^n = |n| + 1, \quad \operatorname{md} d = 1.$$

Let $S = \operatorname{gr}_{\mathrm{md}} U_\lambda(L)$ and let $S' = \operatorname{gr}_{\mathrm{md}} U_\lambda(L')$. As $\operatorname{md}([gt^a, ht^b]) \leq |a+b|+1 < |a|+|b|+2 = \operatorname{md}(gt^a) + \operatorname{md}(ht^b)$, and $\operatorname{md}([d, gt^a]) < \operatorname{md}(gt^a) + \operatorname{md}(d)$, thus $U_\lambda(L)$ is almost commutative with respect to the filtration induced by md, and so, as a ring, S is isomorphic to a polynomial ring in the variables d and $gt^{ra+|g|}$, graded by $\deg(gt^a) = |a|+1$ and $\deg d = 1$. The subring S' is isomorphic to a polynomial ring in the variables $gt^{ra+|g|}$. Thus S is connected graded. Further, S has a Poisson bracket

$$\{gt^a, ht^b\} = \begin{cases} [g,h]t^{a+b} & \text{if } ab \geq 0, \\ 0 & \text{else,} \end{cases} \quad \text{and} \quad \{d, gt^a\} = agt^a$$

induced from the commutator in $U_\lambda(L)$, as in the proof of Theorem 3.9, and S' is a Poisson subalgebra of S.

Let $B = S/\operatorname{gr}_{\mathrm{md}} J$, and note that B is the associated graded ring of A with respect to the filtration induced by md. Likewise, let $B' = S'/\operatorname{gr}_{\mathrm{md}} J'$; we have $B' \cong \operatorname{gr}_{\mathrm{md}}(A')$. As usual, $\operatorname{gr}_{\mathrm{md}} J$ is a Poisson ideal of S (respectively, $\operatorname{gr}_{\mathrm{md}} J'$ is a Poisson ideal of S') and so the Poisson bracket on S (respectively, S') descends to B (respectively, B').

The point of introducing the filtration md is that the GK-dimension of A may be computed from the growth of B. For $j \in \mathbb{N}$, let $S_{\leq j}$ denote the span in S of standard monomials of modified degree $\leq j$ and similarly define $S'_{\leq j}$.

Proposition 5.1. *For $j \in \mathbb{N}$, let $B(j)$ be the image of $S_{\leq j}$ in B and let $B'(j)$ be the image of $S'_{\leq j}$ in B'. Then*

$$\operatorname{GKdim} A = \overline{\lim} \log_j \dim B(j) \quad \textit{and} \quad \operatorname{GKdim} A' = \overline{\lim} \log_j \dim B'(j).$$

Proof. This is [Petukhov and Sierra 2020, Proposition 3.14]. We must check that the hypotheses of that result apply; that is, for B, that $\{B(i), B(j)\} \subseteq B(i+j)$ for all i, j, and that for some s, we have $B(j) \subseteq B(s)^{\{j\}}$ for all j, and similarly for B'. Here recall that if V is a subspace of B, then $V^{\{j\}}$ is defined inductively:

- $V^{\{0\}} = \Bbbk$.
- For $j \in \mathbb{N}$, define $V^{\{j+1\}} = VV^{\{j\}} + \{V, V^{\{j\}}\}$.
- In particular $V^{\{1\}} = V$.

The needed properties for B and B' follow immediately from similar properties for $S_{\leq j}$, taking $s = r+1$. That $\{B(i), B(j)\} \subseteq B(i+j)$ is immediate. And by Lemma 5.3,

$$\mathfrak{g}_{\bar{a}} t^{(n-1)r+a} = (\operatorname{ad} \mathfrak{g}_{\bar{0}} t^r)^{n-1}(\mathfrak{g}_{\bar{a}} t^a) \subseteq (S_{\leq r+1})^{\{n\}} \tag{5.2}$$

for $a \in \{0, \ldots, r-1\}$. Thus $S_{\leq j} \subseteq (S_{\leq r-1})^{\{j\}}$ for all j. □

We give the proof of (5.2).

Lemma 5.3. *For $a \in \{0, \dots, r-1\}$, we have*

$$\mathfrak{g}_{\bar{a}} = [\mathfrak{g}_{\bar{0}}, \mathfrak{g}_{\bar{a}}] = \operatorname{span}([x, y] : x \in \mathfrak{g}_{\bar{0}}, y \in \mathfrak{g}_{\bar{a}}).$$

Proof. A basis for $\mathfrak{g}_{\bar{a}}$ is made up of the $g_{\bar{a},\alpha}$ for $\alpha \in \Delta_{\bar{a}}$ and the $h^i_{\bar{a}}$ from Table 1. We show that all these elements are in $[\mathfrak{g}_{\bar{0}}, \mathfrak{g}_{\bar{a}}]$. From the Basic Facts we have $[g_{\overline{-a},-\alpha}, g_{\bar{a},\alpha}] \in \mathfrak{g}_{\bar{0}}$, and $[[g_{\overline{-a},-\alpha}, g_{\bar{a},\alpha}], g_{\bar{a},\alpha}]$ is a nonzero scalar multiple of $g_{\bar{a},\alpha}$. Writing

$$h^i_{\bar{a}} = \sum_{j=0}^{r'} \eta^{aj} h_{\sigma^j(i)},$$

let

$$e = \sum_{j=0}^{r'} e_{\sigma^j(\alpha_i)} \in \mathfrak{g}_{\bar{0}}, \qquad f = \sum_{j=0}^{r'} \eta^{aj} f_{\sigma^j(\alpha_i)} \in \mathfrak{g}_{\bar{a}}.$$

Then $[e, f] = h^i_{\bar{a}}$. The lemma follows. □

Let $0 \neq H^+ \in J \cap U(L_+)$ and $0 \neq H^- \in J \cap U(L_-)$ be the elements produced by Proposition 4.10. Let $G^+ = \operatorname{gr}_{\mathrm{md}} H^+$ and let $G^- = \operatorname{gr}_{\mathrm{md}} H^-$. We now apply the reduction procedure in Proposition 3.1 to obtain a reduction result for B. This is:

Lemma 5.4. *There exist positive integers m, n so that the following hold*:

(1) *Every standard monomial $M = g_{i_1} t^{j_1} \cdots g_{i_m} t^{j_m}$ with $n \leq j_1 \leq \cdots \leq j_m$ satisfies*

$$M = H + \sum c_s M_s,$$

where $H \in \operatorname{gr}_{\mathrm{md}} J \cap S(L_+)$ is homogeneous in modified degree, the sum is finite, $c_s \in \Bbbk^\times$, and the M_s are standard monomials so that for each t we have at least one t-power $< n$ featuring in M_s.

(2) *Every standard monomial $M = g_{i_1} t^{j_1} \cdots g_{i_m} t^{j_m}$ with $j_1 \leq \cdots \leq j_m \leq -n$ satisfies*

$$M = H + \sum c_s M_s,$$

where $H \in \operatorname{gr}_{\mathrm{md}} J \cap S(L_-)$ is homogeneous in modified degree, the sum is finite, $c_s \in \Bbbk^\times$, and the M_s are standard monomials so that for each t we have at least one t-power $> -n$ featuring in M_s.

Proof. It suffices to prove (1). Noting that $S(L_+) \subset S$ is an inclusion of Poisson algebras, we apply Proposition 3.1 to reduce M modulo the Poisson ideal of $S(L_+)$ generated by G^+. This produces an element $H \in \operatorname{gr}_{\mathrm{md}} J \cap S(L_+)$ satisfying all claimed properties but md-homogeneity. But since G^+ is md-homogeneous and the adjoint action of homogeneous elements of L_+, as is used in the proof of Proposition 3.1, preserves md-homogeneity, H is md-homogeneous as claimed. □

We now prove Theorem 1.1(0).

Proof of Theorem 1.1(0). The proofs for $U_\lambda(L)$ and for $U_\lambda(L')$ are very similar; we give the proof for $U_\lambda(L)$. By Proposition 5.1, it suffices to show that $\dim B(j)$ has polynomial growth. (For $U_\lambda(L')$ we show that $\dim B'(j)$ has polynomial growth.)

Let $F \in B(j)$. Repeatedly applying Lemma 5.4 to F, we may rewrite F without increasing $\operatorname{md} F$ so that no monomial in F contains more than $m-1$ t-powers bigger than n or more than $m-1$ t-powers smaller than $-n$. That is, $B(j)$ is spanned by the image of the set of standard monomials M with $\operatorname{md}(M) \le j$ which admit a factorization $M = M_0M_1M_2$, where M_1 is a standard monomial of length $< m$ involving t-powers $\le -n$, M_2 is a standard monomial involving d and t-powers between $1-n$ and $n-1$, and M_3 is a standard monomial of length $< m$ involving t-powers $\ge n$. Let us call such monomials *normal words*, and let

$$r(j) = \#\{\text{normal words } M : \operatorname{md}(M) \le j\}.$$

We have seen that $\dim B(j) \le r(j)$; an argument very similar to the proof of Proposition 4.10 shows that $r(j) \le e(j)c(j)^2$, where

$$e(j) = \#\{M : M \text{ is a standard monomial involving only } d \text{ and } t\text{-powers between } 1-n \text{ and } n-1 \text{ and } \operatorname{md} M \le j\}$$

and

$$c(j) = \#\{M_2 : M_2 \text{ is a standard monomial of length } < m \text{ involving only } t\text{-powers} \ge n \text{ with } \deg M_2 \le j\}.$$

We have seen that $c(j) \le (j \dim \mathfrak{g})^{m-1}$. Similarly to the proofs of Theorem 3.9 and Proposition 3.1, $e(j) \le \binom{(\dim \mathfrak{g})(2n-1)+1+j}{j}$. Thus $r(j)$ is bounded by a polynomial in j of degree $(\dim \mathfrak{g})(2n-1)+2m-1$. □

We also have:

Proposition 5.5. *Let L be an affine Kac–Moody algebra with central element c. Then $U(L) \otimes_{\Bbbk[c]} \Bbbk(c)$ and $U(L') \otimes_{\Bbbk[c]} \Bbbk(c)$ have just-infinite growth as $\Bbbk(c)$-algebras.*

Proof. None of the steps in the proof of Theorem 1.1(0) used that $\Bbbk$ is algebraically closed. Thus we may change the ground field to $\Bbbk(x)$, where x is an indeterminate, to obtain: for any $\lambda \in \Bbbk(x)$, $U_{\Bbbk(x)}(L)/(c-\lambda)$ has just-infinite growth as a $\Bbbk(x)$-algebra. This holds in particular for $\lambda = x$, giving that $U_{\Bbbk(x)}(L)/(c-x) \cong U(L) \otimes_{\Bbbk[c]} \Bbbk(c)$ has just-infinite growth as a $\Bbbk(c)$-algebra, and similarly for L'. □

We will also use modified degree to prove Theorem 1.2(0).

Proof of Theorem 1.2(0). As above, modified degree introduces a filtration on C and on C', and we have

$$\operatorname{gr}_{\operatorname{md}} C \cong S/\operatorname{gr}_{\operatorname{md}} I = R, \quad \operatorname{gr}_{\operatorname{md}}(C') \cong S'/\operatorname{gr}_{\operatorname{md}}(I') = R'.$$

Define $R(j)$, $R'(j)$ to be the elements of R (respectively, R') of modified degree $\le j$. As in the proof of Theorem 1.1, we may use G^+, G^- from Proposition 4.10 to find $m, n \in \mathbb{Z}$ so that for all j, $R(j)$ is spanned by the image of the set of standard monomials M with $\operatorname{md}(M) \le j$ which admit a factorization $M = M_0M_1M_2$, where M_0 is a standard monomial of length $< m$ involving t-powers $\le -n$, M_1 is a

standard monomial involving d and t-powers between $1-n$ and $n-1$, and M_2 is a standard monomial of length $< m$ involving t-powers $\geq n$. Let $D = (\dim \mathfrak{g})(2n-1) + 2m - 1$. The same argument as in the proof of Theorem 1.1 shows that $\dim R(j)$ is bounded by a polynomial in j of degree D and so is bounded by some λj^D.

Let V be a finite-dimensional subspace of C, containing 1, and choose a so that $\mathrm{gr}_{\mathrm{md}} V \subseteq R(a)$. Then $\dim V^j \leq \dim R(aj) \leq \lambda a^D j^D$. Therefore $\mathrm{GKdim}\, C \leq D$.

The proof that $\mathrm{GKdim}\, C' < \infty$ is similar. □

We also have:

Proposition 5.6. *Let I be a nonzero Poisson ideal of $S(L) \otimes_{\Bbbk[c]} \Bbbk(c)$ and let I' be a nonzero Poisson ideal of $S(L') \otimes_{\Bbbk[c]} \Bbbk(c)$. Then $S(L) \otimes_{\Bbbk[c]} \Bbbk(c)/I$ and $S(L') \otimes_{\Bbbk[c]} \Bbbk(c)/I'$ have polynomial growth.*

We leave the proof to the reader.

Remark 5.7. Propositions 5.5 and 5.6 should be viewed as temporary results, as later we will prove, as stated in Theorems 1.1(2) and 1.2(2), that $U(L) \otimes_{\Bbbk[c]} \Bbbk(c)$ and $U(L') \otimes_{\Bbbk[c]} \Bbbk(c)$ are simple, and that $S(L) \otimes_{\Bbbk[c]} \Bbbk(c)$ and $S(L') \otimes_{\Bbbk[c]} \Bbbk(c)$ are Poisson simple.

Remark 5.8. The strategy of proof of Theorems 1.1(0) and 1.2(0) can be modified to give new proofs of [Iyudu and Sierra 2020, Theorems 5.3 and 5.6]. Recall that the *Witt algebra* $W = \Bbbk[t, t^{-1}]\partial$ is the Lie algebra of polynomial vector fields on the punctured line $\left(\text{here } \partial = \frac{d}{dt}\right)$ and the *Virasoro algebra* Vir is the unique nontrivial central extension of W. As a vector space we have

$$\mathrm{Vir} = W \oplus \Bbbk c$$

with Lie bracket

$$[f\partial, g\partial] = (fg' - f'g)\partial + \mathrm{Res}_0(f'g'' - g'f'')c, \quad c \text{ central.}$$

We describe the necessary modifications to prove [loc. cit., Theorem 5.3], that central quotients of $U(\mathrm{Vir})$ have just-infinite growth.

Let $\lambda \in \Bbbk$ and let J be a nonzero ideal of $U_\lambda := U(\mathrm{Vir})/(c-\lambda)$. The proof of Proposition 4.10 may be modified (using the reduction in [loc. cit., Lemma 2.2] instead of Proposition 3.1) to show that there are nonzero $H^+ \in J \cap U(t\Bbbk[t]\partial)$ and $H^- \in J \cap U(t^{-1}\Bbbk[t^{-1}]\partial)$. Similarly to our methods here, filter U_λ by modified degree, where we define $\mathrm{md}(t^{n+1}\partial) = |n| + 1$. Let $G^\pm = \mathrm{gr}_{\mathrm{md}}(H^\pm) \in \mathrm{gr}_{\mathrm{md}}(J)$. The reduction argument in [loc. cit., Lemma 2.2] now gives a version of Lemma 5.4 for $\mathrm{gr}_{\mathrm{md}}(U_\lambda/J)$ and a similar counting argument to the proof of Theorem 1.1 shows that U_λ/J has polynomial growth.

6. Simplicity of nontrivial central quotients

In this section, we prove Theorem 1.1(1,2) and Theorem 1.2(1,2). Throughout the section, let L be an affine Kac–Moody algebra with central element c and derivation d, and let $L'/(c) = \mathfrak{g}[t, t^{-1}]^\sigma$, where $\mathfrak{g}$ is a finite-dimensional simple Lie algebra and $\sigma \in \mathrm{Aut}(\mathfrak{g})$.

One of our main techniques will be to use the following corollary of Theorem 1.1(0). (This was the reason for proving such a general version of this theorem, even though the result for $\lambda \neq 0$ will soon be superseded.)

Proposition 6.1. *Let $\lambda \in \Bbbk^*$ and let B be either $U_\lambda(L)$ or $U_\lambda(L')$. Let A be a $\Bbbk$-subalgebra of B with* $\operatorname{GKdim} A = \infty$. *If J is a nonzero ideal of B then $J \cap A \neq (0)$.*

Proof. Combine Scholium 2.4 with Theorem 1.1(0). □

We will show that we can restrict without loss of generality to the case $\mathfrak{g} = \mathfrak{sl}_2$ (and $\sigma = 1$). Thus to begin we consider this case.

6.1. ***The $\mathfrak{sl}_2$ case.*** In this subsection assume now that L has type $A_1^{(1)}$ (so $\mathfrak{g} = \mathfrak{sl}_2$ and $\sigma = 1$). The derived subalgebra L' of L is isomorphic as a vector space to $\mathfrak{sl}_2[t, t^{-1}] \oplus \Bbbk c$. We fix notation for elements of L': let e, f, h be the standard basis of $\mathfrak{sl}_2$ and for $n \in \mathbb{Z}$ let $e_n = et^n,\ f_n = ft^n,\ h_n = ht^n$. Let $\lambda \in \Bbbk^*$. In this subsection we will show that $U_\lambda(L')$ is simple.

Theorem 6.2. *Let L be an affine Lie algebra of type $A_1^{(1)}$, so $L \cong \widehat{\mathfrak{sl}_2}$:*

(1) *For any $\lambda \in \Bbbk^*$, the algebra $U_\lambda(L')$ is simple.*

(2) *Any nonzero ideal of $U(L')$ contains a nonzero polynomial in c; equivalently, $U(L') \otimes_{\Bbbk[c]} \Bbbk(c)$ is simple.*

Proof. (1) Let $\mathbb{A}_\lambda$ be the subalgebra of $U_\lambda(L')$ generated by $\{h_i : i \in \mathbb{Z} \setminus \{0\}\}$. We claim that $\mathbb{A}_\lambda$ is isomorphic to the infinite Weyl algebra and is thus a simple ring. To see this, let $\mathbb{A}$ be the infinite Weyl algebra of differential operators on $\Bbbk[x_1, x_2, \dots]$. We have

$$\mathbb{A} \cong \Bbbk\langle x_1, \partial_1, x_2, \partial_2, \dots\rangle / \langle x_i x_j = x_j x_i,\ \partial_i \partial_j = \partial_j \partial_i,\ \partial_i x_j - x_j \partial_i = \delta_{i,j} \mid i, j \in \mathbb{Z}_{\geq 1}\rangle.$$

It is well known that $\mathbb{A}$ is simple; for example, $\mathbb{A}$ is a direct limit of the (finite) Weyl algebras A_n generated by $x_1, \dots, x_n, \partial_1, \dots, \partial_n$ which are simple by [Goodearl and Warfield 2004, Corollary 2.2].

A presentation for $\mathbb{A}_\lambda$ is

$$\Bbbk\langle h_i : i \in \mathbb{Z} \setminus \{0\}\rangle / \langle h_i h_j - h_j h_i = 2i\lambda \delta_{i+j,0}\rangle,$$

from (4.7). Thus the map $\partial_i \mapsto h_i/(2i\lambda)$, $x_i \mapsto h_{-i}$ induces an isomorphism $\mathbb{A} \xrightarrow{\sim} \mathbb{A}_\lambda$.

Let J be a nonzero ideal of $U_\lambda(L')$. By Proposition 6.1, $J \cap \mathbb{A}_\lambda \neq (0)$. As $\mathbb{A}_\lambda$ is simple, $1 \in J$.

(2) The proof is similar. Suppose that J is a nonzero ideal of $U(L')$. Then $J(c) := J \otimes_{\Bbbk[c]} \Bbbk(c)$ is a nonzero ideal of $U(L') \otimes_{\Bbbk[c]} \Bbbk(c)$. Combining Scholium 2.4 and Proposition 5.5 we see that $J(c)$ has a nontrivial intersection with the $\Bbbk(c)$-subalgebra of $U(L') \otimes_{\Bbbk[c]} \Bbbk(c)$ generated by $\{h_i : i \in \mathbb{Z} \setminus \{0\}\}$. Observe now that $x_i \mapsto h_{-i}$, $\partial_i \mapsto h_i/(2ic)$ induces an isomorphism between the simple ring $\mathbb{A} \otimes_\Bbbk \Bbbk(c)$ and this subalgebra. Thus $1 \in J(c)$, and it follows that $J \cap \Bbbk[c] \neq (0)$. □

6.2. ***The general case.*** We now let L be an arbitrary affine Lie algebra. To complete the proof of Theorem 1.1, we note that L contains (in fact, many choices of) a subalgebra isomorphic to $\widehat{\mathfrak{sl}_2}$.

Lemma 6.3. *Let L be an affine Kac–Moody algebra, with Chevalley generators e_i, f_i, h_i, central element c, and derivation d. For any f_i, there is a subalgebra $\bar{L}$ of L which contains f_i, c and d and is isomorphic to $\widehat{\mathfrak{sl}_2}$ via an isomorphism which sends the positive Borel of $\widehat{\mathfrak{sl}_2}$ into the positive Borel of L.*

Proof. This is well known, and is proved in [Carbone et al. 2021]. □

We now complete the proof of Theorem 1.1 by proving parts (1) and (2), which are, respectively, parts (1) and (2) of the next result.

Theorem 6.4. *Let L be an affine Lie algebra*:

(1) *For any $\lambda \in \Bbbk^*$, the algebras $U_\lambda(L)$ and $U_\lambda(L')$ are simple.*

(2) *Any nonzero ideal of $U(L)$ or of $U(L')$ contains a nonzero element of $\Bbbk[c]$; equivalently, the algebras $U(L) \otimes_{\Bbbk[c]} \Bbbk(c)$ and $U(L') \otimes_{\Bbbk[c]} \Bbbk(c)$ are simple.*

Proof. By Lemma 6.3, let $\bar{L}$ be a Lie subalgebra of L which is isomorphic to $\widehat{\mathfrak{sl}_2}$.

(1) Let J be a nonzero ideal of $U_\lambda(L)$ or of $U_\lambda(L')$. By Proposition 6.1, $J \cap U_\lambda(\bar{L}') \neq (0)$ and thus $1 \in J$ by Theorem 6.2(1).

(2) This proof is similar, applying Scholium 2.4, Proposition 5.5, and Theorem 6.2(2) to the localized ideal $J \otimes_{\Bbbk[c]} \Bbbk(c)$. □

We next finish the proof of Theorem 1.2 by proving parts (1) and (2).

Theorem 6.5. *Let L be an affine Lie algebra*:

(1) *For any $\lambda \in \Bbbk^*$, the algebras $S_\lambda(L)$ and $S_\lambda(L')$ are Poisson simple.*

(2) *Any nonzero Poisson ideal of $S(L)$ or of $S(L')$ contains a nonzero element of $\Bbbk[c]$; equivalently, $S(L) \otimes_{\Bbbk[c]} \Bbbk(c)$ and $S(L') \otimes_{\Bbbk[c]} \Bbbk(c)$ are Poisson simple.*

Proof. Again, let $\bar{L}$ be a Lie subalgebra of L which is isomorphic to $\widehat{\mathfrak{sl}_2}$. Define $e_n, h_n, f_n \in \bar{L}'$ as in Section 6.1. For $i \in \mathbb{Z}_{\geq 1}$, let $x_i = h_{-i}$ and let $y_i = h_i/(2\lambda i)$, and let $\mathbb{S}$ be the subalgebra of S_λ generated by the x_i and y_i. Here the $\{x_j\}$ Poisson commute, as do the $\{y_i\}$, and $\{y_i, x_j\} = \delta_{i,j}$. Thus $\mathbb{S}$ is isomorphic to the *infinite Poisson Weyl algebra*, which is Poisson simple. One way to see this is that the (Poisson) subalgebras $\Bbbk[x_1, \ldots, x_n, y_1, \ldots, y_n]$ are all Poisson simple; a proof is given in [Bavula 2020].

(1) Let J be a nonzero Poisson ideal of $S_\lambda(L)$ or $S_\lambda(L')$. By Theorem 1.2(0) and Scholium 2.4, $J \cap \mathbb{S} \neq (0)$. As $\mathbb{S}$ is Poisson simple, $1 \in J$.

(2) The proof is similar, using Proposition 5.6 and a modification of the proof of Theorem 6.4(2). We leave the details to the reader. □

7. Other applications

In this section, we give other applications of our growth results.

In Proposition 4.10 we used growth to show that nontrivial ideals of $U_\lambda(L)$ must meet $U(L_+)$. However, we can use Theorem 1.1(0) to obtain much stronger results of a similar flavor.

For simplicity, we state this next result only for untwisted loop algebras, although a similar result clearly holds in the twisted case. Corollary 7.2, which is an immediate consequence, seems rather surprising without the growth context.

Proposition 7.1. *Let $\mathfrak{g}$ be a finite-dimensional simple Lie algebra. Let J be a nonzero ideal of $U(\mathfrak{g}[t, t^{-1}])$ and let $\mathfrak{k}$ be any infinite-dimensional Lie subalgebra of $\mathfrak{g}[t, t^{-1}]$. Then $J \cap U(\mathfrak{k})$ is nonzero.*

Proof. This is a direct application of Proposition 6.1 with $\lambda = 0$. □

Corollary 7.2. *In the situation of Proposition 7.1, if $X \subset \mathbb{Z}$ is any infinite set, for example X consists of all of the powers of 7 or all of the primes, and g is any element of $\mathfrak{g}$, then J contains an element involving only gt^x for $x \in X$.*

Proof. The vector space spanned by $\{gt^x : x \in X\}$ is an infinite-dimensional (abelian) Lie subalgebra of $\mathfrak{g}[t, t^{-1}]$, so this follows directly from Proposition 7.1. □

It is well known that the enveloping algebras $U(\mathfrak{g}[t]^\sigma)$, $U(L)$, etc., are not left or right noetherian. However, the results of this paper, as well as [León Sánchez and Sierra 2023, Corollary 5.14], raise the natural question of whether they satisfy the ascending chain condition on two-sided ideals. We close with two results related to this question.

In the next two results, let $\mathfrak{g}$ be a finite-dimensional simple Lie algebra with diagram automorphism σ, and let L be the affine Kac–Moody algebra associated to $\mathfrak{g}$ and σ.

Proposition 7.3. *The algebras $U(\mathfrak{g}[t]^\sigma)$, $U(L)$, and $U(L')$ satisfy the ascending chain condition (ACC) on completely prime ideals.*

Proof. The proof of [Iyudu and Sierra 2020, Proposition 6.4] works in our setting, appealing to Proposition 5.5 and Theorems 1.1 and 3.9 as necessary. We omit the details. □

Remark 7.4. Let $W_+ = t^2\Bbbk[t]\frac{d}{dt}$ be the *positive Witt algebra*. It is shown in[Iyudu and Sierra 2020, Theorem 1.2] that $U(W_+)$ has just-infinite growth. Further, by [Petukhov and Sierra 2020, Theorem 1.5], the symmetric algebra $S(W_+)$ satisfies the ascending chain condition on radical Poisson ideals. These results are supporting evidence for the conjecture [Petukhov and Sierra 2020, Conjecture 1.3] that the enveloping algebra $U(W_+)$ satisfies the ascending chain condition on two-sided ideals.

It is now known [León Sánchez and Sierra 2023] that symmetric algebras of (twisted) loop algebras satisfy the ascending chain condition on radical Poisson ideals, and this can easily be extended to symmetric algebras of affine Kac–Moody algebras. Combining this result with Theorem 1.1 and Proposition 7.3, it is natural to ask if enveloping algebras of affine Kac–Moody algebras satisfy the ascending chain condition

on two-sided ideals. (It is easy to see that these enveloping algebras are not left or right noetherian.) This is the subject of ongoing research. Note that this question is only really interesting for $U_0(L')$.

A ring R is *Hopfian* if R is not isomorphic to any proper quotient R/J (equivalently, any epimorphism from $R \to R$ is an isomorphism). If R satisfies the ascending chain condition on two-sided ideals, then R must be Hopfian. We do not know if enveloping algebras of affine Lie algebras satisfy this ACC, but it is a consequence of our growth results that they are Hopfian. Further, enveloping algebras of current algebras and central quotients of enveloping algebras of affine algebras satisfy the stronger Bassian property, where a ring R is *Bassian* if there is no injection of R into any proper quotient R/J.

Proposition 7.5. *The algebras $U(\mathfrak{g}[t]^\sigma)$, $U_0(L')$, and $U_0(L)$ are Bassian and Hopfian. Further, $U(L)$ and $U(L')$ are Hopfian.*

Proof. This proof is similar to the proof of [Iyudu and Sierra 2020, Proposition 6.5], but as it is fairly brief we give it here in full.

If R has just infinite growth, then $\operatorname{GKdim} R/J < \operatorname{GKdim} R$ for any proper ideal J of R, so R cannot inject in R/J. Thus the Bassian (and thus Hopfian) property for $U(\mathfrak{g}[t]^\sigma)$ follows from Theorem 3.9, and for $U_0(L')$ and $U_0(L)$ it follows from Theorem 1.1(0).

Let U be either $U(L)$ or $U(L')$. To show that U is Hopfian, let f be a surjective endomorphism of U, with kernel J. As $U/J \cong \operatorname{Im}(f) = U$ is torsion-free as a module over $\Bbbk[c]$, the complex

$$0 \to J \otimes_{\Bbbk[c]} \Bbbk(c) \to U \otimes_{\Bbbk[c]} \Bbbk(c) \xrightarrow{f\otimes 1} U \otimes_{\Bbbk[c]} \Bbbk(c) \to 0$$

is exact. Now by Proposition 5.5, we must have $J \otimes_{\Bbbk[c]} \Bbbk(c) = 0$, as otherwise a $\Bbbk(c)$-algebra of finite GK-dimension would surject onto one of infinite GK-dimension. As U is $\Bbbk[c]$-torsion-free (or by Theorem 1.1(2)), $J = 0$. □

Acknowledgements

Biswal is supported, and Sierra is partially supported, by the EPSRC grant EP/T018844/1. We thank the EPSRC for their support.

We are grateful to Lucas Calixto, Vyjayanthi Chari, Travis Scrimshaw, and Sankaran Viswanath for helpful discussions and ideas. We thank the anonymous referees for the useful comments.

References

[Bavula 2020] V. V. Bavula, "The generalized Weyl Poisson algebras and their Poisson simplicity criterion", *Lett. Math. Phys.* **110**:1 (2020), 105–119. MR Zbl

[Carbone et al. 2021] L. Carbone, K. N. Raghavan, B. Ransingh, K. Roy, and S. Viswanath, "π-systems of symmetrizable Kac–Moody algebras", *Lett. Math. Phys.* **111**:1 (2021), art. id. 5. MR Zbl

[Chari 1985] V. Chari, "Annihilators of Verma modules for Kac–Moody Lie algebras", *Invent. Math.* **81**:1 (1985), 47–58. MR Zbl

[Goodearl and Warfield 2004] K. R. Goodearl and R. B. Warfield, Jr., *An introduction to noncommutative Noetherian rings*, 2nd ed., Lond. Math. Soc. Student Texts **61**, Cambridge Univ. Press, 2004. MR Zbl

[Iyudu and Sierra 2020] N. K. Iyudu and S. J. Sierra, "Enveloping algebras with just infinite Gelfand–Kirillov dimension", *Ark. Mat.* **58**:2 (2020), 285–306. MR Zbl

[Kac 1990] V. G. Kac, *Infinite-dimensional Lie algebras*, 3rd ed., Cambridge Univ. Press, 1990. MR Zbl

[Krause and Lenagan 1985] G. R. Krause and T. H. Lenagan, *Growth of algebras and Gelfand–Kirillov dimension*, Res. Notes in Math. **116**, Pitman, Boston, MA, 1985. MR Zbl

[León Sánchez and Sierra 2023] O. León Sánchez and S. J. Sierra, "A Poisson basis theorem for symmetric algebras of infinite-dimensional Lie algebras", *Ark. Mat.* **61**:2 (2023), 375–412. MR Zbl

[Petukhov and Sierra 2020] A. V. Petukhov and S. J. Sierra, "Ideals in the enveloping algebra of the positive Witt algebra", *Algebr. Represent. Theory* **23**:4 (2020), 1569–1599. MR Zbl

[Smith 1976] M. K. Smith, "Universal enveloping algebras with subexponential but not polynomially bounded growth", *Proc. Amer. Math. Soc.* **60** (1976), 22–24. MR Zbl

Communicated by Jason P. Bell
Received 2022-10-21 Revised 2024-05-27 Accepted 2024-07-15

rekha@niser.ac.in *School of Mathematical Sciences, National Institute of Science Education and Research, Bhubaneswar, India*

s.sierra@ed.ac.uk *School of Mathematics, University of Edinburgh, Edinburgh, United Kingdom*

ALGEBRA AND NUMBER THEORY 19:6 (2025)
https://doi.org/10.2140/ant.2025.19.1231

msp

The integral Chow ring of weighted blow-ups

Veronica Arena and Stephen Obinna

Appendix written jointly with Dan Abramovich

We give a formula for the Chow rings of weighted blow-ups. Along the way, we also compute the Chow rings of weighted projective stack bundles, a formula for the Gysin homomorphism of a weighted blow-up, and a generalization of the splitting principle. In addition, in the Appendix we compute the Chern class of a weighted blow-up.

1. Introduction

A short introduction to weighted blow-ups. The blow-up is an important operation that is ubiquitous in algebraic geometry. When working with algebraic stacks, there is a natural generalization of the blow-up, called a weighted blow-up. Weighted blow-ups appear naturally in the study of moduli spaces, for example in [Inchiostro 2022, Theorem 2.6] $\overline{\mathcal{M}}_{1,2}$ is obtained as the weighted blow-up of the weighted projective plane $\mathcal{P}(2,3,4)$ at a point. This is a particular case of [Arena et al. 2023, Theorem 7.3], where $\overline{\mathcal{M}}_{1,n}$ is obtained as the blow-up of the moduli space of pseudostable curves at the cuspidal locus.

See [Quek and Rydh 2021] for a thorough introduction to weighted blow-ups, or [Arena et al. 2023] for a more condensed version. Intuitively, a weighted blow-up is like an ordinary blow-up, but with positive integer weights on the normal directions at each point of the center. Weighted blow-ups preserve many of the properties of (ordinary) blow-ups such as transforming the center into a divisor or being an isomorphism outside of the center.

For example, the blow-up of $\mathbb{A}^d$ at the origin $\{0\}$ with weights $a_1, \dots, a_d$ replaces the origin with the weighted projective stack $\mathcal{P}(a_1, \dots, a_d)$, which is our exceptional divisor. Moreover, it is isomorphic to $\mathbb{A}^d \smallsetminus \{0\}$ outside $\mathcal{P}(a_1, \dots, a_d)$. Formally, the weights are indicated by using a Rees algebra, as illustrated in the following example.

Example. Suppose we wish to blow up the origin $X = \{0\}$ in $Y = \mathbb{A}^2$ with weights 1 in the x-direction and 2 in the y-direction. This is given by

$$\mathrm{Bl}_X Y := \mathrm{Proj}_Y\Big(\bigoplus I_n\Big) = \Big[\Big(\mathrm{Spec}_Y\Big(\bigoplus I_n\Big) \smallsetminus V(I_+)\Big)/\mathbb{G}_m\Big] \to Y,$$

This research is supported in part by funds from BSF grant 2018193 and NSF grant DMS-2100548.
MSC2020: 14E05, 14F43.
Keywords: weighted projective bundles, weighted affine bundles, Chow groups, intersection theory, weighted blow-ups, algebraic stacks, algebraic geometry, Chow rings, projective stack bundles.

where

$$I_0 = k[x, y] \supset I_1 = (x, y) \supset I_2 = (x^2, y) \supset I_3 = (x^3, xy, y^2) \supset \cdots .$$

The weights of x, y are the same as the maximum degree in the graded algebra $\bigoplus I_n$ in which they appear as linear terms, and I_n consists of polynomials in x, y with weight at least n.

We also assume that all weighted blow-ups are regular in the sense of [Quek and Rydh 2021, Definition 5.2.7] or equivalently [Arena et al. 2023, Definition 2.13].

Content of the paper. Let $f : \widetilde{Y} \to Y$ be a regular weighted blow-up of $X \subset Y$ with positive weights $a_1, \ldots, a_d$ and let $\widetilde{X}$ be the exceptional divisor. For most of the paper we will assume X, Y are smooth algebraic spaces over a field of characteristic 0. In Section 7 we will generalize to the case of $\mathscr{X}$, $\mathscr{Y}$ quotient stacks by a linear algebraic group.

Then we have the commutative diagram

$$\begin{array}{ccc} \widetilde{X} & \xrightarrow{j} & \widetilde{Y} \\ \downarrow{\scriptstyle g} & & \downarrow{\scriptstyle f} \\ X & \xrightarrow{i} & Y \end{array}$$

which is not Cartesian, unlike the ordinary blow-up case (an example of this can be found in [Quek and Rydh 2021, Remark 3.2.10]).

In the case of a classical blow-up, a description of the Chow ring $A^*(\widetilde{Y})$ and of its $A^*(Y)$-module structure is given in [Fulton 1998, Exercise 8.3.9] or [Eisenbud and Harris 2016, Proposition 13.12]. The purpose of this paper is to give a similar description for the Chow ring of a weighted blow-up.

We will use the functoriality of Chow rings including pull-backs, pushforwards and the Gysin map $f^!$ with the key property of making the following diagram commute:

$$\begin{array}{ccc} A^*(\widetilde{X}) & \xrightarrow{j_*} & A^*(\widetilde{Y}) \\ {\scriptstyle f^!}\uparrow & & \uparrow{\scriptstyle f^*} \\ A^*(X) & \xrightarrow{i_*} & A^*(Y) \end{array}$$

The formula for the Chow ring will follow from the exact sequence in the theorem below, generalizing the key sequence in [Fulton 1998, Proposition 6.7(e)].

Theorem 6.1 (key sequence). *Let X, Y, $\widetilde{X}$, $\widetilde{Y}$, f be as above. Then we have the exact sequence of Chow groups*

$$A^*(X) \xrightarrow{(f^!, -i_*)} A^*(\widetilde{X}) \oplus A^*(Y) \xrightarrow{j_* + f^*} A^*(\widetilde{Y}) \to 0.$$

Further, if we use rational coefficients, then this becomes a split short exact sequence with g_* left inverse to $(f^!, -i_*)$:

$$0 \to A^*(X, \mathbb{Q}) \xrightarrow{(f^!, -i_*)} A^*(\widetilde{X}, \mathbb{Q}) \oplus A^*(Y, \mathbb{Q}) \xrightarrow{j_* + f^*} A^*(\widetilde{Y}, \mathbb{Q}) \to 0.$$

Note that since our blow-up diagram is not Cartesian, the codomain of $f^!$ is $A^*(X \times_Y \widetilde{Y})$, but $\widetilde{X}$ is the reduction of $X \times_Y \widetilde{Y}$ so we can identify their Chow groups.

Moreover, when working with integral coefficients, the sequence is no longer exact on the left as shown in Example 6.2. Passing to rational coefficients however, allows us to maintain exactness on the left and to define a left inverse of $(f^!, -i_*)$ via g_*. In fact, it is enough to pass to $\mathbb{Z}[1/a_1, \ldots, 1/a_d]$-coefficients.

From the sequence, we can get the following description of $A^*(\widetilde{Y})$.

Theorem 6.4 (Chow ring of a weighted blow-up). *If $\widetilde{Y} \to Y$ is a weighted blow-up of Y at a closed subvariety X, then the Chow ring $A^*(\widetilde{Y})$ is isomorphic **as a group** to the quotient*

$$A^*(\widetilde{Y}) \cong \frac{(A^*(X)[t]) \cdot t \oplus A^*(Y)}{\big(((P(t) - P(0))\alpha, -i_*(\alpha)), \forall \alpha \in A^*(X)\big)},$$

with $P(t) = c_{\mathrm{top}}^{\mathbb{G}_m}(\mathcal{N}_X Y)(t)$ (defined below) and $[\widetilde{X}] = -t$.

The multiplicative structure on $A^*(\widetilde{Y})$ is induced by the multiplicative structures on $A^*(X)$ and $A^*(Y)$ and by the pull-back map in the following way:

$$(0, \beta) \cdot (t, 0) = (i^*(\beta)t, 0).$$

Equivalently $A^*(\widetilde{Y})$ can be expressed as a quotient of the fiber product

$$\frac{A^*(Y) \times_{A^*(X)} A^*(X)[t]}{((i_*\alpha,\, P(t)\alpha), \forall \alpha \in A^*(X))},$$

with $i^* : A^*(Y) \to A^*(X)$ on the left and $A^*(X)[t] \to A^*(X)$ on the right given by evaluating t at 0.

In order to use the key sequence, we need to give a presentation for the Chow ring of the exceptional divisor $\widetilde{X}$.

In the classical case, $\widetilde{X}$ is a projective bundle over X and the Chow ring of a projective bundle can be described via the formula [Eisenbud and Harris 2016, Theorem 9.6]. In the case of a weighted blow-up, the exceptional divisor is a projective stack bundle, i.e., the projectivization of a weighted affine bundle (Definitions 3.2, 3.4). In Section 3 we define the top $\mathbb{G}_m$-equivariant Chern class for a weighted affine bundle E in terms of its homogeneous pieces as

$$c_{\mathrm{top}}^{\mathbb{G}_m}(E) = \prod c_{\mathrm{top}}^{\mathbb{G}_m}(E_i) = \prod_i (c_{n_i}(E_i) + a_i t c_{n_i-1}(E_i) + \cdots + a_i^{n_i} t^{n_i}),$$

and we give a formula for the integral Chow ring of a projective stack bundle (which was proven for rational coefficients in [Mustaţă and Mustaţă 2012, Lemma 2.10(b)]).

Theorem 3.12 (weighted projective bundle formula). *Let E be a weighted, affine bundle over X of rank n. Let $c_{\mathrm{top}}^{\mathbb{G}_m}(E)(t)$ be its $\mathbb{G}_m$-equivariant top Chern class. Then*

$$A^*(\mathcal{P}(E)) \cong \frac{A^*(X)[t]}{c_{\mathrm{top}}^{\mathbb{G}_m}(E)(t)}.$$

Finally, to have a complete description of the exact sequence in Theorem 6.1, we need the appropriate generalization for the excess intersection formula [Fulton 1998, Theorem 6.3]. Unlike the case of an ordinary blow-up, in the weighted blow-up case we don't have an excess bundle and we describe $f^!$ as the multiplication by a difference quotient of the top $\mathbb{G}_m$-equivariant Chern class of the normal bundle.

Theorem 5.5 (weighted key formula). *Let X, Y, $\widetilde{X}$, $\widetilde{Y}$, f be as above. Let us identify $A^*(\widetilde{X}) \cong A^*(X)[t]/P(t)$ with $P(t) = c_{\mathrm{top}}^{\mathbb{G}_m}(\mathcal{N}_X Y)(t)$. Then we have the following formula for the Gysin homomorphism $f^! : A^*(X) \to A^*(\widetilde{X})$:*

$$f^!(\alpha) = \frac{P(t) - P(0)}{t}\alpha.$$

The proof of our formula for the Gysin homomorphism relies on a generalization of the splitting principle, Theorem 4.9, stated in terms of maps to classifying stacks BT, $B\,\mathrm{GL}_{\boldsymbol{n}}$, $BG_{\boldsymbol{a},\boldsymbol{n}}$.

Theorem 4.9 (the splitting principle). *Let $E \to X$ be a weighted affine bundle defined by a map $X \to BG_{\boldsymbol{a},\boldsymbol{n}}$. Let T be the standard maximal torus in $\mathrm{GL}_{\boldsymbol{n}}$. Then the map $X'' \to X$ in the fiber diagram*

$$\begin{array}{ccc} X'' & \longrightarrow & BT \\ \downarrow & & \downarrow \\ X' & \longrightarrow & B\,\mathrm{GL}_{\boldsymbol{n}} \\ \downarrow & & \downarrow \\ X & \longrightarrow & BG_{\boldsymbol{a},\boldsymbol{n}} \end{array}$$

induces an injection of Chow rings $A^*(X) \hookrightarrow A^*(X'')$ via pull-back.

Here $G_{\boldsymbol{a},\boldsymbol{n}}$ is the structure group of the weighted affine bundle E and $\mathrm{GL}_{\boldsymbol{n}}$ is the structure group of its associated weighted vector bundle. Note that the upper square of the diagram is equivalent to the classical splitting principle in [Fulton 1998, page 51] as in [Totaro 2014, Theorem 2.13].

2. Equivariant intersection theory

From now on Y, X will be smooth quasiseparated algebraic spaces, of finite type over a field k of characteristic 0, with $\mathbb{G}_m$ actions. We will also assume the $\mathbb{G}_m$ action is trivial on X.

Let us first recall the following definitions of equivariant Chow groups for a linear algebraic group G.

Definition 2.1. [Edidin and Graham 1998, Definition-Proposition 1] Let Y be a d-dimensional quasiseparated algebraic space of finite type over a field k, together with a G action. Let g be the dimension of G. The i-th G-equivariant Chow group of Y is defined as

$$A_i^G(Y) := A_{i+l-g}(Y \times U/G),$$

where U is an open subspace of an l-dimensional representation, on which G acts freely and whose complement has codimension greater than $d - i$.

In this article we will mostly use this definition in the particular case of a $\mathbb{G}_m$ action. In particular, this leaves us with very convenient choices for representations: V will be an l-dimensional vector space with the standard $\mathbb{G}_m$ action with weight 1 and $U = V \setminus \{0\}$.

Example 2.2. We have $A^*_{\mathbb{G}_m}(X) \cong A^*(X)[t]$. Indeed, since the $\mathbb{G}_m$ action is trivial on X

$$A^i_{\mathbb{G}_m}(X) = A^{\mathbb{G}_m}_{d-i}(X) = A_{d-i+l-1}(X \times U/\mathbb{G}_m) = A^i(X \times \mathbb{P}^{l-1}) = \bigoplus_{k=0}^{i} A^k(X)t^{i-k},$$

with $t = c_1(\mathcal{O}_{\mathbb{P}^{l-1}}(1))$ and the isomorphism follows.

Definition 2.3 [Edidin and Graham 1998, Definition 1]. Let E be a G equivariant vector bundle over Y. The equivariant Chern classes of E are the operators

$$c^G_j : A^G_i(Y) \to A^G_{i-j}(Y), \quad \text{with } c^G_j(E) \cap \alpha = c_j(E \times U/G) \cap \alpha \in A_{i+l-j-g}(Y \times U/G) = A^G_{i-j}(Y).$$

As mentioned in [Molina Rojas and Vistoli 2006, Section 2] many of the standard properties of Chow groups still hold in the equivariant case. Below we collect some that will be used later.

Proposition 2.4. *The following are true*:

(1) *The first Chern class of the tensor product of line bundles is the sum of the first Chern classes of each line bundle*: $c^G_1(L \otimes L') = c^G_1(L) + c^G_1(L')$.

(2) *Let E be a G-equivariant vector bundle over Y and $f : Y' \to Y$ a map such that f^*E has a filtration of G-equivariant vector bundles $f^*E = F_r \supset \cdots \supset F_0 = 0$. Let $E_i = F_i/F_{i-1}$. Then*

$$c^G(f^*E) = \prod_i c^G(E_i).$$

(3) *If Z is a closed G-invariant subscheme of Y, we have the exact sequence*

$$A^*_G(Z) \to A^*_G(Y) \to A^*_G(Y \smallsetminus Z) \to 0.$$

(4) *Let $\pi : E \to Y$ be a G-equivariant vector bundle over Y and $s_0 : Y \to E$ be the zero section. Then the Gysin pull-back map $s_0^* : A^*_G(E) \to A^*_G(Y)$ is an isomorphism equal to the inverse of π^*.*

Proof. We will prove (3). The proofs of the remaining parts are analogous. Let U have dimension l high enough that $A^G_i(Y)$ is defined as $A_{i+l-g}(Y \times U/G)$. Then, also by definition, we have $A^G_i(Z) := A_{i+l-1}(Z \times U/G)$. In particular, $Z \times U/G$ and $Y \times U/G$ are algebraic spaces, and the localization sequence

$$A^*(Z \times U/G) \to A^*(Y \times U/G) \to A^*(Y \times U/G \smallsetminus Z \times U/G) \to 0$$

is exact. Therefore statement (3) holds as well. □

Another proposition, that will be very useful later, is [Molina Rojas and Vistoli 2006, Lemma 2.2], for which we will quote the statement and the proof.

Lemma 2.5 [Molina Rojas and Vistoli 2006, Lemma 2.2]. *Let G be an affine linear group acting on a smooth scheme Y. Let $\pi : E \to Y$ be a G-equivariant vector bundle of rank n. Call $E_0 \subset E$ the complement of the zero section $s : Y \to E$. Then the pull-back homomorphism $\pi|^*_{E_0} : A^*_G(Y) \to A^*_G(E_0)$ is surjective, and its kernel is generated by the top Chern class $c^G_n(E) \in A^n_G(Y)$.*

Proof. Consider the diagram

$$\begin{array}{ccccccc} & & & & A_G^*(Y) & & \\ & \nearrow & & \nearrow & \uparrow{\scriptstyle s^*} & \searrow{\scriptstyle \pi|_{E_0}^*} & \\ A_G^*(Y) & \xrightarrow{s_*} & A_G^*(E) & \longrightarrow & A_G^*(E_0) & \longrightarrow & 0 \end{array}$$

where the bottom is the localization sequence. Since s^* is an isomorphism, inverse to π^*, we see that $\pi|_{E_0}^*$ is surjective with kernel generated by the image of s^*s_*. By the self-intersection formula, s^*s_* is multiplication by $c_n^G(E)$. □

3. Chow groups of weighted projective stack bundles

In this section we will give a formula for the Chow ring of a weighted projective stack bundle. Weighted projective stack bundles appear as the exceptional divisor of a weighted blow-up. We start by computing the Chern classes of weighted affine bundles, and then show we can apply Lemma 2.5 to them. A similar formula for rational coefficients appears in [Mustaţă and Mustaţă 2012, Theorem 2.10(b)].

Definition 3.1. An affine bundle is a smooth affine morphism $E \to X$ such that E is, locally in the smooth topology, isomorphic to $X \times \mathbb{A}^n$.

Definition 3.2 [Quek and Rydh 2021, Definition 2.1.3]. A weighted affine bundle is a $\mathbb{G}_m$-equivariant affine bundle $E \to X$, where locally in the smooth topology $\mathbb{G}_m$ acts linearly on $\mathbb{A}^n$ with positive weights $a_1, \ldots, a_n \in \mathbb{Z}$.

We note in Remark 4.2 that the structure group is special, which means weighted affine bundles over a scheme are Zariski-locally trivial. This is used to apply [Stacks 2005–, Tag 0GUB] in the proof of Corollary 3.11.

It will sometimes be convenient to emphasize the *distinct* weights of a weighted affine bundle. When we do this we will list the distinct weights as $a_1, \ldots, a_r$, and use n_i to refer to the dimension of the subspace of $\mathbb{A}^n$ where the action has weight a_i. In these cases, we will highlight the fact that the weights are distinct in the relevant statements.

Definition 3.3. A weighted vector bundle is a weighted affine bundle whose underlying $\mathbb{G}_m$ space is a vector bundle. Equivalently, it is a weighted affine bundle with linear transition functions.

Notice that our terminology is slightly different from that of [Quek and Rydh 2021]. What they call twisted/untwisted weighted vector bundles, we call weighted affine/vector bundles respectively. Also note that the $\mathbb{G}_m$ action on a weighted vector bundle must preserve the degree. In particular, the bundle splits into homogeneous vector bundles, where $\mathbb{G}_m$ acts with the same degree.

Definition 3.4 [Quek and Rydh 2021, Definition 2.1.5]. A weighted projective stack bundle over X is the stack-theoretic Proj of a graded algebra corresponding to a weighted affine bundle with strictly positive weights. Precisely, if R is a graded algebra such that $E = \mathrm{Spec}_X(R)$ then $\mathrm{Proj}_X(R) = [\mathrm{Spec}_X(R) \smallsetminus V(R_+)/\mathbb{G}_m]$.

3.1. ***Equivariant Chern classes of a weighted line bundle.*** Let us denote by $L \to X$ a line bundle over X with the trivial action. Let us denote by $L^{(a)}$ the same underlying line bundle, endowed with the weight-a $\mathbb{G}_m$ action. This is a notation we will adopt only for Sections 3.1 and 3.2, but abandon later as the weight of the $\mathbb{G}_m$ action will be clear from context.

Some of the following lemmas are likely already known, but are stated and proven for completeness as we couldn't find specific references.

Lemma 3.5. *Let $L^{(a)}$ be a $\mathbb{G}_m$-equivariant line bundle over X with weight a. Then the first equivariant Chern class of $L^{(a)}$ is $c_1^{\mathbb{G}_m}(L^{(a)}) = c_1(L) + at$ via the identification in Example 2.2.*

Proof. We know that $L^{(a)} = L^{(a)} \otimes \mathcal{O}_X = L \otimes \mathcal{O}_X{}^{(a)}$. In particular,

$$c_1^{\mathbb{G}_m}(L^{(a)}) = c_1^{\mathbb{G}_m}(L \otimes \mathcal{O}_X{}^{(a)}) = c_1^{\mathbb{G}_m}(L) + a c_1^{\mathbb{G}_m}(\mathcal{O}_X{}^{(1)}).$$

Now, since the action on L is trivial, $(L \times U)/\mathbb{G}_m = L \times \mathbb{P}^{l-1}$, and since $A^1(X \times \mathbb{P}^{l-1}) = A^1(X) \oplus A^0(X)t$, we get

$$c_1^{\mathbb{G}_m}(L) = c_1(L \times \mathbb{P}^{l-1}) = c_1(L) \in A^1(X).$$

We only need to prove $c_1(\mathcal{O}_X{}^{(1)}) = t$.

Let us consider the projection to a point P, $f : X \to P$. The map defines a graded ring homomorphism $f^* : A^*_{\mathbb{G}_m}(P) \to A^*_{\mathbb{G}_m}(X)$, i.e., a map $f^* : \mathbb{Z}[t] \to A^*(X)[t]$ defined by $1 \mapsto 1$ and $t \mapsto t$.

Now $\mathcal{O}_X{}^{(1)} = f^*(\mathcal{O}_P{}^{(1)})$ and

$$c_1^{\mathbb{G}_m}(\mathcal{O}_X{}^{(1)}) = c_1^{\mathbb{G}_m}(f^*(\mathcal{O}_P{}^{(1)})) = f^* c_1^{\mathbb{G}_m}(\mathcal{O}_P{}^{(1)}).$$

Therefore it is enough to prove $c_1^{\mathbb{G}_m}(\mathcal{O}_P{}^{(1)}) = t$.

By definition $c_1^{\mathbb{G}_m}(\mathcal{O}_P{}^{(1)}) = c_1(\mathcal{O}_P \times U/\mathbb{G}_m)$ as a bundle over $U/\mathbb{G}_m$, with $U/\mathbb{G}_m = \mathbb{A}^2 \setminus (0,0)/\mathbb{G}_m = \mathbb{P}^1$.

Now, a nonzero section $s : U \to U \times \mathcal{O}_P/\mathbb{G}_m$ is given by $(x_0, x_1) \mapsto (x_0, x_1, x_0)$ and intersects the zero section $(x_0, x_1, 0)$ in $x_0 = 0$, which gives us $\mathcal{O}_{\mathbb{P}^1}(1)$, whose first Chern class is t in $A^1(U \times \mathcal{O}_P/\mathbb{G}_m) = A^1(\mathbb{P}^1)$, as desired. □

3.2. ***Equivariant Chern classes of homogeneous bundles.***

Proposition 3.6 (homogeneous bundles admit a splitting with line bundles). *Let $E^{(a)}$ be a rank-n vector bundle over X with $\mathbb{G}_m$ acting homogeneously with weight a on it. Then there exists $f : X' \to X$ such that $f^*E^{(a)}$ has a filtration*

$$f^*E^{(a)} \supset F_n^{(a)} \supset \cdots \supset F_0^{(a)} = 0$$

with $\mathbb{G}_m$-equivariant line bundle quotients $L_j^{(a)} = F_{j+1}^{(a)}/F_j^{(a)}$ and f^* is injective.

Proof. Let us consider the underlying bundle E. Then by the splitting construction [Fulton 1998, page 51] there is a map $f : X' \to X$ with a filtration $f^*E = F_n \supset \cdots \supset F_0 = 0$ with line bundle quotients.

These bundles naturally have the structure of $\mathbb{G}_m$-equivariant vector bundles with weight 1. By replacing the weight-1 action with a weight-a action we get the desired sequence. □

Corollary 3.7. *Let $E^{(a)}$ be a homogeneous $\mathbb{G}_m$-equivariant vector bundle of rank n with weight a, and E the underlying vector bundle endowed with the trivial $\mathbb{G}_m$ action. Then the top equivariant Chern class in $A^*_{\mathbb{G}_m}(X) = A^*(X)[t]$ is*

$$c_n^{\mathbb{G}_m}(E^{(a)}) = c_n(E) + atc_{n-1}(E) + \cdots + a^n t^n.$$

Proof. Let $f : X' \to X$ as in Proposition 3.6. Then $c_n^{\mathbb{G}_m}(f^*E^{(a)}) = \prod_{i=1}^n c_1^{\mathbb{G}_m}(L_i^{(a)})$. By Lemma 3.5 $c_1^{\mathbb{G}_m}(L_i^{(a)}) = c_1(L_i) + at$.

Therefore $c_n^{\mathbb{G}_m}(f^*E^{(a)}) = c_n(f^*E) + atc_{n-1}(f^*E) + \cdots + a^n t^n$. By the injectivity of f^* we are done. □

3.3. *Chern classes of weighted affine bundles.*

Proposition 3.8. *Let E be a weighted affine bundle over X. Let $0 < a_1 < \cdots < a_r$ be the* ***distinct*** *weights of the $\mathbb{G}_m$ action. Then there exist unique subbundles F_i such that*

$$E \supset F_r \supset \cdots \supset F_1 \supset 0,$$

with well-defined quotients $E_i = F_i/F_{i-1}$ which are homogeneous vector bundles with weights a_i.

Proof. Let $E = \mathrm{Spec}_X(R)$ and $\{U_i\}$ be a cover for X such that $E|_{U_i}$ is the trivial bundle. Then we have graded isomorphisms

$$\alpha_i : R|_{U_i} \to \mathcal{O}_{U_i}[x_{i,1}^{(a_1)}, \ldots, x_{i,n_1}^{(a_1)}, x_{i,1}^{(a_2)}, \ldots, x_{i,n_2}^{(a_2)}, \ldots, x_{i,1}^{(a_r)}, \ldots, x_{i,n_r}^{(a_r)}],$$

with $x_{i,1}^{(a_h)}, \ldots, x_{i,n_h}^{(a_h)}$ having weight a_h. A general transition map

$$\alpha_{ij} = \alpha_j|_{U_{ij}} \circ \alpha_i|_{U_{ij}}^{-1} : \mathcal{O}_{U_i j}[x_{i,1}^{(a_1)}, \ldots, x_{i,n_r}^{(a_{n_r})}] \to \mathcal{O}_{U_{ij}}[x_{j,1}^{(a_1)}, \ldots, x_{j,r}^{(a_{n_r})}]$$

will map $x_{i,l}^{(a_h)}$ to a homogeneous polynomial of degree a_h.

Now let $F_k|_{U_i} := \alpha_i^{-1}(\mathcal{O}_{U_i}[x_{i,1}^{(a_1)}, \ldots, x_{i,n_k}^{(a_k)}])$ be the locus where $\mathbb{G}_m$ acts with weights smaller than or equal to a_k. This defines uniquely r subbundles of E. Moreover the quotients F_k/F_{k-1} are well-defined. Indeed, while these are affine bundles, they are locally isomorphic to vector bundles so we can at least take quotients locally. By construction, taking quotients locally gives us bundles consisting only of the weight-a_k pieces of E, and since the lower-degree pieces have been reduced to 0, we are left with linear transition functions making the F_k/F_{k-1} homogeneous vector bundles of weight a_k, as needed. □

Proposition 3.9. *Let E be a weighted affine bundle over X as in Proposition 3.8. Let $N_X E$ be the (nonweighted) normal bundle of X in E. Then, with $X \hookrightarrow E$ the zero section, we have*

$$N_X E \cong E_1 \oplus \cdots \oplus E_r.$$

Proof. Let I be the ideal sheaf of X in E and let α_{ij} be its transition functions as in Proposition 3.8, with each $x_{i,l}^{(a_h)}$ mapped to a homogeneous polynomial of degree a_h.

When computing the transition functions $\bar{\alpha}_{i,j}$ for $N_X E$, we are taking the quotient by I^2; i.e., $\bar{\alpha}_{i,j}$ will preserve only the linear terms of said polynomials and delete the monomials coming from local coordinates where $\mathbb{G}_m$ acts with lower degree.

In particular, local coordinates on which $\mathbb{G}_m$ acts with a certain degree must be mapped to coordinates in the same degree, and the normal bundle splits into $E_1' \oplus E_2' \oplus \cdots \oplus E_r'$, where $\mathbb{G}_m$ acts on the coordinates of E_i' with weight a_i.

But the way we obtained the transition functions for E_i' is equivalent to considering the locus F_i in Proposition 3.8 and quotient by F_{i-1}, so we obtained the desired decomposition. □

Definition 3.10. For an affine bundle E, with E_i as in Propositions 3.8 and 3.9, we define the $\mathbb{G}_m$-equivariant total Chern class of E as $c^{\mathbb{G}_m}(E) = c^{\mathbb{G}_m}(N_X E) = \prod c^{\mathbb{G}_m}(E_i)$.

Corollary 3.11. *Lemma 2.5 also holds in the case where E is a weighted affine bundle.*

Proof. Note that by Definition 3.10 and Proposition 3.9, we have $c_i^G(E) = c_i^G(N_X E)$. The rest of the proof is the same as in Lemma 2.5, noting that $\pi^* : A^*(X) \to A^*(E)$ is still an isomorphism by [Stacks 2005–, Tag 0GUB], and $s^* : A^*(E) \to A^*(X)$ is still its inverse. □

3.4. *The Chow ring of a weighted projective stack bundle.*

Theorem 3.12 (weighted projective bundle formula). *Let E be a weighted, affine bundle over X of rank n. Let $c_{\text{top}}^{\mathbb{G}_m}(E)(t)$ be its $\mathbb{G}_m$-equivariant top Chern class. Then*

$$A^*(\mathcal{P}(E)) \cong \frac{A^*(X)[t]}{c_{\text{top}}^{\mathbb{G}_m}(E)(t)}.$$

Proof. Note that, by definition, $A^*(\mathcal{P}(E)) = A^*([(E \smallsetminus X)/\mathbb{G}_m]) = A^*_{\mathbb{G}_m}(E \smallsetminus X)$, with $E \smallsetminus X$ being E minus the zero section. By Lemma 2.5 we only need to prove that the image of $c_n^{\mathbb{G}_m}(E)$ via the identification in Example 2.2 is $\prod_i (c_{n_i}(E_i) + a_i t c_{n_i-1}(E_i) + \cdots + a_i^{n_i} t^{n_i})$.

By Corollary 3.7, it follows $c_n^{\mathbb{G}_m}(E) = \prod c_{n_i}^{\mathbb{G}_m}(E_i) = \prod_i (c_{n_i}(E_i) + a_i t c_{n_i-1}(E_i) + \cdots + a_i^{n_i} t^{n_i})$ and we are done. □

Below there are some (familiar) special cases.

Example 3.13 (the Chow ring of $\mathcal{P}(a_1, \ldots, a_n)$ [Inchiostro 2022, Lemma 4.7]). Let $\mathcal{P}(a_1, \ldots, a_n)$ be the weighted projective stack with weights $a_1, \ldots, a_n$. We can consider $\mathcal{P}(a_1, \ldots, a_n)$ as a weighted projective bundle over a point. In particular, we have that $\mathcal{P}(a_1, \ldots, a_n)$ splits into n trivial bundles with weights $a_1, \ldots, a_n$. Each of these line bundles will have the first Chern class equal to zero. It follows that

$$A^*(\mathcal{P}(a_1, \ldots, a_n)) = \frac{\mathbb{Z}[t]}{a_1 \cdots a_n t^n}.$$

Example 3.14 (the Chow ring of a classical projective bundle [Eisenbud and Harris 2016, Theorem 9.6]). Let E be a vector bundle of rank n over X in the classical sense. In this case, we have $\mathbb{G}_m$ acting homogeneously on the whole space with weight 1. In particular

$$A^*(\mathcal{P}(E)) = \frac{A^*(X)[t]}{c_n(E) + c_{n-1}(E)t + \cdots + t^n}.$$

Example 3.15. As a toric example, this can be recovered as a consequence of [Iwanari 2009, Theorem 2.2]. The exceptional divisor $\widetilde{X}$ of the weighted blow-up of $X = V(x_1, \dots, x_n)$ in $\mathbb{P}^{m+n}$, with (possibly equal) weights $a_1, \dots, a_n$.

In this case, the Chow ring of the base will be $A^*(X) = A^*(\mathbb{P}^m) = \mathbb{Z}[x]/(x^{m+1})$.

The normal cone of X in $\mathbb{P}^{n+m}$, $N_X\mathbb{P}^{n+m}$, will split into the sum of n copies of $\mathcal{O}_{\mathbb{P}^m}(1)$, on each one of which $\mathbb{G}_m$ will act with weight a_i. We will denote the normal cone together with the $\mathbb{G}_m$ action by $N_{a_1,\dots,a_n}\mathbb{P}^{n+m}$.

Now $c_1(\mathcal{O}_{\mathbb{P}^m}(1)) = x \in \mathbb{Z}[x]/x^{m+1}$. Therefore $c_n^{\mathbb{G}_m}(N_X\mathbb{P}^{n+m}) = \prod(x + a_i t)$ and

$$A^*(\widetilde{X}) = A^*(\mathcal{P}(N_{a_1,\dots,a_n}\mathbb{P}^{n+m})) = \frac{\mathbb{Z}[x,t]}{\left(x^{m+1}, \prod_{i=1}^n (x + a_i t)\right)}.$$

4. The splitting principle

The goal of this section is to prove Theorem 4.9, an analog of the splitting principle. We start by proving some facts about structure groups and classifying spaces. Using those results we construct, for any weighted affine bundle $E \to X$, a map $X' \to X$ that allows us to pull back our affine bundle to a vector bundle $E' \to X'$ with $A^*(X') \cong A^*(X)$.

4.1. ***Structure groups.*** From here on, let $E \to X$ be an affine bundle with fibers isomorphic to an affine space V, $\boldsymbol{n} = (n_1, \dots, n_r)$ be the dimensions of its homogeneous parts and $\boldsymbol{a} = (a_1, \dots, a_r)$ be their *distinct* weights. Let $G_{\boldsymbol{a},\boldsymbol{n}}$ be the group $\mathrm{Aut}_{\mathbb{G}_m}(V)$ of $\mathbb{G}_m$-equivariant automorphisms of V and $\mathrm{GL}_{\boldsymbol{n}} = \prod \mathrm{GL}(n_i)$. Moreover, define $VG_{\boldsymbol{a},\boldsymbol{n}} = [V/G_{\boldsymbol{a},\boldsymbol{n}}]$ and $V\mathrm{GL}_{\boldsymbol{n}} = [V/\mathrm{GL}_{\boldsymbol{n}}]$.

Lemma 4.1. *There is a surjective group homomorphism $G_{\boldsymbol{a},\boldsymbol{n}} \to \mathrm{GL}_{\boldsymbol{n}}$ and a section $\mathrm{GL}_{\boldsymbol{n}} \to G_{\boldsymbol{a},\boldsymbol{n}}$. The kernel of the surjection is a unipotent group $U_{\boldsymbol{a},\boldsymbol{n}}$.*

Proof. In [Quek and Rydh 2021, Section 2.1.7], the authors offer an explicit description of $G_{\boldsymbol{a},\boldsymbol{n}}$. In fact, they decompose $G_{\boldsymbol{a},\boldsymbol{n}}$ in the recursive semidirect product

$$G_{\boldsymbol{a},\boldsymbol{n}} = (\mathrm{GL}_{n_r} \times G_{\boldsymbol{a}',\boldsymbol{n}'}) \ltimes \mathbb{G}_a^{n_r N_r},$$

with $\boldsymbol{a}' = (a_1, \dots, a_{r-1})$, $\boldsymbol{n}' = (n_1, \dots, n_{r-1})$, and N_r the dimension of a_r-th-degree piece of a graded polynomial algebra with free variables $\{x_{i,j} : 1 \le i \le r-1,\ 1 \le j \le n_i\}$, where $x_{i,j}$ is given weight a_i.

Unraveling the recursion gives

$$\begin{aligned} G_{\boldsymbol{a},\boldsymbol{n}} &= \left(\mathrm{GL}_{n_r} \times \cdots \times ((\mathrm{GL}_{n_1} \times \{1\}) \ltimes \mathbb{G}_a^{n_1 N_1}) \ltimes \cdots\right) \ltimes \mathbb{G}_a^{n_r N_r} \\ &= (\cdots (\mathrm{GL}_{\boldsymbol{n}} \ltimes \mathbb{G}_a^{n_1 N_1}) \ltimes \cdots) \ltimes \mathbb{G}_a^{n_r N_r}. \end{aligned}$$

This expression provides us with the desired surjection and section. The kernel is a successive extension of additive groups and so is unipotent. □

Remark 4.2. Let us recall that a linear algebraic group G is special (in the sense of Serre) when every principal G-bundle is Zariski-locally trivial. The description of $G_{\boldsymbol{a},\boldsymbol{n}}$ above implies that the group $G_{\boldsymbol{a},\boldsymbol{n}}$ is special (as noted in [Quek and Rydh 2021, Remark 2.1.8]).

This was useful in Section 3, while working with weighted affine bundles.

4.2. *Lemmas on classifying spaces.*

Lemma 4.3. *Let G be a linear algebraic group which is a semidirect product of groups $G = L \ltimes U$. Then we get the following Cartesian diagram*:

$$\begin{array}{ccc} \{*\} & \longrightarrow & BL \\ \downarrow & & \downarrow \\ BU & \longrightarrow & BG \end{array}$$

Proof. Consider the following diagram:

$$\begin{array}{ccc} BU \times_{BG} BL & \longrightarrow & BL \\ \downarrow & & \downarrow \\ BU & \longrightarrow & BG \\ \downarrow & & \downarrow \\ \{*\} & \longrightarrow & BL \end{array}$$

The bottom square is a Cartesian square, coming from the short exact sequence $1 \to U \to G \to L \to 1$. Indeed an object over a scheme S of the fiber product $* \times_{BL} BG$ is a principal G-bundle $P_G \to S$ with a trivialization of the associated L-bundle $(P_G \times L)/G \cong P_G/U \cong S \times L$. But then the preimage of $S \times \{\mathrm{id}\}$ in P_G is a principal U-bundle, that is, an object of BU. Conversely, if we have a principal U-bundle $P_U \to S$ then we have an associated G-bundle $P_G = (P_U \times G)/U$ whose associated L-bundle $P_L \cong (P_G \times L)/G \cong P_G/U$ has a canonical trivialization: $P_G/U \cong ((P_U \times G)/U)/U$, where the second U acts on the right, giving us $P_L \cong (P_U \times L)/U \cong S \times L$. One can check that these correspondences are inverse to each other.

So we have that large square is Cartesian and $\{*\} \times_{BL} BL = BU \times_{BG} BL$. Further, the composite map on the right is the identity on BL, so that $\{*\} \times_{BL} BL = \{*\}$. □

Corollary 4.4. *$B\,\mathrm{GL}_{\boldsymbol{n}} \to BG_{\boldsymbol{a},\boldsymbol{n}}$ is a $U_{\boldsymbol{a},\boldsymbol{n}}$ bundle*; *specifically we have the following Cartesian diagram*:

$$\begin{array}{ccc} U_{\boldsymbol{a},\boldsymbol{n}} & \longrightarrow & B\,\mathrm{GL}_{\boldsymbol{n}} \\ \downarrow & & \downarrow \\ \{*\} & \longrightarrow & BG_{\boldsymbol{a},\boldsymbol{n}} \end{array}$$

Consequently, given a morphism $X \to BG_{\boldsymbol{a},\boldsymbol{n}}$ the fiber product $X \times_{BG_{\boldsymbol{a},\boldsymbol{n}}} B\,\mathrm{GL}_{\boldsymbol{n}} \to X$ is a $U_{\boldsymbol{a},\boldsymbol{n}}$ bundle.

Proof. Applying Lemma 4.3 to $G_{\boldsymbol{a},\boldsymbol{n}}$, $\mathrm{GL}_{\boldsymbol{n}}$, $U_{\boldsymbol{a},\boldsymbol{n}}$ as in Lemma 4.1 and appending on the left the Cartesian diagram coming from the standard presentation of $BU_{\boldsymbol{a},\boldsymbol{n}}$

$$\begin{array}{ccc} U_{\boldsymbol{a},\boldsymbol{n}} & \longrightarrow & \{*\} \\ \downarrow & & \downarrow \\ \{*\} & \longrightarrow & BU_{\boldsymbol{a},\boldsymbol{n}} \end{array}$$

we get the Cartesian diagram

$$\begin{array}{ccccc} U_{\boldsymbol{a},\boldsymbol{n}} & \longrightarrow & \{*\} & \longrightarrow & B\,\mathrm{GL}_{\boldsymbol{n}} \\ \downarrow & & \downarrow & & \downarrow \\ \{*\} & \longrightarrow & BU_{\boldsymbol{a},\boldsymbol{n}} & \longrightarrow & BG_{\boldsymbol{a},\boldsymbol{n}} \end{array}$$

as desired.

Then $X \times_{BG_{\boldsymbol{a},\boldsymbol{n}}} B\,\mathrm{GL}_{\boldsymbol{n}} \to X$ is the pull-back of a $U_{\boldsymbol{a},\boldsymbol{n}}$ bundle and hence is a $U_{\boldsymbol{a},\boldsymbol{n}}$ bundle. □

Lemma 4.5. *Let L be a subgroup of a group scheme G acting on a scheme V. Then the following diagram is Cartesian*:

$$\begin{array}{ccc} [V/L] & \longrightarrow & BL \\ \downarrow & & \downarrow \\ [V/G] & \longrightarrow & BG \end{array}$$

Proof. An object over a scheme S in $[V/G] \times_{BG} BL$ is a triple (P, Q, α), where $P \to S$ is a G-torsor, together with a G-equivariant map to V

$$\begin{array}{ccc} P & \xrightarrow{\phi} & V \\ \downarrow & & \\ S & & \end{array}$$

$Q \to S$ is an L-torsor, and α is an isomorphism of G-torsors $P \xrightarrow{\alpha} Q \times G/L$.

Given such an object we can construct an object in $[V/L]$ by considering the L-torsor $Q \to S$ together with the map $\psi : Q \to V$ defined as follows:

$$\begin{array}{ccccccc} Q & \longrightarrow & Q \times G/L & \xrightarrow{\alpha^{-1}} & P & \xrightarrow{\phi} & V \\ \downarrow & & & & & & \\ S & & & & & & \end{array}$$

To verify this is indeed an object of $[V/L]$, we need to prove that ψ is L-equivariant. Now, ϕ and α^{-1} are G-equivariant, and in particular L-equivariant. Moreover the quotient map $Q \to Q \times G/L$ maps an element ql to $[ql, e] = [qll^{-1}, le] = [q, l]$. But L acts on $Q \times G/L$ through its inclusion into G; hence $ql \mapsto [q, e]\, l$ as desired.

On the other hand given an L-torsor $Q \to S$ together with an L-equivariant map $\psi : Q \to V$ in $[V/L]$, we can construct the triple $(Q \times G/L, Q, \mathrm{id})$ as an object of $[V/G] \times_{BG} BL$. In order for $Q \times G/L$ to be an object in $[V/G]$, we must equip it with a G-equivariant map $Q \times G/L \to V$ or, equivalently, with a G-equivariant, L-invariant map $Q \times G \to V$. The map $\Phi : Q \times G \to V$ defined by $(q, g) \mapsto \psi(q)g$ is L-invariant with respect to the action $l \cdot (q, g) = (ql, l^{-1}g)$ we are quotienting by; indeed

$$(ql, l^{-1}g) \mapsto \psi(ql)l^{-1}g = \psi(q)ll^{-1}g = \psi(q)g.$$

Moreover Φ is G-equivariant: $\Phi((q, g) \cdot h) = \psi(q)gh = (\psi(q)g) \cdot h$. The verification that the functors defined are indeed inverses is standard and will be omitted. □

4.3. *The splitting principle.*

Proposition 4.6. *Given a weighted affine bundle $E \to X$ (respectively a weighted vector bundle), we have a natural map $X \to BG_{\boldsymbol{a},\boldsymbol{n}}$ (respectively $X \to B\mathrm{GL}_{\boldsymbol{n}}$) such that E is the pull-back of $VG_{\boldsymbol{a},\boldsymbol{n}}$ (respectively $V\mathrm{GL}_{\boldsymbol{n}}$).*

Proof. We prove the result for $VG_{\boldsymbol{a},\boldsymbol{n}}$, as the result for $V\mathrm{GL}_{\boldsymbol{n}}$ is effectively the same, with the obvious modifications. Let us denote by Isom the sheaf of isomorphisms of affine bundles $\mathrm{Isom}_X(E, V \times X)$

(respectively, the sheaf of isomorphisms of weighted vector bundles). By a straightforward application of the definitions, it can be seen that Isom is a principal $G_{\boldsymbol{a},\boldsymbol{n}}$ bundle over X and $\mathrm{Isom} \times_X E \cong V \times_{\{*\}} \mathrm{Isom}$.

In particular we get the following Cartesian diagram:

$$\begin{array}{ccc} \mathrm{Isom} & \longrightarrow & \{*\} \\ \downarrow & & \downarrow \\ X & \longrightarrow & BG_{\boldsymbol{a},\boldsymbol{n}} \end{array}$$

Then we can fit the spaces above in the commutative cube

$$\begin{array}{ccccc} \mathrm{Isom} \times_X E & \longrightarrow & V & & \\ \downarrow & \searrow & \downarrow & \searrow & \\ & & E & \longrightarrow & VG_{\boldsymbol{a},\boldsymbol{n}} \\ \mathrm{Isom} & \longrightarrow & \downarrow & & \downarrow \\ & \searrow & \{*\} & \searrow & \\ & & X & \longrightarrow & BG_{\boldsymbol{a},\boldsymbol{n}} \end{array}$$

where the bottom, back and side squares are fiber squares.

Note that $\mathrm{Isom} \times_X E \to E$ is a principal $G_{\boldsymbol{a},\boldsymbol{n}}$ bundle. Moreover, the action of $G_{\boldsymbol{a},\boldsymbol{n}}$ on Isom gives a $G_{\boldsymbol{a},\boldsymbol{n}}$-equivariant map $\mathrm{Isom} \times_X E \to V$ via the identification of $\mathrm{Isom} \times_X E$ with $V \times \mathrm{Isom}$. This gives us a map of quotient stacks $E \to VG_{\boldsymbol{a},\boldsymbol{n}}$ which makes the top of the cube Cartesian.

It follows that

$$\begin{array}{ccc} E & \longrightarrow & VG_{\boldsymbol{a},\boldsymbol{n}} \\ \downarrow & & \downarrow \\ X & \longrightarrow & BG_{\boldsymbol{a},\boldsymbol{n}} \end{array}$$

is a fiber square, as needed. □

Corollary 4.7. *Let $E \to X$ be a weighted affine bundle, with corresponding map $X \to BG_{\boldsymbol{a},\boldsymbol{n}}$. Then E', the pull-back of E via the map $X' = X \times_{BG_{\boldsymbol{a},\boldsymbol{n}}} \mathrm{GL}_{\boldsymbol{n}} \to X$, is a weighted **vector** bundle.*

Proof. Consider the following diagram:

$$\begin{array}{ccccc} E' & \longrightarrow & X' & & \\ \downarrow & \searrow & \downarrow & \searrow & \\ & & V\mathrm{GL}_{\boldsymbol{n}} & \longrightarrow & B\mathrm{GL}_{\boldsymbol{n}} \\ E & \longrightarrow & \downarrow & & \downarrow \\ & \searrow & X & \searrow & \\ & & VG_{\boldsymbol{a},\boldsymbol{n}} & \longrightarrow & BG_{\boldsymbol{a},\boldsymbol{n}} \end{array}$$

The back and right squares are Cartesian by construction. By Lemma 4.5 and Proposition 4.6 we have that the front and bottom squares are Cartesian. Any such cube with these sides Cartesian is Cartesian, in particular the top square. Since $E' \to X'$ is the pull-back of the vector bundle $V\mathrm{GL}_{\boldsymbol{n}} \to B\mathrm{GL}_{\boldsymbol{n}}$, it is a vector bundle as desired. □

Lemma 4.8. *Let $X' \to X$ be as in Corollary 4.7. Then the pull-back map of Chow rings $A^*(X) \to A^*(X')$ is an isomorphism.*

Proof. By Corollary 4.4, $X' \xrightarrow{\phi} X$ is a $U_{\boldsymbol{a},\boldsymbol{n}}$-bundle and $U_{\boldsymbol{a},\boldsymbol{n}}$ is a unipotent group. In particular, $U_{\boldsymbol{a},\boldsymbol{n}}$ is a successive extension of the additive group $\mathbb{G}_a$ by itself and being a $U_{\boldsymbol{a},\boldsymbol{n}}$-bundle is equivalent to being a succession of affine bundles; hence by [Stacks 2005–, Tag 0GUB] we obtain an isomorphism of Chow rings $\phi^* : A^*(X) \to A^*(X')$. □

Theorem 4.9 (the splitting principle). *Let $E \to X$ be a weighted affine bundle defined by a map $X \to BG_{\boldsymbol{a},\boldsymbol{n}}$. Let T be the standard maximal torus in $\mathrm{GL}_{\boldsymbol{n}}$ and $BT := [*/T]$ its classifying stack. Then the map $X'' \to X$ in the fiber diagram*

$$\begin{array}{ccc} X'' & \longrightarrow & BT \\ \downarrow & & \downarrow \\ X' & \longrightarrow & B\,\mathrm{GL}_{\boldsymbol{n}} \\ \downarrow & & \downarrow \\ X & \longrightarrow & BG_{\boldsymbol{a},\boldsymbol{n}} \end{array}$$

induces an injection of Chow rings $A^*(X) \hookrightarrow A^*(X'')$ via pull-back.

Proof. By the argument in the proof of [Totaro 2014, Theorem 2.13] we have an injection $A^*(X') \hookrightarrow A^*(X'')$. By composing with the isomorphism in Lemma 4.8, we have the desired map. □

5. The Gysin homomorphism induced by a weighted blow-up

The goal for this section is to prove Theorem 5.5, which replaces the excess bundle formula in the case of weighted blow-ups.

The strategy for the proof is to reduce to the special case of the weighted blow-up of $BT = [\{0\}/T]$ in $[\mathbb{A}^d/T]$ induced by zero section, which will be computed in Section 5.3.

The reduction to the special case is performed in two steps: first we reduce to the case of the blow-up of an affine space (Section 5.2), and then we apply the splitting principle Theorem 4.9.

Some caution is needed when defining $f^!$, as we don't always have the needed Cartesian diagram. In Section 5.1 we address the issue as well as setting some notation for the rest of the paper.

5.1. ***Notation.*** Let $\widetilde{Y} \to Y$ be the weighted blow-up of Y centered at X, and let $\widetilde{X}$ be the exceptional divisor.

As observed in [Quek and Rydh 2021, Remark 3.2.10] the commutative square is not always Cartesian

$$\begin{array}{ccc} \widetilde{X} & \longrightarrow & \widetilde{Y} \\ \downarrow & & \downarrow \\ X & \longrightarrow & Y \end{array}$$

and when defining $f^!$ we have to make sure to define it with respect to the fiber square

$$\begin{array}{ccc} X \times_Y \widetilde{Y} & \longrightarrow & \widetilde{Y} \\ \downarrow & & \downarrow \\ X & \longrightarrow & Y \end{array}$$

Moreover we have $\widetilde{X} = (X \times_Y \widetilde{Y})_{\mathrm{red}}$ and the diagram below commutes:

$$\begin{array}{ccccc} \widetilde{X} & \longrightarrow & X \times_Y \widetilde{Y} & \xrightarrow{j} & \widetilde{Y} \\ & \searrow^{g} & \downarrow{\scriptstyle h} & & \downarrow{\scriptstyle f} \\ & & X & \xrightarrow{i} & Y \end{array}$$

When looking at Chow rings though, we have a natural isomorphism

$$(\mathrm{red})_* : A^*(\widetilde{X}) \to A^*(X \times_Y \widetilde{Y})$$

induced by the reduction map $\mathrm{red} : \widetilde{X} \to X \times_Y \widetilde{Y}$.

In particular, it makes sense to talk about $f^!$ as the composition of $(\mathrm{red})_*^{-1} \circ f^!$. Throughout the rest of the paper, we will refer to it simply as $f^!$.

Lemma 5.1. *The map $f^! : A^*(X) \to A^*(\widetilde{X})$ is of the form $f^!(\alpha) = g^*(\alpha) \cdot \gamma$ for some element $\gamma \in A^*(\widetilde{X})$.*

Proof. In a similar fashion to what we did in Proposition 2.4, we will prove the statement by passing through algebraic spaces.

Precisely, let $\widetilde{X}_U = (\mathcal{N}_X Y \smallsetminus 0) \times U/\mathbb{G}_m$ with U open as in Definition 2.1 inducing isomorphisms of Chow groups for $\widetilde{X}$ of the appropriate degree. In fact, if U is chosen large enough we also get the algebraic space $\widetilde{Y}_U$ with analogous induced isomorphisms of Chow groups for $\widetilde{Y}$.

For the appropriate degrees, the induced maps $f_U : \widetilde{Y}_U \to Y$ with $(\tilde{y}, u) \mapsto f(\tilde{y})$ and $g_U : \widetilde{X}_U \to X$ with $(\tilde{x}, u) \mapsto g(\tilde{x})$ will themselves induce group homomorphisms $f_U^! : A^*(X) \to A^*(\widetilde{X}_U)$ and $g_U^* : A^*(X) \to A^*(\widetilde{X}_U)$.

Now $g_U : \widetilde{X}_U \to X$ is a smooth map and by [Fulton 1998, Theorem 17.4.2] we have isomorphisms

$$A^p(\widetilde{X}_U) \cong A^p(\widetilde{X}_U \xrightarrow{\mathrm{id}} \widetilde{X}_U) \xrightarrow{\cdot[g_U]} A^{p-d}(\widetilde{X}_U \to X).$$

In particular, for the degrees on which $A^p(\widetilde{X}_U) = A^p_{\mathbb{G}_m}(\mathcal{N}_X Y \smallsetminus 0) \cong A^p(\widetilde{X})$, we have that $f_U^! = \gamma_U \cdot [g_U] = \gamma_U \cdot g_U^*$ for some $\gamma_U \in A^p(\widetilde{X}_U)$ is equivalent to saying $f^!(\alpha) = \gamma_U \cdot g^*(\alpha)$ for some element $\gamma_U \in A^p(\widetilde{X})$.

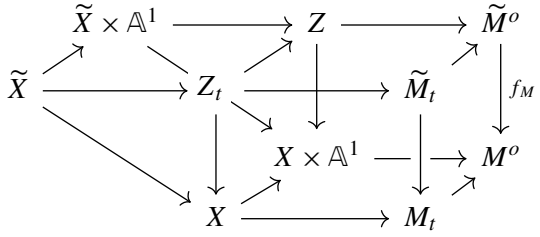
Since the elements γ_U must agree whenever U has high enough dimension, they must coincide. Hence there exists a unique element $\gamma \in A^*(\widetilde{X})$ such that $f^!(\alpha) = \gamma \cdot g^*(\alpha)$. □

5.2. ***Specialization to the weighted normal cone.*** Analogously to [Fulton 1998, Section 5.2], Quek and Rydh [2021, Section 4.3] constructed a deformation to the weighted normal cone of X in Y, which is a weighted affine bundle in our case. We will be using their construction to reduce our argument to the case where Y a weighted affine bundle over X. A similar construction can be found in [Mustaţă and Mustaţă 2012, Section 2.3].

Note that when given a weighted embedding that defines a weighted blow-up of smooth varieties, the weighted normal cone is a weighted affine bundle, which we will denote by $\mathcal{N}_X Y$.

Theorem 5.2. *Let X, Y, $\widetilde{X}$, $\widetilde{Y}$ be as usual. Let $N = \mathcal{N}_X Y$ be the weighted normal affine bundle of X in Y and $f_N : \widetilde{N} \to N$ be the weighted blow-up of the zero section of said bundle, with the same weights as $f : \widetilde{Y} \to Y$. Then the induced maps $f^! : A^*(X) \to A^*(\widetilde{X})$ and $f_N^! : A^*(X) \to A^*(\widetilde{X})$ coincide.*

Proof. Let M^o be the deformation to the weighted normal cone as defined in [Quek and Rydh 2021, Definition 4.3.3] and let $\widetilde{M}^o$ be the weighted blow-up of $X \times \mathbb{A}^1$ in M^o with the same weights as f, i.e., the weighted blow-up induced by the weighted embedding in [loc. cit., Definition 4.3.4,(iv)]. Let M_t and $\widetilde{M}_t$ respectively, be the fibers over t. Let $Z = (X \times \mathbb{A}^1) \times_{M^o} \widetilde{M}^0$. Then we have the following diagram:



By looking at the composition $X \to M_t \to M^o$ in the subdiagram

$$\begin{array}{ccccc} Z_t & \longrightarrow & \widetilde{M}_t & \longrightarrow & \widetilde{M}^o \\ \downarrow & & \downarrow & & \downarrow f_M \\ X & \longrightarrow & M_t & \longrightarrow & M^o \end{array}$$

we see that for $t \neq 0$ we have that $f_M^! : A^*(X) \to A^*(Z_t) = A^*(\widetilde{X})$ is precisely $f^!$, and for $t = 0$ it is precisely $f_N^!$. Now looking at the composition $X \to X \times \mathbb{A}^1 \to M^o$ in the subdiagram

$$\begin{array}{ccccc} Z_t & \longrightarrow & Z & \longrightarrow & \widetilde{M}^o \\ \downarrow & & \downarrow & & \downarrow f_M \\ X & \longrightarrow & X \times \mathbb{A}^1 & \longrightarrow & M^o \end{array}$$

we want to show that $f_M^! : A^*(X) \to A^*(Z_t) = A^*(\widetilde{X})$ is the same for all t.

By [Fulton 1998, Theorem 6.4] the following diagram commutes:

$$\begin{array}{ccc} A^*(Z_t) = A^*(\widetilde{X}) & \xleftarrow{i_t^!} & A^*(Z) = A^*(\widetilde{X} \times \mathbb{A}^1) \\ {\scriptstyle f_M^!}\uparrow & & {\scriptstyle f_M^!}\uparrow \\ A^*(X) & \xleftarrow[i_t^*]{} & A^*(X \times \mathbb{A}^1) \end{array}$$

But the horizontal maps are isomorphisms, inverse to the pull-back along the products with $\mathbb{A}^1$. Since the horizontal maps are independent of t and the map on the right is independent of t, so is the map on the left. □

5.3. ***The special case of*** $[\mathbb{A}^d/T]$**.** Let us now study the particular case of a point in the affine space over the diagonal action of the torus:

$$\begin{array}{ccccc} \mathcal{P}(a_1,\dots,a_d) & \longrightarrow & 0\times_{\mathbb{A}^d}\mathrm{Bl}_{a_1,\dots,a_d}\mathbb{A}^d & \xrightarrow{\ j\ } & \mathrm{Bl}_{a_1,\dots,a_d}\mathbb{A}^d \\ & \searrow^{g} & \downarrow h & & \downarrow f \\ & & 0 & \xrightarrow{\ i\ } & \mathbb{A}^d \end{array}$$

In order to explicitly give a formula for $f^!$ we need presentations for the equivariant Chow rings $A_T^*(-)$.

Now $A_T^*(0)\cong A_T^*(\mathbb{A}^d)\cong\mathbb{Z}[x_1,\dots,x_d]$. Details about $A_T^*(0)$ can be found in [Edidin and Graham 1998] and in [Iwanari 2009] for equivariant Chow rings of toric stacks.

Let us first observe that, since $\mathcal{P}(a_1,\dots,a_d)$ is the reduction of $0\times_{\mathbb{A}^d}\mathrm{Bl}_{a_1,\dots,a_d}\mathbb{A}^d$, there is an isomorphism of Chow rings $A_T^*(0\times_{\mathbb{A}^d}\mathrm{Bl}_{a_1,\dots,a_d}\mathbb{A}^d)\cong A_T^*(\mathcal{P}(a_1,\dots,a_d))$.

Moreover $\mathrm{Bl}_{a_1,\dots,a_d}\mathbb{A}^d$ is a line bundle over $\mathcal{P}(a_1,\dots,a_d)$; in fact it is the total space of $\mathcal{O}_{\mathcal{P}(a_1,\dots,a_d)}(-1)$, and we have the isomorphism $A_T^*(\mathrm{Bl}_{a_1,\dots,a_d}\mathbb{A}^d)\cong A_T^*(\mathcal{P}(a_1,\dots,a_d))$.

We are left with computing $A_T^*(\mathcal{P}(a_1,\dots,a_d))$.

Lemma 5.3. $$A_T^*(\mathcal{P}(a_1,\dots,a_d))\cong\frac{\mathbb{Z}[x_1,\dots,x_d,t]}{P(t)},\quad \textit{where } P(t):=\prod_{i=1}^{d}(x_i+ta_i).$$

Proof. By construction $\mathcal{P}(a_1,\dots,a_d)=[(\mathbb{A}^d\smallsetminus 0)/\mathbb{G}_m]$ and the actions of $\mathbb{G}_m$ and T on $\mathbb{A}^d\smallsetminus 0$ commute. In particular

$$A_T^*(\mathcal{P}(a_1,\dots,a_d))\cong A_T^*([(A^d\smallsetminus 0)/\mathbb{G}_m])\cong A_{T\times\mathbb{G}_m}^*(\mathbb{A}^d\smallsetminus 0).$$

Similarly to the computation above,

$$A_{T\times\mathbb{G}_m}^*(0)\cong A_{T\times\mathbb{G}_m}^*(\mathbb{A}^d)\cong\mathbb{Z}[x_1,\dots,x_d,t],$$

where $x_1,\dots,x_d$ are given by the action of T and t is given by the action of $\mathbb{G}_m$. Finally, the image of the first map in the localization sequence,

$$A_{T\times\mathbb{G}_m}^*(0)\to A_{T\times\mathbb{G}_m}^*(\mathbb{A}^d)\to A_{T\times\mathbb{G}_m}^*(\mathbb{A}^d\smallsetminus 0)\to 0,$$

is generated by $P(t):=\prod_{i=1}^{d}(x_i+ta_i)$. Indeed the top Chern class of the $T\times\mathbb{G}_m$-equivariant bundle splits along each component of $\mathbb{A}^d$. On the i-th component of $\mathbb{A}^d$ the i-th component of T acts with weight 1 and the other components of T act with weight 0, while $\mathbb{G}_m$ acts with weight a_i.

Therefore $A_T^*(\mathcal{P}(a_1,\dots,a_d))$ and $A_T^*(\mathrm{Bl}_{a_1,\dots,a_d}\mathbb{A}^d)$ are both isomorphic to $\mathbb{Z}[x_1,\dots,x_d,t]/P(t)$. □

Theorem 5.4. *Let* $f:[\mathrm{Bl}_{a_1,\dots,a_d}\mathbb{A}^d/T]\to[\mathbb{A}^d/T]$ *be the blow-up of* $[0/T]$ *in* $[\mathbb{A}^d/T]$ *with weights* $a_1,\dots,a_d$*. Then*

$$f^!(1)=\left(\frac{c_{\mathrm{top}}^{\mathbb{G}_m}([\mathbb{A}^n/T])(t)-c_{\mathrm{top}}^{\mathbb{G}_m}([\mathbb{A}^n/T])(0)}{t}\right).$$

Proof. By [Edidin and Graham 1998, Proposition 3] and Lemma 5.1 $f^!$ satisfies $f^*i_* = j_*f^!$, making the following diagram commute:

$$\begin{array}{ccc} A_T^*(\mathcal{P}(a_1,\dots,a_d)) & \xrightarrow{j_*} & A_T^*(\mathrm{Bl}_{a_1,\dots,a_d}\mathbb{A}^d) \\ {\scriptstyle f^!(1)\cdot g^*}\uparrow & & \uparrow{\scriptstyle f^*} \\ A^*(0)_T & \xrightarrow{i_*} & A^*(\mathbb{A}^d)_T \end{array}$$

Since $\mathbb{A}^d$ is a rank-d bundle over 0 and $\mathrm{Bl}_{a_1,\dots,a_d}\mathbb{A}^d$ is the tautological line bundle over $\mathcal{P}(a_1,\dots,a_d)$, the homomorphisms i^* and j^* are isomorphisms of T-equivariant Chow rings, which gives us the following:

$$\begin{array}{ccccc} A_T^*(\mathcal{P}(a_1,\dots,a_d)) & \xrightarrow{j_*} & A_T^*(\mathrm{Bl}_{a_1,\dots,a_d}\mathbb{A}^d) & \xrightarrow{j^*} & A_T^*(\mathcal{P}(a_1,\dots,a_d)) \\ {\scriptstyle f^!(1)\cdot g^*}\uparrow & & \uparrow{\scriptstyle f^*} & & \uparrow{\scriptstyle g^*} \\ A^*(0)_T & \xrightarrow{i_*} & A^*(\mathbb{A}^d)_T & \xrightarrow{i^*} & A^*(0)_T \end{array}$$

Now $i^*i_* : A_T^*(0) \to A_T^*(0)$ is just the multiplication by the top equivariant Chern class of the bundle $\mathbb{A}^d$ over 0. Specifically $i^*i_*(\alpha) = \alpha \cdot x_1 \cdots x_d$.

Similarly, j^*j_* is the image of the top equivariant Chern class of the bundle $\mathrm{Bl}_{a_1,\dots,a_d}\mathbb{A}^d$ over $\mathcal{P}(a_1,\dots,a_d)$. But $\mathrm{Bl}_{a_1,\dots,a_d}\mathbb{A}^d$ is the total space of $\mathcal{O}_{\mathcal{P}(a_1,\dots,a_d)}(-1)$, so we have that j^*j_* is multiplication by $-t$, where $\mathcal{P}(a_1,\dots,a_d)$ has the presentation of Example 3.13. Therefore we must have

$$f^*i_*(\alpha) = \alpha \cdot x_1 \cdots x_d = \alpha \cdot c_{\mathrm{top}}^{\mathbb{G}_m}([\mathbb{A}^d/T])(0) = f^!(1)\alpha(-t) = j_*f^!(\alpha)$$

and since t is not a zero divisor in $A_T^*(\mathcal{P}(a_1,\dots,a_d))$, we must have

$$f^!(1) = \frac{-c_{\mathrm{top}}^{\mathbb{G}_m}([\mathbb{A}^d/T])(0)}{t} = \frac{c_{\mathrm{top}}^{\mathbb{G}_m}([\mathbb{A}^d/T])(t) - c_{\mathrm{top}}^{\mathbb{G}_m}([\mathbb{A}^d/T])(0)}{t},$$

as needed. □

5.4. *A formula for the Gysin homomorphism.*

Theorem 5.5. *Let X, Y, $\widetilde{X}$, $\widetilde{Y}$, f be as usual. Let us identify $A^*(\widetilde{X}) \cong A^*(X)[t]/P(t)$ with $P(t) = c_{\mathrm{top}}^{\mathbb{G}_m}(\mathcal{N}_X Y)(t)$. Then we have the following formula for the Gysin homomorphism $f^! : A^*(X) \to A^*(\widetilde{X})$:*

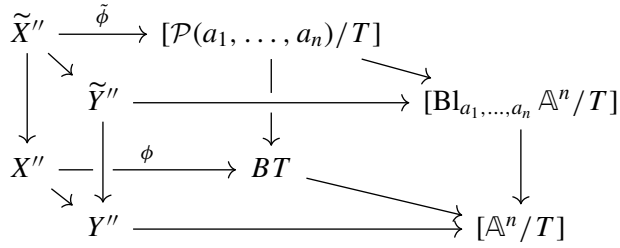
$$f^!(\alpha) = \frac{P(t) - P(0)}{t}\alpha.$$

Proof. With the presentation of $A^*(\widetilde{X})$ above, the map g^* is the natural inclusion of $A^*(X)$ in $A^*(X)[t]/P(t)$ and, by Lemma 5.1 we only need to show

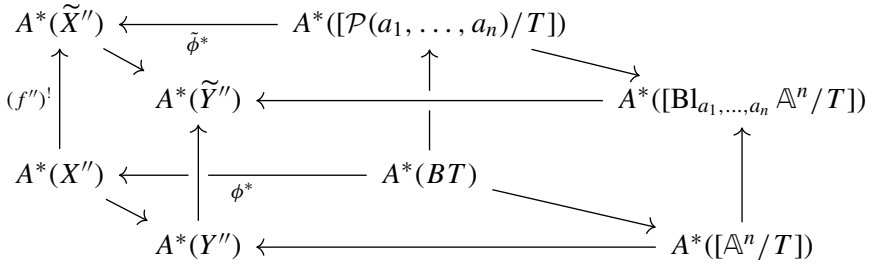
$$f^!(1) = \frac{P(t) - P(0)}{t}.$$

By Theorem 5.2 we can assume that Y is a weighted affine bundle over X. By the splitting principle in Theorem 4.9 it is enough to prove the equality for the pull-back X''. Since weighted blow-ups commute

with base change, the blow-up $f'' : \widetilde{Y}'' \to Y''$ sits in the commutative diagram



which induces the following commutative diagram of Chow groups:



Since equivariant Chern classes commute with pull-backs and Y'' is a vector bundle over X'', by Theorem 5.4 the following holds:

$$(f'')^!(1) = (f'')^!(\phi^*(1)) = \tilde{\phi}^*\left(\frac{c_{\mathrm{top}}^{\mathbb{G}_m}([\mathbb{A}^n/T])(t) - c_{\mathrm{top}}^{\mathbb{G}_m}(\mathbb{A}^n/T)(0)}{t}\right)$$
$$= \left(\frac{c_{\mathrm{top}}^{\mathbb{G}_m}(\tilde{\phi}^*[\mathbb{A}^n/T])(t) - c_{\mathrm{top}}^{\mathbb{G}_m}(\tilde{\phi}^*[\mathbb{A}^n/T])(0)}{t}\right) = \left(\frac{c_{\mathrm{top}}^{\mathbb{G}_m}(Y'')(t) - c_{\mathrm{top}}^{\mathbb{G}_m}(Y'')(0)}{t}\right),$$

which is the desired difference quotient. □

6. The Chow ring of a weighted blow-up

In this section we generalize the key sequence in [Fulton 1998, Proposition 6.7(e)] and then use it to compute a formula for the Chow ring of a weighted blow-up.

Let us recall the notation. Let $f : \widetilde{Y} \to Y$ be the weighted blow-up of Y at X. Let $\widetilde{X}$ be the exceptional divisor. Then we have the commutative diagram

$$\begin{array}{ccc} \widetilde{X} & \xrightarrow{j} & \widetilde{Y} \\ \downarrow{\scriptstyle g} & & \downarrow{\scriptstyle f} \\ X & \xrightarrow{i} & Y \end{array}$$

and the map $f^!$ (computed in 5.5) is

$$f^!(\alpha) = \frac{P(t) - P(0)}{t}\alpha,$$

$P(t) = c_{\mathrm{top}}^{\mathbb{G}_m}(N_X Y)(t)$ and f^* is the Gysin homomorphism defined as in [Vistoli 1989, Definition 3.10].

6.1. *The Grothendieck sequence.*

Theorem 6.1 (key sequence). *Let X, Y, $\widetilde{X}$, $\widetilde{Y}$, f, i, j be as above. Then we have the exact sequence of Chow groups*

$$A^*(X) \xrightarrow{(f^!,-i_*)} A^*(\widetilde{X}) \oplus A^*(Y) \xrightarrow{j_*+f^*} A^*(\widetilde{Y}) \to 0.$$

Further, if we use rational coefficients, then this becomes a split short exact sequence with g_* left inverse to $(f^!, -i_*)$:

$$0 \to A^*(X, \mathbb{Q}) \xrightarrow{(f^!,-i_*)} A^*(\widetilde{X}, \mathbb{Q}) \oplus A^*(Y, \mathbb{Q}) \xrightarrow{j_*+f^*} A^*(\widetilde{Y}, \mathbb{Q}) \to 0.$$

Proof. To prove exactness let us look at the double complex of higher Chow groups given by localization sequence as in [Bloch 1986, Theorem 3.1]

$$\begin{array}{ccccccccc}
\cdots A^*(U,1) & \xrightarrow{\tilde{\delta}_1} & A^*(\widetilde{X}) & \xrightarrow{j_*} & A^*(\widetilde{Y}) & \longrightarrow & A^*(U) & \longrightarrow & 0 \\
\| & & {\scriptstyle f^!}\uparrow & & {\scriptstyle f^*}\uparrow & & \| & & \\
\cdots A^*(U,1) & \xrightarrow{\delta_1} & A^*(X) & \xrightarrow{i_*} & A^*(Y) & \longrightarrow & A^*(U) & \longrightarrow & 0
\end{array}$$

where $U \cong \widetilde{Y} \smallsetminus \widetilde{X} \cong Y \smallsetminus X$.

Since both of the complexes are exact, the total complex

$$\cdots A^*(U,1) \oplus A^*(X) \to A^*(\widetilde{X}) \oplus A^*(Y) \to A^*(\widetilde{Y}) \oplus A^*(U) \to A^*(U) \to 0$$

is also exact.

Let us prove that the map $A^*(\widetilde{X}) \oplus A^*(Y) \xrightarrow{j_*+f^*} A^*(\widetilde{Y})$ is surjective. Let α be any cycle in $A^*(\widetilde{Y})$, $\bar{\alpha}$ be the restriction of α to $A^*(U)$, and $\beta \in A^*(Y)$ be any cycle that restricts to $\bar{\alpha}$ in $A^*(U)$. By commutativity, $\alpha - f^*(\beta)$ restricts to 0 in $A^*(U)$ and must be in the image of j_*. So α is in the image of $j_* + f^*$. Therefore the complex

$$\cdots A^*(U,1) \oplus A^*(X) \to A^*(\widetilde{X}) \oplus A^*(Y) \to A^*(\widetilde{Y}) \to 0$$

is still exact.

Moreover, the image of the map $A^*(U,1) \oplus A^*(X) \xrightarrow{(-\tilde{\delta}_1+f^!,-i_*)} A^*(\widetilde{X}) \oplus A^*(Y)$ coincides with the one of $A^*(X) \xrightarrow{(f^!,-i_*)} A^*(\widetilde{X}) \oplus A^*(Y)$. Indeed, let $(\tilde{x}, y) = (-\tilde{\delta}_1(u) + f^!(x), -i_*(x))$ and let $x' = x + \delta_1(u)$. Then $f^!(x') = f^!(x) - f^!(\delta_1(u)) = f^!(x) - \tilde{\delta}_1(u) = \tilde{x}$ and $i_*(x') = i_*(x) - i_*(\delta_1(u)) = i_*(x) = y$.

It follows that $\ker(j_* + f^*) = \mathrm{Im}(f^!, -i_*)$ and that

$$A^*(X) \xrightarrow{(f^!,-i_*)} A^*(\widetilde{X}) \oplus A^*(Y) \xrightarrow{j_*+f^*} A^*(\widetilde{Y}) \to 0$$

is exact.

Lastly, if we use rational coefficients then there is a left inverse of $(f^!, -i^*)$ given by $(\alpha, \beta) \mapsto g_*(\alpha)$. Indeed, let $x \in A^*(X)$. Then $g_*(f^!(x)) = g_*(\gamma \cdot g^*(x)) = g_*(\gamma) \cdot x$, with γ the difference quotient $(P(t) - P(0))/t$ as in Theorem 5.5. Now γ is a degree-$(n-1)$ polynomial in t, of which only the leading term $a_1 \cdots a_n t^{n-1}$ will survive the pushforward. We only need to show that $g_*(t^{n-1}) = 1/(a_1 \cdots a_n)$.

It is enough to verify this when X is a point and $\widetilde{X}$ is the weighted projective stack $\mathcal{P}(a_1, \ldots, a_n)$. Notice that $a_i t = x_i$, where the x_i are the fundamental classes of the usual coordinate divisors, so $a_1 \cdots a_{n-1} t^{n-1} = x_1 \cdots x_{n-1}$, which is the fundamental class of a stacky point isomorphic to $B\mu_{a_n}$ and so pushes forward to $1/a_n$; thus $g_*(t^{n-1}) = 1/(a_1 \cdots a_n)$. □

Example 6.2. To see why the sequence with integer coefficients is not short exact, let us consider X an elliptic curve in $Y = \mathbb{P}^2$. Let $\widetilde{Y}$ be the blow-up of Y at X with weight 2. Let $P, Q \in X$ be distinct points of order 2 and consider the difference $[P] - [Q] \in A^*(X)$. When pushed forward to $A^*(Y)$ via i_*, all points are rationally equivalent; hence $i_*([P] - [Q]) = 0$. On the other hand, $f^!$ is multiplication by 2, so $f^!([P]) = f^!([Q]) = 0$. But $[P] - [Q]$ is nonzero, so $(f^!, -i_*)$ is not injective.

Remark 6.3. Note that, when looking at the double complex in the proof of Theorem 6.1 and taking the total complex, one defines a long exact sequence of higher Chow groups. The isomorphisms of higher Chow groups

$$\alpha_i : A^*(\widetilde{Y} \setminus \widetilde{X}, i) \to A^*(Y \setminus X, i)$$

allow us to delete $A^*(\widetilde{Y} \setminus \widetilde{X})$ and $A^*(Y \setminus X)$ from the complex via a diagram chase analogous to the ones in the proof. Then we can obtain the long exact sequence

$$\cdots \to A^*(X, i) \to A^*(\widetilde{X}, i) \oplus A^*(Y, i) \to A^*(\widetilde{Y}, i) \to A^*(X, i-1) \to \cdots .$$

6.2. *The Chow ring of a weighted blow-up.*

Theorem 6.4 (Chow ring of a weighted blow-up). *If $\widetilde{Y} \to Y$ is a weighted blow-up of Y at a closed subvariety X, then the Chow ring $A^*(\widetilde{Y})$ is isomorphic* ***as a group*** *to the quotient*

$$A^*(\widetilde{Y}) \cong \frac{(A^*(X)[t]) \cdot t \oplus A^*(Y)}{(((P(t) - P(0))\alpha, -i_*(\alpha)), \forall \alpha \in A^*(X))},$$

with $P(t) = c_{\mathrm{top}}^{\mathbb{G}_m}(\mathcal{N}_X Y)(t)$ and $[\widetilde{X}] = -t$.

The multiplicative structure on $A^*(\widetilde{Y})$ is induced by the multiplicative structures on $A^*(X)$ and $A^*(Y)$ and by the pull-back map in the following way:

$$(0, \beta) \cdot (t, 0) = (i^*(\beta)t, 0).$$

Equivalently $A^*(\widetilde{Y})$ can be expressed as a quotient of the fiber product

$$\frac{A^*(Y) \times_{A^*(X)} A^*(X)[t]}{((i_*\alpha, P(t)\alpha), \forall \alpha \in A^*(X))},$$

with $i^* : A^*(Y) \to A^*(X)$ on the left and $A^*(X)[t] \to A^*(X)$ on the right given by evaluating t at 0.

Proof. The exact sequence in Theorem 6.1 gives us an isomorphism of *groups*

$$A^*(\widetilde{Y}) \cong \frac{A^*(\widetilde{X}) \oplus A^*(Y)}{((f^!(\alpha), -i_*(\alpha)), \forall \alpha \in A^*(X))}.$$

If we use Theorem 3.12 to rewrite $A^*(\widetilde{X})$ and also add an explicit factor of $[\widetilde{X}]$ to represent how $A^*(\widetilde{X})$ is mapped into $A^*(\widetilde{Y})$, then as a *group* we can rewrite $A^*(\widetilde{Y})$ as

$$\frac{((A^*(X)[t])\cdot[\widetilde{X}])\oplus A^*(Y)}{((c_{\mathrm{top}}^{\mathbb{G}_m}(\mathcal{N}_X Y)(t)[\widetilde{X}],0),(f^!(\alpha)[\widetilde{X}],-i^*(\alpha))\forall\alpha\in A^*(X))}.$$

Notice that $t[\widetilde{X}]=-[\widetilde{X}]^2$ (since t is the class of $\mathcal{O}_{\widetilde{X}}(1)$) so that there is an isomorphism between the ring presented above and the ring presented without the symbol $[\widetilde{X}]$ given by replacing $[\widetilde{X}]$ with $-t$.

Now we need to determine the ring structure. Since much of the ring structure is inherited from that of $A^*(Y)$ and $A^*(\widetilde{X})$, what remains is just to determine how to multiply elements coming from $A^*(Y)$ with those coming from $A^*(\widetilde{X})$. Consider the usual commutative square:

$$\begin{array}{ccc}\widetilde{X} & \to & \widetilde{Y}\\ \downarrow & & \downarrow\\ X & \to & Y\end{array}$$

Intersecting some class $\beta\in A^*(Y)$ with $\widetilde{X}$ amounts to pulling it back to $A^*(\widetilde{X})$. By commutativity we have $g^*(i^*(\beta))=j^*(f^*(\beta))$ and by pushforward we obtain $(0,\beta)\times(t,0)=(i^*(\beta)t,0)$.

Finally, notice also that $c_{\mathrm{top}}^{\mathbb{G}_m}(\mathcal{N}_X Y)(t)[\widetilde{X}]$ is now redundant, as for $\alpha=1$ we have

$$\begin{aligned}(t)(f^!(\alpha)(-t)-i_*(\alpha)) &= (t)\left(\frac{c_{\mathrm{top}}^{\mathbb{G}_m}(\mathcal{N}_X Y)(t)-c_{\mathrm{top}}(\mathcal{N}_X Y)}{t}(-t)-i_*(1)\right)\\ &= t(-c_{\mathrm{top}}^{\mathbb{G}_m}(\mathcal{N}_X Y)(t)+c_{\mathrm{top}}(\mathcal{N}_X Y)-i_*(1))=(-t)\cdot c_{\mathrm{top}}^{\mathbb{G}_m}(\mathcal{N}_X Y)(t).\end{aligned}$$

The last equality comes from $t\cdot i_*(1)=g^*(i^*(i_*(1)))=c_{\mathrm{top}}(\mathcal{N}_X Y)$.

Putting everything together, we have that $A^*(\widetilde{Y})$ is the group

$$A^*(\widetilde{Y})\cong\frac{(A^*(X)[t])\cdot t\oplus A^*(Y)}{(((P(t)-P(0))\alpha,-i_*(\alpha)),\forall\alpha\in A^*(X))},$$

with the desired multiplication. □

Corollary 6.5. *If $i^*:A^*(Y)\to A^*(X)$ is surjective, then this formula simplifies to resemble a formula of Keel* [1992, Theorem 1, page 571]

$$A^*(\widetilde{Y})\cong\frac{A^*(Y)[t]}{(t\cdot\ker(i^*),\,Q(t))},$$

where $Q(t)=c_{\mathrm{top}}^{\mathbb{G}_m}(\mathcal{N}_X Y)(t)-c_{\mathrm{top}}^{\mathbb{G}_m}(\mathcal{N}_X Y)(0)+[X]$.

Proof. We first prove that we can suppress $A^*(X)$ from the presentation in Theorem 6.4, i.e., that the elements of the form $(\alpha\cdot t,0)$ with $\alpha\in A^*(X)$ can be described as pairs of the form $(0,\beta)\cdot(t,0)$ for some $\beta\in A^*(Y)$. Let β such that $i^*(\beta)=\alpha$. By the multiplicative-structure condition of Theorem 6.4 gives

$$(\alpha\cdot t,0)=(i^*(\beta)\cdot t,0)=(t,0)$$

as desired.

In particular, the condition $t \cdot (\beta - i^*(\beta)) \forall \beta \in A^*(Y)$ reduces to $t \cdot \ker(i^*)$.

Finally, $t \cdot f^!(\alpha) + i_*(\alpha) = (c_{\text{top}}^{\mathbb{G}_m}(\mathcal{N}_X Y)(t) - c_{\text{top}}(\mathcal{N}_X Y))\alpha + i_*(\alpha)$ and $i_*(\alpha) = i_*(i^*(\beta)) = [X] \cdot \beta$ for some $\beta \in A^*(Y)$. So $t \cdot f^! + i_*$ is multiplication by $(c_{\text{top}}^{\mathbb{G}_m}(\mathcal{N}_X Y)(t) - c_{\text{top}}(\mathcal{N}_X Y) + [X])$, which is precisely the $Q(t)$ desired. □

6.3. ***An example*: *the Chow ring of* $\overline{\mathcal{M}}_{1,2}$.** The Chow ring $A^*(\overline{\mathcal{M}}_{1,2})$ has been computed in [Di Lorenzo et al. 2024; Inchiostro 2022]. The latter uses the construction of $\overline{\mathcal{M}}_{1,2}$ as the weighted blow-up of $\mathcal{P}(2, 3, 4)$. We give yet another computation of the ring, using the same blow-up construction.

We start by recalling the following:

Theorem 6.6 [Inchiostro 2022, Theorem 2.6]. *There exists an isomorphism* $\overline{\mathcal{M}}_{1,2} \cong \text{Bl}_Z^{(4,6)} \mathcal{P}(2, 3, 4)$, *where* $\text{Bl}_Z^{(4,6)} \mathcal{P}(2, 3, 4)$ *is the weighted blow-up of the point* $Z = [s^2 : s^3 : 0]$ *in* $\mathcal{P}(2, 3, 4)$ *with weights* $(4, 6)$.

Given this, $A^*(\overline{\mathcal{M}}_{1,2})$ becomes a straightforward computation,

Proposition 6.7 [Inchiostro 2022, Theorem 4.12].

$$A^*(\overline{\mathcal{M}}_{1,2}) \cong \frac{\mathbb{Z}[y, t]}{(ty, 24(t^2 + y^2))}.$$

Proof. First, since $i : Z \to \mathcal{P}(2, 3, 4)$ is just the inclusion of a point, we have that i is surjective, and since $A^*(\mathcal{P}(2, 3, 4)) \cong \mathbb{Z}[y]/(24y^3)$ we know the kernel is (y). By Corollary 6.5 we then have

$$A^*(\overline{\mathcal{M}}_{1,2}) \cong \frac{A^*(\mathcal{P}(2, 3, 4))[t]}{(ty,\, Q(t))} \cong \frac{\mathbb{Z}[y, t]}{(24y^3, ty,\, Q(t))},$$

where $Q(t)$ restricts to $c_{\text{top}}^{\mathbb{G}_m}(\mathcal{N}_Z Y)$ and has constant term $[Z]$. As $\mathcal{N}_Z Y$ splits into trivial line bundles, we see $c_{\text{top}}^{\mathbb{G}_m}(\mathcal{N}_Z Y) = (4t)(6t)$. Writing $Z = V(x_3) \cap V(x_1^3 - x_2^2)$, we see $[Z] = (4y)(6y)$, so $Q(t) = 24t^2 + 24y^2$. Lastly, the term $24y^3$ is now redundant and we have

$$A^*(\overline{\mathcal{M}}_{1,2}) \cong \frac{\mathbb{Z}[y, t]}{(ty, 24(t^2 + y^2))}.$$ □

7. Generalization to quotient stacks

Let us now consider the case of $\mathscr{Y} = [Y/G]$, where Y is an algebraic space and G is a linear algebraic group; hence it is possible to define the G-equivariant Chow ring $A_G^*(Y)$ as in [Edidin and Graham 1998].

A weighted embedding of $\mathscr{X}$ in $\mathscr{Y}$ defines a weighted embedding of X in Y via pull-back and, since the quotient maps are smooth, we have $\widetilde{\mathscr{Y}} \cong [\widetilde{Y}/G]$ and $\widetilde{\mathscr{X}} \cong [\widetilde{X}/G]$.

Theorem 7.1. *The theorems in Sections 3, 5, and 6 hold for* $f : \widetilde{\mathscr{Y}} \to \mathscr{Y}$ *weighted blow-up of* $\mathscr{Y}$ *at* $\mathscr{X}$.

Proof. Let us prove that $A^*(\widetilde{\mathscr{X}}) \cong A^*(\mathscr{X})[t]/P(t)$ as in Theorem 3.12; the proof for the other results will be almost identical.

For any p let U be as in Definition 2.1 of dimension high enough such that $A^q(X_U) \cong A_G^q(X) \cong A^q(\mathscr{X})$ with $X_U := X \times U/G$, up to degree p.

Since X_U is an algebraic space, by Theorem 3.12

$$A^*(\widetilde{X}_U) \cong A^*(X_U)[t]/P_U(t).$$

Note that $P_U(t) = c_{\mathrm{top}}^{\mathbb{G}_m}(\mathcal{N}_{X_U} Y_U)$ is the pull-back of $P(t) = c_{\mathrm{top}}^{\mathbb{G}_m}(\mathcal{N}_{\mathscr{X}} \mathscr{Y})$, which is a finite-degree polynomial. In particular for large enough p, $P_U(t)$ does not depend on U and it is exactly $P(t)$.

Since $\widetilde{X}_U$ is open inside a vector bundle, we have isomorphisms $A^q(\widetilde{X}_U) \cong A^q(\widetilde{\mathscr{X}})$ up to degree p. Since Theorem 3.12 holds up to degree p, for any p we have the desired isomorphism of Chow rings. □

Appendix: Chern class of weighted blow-up

A1. ***The goal.*** Consider a smooth subvariety X of a smooth variety Y, with blow-up $\widetilde{Y}$ and exceptional divisor, as in the following standard diagram:

$$\begin{array}{ccc} \widetilde{X} & \xrightarrow{j} & \widetilde{Y} \\ \downarrow{\scriptstyle g} & & \downarrow{\scriptstyle f} \\ X & \xrightarrow{i} & Y \end{array}$$

Fulton [1998, Theorem 15.4] provided a formula for the total Chern class $c(\widetilde{Y}) := c(T_{\widetilde{Y}})$ in terms of the blow-up data. The purpose of this note is to revisit that formula and generalize it to the case of a weighted blow-up. Since smoothness is important in these considerations, our weighted blow-ups are always stack-theoretic.

A2. ***Setup and formula.*** In our setup, X and Y are still smooth varieties, and X is the support of a weighted center with weighted normal bundle N of rank $d = \operatorname{codim}(X \subset Y)$. The weighted normal bundle is a weighted affine bundle with total Chern class we denote by $c(N) \in A^*(X)$ and total $\mathbb{G}_m$-equivariant Chern class $c^{\mathbb{G}_m}(N) = Q(t) \in A^*_{\mathbb{G}_m}(X) = A^*(X)[t]$, where t is the equivariant parameter corresponding to the standard character of $\mathbb{G}_m$. In particular $Q(0) = c(N)$.

We recall from Theorem 6.4 in the main text that

$$A^*(\widetilde{Y}) = (A^*(Y) \oplus t A^*(\widetilde{X}))/I,$$

where

$$I = \big(i_*(\alpha) \oplus -(Q(t) - Q(0))\alpha \mid \alpha \in A^*(X)\big).$$

We denote by

$$q : A^*(Y) \oplus t A^*(X)[t] \to (A^*(Y) \oplus t A^*(\widetilde{X}))/I = A^*(\widetilde{Y})$$

the natural quotient map.

Theorem A.1. *We have*

$$\frac{c(\widetilde{Y})}{f^* c(Y)} = q\left(\frac{(1-t)Q(t)}{Q(0)}\right).$$

We note that the right-hand side is of the form $q(1 \oplus t \cdot R(t))$, with some $R(t) \in A^*(X)[\![t]\!]$.

The formula was proved for Chow groups with rational coefficients by Anca and Andrei Mustaţa [2012, Proposition 2.12]. Our proof in essence verifies that their arguments carry over integrally.

A3. ***Approach.*** Our approach combines the equivariant methods used in the main text to study and compute Chow rings of weighted projective stack bundles and weighted blow-ups, combined with ideas in Aluffi's paper [2010] and lecture [2011], especially the user-friendly presentation of the formula in Aluffi's lecture. While Aluffi provides a proof only for complete intersections, the methods of Theorem 6.4 allow us to reduce the general case to a situation where Aluffi's proof applies.

A3.1. *The quotient class.* One first notes that the class $c(\widetilde{Y})/(f^*c(Y))$ appearing on the left-hand side has properties enabling flexible treatment:

Lemma A.2. *The class $c(\widetilde{Y})/(f^*c(Y))$ is of the form $q(1 \oplus t \cdot R(t))$, with some $R(t) \in A^*(X)[\![t]\!]$, and is functorial for smooth morphisms $Y' \to Y$ and closed embeddings $Y' \to Y$ that meet X transversely.*

Proof. To see that it has this form, consider the localization sequence

$$A^*(\widetilde{X}) \to A^*(\widetilde{Y}) \to A^*(\widetilde{Y} \smallsetminus \widetilde{X}) \to 0.$$

Since $c(\widetilde{Y})$ and $f^*c(Y)$ must pull back to the same class on $A^*(\widetilde{Y} \smallsetminus \widetilde{X})$, and their ratio pulls back to 1. In particular we see that $c(\widetilde{Y})/(f^*c(Y)) - 1$ must be in the image of $A^*(\widetilde{X})$, which means $c(\widetilde{Y})/(f^*c(Y))$ is of the desired form.

To see functoriality, consider the following diagram:

$$\begin{array}{ccc} \widetilde{Y}' & \xrightarrow{\tilde{h}} & \widetilde{Y} \\ \downarrow{\scriptstyle f'} & & \downarrow{\scriptstyle f} \\ Y' & \xrightarrow{h} & Y \end{array}$$

We must show

$$\frac{c(\widetilde{Y}')}{f'^*c(Y')} = \tilde{h}^* \frac{c(\widetilde{Y})}{f^*c(Y)},$$

but this is equivalent to

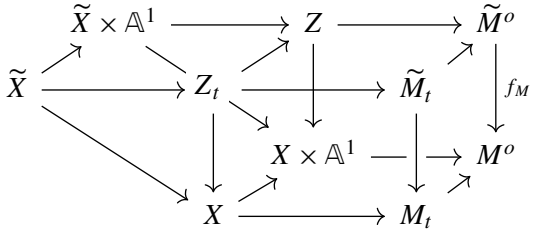
$$\frac{c(\widetilde{Y}')}{\tilde{h}^*c(\widetilde{Y})} = f'^* \frac{c(Y')}{h^*c(Y)}.$$

This is true when h is smooth because the relative tangent bundle of h is compatible with pull-back under f, and true when h is a closed embedding since the normal bundle of h is compatible with pull-back under f. □

A3.2. *Degeneration to the weighted normal bundle.* Applying the lemma to the degeneration to the weighted normal bundle we obtain:

Lemma A.3. *It suffices to prove the theorem, namely to compute $R(t)$ and $c(\widetilde{Y})/(f^*c(Y))$, when $Y = \mathcal{N}_X Y$.*

Proof. Recall the diagram from Theorem 5.2



where $M_t \cong Y$ for $t \neq 0$ and $M_0 = \mathcal{N}_X Y$.

By the previous lemma, the expression $c(\widetilde{M}_t)/(f^* c(M_t))$ can be pulled back from $\widetilde{M}^o$ along the embedding corresponding to t and is determined by a class on $\widetilde{X}$. However, neither $\widetilde{X}$ nor $\widetilde{M}^o$ depend on t so it is enough to compute things when $t = 0$, that is, for $\mathcal{N}_X Y$. □

A3.3. *The universal case.* By Theorem 4.9, the homomorphism $A^*(BG_{\boldsymbol{a},\boldsymbol{n}}) \to A^*(BT)$ is injective. Therefore:

Lemma A.4. *It suffices to prove the theorem when $X = BT$ and $Y = [V/T]$. Equivalently, it suffices to prove the theorem T-equivariantly when X is a point, the origin on $Y = \mathbb{A}^d$.*

Proof. This follows from functoriality and Theorem 4.9 since the maps $X \to BG_{\boldsymbol{a},\boldsymbol{n}}$ and $BT \to BG_{\boldsymbol{a},\boldsymbol{n}}$ are smooth. □

A3.4. *The toric case of affine space.* Finally, let X be the origin of $Y = \mathbb{A}^d$. Let

$$A_T^*(0) \cong A_T^*(\mathbb{A}^d) \cong \mathbb{Z}[x_1, \ldots, x_d]$$

and

$$A_T^*(\mathcal{P}(a_1, \ldots, a_d)) \cong A_T^*(\mathrm{Bl}_{(a_1,\ldots,a_d)}\, \mathbb{A}^d) \cong \frac{\mathbb{Z}[x_1, \ldots, x_d, t]}{\left(\prod (x_i + a_i t)\right)}$$

as in Section 5.3.

Proposition A.5. *We have $c^T(Y) = Q(0)$ and $c^T(\widetilde{Y}) = q((1-t)Q(t))$.*

Proof. This is essentially the same argument as [Aluffi 2006, Theorem 4.2].

Let $D = \Sigma \widetilde{X}_i + \widetilde{X}$ be the sum of all the irreducible toric divisors on $\widetilde{Y}$. By repeating the argument in [Fulton 1993, Proposition p. 87], we have the exact sequence

$$0 \to \Omega^1_{\widetilde{Y}} \to \Omega^1_{\widetilde{Y}}(\log D) \to \left(\bigoplus_{i=1}^{d} \mathcal{O}_{X_i}\right) \oplus \mathcal{O}_{\widetilde{X}} \to 0,$$

and $\Omega^1_{\widetilde{Y}}(\log D)$ is trivial. By Whitney's formula,

$$c^T(\Omega^1_{\widetilde{Y}}) = \frac{1}{c^T(\mathcal{O}_{\widetilde{X}}) c^T\left(\bigoplus \mathcal{O}_{\widetilde{X}_i}\right)} = (1+t) \prod_i (1 - a_i t - x_i).$$

By taking the dual we obtain

$$c^T(T\widetilde{Y}) = (1-t)\prod_i (1 + a_i t + x_i) = Q(t).$$

The same argument works to prove $c^T(Y) = Q(0)$. □

The theorem follows.

Acknowledgements

We would like to thank Dan Abramovich for his invaluable help and guidance, as well as Jarod Alper, Paolo Aluffi, Martin Bishop, Samir Canning, Andrea Di Lorenzo, Giovanni Inchiostro, Patrick Jefferson, Michele Pernice and Ming Hao Quek for insightful conversations. We would like to thank the referees and Melody Chan, Brendan Hassett, Eric Larson and Isabel Vogt for their precious advice. We thank Brown University for the support received during the development of this research.

References

[Aluffi 2006] P. Aluffi, "Classes de Chern des variétés singulières, revisitées", *C. R. Math. Acad. Sci. Paris* **342**:6 (2006), 405–410. MR Zbl

[Aluffi 2010] P. Aluffi, "Chern classes of blow-ups", *Math. Proc. Cambridge Philos. Soc.* **148**:2 (2010), 227–242. MR Zbl

[Aluffi 2011] P. Aluffi, "Chern classes of blow-ups", lecture video, Worldwide Center of Math., 2011, available at https://youtu.be/oOmEWwJbJJg?si=AXfCq8if3I7y4LoN.

[Arena et al. 2023] V. Arena, A. Di Lorenzo, G. Inchiostro, S. Mathur, S. Obinna, and M. Pernice, "A criterion for smooth weighted blow-downs", preprint, 2023. arXiv 2310.15076

[Bloch 1986] S. Bloch, "Algebraic cycles and higher K-theory", *Adv. Math.* **61**:3 (1986), 267–304. MR Zbl

[Di Lorenzo et al. 2024] A. Di Lorenzo, M. Pernice, and A. Vistoli, "Stable cuspidal curves and the integral Chow ring of $\bar{\mathscr{M}}_{2,1}$", *Geom. Topol.* **28**:6 (2024), 2915–2970. MR Zbl

[Edidin and Graham 1998] D. Edidin and W. Graham, "Equivariant intersection theory", *Invent. Math.* **131**:3 (1998), 595–634. MR

[Eisenbud and Harris 2016] D. Eisenbud and J. Harris, 3264 *and all that*: *a second course in algebraic geometry*, Cambridge Univ. Press, 2016. MR Zbl

[Fulton 1993] W. Fulton, *Introduction to toric varieties*, Ann. of Math. Stud. **131**, Princeton Univ. Press, 1993. MR Zbl

[Fulton 1998] W. Fulton, *Intersection theory*, 2nd ed., Ergebnisse der Math. (3) **2**, Springer, 1998. MR Zbl

[Inchiostro 2022] G. Inchiostro, "Moduli of genus one curves with two marked points as a weighted blow-up", *Math. Z.* **302**:3 (2022), 1905–1925. MR Zbl

[Iwanari 2009] I. Iwanari, "Integral chow rings of toric stacks", *Int. Math. Res. Not.* **2009**:24 (2009), 4709–4725. MR Zbl

[Keel 1992] S. Keel, "Intersection theory of moduli space of stable n-pointed curves of genus zero", *Trans. Amer. Math. Soc.* **330**:2 (1992), 545–574. MR Zbl

[Molina Rojas and Vistoli 2006] L. A. Molina Rojas and A. Vistoli, "On the Chow rings of classifying spaces for classical groups", *Rend. Sem. Mat. Univ. Padova* **116** (2006), 271–298. MR Zbl

[Mustaţă and Mustaţă 2012] A. M. Mustaţă and A. Mustaţă, "The structure of a local embedding and Chern classes of weighted blow-ups", *J. Eur. Math. Soc.* **14**:6 (2012), 1739–1794. MR Zbl

[Quek and Rydh 2021] M. H. Quek and D. Rydh, "Weighted blow-ups", draft preprint, 2021, available at https://people.kth.se/~dary/weighted-blowups20220329.pdf.

[Stacks 2005–] "The Stacks project", electronic reference, 2005–, available at http://stacks.math.columbia.edu.

[Totaro 2014] B. Totaro, *Group cohomology and algebraic cycles*, Cambridge Tracts in Math. **204**, Cambridge Univ. Press, 2014. MR Zbl

[Vistoli 1989] A. Vistoli, "Intersection theory on algebraic stacks and on their moduli spaces", *Invent. Math.* **97**:3 (1989), 613–670. MR Zbl

Communicated by Christopher Hacon
Received 2023-08-17 Revised 2024-05-29 Accepted 2024-07-16

veronica_arena@brown.edu *Department of Mathematics, Brown University, Providence, RI, United States*

stephen_obinna@brown.edu *Department of Mathematics, Brown University, Providence, RI, United States*

dan_abramovich@brown.edu *Department of Mathematics, Brown University, Providence, RI, United States*

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. 6 2025