

Algebra & Number Theory

Volume 13

2019

No. 5



Algebra & Number Theory

msp.org/ant

EDITORS

MANAGING EDITOR

Bjorn Poonen
Massachusetts Institute of Technology
Cambridge, USA

EDITORIAL BOARD CHAIR

David Eisenbud
University of California
Berkeley, USA

BOARD OF EDITORS

Richard E. Borcherds	University of California, Berkeley, USA	Raman Parimala	Emory University, USA
Antoine Chambert-Loir	Université Paris-Diderot, France	Jonathan Pila	University of Oxford, UK
J-L. Colliot-Thélène	CNRS, Université Paris-Sud, France	Anand Pillay	University of Notre Dame, USA
Brian D. Conrad	Stanford University, USA	Michael Rapoport	Universität Bonn, Germany
Samit Dasgupta	University of California, Santa Cruz, USA	Victor Reiner	University of Minnesota, USA
Hélène Esnault	Freie Universität Berlin, Germany	Peter Sarnak	Princeton University, USA
Gavril Farkas	Humboldt Universität zu Berlin, Germany	Joseph H. Silverman	Brown University, USA
Hubert Flenner	Ruhr-Universität, Germany	Michael Singer	North Carolina State University, USA
Sergey Fomin	University of Michigan, USA	Christopher Skinner	Princeton University, USA
Edward Frenkel	University of California, Berkeley, USA	Vasudevan Srinivas	Tata Inst. of Fund. Research, India
Andrew Granville	Université de Montréal, Canada	J. Toby Stafford	University of Michigan, USA
Joseph Gubeladze	San Francisco State University, USA	Pham Huu Tiep	University of Arizona, USA
Roger Heath-Brown	Oxford University, UK	Ravi Vakil	Stanford University, USA
Craig Huneke	University of Virginia, USA	Michel van den Bergh	Hasselt University, Belgium
Kiran S. Kedlaya	Univ. of California, San Diego, USA	Akshay Venkatesh	Institute for Advanced Study, USA
János Kollár	Princeton University, USA	Marie-France Vignéras	Université Paris VII, France
Philippe Michel	École Polytechnique Fédérale de Lausanne	Kei-Ichi Watanabe	Nihon University, Japan
Susan Montgomery	University of Southern California, USA	Melanie Matchett Wood	University of Wisconsin, Madison, USA
Shigefumi Mori	RIMS, Kyoto University, Japan	Shou-Wu Zhang	Princeton University, USA
Martin Olsson	University of California, Berkeley, USA		

PRODUCTION

production@msp.org
Silvio Levy, Scientific Editor


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

The subscription price for 2019 is US \$385/year for the electronic version, and \$590/year (+\$60, 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. Periodical rate postage paid at Berkeley, CA 94704, and additional mailing offices.

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

PUBLISHED BY

 **mathematical sciences publishers**
nonprofit scientific publishing

<http://msp.org/>

© 2019 Mathematical Sciences Publishers

Surjectivity of Galois representations in rational families of abelian varieties

Aaron Landesman, Ashvin A. Swaminathan, James Tao and Yujie Xu
Appendix by Davide Lombardo

In this article, we show that for any nonisotrivial family of abelian varieties over a rational base with big monodromy, those members that have adelic Galois representation with image as large as possible form a density-1 subset. Our results can be applied to a number of interesting families of abelian varieties, such as rational families dominating the moduli of Jacobians of hyperelliptic curves, trigonal curves, or plane curves. As a consequence, we prove that for any dimension $g \geq 3$, there are infinitely many abelian varieties over \mathbb{Q} with adelic Galois representation having image equal to all of $\mathrm{GSp}_{2g}(\widehat{\mathbb{Z}})$.

1. Introduction and statement of results

1A. Background. One of the most significant breakthroughs in the theory of Galois representations came in 1972, when Serre proved the open image theorem for elliptic curves in his seminal paper [Serre 1972]. Serre’s theorem states that for any elliptic curve E over a number field K without complex multiplication, the image of the associated *adelic* Galois representation ρ_E is an open subgroup of the general symplectic group $\mathrm{GSp}_2(\widehat{\mathbb{Z}})$.¹ The Open Image Theorem not only gives rise to many important corollaries — from the simple consequence that the image of ρ_E has finite index in $\mathrm{GSp}_2(\widehat{\mathbb{Z}})$, to the intriguing result that the density of supersingular primes of E is 0 — but recently, within the past two decades, the theorem has also inspired a body of research concerning the following question:

Question. How large can the image of the adelic Galois representation associated to an elliptic curve be, and how often do elliptic curves attain this largest possible Galois image?

The first major result addressing the above question was achieved by Duke [1997]. He proved that for “most” elliptic curves E over \mathbb{Q} in the standard family with Weierstrass equation $y^2 = x^3 + ax + b$, the image of the *mod- ℓ reduction* of ρ_E is all of $\mathrm{GSp}_2(\mathbb{Z}/\ell\mathbb{Z})$ for every prime number ℓ ; here and in what follows, “most” means a density-1 subset of curves ordered by naïve height. Duke’s result does not imply, however, that ρ_E surjects onto $\mathrm{GSp}_2(\widehat{\mathbb{Z}})$ for most E . In fact, as Serre [1972] observes, the image of ρ_E has index divisible by 2 in $\mathrm{GSp}_2(\widehat{\mathbb{Z}})$ for every elliptic curve E/\mathbb{Q} . Nonetheless, Jones [2010, Theorem 4]

MSC2010: primary 11F80; secondary 11G10, 11G30, 11N36, 11R32, 12E25.

Keywords: Galois representation, abelian variety, étale fundamental group, large sieve, big monodromy, Hilbert irreducibility theorem.

¹Recall that $\mathrm{GSp}_2(\widehat{\mathbb{Z}}) = \mathrm{GL}_2(\widehat{\mathbb{Z}})$; here, we prefer to use the less common symplectic notation so as to highlight the analogy between the elliptic curve case and that of higher dimensional abelian varieties.

proves that most elliptic curves E in the standard family over \mathbb{Q} have *adelic* Galois representations with image as large as possible (i.e., with index 2 in $\mathrm{GSp}_2(\widehat{\mathbb{Z}})$).

The obstruction to having surjective adelic Galois representation faced by elliptic curves over \mathbb{Q} does not occur over other number fields. Greicius [2010, Theorem 1.5] constructed the first explicit example of an elliptic curve over a number field with Galois image equal to all of $\mathrm{GSp}_2(\widehat{\mathbb{Z}})$. Greicius' example is not the only elliptic curve with this property: Zywina [2010a, Theorem 1.2] employs the above result of Jones to show that most elliptic curves in the standard family over a number field $K \neq \mathbb{Q}$ have Galois image equal to all of $\mathrm{GSp}_2(\widehat{\mathbb{Z}})$ as long as $K \cap \mathbb{Q}^{\mathrm{cyc}} = \mathbb{Q}$, where $\mathbb{Q}^{\mathrm{cyc}}$ is the maximal cyclotomic extension of \mathbb{Q} . Subsequently, Zywina [2010b, Theorem 1.15] achieves an intriguing generalization of this result: using a variant of Hilbert's irreducibility theorem, he shows that most members of *every* nonisotrivial rational family of elliptic curves over *any* number field have Galois image as large as possible given the constraints imposed by the arithmetic and geometric properties of the family. Further results over \mathbb{Q} were obtained in [Grant 2000; Cojocaru and Hall 2005; Cojocaru et al. 2011] (see [Zywina 2010b, p. 6] for a more detailed overview).

Given that the above question is so well-studied in the context of elliptic curves, it is natural to ask whether any of the aforementioned theorems extend to abelian varieties of higher dimension. As it happens, explicit examples of curves whose Jacobians have maximal Galois image have been constructed: it follows from the results of [Dieulefait 2002; Zywina 2015] that one can algorithmically write down equations of abelian surfaces and three-folds over \mathbb{Q} with Galois image as large as possible. Moreover, there are several results showing that in a family of abelian varieties, "most" fibers lying over closed points of the base have Galois image with finite index in the Galois image of the family. For instance, in [Cadoret 2015] (see also [Cadoret and Moonen 2018]), the author shows that the set of fibers lying over K -points of the base for which the associated Galois image does *not* have finite index in that of the family is a thin set. Furthermore, in [Cadoret and Tamagawa 2012; 2013], the authors show that when the base of the family is a curve, the set of fibers lying over K -points of the base (and more generally closed points of bounded degree) for which the associated Galois image does *not* have finite index in that of the family is a finite set. However, we are not aware of any results in the literature describing the density of higher-dimensional abelian varieties whose adelic Galois representations have maximal image (as opposed to merely having finite index) in that of the family.

1B. Main result. The primary objective of this article is to prove that an analogue of Zywina's result for rational families of elliptic curves in [Zywina 2010b, Theorem 1.15] holds for abelian varieties of arbitrary dimension, subject to a mild hypothesis on the *monodromy* (i.e., Galois image) of the family under consideration. Before stating our theorems, we must establish some of the requisite notation; we expatiate upon this and other important background material in Section 3A, where precise definitions are provided.

Let K be a number field with fixed algebraic closure \bar{K} , let $U \subset \mathbb{P}_K^r$ be a dense open subscheme, and let $A \rightarrow U$ be a family of g -dimensional principally polarized abelian varieties (henceforth, PPAVs). Let $H_A \subset \mathrm{GSp}_{2g}(\widehat{\mathbb{Z}})$ be the monodromy of the family and let $H_{A_u} \subset H_A$ be the monodromy of the fiber A_u

over $u \in U$. Finally, to facilitate our enumeration of PPAVs, let $\text{Ht} : \mathbb{P}^r(\bar{K}) \rightarrow \mathbb{R}_{>0}$ denote the absolute multiplicative height on projective space,² and define a height function $\| - \|$ on the lattice \mathcal{O}_K^r sending (t_1, \dots, t_r) to $\max_{\sigma,i} |\sigma(t_i)|$, where σ varies over all field embeddings $\sigma : K \hookrightarrow \mathbb{C}$. Our main result is stated as follows:

Theorem 1.1. *Let B, n be arbitrary positive real numbers, and suppose that the rational family $A \rightarrow U$ is nonisotrivial and has big monodromy, meaning that H_A is open in $\text{GSp}_{2g}(\widehat{\mathbb{Z}})$. Let $\delta_{\mathbb{Q}}$ be the index of the closure of the commutator subgroup of H_A in $H_A \cap \text{Sp}_{2g}(\widehat{\mathbb{Z}})$, and let $\delta_K = 1$ for $K \neq \mathbb{Q}$. Then $[H_A : H_{A_u}] \geq \delta_K$ for all $u \in U(K)$, and we have the following asymptotic statements:*

$$\frac{|\{u \in U(K) \cap \mathcal{O}_K^r : \|u\| \leq B, [H_A : H_{A_u}] = \delta_K\}|}{|\{u \in U(K) \cap \mathcal{O}_K^r : \|u\| \leq B\}|} = 1 + O((\log B)^{-n}), \quad \text{and}$$

$$\frac{|\{u \in U(K) : \text{Ht}(u) \leq B, [H_A : H_{A_u}] = \delta_K\}|}{|\{u \in U(K) : \text{Ht}(u) \leq B\}|} = 1 + O((\log B)^{-n}),$$

where the implied constants depend only on $A \rightarrow U$ and n .

Remark 1.2. Notice that Theorem 1.1 holds trivially in dimension 0. In [Zywina 2010b, Theorem 1.15], where the 1-dimensional case of Theorem 1.1 is treated, Zywina bounds the error more sharply, by $O((\log B)B^{-\frac{1}{2}})$ as opposed to our bound of $O((\log B)^{-n})$. In what follows, we shall primarily restrict ourselves to the case where the dimension g is at least 2.

Remark 1.3. Wallace [2014] studies a variant of Theorem 1.1 in the 2-dimensional case. Unfortunately, his argument relies upon a mistaken Masser–Wüstholz-type result of Kawamura, [2003, Main Theorem 2]. Although Wallace [2014, p. 468] describes how to correct some of the errors in Kawamura’s proof, the modified argument still appears to be mistaken; see [Lombardo 2016b, p. 27] for a description of one error in Kawamura’s argument that Wallace does not adequately address. Using the result stated in the Appendix, written by Davide Lombardo, we are able to patch this error in Wallace’s argument.

Remark 1.4. The locus of $u \in U(K)$ with $[H_A : H_{A_u}] > \delta_K$ will not in general be Zariski-closed, so the “sparseness” of this locus can only be quantified by an asymptotic statement. To see why, consider the family of elliptic curves over K given by the Weierstrass equations $y^2 = x^3 + x + a$ for $a \in K$. Note that the mod-2 reduction of the monodromy is nontrivial for the family but is trivial for infinitely many members of the family, namely those for which the defining polynomial $x^3 + x + a$ factors completely over K .

We now outline the proof of Theorem 1.1. Hilbert’s irreducibility theorem is the prototype for results like Theorem 1.1, but it only applies in the setting of finite groups. Indeed, the phenomenon that Galois representations associated to elliptic curves over \mathbb{Q} never surject onto $\text{GSp}_2(\widehat{\mathbb{Z}})$ shows that Hilbert’s irreducibility theorem cannot hold for infinite groups. However, when $A \rightarrow U$ has big monodromy, in the sense that H_A is open in $\text{GSp}_{2g}(\widehat{\mathbb{Z}})$, the problem is essentially reduced to showing that, for most

²See [Hindry and Silverman 2000, Section B.2, p. 174] for the definition.

$u \in U(K)$, the mod- ℓ reduction of H_{A_u} contains $\mathrm{GSp}_{2g}(\mathbb{Z}/\ell\mathbb{Z})$ for each sufficiently large prime ℓ . This reduction uses an infinite version of Goursat’s lemma. Since these mod- ℓ reductions are *finite* groups, the naïve expectation is that Hilbert’s irreducibility theorem can be applied once for each ℓ . Unfortunately, the sum of the resulting error terms does not *a priori* converge to zero.

To overcome this problem, we divide the primes ℓ into three regions.

- (a) We handle all sufficiently large primes by means of a delicate argument involving the large sieve that allows us to apply a recent result of Lombardo (namely, [Lombardo 2016a, Theorem 1.2] and Proposition A.2).
- (b) For the smaller primes, Wallace’s effective version of the Hilbert irreducibility theorem gives sufficiently good error terms. His approach is to complete $\phi : U \rightarrow \mathrm{Spec} K$ to a map $\tilde{\phi} : \mathcal{U} \rightarrow \mathrm{Spec} \mathcal{O}_K$ (see Section 3B), and then to apply the large sieve using information gleaned from the special fibers of $\tilde{\phi}$. To ensure that the monodromy maps associated to special fibers of $\tilde{\phi}$ capture enough information about the monodromy of the whole family, we assume the family is nonisotrivial and has big monodromy. Our main contribution to this step is an application of the Grothendieck specialization theorem, which shows that Wallace’s Property (A2) — concerning the relation between the monodromy maps associated to a geometric *special* fiber and to a geometric *generic* fiber — holds in a very general setting.
- (c) Lastly, to handle the finitely many primes that remain, the Cohen–Serre version of the Hilbert irreducibility theorem suffices.

We encourage the reader to refer to Section 4A for a more detailed discussion of the intricate arguments outlined above.

Remark 1.5. Note that the proof strategy outlined above is greatly influenced by the methods that Zywina [2010b] employed to handle the case where $g = 1$ and also by unpublished work of Zureick-Brown and Zywina. In particular, the idea of formulating the problem in terms of monodromy groups and solving it by applying effective versions of Hilbert’s irreducibility theorem and Serre’s open image theorem is largely due to them.

Zureick-Brown and Zywina were the first to state a version of Theorem 1.1. Indeed, in a 2013 talk at the Institute for Advanced Study, Zywina announced that he and Zureick-Brown had proven a result very much like Theorem 1.1 using a strategy similar to that outlined above. Following this talk, Deligne suggested a potential way to strengthen the result by removing the hypothesis that the family has big monodromy, and it is our understanding that Zywina has been attempting to remove this hypothesis by following Deligne’s suggestion and that his work is still in progress. As the details of the work of Zureick-Brown and Zywina are not available, we have worked out a modified approach that utilizes recent results of Wallace [2014] and Lombardo [2016a] that had not been published at the time of Zywina’s talk. In light of the above, we would like to extend a special acknowledgment to Zureick-Brown and

Zywina for formulating the questions that motivated our work and for introducing the ideas that inspired our proof of Theorem 1.1.

1C. Applications. We record a number of interesting applications of our main result. These and several further applications are stated and proven in Theorem 5.5.

Theorem 1.6 (Abbreviation of Theorem 5.5). *Let \mathcal{A}_g denote the moduli stack of g -dimensional PPAVs, suppose $A \rightarrow U$ is a rational family, and let V be the smallest locally closed substack of \mathcal{A}_g through which $U \rightarrow \mathcal{A}_g$ factors. The conclusion of Theorem 1.1 holds if V is normal and contains a dense open substack of any of the following loci:*

- (a) *the substack of Jacobians of hyperelliptic curves, or*
- (b) *the substack of Jacobians of trigonal curves, or*
- (c) *the substack of Jacobians of plane curves of degree d (see Remark 5.4 for a more precise description of this substack), or*
- (d) *the substack of Jacobians of all curves in \mathcal{M}_g , or*
- (e) *the moduli stack \mathcal{A}_g .*

Theorem 1.6 has the following noteworthy corollary:

Corollary 1.7. *For every $g > 2$, there exist infinitely many PPAVs A over \mathbb{Q} with the property that $\rho_A(G_{\mathbb{Q}}) = \mathrm{GSp}_{2g}(\widehat{\mathbb{Z}})$.*

Proof. Let $\mathcal{T}^s(g \bmod 2) \subset \mathcal{A}_g$ denote the locus of trigonal curves over \mathbb{Q} of lowest Maroni invariant (as defined at the beginning of Section 5B). We have that $\mathcal{T}^s(g \bmod 2)$ is rational and normal when $g > 2$ (by Theorem 5.5(b)) and has monodromy equal to all of $\mathrm{GSp}_{2g}(\widehat{\mathbb{Z}})$ when $g > 2$ (by Remark 5.6). Since $\mathcal{T}^s(g \bmod 2)$ is a dense open substack of the locus Jacobians of trigonal curves, Theorem 1.6 implies that Theorem 1.1 applies to $\mathcal{T}^s(g \bmod 2)$. \square

Remark 1.8. The above proof of Corollary 1.7 is not constructive. For explicit examples of 1-, 2-, and 3-dimensional PPAVs with maximal adelic Galois representations, see [Greicius 2010, Theorem 1.5; Serre 1972, Sections 5.5.6–8; Landesman et al. 2017a; Zywina 2015, Theorem 1.1].

We conclude this section with a representative example, which has incidentally enjoyed significant discussion in the literature.

Example 1.9. In this example, we take our family to be the Hilbert scheme \mathcal{H}_4 of plane curves of degree 4 over \mathbb{Q} . There is quite a bit of earlier work concerning Galois representations associated to Jacobians of such curves. For instance, a single example of a plane quartic such that the adelic Galois representation associated to its Jacobian has image equal to $\mathrm{GSp}_6(\widehat{\mathbb{Z}})$ is given in [Zywina 2015, Theorem 1.1]. In [Anni et al. 2016, Corollary 1.1], an example of a genus-3 hyperelliptic curve whose Jacobian has mod- ℓ monodromy equal to $\mathrm{GSp}_6(\mathbb{Z}/\ell\mathbb{Z})$ for primes $\ell \geq 3$ is constructed. For any $\ell \geq 13$, [Arias-de Reyna et al. 2016, Theorem 0.1] gives an infinite family of 3-dimensional PPAVs with mod- ℓ monodromy equal to

$\mathrm{GSp}_6(\mathbb{Z}/\ell\mathbb{Z})$. All of these existence statements are subsumed by the main results of the present article; indeed, from Remark 5.6 and Theorem 1.6, we obtain the considerably stronger statement that a density-1 subset of this family has Galois representation with image equal to $\mathrm{GSp}_6(\widehat{\mathbb{Z}})$.

The rest of this paper is organized as follows. In Section 2, we define the symplectic group and prove properties concerning its open and closed subgroups. In Section 3, we introduce the basic definitions and properties associated to Galois representations of abelian varieties and families thereof. These definitions and properties are used heavily in Section 4, which is devoted to proving the main theorem of this article, Theorem 1.1. In Section 5, we show that Theorem 1.1 can be applied to study many interesting families of PPAVs, and in so doing, we prove a result that implies Theorem 1.6. Finally, in the Appendix, Davide Lombardo proves a key input that we employ in Section 4 to handle the genus-2 case of Theorem 1.1.

2. Definitions and properties of symplectic groups

In this section, we first detail the basic definitions and properties of symplectic groups, and we then proceed to prove a few group-theoretic lemmas that are used in our proof of the main result of this paper, Theorem 1.1. The reader should feel free to proceed to Section 3 upon reading the statements of Propositions 2.5 and 2.6.

2A. Symplectic groups. Fix a commutative ring R , a free R -module M of rank $2g$ for some positive integer g , and a nondegenerate alternating bilinear form $\langle -, - \rangle : M \times M \rightarrow R$. Define the *general symplectic group* (alternatively, the *group of symplectic similitudes*) $\mathrm{GSp}(M)$ to be the subgroup of $\mathrm{GL}(M)$ consisting of all R -automorphisms S such that there exists some $m_S \in R^\times$, called the *multiplier* of S , satisfying $\langle Sv, Sw \rangle = m_S \cdot \langle v, w \rangle$ for all $v, w \in M$. One readily observes that the *mult* map

$$\mathrm{mult} : \mathrm{GSp}(M) \rightarrow R^\times, \quad S \mapsto m_S$$

is a group homomorphism, and its kernel is the *symplectic group*, denoted by $\mathrm{Sp}(M)$.

By choosing a suitable R -basis for M , we can arrange for the corresponding matrix of the inner product $\langle -, - \rangle$ to be given by

$$\Omega_{2g} = \left[\begin{array}{c|c} 0 & \mathrm{id}_g \\ \hline -\mathrm{id}_g & 0 \end{array} \right],$$

where id_g denotes the $g \times g$ identity matrix. From this choice of basis we obtain an identification $\mathrm{GL}(M) \simeq \mathrm{GL}_{2g}(R)$. We then define $\mathrm{GSp}_{2g}(R)$ to be the image of $\mathrm{GSp}(M)$ and $\mathrm{Sp}_{2g}(R)$ to be the image of $\mathrm{Sp}(M)$ under this identification. Let $\det : \mathrm{GL}_{2g}(R) \rightarrow R^\times$ be the determinant map. Since the diagram

$$\begin{array}{ccc} \mathrm{GSp}(M) & \xrightarrow{\sim} & \mathrm{GSp}_{2g}(R) \\ & \searrow \mathrm{mult}^g & \downarrow \det \\ & & R^\times \end{array}$$

commutes, where the diagonal map is the multiplier map raised to the g -th power, one deduces that $\mathrm{GSp}_{2g}(R)$ is in fact the subgroup of $\mathrm{GL}_{2g}(R)$ consisting of all invertible matrices S satisfying $S^T \Omega_{2g} S = (\mathrm{mult} S) \Omega_{2g}$ and that $\mathrm{Sp}_{2g}(R) = \ker(\mathrm{mult} : \mathrm{GSp}_{2g}(R) \rightarrow R^\times)$.

Let $\mathrm{Mat}_{2g \times 2g}(R)$ denote the space of $2g \times 2g$ matrices with entries in R . In subsequent subsections, we will make heavy use of the ‘‘Lie algebra’’ $\mathfrak{sp}_{2g}(R)$, which is defined by

$$\mathfrak{sp}_{2g}(R) := \{M \in \mathrm{Mat}_{2g \times 2g}(R) : M^T \Omega_{2g} + \Omega_{2g} M = 0\}.$$

It is easy to see that $M^T \Omega_{2g} + \Omega_{2g} M = 0$ is equivalent to M being a block matrix with $g \times g$ blocks of the form

$$M = \left[\begin{array}{c|c} A & B \\ \hline C & -A^T \end{array} \right],$$

where B and C are symmetric.

For the purpose of studying Galois representations associated to PPAVs, we will be primarily interested in the cases where the ring R is the profinite completion $\widehat{\mathbb{Z}}$ of \mathbb{Z} , the ring of ℓ -adic integers \mathbb{Z}_ℓ for a prime number ℓ , or the finite cyclic ring $\mathbb{Z}/m\mathbb{Z}$ for a positive integer m . Note in particular that we have the identifications

$$\mathrm{GSp}_{2g}(\mathbb{Z}_\ell) \simeq \varprojlim_k \mathrm{GSp}_{2g}(\mathbb{Z}/\ell^k\mathbb{Z}) \quad \text{and} \tag{2-1}$$

$$\prod_{\text{prime } \ell} \mathrm{GSp}_{2g}(\mathbb{Z}_\ell) \simeq \mathrm{GSp}_{2g}(\widehat{\mathbb{Z}}) \simeq \varprojlim_m \mathrm{GSp}_{2g}(\mathbb{Z}/m\mathbb{Z}). \tag{2-2}$$

From (2-1) and (2-2), we obtain the ℓ -adic projection map $\pi_\ell : \mathrm{GSp}_{2g}(\widehat{\mathbb{Z}}) \rightarrow \mathrm{GSp}_{2g}(\mathbb{Z}_\ell)$ and the mod- m reduction map $r_m : \mathrm{GSp}_{2g}(\widehat{\mathbb{Z}}) \rightarrow \mathrm{GSp}_{2g}(\mathbb{Z}/m\mathbb{Z})$. Observe that (2-1) and (2-2) both hold with GSp_{2g} replaced by Sp_{2g} .

2B. Notation. In what follows, we study subquotients of $\mathrm{Sp}_{2g}(\widehat{\mathbb{Z}})$, $\mathrm{Sp}_{2g}(\mathbb{Z}_\ell)$, and $\mathrm{Sp}_{2g}(\mathbb{Z}/\ell^k\mathbb{Z})$ for ℓ a prime number and k a positive integer. We use the following notational conventions:

- Let $H \subset \mathrm{Sp}_{2g}(\widehat{\mathbb{Z}})$ be a closed subgroup.
- Let $H_\ell := \pi_\ell(H) \subset \mathrm{Sp}_{2g}(\mathbb{Z}_\ell)$ be the ℓ -adic reduction of H . More generally, for any set S of prime numbers, let H_S denote the projection of H onto $\prod_{\ell \in S} \mathrm{Sp}_{2g}(\mathbb{Z}_\ell)$.
- Let $H(m) = r_m(H) \subset \mathrm{Sp}_{2g}(\mathbb{Z}/m\mathbb{Z})$ be the mod- m reduction of H . We often take $m = \ell^k$.
- Let $\Gamma_{\ell^k} = \ker(\mathrm{Sp}_{2g}(\mathbb{Z}_\ell) \rightarrow \mathrm{Sp}_{2g}(\mathbb{Z}/\ell^k\mathbb{Z}))$. Notice that the map $M \mapsto \mathrm{id}_{2g} + \ell^k M$ gives an isomorphism of groups

$$\mathfrak{sp}_{2g}(\mathbb{Z}/\ell\mathbb{Z}) \simeq \ker(\mathrm{Sp}_{2g}(\mathbb{Z}/\ell^{k+1}\mathbb{Z}) \rightarrow \mathrm{Sp}_{2g}(\mathbb{Z}/\ell^k\mathbb{Z}))$$

for every $k \geq 1$, so we will use $\mathfrak{sp}_{2g}(\mathbb{Z}/\ell\mathbb{Z})$ to denote the above kernel.

- For any group G , let $[G, G]$ be its commutator subgroup, and let $G^{\mathrm{ab}} := G/[G, G]$ be its abelianization.

- For any group G , let $\text{Quo}(G)$ be the set of isomorphism classes of finite nonabelian simple quotients of G , and let $\text{Occ}(G)$ be the set of isomorphism classes of finite nonabelian simple *sub*quotients of G .
- For any positive integer m , let S_m denote the symmetric group on m letters.

2C. Generalizing Goursat’s lemma. In Sections 2D and 2E, it will be crucial for us to have a theorem that allows us to express a subgroup of $\text{Sp}_{2g}(\widehat{\mathbb{Z}})$ as (roughly) the product of its ℓ -adic projections. A natural tool for doing this is Goursat’s lemma, but in much of the literature (e.g., [Ribet 1976, Lemma 5.2.1; Zywna 2010a, Lemma A.4]), this result is stated for *finite* products or for *finite* groups. This section is devoted to proving Lemma 2.2, which generalizes Goursat’s lemma to apply in the setting that we need, namely for *countable* products of *profinite* groups.

Lemma 2.1. *Let $G = \prod_{i=1}^n G_i$ be a product of profinite groups. Then every finite simple quotient of G is a finite simple quotient of G_i for some i , and vice versa.*

Proof. Consider a finite simple quotient $\phi : G \twoheadrightarrow H$. Since each $G_i \subset G$ is normal, the image $\phi(G_i) \subset H$ is also normal. For any i , if $\phi(G_i)$ is larger than $\{1\}$, then it equals H since H is simple, and the composition $G_i \hookrightarrow G \twoheadrightarrow H$ expresses H as a quotient of G_i . If no such i exists, then $\ker \phi = G$, contradiction. The “vice versa” statement is obvious. \square

Lemma 2.2 (generalized Goursat’s lemma). *Let A be a countable set, and suppose $\{G_\alpha\}_{\alpha \in A}$ is a collection of profinite groups such that, for all pairs $\alpha, \beta \in A$ with $\alpha \neq \beta$, the groups G_α and G_β have no finite simple quotients in common. Let $G := \prod_{\alpha \in A} G_\alpha$, and let $\pi_\alpha : G \rightarrow G_\alpha$ be the natural projections. If $H \subset G$ is a closed subgroup with $\pi_\alpha(H) = G_\alpha$ for all $\alpha \in A$, then $H = G$.*

Proof. First take $A = \{1, 2\}$, so that $G = G_1 \times G_2$. The subgroup $N_1 \times \{1\} := (G_1 \times \{1\}) \cap H \subset G$ is normal because $\pi_1(H) = G_1$. This means N_1 is a normal subgroup of G_1 . Similarly for the subgroup $\{1\} \times N_2$. With these definitions, the closed subgroup $H/(N_1 \times N_2) \subset (G_1/N_1) \times (G_2/N_2)$ surjects onto each factor via the natural projections. We have thereby reduced to the case $N_1 = N_2 = 0$. By [Ribet 1976, Lemma 5.2.1], we know that $G_1 \simeq G_2$ as profinite groups. The result follows because two isomorphic profinite groups have a nontrivial finite simple quotient in common (and any quotient of G_i/N_i is *a priori* a quotient of G_i).

Now take $A = \{1, 2, \dots, n\}$ for $n \geq 3$, and suppose (by induction) that the result has been proven for $n - 1$. For any $H \subset G = \prod_{i=1}^n G_i$ satisfying the hypotheses of the theorem, let H' be the image of H under the projection $G \twoheadrightarrow \prod_{i=1}^{n-1} G_i$. Then H' satisfies the hypotheses for $n - 1$, so we conclude that $H' = \prod_{i=1}^{n-1} G_i$. By Lemma 2.1, the groups $\prod_{i=1}^{n-1} G_i$ and G_n have no finite simple quotients in common, so the $n = 2$ case tells us that $H = G$.

The only remaining case is $A = \{1, 2, \dots\}$. Consider $H \subset G$ satisfying the hypotheses of the theorem. For each n , let $H_{\{1,2,\dots,n\}}$ be the image of H under the projection $G \twoheadrightarrow \prod_{i=1}^n G_i$. By the finite case proved above, we know that $H_{\{1,2,\dots,n\}} = \prod_{i=1}^n G_i$ for each $n \geq 1$. Fix an element $g := (g_i)_{i \geq 1} \in G$, and define a sequence $\{h_1, h_2, \dots\}$ of elements of H as follows: let h_n be any element of H whose image in $\prod_{i=1}^n G_i$

equals (g_1, \dots, g_n) . In the product topology, $h_n \rightarrow g$ as $n \rightarrow \infty$, so $g \in H$ since H is closed. Since $g \in G$ was arbitrary, we conclude that $H = G$. \square

2D. Closed subgroups of $\mathrm{Sp}_{2g}(\widehat{\mathbb{Z}})$. As before, let $H \subset \mathrm{Sp}_{2g}(\widehat{\mathbb{Z}})$ be a closed subgroup. The main result of this section is Proposition 2.5, which shows that properties of H can be deduced from corresponding properties of the ℓ -adic projections $H_\ell \subset \mathrm{Sp}_{2g}(\mathbb{Z}_\ell)$ as ℓ ranges over the prime numbers. We use Proposition 2.5 crucially in our proof of the main theorem, Theorem 1.1, and more specifically in the proof of Proposition 4.2.

The next lemma enables us to verify the conditions required for applying Lemma 2.2:

Lemma 2.3. *If $g > 2$ or $\ell > 2$, we have $\mathrm{Quo}(\mathrm{Sp}_{2g}(\mathbb{Z}_\ell)) = \{\mathrm{P}\mathrm{Sp}_{2g}(\mathbb{Z}/\ell\mathbb{Z})\}$. Moreover, for all $g \geq 2$, we have $\mathrm{Quo}(\mathrm{Sp}_{2g}(\mathbb{Z}_\ell)) \cap \mathrm{Quo}(\mathrm{Sp}_{2g}(\mathbb{Z}_{\ell'})) = \emptyset$ if $\ell \neq \ell'$.*

Proof. Since Γ_ℓ is a pro- ℓ group, we have that $\mathrm{Quo}(\mathrm{Sp}_{2g}(\mathbb{Z}_\ell)) = \mathrm{Quo}(\mathrm{Sp}_{2g}(\mathbb{Z}/\ell\mathbb{Z}))$. Furthermore, quotienting by $\{\pm \mathrm{id}_{2g}\}$, we have that $\mathrm{Quo}(\mathrm{Sp}_{2g}(\mathbb{Z}/\ell\mathbb{Z})) = \mathrm{Quo}(\mathrm{Sp}_{2g}(\mathbb{Z}/\ell\mathbb{Z})/\{\pm \mathrm{id}_{2g}\})$. By [O’Meara 1978, Theorem 3.4.1], we have that $\mathrm{Sp}_{2g}(\mathbb{Z}/\ell\mathbb{Z})/\{\pm \mathrm{id}_{2g}\} = \mathrm{P}\mathrm{Sp}_{2g}(\mathbb{Z}/\ell\mathbb{Z})$ is simple for $g > 2$ or $\ell > 2$. It follows that $\mathrm{Quo}(\mathrm{Sp}_{2g}(\mathbb{Z}_\ell)) = \{\mathrm{P}\mathrm{Sp}_{2g}(\mathbb{Z}/\ell\mathbb{Z})\}$ in this case.

To finish the proof, note that $\mathrm{Quo}(\mathrm{Sp}_{2g}(\mathbb{Z}_\ell)) \cap \mathrm{Quo}(\mathrm{Sp}_{2g}(\mathbb{Z}_{\ell'})) = \emptyset$ for $g > 2$ or $\ell, \ell' > 2$ because $\mathrm{P}\mathrm{Sp}_{2g}(\mathbb{Z}/\ell\mathbb{Z}) \neq \mathrm{P}\mathrm{Sp}_{2g}(\mathbb{Z}/\ell'\mathbb{Z})$ for $\ell \neq \ell'$ because their orders are different. The only remaining case is where $g = 2$, $\ell = 2$, and $\ell' > 2$. In this case, observe that $\mathrm{P}\mathrm{Sp}_{2g}(\mathbb{Z}/\ell'\mathbb{Z}) \notin \mathrm{Quo}(\mathrm{Sp}_{2g}(\mathbb{Z}/2\mathbb{Z}))$ for $\ell' > 2$, since the order of $\mathrm{P}\mathrm{Sp}_{2g}(\mathbb{Z}/\ell'\mathbb{Z})$ exceeds that of $\mathrm{Sp}_{2g}(\mathbb{Z}/2\mathbb{Z})$. \square

We next prove Proposition 2.4, which we then use to deduce the main result of this section, Proposition 2.5.

Proposition 2.4. *Let $g \geq 2$ and let $H \subset \mathrm{Sp}_{2g}(\widehat{\mathbb{Z}})$ be a closed subgroup. Suppose there is a prime number $p \geq 2$ such that $H(\ell) = \mathrm{Sp}_{2g}(\mathbb{Z}/\ell\mathbb{Z})$ for all $\ell > p$. Then we have that*

$$H = H_{\{\ell \leq p\}} \times \prod_{\ell > p} \mathrm{Sp}_{2g}(\mathbb{Z}_\ell). \tag{2-3}$$

The idea of the proof is to apply Lemma 2.2 to conclude that if the group surjects onto each factor, then it surjects onto the product. We verify the hypotheses of Lemma 2.2 using Lemma 2.3 and the fact that all simple quotients of $H_{\{\ell \leq p\}}$ have smaller order than $\mathrm{P}\mathrm{Sp}_{2g}(\mathbb{Z}_\ell)$ for $\ell > p$.

Proof. The case where $g = 1$ is handled by [Zywina 2010b, Lemma 7.6], so take $g \geq 2$. By [Landesman et al. 2017b, Theorem 1], the fact that $H(\ell) = \mathrm{Sp}_{2g}(\mathbb{Z}/\ell\mathbb{Z})$ implies that $H_\ell = \mathrm{Sp}_{2g}(\mathbb{Z}_\ell)$ for all $\ell > p$.

The proposition follows upon applying Lemma 2.2 to the product $H_{\{\ell \leq p\}} \times \prod_{\ell > p} \mathrm{Sp}_{2g}(\mathbb{Z}_\ell)$. However, to apply it, we must check that no two of the groups $H_{\{\ell \leq p\}}$ and $\mathrm{Sp}_{2g}(\mathbb{Z}_\ell)$ for $\ell > p$ have any finite simple quotients in common. From [Landesman et al. 2017b, Proposition 1(a)], we have that the group $\mathrm{Sp}_{2g}(\mathbb{Z}_\ell)$ has trivial abelianization for $\ell > 2$ and thus has no finite abelian simple quotients. Thus, it remains to verify that the sets of nonabelian simple quotients $\mathrm{Quo}(H_{\{\ell \leq p\}})$ and $\mathrm{Quo}(\mathrm{Sp}_{2g}(\mathbb{Z}_\ell))$ for $\ell > p$ are all pairwise disjoint. Our strategy for checking this condition is to bound the sizes of the groups appearing

in $\text{Quo}(H_{\{\ell \leq p\}})$. First, observe that

$$\text{Quo}(H_{\{\ell \leq p\}}) \subset \text{Occ}\left(\prod_{\ell \leq p} \text{Sp}_{2g}(\mathbb{Z}_\ell)\right) = \bigcup_{\ell \leq p} \text{Occ}(\text{Sp}_{2g}(\mathbb{Z}_\ell)),$$

where the last step follows from the first displayed equation of [Serre 1998, p. IV-25]. But $\text{Occ}(\text{Sp}_{2g}(\mathbb{Z}_\ell)) = \text{Occ}(\Gamma_\ell) \cup \text{Occ}(\text{Sp}_{2g}(\mathbb{Z}/\ell\mathbb{Z}))$, and $\text{Occ}(\Gamma_\ell) = \emptyset$ because Γ_ℓ is a pro- ℓ group, so $\text{Occ}(\text{Sp}_{2g}(\mathbb{Z}_\ell)) = \text{Occ}(\text{Sp}_{2g}(\mathbb{Z}/\ell\mathbb{Z}))$. Because $\text{Sp}_{2g}(\mathbb{Z}/\ell\mathbb{Z})$ is not simple, every element of $\text{Occ}(\text{Sp}_{2g}(\mathbb{Z}/\ell\mathbb{Z}))$ is bounded in size by $|\text{Sp}_{2g}(\mathbb{Z}/\ell\mathbb{Z})|/2$, so every element of $\text{Quo}(H_{\{\ell \leq p\}})$ is bounded in size by $|\text{Sp}_{2g}(\mathbb{Z}/p\mathbb{Z})|/2$. Observing that

$$\frac{1}{2} \cdot |\text{Sp}_{2g}(\mathbb{Z}/p\mathbb{Z})| < |\text{PSp}_{2g}(\mathbb{Z}/\ell\mathbb{Z})|$$

for every $\ell > p$, the desired condition follows by applying Lemma 2.3. □

Proposition 2.5. *Let $G \subset \text{Sp}_{2g}(\widehat{\mathbb{Z}})$ be an open subgroup. There exists a positive integer M such that, for every closed subgroup $H \subset G$, we have $H = G$ if and only if $H(M) = G(M)$ and $H(\ell) = \text{Sp}_{2g}(\mathbb{Z}/\ell\mathbb{Z})$ for every prime $\ell \nmid M$.*

The idea of the proof is to find a sufficiently large M so that if $H(M) = G(M)$ then $H_{\{\ell \nmid M\}} = G_{\{\ell \nmid M\}}$, which reduces the problem to the situation of Proposition 2.4.

Proof. Again, the case where $g = 1$ is handled in [Zywina 2010b, Lemma 7.6], so take $g \geq 2$. Let p be any prime such that $G(\ell) = \text{Sp}_{2g}(\mathbb{Z}/\ell\mathbb{Z})$ for all primes $\ell > p$. Observe that the groups Γ_{ℓ^k} are open in $\text{Sp}_{2g}(\mathbb{Z}_\ell)$ because they have finite index in $\text{Sp}_{2g}(\mathbb{Z}_\ell)$. Since $G \subset \text{Sp}_{2g}(\widehat{\mathbb{Z}})$ is open, the group $G_{\{\ell \leq p\}} \subset \prod_{\ell \leq p} \text{Sp}_{2g}(\mathbb{Z}_\ell)$ is open too, so there exist exponents $e(\ell) \geq 1$ with the property that

$$\prod_{\ell \leq p} \Gamma_{\ell^{e(\ell)}} \subset G_{\{\ell \leq p\}}.$$

Since the groups Γ_{ℓ^k} are finitely generated pro- ℓ open normal subgroups of $\text{GSp}_{2g}(\mathbb{Z}_\ell)$, condition (ii) from [Serre 1997, Proposition 10.6] is satisfied. Hence, the equivalence of conditions (ii) and (iv) from [Serre 1997, Proposition 10.6] implies that the Frattini subgroup defined by

$$\Phi(G_{\{\ell \leq p\}}) := \bigcap_{\substack{S \subset G_{\{\ell \leq p\}} \\ S \text{ maximal closed in } G_{\{\ell \leq p\}}}} S$$

is open and normal in $G_{\{\ell \leq p\}}$. This means we can find exponents $e'(\ell) \geq 1$ such that

$$\prod_{\ell \leq p} \Gamma_{\ell^{e'(\ell)}} \subset \Phi(G_{\{\ell \leq p\}}).$$

Define $M := \prod_{\ell \leq p} \ell^{e'(\ell)}$. Then $H(M) = G(M)$ implies that $H_{\{\ell \leq p\}} = G_{\{\ell \leq p\}}$.

Now take H satisfying $H(M) = G(M)$ and $H(\ell) = \text{Sp}_{2g}(\mathbb{Z}/\ell\mathbb{Z})$ for every prime $\ell \nmid M$. We have that

$$H \subset G \subset H_{\{\ell \leq p\}} \times \prod_{\ell > p} \text{Sp}_{2g}(\mathbb{Z}_\ell).$$

To show that $H = G$, we need only verify

$$H = H_{\{\ell \leq p\}} \times \prod_{\ell > p} \mathrm{Sp}_{2g}(\mathbb{Z}_\ell),$$

which follows immediately from Proposition 2.4. □

2E. Open subgroups of $\mathrm{GSp}_{2g}(\widehat{\mathbb{Z}})$. We now return to studying the general symplectic group $\mathrm{GSp}_{2g}(\widehat{\mathbb{Z}})$. The main result of this subsection tells us that the closure of the commutator subgroup of an open subgroup of $\mathrm{GSp}_{2g}(\widehat{\mathbb{Z}})$ is open:

Proposition 2.6. *Let $g \geq 2$, and let $H \subset \mathrm{GSp}_{2g}(\widehat{\mathbb{Z}})$ be an open subgroup. Then the closure of $[H, H]$ is an open subgroup of $\mathrm{Sp}_{2g}(\widehat{\mathbb{Z}})$.*

In order to prove Proposition 2.6, we shall require a number of preliminary lemmas, which are stated and proven in Sections 2E1 and 2E2.

2E1. Openness condition. The next two lemmas give us a criterion for openness in $\mathrm{Sp}_{2g}(\widehat{\mathbb{Z}})$:

Lemma 2.7. *Let S be a finite set of prime numbers, and let $H \subset \prod_{\ell \in S} \mathrm{Sp}_{2g}(\mathbb{Z}_\ell)$ be a closed subgroup. If each $H_\ell \subset \mathrm{Sp}_{2g}(\mathbb{Z}_\ell)$ is open, then $H \subset \prod_{\ell \in S} \mathrm{Sp}_{2g}(\mathbb{Z}_\ell)$ is open.*

Proof. There exists a finite-index subgroup $H' \subset H$ such that $H'(\ell)$ is trivial for every $\ell \in S$, namely the intersection of the kernels of the mod- ℓ reductions maps $H \rightarrow H(\ell)$. Since each H'_ℓ is a pro- ℓ group, Lemma 2.2 implies that $H' = \prod_{\ell \in S} H'_\ell$. Thus, H contains an open subgroup and is therefore itself open. □

Lemma 2.8. *Let $g \geq 2$ and let $H \subset \mathrm{Sp}_{2g}(\widehat{\mathbb{Z}})$ be a closed subgroup. If $H_{\ell'}$ is open in $\mathrm{Sp}_{2g}(\mathbb{Z}_{\ell'})$ for all ℓ' and $H_\ell = \mathrm{Sp}_{2g}(\mathbb{Z}_\ell)$ for all but finitely many ℓ , then H is open in $\mathrm{Sp}_{2g}(\widehat{\mathbb{Z}})$.*

Proof. Let p be the largest prime with $H_p \neq \mathrm{Sp}_{2g}(\mathbb{Z}_p)$. By Lemma 2.7, we have that $H_{\{\ell \leq p\}} \subset \prod_{\ell \leq p} \mathrm{Sp}_{2g}(\mathbb{Z}_\ell)$ is an open subgroup. The result then follows from Proposition 2.4. □

2E2. Two computational lemmas. The next two results are used in the proof of Proposition 2.6. The following lemma describes the commutator of an element of Γ_{ℓ^m} with an element of Γ_{ℓ^n} .

Lemma 2.9. *Let $n \leq m$ be positive integers, and let $\mathrm{id}_{2g} + \ell^n U$ and $\mathrm{id}_{2g} + \ell^m V$ be elements of $\mathrm{GL}_{2g}(\mathbb{Z}_\ell)$. Then we have*

$$(\mathrm{id}_{2g} + \ell^n U)^{-1}(\mathrm{id}_{2g} + \ell^m V)(\mathrm{id}_{2g} + \ell^n U)(\mathrm{id}_{2g} + \ell^m V)^{-1} \equiv \mathrm{id}_{2g} + \ell^{n+m}(VU - UV) \pmod{\ell^{2n+m}}.$$

Proof. We have

$$\begin{aligned} (\mathrm{id}_{2g} + \ell^m V)(\mathrm{id}_{2g} + \ell^n U)(\mathrm{id}_{2g} + \ell^m V)^{-1} &= \mathrm{id}_{2g} + \ell^n (\mathrm{id}_{2g} + \ell^m V)U(\mathrm{id}_{2g} + \ell^m V)^{-1} \\ &= \mathrm{id}_{2g} + \ell^n (\mathrm{id}_{2g} + \ell^m V)U \left(\sum_{i=0}^{\infty} (-1)^i \ell^{im} V^i \right) \\ &= \mathrm{id}_{2g} + \ell^n \sum_{i=0}^{\infty} [(-1)^i \ell^{im} UV^i + (-1)^i \ell^{(i+1)m} VUV^i] \\ &= \mathrm{id}_{2g} + \ell^n U + \ell^{n+m}(VU - UV)(\mathrm{id}_{2g} + \ell^m V)^{-1}. \end{aligned}$$

Multiplying on the left by $(\text{id}_{2g} + \ell^n U)^{-1}$ gives the desired result. \square

In the next proposition, we show the commutator subalgebra of $\mathfrak{sp}_{2g}(\mathbb{Z}/\ell\mathbb{Z})$ is sufficiently large for all primes ℓ .

Proposition 2.10. *We have the following results:*

- (a) *For all $g \geq 1$ and $\ell \geq 3$ we have $[\mathfrak{sp}_{2g}(\mathbb{Z}/\ell\mathbb{Z}), \mathfrak{sp}_{2g}(\mathbb{Z}/\ell\mathbb{Z})] = \mathfrak{sp}_{2g}(\mathbb{Z}/\ell\mathbb{Z})$.*
- (b) *For all $g \geq 1$ we have $[\mathfrak{sp}_{2g}(\mathbb{Z}/4\mathbb{Z}), \mathfrak{sp}_{2g}(\mathbb{Z}/4\mathbb{Z})] \supset 2 \cdot \mathfrak{sp}_{2g}(\mathbb{Z}/2\mathbb{Z})$.*

Proof. Statement (a) follows immediately from [Steinberg 1961, Theorem 2.6], which states that $\mathfrak{sp}_{2g}(\mathbb{Z}/\ell\mathbb{Z})$ is simple for $\ell \geq 3$. It remains to prove statement (b). For this, we compute several commutators and make deductions based on each one. For convenience, let $\mathfrak{g} = [\mathfrak{sp}_{2g}(\mathbb{Z}/4\mathbb{Z}), \mathfrak{sp}_{2g}(\mathbb{Z}/4\mathbb{Z})]$, let A, D denote arbitrary $g \times g$ matrices, and let B, C, E, F denote symmetric $g \times g$ matrices. Since

$$\left[\left[\begin{array}{c|c} A & 0 \\ \hline 0 & -A^T \end{array} \right], \left[\begin{array}{c|c} D & 0 \\ \hline 0 & -D^T \end{array} \right] \right] = \left[\begin{array}{c|c} AD - DA & 0 \\ \hline 0 & A^T D^T - D^T A^T \end{array} \right], \tag{2-4}$$

all block-diagonal matrices in $\mathfrak{sp}_{2g}(\mathbb{Z}/4\mathbb{Z})$ with every diagonal entry equal to 0 are contained in \mathfrak{g} . This can be seen by taking A and D to be various elementary matrices. Furthermore,

$$\left[\left[\begin{array}{c|c} 0 & B \\ \hline C & 0 \end{array} \right], \left[\begin{array}{c|c} 0 & E \\ \hline F & 0 \end{array} \right] \right] = \left[\begin{array}{c|c} BF - EC & 0 \\ \hline 0 & CE - FB \end{array} \right], \tag{2-5}$$

so we can arrange that $BF - EC$ is an elementary matrix with a single nonzero entry on the diagonal. Summing matrices from (2-4) and (2-5) tells us that all block-diagonal matrices are contained in \mathfrak{g} . Additionally,

$$\left[\left[\begin{array}{c|c} \text{id}_g & 0 \\ \hline 0 & -\text{id}_g \end{array} \right], \left[\begin{array}{c|c} 0 & B \\ \hline 0 & 0 \end{array} \right] \right] = \left[\begin{array}{c|c} 0 & 2B \\ \hline 0 & 0 \end{array} \right]. \tag{2-6}$$

Repeating the computation from (2-6) with the other off-diagonal block nonzero implies that 2 times any matrix in $\mathfrak{sp}_{2g}(\mathbb{Z}/2\mathbb{Z})$ whose diagonal blocks are 0 is an element of \mathfrak{g} . The desired result follows because $2 \cdot \mathfrak{sp}_{2g}(\mathbb{Z}/2\mathbb{Z})$ is contained in the subspace generated by the matrices from (2-4), (2-5), and (2-6). \square

2E3. Completing the proof. In order to prove Proposition 2.6, we require the following lemma, which states that the closure of the commutator $[\Gamma_{\ell^k}, \Gamma_{\ell^k}]$ is large.

Lemma 2.11. *Fix $k \geq 1$. Then if $\ell \neq 2$, the closure of $[\Gamma_{\ell^k}, \Gamma_{\ell^k}]$ contains $\Gamma_{\ell^{2k}}$ and if $\ell = 2$, the closure of $[\Gamma_{\ell^k}, \Gamma_{\ell^k}]$ contains $\Gamma_{\ell^{2k+1}}$.*

Proof. First suppose $\ell \geq 3$. Statement (a) of Proposition 2.10 implies that for any $W' \in \mathfrak{sp}_{2g}(\mathbb{Z}/\ell\mathbb{Z})$, there exist $U', V' \in \mathfrak{sp}_{2g}(\mathbb{Z}/\ell\mathbb{Z})$ such that $V'U' - U'V' = W'$. Choosing lifts W, U, V of W', U', V' , it follows from Lemma 2.9 that for every i and for every such

$$\text{id}_{2g} + \ell^{2k+i} W \in \Gamma_{\ell^{2k+i}}, \quad \text{id}_{2g} + \ell^k U \in \Gamma_{\ell^k}, \quad \text{and} \quad \text{id}_{2g} + \ell^{k+i} V \in \Gamma_{\ell^{k+i}},$$

we have that

$$(\mathrm{id}_{2g} + \ell^k U)^{-1} (\mathrm{id}_{2g} + \ell^{k+i} V) (\mathrm{id}_{2g} + \ell^k U) (\mathrm{id}_{2g} + \ell^{k+i} V)^{-1} \equiv \mathrm{id}_{2g} + \ell^{2k+i} W \pmod{\ell^{2k+i+1}}.$$

Take $M_0 \in \Gamma_{\ell^{2k}}$. There exists $X_1 \in [\Gamma_{\ell^{2k}}, \Gamma_{\ell^{2k}}]$ and $M_1 \in \Gamma_{\ell^{2k+1}}$ with the property that $M_0 = X_1 M_1$. Proceeding inductively in this manner, we obtain sequences

$$\{X_i : i = 1, 2, \dots\} \subset [\Gamma_{\ell^k}, \Gamma_{\ell^k}] \quad \text{and} \quad \{M_i : i = 0, 1, 2, \dots\} \quad \text{with} \quad M_i \in \Gamma_{\ell^{2k+i}}$$

such that $M_i = X_{i+1} M_{i+1}$ for each i . Then we have the following equalities of matrices in $\mathrm{Sp}_{2g}(\mathbb{Z}_\ell)$:

$$M_0 = \lim_{i \rightarrow \infty} \left(\prod_{j=1}^i X_j \right) M_i = \prod_{j=1}^{\infty} X_j.$$

It follows that $\Gamma_{\ell^{2k}}$ is contained in the closure of $[\Gamma_{\ell^k}, \Gamma_{\ell^k}]$.

Now suppose $\ell = 2$. Observe that for each $k \geq 2$ we have

$$\mathrm{id}_{2g} + 2^k \cdot \mathfrak{sp}_{2g}(\mathbb{Z}/4\mathbb{Z}) = \ker(\mathrm{Sp}_{2g}(\mathbb{Z}/2^{k+2}\mathbb{Z}) \rightarrow \mathrm{Sp}_{2g}(\mathbb{Z}/2^k\mathbb{Z})).$$

It follows from statement (b) of Proposition 2.10 and Lemma 2.9 that for every choice of $\mathrm{id}_{2g} + 2^{2k+i+1} W \in \Gamma_{2^{2k+i+1}}$ and for each nonnegative integer i , there exist $\mathrm{id}_{2g} + 2^k U \in \Gamma_{2^k}$ and $\mathrm{id}_{2g} + 2^{k+i} V \in \Gamma_{2^{k+i}}$ with the property that

$$(\mathrm{id}_{2g} + 2^k U)^{-1} (\mathrm{id}_{2g} + 2^{k+i} V) (\mathrm{id}_{2g} + 2^k U) (\mathrm{id}_{2g} + 2^{k+i} V)^{-1} \equiv \mathrm{id}_{2g} + 2^{2k+i+1} W \pmod{\ell^{2k+i+2}}.$$

One may now finish the proof by applying a similar inductive argument to the one used in the case $\ell \geq 3$. \square

We are finally in position to prove the main result of this section.

Proof of Proposition 2.6. By Lemma 2.8, it suffices to prove the following two statements:

- (a) The closure of $[H, H]$ surjects onto $\mathrm{Sp}_{2g}(\mathbb{Z}_\ell)$ for all but finitely many ℓ .
- (b) The closure of $[H, H]$ maps onto an open subgroup of $\mathrm{Sp}_{2g}(\mathbb{Z}_\ell)$ for each ℓ .

For statement (a), notice that H surjects onto $\mathrm{GSp}_{2g}(\mathbb{Z}_\ell)$ for all but finitely many ℓ . Note that for $\ell \geq 3$, we have $[\mathrm{GSp}_{2g}(\mathbb{Z}_\ell), \mathrm{GSp}_{2g}(\mathbb{Z}_\ell)] = \mathrm{Sp}_{2g}(\mathbb{Z}_\ell)$ because, by [Landesman et al. 2017b, Proposition 1], we have that

$$\mathrm{Sp}_{2g}(\mathbb{Z}_\ell) = [\mathrm{Sp}_{2g}(\mathbb{Z}_\ell), \mathrm{Sp}_{2g}(\mathbb{Z}_\ell)] \subset [\mathrm{GSp}_{2g}(\mathbb{Z}_\ell), \mathrm{GSp}_{2g}(\mathbb{Z}_\ell)] \subset \mathrm{Sp}_{2g}(\mathbb{Z}_\ell).$$

Thus, $[H, H]$ itself surjects onto $[\mathrm{GSp}_{2g}(\mathbb{Z}_\ell), \mathrm{GSp}_{2g}(\mathbb{Z}_\ell)] = \mathrm{Sp}_{2g}(\mathbb{Z}_\ell)$ for all $\ell \geq 3$.

To show statement (b), we prove that the closure of $[H', H']$ is open in $\mathrm{Sp}_{2g}(\mathbb{Z}_\ell)$ for any open subgroup $H' \subset \mathrm{GSp}_{2g}(\mathbb{Z}_\ell)$. Since H' is open, there exists some $k \geq 1$ such that $\Gamma_{\ell^k} \subset H'$, so by Lemma 2.11, there exists $m \geq 2k$ such that $\Gamma_{\ell^m} \subset [\Gamma_{\ell^k}, \Gamma_{\ell^k}] \subset [H', H']$. Thus, $[H', H']$ contains an open subgroup and must therefore itself be open, as desired. \square

3. Background on Galois representations of PPAVs

This section is devoted to describing the basic definitions and properties concerning Galois representations associated to families of PPAVs. Specifically, in Section 3A, we construct these Galois representations and provide precise definitions for the various monodromy groups discussed in Section 1B. Then, in Section 3B, we explain how a family of PPAVs over a number field K may be extended to a family over the number ring \mathcal{O}_K . The notation introduced in this section will be utilized throughout the rest of the paper.

3A. Defining Galois representations for families of PPAVs. Let K be a number field, and let $g \geq 0$ be an integer. Fix a base scheme T (we usually take T to be $\text{Spec } K$ or an open subscheme of $\text{Spec } \mathcal{O}_K$), and let U be an integral T -scheme with generic point η (we usually take U to be an open subscheme of \mathbb{P}_K^r or $\mathbb{P}_{\mathcal{O}_K}^r$). Let $A \rightarrow U$ be a *family* of g -dimensional PPAVs, by which we mean the following:

- The morphism $A \rightarrow U$ is flat, proper, and finitely presented with smooth geometrically connected fibers of dimension g .
- A is a group scheme over U , and the resulting abelian scheme is equipped with a principal polarization.

Note that $A \rightarrow U$ is automatically abelian, smooth, and projective, and further observe that the fiber A_u over any point $u \in U$ is a PPAV of dimension g over the residue field $\kappa(u)$ of u .

Choose a geometric generic point $\bar{\eta}$ for U . If $\kappa(\eta)$ has characteristic prime to m , the action of the étale fundamental group $\pi_1(U, \bar{\eta})^3$ on the geometric generic fiber $A_{\bar{\eta}}[m]$ gives rise to a continuous linear representation whose image is constrained by the Weil pairing to lie in the general symplectic group $\text{GSp}_{2g}(\mathbb{Z}/m\mathbb{Z})$. We denote this *mod- m representation* by

$$\rho_{A,m} : \pi_1(U, \bar{\eta}) \rightarrow \text{GSp}_{2g}(\mathbb{Z}/m\mathbb{Z}). \tag{3-1}$$

The map in (3-1) is well-defined up to the choice of base-point $\bar{\eta}$, and choosing a different such $\bar{\eta}$ would only alter the image of $\rho_{A,m}$ by an inner automorphism of $\text{GSp}_{2g}(\mathbb{Z}/m\mathbb{Z})$. For this reason, when it will not lead to confusion, we may omit the basepoint from our notation and write $\pi_1(U)$ for $\pi_1(U, \bar{\eta})$.

If ℓ is a prime not dividing the characteristic of $\kappa(\eta)$, then we can take the inverse limit of the mod- ℓ^k representations to obtain the *ℓ -adic representation*

$$\rho_{A,\ell^\infty} : \pi_1(U) \rightarrow \varprojlim_k \text{GSp}_{2g}(\mathbb{Z}/\ell^k\mathbb{Z}). \tag{3-2}$$

Moreover, if $\kappa(\eta)$ has characteristic 0, we can take the inverse limit of all the mod- m representations (or equivalently the product of all the ℓ -adic representations) to obtain an *adelic or global representation*

$$\rho_A : \pi_1(U) \rightarrow \varprojlim_m \text{GSp}_{2g}(\mathbb{Z}/m\mathbb{Z}) \simeq \text{GSp}_{2g}(\widehat{\mathbb{Z}}). \tag{3-3}$$

³For a general foundational reference on the étale fundamental group, see [SGA 1 1971].

Remark 3.1. In the situation that $U = \text{Spec } K$, the choice of $\bar{\eta}$ corresponds to a choice of algebraic closure \bar{K} of K . Taking $G_K := \text{Gal}(\bar{K}/K)$ to be the absolute Galois group, we have that $\pi_1(U, \bar{\eta}) = G_K$. This recovers the notion of a Galois representation of a PPAV over a field as a map $\rho_A : G_K \rightarrow \text{GSp}_{2g}(\widehat{\mathbb{Z}})$.

Remark 3.2. For a commutative ring R , recall from the definition of the general symplectic group that we have a multiplier map $\text{mult} : \text{GSp}_{2g}(R) \rightarrow R^\times$. Let χ_m be the mod- m cyclotomic character, and let χ be the cyclotomic character. If $U = \text{Spec } k$, (with k an arbitrary characteristic 0 field) it follows from G_k -invariance of the Weil pairing that $\chi_m = \text{mult} \circ \rho_{A,m}$ and $\chi = \text{mult} \circ \rho_A$. More generally, if U is normal and integral, and $\phi : \pi_1(U) \rightarrow \pi_1(\text{Spec } K)$, then $\chi \circ \phi = \text{mult} \circ \rho_A$, which holds because it holds for the generic fiber $A_\eta \rightarrow \text{Spec } K(\eta)$, and the map $\pi_1(\eta) \rightarrow \pi_1(U)$ is surjective.

We now define the monodromy groups associated to the representations defined above. We call the image of $\rho_A : \pi_1(U) \rightarrow \text{GSp}_{2g}(\widehat{\mathbb{Z}})$ the *monodromy* of the family $A \rightarrow U$, and we denote it by H_A . When the base scheme is $T = \text{Spec } K$, we also define the *geometric monodromy*, denoted by H_A^{geom} , to be the image of the adelic representation $\rho_{A_{\bar{K}}} : \pi_1(U_{\bar{K}}) \rightarrow \text{GSp}_{2g}(\widehat{\mathbb{Z}})$ associated to the base-changed family $A_{\bar{K}} \rightarrow U_{\bar{K}}$. Since the cyclotomic character is trivial on $G_{\bar{K}}$, it follows that H_A^{geom} is actually a subgroup of $\text{Sp}_{2g}(\widehat{\mathbb{Z}})$. We write $H_A(m)$ and $H_A^{\text{geom}}(m)$ for the mod- m reductions of the above-defined monodromy groups. We say $A \rightarrow U$ has big monodromy if H_A is open in $\text{GSp}_{2g}(\widehat{\mathbb{Z}})$ and $A \rightarrow U$ has big geometric monodromy if H_A^{geom} is open in $\text{Sp}_{2g}(\widehat{\mathbb{Z}})$.

In particular, for each $u \in U$, H_{A_u} and $H_{A_u}^{\text{geom}}$ are the monodromy groups associated to the family $A_u \rightarrow \text{Spec } \kappa(u)$. Since A_u is the pullback of A along $\iota : u \rightarrow U$, $\rho_{A_u} = \iota \circ \rho_A$ and we obtain an inclusion $H_{A_u} \subset H_A$. Note that if U is normal, then the map $\pi_1(\eta) \rightarrow \pi_1(U)$ is surjective, so we have that $H_{A_\eta} = H_A$.

3B. Extending families over K to \mathcal{O}_K . Recall that, for a single abelian variety A_u over $u = \text{Spec } K$, good reduction for A_u at a prime $\mathfrak{p} \in \Sigma_K$ implies that the Galois representation $\rho_{A_u,m} : G_K \rightarrow \text{GSp}(\mathbb{Z}/m\mathbb{Z})$ is unramified at \mathfrak{p} , provided that \mathfrak{p} does not divide m . All but finitely many primes \mathfrak{p} are primes of good reduction for A_u . Similarly, for a family $A \rightarrow U$ over $\text{Spec } K$, extending the definition of this family “across” a prime $\mathfrak{p} \in \Sigma_K$ reveals constraints on the monodromy of that family and its subfamilies. The purpose of this section is to explain why any family $A \rightarrow U$ can be extended across most primes in Σ_K . The constructions introduced here become particularly important in Section 4F, where we apply the results of [Wallace 2014]. A similar treatment of these constructions can be found in [Wallace 2014, pp. 460–462].

Retain the setting of Theorem 1.1. Start with a family $A \rightarrow U$ of PPAVs over $\text{Spec } K$. Using standard spreading out techniques as in [EGA IV₃ 1966, §8] (see in particular [EGA IV₃ 1966, 8.10.5(xii), 9.7.7(ii); EGA IV₄ 1967, 17.7.8(ii)]), we can extend the family $A \rightarrow U$ to a family $\mathcal{A} \rightarrow \mathcal{U}$, where \mathcal{U} is an open subscheme of $\mathbb{P}_{\mathcal{O}_K}^r$, whose generic fiber over $\text{Spec } K \rightarrow \text{Spec } \mathcal{O}_K$ is just $A \rightarrow U$. Recall from Section 3A that the term “family” means that $\mathcal{A} \rightarrow \mathcal{U}$ is smooth and proper with geometrically connected fibers and that \mathcal{A} is an abelian scheme over \mathcal{U} with a principal polarization. This construction is depicted in the

following commutative diagram:

$$\begin{array}{ccc}
 A & \longrightarrow & \mathcal{A} \\
 \downarrow & & \downarrow \\
 U & \longrightarrow & \mathcal{U} \xrightarrow{\text{open emb.}} \mathbb{P}^r_{\mathcal{O}_K} \\
 \downarrow & & \downarrow \\
 \text{Spec } K & \longrightarrow & \text{Spec } \mathcal{O}_K
 \end{array}$$

Let $Z := \mathbb{P}^r_K \setminus U$ be the locus where the original family is not defined, and let \mathcal{Z} denote the closure of Z in $\mathbb{P}^r_{\mathcal{O}_K}$. Since the bottom square in the diagram above is Cartesian, each irreducible component of $\mathbb{P}^r_{\mathcal{O}_K} \setminus \mathcal{U}$ that is not contained in \mathcal{Z} cannot map generically onto $\text{Spec } \mathcal{O}_K$ and must therefore map to a single prime $\mathfrak{p} \in \Sigma_K$. Since there are finitely many irreducible components of \mathcal{Z} , the set S of primes $\mathfrak{p} \in \text{Spec } \mathcal{O}_K$ for which $\mathbb{P}^r_{\mathbb{F}_{\mathfrak{p}}} \setminus \mathcal{U}_{\mathbb{F}_{\mathfrak{p}}} \neq \mathcal{Z}_{\mathbb{F}_{\mathfrak{p}}}$ is a finite set. The primes in S can be thought of as the “bad primes” for the family: the smoothness of $\mathcal{A} \rightarrow \mathcal{U}$ implies that any abelian variety A_u for $u \in U(K)$ will have good reduction away from the primes in S and the primes lying under the (finite) intersection $\overline{\{u\}} \cap \mathcal{Z} \subset \mathbb{P}^r_{\mathcal{O}_K}$.

3B1. Monodromy groups of subfamilies. Let $m \in \mathbb{Z}$, let $P_m \subset \Sigma_K$ be the set of primes dividing m , and let $\text{Spec } \mathcal{O}_{P_m}$ be the complement of P_m in $\text{Spec } \mathcal{O}_K$. Then the base change $\mathcal{U}_{\mathcal{O}_{P_m}}$ of \mathcal{U} from $\text{Spec } \mathcal{O}_K$ to $\text{Spec } \mathcal{O}_{P_m}$ is the open subset of \mathcal{U} on which $\mathcal{A}[m] \rightarrow \mathcal{U}$ is unramified and hence finite étale. Therefore, we obtain a finite étale cover $\mathcal{A}_{\mathcal{O}_{P_m}}[m] \rightarrow \mathcal{U}_{\mathcal{O}_{P_m}}$ and hence a map $\rho : \pi_1(\mathcal{U}_{\mathcal{O}_{P_m}}) \rightarrow \text{GSp}_{2g}(\mathbb{Z}/m\mathbb{Z})$ just as in Section 3A. The original family of interest can be thought of as a subfamily of this one: we have maps $U_{\bar{K}} \rightarrow U \rightarrow \mathcal{U}_{\mathcal{O}_{P_m}}$, from which we obtain maps

$$\pi_1(U_{\bar{K}}) \longrightarrow \pi_1(U) \longrightarrow \pi_1(\mathcal{U}_{\mathcal{O}_{P_m}}) \xrightarrow{\rho} \text{GSp}_{2g}(\mathbb{Z}/m\mathbb{Z}).$$

Lemma 3.3. *The continuous map $\pi_1(U) \rightarrow \pi_1(\mathcal{U}_{\mathcal{O}_{P_m}})$ is surjective.*

Proof. This lemma is a consequence of [SGA 1 1971, exposé V, proposition 8.2]; we nonetheless include a proof because it helps illustrate the constructions introduced in this section. It suffices to show that the composition of this map with any surjective continuous map $\pi_1(\mathcal{U}_{\mathcal{O}_{P_m}}) \rightarrow G$ onto a finite group G is surjective. According to [Stacks 2005–, Tag 03SF], a finite quotient of the étale fundamental group corresponds to a connected finite Galois cover, so let $\mathcal{V}_m \rightarrow \mathcal{U}_{\mathcal{O}_{P_m}}$ be the cover corresponding to our chosen surjection. By [Stacks 2005–, Tag 0DV6], the composed map $\pi_1(U) \rightarrow \pi_1(\mathcal{U}_{\mathcal{O}_{P_m}}) \rightarrow G$ gives a $\pi_1(U)$ -action on G which corresponds to the pulled back cover $(\mathcal{V}_m)_K \rightarrow U$. The latter is connected if and only if the composed map is surjective. Since \mathcal{V}_m is connected and étale over $\text{Spec } \mathcal{O}_{P_m}$, it is irreducible, which implies that $(\mathcal{V}_m)_K$ is irreducible (its generic points correspond to those of \mathcal{V}_m), hence connected. □

By [Stacks 2005–, Tag 0DV6], the resulting monodromy representation $\pi_1(U) \rightarrow \text{GSp}_{2g}(\mathbb{Z}/m\mathbb{Z})$ equals that obtained from the pullback of the finite étale cover $\mathcal{A}_{\mathcal{O}_{P_m}}[m] \rightarrow \mathcal{U}_{\mathcal{O}_{P_m}}$ to U . But the pullback is just

the family $A[m] \rightarrow U$, so this monodromy representation equals $\rho_{A,m}$, and its image equals $H_A(m)$. The lemma therefore implies that the image of the map $\pi_1(\mathcal{U}_{\mathcal{O}_{P_m}}) \rightarrow \mathrm{GSp}_{2g}(\mathbb{Z}/m\mathbb{Z})$ equals $H_A(m)$. Similarly, the map $\pi_1(U_{\bar{K}}) \rightarrow \mathrm{GSp}_{2g}(\mathbb{Z}/m\mathbb{Z})$ has image equal to $H_A^{\mathrm{geom}}(m)$.

Moreover, for $\mathfrak{p} \in \Sigma_K$ not dividing m , we can also consider the subfamilies $\mathcal{U}_{\bar{\mathbb{F}}_{\mathfrak{p}}} \rightarrow \mathcal{U}_{\mathbb{F}_{\mathfrak{p}}} \rightarrow \mathcal{U}_{\mathcal{O}_{P_m}}$ obtained by extending scalars along the maps $\mathcal{O}_{P_m} \rightarrow \mathbb{F}_{\mathfrak{p}} \rightarrow \bar{\mathbb{F}}_{\mathfrak{p}}$ for some algebraic closure $\bar{\mathbb{F}}_{\mathfrak{p}}$ of $\mathbb{F}_{\mathfrak{p}}$. As before, we obtain maps

$$\pi_1(\mathcal{U}_{\bar{\mathbb{F}}_{\mathfrak{p}}}) \longrightarrow \pi_1(\mathcal{U}_{\mathbb{F}_{\mathfrak{p}}}) \longrightarrow \pi_1(\mathcal{U}_{\mathcal{O}_{P_m}}) \xrightarrow{\rho} \mathrm{GSp}_{2g}(\mathbb{Z}/m\mathbb{Z}).$$

We denote by $H_{A,\mathfrak{p}}(m)$ and $H_{A,\mathfrak{p}}^{\mathrm{geom}}$ the images of the maps $\pi_1(\mathcal{U}_{\mathbb{F}_{\mathfrak{p}}}) \rightarrow \mathrm{GSp}_{2g}(\mathbb{Z}/m\mathbb{Z})$ and $\pi_1(\mathcal{U}_{\bar{\mathbb{F}}_{\mathfrak{p}}}) \rightarrow \mathrm{GSp}_{2g}(\mathbb{Z}/m\mathbb{Z})$, respectively.

3B2. Notation for Galois étale covers. As explained in the proof of Lemma 3.3, finite quotients of the étale fundamental group correspond to connected finite Galois étale covers. We now fix notation for the Galois étale covers introduced in the proof of Lemma 3.3 that will be used later in Section 4 to state and verify Wallace’s criteria [2014].

- Let \mathcal{V}_m be the cover of $\mathcal{U}_{\mathcal{O}_{P_m}}$ corresponding to the map $\pi_1(\mathcal{U}_{\mathcal{O}_{P_m}}) \rightarrow \mathrm{GSp}_{2g}(\mathbb{Z}/m\mathbb{Z})$.
- Let V_m be the cover of U corresponding to the map $\pi_1(U) \rightarrow \mathrm{GSp}_{2g}(\mathbb{Z}/m\mathbb{Z})$.

Here, each map from $\pi_1(-)$ to a finite group gives a quotient of $\pi_1(-)$ as its image. By the reasoning of Lemma 3.3, $V_m = (\mathcal{V}_m)_K$.

Remark 3.4. The result of Lemma 3.3 is special to the base change $\mathcal{O}_K \rightarrow K$. In general, the other maps of $\pi_1(-)$ will not be surjective, nor will the finite Galois étale covers $(V_m)_{\bar{K}}$, $(\mathcal{V}_m)_{\mathbb{F}_{\mathfrak{p}}}$, and $(\mathcal{V}_m)_{\bar{\mathbb{F}}_{\mathfrak{p}}}$ be connected.

4. Proof of Theorem 1.1

4A. Outline of the proof. With the view of making the proof of Theorem 1.1 more readily comprehensible, we now briefly describe the key aspects of the argument. We encourage the reader to refer to Figure 1 for a schematic diagram illustrating the argument.

We begin in Section 4B by proving Proposition 4.1, showing that a nonisotrivial family with big monodromy also has big geometric monodromy. Then, in Section 4C, we introduce some of the notation and standing assumptions employed in the proof. In particular, since our family has big geometric monodromy, by Proposition 4.1, we are able to define the constant C in point (b) of Section 4C, which will later be needed to apply the results of [Wallace 2014] (see Section 4F1).

Then, in Section 4D, we reduce the problem to checking (1) that for an appropriately chosen integer M' depending on the family, most members of the family have the same mod- M' image as that of the family; and (2) that for all sufficiently large primes ℓ , most members of the family have the same mod- ℓ image as that of the family.

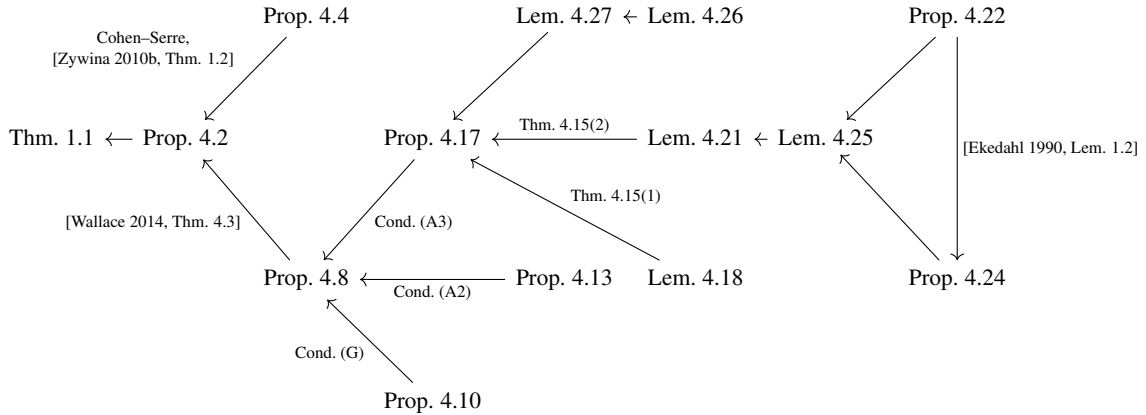


Figure 1. A schematic diagram for the proof of the main theorem, Theorem 1.1.

The mod- M' image is dealt with in Section 4E using Proposition 4.4, which is the Cohen–Serre version of the Hilbert irreducibility theorem. For dealing with the mod- ℓ images, there are two regimes of primes to consider, a medium regime and a high regime, when ℓ is bigger than a suitable power of $\log B$. We handle both of these regimes in Section 4F by applying a result of Wallace [2014, Theorem 3.9], for which we must verify the following four conditions: (G), (A1), (A2), and (A3). The rest of Section 4 is devoted to verifying that these conditions hold in our setting.

Conditions (G) and (A1), which are fairly easy to check, are treated in Sections 4F and 4G. Next, condition (A2) is dealt with in Section 4H by applying the Grothendieck specialization theorem in Proposition 4.13. These first three conditions together essentially yield an effective version of the Hilbert irreducibility theorem, which allows us to check primes ℓ in the medium regime. Finally, in Section 4I, we verify condition (A3), which allows us to dispense with primes in the high regime. The key input to checking this condition is a recent result of Lombardo, stated in Theorem 4.15. In order to apply Lombardo’s result to our setting, as is done in Proposition 4.17, we must verify two hypotheses and relate the naïve height we are using to the Faltings height used in Theorem 4.15. The first hypothesis is verified in Lemma 4.18 using [Ellenberg et al. 2009, Proposition 5]. The second hypothesis is a somewhat trickier condition, and we verify it in Proposition 4.21 using the large sieve, Theorem 4.19. In order to apply the large sieve, we must bound contributions at each prime, which is done in Proposition 4.24 using a general scheme-theoretic result of Ekedahl [1990, Lemma 1.2] together with Proposition 4.22. We conclude the section with a brief appendix concerning the relationship between the naïve height and the Faltings height (see Lemma 4.27).

4B. Equivalence of big geometric monodromy and big monodromy. In the course of the proof, it will be useful to know that our given family $A \rightarrow U$ not only has big monodromy, but also has big geometric monodromy. In particular, this is crucially needed to define the constant C in point (b) of Section 4C,

which is used in applying the results of [Wallace 2014] (see Section 4F1). We now prove the following result, implying that our given family has big geometric monodromy.

Proposition 4.1. *Suppose $A \rightarrow U$ is a nonisotrivial family of abelian varieties of relative dimension $g \geq 2$, with U a smooth geometrically connected scheme over a number field K . Then, A has big geometric monodromy if and only if it has big monodromy.*

Proof. We first show the easier direction: if the family $A \rightarrow U$ has big geometric monodromy then $A \rightarrow U$ also has big monodromy, in the sense that H_A is open in $\mathrm{GSp}_{2g}(\widehat{\mathbb{Z}})$. To see this, consider the exact sequence

$$0 \longrightarrow \mathrm{Sp}_{2g}(\widehat{\mathbb{Z}}) \longrightarrow \mathrm{GSp}_{2g}(\widehat{\mathbb{Z}}) \xrightarrow{\text{mult}} \widehat{\mathbb{Z}}^\times \longrightarrow 0.$$

Since $H_A^{\mathrm{geom}} \subset H_A$, the big geometric monodromy assumption tells us that $H_A \cap \mathrm{Sp}_{2g}(\widehat{\mathbb{Z}})$ is open in $\mathrm{Sp}_{2g}(\widehat{\mathbb{Z}})$. It therefore suffices to show that $\text{mult}(H_A)$ is open in $\widehat{\mathbb{Z}}^\times$. But $\text{mult}(H_A) = \chi(G_K)$, as mentioned in Remark 3.2, and $\chi(G_K)$ has finite index because K/\mathbb{Q} has finite degree.

It only remains to prove that if the family has big monodromy and is nonisotrivial, it has big geometric monodromy. To show this, from the exact sequence

$$1 \longrightarrow \pi_1(U_{\bar{K}}) \longrightarrow \pi_1(U) \longrightarrow \pi_1(K) \longrightarrow 1$$

$\pi_1(U_{\bar{K}}) \subset \pi_1(U)$ is normal. Therefore, H_A^{geom} is a normal subgroup of H_A , and hence also a normal subgroup of $H_A \cap \mathrm{Sp}_{2g}(\widehat{\mathbb{Z}})$. Let $\psi : \mathrm{Sp}_{2g}(\mathbb{Z}) \rightarrow \mathrm{Sp}_{2g}(\widehat{\mathbb{Z}})$ denote the natural profinite completion map. Since $H_A^{\mathrm{geom}} \subset H_A \cap \mathrm{Sp}_{2g}(\widehat{\mathbb{Z}})$ is normal, it follows that $\psi^{-1}(H_A^{\mathrm{geom}}) \subset \psi^{-1}(H_A \cap \mathrm{Sp}_{2g}(\widehat{\mathbb{Z}}))$ is normal. Since H_A has finite index in $\mathrm{GSp}_{2g}(\widehat{\mathbb{Z}})$, $\psi^{-1}(H_A \cap \mathrm{Sp}_{2g}(\widehat{\mathbb{Z}}))$ has finite index in $\mathrm{Sp}_{2g}(\mathbb{Z})$. Since $g \geq 2$ (so that $\mathrm{Sp}_{2g}(\mathbb{Z})$ has rank at least 2), by the Margulis normal subgroup theorem (see, for example [Morris 2015, Theorem 17.1.1]), $\psi^{-1}(H_A^{\mathrm{geom}})$ either has finite index in $\psi^{-1}(H_A \cap \mathrm{Sp}_{2g}(\widehat{\mathbb{Z}}))$ or is finite. We will show that in the first case A has big geometric monodromy and in the second case A is isotrivial.

In the case that $\psi^{-1}(H_A^{\mathrm{geom}})$ has finite index in $\psi^{-1}(H_A \cap \mathrm{Sp}_{2g}(\widehat{\mathbb{Z}}))$, $\psi^{-1}(H_A^{\mathrm{geom}})$ also has finite index in $\mathrm{Sp}_{2g}(\mathbb{Z})$. Then, since H_A^{geom} is closed, the finite set $\mathrm{Sp}_{2g}(\mathbb{Z})/\psi^{-1}(H_A^{\mathrm{geom}})$ is dense in the profinite space $\mathrm{Sp}_{2g}(\widehat{\mathbb{Z}})/H_A^{\mathrm{geom}}$. It follows that H_A^{geom} also has finite index in $\mathrm{Sp}_{2g}(\widehat{\mathbb{Z}})$, meaning A has big geometric monodromy.

To conclude the proof, it only remains to show that if $\psi^{-1}(H_A^{\mathrm{geom}})$ is finite, then A is isotrivial. In this case, let M_A^{geom} denote the image of the topological monodromy representation $\pi_1^{\mathrm{top}}(U_{\mathbb{C}}) \rightarrow \mathrm{Sp}_{2g}(\mathbb{Z})$. By [SGA 1 1971, exposé XIII, proposition 4.6], we have $\pi_1(U_{\mathbb{C}}) \simeq \pi_1(U_{\bar{K}})$, and therefore the comparison theorem tells us that H_A^{geom} is the profinite completion of M_A^{geom} . This implies $M_A^{\mathrm{geom}} \subset \psi^{-1}(H_A^{\mathrm{geom}})$ and so M_A^{geom} is finite. It follows that H_A^{geom} is finite, being the profinite completion of M_A^{geom} . After making a finite base change, we may assume H_A^{geom} is trivial. Then, it is a standard fact that A is isotrivial when its monodromy representation is trivial. For example, this follows from [Grothendieck 1966]. \square

4C. Notation and standing assumptions. Before proceeding with the proof, we set some notation and assumptions, which will remain in place for the remainder of this section.

- (a) As mentioned in Remark 1.2, the case where $g = 1$ is handled in [Zywina 2010b, Theorem 7.1], so we will restrict our consideration to the case where $g \geq 2$.
- (b) Since we are assuming that $A \rightarrow U$ has big monodromy, it follows that $A \rightarrow U$ has big geometric monodromy, by Proposition 4.1. Define C to be the smallest integer bigger than 2, depending only on U , with the property that for all primes $\ell > C$ we have $H_A^{\text{geom}}(\ell) = \text{Sp}_{2g}(\mathbb{Z}/\ell\mathbb{Z})$ and $H_A(\ell) = \text{GSp}_{2g}(\mathbb{Z}/\ell\mathbb{Z})$.
- (c) Using [Zywina 2010b, Proposition 6.1] and the explanation given after the statement of [Zywina 2010b, Theorem 7.1], one readily checks that in Theorem 1.1, the asymptotic statement for K -valued points (i.e., points in $U(K)$) can be deduced immediately from the statement for *lattice points* (i.e., points in $U(K) \cap \mathcal{O}_K^r$). In what follows, we will work with K -valued points or lattice points depending on what is most convenient.
- (d) Let $K^{\text{cyc}} \subset \bar{K}$ denote the maximal cyclotomic extension of K , and let $K^{\text{ab}} \subset \bar{K}$ denote the maximal abelian extension of K .
- (e) In what follows, for a subgroup H of a topological group G , let $[H, H]$ denote the *closure* of the usual commutator subgroup.

4D. Main body of the proof. We begin by reducing the proof of Theorem 1.1 to proving Proposition 4.2.

Proof of Theorem 1.1 assuming Proposition 4.2. As argued in [Zywina 2010b, Proof of Theorem 7.1], for any $u \in U(K)$ we have

$$[H_A : H_{A_u}] = [H_A \cap \text{Sp}_{2g}(\widehat{\mathbb{Z}}) : \rho_{A_u}(\text{Gal}(\bar{K}/K^{\text{cyc}}))].$$

In the case that $K = \mathbb{Q}$, the Kronecker–Weber Theorem tells us that $\mathbb{Q}^{\text{cyc}} = \mathbb{Q}^{\text{ab}}$, so we have

$$[H_A : H_{A_u}] = \delta_{\mathbb{Q}} \cdot [[H_A, H_A] : \rho_{A_u}(\text{Gal}(\bar{\mathbb{Q}}/\mathbb{Q}^{\text{ab}}))],$$

where $\delta_{\mathbb{Q}}$ is the index of $[H_A, H_A]$ in $H_A \cap \text{Sp}_{2g}(\widehat{\mathbb{Z}})$. Then Theorem 1.1 follows immediately from point (c) of Section 4C and the following proposition. \square

Proposition 4.2. *Let $B, n > 0$. We have the following asymptotic statements, where the implied constants depend only on U and n :*

- (1) *For every number field K ,*

$$\frac{|\{u \in U(K) \cap \mathcal{O}_K^r : \|u\| \leq B, \rho_{A_u}(\text{Gal}(\bar{K}/K^{\text{ab}})) = [H_A, H_A]\}|}{|\{u \in U(K) \cap \mathcal{O}_K^r : \|u\| \leq B\}|} = 1 + O((\log B)^{-n}).$$

- (2) *Furthermore, if $K \neq \mathbb{Q}$,*

$$\frac{|\{u \in U(K) \cap \mathcal{O}_K^r : \|u\| \leq B, \rho_{A_u}(\text{Gal}(\bar{K}/K^{\text{cyc}})) = H_A \cap \text{Sp}_{2g}(\widehat{\mathbb{Z}})\}|}{|\{u \in U(K) \cap \mathcal{O}_K^r : \|u\| \leq B\}|} = 1 + O((\log B)^{-n}).$$

Remark 4.3. Proposition 4.2 is a generalization of [Zywina 2010b, Proposition 7.9] from the case $g = 1$ to all dimensions. We shall prove it assuming Proposition 4.4 and Proposition 4.8. The basic idea behind the argument is to reduce the problem of studying the (global) monodromy groups to one of studying the mod- M' and mod- ℓ monodromy groups.

Proof assuming Proposition 4.4 and Proposition 4.8. Assuming point (1), the proof of point (2) is completely analogous to the proof of [Zywina 2010b, Proposition 7.9(ii)], which consists of two key steps. The first is the fact that $[H_A, H_A]$ is an open normal subgroup of $H_A \cap \mathrm{Sp}_{2g}(\widehat{\mathbb{Z}})$, which follows from Proposition 2.6. The second is [Zywina 2010b, Proposition 7.7], which is a variant of Hilbert’s irreducibility theorem and does not depend in any way on the context of elliptic curves (with which [Zywina 2010b, Section 7] is concerned). It therefore suffices to prove point (1).

Since $\mathrm{Gal}(\bar{K}/K^{\mathrm{ab}}) = [G_K, G_K]$, it follows by the continuity of ρ_{A_u} and the compactness of profinite groups that $\rho_{A_u}(\mathrm{Gal}(\bar{K}/K^{\mathrm{ab}})) = [H_{A_u}, H_{A_u}]$. Thus $\rho_{A_u}(\mathrm{Gal}(\bar{K}/K^{\mathrm{ab}}))$ is a closed subgroup of $[H_A, H_A]$. Moreover, by Proposition 2.6, $[H_A, H_A]$ is an open subgroup of $\mathrm{Sp}_{2g}(\widehat{\mathbb{Z}})$, so we may apply Proposition 2.5 with $G = [H_A, H_A]$ and $H = \rho_{A_u}(\mathrm{Gal}(\bar{K}/K^{\mathrm{ab}}))$. In so doing, we obtain a positive integer M so that the only closed subgroup of $[H_A, H_A]$ whose mod- M reduction equals $[H_A, H_A](M) = [H_A(M), H_A(M)]$ and whose mod- ℓ reduction equals $\mathrm{Sp}_{2g}(\mathbb{Z}/\ell\mathbb{Z})$ for every prime number $\ell \nmid M$ is $[H_A, H_A]$ itself. The same property is true when M is replaced by any multiple M' of M , and we choose a multiple M' which is divisible by all primes less than C , where C is defined as in point (b) of Section 4C. The defining property of M' then implies that

$$\frac{|\{u \in U(K) \cap \mathcal{O}_K^r : \|u\| \leq B, \rho_{A_u}(\mathrm{Gal}(\bar{K}/K^{\mathrm{ab}})) \neq [H_A, H_A]\}|}{|\{u \in U(K) \cap \mathcal{O}_K^r : \|u\| \leq B\}|} \leq \frac{|\{u \in U(K) \cap \mathcal{O}_K^r : \|u\| \leq B, \rho_{A_u, M'}(\mathrm{Gal}(\bar{K}/K^{\mathrm{ab}})) \neq [H_A(M'), H_A(M')]\}|}{|\{u \in U(K) \cap \mathcal{O}_K^r : \|u\| \leq B\}|} \tag{4-1}$$

$$+ \frac{|\{u \in U(K) \cap \mathcal{O}_K^r : \|u\| \leq B, \rho_{A_u, \ell}(\mathrm{Gal}(\bar{K}/K^{\mathrm{ab}})) \neq \mathrm{Sp}_{2g}(\mathbb{Z}/\ell\mathbb{Z}) \text{ for some } \ell \nmid M'\}|}{|\{u \in U(K) \cap \mathcal{O}_K^r : \|u\| \leq B\}|}. \tag{4-2}$$

The rest of this section is devoted to finding upper bounds for (4-1) and (4-2). To bound (4-1), notice that we have

$$\rho_{A_u, M'}(\mathrm{Gal}(\bar{K}/K^{\mathrm{ab}})) \neq [H_A(M'), H_A(M')] \implies H_{A_u}(M') \neq H_A(M').$$

It then follows from Proposition 4.4 that (4-1) is bounded by $O((\log B)/B^{[K:\mathbb{Q}]/2})$. To bound (4-2), notice that for $\ell \geq 3$ we have

$$\rho_{A_u, \ell}(\mathrm{Gal}(\bar{K}/K^{\mathrm{ab}})) \neq \mathrm{Sp}_{2g}(\mathbb{Z}/\ell\mathbb{Z}) \implies H_{A_u}(\ell) \not\supseteq \mathrm{Sp}_{2g}(\mathbb{Z}/\ell\mathbb{Z}),$$

because [Landesman et al. 2017b, Proposition 1(a)] tells us that $\mathrm{Sp}_{2g}(\mathbb{Z}/\ell\mathbb{Z})$ has trivial abelianization for $\ell \geq 3$. Since $C \geq 3$ by definition, it follows from Proposition 4.8 that (4-2) is $O((\log B)^{-n})$, since $\ell \nmid M'$ implies that $\ell > C$. Combining the above estimates completes the proof of point (1). \square

It now remains to bound the terms (4-1) and (4-2).

4E. Bounding the contribution of (4-1). The next result is the means by which we bound (4-1); it is an immediate corollary of the Cohen–Serre version of Hilbert’s irreducibility theorem (see [Zywina 2010b, Theorem 1.2]) since the set in the numerator of (4-3) is a “thin set.”

Proposition 4.4. *For every integer $M' \geq 2$, we have*

$$\frac{|\{u \in U(K) \cap \mathcal{O}_K^r : \|u\| \leq B, H_{A_u}(M') \neq H_A(M')\}|}{|\{u \in U(K) \cap \mathcal{O}_K^r : \|u\| \leq B\}|} \ll \frac{\log B}{B^{[K:\mathbb{Q}]/2}}, \tag{4-3}$$

where the implied constant depends only in U and M' .⁴

4F. Bounding the contribution of (4-2). To complete the proof of Theorem 1.1, it remains to bound (4-2). We do this in Proposition 4.8, which relies on a strong version of Hilbert’s irreducibility theorem due to Wallace [2014, Theorem 3.9]. Before we can state and apply Wallace’s result, we must introduce the various conditions upon which it depends. The setup detailed in [Wallace 2014, Section 3.2] applies in a more general context than the one described below, but we specialize our discussion for the sake of brevity.

4F1. Setup and statement of [Wallace 2014, Theorem 3.9]. We start by introducing some notation to help us count points $u \in U(K)$ whose associated monodromy groups H_{A_u} are not maximal. Let $B > 0$, and make the following two definitions:

$$E_\ell(B) := \{u \in U(K) : \text{Ht}(u) \leq B, H_A^{\text{geom}}(\ell) \not\subset H_u(\ell)\}, \quad \text{and}$$

$$E(B) := \bigcup_{\text{prime } \ell > C} E_\ell(B),$$

where C is defined as in point (b) of Section 4C. The set $E_\ell(B)$ should be thought of as the set of exceptional points of height bounded by B for the ℓ -adic representation, and the set $E(B)$ should likewise be thought of as the set of points of height bounded by B that are exceptional for some $\ell > C$. Note in particular that for any $\ell > C$ we have $H_A(\ell)/H_A^{\text{geom}}(\ell) \simeq (\mathbb{Z}/\ell\mathbb{Z})^\times$; this condition is important for the proof of [Wallace 2014, Theorem 3.9] to go through, so we impose the following restriction:

$$\text{For the rest of this section, we will maintain } \ell > C \text{ as a standing assumption.} \tag{4-4}$$

For ease of notation, we redefine the set $S \subset \Sigma_K$ of “bad” primes, defined in Section 3B, by adjoining to it all primes $\ell < C$.

Remark 4.5. Note that our definition of the exceptional set $E(B)$ differs slightly from that given in [Wallace 2014, Theorem 1.1], where it is defined to be the union over *all* primes ℓ of the ℓ -adic exceptional sets $E_\ell(B)$. This difference is inconsequential, as we can always deal with a finite collection of primes using Proposition 4.4. Indeed, this is exactly why we replace M by a multiple M' divisible by all primes $\ell < C$ in the proof of Proposition 4.2.

⁴For functions f, g in the variable B , we say that $f(B) \ll g(B)$ if there exists a constant $c > 0$ such that $|f(B)| \leq c \cdot |g(B)|$ for all sufficiently large B .

Now that we have introduced the setup needed for stating [Wallace 2014, Theorem 3.9], we declare the four criteria required for the theorem to be applied. For this, it will now be crucial to recall notation from the geometric setup detailed in Section 3B.

Conditions 4.6. Recall from Section 3B that P_m denotes the set of primes of \mathcal{O}_K dividing an integer m and that V_ℓ denotes the connected Galois étale cover of U giving rise to the monodromy group $H_A(\ell)$ for a prime ℓ . In order to apply [Wallace 2014, Theorem 3.9], we need to verify the following geometric condition on the covers $V_\ell \rightarrow U$ as ℓ ranges through the primes greater than C :

- (G) Let ζ_ℓ denote a primitive ℓ^{th} root of unity. Each connected component of the base-change $(V_\ell)_{K(\zeta_\ell)}$ is geometrically irreducible.

We also need the following three asymptotic conditions concerning the monodromy groups $H_A(\ell)$, $H_A^{\text{geom}}(\ell)$, and $H_{A,\mathfrak{p}}(\ell)$ for [Wallace 2014, Theorem 3.9] to be applied:

- (A1) There exist constants $\beta_1, \beta_2 > 0$ such that

$$|H_A(\ell)| \ll \ell^{\beta_1} \quad \text{and} \quad |\{\text{conjugacy classes of } H_A(\ell)\}| \ll \ell^{\beta_2},$$

where the implied constants depend only on U .

- (A2) There exists a constant $\beta_3 > 0$ such that

$$\mathcal{T}_\ell := \left| \left\{ \text{prime } \mathfrak{p} \subset \mathcal{O}_K : \mathfrak{p} \in S \cup P_\ell \text{ or } H_{A,\mathfrak{p}}^{\text{geom}}(\ell) \neq H_A^{\text{geom}}(\ell) \right\} \right| \ll \ell^{\beta_3},$$

where the implied constant depends only on $A \rightarrow U$.

- (A3) For each $B > 0$, there exists a subset

$$F(B) \subset \{u \in U(K) : \text{Ht}(u) \leq B\}$$

and constants $c, \gamma > 0$ depending only on $A \rightarrow U$ such that

$$\lim_{B \rightarrow \infty} \frac{|F(B)|}{|\{u \in U(K) : \text{Ht}(u) \leq B\}|} = 1 \quad \text{and} \quad F(B) \cap E(B) \subset \bigcup_{\ell \leq c(\log B)^\gamma} E_\ell(B).$$

We are now in a position to state Wallace’s main result:

Theorem 4.7 [Wallace 2014, Theorem 3.9]. *Suppose that condition (G) holds and that conditions (A1)–(A3) hold with the values $\beta_1, \beta_2, \beta_3, \gamma$.⁵ Then we have the following bound on the proportion of exceptional points of height bounded by B :*

$$\frac{|E(B)|}{|\{u \in U(K) : \text{Ht}(u) \leq B\}|} \ll \frac{|\{u \in U(K) : \text{Ht}(u) \leq B\} \setminus F(B)|}{|\{u \in U(K) : \text{Ht}(u) \leq B\}|} + \frac{(\log B)^{(\beta_1 + \beta_2 + 2)\gamma + 1}}{B^{\frac{1}{2}}}, \quad (4-5)$$

where the implied constant depends only on U .

⁵The constant c from condition (A3) is absorbed into the implied constant in (4-5).

4F2. *Bounding (4-2), conditional on verifying (G), (A2), and (A3).* We have not yet determined that Conditions 4.6 hold in our setting. We defer the verification of these conditions to Sections 4G, 4H, and 4I. Nevertheless, assuming that these conditions hold, we obtain the following consequence:

Proposition 4.8. *Let $n > 0$. Then we have*

$$\frac{|\{u \in U(K) \cap \mathcal{O}_K^r : \|u\| \leq B, H_{A_u}(\ell) \not\subset \mathrm{Sp}_{2g}(\mathbb{Z}/\ell\mathbb{Z}) \text{ for some } \ell > C\}|}{|\{u \in U(K) \cap \mathcal{O}_K^r : \|u\| \leq B\}|} \ll (\log B)^{-n}, \quad (4-6)$$

where the implied constant depends only on U and n .

Proof assuming Propositions 4.10, 4.13, and 4.17. Note that condition (A1) holds trivially in our setting, because

$$\max\{|H_A(\ell)|, |\{\text{conjugacy classes of } H_A(\ell)\}|\} \leq |\mathrm{GSp}_{2g}(\mathbb{Z}/\ell\mathbb{Z})|,$$

and $|\mathrm{GSp}_{2g}(\mathbb{Z}/\ell\mathbb{Z})| = O(\ell^\beta)$ for some positive constant β depending only on g because $\mathrm{GSp}_{2g}(\mathbb{Z}/\ell\mathbb{Z}) \subset \mathrm{GL}_{2g}(\mathbb{Z}/\ell\mathbb{Z})$.

Condition (G) holds by Proposition 4.10, and condition (A2) holds by Proposition 4.13. Proposition 4.17 constructs $F(B)$ that not only satisfy condition (A3), but also have the property that

$$\frac{|\{u \in U(K) : \mathrm{Ht}(u) \leq B\} \setminus F(B)|}{|\{u \in U(K) : \mathrm{Ht}(u) \leq B\}|} \ll (\log B)^{-n}$$

for every $n > 0$. Upon applying the argument in point (c) of Section 4C, which relates the left-hand sides of (4-5) and (4-6), the proposition follows from Theorem 4.7. \square

The rest of this section is devoted to verifying the conditions necessary for the proof of Proposition 4.8.

4G. Verifying condition (G). In this section, we will consider the base-change of the setting established in 3B from K to a finite extension $L \subset \bar{K}$ of K ; in this setting, we obtain a family $A_L \rightarrow U_L$ and a (not necessarily connected) finite Galois étale cover $(V_\ell)_L \rightarrow U_L$. To verify condition (G), we employ the following lemma:

Lemma 4.9. *Let $L \subset \bar{K}$ be a finite extension of K . We have that $H_{A_L}(m) \simeq H_{A_L}^{\mathrm{geom}}(m)$ if and only if all connected components of $(V_m)_L$ are geometrically connected over L .*

Proof. Observe that $(V_m)_L$ and $(V_m)_{\bar{K}}$ are finite Galois étale covers of U_L and $U_{\bar{K}}$, which need not be connected.

Let $W \subset (V_m)_L$ be a connected component, and let $\tilde{W} \subset (V_m)_{\bar{K}}$ be a connected component mapping to W . By construction, $W \rightarrow U_L$ is the connected Galois étale cover corresponding to the surjection $\pi_1(U_L) \twoheadrightarrow H_{A_L}(m)$. Likewise, $\tilde{W} \rightarrow U_{\bar{K}}$ corresponds to $\pi_1(U_{\bar{K}}) \twoheadrightarrow H_A^{\mathrm{geom}}(m) = H_{A_L}^{\mathrm{geom}}(m)$. This implies that:

- The degree d_1 of $W \rightarrow U_L$ equals $|H_{A_L}(m)|$.
- The degree d_2 of $\tilde{W} \rightarrow U_{\bar{K}}$ equals $|H_{A_L}^{\mathrm{geom}}(m)|$.

On the other hand, the maps $(V_m)_L \rightarrow U_L$ and $(V_m)_{\bar{K}} \rightarrow U_{\bar{K}}$ have equal degrees. Therefore $d_1 = d_2$ if and only if all connected components of $(V_m)_L$ are geometrically connected. \square

We are now in position to prove condition (G).

Proposition 4.10. *Condition (G) holds in the setting of Section 3B.*

Proof. Let $L = K(\zeta_\ell)$, and recall the assumption (4-4). Since $(V_\ell)_L \rightarrow U_L$ is étale and U_L is smooth over L , it follows that $(V_\ell)_L$ is smooth over L . Therefore $(V_\ell)_L$ is geometrically irreducible over L if and only if it is geometrically connected over L . Now, by Lemma 4.9, it suffices to show that $H_{A_L}(\ell) = H_A^{\text{geom}}(\ell)$. Since we always have $H_{A_L}(\ell) \supset H_A^{\text{geom}}(\ell)$, it suffices to prove the reverse inclusion $H_{A_L}(\ell) \subset H_A^{\text{geom}}(\ell) = \text{Sp}_{2g}(\mathbb{Z}/\ell\mathbb{Z})$. Since χ_ℓ is trivial on $G_L = \pi_1(\text{Spec } K(\zeta_\ell))$, it follows from Remark 3.2 that $H_{A_L}(\ell) \subset \text{Sp}_{2g}(\mathbb{Z}/\ell\mathbb{Z})$. \square

4H. Verifying condition (A2). Before we carry out the verification of condition (A2) in Proposition 4.13, we need to introduce a modified version of the geometric setup developed in [Zywina 2010b, Section 5.2] and in the proof of [Zywina 2010b, Theorem 5.3].

4H1. Geometric setup from [Zywina 2010b]. Fix the following notation: for a prime $\mathfrak{p} \subset \mathcal{O}_K$, let $K_{\mathfrak{p}}$ be the completion of K at \mathfrak{p} , let $K_{\mathfrak{p}}^{\text{un}}$ be the maximal unramified extension of $K_{\mathfrak{p}}$, let $\mathcal{O}_{\mathfrak{p}}$ be the ring of integers of $K_{\mathfrak{p}}$, and let $\mathcal{O}_{\mathfrak{p}}^{\text{un}}$ be the ring of integers of $K_{\mathfrak{p}}^{\text{un}}$. For a ring R , define $\text{Gr}_R(1, r)$ to be the Grassmannian of lines in \mathbb{P}_R^r and let $\mathcal{L}_R \subset \mathbb{P}_R^r \times \text{Gr}_R(1, r)$ denote the universal line over $\text{Gr}_R(1, r)$. Let Z and \mathcal{Z} be as defined in Section 3B.

We now construct a closed subscheme \mathcal{W} of the Grassmannian parametrizing all lines whose intersections with \mathcal{Z} are not étale over the base. Define the projection $p : \mathcal{L}_{\mathcal{O}_K} \cap (\mathcal{Z} \times \text{Gr}_{\mathcal{O}_K}(1, r)) \rightarrow \text{Gr}_{\mathcal{O}_K}(1, r)$. Let \mathcal{X}_1 be the open subscheme of $\mathcal{L}_{\mathcal{O}_K} \cap (\mathcal{Z} \times \text{Gr}_{\mathcal{O}_K}(1, r))$ on which p is étale with nonempty fibers. Define $\mathcal{W} := p(\mathcal{L}_{\mathcal{O}_K} \cap (\mathcal{Z} \times \text{Gr}_{\mathcal{O}_K}(1, r)) \setminus \mathcal{X}_1)$ with reduced subscheme structure and define $\mathcal{X} := \text{Gr}_{\mathcal{O}_K}(1, r) \setminus \mathcal{W}$. Note that \mathcal{W} is closed because p is proper. Considering \mathcal{W} and \mathcal{X} as schemes over \mathcal{O}_K , let W and X denote their fibers over K .

Lemma 4.11. *The scheme \mathcal{W} , as defined above, is a proper closed subscheme of $\text{Gr}_{\mathcal{O}_K}(1, r)$.*

Proof. It suffices to show that \mathcal{X} is nonempty. In turn, it suffices to show X is nonempty. Since X is the set of points in $\text{Gr}_K(1, r)$ over which p is étale, by generic flatness, we need only verify that there is an open subscheme of $\text{Gr}_K(1, r)$ on which the fibers of p_K are étale. Since Z is reduced, hence generically smooth, and the fiber of p_K over $[L]$ is identified with $Z \cap L$, a Bertini theorem (specifically [Jouanolou 1983, Theoreme I.6.10(2)] applied to the smooth locus of Z over K) implies that $Z \cap L$ is indeed étale over $\kappa([L])$ for $[L]$ general in $\text{Gr}_K(1, r)$. \square

Remark 4.12. By Lemma 4.11, \mathcal{W} is a proper closed subscheme of $\text{Gr}_{\mathcal{O}_K}(1, r)$. Observe that for any line $[L] \in (\text{Gr}_{\mathcal{O}_K}(1, r) \setminus \mathcal{W})(\mathbb{F}_{\mathfrak{p}})$, there exists a lift $[\mathcal{L}] \in (\text{Gr}_{\mathcal{O}_K}(1, r) \setminus \mathcal{W})(\mathcal{O}_{\mathfrak{p}})$. The purpose of the above construction is to ensure that $\mathcal{L} \cap \mathcal{Z}_{\mathcal{O}_{\mathfrak{p}}}$ is étale over $\mathcal{O}_{\mathfrak{p}}$, which we use in the proof of Proposition 4.13.

4H2. *Applying the setup to check (A2).* In the following proposition, we use the Grothendieck specialization theorem to verify that condition (A2) holds in our situation:

Proposition 4.13. *For a prime ideal $\mathfrak{p} \subset \mathcal{O}_K$ let $N(\mathfrak{p})$ denote its norm and define S' to be the finite set of primes over which the fiber of \mathcal{W} is empty. Then,*

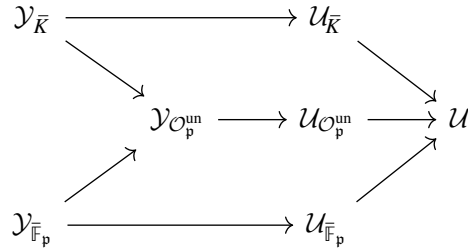
$$\mathcal{T}_\ell \leq |S' \cup P_\ell| + |\{\text{primes } \mathfrak{p} \subset \mathcal{O}_K : \gcd(N(\mathfrak{p}), |\mathrm{Sp}_{2g}(\mathbb{Z}/\ell\mathbb{Z})|) \neq 1\}|.$$

In particular, we have that \mathcal{T}_ℓ is bounded by a fixed power of ℓ , so condition (A2) holds in the setting of Section 3B.

Remark 4.14. In fact, it is true that $\mathcal{T}_\ell \ll \log \ell$. Apart from a finite number of primes depending only on the family $A \rightarrow U$, we need only throw out those primes whose norms are not coprime to $|\mathrm{GSp}_{2g}(\mathbb{Z}/\ell\mathbb{Z})|$. Since $|\mathrm{GSp}_{2g}(\mathbb{Z}/\ell\mathbb{Z})|$ grows polynomially in ℓ , the number of distinct primes dividing $|\mathrm{GSp}_{2g}(\mathbb{Z}/\ell\mathbb{Z})|$ is at most logarithmic in ℓ .

Proof of Proposition 4.13. Take a prime ideal $\mathfrak{p} \notin S' \cup P_\ell$ so that $\gcd(N(\mathfrak{p}), |\mathrm{GSp}_{2g}(\mathbb{Z}/\ell\mathbb{Z})|) = 1$. It suffices to show $H_{A,\mathfrak{p}}^{\mathrm{geom}}(\ell) = \mathrm{Sp}_{2g}(\mathbb{Z}/\ell\mathbb{Z}) = H_A^{\mathrm{geom}}(\ell)$.

Choose $[\mathcal{L}] \in (\mathrm{Gr}_{\mathcal{O}_K}(1, r) \setminus \mathcal{W})(\mathcal{O}_\mathfrak{p})$, which exists by Remark 4.12. Furthermore, define $\mathcal{D} := \mathcal{L} \cap \mathcal{Z}_{\mathcal{O}_\mathfrak{p}}$ and $\mathcal{Y} := \mathcal{L} \setminus \mathcal{D}$. We have the commutative diagram



where all of the horizontal arrows are embeddings. Let $\pi_1^{(p)}$ denote the largest prime to p quotient of the fundamental group. Note that $\rho_{A,\ell}$ factors through $\pi_1^{(p)}(\mathcal{U})$ because we are assuming $\gcd(N(\mathfrak{p}), |\mathrm{GSp}_{2g}(\mathbb{Z}/\ell\mathbb{Z})|) = 1$. By applying the prime to $N(\mathfrak{p})$ étale fundamental group functor to the above diagram, we obtain

$$\begin{array}{ccccccc}
 \pi_1^{(N(\mathfrak{p}))}(\mathcal{Y}_{\bar{K}}) & \xrightarrow{\iota_{\bar{K}}} & \pi_1^{(N(\mathfrak{p}))}(\mathcal{U}_{\bar{K}}) & & & & \\
 \downarrow \phi & \searrow \alpha_{\bar{K}} & & \searrow \beta_{\bar{K}} & & & \\
 & & \pi_1^{(N(\mathfrak{p}))}(\mathcal{Y}_{\mathcal{O}_\mathfrak{p}^{\mathrm{un}}}) & \xrightarrow{\iota_{\mathcal{O}_\mathfrak{p}^{\mathrm{un}}}} & \pi_1^{(N(\mathfrak{p}))}(\mathcal{U}_{\mathcal{O}_\mathfrak{p}^{\mathrm{un}}}) & \xrightarrow{\beta_{\mathcal{O}_\mathfrak{p}^{\mathrm{un}}}} & \pi_1^{(N(\mathfrak{p}))}(\mathcal{U}) \xrightarrow{\rho_{A,\ell}} \mathrm{GSp}_{2g}(\mathbb{Z}/\ell\mathbb{Z}) \quad (4-7) \\
 & \nearrow \alpha_{\bar{\mathbb{F}}_\mathfrak{p}} & & \nearrow \beta_{\bar{\mathbb{F}}_\mathfrak{p}} & & & \\
 \pi_1^{(N(\mathfrak{p}))}(\mathcal{Y}_{\bar{\mathbb{F}}_\mathfrak{p}}) & \xrightarrow{\iota_{\bar{\mathbb{F}}_\mathfrak{p}}} & \pi_1^{(N(\mathfrak{p}))}(\mathcal{U}_{\bar{\mathbb{F}}_\mathfrak{p}}) & & & &
 \end{array}$$

By Remark 4.12, \mathcal{D} is étale over $\mathcal{O}_\mathfrak{p}$. By the Grothendieck specialization theorem, [Orgogozo and Vidal 2000, théorème 4.4], there is a map $\phi : \pi_1^{(N(\mathfrak{p}))}(\mathcal{Y}_{\bar{K}}) \xrightarrow{\sim} \pi_1^{(N(\mathfrak{p}))}(\mathcal{Y}_{\bar{K}_\mathfrak{p}}) \rightarrow \pi_1^{(N(\mathfrak{p}))}(\mathcal{Y}_{\bar{\mathbb{F}}_\mathfrak{p}})$ which makes the

triangle on the left in (4-7) commute and induces an isomorphism on the largest prime-to- $N(\mathfrak{p})$ quotients of the source and target. Note that $\pi_1^{(N(\mathfrak{p}))}(\mathcal{Y}_{\bar{K}}) \xrightarrow{\sim} \pi_1^{(N(\mathfrak{p}))}(\mathcal{Y}_{\bar{K}_{\mathfrak{p}}})$ is an isomorphism by [SGA 1 1971, exposé XIII, proposition 4.6]. Since the rest of the diagram (4-7) commutes, the entire diagram commutes.

Now, observe that we have

$$(\rho_{A,\ell} \circ \beta_{\bar{K}})(\pi_1^{(N(\mathfrak{p}))}(\mathcal{U}_{\bar{K}})) = H_A^{\text{geom}}(\ell) = \text{Sp}_{2g}(\mathbb{Z}/\ell\mathbb{Z})$$

where the last step follows from the Equation (4-4). By [Zywina 2010b, Lemma 5.2], (since the scheme W used in [Zywina 2010b, Lemma 5.2] is contained in the scheme W we have constructed above) we have that

$$(\rho_{A,\ell} \circ \beta_{\bar{K}} \circ \iota_{\bar{K}})(\pi_1^{(N(\mathfrak{p}))}(\mathcal{Y}_{\bar{K}})) = H_A^{\text{geom}}(\ell) = \text{Sp}_{2g}(\mathbb{Z}/\ell\mathbb{Z}).$$

Since ϕ is an isomorphism, we deduce that

$$\begin{aligned} (\rho_{A,\ell} \circ \beta_{\bar{\mathbb{F}}_{\mathfrak{p}}} \circ \iota_{\bar{\mathbb{F}}_{\mathfrak{p}}})(\pi_1^{(N(\mathfrak{p}))}(\mathcal{Y}_{\bar{\mathbb{F}}_{\mathfrak{p}}})) &= (\rho_{A,\ell} \circ \beta_{\bar{\mathbb{F}}_{\mathfrak{p}}} \circ \iota_{\bar{\mathbb{F}}_{\mathfrak{p}}} \circ \phi)(\pi_1^{(N(\mathfrak{p}))}(\mathcal{Y}_{\bar{K}})) \\ &= (\rho_{A,\ell} \circ \beta_{\bar{K}} \circ \iota_{\bar{K}})(\pi_1^{(N(\mathfrak{p}))}(\mathcal{Y}_{\bar{K}})) \\ &= \text{Sp}_{2g}(\mathbb{Z}/\ell\mathbb{Z}). \end{aligned}$$

Therefore, $\text{Sp}_{2g}(\mathbb{Z}/\ell\mathbb{Z}) \subset (\rho_{A,\ell} \circ \beta_{\bar{\mathbb{F}}_{\mathfrak{p}}})(\pi_1^{(N(\mathfrak{p}))}(\mathcal{U}_{\bar{\mathbb{F}}_{\mathfrak{p}}})) = H_{A,\mathfrak{p}}^{\text{geom}}(\ell)$. Since $\ell \nmid N(\mathfrak{p})$, we have that $\bar{\mathbb{F}}_{\mathfrak{p}}$ contains nontrivial ℓ^{th} roots of unity. Thus, the mod- ℓ cyclotomic character is trivial on $\pi_1^{(N(\mathfrak{p}))}(\mathcal{U}_{\bar{\mathbb{F}}_{\mathfrak{p}}})$, and so $\text{Sp}_{2g}(\mathbb{Z}/\ell\mathbb{Z}) \supset H_{A,\mathfrak{p}}^{\text{geom}}(\ell)$. Hence, we have that

$$H_{A,\mathfrak{p}}^{\text{geom}}(\ell) = \text{Sp}_{2g}(\mathbb{Z}/\ell\mathbb{Z}) = H_A^{\text{geom}}(\ell). \quad \square$$

4I. Verifying condition (A3). It remains to check that condition (A3) is satisfied in our setting. As usual, before carrying out the argument, we must fix some notation. Let Σ_K denote the set of nonzero prime ideals of \mathcal{O}_K , and for a prime $\mathfrak{p} \in \Sigma_K$ of good reduction, let $\text{Frob}_{\mathfrak{p}} \in G_K$ denote a corresponding Frobenius element.

Given a PPAV A/K , let $\text{ch}_A(\text{Frob}_{\mathfrak{p}})$ denote the characteristic polynomial of $\rho_A(\text{Frob}_{\mathfrak{p}}) \in \text{GSp}_{2g}(\widehat{\mathbb{Z}})$, and observe that $\text{ch}_A(\text{Frob}_{\mathfrak{p}})$ has coefficients in \mathbb{Z} . Finally, let $h(A)$ denote the absolute logarithmic Faltings height of A , obtained by passing to any field extension over which A has semi-stable reduction.

4I1. Applying Lombardo's result. The key input for our proof of this condition is the following theorem of Lombardo, which is an effective version of the open image theorem:

Theorem 4.15 ([Lombardo 2016a, Theorem 1.2] and Proposition A.2 in the Appendix). *Let A/K be a PPAV of dimension $g \geq 2$. Suppose that we have the following two conditions:*

- (1) $\text{End}_{\bar{K}}(A) = \mathbb{Z}$.
- (2) *There exists a prime $\mathfrak{p} \in \Sigma_K$ at which A has good reduction and such that the splitting field of $\text{ch}_A(\text{Frob}_{\mathfrak{p}})$ has Galois group isomorphic to $(\mathbb{Z}/2\mathbb{Z})^g \rtimes S_g$.*

Then there are constants $c_1, c_2 > 0$ and γ_1, γ_2 , depending only on g and K , for which the following statement is true: For every prime ℓ unramified in K and strictly larger than

$$\max\{c_1(N(\mathfrak{p}))^{\gamma_1}, c_2(h(A))^{\gamma_2}\},$$

the ℓ -adic Galois representation surjects onto $\mathrm{GSp}_{2g}(\mathbb{Z}_\ell)$.

Remark 4.16. The group structure of $(\mathbb{Z}/2\mathbb{Z})^g \rtimes S_g$ is defined by how S_g acts on $(\mathbb{Z}/2\mathbb{Z})^g$, namely by permuting the g factors. This group appears because it is the largest possible Galois group of a reciprocal polynomial, by which we mean a polynomial $P(T)$ satisfying $P(T) = P(1/T) \cdot T^{\deg P}$.

Now, the proof of condition (A3) will follow from Theorem 4.15 once we know that the two hypotheses of Theorem 4.15 hold for a density-1 subset of the K -valued points of the family. We shall first check condition (A3) under the assumption that these hypotheses hold most of the time. To this end, it will be convenient to introduce notation to help us count the points that fail to satisfy one of the hypotheses in Theorem 4.15. For a given family $A \rightarrow U$, define the following two sets:

$$D_1(B) := \{u \in U(K) : \mathrm{Ht}(u) \leq B, A_u \text{ fails hypothesis (1)}\}, \text{ and}$$

$$D_2(B) := \{u \in U(K) : \mathrm{Ht}(u) \leq B, A_u \text{ fails hypothesis (2) for all } \mathfrak{p} \text{ with } N(\mathfrak{p}) \leq (\log B)^{n+1}\}.$$

In the next proposition, we verify condition (A3), conditional upon the assumptions that sets $D_1(B)$ and $D_2(B)$ are sufficiently small (these assumptions are proven in Lemma 4.18 and Proposition 4.21 respectively):

Proposition 4.17. *Let $n > 0$. There are constants c, γ depending only on U such that the following holds: if we define*

$$F(B) := \{u \in U(K) : \mathrm{Ht}(u) \leq B, H_{A_u}(\ell) \supset \mathrm{Sp}_{2g}(\mathbb{Z}/\ell\mathbb{Z}) \text{ for all } \ell > c(\log B)^\gamma\},$$

then we have

$$\frac{|F(B)|}{|\{u \in U(K) : \mathrm{Ht}(u) \leq B\}|} = 1 + O((\log B)^{-n}), \tag{4-8}$$

where the implied constant depends only on U and n .

Proof assuming Lemma 4.18, Proposition 4.21, and Lemma 4.27. Let c_1, c_2 and γ_1, γ_2 be as in Theorem 4.15. There exist constants c'_2, γ'_2 , chosen appropriately in terms of the constants c_0, d_0 provided by Lemma 4.27, such that the following holds: for $u \in U(K)$ with $\mathrm{Ht}(u) > B_0$, where B_0 is a positive constant depending only on U , we have that

$$c_2(h(A_u))^{\gamma_2} \leq c'_2(\log \mathrm{Ht}(u))^{\gamma'_2}.$$

The requirement that $\mathrm{Ht}(u)$ be sufficiently large is insignificant because

$$\frac{|\{u \in U(K) : \mathrm{Ht}(u) \leq B_0\}|}{|\{u \in U(K) : \mathrm{Ht}(u) \leq B\}|} \ll \frac{1}{B^{[K:\mathbb{Q}](r+1)}}, \tag{4-9}$$

and the right-hand side of (4-9) is dominated by the right-hand side of (4-8). If we take

$$c = \max\{c_1, c'_2\} \quad \text{and} \quad \gamma = \max\{(n + 1)\gamma_1, \gamma'_2\},$$

Theorem 4.15 tells us that

$$\{u \in U(K) : \text{Ht}(u) \leq B\} \setminus F(B) \subset D_1(B) \cup D_2(B).$$

The desired result follows from Lemmas 4.18 and 4.21, from which we deduce that

$$\frac{|D_1(B) \cup D_2(B)|}{|\{u \in U(K) : \text{Ht}(u) \leq B\}|} \ll (\log B)^{-n}. \quad \square$$

In what follows, we prove the results upon which the above proof of Proposition 4.17 depends. To begin with, we check that hypotheses (1) and (2) from Theorem 4.15 hold in our setting by bounding D_1 in Lemma 4.18 (thus verifying hypothesis (1)) and bounding D_2 in Proposition 4.21 (thus verifying hypothesis (2)).

4I2. Verifying hypothesis (1). We check that hypothesis (1) holds in our setting via the following:

Lemma 4.18. *We have that*

$$\frac{|D_1(B)|}{|\{u \in U(K) : \text{Ht}(u) \leq B\}|} \ll \frac{\log B}{B^{[K:\mathbb{Q}]/2}}, \quad (4-10)$$

where the implied constant depends only on U .

Proof. Choose $\ell > \max\{C, \ell_1(g)\}$, where C is defined in (4-4) and $\ell_1(g)$ is the constant, depending only on the dimension g , given in [Ellenberg et al. 2009, Proposition 4]. By that proposition we have that $|D_1(B)|$ is bounded above by $|\{u \in U(K) : H_{A_u}(\ell) \supset \text{Sp}_{2g}(\mathbb{Z}/\ell\mathbb{Z})\}|$. The lemma then follows from Proposition 4.4, where we are using point (c) of Section 4C to pass from lattice points to K -valued points. \square

4I3. Verifying hypothesis (2). In Proposition 4.21 we complete the verification of hypothesis (2) by means of an argument involving the large sieve, which lets one bound a set in terms of its reduction modulo primes. The large sieve is stated as follows:

Theorem 4.19 (large sieve, [Zywina 2010a, Theorem 4.1]). *Let $\|\cdot\|$ be a norm on $\mathbb{R} \otimes_{\mathbb{Z}} \mathcal{O}_K^r$, and fix a subset $Y \subset \mathcal{O}_K^r$. Let $B \geq 1$ and $Q > 0$ be real numbers, and for every prime $\mathfrak{p} \in \Sigma_K$, let $0 \leq \omega_{\mathfrak{p}} < 1$ be a real number. Suppose that we have the following two conditions:*

- (a) *The image of Y in $\mathbb{R} \otimes_{\mathbb{Z}} \mathcal{O}_K^r$ is contained in a ball of radius B .*
- (b) *For every $\mathfrak{p} \in \Sigma_K$ with $N(\mathfrak{p}) < Q$, we have $|Y_{\mathfrak{p}}| \leq (1 - \omega_{\mathfrak{p}}) \cdot N(\mathfrak{p})^r$, where $Y_{\mathfrak{p}}$ is the image of Y under reduction modulo \mathfrak{p} .*

Then we have that

$$|Y| \ll \frac{B^{[K:\mathbb{Q}]r} + Q^{2r}}{L(Q)}, \quad \text{where } L(Q) := \sum_{\substack{\mathfrak{a} \subset \mathcal{O}_K \text{ squarefree} \\ N(\mathfrak{a}) \leq Q}} \prod_{\text{prime } \mathfrak{p} | \mathfrak{a}} \frac{\omega_{\mathfrak{p}}}{1 - \omega_{\mathfrak{p}}},$$

and the implied constant depends only on K, r , and $\| - \|$.

We must now specialize the abstract setup in Theorem 4.19 to our setting. To do so, we define the various objects at play in the large sieve as follows:

Definition 4.20. Introduce the following notation:

- Let $\| - \|$ be the norm defined in Section 1B.
- Let $B \geq 1$, take $Q := (\log B)^{n+1}$.
- Let m be the positive integer produced by Proposition 4.22, let ζ_m denote a primitive m -th root of unity, and let $\Sigma_K^m \subset \Sigma_K$ be the set of $\mathfrak{p} \in \Sigma_K$ which split completely in $K(\zeta_m)$. Now, with σ, τ as in Lemma 4.25, we may take $\omega_{\mathfrak{p}} = \sigma$ for all $\mathfrak{p} \in \Sigma_K^m$ with $N(\mathfrak{p}) > \tau$ and $\omega_{\mathfrak{p}} = 0$ for all other $\mathfrak{p} \in \Sigma_K$.
- We take Y to be the following set:

$$Y := \{u \in U(K) \cap \mathcal{O}_K^r : \|u\| \leq B, A_u \text{ fails hypothesis (2) for all } \mathfrak{p} \text{ with } N(\mathfrak{p}) \leq (\log B)^{n+1}\}.$$

As above, $Y_{\mathfrak{p}}$ denotes the mod- \mathfrak{p} reduction of Y .

- Define $T_{\mathfrak{p}}$ by

$$T_{\mathfrak{p}} := \{x \in \mathcal{U}_{\mathbb{F}_{\mathfrak{p}}} : \text{splitting field of } \text{ch}_A(\text{Frob}_{\mathfrak{p}}) \text{ has Galois group } (\mathbb{Z}/2\mathbb{Z})^g \rtimes S_g\}.$$

The motivation for defining $T_{\mathfrak{p}}$ is that its complement contains $Y_{\mathfrak{p}}$.

To ensure that the choices made in Definition 4.20 are suitable, we must prove Proposition 4.22 and Lemma 4.25, which when taken together assert that there exist a positive integer m and $\sigma, \tau > 0$ such that $|Y_{\mathfrak{p}}| \leq (1 - \sigma) \cdot N(\mathfrak{p})^r$ for all $\mathfrak{p} \in \Sigma_K^m$. However, the proof of this result is rather laborious, and stating it now would serve to distract the reader from the primary thrust of the argument. We therefore defer the proof of Lemma 4.25 to Section 4I4, and conditional upon this, we now use the large sieve to check that hypothesis (2) holds in our setting.

Proposition 4.21. For $n > 0$, we have that

$$\frac{|D_2(B)|}{|\{u \in U(K) : \text{Ht}(u) \leq B\}|} \ll (\log B)^{-n}.$$

Proof assuming Proposition 4.22 and Lemma 4.25. Theorem 4.19 yields the estimate

$$|Y| \ll \frac{B^{[K:\mathbb{Q}]r} + (\log B)^{2n(n+1)}}{L((\log B)^{n+1})},$$

whose denominator is bounded below by

$$\begin{aligned} L((\log B)^n) &> \sum_{\substack{\mathfrak{p} \in \Sigma_K^m \\ \tau < N(\mathfrak{p}) < (\log B)^{n+1}}} \frac{\sigma}{1 - \sigma} \\ &> \sigma \cdot |\{\mathfrak{p} \in \Sigma_K^m : \tau < N(\mathfrak{p}) \leq (\log B)^{n+1}\}|. \end{aligned}$$

Applying the Chebotarev Density Theorem yields that

$$|\{\mathfrak{p} \in \Sigma_K^m : \tau < N(\mathfrak{p}) \leq (\log B)^{n+1}\}| \gg |\{\mathfrak{p} \in \Sigma_K : \tau < N(\mathfrak{p}) \leq (\log B)^{n+1}\}|.$$

Applying the Prime Number Theorem yields that

$$|\{\mathfrak{p} \in \Sigma_K : \tau < N(\mathfrak{p}) \leq (\log B)^{n+1}\}| \gg \frac{(\log B)^{n+1}}{\log((\log B)^{n+1})}.$$

Combining the above estimates, we deduce that

$$\begin{aligned} \frac{|Y|}{|\{u \in U(K) \cap \mathcal{O}_K^r : \|u\| \leq B\}|} &\ll \frac{B^{[K:\mathbb{Q}]r} + (\log B)^{2n(n+1)}}{(\log B)^{n+1}/\log((\log B)^{n+1})} \cdot \frac{1}{B^{[K:\mathbb{Q}]r}} \\ &\ll \frac{\log((\log B)^{n+1})}{(\log B)^{n+1}} \ll (\log B)^{-n}. \end{aligned}$$

Finally, employing point (c) of Section 4C to translate the above estimate from lattice points to K -valued points yields the desired result. \square

4I4. Validating the sieve setup. This section is devoted to proving Proposition 4.22 and Lemma 4.25, which together verify that the sieve setup introduced in Definition 4.20 satisfies the necessary conditions for applying the large sieve as we did in the proof of Proposition 4.21. We start by constructing the value of m that we use in our application of the large sieve:

Proposition 4.22. *There is a positive integer m and a subset $\mathcal{C} \subset \mathrm{Sp}_{2g}(\mathbb{Z}/m\mathbb{Z})$ invariant under conjugation in $\mathrm{Sp}_{2g}(\mathbb{Z}/m\mathbb{Z})$, and hence in $\mathrm{GSp}_{2g}(\mathbb{Z}/m\mathbb{Z})$, such that the following holds:*

- (a) *We have $H_A(m) = \mathrm{GSp}_{2g}(\mathbb{Z}/m\mathbb{Z})$ and $H_A^{\mathrm{geom}}(m) = \mathrm{Sp}_{2g}(\mathbb{Z}/m\mathbb{Z})$.*
- (b) *For any $\mathfrak{p} \notin S$ and any closed point $x \in \mathcal{U}_{\mathbb{F}_p}$, if $\rho_{A,m}(\mathrm{Frob}_x) \in \mathcal{C}$, then the splitting field of $\mathrm{ch}(\mathrm{Frob}_x)$ has Galois group $(\mathbb{Z}/2\mathbb{Z})^g \rtimes S_g$.⁶*

Note that it is easy to construct many m satisfying (a) by the big monodromy hypothesis. The main point of this proposition is to show there is an m which also satisfies (b).

Proof. We construct the desired m as a product of four appropriate primes, depending on the family $A \rightarrow U$. By, for example, Hilbert irreducibility, or more precisely [Serre 1997, §9.2, Proposition 1] in conjunction with [Serre 1997, §13.1, Theorem 3] applied to the extension

$$\mathbb{Q}(x_1, \dots, x_g)[T] / \left(T^{2g} + \sum_{i=1}^{g-1} (-1)^i x_i (T^{2g-i} + T^i) + (-1)^g x_g T^g + 1 \right) \quad \text{over } \mathbb{Q}(x_1, \dots, x_g),$$

there exists a degree- $2g$ polynomial $P(T) \in \mathbb{Z}[T]$ satisfying $P(T) = P(1/T) \cdot T^{\deg P}$ with Galois group $(\mathbb{Z}/2\mathbb{Z})^g \rtimes S_g$. It is easy to exhibit elements of $(\mathbb{Z}/2\mathbb{Z})^g \rtimes S_g$ whose left-action on $(\mathbb{Z}/2\mathbb{Z})^g \rtimes S_g$ is described by one of the following four cycle types:

$$2 + 1 + \dots + 1, \quad 4 + 1 + \dots + 1, \quad (2g - 2) + 1 + 1, \quad 2g. \tag{4-11}$$

⁶For the definition of S , see the sentence immediately preceding Remark 4.5.

We choose these cycle types because any subgroup of $(\mathbb{Z}/2\mathbb{Z})^g \rtimes S_g$ containing an element with each of these cycle types is in fact all of $(\mathbb{Z}/2\mathbb{Z})^g \rtimes S_g$ by [Kowalski 2006, Lemma 7.1]. For each such partition, the Chebotarev density theorem tells us that there are infinitely many primes ℓ such that $P(T) \pmod{\ell}$ splits according to the chosen partition. For $\ell > C$ we have $\rho_{A,\ell}(\pi_1(U)) = \mathrm{GSp}_{2g}(\mathbb{Z}/\ell\mathbb{Z})$ and $\rho_{A,\ell}(\pi_1(U_{\bar{K}})) = \mathrm{Sp}_{2g}(\mathbb{Z}/\ell\mathbb{Z})$. So, for $i \in \{1, 2, 3, 4\}$ we can find $\ell_i > C$ such that $P(T) \pmod{\ell_i}$ splits according to the i -th partition above. By the Chinese remainder theorem, (a) holds.

To complete the proof, we construct \mathcal{C} and verify (b). Since characteristic polynomials are conjugacy-invariant, the set

$$\mathcal{C} := \{M' \in \mathrm{GSp}_{2g}(\mathbb{Z}/m\mathbb{Z}) : \mathrm{ch}(M') \pmod{\ell_i} \text{ splits as in (4-11) for all } i \in \{1, 2, 3, 4\}\}$$

is a union of conjugacy classes of $\mathrm{GSp}_{2g}(\mathbb{Z}/m\mathbb{Z})$. By [Rivin 2008, Theorem A.1] there exists an $M \in \mathrm{Sp}_{2g}(\mathbb{Z})$ such that $\mathrm{ch}(M)(T) = P(T)$, which shows that \mathcal{C} is nonempty. For this choice of \mathcal{C} , conclusion (b) follows from [Kowalski 2006, Lemma 7.1], which says that any subgroup of $(\mathbb{Z}/2\mathbb{Z})^g \rtimes S_g$ that contains elements realizing all four cycle types in (4-11) must actually equal all of $(\mathbb{Z}/2\mathbb{Z})^g \rtimes S_g$. \square

The reason why we constructed m in Proposition 4.22 in the way that we did is that it allows us to apply the following theorem, which is a crucial tool for bounding the set of Frobenius elements with certain Galois groups modulo each prime.

Theorem 4.23 [Ekedahl 1990, Lemma 1.2]. *Let X be a scheme, and let $\pi : X \rightarrow \mathrm{Spec} \mathcal{O}_K$ be a morphism of finite type. Let $\phi : Y \rightarrow X$ be a connected finite Galois étale cover with Galois group G , and let $\rho : \pi_1(X) \rightarrow G$ denote the corresponding finite quotient. Suppose that $\pi \circ \phi$ has a geometrically irreducible generic fiber, and let \mathcal{C} be a conjugacy-invariant subset of G . For every $\mathfrak{p} \in \Sigma_K$, we have*

$$\frac{|\{x \in X(\mathbb{F}_{\mathfrak{p}}) : \rho(\mathrm{Frob}_x) \in \mathcal{C}\}|}{|X(\mathbb{F}_{\mathfrak{p}})|} = \frac{|\mathcal{C}|}{|G|} + O((N(\mathfrak{p}))^{-\frac{1}{2}}),$$

with implicit constants depending only on the family $Y \rightarrow X$. By Frob_x we mean the Frobenius element in $\pi_1(X)$ corresponding to $x \in X$.

We now apply Theorem 4.23 to the conjugacy-invariant set \mathcal{C} from Proposition 4.22 in order to obtain a lower bound on $|T_{\mathfrak{p}}|$, the number of points $u \in U(K)$ with the splitting field of $\mathrm{ch}_{A_u}(\mathrm{Frob}_{\mathfrak{p}})$ having Galois group equal to $(\mathbb{Z}/2\mathbb{Z})^g \rtimes S_g$.

Proposition 4.24. *As \mathfrak{p} ranges through the elements of Σ_K^m , where m is defined as in Proposition 4.22, we have that $|T_{\mathfrak{p}}| \gg (N(\mathfrak{p}))^f$.*

Proof. Let $L := K(\zeta_m)$. As in Section 3B, let $\mathcal{V}_m \rightarrow \mathcal{U}_{\mathcal{O}_{P_m}}$ be the connected Galois étale cover associated to the mod- m Galois representation $\rho : \pi_1(\mathcal{U}_{\mathcal{O}_{P_m}}) \rightarrow \mathrm{GSp}_{2g}(\mathbb{Z}/m\mathbb{Z})$, and let \mathcal{X} be one of the connected components of $(\mathcal{V}_m)_L$. The map $\mathcal{X} \rightarrow (\mathcal{U}_{\mathcal{O}_{P_m}})_L$ is the connected Galois étale cover associated to the map

$$\rho' : \pi_1((\mathcal{U}_{\mathcal{O}_{P_m}})_L) \longrightarrow \pi_1(\mathcal{U}_{\mathcal{O}_{P_m}}) \xrightarrow{\rho} \mathrm{GSp}_{2g}(\mathbb{Z}/m\mathbb{Z});$$

note that the image of this composite map equals $\rho(\pi_1(\mathcal{U}_{\mathcal{O}_{p_m}})) \cap \mathrm{Sp}_{2g}(\mathbb{Z}/m\mathbb{Z})$ by Remark 3.2, since χ_m is trivial on $K(\zeta_m)$. By Proposition 4.22(a), we have $\rho(\pi_1(\mathcal{U}_{\mathcal{O}_{p_m}})) = \mathrm{GSp}_{2g}(\mathbb{Z}/m\mathbb{Z})$, so we conclude that $\rho'(\pi_1((\mathcal{U}_{\mathcal{O}_{p_m}})_L)) = \mathrm{Sp}_{2g}(\mathbb{Z}/m\mathbb{Z})$.

We seek to apply Theorem 4.23 with

$$\mathcal{X} \rightarrow (\mathcal{U}_{\mathcal{O}_{p_m}})_L \rightarrow \mathrm{Spec} \mathcal{O}_L \quad \text{in place of} \quad Y \rightarrow X \rightarrow \mathrm{Spec} \mathcal{O}_K.$$

To do so, we must check that this composition has geometrically irreducible generic fiber, which follows from the second part of Proposition 4.22(a) in conjunction with Lemma 4.9.

Now let $\mathcal{C} \subset \mathrm{Sp}_{2g}(\mathbb{Z}/m\mathbb{Z})$ be as in Proposition 4.22(b). For any $\mathfrak{p} \in \Sigma_K^m \setminus S$ and $\mathfrak{p}' \in \Sigma_L$ lying over \mathfrak{p} , we have $(\mathcal{U}_L)_{\mathbb{F}_{\mathfrak{p}'}} \simeq \mathcal{U}_{\mathbb{F}_{\mathfrak{p}}}$, and so there is a bijection between

$$\{x \in \mathcal{U}_L(\mathbb{F}_{\mathfrak{p}'}) : \rho'(\mathrm{Frob}_x) \in \mathcal{C}\} \quad \text{and} \quad \{x \in \mathcal{U}(\mathbb{F}_{\mathfrak{p}}) : \rho(\mathrm{Frob}_x) \in \mathcal{C}\}.$$

By Proposition 4.22(b), $T_{\mathfrak{p}}$ contains the latter set, so we have

$$\begin{aligned} |T_{\mathfrak{p}}| &\geq |\{x \in \mathcal{U}(\mathbb{F}_{\mathfrak{p}}) : \rho(\mathrm{Frob}_x) \in \mathcal{C}\}| = |\{x \in \mathcal{U}_L(\mathbb{F}_{\mathfrak{p}'}) : \rho'(\mathrm{Frob}_x) \in \mathcal{C}\}| \\ &= \left(\frac{|\mathcal{C}|}{|G|} + O((N(\mathfrak{p}'))^{-\frac{1}{2}}) \right) \cdot |\mathcal{U}_L(\mathbb{F}_{\mathfrak{p}'})|, \end{aligned}$$

where the last step above follows from Theorem 4.23. Now, we have the estimate

$$|\mathcal{U}_L(\mathbb{F}_{\mathfrak{p}'})| \gg (N(\mathfrak{p}'))^r,$$

because the complement of $(\mathcal{U}_L)_{\mathbb{F}_{\mathfrak{p}'}}$ in $(\mathbb{P}^r_{\mathcal{O}_L})_{\mathbb{F}_{\mathfrak{p}'}}$ has codimension at least 1, since $\mathfrak{p} \notin S$. Combining our results, and using that S is a finite set, we find that

$$|T_{\mathfrak{p}}| \geq \left(\frac{|\mathcal{C}|}{|G|} + O(N(\mathfrak{p}')^{-\frac{1}{2}}) \right) \cdot |\mathcal{U}_L(\mathbb{F}_{\mathfrak{p}'})| \gg N(\mathfrak{p}')^r = N(\mathfrak{p})^r. \quad \square$$

The following lemma completes our verification of the sieve setup by constructing the necessary constants σ, τ .

Lemma 4.25. *There are constants $\sigma, \tau > 0$ such that for all $\mathfrak{p} \in \Sigma_K^m$ with $N(\mathfrak{p}) > \tau$, we have $|Y_{\mathfrak{p}}| \leq (1 - \sigma) \cdot N(\mathfrak{p})^r$.*

Proof. By Proposition 4.24, there are constants $\sigma', \tau' > 0$ such that, for all $\mathfrak{p} \in \Sigma_K^m$ with $N(\mathfrak{p}) > \tau'$, we have $|T_{\mathfrak{p}}| \geq \sigma' \cdot (N(\mathfrak{p}))^r$. For such \mathfrak{p} , we have that

$$|Y_{\mathfrak{p}}| \leq (1 - \sigma') \cdot (N(\mathfrak{p}))^r + O((N(\mathfrak{p}))^{r-1}),$$

where the error term is on order of $N(\mathfrak{p})$ smaller than the main term because \mathcal{Z} has codimension at least 1 in $\mathbb{P}^r_{\mathcal{O}_K}$. By replacing σ' with a slightly smaller σ and τ' with a slightly larger τ , we may write

$$|Y_{\mathfrak{p}}| \leq (1 - \sigma) \cdot (N(\mathfrak{p}))^r. \quad \square$$

4I5. Discussion of heights. In this section, we prove a result that describes the relationship between the absolute multiplicative height on projective space and the absolute logarithmic Faltings height. Let Ht be the height on \mathbb{P}^r_K as defined in 1B, and let h be the Faltings height. Let $\log Ht$ be the absolute logarithmic height on $\mathbb{P}^r(\bar{K})$, and note that $\log Ht$ naturally restricts to a logarithmic height function defined on the open subscheme $U \subset \mathbb{P}^r_K$.

Let \mathcal{A}_g be the moduli stack of g -dimensional PPAVs, and let $p : \mathcal{U}_g \rightarrow \mathcal{A}_g$ be the universal family of abelian varieties. Let $\pi : \mathcal{A}_g \rightarrow A_g$ be its coarse moduli space, and let $j(A) \in A_g(K)$ be the closed point represented by A . As in [Faltings 1983, Section 2], we choose $n \in \mathbb{N}$ such that the line bundle $\mathcal{L} = ((\pi \circ p)_* \omega_{\mathcal{U}_g/\mathcal{A}_g})^{\otimes n}$ is very ample, where $\omega_{\mathcal{U}_g/\mathcal{A}_g}$ is the canonical sheaf of $p : \mathcal{U}_g \rightarrow \mathcal{A}_g$. Fix an embedding $i : A_g \hookrightarrow \mathbb{P}^N$ with $i^* \mathcal{O}_{\mathbb{P}^N}(1) \simeq \mathcal{L}$. The modular height $\log Ht(j(A))$ of A is then the restriction along i of the absolute logarithmic height (i.e., the absolute logarithmic height of $j(A)$ considered as a point of $\mathbb{P}^N(K)$). On the other hand, $\mathcal{O}_{\mathbb{P}^N}(1)$ is a metrized line bundle and restricts to give a metric on \mathcal{L} [Faltings et al. 1992, p. 36]; we denote by $\log Ht_{\mathcal{L}}$ the corresponding height function on A_g .

We now relate the height on projective space and the Faltings height by piecing together results from the literature on heights:

Lemma 4.26. *Let g be a positive integer, K a number field, and let $n \in \mathbb{N}$ be as in the definition of the modular height. Then there exist constants α and β such that for every principally polarized abelian variety A over K , we have*

$$|n \cdot h(A) - \log Ht(j(A))| \leq \alpha \cdot \log \max\{1, \log Ht(j(A))\} + \beta.$$

Proof. By [Faltings 1983, Proof of Lemma 3], there exist constants α_1 and β_1 such that for all abelian varieties A/K , we have

$$|n \cdot h(A) - \log Ht_{\mathcal{L}}(j(A))| \leq \alpha_1 \cdot \log(\log Ht_{\mathcal{L}}(j(A))) + \beta_1.$$

By [Hindry and Silverman 2000, B.3.2(b)], there is a constant β_2 such that

$$|\log Ht_{\mathcal{L}}(j(A)) - \log Ht(j(A))| \leq \beta_2. \quad \square$$

Lemma 4.27. *There exist constants c_0 and d_0 depending only on $A \rightarrow U$ such that*

$$h(A_u) \leq c_0 \log Ht(u) + d_0$$

for all $u \in U(K)$.

Proof. By [Serre 1997, p. 19, Section 2.6, Theorem], $Ht(j(A_u)) \ll Ht(u)$ and $Ht(u) \ll Ht(j(A_u))$ for all $u \in U$. The result then follows from Lemma 4.26. □

5. Applications of Theorem 1.1

The purpose of this section is to demonstrate that the main result, Theorem 1.1, can be applied to a number of interesting families of PPAVs, such as families containing a dense open substack of the locus

of Jacobians of hyperelliptic curves, trigonal curves, or plane curves. In Section 5A, we prove a general tool that is needed to guarantee big monodromy for the loci in our applications, and in Section 5B, we examine each of these applications in detail.

5A. Finite-index criterion. In this section we prove Proposition 5.2, which will be applied in the setting of Theorem 1.1 to determine that U has big monodromy when its image in the moduli stack of abelian varieties has big monodromy. We begin by recalling an elementary criterion giving surjectivity for the map on étale fundamental groups induced by a morphism of Deligne–Mumford stacks. By *Deligne–Mumford stack*, we mean a stack in the étale topology with representable diagonal (i.e., representable by algebraic spaces), which has an étale surjective morphism from a scheme. For a general reference on stacks, see [Olsson 2016] or [Laumon and Moret-Bailly 2000]; also, see [Stacks 2005–] for a more comprehensive reference.

Lemma 5.1. *Suppose $f : X \rightarrow Y$ is a map of Deligne–Mumford stacks. The fiber product $U \times_Y X$ is connected for all finite connected étale maps $U \rightarrow Y$ if and only if the induced map $\pi_1(X) \rightarrow \pi_1(Y)$ is surjective. In particular, if X and Y are normal, integral, and Noetherian, and $f : X \rightarrow Y$ is a flat map with connected geometric generic fiber, then the induced map $\pi_1(X) \rightarrow \pi_1(Y)$ is surjective.*

Proof. The first part holds in greater generality as a statement about Galois categories; see [Stacks 2005–, Tag 0BN6]. As for the second part, we only need verify that a connected finite étale cover $U \rightarrow Y$ pulls back to a connected cover of X . Note that because X and Y are normal and integral, étale covers of X and Y are connected if and only if they are irreducible. Here, we are using that normal and connected implies irreducible and that normality is local in the étale topology over Noetherian stacks. To see why normality is local in the étale topology over a Deligne–Mumford stack, note first that normality is local in the étale topology over any base scheme by [Stacks 2005–, Tag 03E7]. Using this, one defines a Deligne–Mumford stack to be normal if any étale cover by a scheme is normal. From this definition, it follows that normality of a Deligne–Mumford stack is equivalent to normality of any étale cover.

Thus, we only need show that if $U \rightarrow Y$ is any irreducible finite étale cover, then so is $X \times_Y U \rightarrow X$. But this follows from the assumptions that f is flat and U is integral, which implies all generic points of $X \times_Y U$ map to the generic point of U . So, if $X \times_Y U$ were reducible, the geometric generic fiber over U would also be reducible, which contradicts the assumption that f has connected geometric generic fiber, since a geometric generic fiber of $X \times_Y U$ is also a geometric generic fiber of f . \square

Proposition 5.2. *Let k be an arbitrary field of characteristic 0. Suppose X is a scheme and Y is a Deligne–Mumford stack over k , both of which are normal, integral, separated, and finite type over k , and let $f : X \rightarrow Y$ be a dominant map. Then, the image of the induced map $\pi_1(X) \rightarrow \pi_1(Y)$ has finite index in $\pi_1(Y)$. If, in addition, the geometric generic fiber of f is connected, then the map $\pi_1(X) \rightarrow \pi_1(Y)$ is surjective.*

Proof. To begin, we reduce to the case in which f is smooth. By generic smoothness, we may replace X by a dense open $X' \subset X$ so that $f|_{X'}$ is smooth. Since, $\pi_1(X') \rightarrow \pi_1(X)$ is a surjection by Lemma 5.1, in order to prove the proposition, we may replace X by X' .

The last sentence of this proposition follows from Lemma 5.1 (here we only needed that the map be f be flat, but we have already reduced to the case it is smooth). To conclude, we only need prove that the image of $\pi_1(X) \rightarrow \pi_1(Y)$ has finite index in $\pi_1(Y)$, without the assumption that the geometric generic fiber of f is connected. Since f is smooth and Y is Deligne–Mumford, we can find a scheme U and a dominant étale map $U \rightarrow X$ such that $U \rightarrow Y$ factors through \mathbb{A}_Y^N , where N is the dimension of the geometric generic fiber of f and $U \rightarrow \mathbb{A}_Y^N$ étale. So, after passing to a dense open substack $W \subset \mathbb{A}_Y^N$ and a dense open subscheme $U' \subset U$, we may assume that $U' \rightarrow W$ is a finite étale cover: To see why, take a smooth cover of \mathbb{A}_Y^N by a scheme. The pullback to U is a separated algebraic space, so it has a dense open subspace that is a scheme by [Olsson 2016, Theorem 6.4.1]. The finiteness claim then follows because the resulting étale morphism of schemes is locally quasifinite, of finite type, and quasiseparated, hence generically finite on the target. Since $U' \rightarrow W$ is finite étale, $\pi_1(U') \rightarrow \pi_1(W)$ has finite index. Because the maps $\pi_1(W) \rightarrow \pi_1(\mathbb{A}_Y^N)$ and $\pi_1(\mathbb{A}_Y^N) \rightarrow \pi_1(Y)$ are surjective by Lemma 5.1, the composition $\pi_1(U') \rightarrow \pi_1(Y)$ has finite index in $\pi_1(Y)$, and hence so does $\pi_1(X) \rightarrow \pi_1(Y)$. \square

5B. Applications. Let K be a number field with fixed algebraic closure \bar{K} , let \mathcal{M}_g denote the moduli stack of curves of genus g over K , and let \mathcal{A}_g denote the moduli stack of PPAVs of dimension g over K . We have a natural map $\tau_g : \mathcal{M}_g \rightarrow \mathcal{A}_g$ given by the Torelli map, which sends a curve to its Jacobian. Let \mathcal{U}_g denote the universal family over \mathcal{A}_g . Note that if U is any scheme and $A \rightarrow U$ is a family of PPAVs, then there exist maps $A \rightarrow \mathcal{U}_g$ and $U \rightarrow \mathcal{A}_g$ such that A equals the fiber product $U \times_{\mathcal{A}_g} \mathcal{U}_g$.

We will also be interested in the locus of smooth hyperelliptic curves of genus g , $\mathcal{H}_g \subset \mathcal{M}_g$, and the locus of trigonal curves of genus g , $\mathcal{T}^g \subset \mathcal{M}_g$. If a curve C is trigonal, there exists a unique nonnegative integer M , called the Maroni invariant, with the property that there is a canonical embedding into the Hirzebruch surface $\mathbb{F}_M := \mathbb{P}_{\mathbb{P}^1}(\mathcal{O}_{\mathbb{P}^1} \oplus \mathcal{O}_{\mathbb{P}^1}(M))$. As mentioned in [Patel and Vakil 2015], the Maroni invariant takes on all integer values between 0 and $(g+2)/3$ with the same parity as g . Let $\mathcal{T}^g(M) \subset \mathcal{M}_g$ denote the substack of trigonal curves of Maroni invariant M .

In order to more easily utilize Proposition 5.2 for the purpose of giving interesting examples of Theorem 1.1, we record the following easy consequence of Proposition 5.2:

Corollary 5.3. *Let $U \subset \mathbb{P}_K^r$ be an open subscheme, and let $A \rightarrow U$ be a family of g -dimensional PPAVs. Let $\phi : U \rightarrow \mathcal{A}_g$ be the map induced by the universal property of \mathcal{A}_g . Let V be the smallest locally closed substack of \mathcal{A}_g through which U factors, and let $W \subset \mathcal{A}_g$ be a normal integral substack. Suppose further that $W \cap V$ is dense in W and that V is normal. Then, if W has big monodromy, so do V and U . Furthermore, if the geometric generic fiber of ϕ is irreducible, then the monodromy of V agrees with that of U . In particular, the conclusion of Theorem 1.1 holds for U .*

Proof. By Lemma 5.1, if W has big monodromy so does the dense open subset $W \cap V \subset W$. Therefore, V has big monodromy, because it contains $W \cap V$, which has big monodromy. The result then follows from Proposition 5.2, once we verify that both U and V are normal, irreducible, separated, and finite type over K , with V Deligne–Mumford. All of these conditions are immediate except possibly that V is generically smooth, which holds by generic smoothness on a smooth cover of V by a scheme. \square

Before stating the main theorem of this section, we pause to describe more precisely what we mean by “the locus of plane curves.”

Remark 5.4. In Theorem 1.6(c) and Theorem 5.5(d), we refer to the “substack of Jacobians of plane curves of degree d ,” for $d \geq 3$, and we now make more precise what we mean by this locus. When $d = 3$, all 1-dimensional abelian varieties can be realized as the Jacobian of a degree-3 plane curve, so in this case we take the locus to be all of $\mathcal{M}_{1,1}$. For $d \geq 4$, we will define a locally closed substack of \mathcal{M}_g , where $g = \binom{d-1}{2}$, and the locus of Jacobians of plane curves of degree d will denote the image of this under the Torelli map. For $d \geq 4$, let $\pi_d : \mathcal{V}_d \rightarrow \mathbb{P}^{\binom{d+2}{2}-1}$ denote the universal family over the Hilbert scheme of plane curves of degree d , and let $U_d \subset \mathbb{P}^{\binom{d+2}{2}-1}$ denote the dense open subscheme over which π_d is smooth. Since $\mathcal{V}_d|_{U_d} \subset U_d \times \mathbb{P}^2$, the action of PGL_3 on \mathbb{P}^2 induces an action on $\mathcal{V}_d|_{U_d}$ and hence on U_d . Then, we define the substack of Jacobians of plane curves of degree d to be the stack theoretic quotient $[U_d/\mathrm{PGL}_3]$.

Note that there is a natural map $[U_d/\mathrm{PGL}_3] \rightarrow \mathcal{M}_g$. It can be verified that this map is a locally closed immersion of stacks. Further, one can show $[U_d/\mathrm{PGL}_3]$ represents the functor associating to any base scheme T projective flat morphisms $f : C \rightarrow T$ where each geometric fiber is a proper smooth curve of genus $g := \binom{d-1}{2}$ with a degree d invertible sheaf on C which commutes with base change. In this sense, $[U_d/\mathrm{PGL}_3]$ may naturally be referred to as “the locus of plane curves of degree d ” and it is evidently smooth, since U_d is smooth, being a dense open subscheme of projective space.

Let us now briefly sketch the proof of the two facts claimed above. First, one can first see that $[U_d/\mathrm{PGL}_3]$ represents the claimed functor by defining natural maps both ways and verifying they are mutually inverse. To show $[U_d/\mathrm{PGL}_3] \rightarrow \mathcal{M}_g$ is a locally closed immersion, one can factor $[U_d/\mathrm{PGL}_3] \rightarrow \mathcal{M}_g$ through the stack G_d^2 parametrizing the g_d^2 on the universal curve over \mathcal{M}_g , via a natural generalization of the definition given in [Arbarello et al. 2011, Chapter XXI, Definition 3.12]. One can check the map $[U_d/\mathrm{PGL}_3] \rightarrow G_d^2$ is an open immersion from the definitions. Finally, one can verify that the map $G_d^2 \rightarrow \mathcal{M}_g$ is a locally closed immersion, using that every smooth plane curve of degree at least 4 has a unique g_d^2 , see [Arbarello et al. 1985, Appendix A, Exercises 17 and 18], and the valuative criterion for locally closed immersions [Mochizuki 1999, Chapter 1, Corollary 2.13].

We are now in position to state and prove the main theorem of this section:

Theorem 5.5. *Suppose $A \rightarrow U$ is a rational family of principally polarized abelian varieties and define V to be the smallest locally closed substack of \mathcal{A}_g through which U factors. The conclusion of Theorem 1.1 holds whenever V is normal and contains a dense open substack of one of the following loci:*

- (a) *The locus $\tau_g(\mathcal{H}_g)$ for any $g \geq 1$. For every $g \geq 1$, there exists a U dominating $\tau_g(\mathcal{H}_g)$ because \mathcal{H}_g is unirational.*
- (b) *The locus $\tau_g(\mathcal{F}^g(M))$ of Jacobians of trigonal curves with Maroni invariant $M < \frac{g}{3} - 1$ for any $g \geq 5$. In this case, there exists U dominating $\tau_g(\mathcal{F}^g(M))$ because $\mathcal{F}^g(M)$ is unirational.*
- (c) *The locus of trigonal curves \mathcal{F}^g in any $g \geq 3$. We can take U to be any open subscheme of \mathcal{F}^g , as \mathcal{F}^g is rational.*

- (d) *The locus of Jacobians of degree- d plane curves for any $d \geq 3$. In this case, the open subscheme of the Hilbert scheme of degree- d plane curves parametrizing smooth curves is rational and dominates the locus of Jacobians of degree- d plane curves.*
- (e) *The locus $\tau_g(\mathcal{M}_g)$ for any $g \geq 1$. In this case, when $1 \leq g \leq 14$, \mathcal{M}_g is unirational, so there exists a U dominating \mathcal{M}_g . Moreover, when $3 \leq g \leq 6$, \mathcal{M}_g is rational, and so we may take U to be any open subscheme of \mathcal{M}_g .*
- (f) *The locus \mathcal{A}_g for any $g \geq 1$. When $1 \leq g \leq 5$, \mathcal{A}_g is unirational, so such a U exists.*

Proof. By Corollary 5.3, it suffices to check that each of the families enumerated above has a dense open substack which has big monodromy, is irreducible, and is normal, and to verify the rationality and unirationality claims made above. Irreducibility of these loci is well-known. Note that in the first five cases, if we denote the locus in question by $\tau_g(W) \subset \mathcal{A}_g$, it suffices to verify that $W \subset \mathcal{M}_g$ is smooth as a substack of \mathcal{M}_g , as we now explain. First, $\tau_g(W) \subset \mathcal{A}_g$ is generically smooth because it is reduced, since it is the image of W , which is reduced. Taking a smooth dense open $Z' \subset \tau_g(W)$, we have that $\tau_g^{-1}(Z') \subset W$ is a dense open substack, hence it is also smooth and has big monodromy. This implies Z' also has big monodromy since the monodromy of a locus in \mathcal{M}_g agrees with the monodromy of its image in \mathcal{A}_g under τ_g , as both can be identified with the monodromy action on the first cohomology group. We now conclude the proof by verifying that each locus in \mathcal{M}_g (in the first five cases) is normal, has big monodromy, and is rational or unirational when claimed. In fact, we just show the substack has big geometric monodromy, since this implies it has big monodromy by Proposition 4.1.

- (a) The hyperelliptic locus, \mathcal{H}_g , has big geometric monodromy as was shown independently in [Mumford 2007, Lemma 8.12; A'Campo 1979, théorème 1]. The hyperelliptic locus \mathcal{H}_g is smooth and unirational because it is the quotient of an open subscheme of \mathbb{P}_K^{2g+2} by the smooth action of PGL_2 .
- (b) By [Bolognesi and Lönne 2016, Theorem, p. 2], $\mathcal{T}^g(M)$ has big geometric monodromy when $M < \frac{g}{3} - 1$. Additionally, $\mathcal{T}^g(M)$ is smooth and unirational because it can be expressed as a quotient $[U/G]$ of a smooth rational scheme U by a smooth group scheme G . Here, G is the group of automorphisms of the Hirzebruch surface \mathbb{F}_M and U is an open subscheme of the projectivization of the linear system of class $3e + ((g + 3M + 2)/2)f$ on \mathbb{F}_M , where f is the class of the fiber over \mathbb{P}^1 and e is the unique section with negative self-intersection (see [Bolognesi and Lönne 2016, p. 8] for an explanation of this description of U). Note that in this application, we are implicitly translating between the topological monodromy representation of \mathcal{M}_g described in [Bolognesi and Lönne 2016, Theorem, p. 2] and the algebraic Galois representation in \mathcal{A}_g , but these two representations are compatible, essentially because both are given by the action of the fundamental group on the first cohomology group.
- (c) In the case that $g \geq 5$, we have $\mathcal{T}^g(g \bmod 2)$ is birational to \mathcal{T}^g , so \mathcal{T}^g has a smooth dense open with big geometric monodromy by the previous part. Next, \mathcal{T}^g is rational for $g \geq 5$ by [Ma 2015, Theorem, p. 1]. The cases $g = 3, 4$ hold because for such g , \mathcal{T}^g forms a dense open in \mathcal{M}_g , which is itself rational and smooth, as shown in the proof of part (e) below.

(d) By Remark 5.4, the locus of plane curves (as was also defined in Remark 5.4) in \mathcal{M}_g is smooth. By [Beauville 1986, théorème 4], the locus of smooth degree- d plane curves in the Hilbert scheme has big geometric monodromy. It follows from Lemma 5.1 that the locus of plane curves has big monodromy. The locus of *smooth* degree- d plane curves in the Hilbert scheme is certainly rational, as it is an open subscheme of the Hilbert scheme of degree- d plane curves, which is itself isomorphic to $\mathbb{P}_K^{\binom{d+2}{2}-1}$.

(e) By [Deligne and Mumford 1969, (5.12)], the geometric monodromy of \mathcal{M}_g is all of $\mathrm{Sp}_{2g}(\widehat{\mathbb{Z}})$ for every $g \geq 1$. (Alternatively, the fact that \mathcal{M}_g has big geometric monodromy follows immediately from the corresponding fact for any one of parts (a)–(d).) Next, \mathcal{M}_g is smooth by [Deligne and Mumford 1969, Theorem (5.2)]. We have that \mathcal{M}_g is unirational for $1 \leq g \leq 14$ by [Verra 2005]. Moreover, when $3 \leq g \leq 6$, we have that \mathcal{M}_g is rational; see [Casnati and Fontanari 2007, p. 2] for comprehensive references.

(f) Note that \mathcal{A}_g has geometric big monodromy because \mathcal{A}_g contains \mathcal{M}_g and \mathcal{M}_g has monodromy $\mathrm{Sp}_{2g}(\widehat{\mathbb{Z}})$, as argued in point (d). Further, \mathcal{A}_g is smooth by [Oort 1971, Theorem 2.4.1]. We have that \mathcal{A}_g is unirational for $1 \leq g \leq 5$ as shown in [Verra 2005, p. 1]. \square

Remark 5.6. In most of the cases enumerated in Theorem 5.5, we actually know that the geometric monodromy is not only big, but also equal to $\mathrm{Sp}_{2g}(\widehat{\mathbb{Z}})$. By Corollary 5.3, this occurs when U has irreducible geometric generic fiber over any of the following loci:

- (a) the locus $\mathcal{T}^g(M)$ for any $M < \frac{g}{3} - 1$, by [Bolognesi and Lönne 2016, Theorem, p. 2];
- (b) the locus of plane curves of degree d with d even, by [Beauville 1986, théorème 4(i)];
- (c) the locus \mathcal{M}_g for any g , by [Deligne and Mumford 1969, (5.12)];
- (d) the locus \mathcal{A}_g for any g , because $\mathcal{M}_g \subset \mathcal{A}_g$ and \mathcal{M}_g has full monodromy by point (d).

Remark 5.7. If $A \rightarrow U$ is a family with $H_A^{\mathrm{geom}} = \mathrm{Sp}_{2g}(\widehat{\mathbb{Z}})$, then the group H_A can be determined as follows. The intersection $K \cap \mathbb{Q}^{\mathrm{cyc}}$ is of the form $\mathbb{Q}(\zeta_n)$ for some $n \geq 2$. Let $r_n : \widehat{\mathbb{Z}} \rightarrow \mathbb{Z}/n\mathbb{Z}$ be the reduction map. Then

$$H_A = \ker(r_n \circ \mathrm{mult}) = \{M \in \mathrm{Sp}_{2g}(\widehat{\mathbb{Z}}) : \mathrm{mult} M \equiv 1 \pmod{n}\},$$

which follows from Remark 3.2. Thus, when the conclusion of the preceding remark holds, Theorem 1.1 tells us the following:

- If $K \neq \mathbb{Q}$, or if $K = \mathbb{Q}$ and $g \geq 3$, then most $u \in U(K)$ have $H_{A_u} = \ker(r_n \circ \mathrm{mult})$.
- If $K = \mathbb{Q}$ and $g \in \{1, 2\}$, then most $u \in U(K)$ are such that $[\mathrm{GSp}_{2g}(\widehat{\mathbb{Z}}) : H_{A_u}] = 2$.

Remark 5.8. Theorem 5.5(a) tells us that if U dominates \mathcal{H}_g , then the conclusion of Theorem 1.1 holds for U . In the case where U has irreducible geometric generic fiber, we can say explicitly what the monodromy group of the family is and what its commutator is. For example, let $\mathcal{B}_{2g+2, K}$ denote the family of genus- g hyperelliptic curves over K with Weierstrass equation given by

$$y^2 = x^{2g+2} + a_{2g+1}x^{2g+1} + \dots + a_0.$$

We show in [Landesman et al. 2017a, Theorem 1.2] that most members of $\mathcal{A}_{2g+2,K}$ have monodromy equal to $H_{\mathcal{A}_{2g+2,K}}$ (which we explicitly compute) over $K \neq \mathbb{Q}$, and have index-2 monodromy when $K = \mathbb{Q}$. We neither prove nor state this result precisely here, but a complete statement and proof is given in [Landesman et al. 2017a].

Appendix: Explicit surjectivity for abelian surfaces

By Davide Lombardo

Let K be a number field and A/K be an abelian surface such that $\text{End}_{\bar{K}}(A) = \mathbb{Z}$. For every place w of K at which A has good reduction, let Frob_w be the corresponding Frobenius element of $\text{Gal}(\bar{K}/K)$ and let $f_w(x)$ be the characteristic polynomial of Frob_w acting on $T_\ell A$, where ℓ is any prime different from the residual characteristic of w (as is well known, this definition is well-posed). Let $F(w)$ be the splitting field over \mathbb{Q} of $f_w(x)$. By Remark 4.16, the Galois group of $F(w)/\mathbb{Q}$ is isomorphic to a subgroup of $(\mathbb{Z}/2\mathbb{Z})^2 \rtimes S_2 \simeq D_4$, the dihedral group on 4 points.

To state our result we need the following function:

Definition A.1. Let $\alpha(g) = 2^{10}g^3$ and set $b(d, g, h) = ((14g)^{64g^2} d(\max\{h, \log d, 1\})^2)^{\alpha(g)}$.

We shall show the following result, which extends [Lombardo 2016a, Theorem 1.2] to the case of abelian surfaces:

Proposition A.2. *Let v be a place of K , of good reduction for A , such that the Galois group of $f_v(x)$ is isomorphic to D_4 . Let q_v be the order of the residue field at v . For all primes ℓ , let*

$$\rho_{\ell^\infty} : \text{Gal}(\bar{K}/K) \rightarrow \text{Aut}(T_\ell A) \cong \text{GL}_4(\mathbb{Z}_\ell)$$

be the natural ℓ -adic Galois representation attached to A/K . We have $\text{Im } \rho_{\ell^\infty} = \text{GSp}_4(\mathbb{Z}_\ell)$ for all primes ℓ that are unramified in K and strictly larger than

$$\max\{b(2[K : \mathbb{Q}], 4, 2h(A))^{\frac{1}{4}}, (2q_v)^8\}.$$

From now on, let v be a place as in the statement of Proposition A.2. Notice that $f_v(x)$ is irreducible by assumption, hence all its roots are simple. Moreover, $f_v(x)$ doesn't have any real roots, because (by the Weil conjectures) every root of $f_v(x)$ has absolute value $\sqrt{q_v}$, hence its only possible real roots are $\pm\sqrt{q_v}$. But these are algebraic numbers of degree at most 2 over \mathbb{Q} , while $f_v(x)$ is irreducible of degree 4, contradiction. In particular, the roots of $f_v(x)$ come in complex conjugate pairs, so we shall denote them by $\mu_1, \mu_2, \iota(\mu_1), \iota(\mu_2)$, where $\iota : \mathbb{C} \rightarrow \mathbb{C}$ is complex conjugation. We shall need the following lemma:

Lemma A.3. *Let x, y, z be three distinct eigenvalues of Frob_v . We have $y^2 \neq xz$.*

Proof. Suppose first that $z = \iota(x)$. Then $y^2 = x\iota(x) = q_v$, which implies that $y = \pm\sqrt{q_v}$ is a root of $f_v(x)$. As we have already seen, this is a contradiction. Hence, up to renaming the eigenvalues of Frob_v if necessary, we can assume $x = \mu_1, z = \mu_2$ and $y = \iota(\mu_1)$. Since $\text{Gal}(F(v)/\mathbb{Q})$ is isomorphic to D_4 by assumption, there is a $\sigma \in \text{Gal}(F(v)/\mathbb{Q})$ such that $\sigma(\mu_1) = \mu_1, \sigma(\iota(\mu_1)) = \iota(\mu_1), \sigma(\mu_2) = \iota(\mu_2)$ and

$\sigma(\iota(\mu_2)) = \mu_2$. Applying σ to the equality $y^2 = xz$, that is, $\iota(\mu_1)^2 = \mu_1\mu_2$, we get $\iota(\mu_1)^2 = \mu_1\iota(\mu_2)$, whence $\iota(\mu_2) = \mu_2$. But this implies that μ_2 is real, which is once again a contradiction. \square

Proof of Proposition A.2. Let ℓ be a prime unramified in K and strictly larger than $b(2[K : \mathbb{Q}], 4, 2h(A))^{\frac{1}{4}}$. Let $\rho_\ell : \text{Gal}(\bar{K}/K) \rightarrow \text{Aut } A[\ell]$ be the natural Galois representation associated with the ℓ -torsion of A .

Much of the proof of [Lombardo 2016b, Theorem 3.19] still applies in the current setting, and shows that one of the following holds:

- (a) $\text{Im}(\rho_{\ell^\infty}) = \text{GSp}_4(\mathbb{Z}_\ell)$,
- (b) the image of ρ_ℓ is contained in a maximal subgroup of $\text{GSp}_4(\mathbb{F}_\ell)$ of type (2) in the sense of Theorem 3.3 in [Lombardo 2016b].

If we are in case (a) we are done, so assume we are in case (b). To conclude the proof, we shall show that $\ell \leq (2q_v)^8$. If ℓ is equal to the residual characteristic of v this inequality is obvious, so we can assume that $v \nmid \ell$. In this case, the characteristic polynomial of the action of Frob_v on $T_\ell A$ is $f_v(x)$. By [Lombardo 2016b, Lemma 3.4], the eigenvalues of any $x \in \text{Im}(\rho_\ell)$ can be written as $\lambda \cdot \lambda_1^3, \lambda \cdot \lambda_1^2\lambda_2, \lambda \cdot \lambda_1\lambda_2^2, \lambda \cdot \lambda_2^3$ for some $\lambda, \lambda_1, \lambda_2 \in \mathbb{F}_{\ell^2}^\times$. Taking $g := \rho_\ell(\text{Frob}_v)$, we may assume the four eigenvalues v_1, \dots, v_4 of g satisfy $v_2^2 = v_1v_3$.

Let λ be a place of $F(v)$ of characteristic ℓ and identify λ with a maximal ideal of $\mathcal{O}_{F(v)}$. Since $f_v(x)$ splits completely in $F(v)$ by definition, its four roots $\mu_1, \mu_2, \iota(\mu_1), \iota(\mu_2)$ all belong to $\mathcal{O}_{F(v)}$. Upon reduction modulo λ , these four roots yield four elements of $\mathcal{O}_{F(v)}/\lambda$, which is a finite field of characteristic ℓ . Moreover, as $\{\mu_1, \mu_2, \iota(\mu_1), \iota(\mu_2)\}$ is a Galois-stable set, its image in $\bar{\mathbb{F}}_\ell$ is independent of the choice embedding of $\mathcal{O}_{F(v)}/\lambda$ into $\bar{\mathbb{F}}_\ell$, and hence well defined. Denote by $\bar{\mu}_1, \bar{\mu}_2, \overline{\iota(\mu_1)}, \overline{\iota(\mu_2)}$ the images of $\mu_1, \mu_2, \iota(\mu_1), \iota(\mu_2)$ in $\bar{\mathbb{F}}_\ell$.

Now observe that the characteristic polynomial of g is the reduction modulo ℓ of $f_v(x)$, so its roots $v_1, \dots, v_4 \in \bar{\mathbb{F}}_\ell^\times$ must coincide with $\bar{\mu}_1, \bar{\mu}_2, \overline{\iota(\mu_1)}, \overline{\iota(\mu_2)}$ in some order. Given that $v_2^2 = v_1v_3$, there are three (necessarily distinct) eigenvalues of Frob_v , call them x, y, z , that satisfy $y^2 - xz \equiv 0 \pmod{\lambda}$. By Lemma A.3, $N_{F(v)/\mathbb{Q}}(y^2 - xz)$ is a nonzero integer. Therefore, $N_{F(v)/\mathbb{Q}}(y^2 - xz)$ has positive valuation at λ , hence it is divisible by ℓ . In turn, this gives

$$\ell \leq |N_{F(v)/\mathbb{Q}}(y^2 - xz)| = \prod_{\sigma \in \text{Gal}(F(v)/\mathbb{Q})} |\sigma(y)^2 - \sigma(x)\sigma(z)| \leq (2q_v)^8,$$

where the inequality $|\sigma(y)^2 - \sigma(x)\sigma(z)| \leq 2q_v$ follows immediately from the triangle inequality and the Weil conjectures. \square

Acknowledgments

This research was supervised by Ken Ono and David Zureick-Brown at the Emory University Mathematics REU and was supported by the National Science Foundation (grant number DMS-1557960). We would like to thank David Zureick-Brown for suggesting the problem that led to the present article and for offering us his invaluable advice and guidance; in particular, we acknowledge David Zureick-Brown

for providing a detailed outline of the material on heights in Section 4I5. and for much help proving Proposition 5.2. We would like to acknowledge Brian Conrad for his meticulous efforts in providing us with enlightening comments, corrections, and suggestions on nearly every part of this paper. In addition, we would like to thank Davide Lombardo for writing an appendix for the present article and for many fruitful conversations. We thank Daniel Litt for much help proving Proposition 4.1. We also thank the referees for their helpful comments and suggestions. Finally, we would like to thank Jeff Achter, Jarod Alper, Michael Aschbacher, Anna Cadoret, Alina Cojocaru, John Cullinan, Dougal Davis, Anand Deopurkar, Noam Elkies, Jordan Ellenberg, Tony Feng, Nick Gill, Jack Hall, Joe Harris, Eric Katz, Mark Kisin, Ben Moonen, Jackson Morrow, Anand Patel, Bjorn Poonen, Jeremy Rickard, Eric Riedl, Simon Rubinstein-Salzedo, David Rydh, Jesse Silliman, Jacob Tsimerman, Evelina Viada, Erik Wallace, and Alex Wright for their helpful advice. We used *magma* and *Mathematica* for explicit calculations.

References

- [A'Campo 1979] N. A'Campo, "Tresses, monodromie et le groupe symplectique", *Comment. Math. Helv.* **54**:2 (1979), 318–327. MR Zbl
- [Anni et al. 2016] S. Anni, P. Lemos, and S. Siksek, "Residual representations of semistable principally polarized abelian varieties", *Res. Number Theory* **2**:1 (2016), art. 1, 12 pp. MR Zbl
- [Arbarello et al. 1985] E. Arbarello, M. Cornalba, P. A. Griffiths, and J. Harris, *Geometry of algebraic curves, I*, Grundlehren der Math. Wissenschaften **267**, Springer, 1985. MR Zbl
- [Arbarello et al. 2011] E. Arbarello, M. Cornalba, and P. A. Griffiths, *Geometry of algebraic curves, II*, Grundlehren der Math. Wissenschaften **268**, Springer, 2011. MR Zbl
- [Beauville 1986] A. Beauville, "Le groupe de monodromie des familles universelles d'hypersurfaces et d'intersections complètes", pp. 8–18 in *Complex analysis and algebraic geometry* (Göttingen, 1985), edited by H. Grauert, Lecture Notes in Math. **1194**, Springer, 1986. MR Zbl
- [Bolognesi and Lönne 2016] M. Bolognesi and M. Lönne, "Mapping class groups of trigonal loci", *Selecta Math. (N.S.)* **22**:1 (2016), 417–445. MR Zbl
- [Cadoret 2015] A. Cadoret, "An open adelic image theorem for abelian schemes", *Int. Math. Res. Not.* **2015**:20 (2015), 10208–10242. MR Zbl
- [Cadoret and Moonen 2018] A. Cadoret and B. Moonen, "Integral and adelic aspects of the Mumford–Tate conjecture", *J. Inst. Math. Jussieu* (online publication June 2018).
- [Cadoret and Tamagawa 2012] A. Cadoret and A. Tamagawa, "A uniform open image theorem for ℓ -adic representations, I", *Duke Math. J.* **161**:13 (2012), 2605–2634. MR Zbl
- [Cadoret and Tamagawa 2013] A. Cadoret and A. Tamagawa, "A uniform open image theorem for ℓ -adic representations, II", *Duke Math. J.* **162**:12 (2013), 2301–2344. MR Zbl
- [Casnati and Fontanari 2007] G. Casnati and C. Fontanari, "On the rationality of moduli spaces of pointed curves", *J. Lond. Math. Soc. (2)* **75**:3 (2007), 582–596. MR Zbl
- [Cojocaru and Hall 2005] A. C. Cojocaru and C. Hall, "Uniform results for Serre's theorem for elliptic curves", *Int. Math. Res. Not.* **2005**:50 (2005), 3065–3080. MR Zbl
- [Cojocaru et al. 2011] A.-C. Cojocaru, D. Grant, and N. Jones, "One-parameter families of elliptic curves over \mathbb{Q} with maximal Galois representations", *Proc. Lond. Math. Soc. (3)* **103**:4 (2011), 654–675. MR Zbl
- [Deligne and Mumford 1969] P. Deligne and D. Mumford, "The irreducibility of the space of curves of given genus", *Inst. Hautes Études Sci. Publ. Math.* **36** (1969), 75–109. MR Zbl
- [Dieulefait 2002] L. V. Dieulefait, "Explicit determination of the images of the Galois representations attached to abelian surfaces with $\text{End}(A) = \mathbb{Z}$ ", *Experiment. Math.* **11**:4 (2002), 503–512. MR Zbl

- [Duke 1997] W. Duke, “Elliptic curves with no exceptional primes”, *C. R. Acad. Sci. Paris Sér. I Math.* **325**:8 (1997), 813–818. MR Zbl
- [EGA IV₃ 1966] A. Grothendieck, “Eléments de géométrie algébrique, IV: Étude locale des schémas et des morphismes de schémas, III”, *Inst. Hautes Études Sci. Publ. Math.* **28** (1966), 5–255. MR Zbl
- [EGA IV₄ 1967] A. Grothendieck, “Eléments de géométrie algébrique, IV: Étude locale des schémas et des morphismes de schémas, IV”, *Inst. Hautes Études Sci. Publ. Math.* **32** (1967), 5–361. MR Zbl
- [Ekedahl 1990] T. Ekedahl, “An effective version of Hilbert’s irreducibility theorem”, pp. 241–249 in *Séminaire de Théorie des Nombres* (Paris 1988–1989), edited by C. Goldstein, Progr. Math. **91**, Birkhäuser, Boston, 1990. MR Zbl
- [Ellenberg et al. 2009] J. S. Ellenberg, C. Elsholtz, C. Hall, and E. Kowalski, “Non-simple abelian varieties in a family: geometric and analytic approaches”, *J. Lond. Math. Soc.* (2) **80**:1 (2009), 135–154. MR Zbl
- [Faltings 1983] G. Faltings, “Endlichkeitssätze für abelsche Varietäten über Zahlkörpern”, *Invent. Math.* **73**:3 (1983), 349–366. Translated in *Arithmetic geometry, 1986, Springer, 9–26*. MR Zbl
- [Faltings et al. 1992] G. Faltings, G. Wüstholz, F. Grunewald, N. Schappacher, and U. Stuhler, *Rational points* (Bonn/Wuppertal, 1983/1984), 3rd ed., Aspects of Mathematics **E6**, Friedr. Vieweg & Sohn, Braunschweig, 1992. MR
- [Grant 2000] D. Grant, “A formula for the number of elliptic curves with exceptional primes”, *Compositio Math.* **122**:2 (2000), 151–164. MR Zbl
- [Greicius 2010] A. Greicius, “Elliptic curves with surjective adelic Galois representations”, *Experiment. Math.* **19**:4 (2010), 495–507. MR Zbl
- [Grothendieck 1966] A. Grothendieck, “Un théorème sur les homomorphismes de schémas abéliens”, *Invent. Math.* **2** (1966), 59–78. MR Zbl
- [Hindry and Silverman 2000] M. Hindry and J. H. Silverman, *Diophantine geometry: an introduction*, Graduate Texts in Mathematics **201**, Springer, 2000. MR Zbl
- [Jones 2010] N. Jones, “Almost all elliptic curves are Serre curves”, *Trans. Amer. Math. Soc.* **362**:3 (2010), 1547–1570. MR Zbl
- [Jouanolou 1983] J.-P. Jouanolou, *Théorèmes de Bertini et applications*, Progress in Mathematics **42**, Birkhäuser, Boston, 1983. MR Zbl
- [Kawamura 2003] T. Kawamura, “The effective surjectivity of mod l Galois representations of 1- and 2-dimensional abelian varieties with trivial endomorphism ring”, *Comment. Math. Helv.* **78**:3 (2003), 486–493. MR Zbl
- [Kowalski 2006] E. Kowalski, “The large sieve, monodromy and zeta functions of curves”, *J. Reine Angew. Math.* **601** (2006), 29–69. MR Zbl
- [Landesman et al. 2017a] A. Landesman, A. Swaminathan, J. Tao, and Y. Xu, “Hyperelliptic curves with maximal Galois action on the torsion points of their jacobians”, preprint, 2017. arXiv
- [Landesman et al. 2017b] A. Landesman, A. A. Swaminathan, J. Tao, and Y. Xu, “Lifting subgroups of symplectic groups over $\mathbb{Z}/\ell\mathbb{Z}$ ”, *Res. Number Theory* **3** (2017), art. id. 14, 12 pp. MR
- [Laumon and Moret-Bailly 2000] G. Laumon and L. Moret-Bailly, *Champs algébriques*, Ergebnisse der Mathematik (3) **39**, Springer, 2000. MR Zbl
- [Lombardo 2016a] D. Lombardo, “Explicit open image theorems for abelian varieties with trivial endomorphism ring”, preprint, 2016. arXiv
- [Lombardo 2016b] D. Lombardo, “Explicit surjectivity of Galois representations for abelian surfaces and GL_2 -varieties”, *J. Algebra* **460** (2016), 26–59. MR Zbl
- [Ma 2015] S. Ma, “The rationality of the moduli spaces of trigonal curves”, *Int. Math. Res. Not.* **2015**:14 (2015), 5456–5472. MR Zbl
- [Mochizuki 1999] S. Mochizuki, *Foundations of p -adic Teichmüller theory*, AMS/IP Studies in Advanced Mathematics **11**, American Mathematical Society, Providence, RI, 1999. MR Zbl
- [Morris 2015] D. W. Morris, *Introduction to arithmetic groups*, Deductive Press, 2015. MR Zbl
- [Mumford 2007] D. Mumford, *Tata lectures on theta, II: Jacobian theta functions and differential equations*, Birkhäuser, Boston, 2007. MR Zbl

- [Olsson 2016] M. Olsson, *Algebraic spaces and stacks*, American Mathematical Society Colloquium Publications **62**, American Mathematical Society, Providence, RI, 2016. MR Zbl
- [O’Meara 1978] O. T. O’Meara, *Symplectic groups*, Mathematical Surveys **16**, American Mathematical Society, Providence, R.I., 1978. MR Zbl
- [Oort 1971] F. Oort, “Finite group schemes, local moduli for abelian varieties, and lifting problems”, *Compositio Math.* **23** (1971), 265–296. MR Zbl
- [Orgogozo and Vidal 2000] F. Orgogozo and I. Vidal, “Le théorème de spécialisation du groupe fondamental”, pp. 169–184 in *Courbes semi-stables et groupe fondamental en géométrie algébrique* (Luminy, 1998), Progr. Math. **187**, Birkhäuser, Basel, 2000. MR Zbl
- [Patel and Vakil 2015] A. Patel and R. Vakil, “On the Chow ring of the Hurwitz space of degree three covers of \mathbf{P}^1 ”, preprint, 2015. arXiv
- [Arias-de Reyna et al. 2016] S. Arias-de Reyna, C. Armana, V. Karemaker, M. Rebolledo, L. Thomas, and N. Vila, “Large Galois images for Jacobian varieties of genus 3 curves”, *Acta Arith.* **174**:4 (2016), 339–366. MR
- [Ribet 1976] K. A. Ribet, “Galois action on division points of Abelian varieties with real multiplications”, *Amer. J. Math.* **98**:3 (1976), 751–804. MR Zbl
- [Rivin 2008] I. Rivin, “Walks on groups, counting reducible matrices, polynomials, and surface and free group automorphisms”, *Duke Math. J.* **142**:2 (2008), 353–379. MR Zbl
- [Serre 1972] J.-P. Serre, “Propriétés galoisiennes des points d’ordre fini des courbes elliptiques”, *Invent. Math.* **15**:4 (1972), 259–331. MR Zbl
- [Serre 1997] J.-P. Serre, *Lectures on the Mordell–Weil theorem*, 3rd ed., Aspects of Mathematics **E15**, Friedr. Vieweg & Sohn, Braunschweig, 1997. MR Zbl
- [Serre 1998] J.-P. Serre, *Abelian l -adic representations and elliptic curves*, Research Notes in Mathematics **7**, A K Peters, Wellesley, MA, 1998. MR Zbl
- [SGA 1 1971] A. Grothendieck, *Revêtements étales et groupe fondamental* (Séminaire de Géométrie Algébrique du Bois Marie 1960–1961), Lecture Notes in Math. **224**, Springer, 1971. MR Zbl
- [Stacks 2005–] P. Belmans, A. J. de Jong, et al., “The Stacks project”, electronic reference, 2005–, Available at <http://stacks.math.columbia.edu>.
- [Steinberg 1961] R. Steinberg, “Automorphisms of classical Lie algebras”, *Pacific J. Math.* **11** (1961), 1119–1129. MR Zbl
- [Verra 2005] A. Verra, “The unirationality of the moduli spaces of curves of genus 14 or lower”, *Compos. Math.* **141**:6 (2005), 1425–1444. MR Zbl
- [Wallace 2014] E. Wallace, “Principally polarized abelian surfaces with surjective Galois representations on l -torsion”, *J. Lond. Math. Soc.* (2) **90**:2 (2014), 451–471. MR Zbl
- [Zywina 2010a] D. Zywina, “Elliptic curves with maximal Galois action on their torsion points”, *Bull. Lond. Math. Soc.* **42**:5 (2010), 811–826. MR Zbl
- [Zywina 2010b] D. Zywina, “Hilbert’s irreducibility theorem and the larger sieve”, preprint, 2010. arXiv
- [Zywina 2015] D. Zywina, “An explicit Jacobian of dimension 3 with maximal Galois action”, preprint, 2015. arXiv

Communicated by Bjorn Poonen

Received 2017-10-26 Revised 2018-10-01 Accepted 2019-02-22

aaronlandesman@stanford.edu

Department of Mathematics, Stanford University, CA, United States

ashvins@math.princeton.edu

Department of Mathematics, Princeton University, NJ, United States

jamestao@mit.edu

Department of Mathematics, Massachusetts Institute of Technology, Cambridge, MA, United States

yujie@math.harvard.edu

Department of Mathematics, Harvard University, Cambridge, MA, United States

davide.lombardo@unipi.it

Dipartimento di Matematica, Università di Pisa, Italy

A unified and improved Chebotarev density theorem

Jesse Thorner and Asif Zaman

We establish an unconditional effective Chebotarev density theorem that improves uniformly over the well-known result of Lagarias and Odlyzko. As a consequence, we give a new asymptotic form of the Chebotarev density theorem that can count much smaller primes with arbitrary log-power savings, even in the case where a Landau–Siegel zero is present. Our main theorem also interpolates the strongest unconditional upper bound for the least prime ideal with a given Artin symbol as well as the Chebotarev analogue of the Brun–Titchmarsh theorem proved by the authors.

1. Introduction and statement of results

1A. Introduction. Let L/F be a Galois extension of number fields with Galois group G . For each prime ideal \mathfrak{p} of F that is unramified in L , we use the Artin symbol $\left[\frac{L/F}{\mathfrak{p}}\right]$ to denote the conjugacy class of G consisting of the set of Frobenius automorphisms attached to the prime ideals \mathfrak{P} of L which lie over \mathfrak{p} . For any conjugacy class $C \subseteq G$, define the function

$$\pi_C(x) = \pi_C(x, L/F) = \#\{N_{F/\mathbb{Q}}\mathfrak{p} \leq x : \mathfrak{p} \text{ unramified in } L, \left[\frac{L/F}{\mathfrak{p}}\right] = C\}, \quad (1-1)$$

where $N_{F/\mathbb{Q}}$ is the absolute norm of F/\mathbb{Q} . The Chebotarev density theorem states that

$$\pi_C(x) \sim \frac{|C|}{|G|} \text{Li}(x) \quad \text{as } x \rightarrow \infty.$$

It follows from work of V.K. Murty [1997, Section 4] that there exists an absolute, effective, and positive constant c_1 such that

$$\pi_C(x) = \frac{|C|}{|G|} (\text{Li}(x) - \theta_1 \text{Li}(x^{\beta_1}) + O(xe^{-c_1\sqrt{\frac{\log x}{n_L}}})) , \quad \log x \gg \frac{(\log D_L)^2}{n_L} + n_L(\log n_L)^2, \quad (1-2)$$

which refines a well-known result of Lagarias and Odlyzko [1977, Theorem 1.2]. Here, D_L is the absolute discriminant of L , $n_L = [L : \mathbb{Q}]$ is the degree of L over \mathbb{Q} , β_1 is a possible Landau–Siegel zero of the Dedekind zeta function $\zeta_L(s)$ of L , and $\theta_1 = \theta_1(C) \in \{-1, 0, 1\}$ depends on C ; in particular, $\theta_1(C) = 0$ if and only if β_1 does not exist. For comparison, Lagarias and Odlyzko [1977, Theorem 1.1] proved that the generalized Riemann hypothesis for $\zeta_L(s)$ implies the more uniform result

$$\pi_C(x) = \frac{|C|}{|G|} (\text{Li}(x) + O(\sqrt{x} \log(D_L x^{n_L}))), \quad x \gg (\log D_L)^2 (\log \log D_L)^4. \quad (1-3)$$

MSC2010: 11R44.

Keywords: distribution of primes, Chebotarev density theorem, effective, uniform, binary quadratic forms.

As of now, the best bound for β_1 is due to Stark [1974, Theorem 1', p. 148]; it implies that

$$1 - \beta_1 \gg (n_L^{n_L} \log D_L + D_L^{1/n_L})^{-1}. \tag{1-4}$$

Therefore, in order to ensure that $\frac{|C|}{|G|} \text{Li}(x)$ dominates all other terms in (1-2), one must take the range of x to be

$$\log x \gg n_L^{-1} (\log D_L)^2 + n_L (\log n_L)^2 + (1 - \beta_1)^{-1} \tag{1-5}$$

and apply (1-4) if β_1 exists. Otherwise, one omits the last term in (1-5) if β_1 does not exist. Regardless, (1-5) is very prohibitive in many applications where uniformity in L/F is crucial. Thus it often helps in applications to have upper and lower bounds for $\pi_C(x)$ of order $\text{Li}(x)$ in ranges of x which are more commensurate with (1-3). Lagarias, Montgomery, and Odlyzko [1979] made substantial progress on these problems; their work has been improved upon by Weiss [1983], the authors [Thorner and Zaman 2017; 2018], and Zaman [2017]. In particular, it follows from the joint work of the authors [Thorner and Zaman 2017; 2018] that there exist absolute, effective constants $A > 2$ and $B > 2$ such that if D_L is sufficiently large, then

$$\frac{1}{(D_L n_L^{n_L})^A} \frac{|C|}{|G|} \text{Li}(x) \ll \pi_C(x) < (2 + o(1)) \frac{|C|}{|G|} \text{Li}(x) \quad \text{for } x \geq (D_L n_L^{n_L})^B, \tag{1-6}$$

where the $o(1)$ term tends to zero as $(\log x) / \log(D_L n_L^{n_L})$ tends to infinity.¹

To summarize the above discussion, suppose that we are in the worst case scenario with $\theta_1 = 1$ and β_1 is as bad as (1-4) permits. If one is willing to sacrifice an asymptotic equality for $\pi_C(x)$ in order to obtain estimates in noticeably better ranges than (1-5), then one might use (1-6). On the other hand, if one needs an asymptotic equality for $\pi_C(x)$, then one uses (1-2) in the prohibitive range (1-5).

1B. Results. Our main result, Theorem 1.4, is a new asymptotic equality for $\pi_C(x)$ which interpolates both of the aforementioned options while providing several new options. In other words, we prove a new asymptotic equality for $\pi_C(x)$ from which one may deduce both (1-2) and (1-6). First, we present a simplified version of the main result.

Theorem 1.1. *Let L/F be a Galois extension of number fields with Galois group G , and let $C \subseteq G$ be a conjugacy class. Let β_1 denote the Landau–Siegel zero of the Dedekind zeta function $\zeta_L(s)$, if it exists. There exist absolute and effective constants $c_2 > 0$ and $c_3 > 0$ such that if $L \neq \mathbb{Q}$ and $x \geq (D_L n_L^{n_L})^{c_2}$, then*

$$\pi_C(x) = \frac{|C|}{|G|} (\text{Li}(x) - \theta_1 \text{Li}(x^{\beta_1})) \left(1 + O \left(\exp \left[-\frac{c_3 \log x}{\log(D_L n_L^{n_L})} \right] + \exp \left[-\frac{(c_3 \log x)^{1/2}}{n_L^{1/2}} \right] \right) \right),$$

where $\theta_1 = \theta_1(C) \in \{-1, 0, 1\}$. In particular, $\theta_1 = 0$ precisely when β_1 does not exist.

¹The term $n_L^{n_L}$ is usually negligible compared to a power of D_L . If not, one might appeal to [Zaman 2017, Theorem 1.3.1] which states that $\pi_C(x) \gg D_L^{-A} \frac{|C|}{|G|} \text{Li}(x)$ for $x \geq D_L^B$.

The inequality

$$\exp\left[-\frac{(c_3 \log x)^{1/2}}{n_L^{1/2}}\right] \gg \exp\left[-\frac{c_3 \log x}{\log(D_L n_L^{n_L})}\right],$$

holds when $\log x \gg (\log D_L)^2/n_L + n_L(\log n_L)^2$, so we see that Theorem 1.1 recovers (1-2) and is therefore a uniform improvement over it. Also, it follows from the mean value theorem and (1-4) that

$$\text{Li}(x) - \theta_1 \text{Li}(x^{\beta_1}) \gg ((1 - \beta_1) \log(D_L n_L^{n_L})) \text{Li}(x) \gg \frac{\log(D_L n_L^{n_L})}{D_L^{1/n_L} + n_L^{n_L} \log D_L} \text{Li}(x). \tag{1-7}$$

With this lower bound at our disposal, one can see that Theorem 1.1 recovers (1-6). Thus Theorem 1.1 unifies and improves both (1-2) and (1-6).

As noted above, if one wants $\frac{|C|}{|G|} \text{Li}(x)$ to dominate all other terms in (1-2), then one must take x in the range (1-5). However, one can plainly see that

$$\frac{|C|}{|G|}(\text{Li}(x) - \theta_1 \text{Li}(x^{\beta_1})) \tag{1-8}$$

dominates all other terms in Theorem 1.1 for all x in the claimed range, provided that c_2 is suitably large compared to c_3 . At first glance, it may seem awkward that we adjoin the contribution from β_1 to the “main term” when it is classically viewed as an error term. But without eliminating the existence of β_1 , it is well known that in situations where $\theta_1 \neq 0$ and x is small, say $\log x \ll \log(D_L n_L^{n_L})$, the term $-\theta_1 \frac{|C|}{|G|} \text{Li}(x^{\beta_1})$ is more properly treated as a secondary term than an error term. When $\theta_1 = 1$ and β_1 is especially close to 1, this secondary term causes serious difficulties in the proof of Linnik’s bound [1944] for the least prime in an arithmetic progression. Fortunately, it follows from (1-7) that regardless of whether β_1 exists, we have

$$\text{Li}(x) \ll_L \text{Li}(x) - \theta_1 \text{Li}(x^{\beta_1}) < 2 \text{Li}(x). \tag{1-9}$$

Therefore, in the range of x where $-\frac{|C|}{|G|} \theta_1 \text{Li}(x^{\beta_1})$ acts like a secondary term, (1-9) shows that Theorem 1.1 recovers upper and lower bounds of order $\text{Li}(x)$ precisely because (1-8) dominates all other terms in Theorem 1.1. This perspective is implicit in Linnik’s work. On the other hand, when x is sufficiently large in terms of L/F per (1-5), the contribution from β_1 can be safely absorbed into the O -term in Theorem 1.1. In light of these observations, we believe that viewing (1-8) as the “main term” in Theorem 1.1 helps to clarify the role of the contribution from β_1 when one transitions from small values of x to large values of x .

Upon considering the O -term in Theorem 1.1, we see that Theorem 1.1 noticeably improves the range of x in which we have an asymptotic equality for $\pi_C(x)$.

Corollary 1.2. *If $\log x / \log(D_L n_L^{n_L}) \rightarrow \infty$, then $\pi_C(x) \sim \frac{|C|}{|G|}(\text{Li}(x) - \theta_1 \text{Li}(x^{\beta_1}))$.*

Theorem 1.1 also produces a new asymptotic equality in which the error term saves an arbitrarily large power of $\log x$ in a much stronger range of x than (1-2).

Corollary 1.3. *Let $A > 1$. If $\log x \gg_A (\log D_L)(\log \log D_L) + n_L(\log n_L)^2$, then*

$$\pi_C(x) = \frac{|C|}{|G|}(\text{Li}(x) - \theta_1 \text{Li}(x^{\beta_1}))(1 + O_A((\log x)^{-A})). \tag{1-10}$$

In order to state the main result from which Theorem 1.1 follows, we introduce some additional notation. Let $H \subseteq G$ be an abelian subgroup of G such that $H \cap C$ is nonempty, and let $K = L^H$ be the fixed field of H . The characters χ in the dual group \hat{H} are Hecke characters; we write the conductor of χ as f_χ . Define

$$\mathcal{Q} = \mathcal{Q}(L/K) = \max_{\chi \in \hat{H}} N_{K/\mathbb{Q}} f_\chi. \tag{1-11}$$

We write the L -function associated to such a Hecke character as $L(s, \chi, L/K)$. From work of Stark [1974], at most one real Hecke character $\chi_1 \in \hat{H}$ has an associated Hecke L -function $L(s, \chi_1, L/K)$ with a Landau–Siegel zero $\beta_1 = 1 - \lambda_1 / \log(D_K \mathcal{Q} n_K^{n_K})$, where $0 < \lambda_1 < \frac{1}{8}$.

Theorem 1.4. *Let L/F be a Galois extension of number fields with Galois group G , and let $C \subseteq G$ be a conjugacy class. Let $H \subseteq G$ be an abelian subgroup such that $C \cap H$ is nonempty, let K be the fixed field of H , and choose $g_C \in C \cap H$. If $x \geq (D_K \mathcal{Q} n_K^{n_K})^{c_2}$, then*

$$\pi_C(x) = \frac{|C|}{|G|} (\text{Li}(x) - \theta_1 \text{Li}(x^{\beta_1})) \left(1 + O \left(\exp \left[-\frac{c_3 \log x}{\log(D_K \mathcal{Q} n_K^{n_K})} \right] + \exp \left[-\frac{(c_3 \log x)^{1/2}}{n_K^{1/2}} \right] \right) \right),$$

where $\theta_1 = \chi_1(g_C)$ if β_1 exists and $\theta_1 = 0$ otherwise and \mathcal{Q} is given by (1-11). The constants c_2 and c_3 are the same as in Theorem 1.1.

Remark 1.5. As a group-theoretic quantity, θ_1 depends on the choice of $g_C \in C \cap H$. However, if $\theta_1 \neq 0$, then the existence of β_1 implies that θ_1 is well defined.

1C. An application. While it is aesthetically appealing to be able to encapsulate the work in [Lagarias et al. 1979; Lagarias and Odlyzko 1977; Murty 1997; Thorner and Zaman 2017; 2018; Weiss 1983] with a single asymptotic equality, Theorem 1.4 can make progress in certain sieve-theoretic problems when one must compute the local densities. As an example, we prove a new result in the study of primes represented by binary quadratic forms. Let

$$f(u, v) = au^2 + buv + cv^2 \in \mathbb{Z}[u, v]$$

be a positive definite binary quadratic form of discriminant $D = b^2 - 4ac < 0$. We do not assume that D is fundamental. The group $\text{SL}_2(\mathbb{Z})$ naturally acts on such forms by $(T \cdot f)(\mathbf{x}) = f(T\mathbf{x})$ for $T \in \text{SL}_2(\mathbb{Z})$. The class number $h(D)$ is the number of such forms up to SL_2 -equivalence. If f is primitive (that is, $(a, b, c) = 1$) then it is a classical consequence of the Chebotarev density theorem and class field theory that

$$\frac{1}{|\text{stab}(f)|} \sum_{\substack{u, v \in \mathbb{Z} \\ au^2 + buv + cv^2 \leq x}} \mathbf{1}_{\mathbb{P}}(au^2 + buv + cv^2) \sim \frac{\text{Li}(x)}{h(D)} \quad \text{as } x \rightarrow \infty, \tag{1-12}$$

where $\mathbf{1}_{\mathbb{P}}$ is the indicator function for the odd primes and

$$\text{stab}(f) = \{T \in \text{SL}_2(\mathbb{Z}) : T \cdot f = f\}.$$

Note $|\text{stab}(f)| = 2$ unless $D = -3$ or -4 in which case it equals 6 and 4 respectively.

We consider the question of imposing restrictions on the integers u and v which comprise a solution to the equation $p = f(u, v)$. In the special case of $f(u, v) = u^2 + v^2$, Fouvry and Iwaniec [1997] proved that there are infinitely many primes p such that $p = u^2 + v^2$ and u is prime. Their proof, which relies on sieve methods, enables them to asymptotically count such primes.

One might ask whether their methods extend to all positive definite primitive $f(u, v)$ with strong uniformity in the discriminant D .² The answer is not clear to the authors. Nevertheless, Theorem 1.4 enables us to study the distribution of primes $p = f(u, v)$ with some control over the divisors of u and v while maintaining strong uniformity in D . We prove the following result in Section 7.

Theorem 1.6. *Let $D \leq -3$ be an integer and let $f(u, v) = au^2 + buv + cv^2$ be a positive definite primitive integral binary quadratic form with discriminant $D = b^2 - 4ac$. Let P be any integer dividing the product of primes $p \leq z$. For all $A \geq 1$, there exists a sufficiently small constant $\eta = \eta(A) > 0$ such that if $3 \leq z \leq x^{\eta/\log \log x}$ and $3 \leq |D| \leq x^{\eta/\log \log z}$, then*

$$\frac{1}{|\text{stab}(f)|} \sum_{\substack{u, v \in \mathbb{Z} \\ au^2 + buv + cv^2 \leq x \\ (uv, P) = 1}} \mathbf{1}_{\mathbb{P}}(au^2 + buv + cv^2) = \delta_f(P) \frac{\text{Li}(x) - \text{Li}(x^{\beta_1})}{h(D)} \{1 + O_A((\log z)^{-A})\}. \quad (1-13)$$

Here, β_1 is a real simple zero of the Dedekind zeta function $\zeta_{\mathbb{Q}(\sqrt{D})}(s)$ (if it exists),

$$\delta_f(P) = \prod_{p|P} \left(1 - \frac{2 - \mathbf{1}_{p|a}(p) - \mathbf{1}_{p|c}(p)}{p - \left(\frac{D}{p}\right)} \right), \quad (1-14)$$

$\left(\frac{D}{p}\right)$ is the Legendre symbol for $p \neq 2$, $\left(\frac{D}{2}\right)$ is defined by (7-6), and the term $\text{Li}(x^{\beta_1})$ is omitted if β_1 does not exist.

Remark 1.7. The constant $\delta_f(P)$ is always nonnegative. It is possible that $\delta_f(P) = 0$ due to the local factor at $p = 2$ in the product but this occurs precisely when the form $f(u, v)$ does not represent any odd primes. Since $\mathbf{1}_{\mathbb{P}}$ is the indicator function for the *odd* primes, (1-13) trivially holds in this case. The details of this casework are verified in Section 7A1.

While it is natural to think of P as equal to the product of primes up to z , we immediately obtain from Theorem 1.6 the following corollary when P is a *fixed* divisor of the product of primes up to z and $z \rightarrow \infty$ arbitrarily slowly.

Corollary 1.8. *Keep the assumptions of Theorem 1.6. If the integer $P \geq 1$ is fixed, then*

$$\frac{1}{|\text{stab}(f)|} \sum_{\substack{u, v \in \mathbb{Z} \\ au^2 + buv + cv^2 \leq x \\ (uv, P) = 1}} \mathbf{1}_{\mathbb{P}}(au^2 + buv + cv^2) \sim \delta_f(P) \frac{\text{Li}(x) - \text{Li}(x^{\beta_1})}{h(D)} \quad \text{as } \frac{\log x}{\log |D|} \rightarrow \infty.$$

²Added in proof, 17 June 2019: Lam, Schindler, and Xiao [2018] recently extended Fouvry and Iwaniec’s result to all positive-definite primitive binary quadratic forms. However, their error terms do not possess uniformity in the discriminant.

In particular, there exists a prime $p \leq |D|^\alpha$ and $u, v \in \mathbb{Z}$ such that $p = f(u, v)$, $p \nmid D$, and $(uv, P) = 1$, where $\alpha = \alpha(P) > 0$ is a sufficiently large constant depending only on P .

In order to prove Theorem 1.6 with strong uniformity in z and $|D|$, one needs asymptotic control over sums like (1-12) (see (7-4) below) when x is as small as a polynomial in the discriminant, regardless of whether $\zeta_{\mathbb{Q}(\sqrt{D})}(s)$ has a Landau–Siegel zero. This is precisely what Theorem 1.4 provides. For comparison, a slightly stronger version of (1-2) that follows from [Murty 1997] along with the effective bound $(1 - \beta_1)^{-1} \ll |D|^{1/2} \log |D|$ can produce (1-13) with the inferior ranges

$$3 \leq |D| \ll (\log x)^2 / (\log \log x)^2 \quad \text{and} \quad 3 \leq z \leq \exp(c\sqrt{\log x})$$

where $c > 0$ is an absolute constant. As one can plainly see, Theorem 1.4 yields substantial gains over earlier versions of the Chebotarev density theorem. See Remark 7.3 for further discussion.

1D. Overview of the methods. We now give an overview of how the proof of Theorem 1.4 differs from the proofs in [Lagarias et al. 1979; Lagarias and Odlyzko 1977; Murty 1997; Thorner and Zaman 2017; 2018; Weiss 1983]. For convenience, we refer to

$$\frac{|C|}{|G|} (\text{Li}(x) - \theta_1 \text{Li}(x^{\beta_1}))$$

as the “main term” in Theorem 1.4 and all other terms as the “error term”.

The key difference between the proof of (1-2) and the proof of Theorem 1.4 lies in the study of the nontrivial low-lying zeros of $\zeta_L(s)$. The standard zero-free region for $\zeta_L(s)$ indicates that the low-lying zeros of $\zeta_L(s)$ lie further away from the edge of the critical strip $\{s \in \mathbb{C} : 0 < \text{Re}(s) < 1\}$ than zeros of large height. However, the treatments in [Lagarias and Odlyzko 1977; Murty 1997] handle the contribution from the all of the nontrivial zeros by assuming that the low-lying zeros (other than β_1 , if it exists) lie just as close to the edge of the critical strip as zeros of large height. This unduly inflates the contribution from the low-lying zeros, leading to the poor field uniformity in (1-2) along with the poor dependence on the Landau–Siegel zero β_1 if it exists. Consequently, both the range of x and the quality the error term in (1-2) directly depend on the quality of zero-free region available for $\zeta_L(s)$.

In order to efficiently handle the contribution to $\pi_C(x)$ which arises from the low-lying zeros of $\zeta_L(s)$, we factor $\zeta_L(s)$ as a product of Hecke L -functions associated to the Hecke characters of the abelian extension L/K and apply a log-free zero density estimate and the zero repulsion phenomenon for these L -functions. As in Linnik’s work on arithmetic progressions, one typically uses these tools to establish upper and lower bounds of $\pi_C(x)$ when x is small instead of asymptotic equalities [Thorner and Zaman 2017; 2018; Weiss 1983]. In order to facilitate the analysis involving the log-free zero density estimate, we weigh the contribution of each prime ideal counted by $\pi_C(x)$ with a weight whose Mellin transform has carefully chosen decay properties (Lemma 2.2). Similar variations are a critical component in the proofs of (1-6) in [Thorner and Zaman 2017; 2018; Weiss 1983].

By using a log-free zero density estimate and the zero repulsion phenomenon, we ensure that the main term in Theorem 1.4 *always* dominates the error term in Theorem 1.4 when x is at least a polynomial in

$D_K Q n_K^{n_K}$, regardless of whether β_1 exists. As one can see from the ensuing analysis, the quality of the zero-free region dictates the quality of the error term but has no direct impact on the valid range of x . This “decoupling” feature contrasts with the proof of (1-2), where the quality of the zero-free region simultaneously determines both the quality of the error term and the range of x in which the main term dominates.

After we “decouple” the range of x from the influence of the zero-free region, we are finally prepared to separate the contribution of the low-lying zeros from the contribution of the zeros with large height using a dyadic decomposition. This leads to savings over (1-2) only because we have already ensured via the log-free zero density estimate and zero repulsion that the main term in Theorem 1.4 dominates the error term regardless of whether β_1 exists. An additional benefit of this argument is an expression for the error term in Theorem 1.4 as a straightforward single-variable optimization problem involving x and the zero-free region (Lemma 4.5 and (4-13)). This simplification allows us to easily determine the error term with complete uniformity in D_K , $[K : \mathbb{Q}]$, Q , and x (Lemma 4.6).

The fact that Theorem 1.4 holds for *all* Galois extensions L/F is a fairly subtle matter. In the case where $F = \mathbb{Q}$ and L/\mathbb{Q} is a cyclotomic extension, the Chebotarev density theorem reduces to the prime number theorem for arithmetic progressions. Stark’s bound for β_1 (Theorem 3.3, a refinement of (1-4)) recovers a lower bound for $1 - \beta_1$ which is commensurate with the lower bound for $1 - \beta_1$ that follows from Dirichlet’s analytic class number formula for cyclotomic extensions; this suffices for our purposes. In the cyclotomic setting, our proofs only need to quantify the zero repulsion from a Landau–Siegel zero with a strong zero-free region for low-lying zeros (Theorem A.1 with $t \leq 4$). However, if L/F is a Galois extension where the root discriminant of L is especially small, which can happen in infinite class field towers, then Stark’s lower bound for $1 - \beta_1$ is quite small. In this case, the approach which worked well for cyclotomic extensions of \mathbb{Q} appears insufficient to prove Theorem 1.1 for all x in our claimed range.

To address this problem, we use a log-free zero density estimate for Hecke L -functions that naturally incorporates the zero repulsion phenomenon. Roughly speaking, when β_1 is especially close to 1, the quality of the log-free zero density estimate improves by a factor of $1 - \beta_1$; this is stronger than the classical formulation of the zero repulsion phenomenon. Therefore, if $1 - \beta_1$ happens to be as small as Stark’s lower bound allows, the quality of the log-free zero density estimate increases dramatically. This offsets the adverse effect of β_1 in the small root discriminant case. The idea of incorporating the zero repulsion phenomenon directly into the log-free zero density estimate goes back to Bombieri [1987] in the case of Dirichlet characters. For Hecke L -functions over number fields, this was first proved by Weiss (see Theorem 3.2 below). The details of this obstacle and why we genuinely need the particular log-free zero density estimate in Theorem 3.2 are contained in the Appendix, especially Remark A.3.

2. Setup and notation

Throughout the paper, let c_1, c_2, c_3, \dots be a sequence of absolute, effective, and positive constants. All implied constants in the inequalities $f \ll g$ and $f = O(g)$ are absolute and effective unless noted otherwise.

Recall F is a number field with ring of integers \mathcal{O}_F , absolute norm $N = N_{F/\mathbb{Q}}$, absolute discriminant $D_F = |\text{disc}(F/\mathbb{Q})|$, and degree $n_F = [F : \mathbb{Q}]$. Integral ideals will be denoted by \mathfrak{n} and prime ideals by \mathfrak{p} . Moreover, L/F is a Galois extension of number fields with Galois group $G = \text{Gal}(L/F)$. For prime ideals \mathfrak{p} of F unramified in L , the Artin symbol $\left[\frac{L/F}{\mathfrak{p}}\right]$ is the conjugacy class of Frobenius automorphisms of G associated to prime ideals \mathfrak{P} of L lying above \mathfrak{p} .

2A. Prime counting functions. For a conjugacy class C of G and $x \geq 2$, let $\pi_C(x)$ be as in (1-1) and define

$$\psi_C(x) = \psi_C(x, L/F) = \frac{|C|}{|G|} \sum_{\psi} \bar{\psi}(C) \frac{1}{2\pi i} \int_{2-i\infty}^{2+i\infty} -\frac{L'}{L}(s, \psi, L/F) \frac{x^s}{s} ds, \tag{2-1}$$

where ψ runs over the irreducible Artin characters of $G = \text{Gal}(L/F)$ and $L(s, \psi, L/F)$ is the Artin L -function of ψ . It follows from Mellin inversion [Lagarias et al. 1979, p.283] that

$$\psi_C(x) = \sum_{N\mathfrak{n} \leq x} \Lambda_F(\mathfrak{n}) \mathbf{1}_C(\mathfrak{n}), \tag{2-2}$$

where

$$\Lambda_F(\mathfrak{n}) = \begin{cases} \log N\mathfrak{p} & \text{if } \mathfrak{n} = \mathfrak{p}^j \text{ for some prime ideal } \mathfrak{p} \text{ and some integer } j \geq 1, \\ 0 & \text{otherwise.} \end{cases} \tag{2-3}$$

Here, $0 \leq \mathbf{1}_C(\mathfrak{n}) \leq 1$ for all ideals \mathfrak{n} and for prime ideals \mathfrak{p} unramified in L and $j \geq 1$,

$$\mathbf{1}_C(\mathfrak{p}^j) = \begin{cases} 1 & \text{if } \left[\frac{L/F}{\mathfrak{p}}\right]^j \subseteq C, \\ 0 & \text{otherwise.} \end{cases} \tag{2-4}$$

The prime counting functions π_C and ψ_C are related via partial summation.

Lemma 2.1. For $x \geq 2$,

$$\pi_C(x) = \frac{\psi_C(x)}{\log x} + \int_{\sqrt{x}}^x \frac{\psi_C(t)}{t(\log t)^2} dt + O\left(\log D_L + \frac{n_F x^{1/2}}{\log x}\right).$$

Proof. Note the norm of the product of ramified prime ideals divides D_L and the number of prime ideals \mathfrak{p} with norm equal to a given rational prime p is at most n_F . Thus,

$$\pi_C(x) = \sum_{\sqrt{x} < N\mathfrak{p} \leq x} \mathbf{1}_C(\mathfrak{p}) + O\left(\frac{n_F x^{1/2}}{\log x} + \log D_L\right).$$

Define $\theta_C(x) = \sum_{N\mathfrak{p} \leq x} \mathbf{1}_C(\mathfrak{p}) \log N\mathfrak{p}$. It follows by partial summation as well as the previous observations that

$$\sum_{\sqrt{x} < N\mathfrak{p} \leq x} \mathbf{1}_C(\mathfrak{p}) = \int_{\sqrt{x}}^x \frac{\theta_C(t)}{t(\log t)^2} dt + \frac{\theta_C(x)}{\log x}$$

Finally, one can verify that $|\theta_C(x) - \psi_C(x)| \ll n_F x^{1/2}$ by trivially estimating the number of prime ideal powers with norm at most x . Collecting all of these estimates yields the lemma. □

2B. Choice of weight. We now define a weight function which will be used to count prime ideals with norm between \sqrt{x} and x .

Lemma 2.2. *Choose $x \geq 3$, $\varepsilon \in (0, \frac{1}{4})$, and a positive integer $\ell \geq 1$. Define $A = \varepsilon/(2\ell \log x)$. There exists a continuous function $f(t) = f(t; x, \ell, \varepsilon)$ of a real variable t such that:*

- (i) $0 \leq f(t) \leq 1$ for all $t \in \mathbb{R}$, and $f(t) \equiv 1$ for $\frac{1}{2} \leq t \leq 1$.
- (ii) The support of f is contained in the interval $[\frac{1}{2} - \frac{\varepsilon}{\log x}, 1 + \frac{\varepsilon}{\log x}]$.
- (iii) Its Laplace transform $F(z) = \int_{\mathbb{R}} f(t)e^{-zt} dt$ is entire and is given by

$$F(z) = e^{-(1+2\ell A)z} \cdot \left(\frac{1 - e^{(1/2+2\ell A)z}}{-z} \right) \left(\frac{1 - e^{2Az}}{-2Az} \right)^\ell. \tag{2-5}$$

- (iv) Let $s = \sigma + it$, $\sigma > 0$, $t \in \mathbb{R}$ and α be any real number satisfying $0 \leq \alpha \leq \ell$. Then

$$|F(-s \log x)| \leq \frac{e^{\sigma\varepsilon} x^\sigma}{|s| \log x} \cdot (1 + x^{-\sigma/2}) \cdot \left(\frac{2\ell}{\varepsilon|s|} \right)^\alpha.$$

Moreover, $|F(-s \log x)| \leq e^{\sigma\varepsilon} x^\sigma$ and $\frac{1}{2} < F(0) < \frac{3}{4}$.

- (v) If $\frac{3}{4} < \sigma \leq 1$ and $x \geq 10$, then

$$F(-\log x) \pm F(-\sigma \log x) = \left(\frac{x}{\log x} \pm \frac{x^\sigma}{\sigma \log x} \right) \{1 + O(\varepsilon)\} + O\left(\frac{x^{1/2}}{\log x} \right). \tag{2-6}$$

- (vi) Let $s = -\frac{1}{2} + it$ with $t \in \mathbb{R}$. Then

$$|F(-s \log x)| \leq \frac{5x^{-1/4}}{\log x} \left(\frac{2\ell}{\varepsilon} \right)^\ell \left(\frac{1}{4} + t^2 \right)^{-\ell/2}.$$

Proof. These are the contents of [Thorner and Zaman 2018, Lemma 2.2] except for (2-6), which we now prove. Let $\frac{3}{4} < \sigma \leq 1$. From (iii), we observe that

$$F(-\sigma \log x) = \frac{x^\sigma}{\sigma \log x} \left(\frac{e^{\varepsilon\sigma/\ell} - 1}{\varepsilon\sigma/\ell} \right)^\ell + O\left(\frac{x^{\sigma/2}}{\sigma \log x} \right). \tag{2-7}$$

The two cases of $F(-\log x) \pm F(-\sigma \log x)$ are proved differently; we first handle the + case. It follows from (2-7) that

$$F(-\log x) + F(-\sigma \log x) = \frac{x}{\log x} \left(\frac{e^{\varepsilon/\ell} - 1}{\varepsilon/\ell} \right)^\ell + \frac{x^\sigma}{\sigma \log x} \left(\frac{e^{\varepsilon\sigma/\ell} - 1}{\varepsilon\sigma/\ell} \right)^\ell + O\left(\frac{x^{\sigma/2}}{\sigma \log x} \right).$$

The desired asymptotic for $F(-\log x) + F(-\sigma \log x)$ now follows from the Taylor series expansion

$$\left(\frac{e^{\varepsilon\sigma/\ell} - 1}{\varepsilon\sigma/\ell} \right)^\ell = 1 + O(\sigma\varepsilon),$$

which is valid for $0 < \sigma \leq 1$.

For the case of $F(-\log x) - F(-\sigma \log x)$, we first observe that (2-7) implies

$$(\log x)(F(-\log x) - F(-\sigma \log x)) = x \left(\frac{e^{\varepsilon/\ell} - 1}{\varepsilon/\ell} \right)^\ell - \frac{x^\sigma}{\sigma} \left(\frac{e^{\varepsilon\sigma/\ell} - 1}{\varepsilon\sigma/\ell} \right)^\ell + O(x^{1/2}). \tag{2-8}$$

Set

$$a = \frac{e^{\varepsilon/\ell} - 1}{\varepsilon/\ell}, \quad b = \frac{e^{\varepsilon\sigma/\ell} - 1}{\varepsilon\sigma/\ell}$$

so that $a > b \geq 1$. With this convention, we rewrite (2-8) as

$$(\log x)(F(-\log x) - F(-\sigma \log x)) = xa^\ell - \frac{x^\sigma}{\sigma}b^\ell + O(x^{1/2}). \tag{2-9}$$

Since $a > b \geq 1$, it follows from the bound $a^\ell - b^\ell \ll (a - b) \cdot \ell a^{\ell-1}$ that

$$xa^\ell - \frac{x^\sigma}{\sigma}b^\ell = \left(x - \frac{x^\sigma}{\sigma}\right)a^\ell + \frac{x^\sigma}{\sigma}(a^\ell - b^\ell) = \left(x - \frac{x^\sigma}{\sigma}\right)a^\ell + O\left(\frac{x^\sigma}{\sigma}(a - b)\ell a^{\ell-1}\right). \tag{2-10}$$

Since $\frac{3}{4} < \sigma \leq 1$, it follows from taking Taylor series expansions that $a^\ell = 1 + O(\varepsilon)$ and

$$a - b = \sum_{n=1}^{\infty} \frac{(1 - \sigma^n)(\varepsilon/\ell)^n}{(n + 1)!} \leq \sum_{n=1}^{\infty} \frac{n(1 - \sigma)(\varepsilon/\ell)^n}{(n + 1)!} \ll (1 - \sigma)\frac{\varepsilon}{\ell}.$$

We apply these two Taylor expansions to (2-9) and (2-10) to obtain

$$(\log x)(F(-\log x) - F(-\sigma \log x)) = \left(x - \frac{x^\sigma}{\sigma}\right)(1 + O(\varepsilon)) + O\left(\frac{x^\sigma}{\sigma}(1 - \sigma)\varepsilon\right) + O(x^{1/2}). \tag{2-11}$$

Finally, we observe that since $\sigma^{-2}x^\sigma \leq x$ for $\sigma > \frac{3}{4}$ and $x \geq 10$, we have that

$$\frac{x^\sigma}{\sigma}(1 - \sigma) = \sigma \left(\frac{x^\sigma}{\sigma^2} - \frac{x^\sigma}{\sigma} \right) \leq \sigma \left(x - \frac{x^\sigma}{\sigma} \right).$$

We apply this observation to (2-11) to obtain

$$(\log x)(F(-\log x) - F(-\sigma \log x)) = \left(x - \frac{x^\sigma}{\sigma}\right)(1 + O(\varepsilon)) + O(x^{1/2}). \tag{2-12}$$

The desired result follows by dividing both sides of (2-12) by $\log x$. □

Let $\ell \geq 2$ be an integer, $x \geq 3$, and $\varepsilon \in (0, \frac{1}{4})$. Define

$$\tilde{\psi}_C(x; f) = \tilde{\psi}_C(x, L/F; f) = \sum_n \Lambda_F(n) \mathbf{1}_C(n) f\left(\frac{\log Nn}{\log x}\right), \tag{2-13}$$

where $f = f(\cdot; x, \ell, \varepsilon)$ is given by Lemma 2.2. To understand ψ_C , it suffices to study the smooth variant $\tilde{\psi}_C$.

Lemma 2.3. *Let $\ell \geq 2$ be an integer, $x \geq 3$, and $\varepsilon \in (0, \frac{1}{4})$. Then*

$$\psi_C(x) \leq \tilde{\psi}_C(x; f) + O(n_F x^{1/2}) \leq \psi_C(xe^\varepsilon).$$

Moreover, $\tilde{\psi}_C(x; f) = \psi_C(x) + O(n_F x^{1/2} + \varepsilon x)$.

Proof. By Lemma 2.2(i,ii) and definitions (2-2) and (2-13), we observe that

$$\sum_{\sqrt{x} \leq \mathbf{Nn} \leq x} \Lambda_F(\mathbf{n}) \mathbf{1}_C(\mathbf{n}) \leq \tilde{\psi}_C(x; f) \leq \psi_C(xe^\varepsilon).$$

The lemma now follows from (2-2) and the trivial estimate

$$\sum_{z \leq \mathbf{Nn} \leq y} \Lambda_F(\mathbf{n}) \mathbf{1}_C(\mathbf{n}) \leq n_F \sum_{z \leq n \leq y} \Lambda_{\mathbb{Q}}(n) \ll n_F(y - z) \quad \text{for } 2 \leq z \leq y. \quad \square$$

2C. Dedekind zeta functions and Hecke L-functions. Now, assume L/K is an abelian extension of number fields. The Dedekind zeta function $\zeta_L(s)$ satisfies

$$\zeta_L(s) = \prod_{\chi} L(s, \chi, L/K), \tag{2-14}$$

where χ runs over the irreducible 1-dimensional Artin characters of $\text{Gal}(L/K)$. By class field theory, each Artin L -function $L(s, \chi, L/K)$ is equal to a Hecke L -function $L(s, \chi, K)$, where (abusing notation) χ is a certain primitive Hecke character of K . For simplicity, write $L(s, \chi)$ in place of $L(s, \chi, L/K)$ or $L(s, \chi, K)$. Let the integral $\mathfrak{f}_{\chi} \subseteq \mathcal{O}_K$ denote the conductor associated to χ . For each χ , there exist nonnegative integers $a(\chi)$ and $b(\chi)$ satisfying $a(\chi) + b(\chi) = n_K$ such that if we define

$$\gamma(s, \chi) = \left[\pi^{-\frac{s}{2}} \Gamma\left(\frac{s}{2}\right) \right]^{a(\chi)} \left[\pi^{-(s+1)/2} \Gamma\left(\frac{s+1}{2}\right) \right]^{b(\chi)}$$

and

$$\delta(\chi) = \begin{cases} 1 & \text{if } \chi \text{ is trivial,} \\ 0 & \text{otherwise,} \end{cases}$$

then $\xi(s, \chi) := [s(1-s)]^{\delta(\chi)} (D_K \mathfrak{N}\mathfrak{f}_{\chi})^{s/2} \gamma(s, \chi) L(s, \chi)$ satisfies the functional equation

$$\xi(s, \chi) = \varepsilon(\chi) \xi(1-s, \bar{\chi}), \tag{2-15}$$

where $\varepsilon(\chi)$ is a complex number with unit modulus. Furthermore, $\xi(s, \chi)$ is an entire function of order 1 which does not vanish at $s = 0$. Note $L(s, \chi)$ has a simple pole at $s = 1$ if and only if χ is trivial. The nontrivial zeros ρ of $L(s, \chi)$ (which are the zeros of $\xi(s, \chi)$) satisfy $0 < \text{Re}(\rho) < 1$, and the trivial zeros ω of $L(s, \chi)$ (which offset the poles of $\gamma(s, \chi)$) are at the nonnegative integers, each with order at most n_K .

The Dedekind zeta function $\zeta_L(s)$ possesses the same qualities (by considering the case $K = L$ and χ trivial). Namely, its completed L -function is

$$\xi_L(s) = [s(1-s)] D_L^{s/2} \left[\pi^{-s/2} \Gamma\left(\frac{s}{2}\right) \right]^{a_L} \left[(2\pi)^{-s} \Gamma\left(\frac{s+1}{2}\right) \right]^{b_L} \zeta_L(s) \tag{2-16}$$

for certain integers $a_L, b_L \geq 0$ satisfying $a_L + b_L = [L : \mathbb{Q}]$. The trivial zeros ω of $\zeta_L(s)$ are at the nonnegative integers with orders

$$\text{ord}_{s=\omega} \zeta_L(s) = \begin{cases} a_L & \omega = -2, -4, \dots, \\ b_L & \omega = -1, -3, \dots, \\ a_L - 1 & \omega = 0. \end{cases} \tag{2-17}$$

Moreover, the conductor-discriminant formula states that

$$\log D_L = \sum_{\chi} \log(D_K \text{Nf}_{\chi}). \tag{2-18}$$

From (1-11) with $\mathcal{Q} = \mathcal{Q}(L/K)$, it follows that

$$\log D_L \leq [L : K] \log(D_K \mathcal{Q}). \tag{2-19}$$

From this we deduce a somewhat crude bound for $\log D_L$ in terms of D_K, \mathcal{Q} , and n_K .

Lemma 2.4. *If L/K is abelian, then $\log D_L \ll (D_K \mathcal{Q} n_K^{n_K})^2$.*

Proof. By class field theory, L is contained in some ray class field L' of K whose Artin conductor has norm at most \mathcal{Q} . From [Weiss 1983, Lemma 1.16], it follows that $[L : K] \leq [L' : K] \leq D_K \mathcal{Q} e^{O(n_K)}$. The result now follows from (2-19). □

We also record a few standard estimates for Hecke L -functions.

Lemma 2.5 [Lagarias and Odlyzko 1977, Lemma 5.4]. *If $t \in \mathbb{R}$ and χ is a Hecke character of K , then*

$$\#\{\rho = \beta + i\gamma : L(\rho, \chi) = 0, 0 < \beta < 1, |\gamma - t| \leq 1\} \ll \log(D_K \text{Nf}_{\chi}) + n_K \log(|t| + 3),$$

where the zeros ρ are counted with multiplicity.

Lemma 2.6 [Lagarias and Odlyzko 1977, Lemma 5.6]. *Let χ be a Hecke character of K . Then*

$$-\frac{L'}{L}(s, \chi) \ll \log(D_K \text{Nf}_{\chi}) + n_K \log(|\text{Im}(s)| + 3)$$

uniformly for $\text{Re}(s) = -\frac{1}{2}$.

3. The distribution of zeros

For Sections 3 and 4, we will assume that the extension L/K is abelian. For notational simplicity, define

$$\mathcal{Q} = \mathcal{Q}(L/K) := D_K \mathcal{Q} n_K^{n_K}, \tag{3-1}$$

where $\mathcal{Q} = \mathcal{Q}(L/K)$ is given by (1-11). Any sum \sum_{χ} or product \prod_{χ} is over the primitive Hecke characters χ associated with L/K per the factorization in (2-14). Here we list three key results regarding the distribution of zeros of Hecke L -functions.

Theorem 3.1 (zero-free region). *There exists $c_4 > 0$ such that the Dedekind zeta function*

$$\zeta_L(s) = \prod_{\chi} L(s, \chi, L/K)$$

has at most one zero in the region $\operatorname{Re}(s) > 1 - \Delta(|\operatorname{Im}(s)| + 3)$, where the function Δ satisfies

$$\Delta(t) \geq \frac{c_4}{\log(Qt^{n_K})} \quad \text{for } t \geq 3. \quad (3-2)$$

If such an exceptional zero β_1 exists then it is real, simple, and attached to the L -function of a real Hecke character χ_1 .

Proof. This is well known; see, for example, [Weiss 1983, Theorem 1.9]. □

We also refer to the exceptional zero β_1 as a Landau–Siegel zero. Now, for $0 \leq \sigma \leq 1$, $T \geq 1$ and any Hecke character χ , define

$$N(\sigma, T, \chi) = \#\{\rho = \beta + i\gamma : L(\rho, \chi) = 0, \sigma < \beta < 1, |\gamma| \leq T\}, \quad (3-3)$$

where the zeros ρ are counted with multiplicity.

Theorem 3.2 (log-free zero density estimate). *There exists an integer $c_5 \geq 1$ such that*

$$\sum_{\chi} N(\sigma, T, \chi) \ll B_1(QT^{n_K})^{c_5(1-\sigma)} \quad (3-4)$$

uniformly for any $0 < \sigma < 1$ and $T \geq 1$, where

$$B_1 = B_1(T) = \min\{1, (1 - \beta_1) \log(QT^{n_K})\}. \quad (3-5)$$

Proof. Let $\varepsilon_0 > 0$ be a sufficiently small absolute and effective constant. It follows from [Thorner and Zaman 2017, Theorem 3.2] or its variant [Thorner and Zaman 2018, Theorem 4.5] that if $1 - \varepsilon_0 < \sigma < 1$ and $T \geq 1$, then

$$\sum_{\chi} N(\sigma, T, \chi) \ll (QT^{n_K})^{c_5(1-\sigma)}$$

regardless of whether β_1 exists. Weiss [1983, Theorem 4.3] proved that if β_1 exists, then for $1 - \varepsilon_0 < \sigma < 1$ and $T \geq 1$,

$$\sum_{\chi} N(\sigma, T, \chi) \ll (1 - \beta_1) \log(QT^{n_K}) (QT^{n_K})^{c_5(1-\sigma)}.$$

Thus for $T \geq 1$, (3-4) holds with B_1 given by (3-5) in the range $1 - \varepsilon_0 < \sigma < 1$. By enlarging c_5 if necessary and using Stark's bound from Theorem 3.3, one can extend (3-4) to the remaining interval $0 < \sigma < 1 - \varepsilon_0$ by employing the trivial bound that follows from Lemma 2.5. □

Theorems 3.1 and 3.2 comprise the three principles used to prove Linnik's theorem on the least prime in an arithmetic progression: a zero-free region, a log-free zero density estimate, and a quantitative form of the zero repulsion phenomenon. Theorem 3.2 combines the second and third principles by following

the ideas of Bombieri [1987], and this is crucial to our arguments for certain choices of Galois extensions (see the Appendix).

We record an effective lower bound for the size of $1 - \beta_1$ which follows from [Stark 1974, Theorem 1', p.148].

Theorem 3.3 (Stark’s bound). *Let $\beta_1 = 1 - \lambda_1/\log Q$ be a real zero of a real Hecke character χ of the abelian extension L/K . Then $\lambda_1 \gg Q^{-2}$.*

Proof. This follows readily from (1-4) for $1 - \beta$ when β is the real zero of a Dedekind zeta function. If χ is trivial then consider the Dedekind zeta function $\zeta_K(s)$. If χ is quadratic then consider the Dedekind zeta function $\zeta_K(s)L(s, \chi, L/K)$ corresponding to the quadratic extension of K defined by χ . □

As we shall see, these three theorems yield a unified Chebotarev density theorem which produces an asymptotic count for primes even in the presence of a Landau–Siegel zero.

4. Weighted counts of primes in abelian extensions

4A. Main technical result. The proof of Theorem 1.4 rests on the analysis on the weighted prime counting function $\tilde{\psi}_C(x; f) = \tilde{\psi}_C(x, L/K; f)$ given by (2-13), where f is given by Lemma 2.2 and L/K is abelian. The goal of this section is to prove the following proposition.

Proposition 4.1. *Assume L/K is abelian with Galois group G . Let $C \subseteq G$ be a conjugacy class of G . Let $f = f(\cdot; x, \ell, \varepsilon)$ be defined as in Lemma 2.2 with*

$$\varepsilon = 8\ell x^{-1/8\ell}, \quad \ell = 4c_5 n_K. \tag{4-1}$$

If $2 \leq Q \leq x^{1/(36c_5)}$ and $\varepsilon < \frac{1}{4}$, then

$$\frac{|G|}{|C|} \tilde{\psi}_C(x; f) = \left(x - \chi_1(C) \frac{x^{\beta_1}}{\beta_1} \right) \left(1 + O\left(e^{-\frac{c_4}{2} \frac{\log x}{\log Q}} + e^{-\sqrt{c_4(\log x)/4n_K}} \right) \right). \tag{4-2}$$

Remark 4.2. The constants c_4 and c_5 are defined in Theorems 3.1 and 3.2 respectively.

While f and its parameters are chosen in Proposition 4.1, we will assume throughout this section that $\varepsilon \in (0, \frac{1}{4})$ and $\ell \geq 2$ are arbitrary, unless otherwise specified. The arguments leading to Proposition 4.1 are divided into natural steps: shifting a contour, estimating the arising zeros with the log-free zero density estimate, and optimizing the error term with a classical zero-free region.

4B. Shifting the contour.

Lemma 4.3. *If $x \geq 3$, then*

$$\begin{aligned} & \frac{|G|}{|C|} \frac{\tilde{\psi}_C(x; f)}{\log x} \\ &= F(-\log x) - \chi_1(C)F(-\beta_1 \log x) - \sum_{\chi} \bar{\chi}(C) \sum_{\rho_{\chi}}^* F(-\rho_{\chi} \log x) + O\left(\frac{(2\ell/\varepsilon)^{\ell} \log D_L}{x^{1/4} \log x} + \frac{n_L}{\log x} \right), \end{aligned}$$

where the sum \sum^* is over all nontrivial zeros $\rho_\chi \neq \beta_1$ of $L(s, \chi)$, counted with multiplicity. Here the term $F(-\beta_1 \log x)$ may be omitted if the exceptional zero β_1 does not exist.

Proof. By (2-1), (2-13), Lemma 2.2 and a standard Mellin inversion calculation,

$$\frac{|G|}{|C|} \tilde{\psi}_C(x; f) = \sum_\chi \bar{\chi}(C) I_\chi, \quad \text{where } I_\chi = \frac{\log x}{2\pi i} \int_{2-i\infty}^{2+i\infty} -\frac{L'}{L}(s, \chi) F(-s \log x) ds. \quad (4-3)$$

For each Hecke character χ , shift the contour I_χ to the line $\text{Re}(s) = -\frac{1}{2}$. Note F is entire by Lemma 2.2(iii), so we need only consider the zeros and poles of $L(s, \chi)$. We pick up the simple pole at $s = 1$ of $L(s, \chi)$ when χ is trivial and the trivial zero at $s = 0$ of $L(s, \chi)$ of order at most n_K . Moreover, we also pick up all of the nontrivial zeros ρ_χ of $L(s, \chi)$. For the remaining contour along $\text{Re}(s) = -\frac{1}{2}$, we apply Lemma 2.6, Minkowski's estimate $n_K \ll \log D_K$, and Lemma 2.2(vi) to deduce that

$$-\frac{\log x}{2\pi i} \int_{-1/2-i\infty}^{-1/2+i\infty} \frac{L'}{L}(s, \chi, L/K) F(-s \log x) ds \ll \frac{(2\ell/\varepsilon)^\ell \log(D_K N\mathfrak{f}_\chi)}{x^{1/4}}.$$

Combining all of these observations yields

$$(\log x)^{-1} I_\chi = \delta(\chi) F(-\log x) - \sum_{\rho_\chi} F(-\rho_\chi \log x) + O\left(F(0)n_K + \frac{(2\ell/\varepsilon)^\ell \log(D_K N\mathfrak{f}_\chi)}{x^{1/4} \log x}\right). \quad (4-4)$$

Here, ρ_χ runs over all nontrivial zeros of $L(s, \chi)$, including β_1 if it exists. Substituting (4-4) into (4-3) and dividing through by $\log x$, we obtain the desired result but with an error term of

$$O\left(\frac{|F(0)|n_K}{\log x} \sum_\chi |\bar{\chi}(C)| + \frac{(2\ell/\varepsilon)^\ell}{x^{1/4} \log x} \sum_\chi |\bar{\chi}(C)| \log(D_K N\mathfrak{f}_\chi)\right).$$

As L/K is abelian, the characters χ are 1-dimensional so $|\bar{\chi}(C)| = 1$. Thus, applying the conductor-discriminant formula (2-18), the observation $n_K \sum_\chi 1 = [L : K]n_K = n_L$, and Lemma 2.2(iv), we obtain the desired error term. \square

4C. Estimating the zeros. Now we estimate the sum over nontrivial zeros ρ in Lemma 4.3, beginning with those ρ of small modulus.

Lemma 4.4. *If $x \geq 3$, then*

$$\sum_\chi \sum_{\substack{\rho_\chi \\ |\rho_\chi| \leq 1/4}} |F(-\rho_\chi \log x)| \ll x^{1/4} \log D_L.$$

Proof. From Lemmas 2.2(iv) and 2.5,

$$\sum_\chi \sum_{\substack{\rho_\chi \\ |\rho_\chi| \leq 1/4}} |F(-\rho_\chi \log x)| \ll \sum_\chi \sum_{\substack{\rho_\chi \\ |\rho_\chi| \leq 1/4}} x^{1/4} \ll x^{1/4} \sum_\chi (\log(D_K N\mathfrak{f}_\chi) + n_K).$$

The result now follows from Minkowski's estimate $n_K \ll \log D_K$ and (2-18). \square

Next, we use the log-free zero density estimate to analyze the remaining contribution.

Lemma 4.5. *Keep the assumptions and notation of Lemma 4.3. Select ε and ℓ as in (4-1) and assume $\varepsilon < \frac{1}{4}$. For $2 \leq Q \leq x^{1/(8c_5)}$,*

$$\log x \sum_x \sum_{\substack{\rho_x \\ |\rho_x| \geq 1/4}}^* |F(-\rho_x \log x)| \ll \nu_1 x e^{-\eta(x)/2}, \tag{4-5}$$

where

$$\nu_1 = \begin{cases} (1 - \beta_1) \log Q & \text{if } \beta_1 \text{ exists,} \\ 1 & \text{otherwise,} \end{cases} \tag{4-6}$$

and η is given by

$$\eta(x) = \inf_{t \geq 3} [\Delta(t) \log x + \log t]. \tag{4-7}$$

Proof. We dyadically estimate the zeros. For $j \geq 1$, set $T_0 = 0$ and $T_j = 2^{j-1}$ for $j \geq 1$. Consider the sum

$$Z_j := \frac{\log x}{x} \sum_x \sum_{\substack{\rho_x = \beta_x + i\gamma_x \\ T_{j-1} \leq |\gamma_x| \leq T_j \\ |\rho_x| \geq 1/4}} |F(-\rho_x \log x)| \tag{4-8}$$

for $j \geq 1$. First, we estimate the contribution of each zero $\rho = \rho_x$ appearing in Z_j . Let $\rho = \beta + i\gamma$ satisfy $T_{j-1} \leq |\gamma| \leq T_j$ and $|\rho| \geq \frac{1}{4}$, so $|\rho| \geq \max\{T_{j-1}, \frac{1}{4}\} \geq T_j/4$ and $|\rho| \gg |\gamma| + 3$. Thus, Lemma 2.2(iv) with $\alpha = \ell(1 - \beta)$ and our choice of ε imply that

$$\frac{\log x}{x} |F(-\rho \log x)| \ll \frac{x^{\beta-1}}{|\rho|} \left(\frac{2\ell}{\varepsilon|\rho|} \right)^{\ell(1-\beta)} \ll T_j^{-1/2} (|\gamma| + 3)^{-1/2} \cdot x^{-(1-\beta)/2} \cdot (x^{3/8} T_j^\ell)^{-(1-\beta)}.$$

Since $Q \leq x^{1/(8c_5)}$ and $\ell = 4c_5 n_K$, it follows that

$$\frac{\log x}{x} |F(-\rho \log x)| \ll T_j^{-1/2} \cdot (|\gamma| + 3)^{-1/2} x^{-(1-\beta)/2} (QT_j^{n_K})^{-2c_5(1-\beta)}. \tag{4-9}$$

From Theorem 3.1 and (4-7), we deduce

$$(|\gamma| + 3)^{-1/2} x^{-(1-\beta)/2} \leq (|\gamma| + 3)^{-1/2} x^{-\Delta(|\gamma|+3)/2} \leq e^{-\eta(x)/2}.$$

Note the right-hand side is uniform over all nontrivial zeros ρ appearing in (4-5). Combining (4-9) and the above inequality with (4-8), we deduce that

$$Z_j \ll e^{-\eta(x)/2} T_j^{-1/2} \sum_x \sum_{\substack{\rho_x = \beta_x + i\gamma_x \\ T_{j-1} \leq |\gamma_x| \leq T_j}} (QT_j^{n_K})^{-2c_5(1-\beta)}.$$

Defining $N(\sigma, T) = \sum_{\chi} N(\sigma, T, \chi)$, we use partial summation and Theorem 3.2 to see that

$$\begin{aligned} e^{\eta(x)/2} T_j^{1/2} Z_j &\ll \int_0^1 (QT_j^{n_K})^{-2c_5\alpha} dN(1-\alpha, T_j) \\ &\ll \left[(QT_j^{n_K})^{-2c_5} N(0, T_j) + \log(QT_j^{n_K}) \int_0^1 (QT_j^{n_K})^{-2c_5\alpha} N(1-\alpha, T_j) d\alpha \right] \\ &\ll B_1(T_j) \left[(QT_j^{n_K})^{-c_5} + \log(QT_j^{n_K}) \int_0^1 (QT_j^{n_K})^{-c_5\alpha} d\alpha \right] \\ &\ll B_1(T_j). \end{aligned}$$

If a Landau–Siegel zero does not exist then $B_1(T_j) = 1 = \nu_1$. Otherwise, if a Landau–Siegel zero exists then one can verify by (3-5) and a direct calculation that

$$B_1(T_j) T_j^{-1/4} \leq (1 - \beta_1) \cdot \sup_{t \geq 1} [\log(Qt^{n_K})t^{-1/4}] \ll (1 - \beta_1) \log Q = \nu_1.$$

The supremum occurs at $t \ll 1$ since $n_K \leq \log Q$. Therefore,

$$\sum_{j \geq 1} Z_j \ll e^{-\eta(x)/2} \sum_{j \geq 1} \frac{B_1(T_j)}{T_j^{1/4}} \cdot \frac{1}{T_j^{1/4}} \ll \nu_1 e^{-\eta(x)/2} \sum_{j \geq 1} 2^{-j/4} \ll \nu_1 e^{-\eta(x)/2},$$

which yields the lemma by definition (4-8). □

4D. Error term with a classical zero-free region. The quality of the error term in Lemma 4.5, and hence in Proposition 4.1, is reduced to computing $\eta(x)$. This is a single-variable optimization problem.

Lemma 4.6. *Let η be defined by (4-7). If $x \geq 2$ then $e^{-\eta(x)} \leq e^{-c_4 \frac{\log x}{\log Q}} + e^{-\sqrt{c_4(\log x)/n_K}}$.*

Proof. It follows from Theorem 3.1, (4-7), and a change of variables $t = e^u$ that

$$\eta(x) \geq \inf_{u \geq 0} \phi_x(u) \quad \text{where } \phi_x(u) = \frac{c_4 \log x}{\log Q + n_K u} + u.$$

Note that $\phi_x(u) \rightarrow \infty$ as $u \rightarrow \infty$. By standard calculus arguments, one can verify that

$$\eta(x) \geq \begin{cases} \frac{c_4 \log x}{\log Q} & \text{if } 2 \leq x \leq \exp\left(\frac{(\log Q)^2}{c_4 n_K}\right), \\ \sqrt{\frac{c_4 \log x}{n_K}} & \text{if } x \geq \exp\left(\frac{(\log Q)^2}{c_4 n_K}\right). \end{cases} \tag{4-10}$$

This proves the lemma. □

4E. Proof of Proposition 4.1. Choose ε and ℓ as in (4-1) and continue to assume $\varepsilon < \frac{1}{4}$. By Lemmas 4.3–4.5, it follows for $2 \leq Q \leq x^{1/(36c_5)}$ that

$$\frac{|G|}{|C|} \tilde{\psi}_C(x; f) = (\log x)[F(-\log x) - \chi_1(C)F(-\beta_1 \log x)] + O(\nu_1 x e^{-\eta(x)/2} + \mathcal{E}(x)),$$

where $\mathcal{E}(x) = x^{-1/4}(2\ell/\varepsilon)^\ell \log D_L + n_L + x^{1/4}(\log x)(\log D_L)$. From (4-1) and Minkowski’s estimate $n_L \ll \log D_L$, we see that $\mathcal{E}(x) \ll x^{1/4}(\log D_L)(\log x)$. From Lemma 2.4, $\log D_L \ll Q^2 \ll x^{1/10}$ since

$x \geq Q^{36c_5}$ and $c_5 \geq 1$. Hence, $\mathcal{E}(x) \ll x^{1/2}$. Using Lemma 2.2(v), (4-1), and noting $\beta_1 > \frac{1}{2}$, we deduce that

$$\frac{|G|}{|C|} \tilde{\psi}_C(x; f) = \left(x - \chi_1(C) \frac{x^{\beta_1}}{\beta_1} \right) (1 + O(n_K x^{-1/(32c_5 n_K)})) + O(v_1 x e^{-\eta(x)/2} + x^{1/2}) \tag{4-11}$$

for $2 \leq Q \leq x^{1/36c_5}$. Now, we claim that

$$x - \chi_1(C) \frac{x^{\beta_1}}{\beta_1} \gg v_1 x \gg x^{3/4}. \tag{4-12}$$

If β_1 does not exist, then $v_1 = 1$ and (4-12) is immediate. If β_1 exists and $(1 - \beta_1) \log x < 1$, then since $x \geq Q^{36c_5}$ and $e^{-t} \geq 1 - t$ for $0 < t < 1$, we have

$$x - \chi_1(C) \frac{x^{\beta_1}}{\beta_1} \geq x \left(1 - \frac{x^{-(1-\beta_1)}}{\beta_1} \right) \geq (1 - \beta_1) x \log \left(\frac{x}{e} \right) \gg (1 - \beta_1) x \log Q = v_1 x.$$

Otherwise, β_1 exists and $(1 - \beta_1) \log x \geq 1$ so $\beta_1 > \frac{1}{2}$ implies that

$$x - \chi_1(C) \frac{x^{\beta_1}}{\beta_1} \geq x \left(1 - \frac{x^{-(1-\beta_1)}}{\beta_1} \right) \geq x(1 - 2e^{-1}) \gg x \gg v_1 x,$$

Thus, the claim (4-12) follows upon noting that $v_1 \gg Q^{-2} \gg x^{-1/4}$ by Stark’s bound Theorem 3.3 and the condition $x \geq Q^{36c_5}$. Combining (4-12) with (4-11), it follows that

$$\frac{|G|}{|C|} \tilde{\psi}_C(x; f) = \left(x - \chi_1(C) \frac{x^{\beta_1}}{\beta_1} \right) (1 + O(e^{-\eta(x)/2} + n_K x^{-1/(32c_5 n_K)})). \tag{4-13}$$

Finally, we apply Lemma 4.6 and note $n_K x^{-1/(32c_5 n_K)} \ll x^{-1/(300c_5 n_K)} \ll e^{-\sqrt{c_4(\log x)/(4n_K)}}$ for $x \geq Q^{36c_5}$. This completes the proof of Proposition 4.1. □

5. Proof of Theorems 1.1 and 1.4

5A. Abelian extensions. First, we prove Theorem 1.4 in the case of abelian extensions.

Theorem 5.1. *Assume L/K is abelian with Galois group G . Let $C \subseteq G$ be a conjugacy class. Define Q by (3-1), for $2 \leq Q \leq x^{1/c_2}$,*

$$\pi_C(x, L/K) = \frac{|C|}{|G|} (\text{Li}(x) - \chi_1(C) \text{Li}(x^{\beta_1})) (1 + O(e^{-\frac{c_4}{4} \frac{\log x}{\log Q}} + e^{-\sqrt{c_4(\log x)/8n_K}})). \tag{5-1}$$

Here β_1 is a putative exceptional zero with associated real Hecke character χ_1 of L/K .

Proof. Write $g(x) = x - \chi_1(C) x^{\beta_1} / \beta_1$. Select ε as in (4-1). Note the assumption $2 \leq Q \leq x^{1/c_2}$ guarantees $\varepsilon < \frac{1}{4}$ provided c_2 is sufficiently large. From Proposition 4.1 and Lemma 2.3, it follows that

$$\psi_C(x) \leq \frac{|C|}{|G|} g(x) (1 + O(e^{-\frac{c_4}{2} \frac{\log x}{\log Q}} + e^{-\sqrt{c_4(\log x)/4n_K}})) \quad \text{for } x \geq Q^{36c_5}. \tag{5-2}$$

On the other hand, writing $y = xe^\varepsilon$, Proposition 4.1 and Lemma 2.3 also imply

$$\psi_C(y) \geq \frac{|C|}{|G|} g(ye^{-\varepsilon}) (1 + O(e^{-\frac{c_4}{2} \frac{\log y}{\log Q}} + e^{-\sqrt{c_4(\log y)/4n_K}}))$$

for $y \geq 2Q^{36c_5}$. By (4-12) and elementary arguments,

$$|g(ye^{-\varepsilon}) - g(y)e^{-\varepsilon}| \leq \frac{y^{\beta_1}}{\beta_1} (e^{-\varepsilon\beta_1} - e^{-\varepsilon}) \ll y\varepsilon(1 - \beta_1) \ll \varepsilon g(y).$$

In particular, $g(ye^{-\varepsilon}) = g(y)(1 + O(\varepsilon))$. From our choice of ε in (4-1) and the condition $y \geq 2Q^{36c_5}$, one can see that $\varepsilon \ll n_K y^{-1/32c_5 n_K} \ll y^{-1/300c_5 n_K} \ll e^{-\sqrt{c_4(\log y)/4n_K}}$ so

$$\psi_C(y) \geq \frac{|C|}{|G|} g(y) (1 + O(e^{-\frac{c_4}{2} \frac{\log y}{\log Q}} + e^{-\sqrt{c_4(\log y)/4n_K}})) \quad \text{for } y \geq 2Q^{36c_5}.$$

Comparing the above with (5-2), we conclude that

$$\psi_C(x) = \frac{|C|}{|G|} g(x) (1 + O(e^{-\frac{c_4}{2} \frac{\log x}{\log Q}} + e^{-\sqrt{c_4(\log x)/4n_K}}))$$

for $x \geq Q^{40c_5}$. By partial summation (Lemma 2.1) and the observation that, for $\frac{1}{2} < \sigma \leq 1$,

$$\frac{x^\sigma}{\sigma \log x} + \int_{\sqrt{x}}^x \frac{t^{\sigma-1}}{\sigma(\log t)^2} dt = \int_{x^{\sigma/2}}^{x^\sigma} \frac{1}{\log t} dt = \text{Li}(x^\sigma) + O\left(\frac{x^{1/2}}{\log x}\right), \tag{5-3}$$

it follows for $x \geq Q^{40c_5}$ that

$$\frac{|G|}{|C|} \pi_C(x) = (\text{Li}(x) - \chi_1(C) \text{Li}(x^{\beta_1})) (1 + O(e^{-\frac{c_4}{4} \frac{\log x}{\log Q}} + e^{-\sqrt{c_4(\log x)/8n_K}})) + \mathcal{E}_0(x),$$

where $\mathcal{E}_0(x) = \log D_L + n_K x^{1/2} / \log x$. By Lemma 2.4 and the observation that $n_K \ll \log x$, one can verify that $\mathcal{E}_0(x) \ll x^{1/2}$ for $x \geq Q^{40c_5}$. Hence, by (4-12), $\mathcal{E}_0(x)$ can be absorbed into the error term of Section 5A. As c_2 is sufficiently large, this completes the proof of Theorem 5.1. \square

5B. Proof of Theorem 1.4. Now we finish the proof of Theorem 1.4 for any Galois extension L/F with any Galois group G . Using well-known arguments from class field theory, we reduce to the case of abelian extensions.

Lemma 5.2 (Murty, Murty and Saradha). *Let L/F be a Galois extension of number fields with Galois group G , and let $C \subseteq G$ be a conjugacy class. Let H be a subgroup of G such that $C \cap H$ is nonempty, and let K be the fixed field of L by H . Let $g \in C \cap H$, and let $C_H(g)$ denote the conjugacy class of H which contains g . If $x \geq 2$, then*

$$\left| \pi_C(x, L/F) - \frac{|C|}{|G|} \frac{|H|}{|C_H|} \pi_{C_H}(x, L/K) \right| \leq \frac{|C|}{|G|} \left(n_L x^{1/2} + \frac{2}{\log 2} \log D_L \right).$$

Proof. This is carried out during the proof of [Murty et al. 1988, Proposition 3.9]. \square

Now, we apply Lemma 5.2 and subsequently Theorem 5.1 to $\pi_{C_H}(x, L/K)$ of the abelian extension L/K . Consequently, for $2 \leq Q \leq x^{1/c_2}$,

$$\frac{|G|}{|C|} \pi_C(x, L/F) = (\text{Li}(x) - \chi_1(C) \text{Li}(x^{\beta_1}))(1 + O(e^{-\frac{c_4}{4} \frac{\log x}{\log Q}} + e^{-\sqrt{c_4(\log x)/8n_K}})) + O(n_L x^{1/2} + \log D_L), \tag{5-4}$$

where $Q = Q(L/K)$ is defined by (3-1). Since we may assume $c_2 \geq 20$, it follows from Lemma 2.4 and Minkowski’s estimate $n_L \ll \log D_L$ that $n_L x^{1/2} + \log D_L \ll x^{5/8}$ for $x \geq Q^{c_2}$. From (4-12), this estimate may be absorbed into the first error term of (5-4) since $x^{5/8-3/4} = x^{-1/8} \ll e^{-\sqrt{c_4(\log x)/8n_K}}$. This completes the proof of Theorem 1.4. \square

Theorem 1.4 implies Theorem 1.1. Fix $g \in C$, let H in Theorem 1.4 be the cyclic group generated by g , and let K be the fixed field of H . Clearly $n_K \leq n_L$, and the centered equation immediately below [Thorner and Zaman 2017, Equation 1-7] states $D_L^{1/|H|} \leq D_K Q \leq D_L^{1/\varphi(|H|)}$. Theorem 1.1 now follows. \square

6. Reduced composition of beta-sieves

Before proceeding to the proof of Theorem 1.6, we require some sieve machinery that follows from standard results. The setup and discussion here closely follow [Friedlander and Iwaniec 2010, Sections 5.9 and 6.3–6.5]. Let $\lambda' = (\lambda'_d)$ and $\lambda'' = (\lambda''_d)$ be beta sieve weights with the same sifting level z and same level of distribution R . That is, λ'_d and λ''_d satisfy

$$\lambda'_1 = \lambda''_1 = 1, \quad |\lambda'_d| \leq 1, \quad |\lambda''_d| \leq 1,$$

and are supported on squarefree numbers $d < R$ consisting of prime factors $\leq z$. Let

$$s = \frac{\log R}{\log z}$$

be the sifting variable for both sieves. Let g' and g'' be multiplicative functions satisfying

$$0 \leq g'(p) < 1, \quad 0 \leq g''(p) < 1, \quad g'(p) + g''(p) < 1 \quad \text{for all primes } p. \tag{6-1}$$

Assume there exists $K > 1$ and $\kappa > 0$ such that, for all $2 \leq w \leq z$, we have

$$\prod_{w \leq p < z} \left(1 - \frac{g'(p)}{1 - g'(p) - g''(p)}\right)^{-1} \leq K \left(\frac{\log z}{\log w}\right)^\kappa, \tag{6-2}$$

$$\prod_{w \leq p < z} \left(1 - \frac{g''(p)}{1 - g'(p) - g''(p)}\right)^{-1} \leq K \left(\frac{\log z}{\log w}\right)^\kappa.$$

The goal of this section is to estimate the reduced composition given by

$$G := \sum_{(d_1, d_2)=1} \lambda'_{d_1} \lambda''_{d_2} g'(d_1) g''(d_2). \tag{6-3}$$

This expression can arise as the main term when two different sieves are applied to two different sequences that are linearly independent. Keeping this setup, the remainder of this section will be dedicated to the proof of the following theorem.

Theorem 6.1. *Assume $s > 9\kappa + 1 + 10 \log K$, (6-1) holds, and (6-2) holds. If \mathcal{N} and \mathcal{N}' are upper bound beta sieves, then*

$$\sum_{(d_1, d_2)=1} \lambda'_{d_1} \lambda''_{d_2} g'(d_1) g''(d_2) \leq \prod_p (1 - g'(p) - g''(p)) \{1 + e^{9-s} K^{10}\}^2.$$

If \mathcal{N} is a lower bound beta sieve and \mathcal{N}' is an upper bound beta sieve, then

$$\sum_{(d_1, d_2)=1} \lambda'_{d_1} \lambda''_{d_2} g'(d_1) g''(d_2) \geq \prod_p (1 - g'(p) - g''(p)) \{1 - e^{9-s} K^{10}\}.$$

Assume \mathcal{N} is a lower bound beta sieve and \mathcal{N}' is an upper bound beta sieve. The other case is entirely analogous. Thus, if $\theta' = 1 * \lambda'$ and $\theta'' = 1 * \lambda''$ then

$$\theta'_1 = \theta''_1 = 1 \quad \text{and} \quad \theta'_n \leq 0 \leq \theta''_n \quad \text{for } n \geq 2. \tag{6-4}$$

First, we apply [Friedlander and Iwaniec 2010, Lemma 5.6] to (6-3) and, keeping with their notation, we see that

$$G = \sum_{(b_1, b_2)=1} \theta'_{b_1} \theta''_{b_2} g'(b_1) g''(b_2) \prod_{p \nmid b_1 b_2} (1 - g'(p) - g''(p)). \tag{6-5}$$

Define \tilde{h}' , \tilde{h}'' and \tilde{g}' , \tilde{g}'' to be multiplicative functions supported on squarefree numbers with

$$\tilde{h}'(p) = \frac{g'(p)}{1 - g'(p) - g''(p)}, \quad \tilde{h}''(p) = \frac{g''(p)}{1 - g'(p) - g''(p)}, \quad \tilde{g}'(p) = \frac{g'(p)}{1 - g''(p)}, \quad \tilde{g}''(p) = \frac{g''(p)}{1 - g'(p)}.$$

Thus we obtain the usual relations

$$\tilde{h}'(p) = \frac{\tilde{g}'(p)}{1 - \tilde{g}'(p)} \quad \text{and} \quad \tilde{h}''(p) = \frac{\tilde{g}''(p)}{1 - \tilde{g}''(p)}. \tag{6-6}$$

Note $\tilde{h}'(p), \tilde{h}''(p) \geq 0$ and $0 \leq \tilde{g}'(p), \tilde{g}''(p) < 1$ by (6-1). Inserting these definitions into (6-5), we observe that

$$G = \left(\prod_p (1 - g'(p) - g''(p)) \right) \sum_{(b_1, b_2)=1} \theta'_{b_1} \theta''_{b_2} \tilde{h}'(b_1) \tilde{h}''(b_2).$$

If $(b_1, b_2) \neq 1$ then the expression $\theta'_{b_1} \theta''_{b_2} \tilde{h}'(b_1) \tilde{h}''(b_2)$ is nonpositive by (6-4), so we may introduce all of these terms at the cost of a lower bound for G . Thus

$$G \geq \left(\prod_p (1 - g'(p) - g''(p)) \right) \left(\sum_{b_1} \theta'_{b_1} \tilde{h}'(b_1) \right) \left(\sum_{b_2} \theta''_{b_2} \tilde{h}''(b_2) \right). \tag{6-7}$$

The two sums in (6-7) are prepared for standard beta-sieve analysis.

Lemma 6.2. *If \mathcal{N} is a lower bound beta-sieve with $\beta = 9\kappa + 1$ and $s \geq \beta$ then*

$$\sum_b \theta'_b \tilde{h}'(b) \geq 1 - e^{9\kappa-s} K^{10}.$$

If \mathcal{N}' is an upper bound beta-sieve with $\beta = 9\kappa + 1$ and $s \geq \beta$ then

$$\sum_b \theta''_b \tilde{h}''(b) \leq 1 + e^{9\kappa-s} K^{10}.$$

Proof. This statement is essentially the fundamental lemma [Friedlander and Iwaniec 2010, Lemma 6.8]. To make the comparison clear with [loc. cit., Sections 6.3–6.5], one begins with [loc. cit., Equation 6.40] with their D, h, g replaced by our $R, \tilde{h}', \tilde{g}'$ (or $R, \tilde{h}'', \tilde{g}''$, respectively). Per the definition of $V(z)$ on [loc. cit., p. 56], it follows that

$$V(z) = \prod_{p < z} (1 - \tilde{g}'(p)).$$

Thus the assumption [loc. cit., Equation 5.38] corresponds to our (6-2). Next, one defines V_n just as in the equation at the top of [loc. cit., p. 63]; in doing so, we obtain [loc. cit., Equations 6.43 and 6.44]. Finally, using the same truncation parameters, the analysis of [loc. cit., Section 6.5] leading up to [loc. cit., Lemma 6.8] yields our result. \square

Now, we apply Lemma 6.2 to the sum over b_1 (the lower bound sieve \mathcal{N}) in (6-7). Note that the assumption $s > 9\kappa + 1 + 10 \log K$ implies that this sum over b_1 is positive. By the positivity of \tilde{h} and (6-4), we may trivially estimate the sum over b_2 in (6-7) by

$$\sum_{b_2} \tilde{h}''(b_2) \theta''_{b_2} \geq \tilde{h}''(1) \theta''_1 = 1.$$

This proves the lower bound in Theorem 6.1. For the upper bound, we follow the same arguments and apply Lemma 6.2 twice (once to each sieve) in these final steps. \square

7. Restricted primes represented by binary quadratic forms

We recall the setup in Section 1C. Let

$$f(u, v) = au^2 + buv + cv^2 \in \mathbb{Z}[u, v]$$

be a positive definite binary quadratic form of discriminant $D = b^2 - 4ac < 0$, not necessarily fundamental. The group $\text{SL}_2(\mathbb{Z})$ naturally acts on such forms by $(T \cdot f)(\mathbf{x}) = f(T\mathbf{x})$ for $T \in \text{SL}_2(\mathbb{Z})$. The class number $h(D)$ is the number of such forms up to SL_2 -equivalence. We assume that f is primitive (that is, $(a, b, c) = 1$), and we define

$$\text{stab}(f) = \{T \in \text{SL}_2(\mathbb{Z}) : T \cdot f = f\}.$$

Note $|\text{stab}(f)| = 2$ unless $D = -3$ or -4 in which case it equals 6 and 4 respectively.

7A. Proof of Theorem 1.6. Recall by assumption that $3 \leq z \leq x^{\eta/\log \log x}$ where $\eta = \eta(A) > 0$ is sufficiently small. Further, P is any integer dividing the product of primes $\leq z$. Let $1 \leq R \leq x^{1/10}$ be a parameter yet to be specified. Let $\Lambda = (\lambda'_d)$ and $\Lambda'' = (\lambda''_d)$ be sieve weights supported on squarefree integers $d \mid P$ satisfying

$$\lambda'_1 = \lambda''_1 = 1, \quad |\lambda'_d| \leq 1, \quad |\lambda''_d| \leq 1 \quad \text{for } d \geq 1, \quad \lambda'_d = \lambda''_d = 0 \quad \text{for } d \geq R. \quad (7-1)$$

We approximate the condition $(uv, P) = 1$ in (1-13) by considering the sieved sum

$$S(x) = S(x; \Lambda, \Lambda'') := \frac{1}{|\text{stab}(f)|} \sum_{\substack{u, v \in \mathbb{Z} \\ f(u, v) \leq x}} \mathbf{1}_{\mathbb{P}}(f(u, v)) \left(\sum_{d_1 \mid u} \lambda'_{d_1} \right) \left(\sum_{d_2 \mid v} \lambda''_{d_2} \right). \quad (7-2)$$

By swapping the order of summation,

$$S(x) = \sum_{\substack{d_1, d_2 \\ (d_1, d_2) = 1}} \lambda'_{d_1} \lambda''_{d_2} A_{d_1, d_2}(x), \quad (7-3)$$

where

$$A_{d_1, d_2}(x) = \frac{1}{|\text{stab}(f)|} \sum_{\substack{f(u, v) \leq x \\ d_1 \mid u, d_2 \mid v}} \mathbf{1}_{\mathbb{P}}(f(u, v)). \quad (7-4)$$

Before computing the congruence sums $A_{d_1, d_2}(x)$, we introduce the local densities g' and g'' . These are multiplicative functions defined by

$$g'(p) = \begin{cases} \left(p - \left(\frac{D}{p}\right)\right)^{-1} & \text{if } p \mid P \text{ and } p \nmid c, \\ 0 & \text{otherwise,} \end{cases} \quad \text{and} \quad g''(p) = \begin{cases} \left(p - \left(\frac{D}{p}\right)\right)^{-1} & \text{if } p \mid P \text{ and } p \nmid a, \\ 0 & \text{otherwise.} \end{cases} \quad (7-5)$$

Here $\left(\frac{D}{p}\right)$ is the usual Legendre symbol for $p \neq 2$ and

$$\left(\frac{D}{2}\right) = \begin{cases} 0 & \text{if } 2 \mid D, \\ 1 & \text{if } D \equiv 1 \pmod{8}, \\ -1 & \text{if } D \equiv 5 \pmod{8}. \end{cases} \quad (7-6)$$

Our main result on the Chebotarev density theorem, Theorem 1.4, yields the following key lemma whose proof is postponed to Section 7B.

Lemma 7.1. *Let $\gamma > 0$ and $\vartheta > 0$ be a sufficiently small absolute constants, and let d_1, d_2 be relatively prime integers dividing P . If $|d_1 d_2 D| \leq x^\gamma$ then*

$$A_{d_1, d_2}(x) = g'(d_1) g''(d_2) \frac{\text{Li}(x) - \text{Li}(x^{\beta_1})}{h(D)} \{1 + O(\varepsilon_{d_1 d_2}(x))\} + O(\sqrt{x} \log x), \quad (7-7)$$

where β_1 is a simple real zero of the Dedekind zeta function $\zeta_{\mathbb{Q}(\sqrt{D})}(s)$ (if it exists) and

$$\varepsilon_d(x) = \varepsilon_d(x; D) = \exp\left[-\vartheta \frac{\log x}{\log |dD|}\right] + \exp[-(\vartheta \log x)^{1/2}] \quad \text{for } d \geq 1. \quad (7-8)$$

Remark 7.2. For the remainder of the proof of Theorem 1.6, the constant ϑ may be allowed to vary from line-to-line. This will occur finitely many times, so this is no cause for concern.

Remark 7.3. For the sieve to succeed, one crucially requires an asymptotic equality for $A_{d_1, d_2}(x)$ as in (7-7) with small remainder terms. Proceeding via the Chebotarev density theorem, one might use a stronger version of (1-2) in [Murty 1997] to obtain the asymptotic

$$A_{d_1, d_2}(x) = \frac{g'(d_1)g''(d_2)}{h(D)}(\text{Li}(x) + O(xe^{-c_1\sqrt{\log x}})), \quad \text{for } \log x \gg (\log|d_1d_2D|)^2 + \frac{1}{1 - \beta_1}. \quad (7-9)$$

Currently, $(1 - \beta_1)^{-1} \ll |D|^{1/2} \log|D|$ is the best unconditional effective bound for β_1 . Thus x must be quite large with respect to $|D|$, d_1 , and d_2 ; this adversely impacts the permissible ranges of $|D|$ and z in Theorem 1.6. To improve the range of x , one might instead appeal to variants of (1-6) found in [Thorner and Zaman 2017; 2018; Weiss 1983] but this only yields lower and upper bounds for $A_{d_1, d_2}(x)$, rendering the sieve powerless. Fortunately, Theorem 1.4 addresses all of these obstacles simultaneously. Regardless of whether β_1 exists, it maintains an asymptotic with an improved range of x that is polynomial in $|D|$, d_1 , and d_2 while keeping satisfactory control on the error terms. This allows us to strengthen the uniformity of both z and $|D|$ in Theorem 1.6 beyond what earlier versions of the Chebotarev density theorem permit.

Now, set the level of distribution to be

$$R := z^{\frac{1}{\sqrt{\eta}} \log \log z}. \quad (7-10)$$

Recall the constant $\eta = \eta(A) > 0$ should be thought of as very small. Since $z \leq x^{\eta/\log \log x}$ and $|D| \leq x^{\eta/\log \log z}$ by assumption, we have that $R \leq x^{1/10}$ and also $|d_1d_2D| \leq x^{4\sqrt{\eta}}$ for any integers $d_1, d_2 < R$. Thus, by Lemma 7.1 and (7-1), it follows that

$$S(x) = (\mathcal{G} + O(\mathcal{R})) \frac{\text{Li}(x) - \text{Li}(x^{\beta_1})}{h(D)} + O(x^{3/4}), \quad (7-11)$$

where

$$\mathcal{G} = \sum_{\substack{d_1, d_2 \\ (d_1, d_2)=1}} \lambda'_{d_1} \lambda''_{d_2} g'(d_1) g''(d_2), \quad \mathcal{R} = \sum_{\substack{d < R^2 \\ d|P}} \frac{\tau(d)}{\varphi(d)} \varepsilon_d(x).$$

Here τ is the divisor function and φ is Euler phi function. We obtained \mathcal{R} by observing that

$$\sum_{\substack{d_1, d_2 < R \\ d_1, d_2 | P \\ (d_1, d_2)=1}} |\lambda'_{d_1} \lambda''_{d_2}| g'(d_1) g''(d_2) \varepsilon_{d_1 d_2}(x) \leq \sum_{\substack{d < R^2 \\ d|P}} \varepsilon_d(x) \sum_{\substack{d_1 d_2 = d \\ (d_1, d_2)=1}} g'(d_1) g''(d_2) \leq \mathcal{R}$$

since $g'(d_1)g''(d_2) \leq 1/(\varphi(d_1)\varphi(d_2)) = 1/\varphi(d)$ from (7-5) and $\sum_{\substack{d_1 d_2 = d \\ (d_1, d_2)=1}} 1 \leq \tau(d)$. Now, we proceed to calculate the main term \mathcal{G} and remainder terms \mathcal{R} .

7A1. Main term \mathcal{G} . For the main term \mathcal{G} , suppose we have chosen a lower bound sieve for the sum in (1-13); namely, suppose \mathcal{A} is a lower bound beta sieve and \mathcal{A}' is an upper bound beta sieve, each with level of distribution R . Our aim is to apply the Fundamental Lemma in the form of Theorem 6.1. One can see that g' and g'' each satisfy (6-2) with $\kappa = 1$ and K absolutely bounded. Moreover, our choice of sieve has a sufficiently large sifting variable $s = \log R / \log z \gg \eta^{-1}$ because $\eta > 0$ is sufficiently small.

We claim that we may assume

$$g'(p) + g''(p) < 1 \quad \text{for all primes } p$$

and hence g' and g'' also satisfy (6-1). From (7-5), the only concern occurs when $p = 2$ and $2 \mid P$. We prove the claim by checking cases and verifying that $g'(2) + g''(2) \geq 1$ only if Theorem 1.6 is trivially true.

- Suppose $D \equiv 5 \pmod{8}$. By (7-5), we have $g'(2) + g''(2) \leq \frac{1}{3} + \frac{1}{3} < 1$.
- Suppose $D \equiv 1 \pmod{8}$ so $b \equiv 1 \pmod{2}$ and $ac \equiv 0 \pmod{2}$. If $a + b + c \equiv 0 \pmod{2}$ then the sum in (1-13) is necessarily empty because $\mathbf{1}_{\mathbb{P}}$ only detects odd primes. In this case, a and c have opposite parity so $g'(2) + g''(2) = 1$. Hence, $\delta_f(P) = 0$ by (1-14) and Theorem 1.6 is therefore trivially true. Otherwise, if $a + b + c \equiv 1 \pmod{2}$ then a and c have the same parity. As $ac \equiv 0 \pmod{2}$, it must be that $a \equiv c \equiv 0 \pmod{2}$ implying $g'(2) + g''(2) = 0 < 1$ by definition (7-5).
- Suppose $2 \mid D$ so $b \equiv 0 \pmod{2}$. If one of a or c is even then $g'(2) + g''(2) \leq \frac{1}{2} < 1$. Otherwise, if both a and c are odd then $g'(2) + g''(2) = 1$ and $a + b + c \equiv 0 \pmod{2}$. This implies $\delta_f(P) = 0$ and also the sum in (1-13) is necessarily empty so Theorem 1.6 is trivially true.

This proves the claim. Therefore, by Theorem 6.1 and (7-10), it follows that

$$\mathcal{G} \geq \delta_f(P) \{1 + O_A((\log z)^{-A})\} \tag{7-12}$$

since $\eta = \eta(A)$ is sufficiently small. If \mathcal{A} and \mathcal{A}' are both upper bound beta sieves with level of distribution $x^{1/10}$ then one similarly obtains the reverse inequality.

7A2. Remainder terms \mathcal{R} . We estimate \mathcal{R} dyadically. By the Cauchy–Schwarz inequality and standard estimates for τ and φ , we see for $0 \leq N \leq \lceil 2 \log R / \log z \rceil$ that

$$\begin{aligned} \sum_{\substack{z^N \leq d < z^{N+1} \\ d \mid P}} \frac{\tau(d)}{\varphi(d)} \varepsilon_d(x) &\ll \varepsilon_{z^{N+1}}(x) \left(\sum_{\substack{z^N \leq d < z^{N+1} \\ p \mid d \Rightarrow p \leq z}} \frac{1}{d} \right)^{1/2} \left(\sum_{z^N \leq d < z^{N+1}} \frac{\tau(d)^2 d}{\varphi(d)^2} \right)^{1/2} \\ &\ll \varepsilon_{z^{N+1}}(x) ((N + 1) \log z)^{3/2} \left(\sum_{\substack{z^N \leq d < z^{N+1} \\ p \mid d \Rightarrow p \leq z}} \frac{1}{d} \right)^{1/2}. \end{aligned}$$

By (7-10), one has that $R^{\eta' / \log \log R} \leq z \leq R$ where $\eta' > 0$ is sufficiently small depending only on η . In other words, $\log R / \log z \ll \log \log z$. Thus, we may apply Hildebrand’s estimate [1986, Theorem 1] for

z -smooth numbers via partial summation to conclude from (7-8) that the above is

$$\ll (e^{-\vartheta \frac{\log x}{(N+1)\log z}} + e^{-\vartheta \frac{\log x}{\log|D|}} + e^{-\vartheta \sqrt{\log x}}) \rho(N)(N+1)^2 \log^2 z,$$

where ρ is the Dickman–de Bruijn function. Recall we allow the constant $\vartheta > 0$ to change from line-to-line and be replaced by a smaller value if necessary. Summing this estimate over $0 \leq N \leq \lceil 2 \log R / \log z \rceil$ and using the crude estimate $\rho(N) \ll N^{-N}$ for $N \geq 1$, we deduce that

$$\begin{aligned} \mathcal{R} &\ll (\max_{N \geq 1} e^{-\frac{c \log x}{N \log z}} N^{-N+2}) \log^2 z + (e^{-\vartheta \frac{\log x}{\log z}} + e^{-\vartheta \frac{\log x}{\log|D|}} + e^{-\vartheta \sqrt{\log x}}) \log^2 z \\ &\ll (e^{-\vartheta \sqrt{\frac{\log x \log \log x}{\log z}}} + e^{-\vartheta \frac{\log x}{\log z}} + e^{-\vartheta \frac{\log x}{\log|D|}} + e^{-\vartheta \sqrt{\log x}}) \log^2 z. \end{aligned}$$

Since $|D| \leq x^{\eta/\log \log z}$ and $z \leq x^{\eta/\log \log x}$ with $\eta = \eta(A) > 0$ sufficiently small, we have that

$$\mathcal{R} \ll_A (\log z)^{-A}. \tag{7-13}$$

7A3. Concluding the proof. Inserting (7-12) and (7-13) into (7-11) along with the fact that $\delta_f(P) \gg (\log z)^{-2}$ from Mertens’ estimate, we conclude that

$$\sum_{\substack{u, v \in \mathbb{Z} \\ au^2 + buv + cv^2 \leq x \\ (uv, P) = 1}} \frac{\mathbf{1}_{\mathbb{P}}(au^2 + buv + cv^2)}{|\text{stab}(f)|} \geq \delta_f(P) \frac{\text{Li}(x) - \text{Li}(x^{\beta_1})}{h(D)} \{1 + O_A((\log z)^{-A})\} + O(x^{3/4}).$$

By using an upper bound sieve instead (as mentioned at the end of Section 7A1), one also obtains the reverse inequality. Thus, it remains to show the secondary error term $O(x^{3/4})$ may be absorbed into the primary error term. If $\delta_f(P) = 0$ then the arguments in Section 7A1 imply Theorem 1.6 trivially true so we may assume $\delta_f(P) > 0$. By the effective lower bound that $1 - \beta_1 \gg_\varepsilon |D|^{-1/2-\varepsilon}$, the fact that $h(D) \ll_\varepsilon |D|^{1/2+\varepsilon}$, and the assumption that $|D| \leq x^{\eta/\log \log z}$, we see

$$\frac{\text{Li}(x) - \text{Li}(x^{\beta_1})}{h(D)} \gg x^{4/5}.$$

As $\delta_f(P) \gg (\log z)^{-2}$, this implies the claim and hence proves Theorem 1.6. □

7B. Proof of Lemma 7.1. The pair (d_1, d_2) induces another form f_{d_1, d_2} given by

$$f_{d_1, d_2}(s, t) := f(d_1 s, d_2 t).$$

Note its discriminant is $D(d_1 d_2)^2$. With this definition, it follows that

$$A_{d_1, d_2}(x) = \frac{1}{|\text{stab}(f)|} \sum_{p \leq x} \#\{(s, t) \in \mathbb{Z}^2 : p = f_{d_1, d_2}(s, t)\}. \tag{7-14}$$

Observe

$$A_{d_1, d_2}(x) \ll 1 \quad \text{if } (d_1, c) \neq 1 \text{ or } (d_2, a) \neq 1$$

since, in this case, f_{d_1, d_2} is not primitive and hence represents an absolutely bounded number of primes. This trivially establishes Lemma 7.1 in this case. To evaluate $A_{d_1, d_2}(x)$ for all other d_1 and d_2 , we use class field theory.

Lemma 7.4. *Let \mathcal{O}_K be the ring of integers of $K = \mathbb{Q}(\sqrt{D})$. For $d \geq 1$, let \mathcal{O}_d be the order of discriminant $-Dd^2$ in K and let L_d be the ring class field of \mathcal{O}_d . If F is a primitive binary quadratic form of discriminant $-Dd^2$ then*

$$|\mathcal{O}_d^\times| = |\text{stab}(F)|.$$

Moreover, if C_F is the conjugacy class corresponding to F in the Galois group of L_d/K then

$$\#\{(s, t) \in \mathbb{Z}^2 : p = F(s, t)\} = |\mathcal{O}_d^\times| \cdot \#\{\mathfrak{p} \subseteq \mathcal{O}_K : \mathbf{N}\mathfrak{p} = p, \left[\frac{L_d/K}{\mathfrak{p}}\right] = C_F\} \quad \text{for } p \nmid Dd.$$

Here $\left[\frac{L_d/K}{\mathfrak{p}}\right]$ is the Artin symbol of \mathfrak{p} and $\mathbf{N} = \mathbf{N}_{K/\mathbb{Q}}$ is the absolute norm of K/\mathbb{Q} .

Proof. These are straightforward consequences of the theory for positive definite binary quadratic forms, so we only sketch the details. Standard references include for example [Cassels 1978; Cox 1989]. First, one can verify that $\mathcal{O}_d^\times = \{\pm 1\}$ unless \mathcal{O}_d is the ring of integers for $\mathbb{Q}(i)$ or $\mathbb{Q}(\sqrt{-3})$. Similarly, the SL_2 -automorphism group of F is $\left\{\pm \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}\right\}$ unless F is properly equivalent to either $x^2 + y^2$ or $x^2 + xy + y^2$. These are respectively the unique reduced forms of discriminant -4 or -3 . These remaining two cases can be checked by direct calculation.

The second claim follows from the first claim and the one-to-one correspondence between inequivalent representations of a prime p by F and degree 1 prime ideals $\mathfrak{p} \subseteq \mathcal{O}_K$ in the class C_F . For more details, see [Cox 1989, Theorem 7.7]. □

Now, assuming $(d_1, c) = (d_2, a) = 1$, we return to computing $A_{d_1, d_2}(x)$. It follows that f_{d_1, d_2} is primitive so by Lemma 7.4 with $F = f_{d_1, d_2}$ and $d = d_1 d_2$, we deduce that

$$A_{d_1, d_2}(x) = \frac{1}{|\text{stab}(f)|} \sum_{\substack{\mathfrak{N}\mathfrak{p} \leq x \\ \text{deg}(\mathfrak{p})=1}}^\dagger |\mathcal{O}_{d_1 d_2}^\times| + O\left(\sum_{p \mid Dd_1 d_2} 1\right), \tag{7-15}$$

where \sum^\dagger runs over prime ideals \mathfrak{p} in \mathcal{O}_K unramified in $L_{d_1 d_2}$ satisfying $[(L_{d_1 d_2}/K)/\mathfrak{p}] = C_{f_{d_1, d_2}}$. Note, for the primes $p \mid Dd_1 d_2$ in (7-15), we have used that each prime p is represented by f with absolutely bounded multiplicity. We may add the remaining degree 2 prime ideals \mathfrak{p} to the \dagger -marked sum with error at most $O(|\mathcal{O}_{d_1 d_2}^\times| \sqrt{x} \log x) = O(\sqrt{x} \log x)$. Further, we have

$$\sum_{p \mid Dd_1 d_2} 1 \ll \log |Dd_1 d_2| \ll \log x$$

since $|d_1 d_2 D| \leq x^\gamma$. Collecting these observations, it follows that

$$A_{d_1, d_2}(x) = \frac{|\mathcal{O}_{d_1 d_2}^\times|}{|\text{stab}(f)|} \sum_{\mathfrak{N}\mathfrak{p} \leq x}^\dagger 1 + O(\sqrt{x} \log x). \tag{7-16}$$

We invoke Theorem 1.4 to compute the sum in (7-16), thus

$$\sum_{Np \leq x}^{\dagger} 1 = \frac{\text{Li}(x) - \theta_1 \text{Li}(x^{\beta_1})}{h(D(d_1 d_2)^2)} \{1 + O(\varepsilon_{d_1 d_2}(x))\} \quad \text{for } |d_1 d_2 D| \leq x^\gamma, \tag{7-17}$$

where $\varepsilon_{d_1 d_2}(x)$ is defined by (7-8) and $\gamma > 0$ is fixed and sufficiently small. We make two simplifications for (7-17). First, we claim that $\theta_1 = 1$ if the exceptional zero β_1 exists. By a theorem of Heilbronn [1972] generalized by Stark [1974, Theorem 3], since β_1 is a real simple zero of $\zeta_{L_{d_1 d_2}}(s)$ and $L_{d_1 d_2}$ is Galois over \mathbb{Q} with K being its only quadratic subfield, it follows that $\zeta_K(\beta_1) = 0$. Hence, the exceptional Hecke character χ_1 of K from Theorem 1.4 is trivial implying $\theta_1 = 1$. Second, we have for $d \geq 1$ that

$$h(Dd^2) = \frac{h(D)}{[\mathcal{O}^\times : \mathcal{O}_d^\times]} d \prod_{p|d} \left(1 - \left(\frac{D}{p}\right) \frac{1}{p}\right). \tag{7-18}$$

For a proof, see for example [Cox 1989, Theorem 7.4 and Corollary 7.28].

Finally, with these observations, Lemma 7.1 follows by inserting (7-17) and (7-18) into (7-16) and noting that $[\mathcal{O}_1^\times : \mathcal{O}_d^\times] \cdot |\mathcal{O}_d^\times| = |\mathcal{O}_1^\times| = |\text{stab}(f)|$ from Lemma 7.4. □

Appendix: Error term with an exceptional zero

Theorem 3.2 states that if $T \geq 1$, then

$$\sum_{\chi} N(\sigma, T, \chi) \ll B_1 (QT^{n_K})^{c_5(1-\sigma)}, \quad B_1 = \min\{1, (1 - \beta_1) \log(QT^{n_K})\}. \tag{A-1}$$

This clearly implies that regardless of whether β_1 exists, we have

$$\sum_{\chi} N(\sigma, T, \chi) \ll (QT^{n_K})^{c_5(1-\sigma)}. \tag{A-2}$$

If β_1 exists, Theorem 3.2 produces the following strong zero-free region:

Theorem A.1 (zero repulsion). *Suppose the exceptional zero β_1 of Theorem 3.1 exists. There exists $c_6 > 0$ such that if Δ is given in Theorem 3.1, then*

$$\Delta(t) \geq \min \left\{ \frac{1}{2}, \frac{c_6 \log([(1 - \beta_1) \log(Qt^{n_K})]^{-1})}{\log(Qt^{n_K})} \right\}.$$

Let $q \geq 1$ be an integer. In the context of arithmetic progressions, in which case $L = \mathbb{Q}(e^{2\pi i/q})$ and $F = K = \mathbb{Q}$, it is preferable to use (A-2) and Theorem A.1 instead of (A-1), as one can typically obtain numerically superior results with the former. However, in the context of arithmetic progressions, one has the benefit of working with characters of an extension which is abelian over \mathbb{Q} , in which case Theorem 3.3 gives an adequate upper bound for β_1 (should it exist). However, for abelian extensions L/K where the root discriminant of K is rather small, Theorem 3.3 gives an upper bound for β_1 which is not commensurate with the corresponding result for cyclotomic extensions of \mathbb{Q} . In fact, this weak upper bound leads us to actually require a version of the log-free zero density estimate that improves as

β_1 approaches 1 to handle the case when K has a small root discriminant. This is why we use (A-1) in our proofs instead of using (A-2) and Theorem A.1 separately.

For comparison with Lemma 4.6, we quantify the effect of (A-2) and Theorem A.1 on the error term in Lemma 4.5 and subsequently (4-13) in the proof of Proposition 4.1. Since the calculations are tedious, we omit the proof.

Lemma A.2. *Let η be defined by (4-7). Suppose the exceptional zero $\beta_1 = 1 - \lambda_1/\log Q$ of Theorem 3.1 exists. There exists absolute constants $c_7, c_8, c_9 > 0$ such that if $\lambda_1 \leq c_7$ and $Q \leq x^{1/c_9}$,*

$$e^{-\eta(x)} \ll x^{-1/2} + \lambda_1^{10} (e^{-\frac{c_6 \log x}{2 \log Q}} + e^{-c_8 \sqrt{(\log x)/n_K}}) \quad \text{if } \lambda_1 \geq Q^{-20/n_K}, \quad (\text{A-3})$$

$$e^{-\eta(x)} \ll x^{-1/2} + e^{-10 \sqrt{\log(1/\lambda_1)}} (e^{-\frac{c_6 \log x}{2 \log Q}} + e^{-c_8 \sqrt{(\log x)/n_K}}) \quad \text{if } \lambda_1 < Q^{-20/n_K}. \quad (\text{A-4})$$

Remark A.3. Recall the definition of v_1 in (4-6). From (4-11) and (4-12), one can see it is critical to prove an estimate at least as strong as

$$v_1 x e^{-\eta(x)} = o(\lambda_1 x). \quad (\text{A-5})$$

Notice that the density estimate in (A-1) decays linearly with respect to $1 - \beta_1$ (that is, $v_1 = \lambda_1$), so we easily obtain (A-5). Suppose we instead use (A-2), which is tantamount to the trivial estimate $v_1 \leq 1$ when β_1 exists. From (A-3), one obtains (A-5) when $\lambda_1 \geq Q^{-20/n_K}$. Otherwise, from (A-4), if $\lambda_1 < Q^{-20/n_K}$ then we can at best show $x e^{-\eta(x)} = o(e^{-10 \sqrt{\log(1/\lambda_1)}} x)$. The situation $\lambda_1 < Q^{-20/n_K}$ is not uniformly excluded by Stark’s bound (1-4). For example, when the root discriminant D_K^{1/n_K} is bounded and the extension L/K is unramified (that is, $Q = 1$), then

$$Q^{100/n_K} = (D_K Q)^{100/n_K} n_K^{100} \ll n_K^{100}$$

and Stark’s bound (1-4) implies $\lambda_1^{-1} \ll n_K^{n_K} \log D_K$ so it may very well be the case that $\lambda_1^{-1} \gg n_K^{100} \gg Q^{100/n_K}$. This situation with a bounded root discriminant is entirely possible as Minkowski’s unconditional estimate $n_K \ll \log D_K$ is tight when varying over all number fields K . Infinite class field towers are well known sources of this scenario. Thus, we cannot see how to unconditionally obtain the desired linear decay demanded by (A-5) with only (A-2) and Theorem A.1.

Acknowledgements

We thank Kannan Soundararajan for helpful discussions and the anonymous referee for providing very thorough comments on our initial submission. Jesse Thorner is partially supported by a NSF Mathematical Sciences Postdoctoral Fellowship, and Asif Zaman is partially supported by a NSERC Postdoctoral Fellowship.

References

[Bombieri 1987] E. Bombieri, “Le grand crible dans la théorie analytique des nombres”, *Astérisque* 18 (1987), 103. MR Zbl

- [Cassels 1978] J. W. S. Cassels, *Rational quadratic forms*, London Mathematical Society Monographs **13**, Academic Press, London, 1978. MR Zbl
- [Cox 1989] D. A. Cox, *Primes of the form $x^2 + ny^2$* , Wiley, New York, 1989. MR Zbl
- [Fouvry and Iwaniec 1997] E. Fouvry and H. Iwaniec, “Gaussian primes”, *Acta Arith.* **79**:3 (1997), 249–287. MR Zbl
- [Friedlander and Iwaniec 2010] J. Friedlander and H. Iwaniec, *Opera de cribro*, American Mathematical Society Colloquium Publications **57**, American Mathematical Society, Providence, RI, 2010. MR Zbl
- [Heilbronn 1972] H. Heilbronn, “On real simple zeros of Dedekind ζ -functions”, pp. 108–110 in *Proceedings of the Number Theory Conference* (Univ. Colorado, Boulder, Colo., 1972), Univ. Colorado, Boulder, Colo., 1972. MR Zbl
- [Hildebrand 1986] A. Hildebrand, “On the number of positive integers $\leq x$ and free of prime factors $> y$ ”, *J. Number Theory* **22**:3 (1986), 289–307. MR Zbl
- [Lagarias and Odlyzko 1977] J. C. Lagarias and A. M. Odlyzko, “Effective versions of the Chebotarev density theorem”, pp. 409–464 in *Algebraic number fields: L -functions and Galois properties* (Durham, UK, 1975), edited by A. Fröhlich, Academic Press, London, 1977. MR Zbl
- [Lagarias et al. 1979] J. C. Lagarias, H. L. Montgomery, and A. M. Odlyzko, “A bound for the least prime ideal in the Chebotarev density theorem”, *Invent. Math.* **54**:3 (1979), 271–296. MR Zbl
- [Lam et al. 2018] C. Lam, D. Schindler, and S. Xiao, “On prime values of binary quadratic forms with a thin variable”, preprint, 2018. arXiv
- [Linnik 1944] U. V. Linnik, “On the least prime in an arithmetic progression, I: The basic theorem”, *Rec. Math. [Mat. Sbornik] N.S.* **15**(57) (1944), 139–178. MR Zbl
- [Murty 1997] V. K. Murty, “Modular forms and the Chebotarev density theorem, II”, pp. 287–308 in *Analytic number theory* (Kyoto, 1996), edited by Y. Motohashi, London Math. Soc. Lecture Note Ser. **247**, Cambridge Univ. Press, 1997. MR Zbl
- [Murty et al. 1988] M. R. Murty, V. K. Murty, and N. Saradha, “Modular forms and the Chebotarev density theorem”, *Amer. J. Math.* **110**:2 (1988), 253–281. MR Zbl
- [Stark 1974] H. M. Stark, “Some effective cases of the Brauer–Siegel theorem”, *Invent. Math.* **23** (1974), 135–152. MR Zbl
- [Thorner and Zaman 2017] J. Thorner and A. Zaman, “An explicit bound for the least prime ideal in the Chebotarev density theorem”, *Algebra Number Theory* **11**:5 (2017), 1135–1197. MR Zbl
- [Thorner and Zaman 2018] J. Thorner and A. Zaman, “A Chebotarev variant of the Brun–Titchmarsh theorem and bounds for the Lang–Trotter conjectures”, *Int. Math. Res. Not.* **2018**:16 (2018), 4991–5027. MR Zbl
- [Weiss 1983] A. Weiss, “The least prime ideal”, *J. Reine Angew. Math.* **338** (1983), 56–94. MR Zbl
- [Zaman 2017] A. A. Zaman, *Analytic estimates for the Chebotarev density theorem and their applications*, Ph.D. thesis, University of Toronto, 2017, Available at <https://search.proquest.com/docview/1993455319>. MR

Communicated by Roger Heath-Brown

Received 2018-03-22 Revised 2018-11-29 Accepted 2019-01-30

jthorner@stanford.edu

Department of Mathematics, Stanford University, Stanford, CA, United States

aazaman@stanford.edu

Department of Mathematics, Stanford University, Stanford, CA, United States

On the Brauer–Siegel ratio for abelian varieties over function fields

Douglas Ulmer

Hindry has proposed an analog of the classical Brauer–Siegel theorem for abelian varieties over global fields. Roughly speaking, it says that the product of the regulator of the Mordell–Weil group and the order of the Tate–Shafarevich group should have size comparable to the exponential differential height. Hindry–Pacheco and Griffon have proved this for certain families of elliptic curves over function fields using analytic techniques. Our goal in this work is to prove similar results by more algebraic arguments, namely by a direct approach to the Tate–Shafarevich group and the regulator. We recover the results of Hindry–Pacheco and Griffon and extend them to new families, including families of higher-dimensional abelian varieties.

1. Introduction

The classical Brauer–Siegel theorem [Brauer 1950] says that if K runs through a sequence of Galois extensions of \mathbb{Q} with discriminants $d = d_K$ satisfying $[K : \mathbb{Q}] / \log d \rightarrow 0$, then

$$\frac{\log(Rh)}{\log \sqrt{d}} \rightarrow 1$$

where $R = R_K$ and $h = h_K$ are the regulator and class number of K . The proof uses the class number formula

$$\operatorname{Res}_{s=1} \zeta_K(s) = \frac{2^{r_1} (2\pi)^{r_2} Rh}{w\sqrt{d}}$$

and analytic methods.

Hindry [2007] conjectured an analog of the Brauer–Siegel theorem for abelian varieties. If A is an abelian variety over a global field K with regulator R , Tate–Shafarevich group III (assumed to be finite), and exponential differential height H (definitions below), Hindry proposed that the Brauer–Siegel ratio

$$\text{BS}(A) := \frac{\log(R|\text{III}|)}{\log(H)}$$

should tend to 1 for any sequence of abelian varieties over a fixed K with $H \rightarrow \infty$.

MSC2010: primary 11G05; secondary 11G10, 11G40.

Keywords: abelian variety, Tate–Shafarevich group, regulator, height, Brauer–Siegel ratio, function field.

Hindry and Pacheco [2016] considered the case where K is a global function field of characteristic $p > 0$. Assuming the finiteness of III, they proved (Corollary 1.13) that

$$0 \leq \liminf_A \text{BS}(A) \leq \limsup_A \text{BS}(A) = 1, \quad (1.1)$$

where the limits are over the family of all nonconstant abelian varieties of a fixed dimension over K ordered by height. Note that this leaves open the possibility of a sequence of abelian varieties with Brauer–Siegel ratio tending to a limit < 1 , a possibility not envisioned in Hindry’s earlier paper. Hindry and Pacheco also conjectured and gave evidence for the claim that the lower bound $0 \leq \liminf_A \text{BS}(A)$ should be an equality when A runs through the family of quadratic twists of a fixed elliptic curve. Moreover, they gave an example (Theorem 1.4) of a family of elliptic curves E with $H \rightarrow \infty$ and proved $\lim_E \text{BS}(E) = 1$ without having to assume any unproven conjectures. In his Paris VII thesis, Griffon [2016] gave several other examples of families of elliptic curves where $\lim_E \text{BS}(E) = 1$ again without assuming unproven conjectures.

As with the original Brauer–Siegel theorem, the analyses of Hindry–Pacheco and Griffon use analytic techniques. More precisely, finiteness of the Tate–Shafarevich group implies the conjecture of Birch and Swinnerton-Dyer (in its strong form), and so a class number formula of the shape

$$L^*(A) = \alpha \frac{|\text{III}|R}{H}$$

where $L^*(A)$ is the leading Taylor coefficient of the L -function at $s = 1$ and α is a relatively innocuous, nonzero factor. (We will give the precise statement below.) Hindry–Pacheco and Griffon then prove their results by estimating (and in some cases calculating quite explicitly) $L^*(A)$.

Our goal in this work is to prove several results about Brauer–Siegel ratios by more algebraic arguments, in other words through a direct approach to the Tate–Shafarevich group and the regulator. More precisely, we prove the following results without recourse to L -functions:

- (1) a transparent and conceptual proof that $\liminf_A \text{BS}(A) \geq 0$ via a lower bound on the regulator;
- (2) a new connection between the growth of $|\text{III}|$ as the finite ground field is extended and the number $R|\text{III}|$ over the given field;
- (3) a general calculation of the limiting Brauer–Siegel ratio for the sequence $E^{(p^n)}$ of Frobenius pull-backs of an elliptic curve E ;
- (4) a new proof that $\lim_d \text{BS}(E_d) = 1$ in the families of elliptic curves studied by Hindry–Pacheco and Griffon;
- (5) proofs that $\lim_d \text{BS}(J_d) = 1$ for families of Jacobians of all dimensions;
- (6) and results on quadratic twists that illustrate the limitations of our p -adic techniques.

“Without recourse to L -functions” means by algebraic methods. We do use the BSD formula, but this is just a bookkeeping device for the connections between cohomology and other invariants. We do not use the Euler product or any properties of $L(A, s)$ as a function of s . That said, we have not eliminated

analysis entirely: points (4–6) above all require an equidistribution result for the action of multiplication by p on $\mathbb{Z}/d\mathbb{Z}$.

The plan of the paper is as follows: In Section 2, we set up notation, review and extend certain auxiliary results of Hindry–Pacheco on component groups, and prove a lemma useful for estimating heights. In Section 3, we prove a general integrality result on regulators of abelian varieties which leads immediately to a lower bound on the Brauer–Siegel ratio. In Section 4, we introduce “ $\dim \text{III}(A)$ ”, a new and extremely useful technical device which is closely related to slopes of L -functions and which is computable in many interesting situations. As a first application, in Section 5 we compute the limiting Brauer–Siegel ratio for the sequence of Frobenius pull-backs of an elliptic curve. Sections 6–9 develop p -adic cohomological machinery that allows one to compute $\dim \text{III}(A)$ and estimate $\text{BS}(A)$ for Jacobians of curves with Néron models related to products of Fermat curves. In the rest of the paper, we use this machinery to recover the results of Hindry–Pacheco and Griffon and to extend them to higher genus Jacobians. Section 10 discusses curves defined by equations involving 4 monomials. Section 11 discusses curves coming from Berger’s construction [2008]. Finally, in Section 12 we consider twists of constant elliptic curves.

It is a pleasure to thank Richard Griffon for several helpful comments and an anonymous referee for his or her careful reading of the paper and valuable suggestions.

2. Preliminaries

2.1. Notation and definitions. We set notation and recall definitions which will be used throughout the paper.

Fix a prime number p , a power q of p , and a smooth, projective, absolutely irreducible curve \mathcal{C} of genus $g_{\mathcal{C}}$ over $k = \mathbb{F}_q$, the field of q elements. Let K be the function field $\mathbb{F}_q(\mathcal{C})$. We write v for a place of K , d_v for the degree of v , K_v for the completion of K at v , \mathcal{O}_v for the ring of integers in K_v , and k_v for the residue field, a finite extension of k of degree d_v .

Let A be an abelian variety over K with dual \hat{A} . A theorem of Lang and Néron guarantees that the *Mordell–Weil groups* $A(K)$ and $\hat{A}(K)$ are finitely generated abelian groups. (See [Lang and Néron 1959], or [Conrad 2006] for a more modern account.)

There is a bilinear pairing

$$\langle \cdot, \cdot \rangle : A(K) \times \hat{A}(K) \rightarrow \mathbb{Q}$$

which is nondegenerate modulo torsion. (This is the canonical Néron–Tate height divided by $\log q$. See [Néron 1965] for the definition and [Hindry and Silverman 2000, B.5] for a friendly introduction.) Choosing a basis P_1, \dots, P_r for $A(K)$ modulo torsion and a basis $\hat{P}_1, \dots, \hat{P}_r$ for $\hat{A}(K)$ modulo torsion, we define the *regulator* of A as

$$\text{Reg}(A) := |\det \langle P_i, \hat{P}_j \rangle_{1 \leq i, j \leq r}|.$$

The regulator is a positive rational number, well-defined independently of the choice of bases.

We write $H^1(K, A)$ for the étale cohomology of K with coefficients in A and similarly for $H^1(K_v, A)$. The *Tate–Shafarevich group* of A is defined as

$$\text{III}(A) := \ker \left(H^1(K, A) \rightarrow \prod_v H^1(K_v, A) \right),$$

where the product of over the places of K and the map is the product of the restriction maps. This group is conjectured to be finite, and we assume this conjecture throughout the paper. However, in all of the explicit calculations below, we can in fact prove that $\text{III}(A)$ is finite without additional assumptions.

Let $\mathcal{A} \rightarrow \mathcal{C}$ be the Néron model of A/K . This is a smooth group scheme over \mathcal{C} with a certain universal property whose generic fiber is A/K . See [Bosch et al. 1990] for a modern account. Let $s : \mathcal{C} \rightarrow \mathcal{A}$ be the zero-section. We define an invertible sheaf ω on \mathcal{C} by

$$\omega := s^*(\Omega_{\mathcal{A}/\mathcal{C}}^{\dim(A)}) = \bigwedge^{\dim(A)} s^*(\Omega_{\mathcal{A}/\mathcal{C}}^1).$$

The *exponential differential height* of A (which we often refer to simply as the *height*) is

$$H(A) := q^{\deg \omega}.$$

If A is an elliptic curve and $\mathcal{C} = \mathbb{P}^1$, then $\deg \omega$ has simple interpretation in terms of the degrees of the coefficients in a Weierstrass equation defining A . See [Ulmer 2011, Lecture 3] for details.

For each place v of K , we write c_v for the number of connected components of the special fiber of \mathcal{A} at v which are defined over the residue field. We define the *Tamagawa number* of A as

$$\tau(A) := \prod_v c_v.$$

(This usage is in conflict with our earlier papers, in particular [Ulmer 2014a], where the Tamagawa number is defined to be

$$\frac{\tau(A)}{H(A)q^{\dim(A)(g_{\mathcal{C}}-1)}}.$$

The earlier usage is historically more appropriate, as the definition there is a volume defined in close analogy with Tamagawa’s work on linear algebraic groups, see [Weil 1982], but the terminology we adopt here is more convenient for our current purposes.)

Next we consider the Hasse–Weil L -function of A over K , denoted $L(A, s)$. It is a function of a complex variable s defined as an Euler product over the places of K which is convergent in the half-plane $\Re s > \frac{3}{2}$ and which is known to have a meromorphic continuation to the whole s -plane. More precisely, $L(A, s)$ is a rational function in q^{-s} , and if the K/k -trace of A is trivial, then $L(A, s)$ is in fact a polynomial in q^{-s} of the form

$$\prod_i (1 - \alpha_i q^{-s}),$$

where the inverse root α_i are Weil integers of size q .

We define the leading coefficient of the L -function as

$$L^*(A) := \frac{1}{(\log q)^r} \frac{1}{r!} \left(\frac{d}{ds} \right)^r L(A, s) \Big|_{s=1}$$

where r is the order of vanishing $r := \text{ord}_{s=1} L(A, s)$. (With the factor $1/(\log q)^r$, this is the leading coefficient of L as a rational function in $T = q^{-s}$, and with this normalization, it has the virtue of being a rational number.)

All of the invariants mentioned above are connected by the conjecture of Birch and Swinnerton-Dyer (“BSD conjecture”), which we take to be the conjunction of the following three statements:

- (1) $\text{ord}_{s=1} L(A, s) = \text{Rank } A(K)$.
- (2) $\text{III}(A)$ is finite (with order denoted $|\text{III}(A)|$).
- (3) We have an equality

$$L^*(A) = \frac{\text{Reg}(A)|\text{III}(A)|}{H(A)} \frac{\tau(A)}{q^{\dim(A)(gc-1)} |A(K)_{\text{tor}}| \cdot |\hat{A}(K)_{\text{tor}}|}.$$

It is known that parts (1) and (2) are equivalent, and when they hold, part (3) holds as well. (See [Kato and Trihan 2003] for the end of a long line of reasoning leading to these results.)

From the point of view of the Brauer–Siegel ratio, the main terms of interest in the third part of the BSD conjecture are $\text{Reg}(A)$, $|\text{III}(A)|$, and $H(A)$, whereas the other factors are either constant ($q^{\dim(A)(gc-1)}$) or turn out to be negligible ($\tau(A)$ and $|A(K)_{\text{tor}} \times \hat{A}(K)_{\text{tor}}|$). We will discuss the Tamagawa number and the results of Hindry and Pacheco on it in the next section, whereas the torsion subgroups $A(K)_{\text{tor}}$ and $\hat{A}(K)_{\text{tor}}$ will play almost no role in our analysis.

2.2. Bounds on Tamagawa numbers (1). Hindry and Pacheco [2016, Proposition 6.8] bound the Tamagawa number in terms of the height under certain tameness assumptions. More precisely, they showed that for a fixed global field K , as A varies over all abelian varieties of fixed dimension d over K , we have

$$\tau(A) = O(H^\epsilon)$$

for all $\epsilon > 0$, provided that $p > 2 \dim(A) + 1$ or A has everywhere semistable reduction.

In this section and Sections 2.5 and 2.6, we outline three improvements of this result, all motivated by applications later in the paper.

Lemma 2.2.1. *Let E run through the set of all elliptic curves over a global function field K . Then*

$$\tau(E) = O(H(E)^\epsilon)$$

for every $\epsilon > 0$.

The point is that we allow arbitrary characteristic and make no semistability hypothesis. This result was also proven by Griffon [2016, Theorem 1.5.4], but we include a proof here for the convenience of the reader.

Proof. This follows easily from Ogg’s formula [1967] (see also [Saito 1988] for a more general result proven with modern methods). Indeed, if Δ_v is a minimal discriminant for E at the place v , Ogg’s formula says that

$$\text{ord}_v(\Delta_v) = c_v + f_v - 1$$

where f_v is the exponent of the conductor of E at v . Summing over places where E has bad reduction (i.e., where $\text{ord}_v(\Delta_v) \geq 1$) and using that $f_v - 1 \geq 0$ at these places, we have

$$\sum_v c_v d_v \leq \sum_v \text{ord}_v(\Delta_v) d_v \leq 12 \deg(\omega)$$

where d_v is the degree of v and where the last inequality holds because Δ can be interpreted as a section of $\omega^{\otimes 12}$. This recovers the main bound (Theorem 6.5 of [Hindry and Pacheco 2016]), and the rest of the argument — converting this additive bound to a multiplicative bound — proceeds exactly as in [Hindry and Pacheco 2016, Proposition 6.8]. \square

2.3. Families from towers of fields. Let A be an abelian variety over a function field K . For each positive integer d (or positive integer d prime to p), let K_d be a geometric extension of K , and let $A_d = A \times_K K_d$. This gives a sequence of abelian varieties and one may ask about the behavior of $\text{BS}(A_d)$ as $d \rightarrow \infty$.

For most of the paper, we will be concerned with the special case where there are isomorphisms $K_d \cong K$ for all d . In this case, we may view the sequence A_d as a sequence of abelian varieties over a fixed function field. This is the context of the results and conjectures of Hinry and Pacheco, and we will give four examples in the rest of this section. Nevertheless, the general case is also interesting, and we will give develop foundational results in a more general context in Section 2.4.

2.3.1. Kummer families. Let $K = \mathbb{F}_q(t)$, and for each positive integer d prime to p , let $K_d = \mathbb{F}_q(u)$ where $u^d = t$. Note that the extension K_d/K is unramified away from the places $t = 0$ and $t = \infty$ of K . Let A be an abelian variety over K , and let A_d be the abelian variety over K obtained by base change to K_d , followed by the isomorphism of fields $\mathbb{F}_q(u) \cong \mathbb{F}_q(t)$, $u \mapsto t$. (In more vivid terms, A_d is the result of substituting t^d for each appearance of t in the equations defining A .) We say that the sequence of abelian varieties A_d is the *family associated to A and the Kummer tower*. Such families have been a prime source of examples for the Brauer–Siegel ratio.

2.3.2. Artin–Schreier families. We may proceed analogously with the tower of Artin–Schreier extensions. Again, let $K = \mathbb{F}_q(t)$, and for each positive integer d , let $K_d = \mathbb{F}_q(u)$ where $u^{p^d} - u = t$. Note that the extension K_d/K is unramified away from the place $t = \infty$ of K . Let A be an abelian variety over K , and let A_d be the abelian variety over K obtained by base change to K_d followed by the isomorphism of fields $\mathbb{F}_q(u) \cong \mathbb{F}_q(t)$, $u \mapsto t$. (In more vivid terms, A_d is the result of substituting $t^{p^d} - t$ for each appearance of t in the equations defining A .) We say that the sequence of abelian varieties A_d is the *family associated to A and the Artin–Schreier tower*.

2.3.3. Division tower families. One may also consider an elliptic curve variant: Let K be the function field $\mathbb{F}_q(E)$ where E is an elliptic curve over \mathbb{F}_q . For each positive integer d prime to p , consider the field extension K_d/K associated to the multiplication map $d : E \rightarrow E$. Thus $[K_d : K] = d^2$, but K_d is canonically isomorphic as a field (even as an \mathbb{F}_q -algebra) to K . Given an abelian variety A over K , let A_d be the abelian variety over K obtained by base-changing A to K_d and then using the isomorphism of fields $K_d \cong K$. We say that the sequence A_d of abelian varieties over K is the *family associated to a division tower*. Everything we say about Kummer and Artin–Schreier towers has an obvious analog for division towers. In most cases the latter is simpler because in the division case, K_d/K is unramified.

2.3.4. PGL_2 families. Let $K = \mathbb{F}_q(t)$ and for each positive integer d let $K_d = \mathbb{F}_q(u)$ where $\mathbb{F}_q(u)/\mathbb{F}_q(t)$ is the field extension associated to the quotient morphism

$$\mathbb{P}^1 \rightarrow \mathbb{P}^1 / \mathrm{PGL}_2(\mathbb{F}_{p^d}) \cong \mathbb{P}^1.$$

We normalize the isomorphism so that the \mathbb{F}_{p^d} -rational points on the upper \mathbb{P}^1 map to 0 and $\mathbb{P}^1(\mathbb{F}_{p^{2d}}) \setminus \mathbb{P}^1(\mathbb{F}_{p^d})$ maps to 1. Then the extension K_d/K is unramified away from the places $t = 0$ and $t = 1$ of K , and it is tamely ramified over $t = 1$. Given an abelian variety A over K , let A_d be the abelian variety over K obtained by base-changing A to K_d and then using the isomorphism of fields $\mathbb{F}_q(u) \cong \mathbb{F}_q(t)$, $u \mapsto t$. We say that the sequence A_d of abelian varieties over K is the *family associated to the PGL_2 tower*.

The discussion above gives four different meanings to the notations K_d and A_d ! Which meaning is intended in each use below should be clear from the context.

We end this section with a simple lemma that plays a key role in our analysis of Tamagawa numbers in families associated to towers.

Lemma 2.3.5. *Let $K = \mathbb{F}_q(C)$ be a function field, and let K_d be a sequence of geometric extensions of K such that the genus of (the curve associated to) K_d is ≤ 1 for all d . Then for every place v of K , there is a constant C_v depending only on q and $\deg v$ such that for all d , the number of places of K_d dividing v is at most $C_v[K_d : K] / \log[K_d : K]$.*

Proof. Write $D = [K_d : K]$ and set $x = \log D / \log q$. Fix a place v of K . Then the number of places w of K_d dividing v and of absolute degree $\geq x$ is at most

$$\frac{D}{x / \deg v} = \deg v \log q \frac{D}{\log D}.$$

On the other hand, by the Weil bound, the total number of places of K_d of degree $\leq x$ is bounded by $Cq^x/x = C'D/\log D$ where C and C' depend only on q , $\deg v$ and the genus of K_d . Since the latter is either 0 or 1, the constant can be taken to depend only on q and $\deg v$. This shows that the total number of places of K_d dividing v is $\leq C_v D / \log D$ where C_v depends only on q and $\deg v$. \square

2.4. Towers of geometrically Galois extensions. In this section, we discuss a more general class of towers of fields K_d where we are able to bound Tamagawa numbers of the associated sequences of abelian varieties. This additional generality was suggested by the anonymous referee, to whom we are grateful.

Readers who are mainly interested in the applications to the Kummer tower later in the paper are invited to skip ahead to Section 2.5

2.4.1. Geometrically Galois extensions. Let k be a field and let $K = k(\mathcal{C})$ be the function field of a smooth, projective, geometrically irreducible curve over k . We say that a finite, geometric extension K_d/K is *geometrically Galois* if the Galois closure L_d of K_d over K has the form $L_d = k_d K_d$ where k_d is a finite Galois extension of k . Equivalently, there is a finite Galois extension k_d of k such that $k_d K_d$ is Galois over $k_d K$. (We take k_d to be minimal such extension.) Let $G_d = \text{Gal}(L_d/k_d K)$ and $\Gamma_d = \text{Gal}(k_d/k) \cong \text{Gal}(k_d K/K) \cong \text{Gal}(L_d/K_d)$, so that Γ_d acts on G_d by conjugation and $\text{Gal}(L_d/K)$ is the semidirect product $G_d \rtimes \Gamma_d$. We call G_d , with its action of Γ_d , the *geometric Galois group* of K_d/K and we call k_d the *splitting field* of G_d . (We remark that there is a finite étale group scheme \underline{G}_d over k attached to G_d with its Γ_d action, and \underline{G}_d becomes a constant group over k_d , see [Milne 1980, §II.1].)

2.4.2. Towers of geometrically Galois extensions. We now consider a tower of geometrically Galois extensions K_d/K indexed by positive integers d (or positive integers relatively prime to p) with containments $K_d \subset K_{d'}$ whenever d divides d' . These containments induce surjections $G_{d'} \rightarrow G_d$ and $\Gamma_{d'} \rightarrow \Gamma_d$ which are compatible in the obvious sense with the actions of Γ_d and $\Gamma_{d'}$ on G_d and $G_{d'}$ respectively.

Each of the families of towers in Section 2.3 gives an example of a tower of geometrically Galois extensions.

In the case of the Kummer tower, the geometric Galois group is $G_d = \mu_d(\overline{\mathbb{F}_q})$, the splitting field k_d is $\mathbb{F}_q(\mu_d)$, and $\Gamma_d = \text{Gal}(\mathbb{F}_q(\mu_d)/\mathbb{F}_q)$ is the subgroup of $(\mathbb{Z}/d\mathbb{Z})^\times$ generated by q .

In the Artin–Schreier tower, the geometric Galois group is $G_d = \mathbb{F}_{p^d}$, the splitting field k_d is the compositum $\mathbb{F}_q \mathbb{F}_{p^d}$, and $\Gamma_d = \text{Gal}(\mathbb{F}_q \mathbb{F}_{p^d}/\mathbb{F}_q)$ is the cyclic group generated by the q -power Frobenius.

In the division tower corresponding to an elliptic curve E over \mathbb{F}_q , the geometric Galois group is $E[d]$, the splitting field k_d is $\mathbb{F}_q(E[d])$, and $\Gamma_d = \text{Gal}(k_d/\mathbb{F}_q)$ is the cyclic group generated by the action of the q -power Frobenius on the d torsion points.

In the PGL_2 tower, the geometric Galois group is $G_d = \text{PGL}_2(\mathbb{F}_{p^d})$, the splitting field k_d is $\mathbb{F}_q \mathbb{F}_{p^d}$, and $\Gamma_d = \text{Gal}(\mathbb{F}_q \mathbb{F}_{p^d}/\mathbb{F}_q)$ is the cyclic group generated by the q -power Frobenius.

For a more general class of examples, let K_d/K be any of the towers above, and fix an extension F/K which is linearly disjoint from each K_d over K . Then the fields $F_d := F K_d$ form a tower of geometrically Galois extensions with the geometric Galois group of F_d/F isomorphic to that of K_d/K . Note however, that in general the genus of F_d tends to infinity with d .

We next consider two group-theoretic results related to these towers, both concerning the number of orbits of Γ_d acting on G_d . (As motivation, we note that the orbits of Γ_d on G_d are in bijection with the closed points of the scheme \underline{G}_d .)

To state the first result, we make a somewhat elaborate hypothesis on the system of groups G_d with their Γ_d actions.

Hypothesis 2.4.3. (1) There exists a function ϕ of positive integers such that $|G_d| = \sum_{e|d} \phi(e)$ for all d .

(2) There a decomposition $G_d = \cup_{e|d} G'_{d,e}$ such that $|G'_{d,e}| = \phi(e)$.

- (3) The action of Γ_d on G_d respects the decomposition above, and the orbits of Γ_d on $G'_{d,e}$ have cardinality $\geq C \log|G_e|$ for some constant C independent of d and e .

This hypothesis clearly implies that the splitting field k_d has degree $[k_d : k] = |\Gamma_d| \geq C \log|G_d|$. It would be interesting to know whether the converse holds.

Lemma 2.4.4. *Hypothesis 2.4.3 is satisfied by the Kummer, Artin–Schreier, division, and PGL_2 towers.*

Proof. In the Kummer case, G_d consists of the d -th roots of unity in $\overline{\mathbb{F}_q}$, and we let $G'_{d,e}$ be those of order exactly e . Then $|G'_{d,e}|$ is independent of d , and we set $\phi(e) = |G'_{d,e}|$. The orbit of Γ through $\zeta \in G'_{d,e}$ has size f where f is the smallest positive integer such that $\zeta^{q^f} = \zeta$. Since ζ has order exactly e , this is the smallest f such that $q^f \equiv 1 \pmod{e}$. Clearly this f satisfies $f \geq \log e / \log q$ and this establishes Hypothesis 2.4.3.

In the Artin–Schreier case, G_d is the additive group of \mathbb{F}_{p^d} , and we let $G'_{d,e}$ consists of those elements of $\mathbb{F}_{p^e} \subset \mathbb{F}_{p^d}$ which do not lie in any smaller extension of \mathbb{F}_p , i.e., $\alpha \in G'_{d,e}$ if and only if $\mathbb{F}_p(\alpha) = \mathbb{F}_{p^e}$. We set $\phi(e) = |G'_{d,e}|$ (which is independent of d). Since $\alpha^{p^f} \neq \alpha$ for $0 < f < e$, it follows immediately that the orbit of the q -power Frobenius through $\alpha \in G'_{d,e}$ has size at least $e / (\log q / \log p)$, and this establishes Hypothesis 2.4.3.

In the division case, G_d consists of the $\overline{\mathbb{F}_q}$ -points of E of order dividing d . We let $G'_{d,e}$ be the subset of points of order exactly e , and $\phi(e) = |G'_{d,e}|$ (which is independent of d). If $P \in G'_{d,e}$ and $\text{Fr}_q^f(P) = P$, then $P \in E(\mathbb{F}_{q^f})$, and this implies that $|E(\mathbb{F}_{q^f})| \geq e$. But the Weil bound implies that $|E(\mathbb{F}_{q^f})| \leq (q^{f/2} + 1)^2$ which in turn implies that $f \geq C \log e$ for some constant C independent of e .

In the PGL_2 case, G_d is $\text{PGL}_2(\mathbb{F}_{p^d})$. For $g \in G_d$, let $\mathbb{F}_p(g)$ be defined as follows: choose a representative of g in $\text{GL}_2(\mathbb{F}_{p^d})$ one of whose entries is 1, and let $\mathbb{F}_p(g)$ be the extension of \mathbb{F}_p generated by the other entries. It is easy to see that $\mathbb{F}_p(g)$ is well defined independent of the choice of representative and that $\text{Fr}_p^f(g) = g$ if and only if Fr_p^f fixes $\mathbb{F}_p(g)$. We let $G'_{d,e}$ consists of those elements $g \in G_d$ with $\mathbb{F}_p(g) = \mathbb{F}_{p^e}$. We set $\phi(e) = |G'_{d,e}|$ (which is independent of d). Since $\text{Fr}_p^f(g) \neq g$ for $0 < f < e$, it follows immediately that the orbit of the q -power Frobenius through $g \in G'_{d,e}$ has size at least $e / (\log q / \log p)$, and this establishes Hypothesis 2.4.3. □

Remark 2.4.5. A “dual” perspective makes Hypothesis 2.4.3 more transparent in the cases considered in Lemma 2.4.4. Namely, let $F = \mathbb{F}_q(\mathcal{C})$ be the function field of a curve of genus 0 or 1 over \mathbb{F}_q . (These are the cases where $\text{Aut}_{\overline{\mathbb{F}_q}}(\mathcal{C})$ is infinite.) For each d , let G_d be a subgroup of $\text{Aut}_{\overline{\mathbb{F}_q}}(\mathcal{C})$ which is stable under the q -power Frobenius, and let Γ_d be the group of automorphisms of G_d generated by Frobenius. The quotient $(\mathcal{C} \times \overline{\mathbb{F}_q}) / G_d$ has a canonical model over \mathbb{F}_q ; let F_d be its function field. With this notation, the extension F / F_d is geometrically Galois with group (G_d, Γ_d) . Suppose further that if $e \mid d$ then $G_e \subset G_d$, so that $F_d \subset F_e$. Then it is natural to define G'_d as the set of elements in G_d which are not in G_e for any divisor of d with $e < d$. Clearly G'_e depends only on e , and the decomposition $G_d = \cup_{e \mid d} G'_e$ is evident. All of the examples of Lemma 2.4.4 can be recast in this form.

The following lemma is modeled on [Griffon 2016, Lemme 3.1.1].

Lemma 2.4.6. *Let K_d/K be a tower of geometrically Galois extensions such that for all d , $|G_d| \geq d$ and such that Hypothesis 2.4.3 holds. Then there is a constant C_1 such that the number of orbits of Γ_d on G_d satisfies*

$$|G_d/\Gamma_d| \leq C_1 \frac{|G_d|}{\log|G_d|}$$

for all $d > 1$.

Proof. Let $\psi(d) = |G_d|$, so that $\psi(d) = \sum_{e|d} \phi(e)$. Extend ψ to a function of real numbers which is continuous, increasing, and satisfies $\psi(x) \geq x$ for all x . By Hypothesis 2.4.3, for all $d > 1$ the number of orbits of Γ_d on $G'_{d,e}$ satisfies

$$|G'_{d,e}/\Gamma_d| \leq C^{-1} \frac{\phi(e)}{\log \psi(e)}.$$

Let $x > 1$ be a parameter to be chosen later. We have

$$\begin{aligned} |\Gamma_d| &\leq C_2 \sum_{1 < e|d} \frac{\phi(e)}{\log \psi(e)} && (C_2 \text{ to compensate for omitting } e = 1) \\ &= C_2 \sum_{\substack{1 < e|d \\ e \leq x}} \frac{\phi(e)}{\log \psi(e)} + C_2 \sum_{\substack{1 < e|d \\ e > x}} \frac{\phi(e)}{\log \psi(e)} \\ &\leq C_2 \sum_{\substack{1 < e|d \\ e \leq x}} \frac{\phi(e)}{\log \psi(e)} + C_2 \frac{\psi(d)}{\log \psi(x)} && (\sum \phi(e) = \psi(d) \text{ and } \psi \text{ increasing}) \\ &\leq C_2 \sum_{\substack{1 < e|d \\ e \leq x}} \frac{\psi(e)}{\log \psi(e)} + C_2 \frac{\psi(d)}{\log \psi(x)} && (\phi(e) \leq \psi(e)) \\ &\leq C_3 \frac{\psi(x)}{\log \psi(x)} \sum_{\substack{1 < e|d \\ e \leq x}} 1 + C_2 \frac{\psi(d)}{\log \psi(x)} && (x \mapsto \psi(x) \mapsto \psi(x)/\log \psi(x), \text{ increasing for } x > 2.72) \\ &\leq C_3 \frac{x\psi(x)}{\log \psi(x)} + C_2 \frac{\psi(d)}{\log \psi(x)} \\ &\leq C_3 \frac{\psi(x)^2}{\log \psi(x)} + C_2 \frac{\psi(d)}{\log \psi(x)} && (\psi(x) \geq x) \end{aligned}$$

Now since ψ is increasing and continuous, we may choose x so that $\psi(x)^2 = \psi(d)$, and for this choice we have

$$|G_d/\Gamma_d| \leq (2C_3 + 2C_2) \frac{\psi(d)}{\log \psi(d)}.$$

Thus setting $C_1 = 2C_3 + 2C_2$ completes the proof. □

We now consider the set of orbits of Γ on a homogeneous space for G .

Lemma 2.4.7. *Let G be a finite group and let T be a principal homogeneous space for G . Let Γ be a group acting on G (by group automorphisms) and on T (by permutations), and suppose that the actions*

of Γ on G and T are compatible with the action of G on T (i.e., for all $\gamma \in \Gamma$, $g \in G$, and $t \in T$, $\gamma(gt) = \gamma(g)\gamma(t)$). Then

$$|T/\Gamma| \leq |G/\Gamma|.$$

Proof. We use the orbit counting lemma:

$$|G/\Gamma| = \frac{1}{|\Gamma|} \sum_{\gamma \in \Gamma} |G^\gamma|$$

where G^γ denotes the set of fixed points of γ acting on G . Similarly,

$$|T/\Gamma| = \frac{1}{|\Gamma|} \sum_{\gamma \in \Gamma} |T^\gamma|$$

where T^γ denotes the set of fixed points of γ acting on T . We claim that if T^γ is not empty, then it is a principal homogeneous space for G^γ . Indeed, it is clear that if $g \in G^\gamma$ and $t \in T^\gamma$, then $gt \in T^\gamma$. Conversely, if $t, t' \in T^\gamma$ and $g \in G$ is the unique element such that $gt = t'$, then

$$\gamma(g)t = \gamma(g)\gamma(t) = \gamma(gt) = \gamma(t') = t' = gt,$$

and so $\gamma(g) = g$. Therefore, for each $\gamma \in \Gamma$, $|T^\gamma| \leq |G^\gamma|$. We conclude that

$$|T/\Gamma| = \frac{1}{|\Gamma|} \sum_{\gamma \in \Gamma} |T^\gamma| \leq \frac{1}{|\Gamma|} \sum_{\gamma \in \Gamma} |G^\gamma| = |G/\Gamma|,$$

and this completes the proof of the lemma. □

Remark 2.4.8. In fact, the conclusion of the lemma holds when we assume only that G acts transitively on T . To see this, it suffices to check that for all $\gamma \in \Gamma$, $|T^\gamma| \leq |G^\gamma|$. If T^γ is empty, there is nothing to prove. If not, choose $t_0 \in T^\gamma$, let G_0 be the stabilizer of t_0 in G , and set

$$F(\gamma) = \{g \in G \mid \gamma(gt_0) = gt_0\} = \{g \in G \mid g^{-1}\gamma(g) \in G_0\}.$$

Then G_0 acts freely on $F(\gamma)$ by right multiplication, and the quotient is T^γ . Thus $|F(\gamma)| = |G_0| \cdot |T^\gamma|$. On the other hand, G^γ acts freely on $F(\gamma)$ by left multiplication, and the quotient maps injectively to G_0 by $g \mapsto g^{-1}\gamma(g)$. Thus we find

$$|G_0| \cdot |T^\gamma| = |F(\gamma)| \leq |G^\gamma| \cdot |G_0|$$

and so $|T^\gamma| \leq |G^\gamma|$. It is also clear that $G^\gamma G_0 \subset F(\gamma)$ so in all we have

$$\frac{|G^\gamma|}{|G_0^\gamma|} \leq |T^\gamma| \leq |G^\gamma|.$$

Simple examples show that both bounds are sharp. Thanks to Alex Ryba for the proofs in this remark and the preceding lemma.

Corollary 2.4.9. *Suppose that K_d is a tower of geometrically Galois extensions of K such that $[K_d : K] \geq d$ and such that Hypothesis 2.4.3 holds. Let v be a place of K . Then there is a constant C_v depending only on K and v such that for all d the number of place of K_d over v is at most $C_v[K_d : K] / \log[K_d : K]$.*

Proof. First assume that v is unramified in K_d . Let T_d be the set of geometric points in the fiber over v (i.e., in the fiber of the map of curves corresponding to the extension K_d/K) and let $G = G_d$ be the geometric Galois group of K_d over K . Let k_v be the residue field at v and let $\Gamma_d = \text{Gal}(k_d/k_v)$, a subgroup of the Galois group of the splitting field of G_d . Then T_d is a principal homogeneous space for G_d , and Γ_d acts on G_d and T_d compatibly with the action of G_d on T_d . By Lemma 2.4.7, $|T_d/\Gamma_d| \leq |G_d/\Gamma_d|$. But T_d/Γ_d is in bijection with the set of places of K_d over v , and by Lemma 2.4.6 (applied to the extensions $k_v K_d/k_v K$), there is a constant C_v (depending on v because the tower in question depends on v) such that

$$|G_d/\Gamma_d| \leq C_v \frac{[K_d : K]}{\log[K_d : K]}.$$

This completes the proof of the corollary when v is unramified in K_d . The general case follows from the same argument using Remark 2.4.8 in place of Lemma 2.4.7. \square

2.5. Bounds on Tamagawa numbers (2). We now turn to a second improvement on the Hindry–Pacheco bound on Tamagawa numbers. We consider towers of fields satisfying the conclusions of Lemma 2.3.5 and Corollary 2.4.9, and we bound Tamagawa numbers using only a mild (local) semistability hypothesis and no restriction on the characteristic of the ground field.

Recall the line bundle ω_A associated to an abelian variety A defined in Section 2.1.

Proposition 2.5.1. *Let K be a global function field of characteristic p , let Z be a finite set of places of K , and let K_d be a tower of geometrically Galois extensions of K . Assume that $[K_d : K] \geq d$ and that for each place v of K there is a constant C_v such that the number of places of K_d dividing v is $\leq C_v[K_d : K] / \log[K_d : K]$ for all d . Suppose that each K_d/K is unramified outside Z . Let A be an abelian variety over K which has semistable reduction at each place $v \in Z$ and such that $\deg \omega_A > 0$. Let $A_d = A \times_K K_d$. Then*

$$\tau(A_d) = O(H(A_d)^\epsilon)$$

for every $\epsilon > 0$.

Proof. To lighten notation, let $D = [K_d : K]$. Since A has semistable reduction at the possibly ramified places Z , we have $\deg \omega_{A_d} = D \deg \omega_A \geq D$, so it will suffice to show that

$$\tau(A_d) = O(q^{D\epsilon})$$

for all $\epsilon > 0$.

For each place v of K , let c_v be the order of the group of connected components of the special fiber of the Néron model of A at v . Let \bar{c}_v be the order of the group of connected components of the special fiber of the Néron model of A at a place of $\overline{\mathbb{F}_q}K$ over v . (The order is independent of the choice.) Since the

former group is a subgroup of the latter, c_v divides \bar{c}_v . If w is a place of K_d over v , let c_w be the order of the component group of the Néron model of A over K_d .

Consider a place $v \notin Z$. Since K_d/K is unramified at v , c_w divides \bar{c}_v . By assumption, the number of places w over v is bounded by $C_v D/\log D$. Since there are only finitely many places of K where A has bad reduction, we may set $C_1 = \max\{\bar{c}_v^{C_v} \mid v \text{ of bad reduction}\}$ and conclude that

$$\prod_{w \mid v \notin Z} c_w \leq \prod_{v \notin Z} \bar{c}_v^{C_v D/\log D} \leq C_1^{D/\log D}.$$

Now consider a place $v \in Z$, let w be a place of K_d over v , and let r be the ramification index of w over v . Since K_d/K is geometrically Galois, r depends only on v . Since A is assumed to have semistable reduction, [Halle and Nicaise 2010, Theorem 5.7] implies that

$$c_w \leq \bar{c}_v r^{\dim(A)}.$$

Moreover, by assumption, the number of places of K_d over v is at most $\min\{D/r, C_v D/\log D\}$ for some constant C_v which is independent of D . If $r \leq (\log D)/C_v$, we have

$$\prod_{w \mid v} c_w \leq (\bar{c}_v r^{\dim(A)})^{C_v D/\log D} \leq C_2^{D/(\log D/\log \log D)}$$

where C_2 depends only on v and A . If $r \geq (\log D)/C_v$, we have

$$\prod_{w \mid v} c_w \leq (\bar{c}_v r^{\dim(A)})^{D/r} \leq C_3^{D \log r/r} \leq C_4^{D/(\log D/\log \log D)}$$

where again C_3 and C_4 depend only on v and A .

Taking the product over all place w of K_d and setting $C_5 = \max\{C_2, C_4\}$, we have

$$\prod_w c_w = \left(\prod_{w \mid v \notin Z} c_w \right) \left(\prod_{w \mid v \in Z} c_w \right) \leq (C_1^{D/\log D}) (C_5^{D/(\log D/\log \log D)})^{|Z|}$$

and this is clearly $O(q^{D^\epsilon})$ as d (and therefore D) tends to infinity. □

We now give the main application of the results in this section. Assume $K = \mathbb{F}_q(t)$ or $K = \mathbb{F}_q(E)$ for an elliptic curve E , and consider a family of abelian varieties A_d over K associated to the Kummer, Artin–Schreier, division, or PGL_2 towers. Recall the line bundle $\omega = \omega_A$ defined in the Section 2.1.

Corollary 2.5.2. *As d runs through positive integers prime to p (or all positive integers in the Artin–Schreier case), we have*

$$\tau(A_d) = O(H(A_d)^\epsilon)$$

for every $\epsilon > 0$ in any of the following situations:

- (1) A is an abelian variety over $K = \mathbb{F}_q(t)$, A_d is the family associated to the Kummer tower, $\deg(\omega) > 0$, and A has semistable reduction at $t = 0$ and $t = \infty$.

- (2) A is an abelian variety over $K = \mathbb{F}_q(t)$, A_d is the family associated to the Artin–Schreier tower, $\deg(\omega) > 0$, and A has semistable reduction at $t = \infty$.
- (3) A is an abelian variety over $K = \mathbb{F}_q(E)$, A_d is the family associated to the division tower, and $\deg(\omega) > 0$.
- (4) A is an abelian variety over $K = \mathbb{F}_q(t)$, A_d is the family associated to the PGL_2 tower, $\deg(\omega) > 0$, and A has semistable reduction at $t = 0$ and $t = 1$.

Proof. This is an immediate consequence of Proposition 2.5.1 together with Lemma 2.3.5. □

2.6. Bounds on Tamagawa numbers (3). Our third improvement on the Hindry–Pacheco bound on Tamagawa numbers is to note that we can get by with a weaker hypotheses in case (1) of Corollary 2.5.2. Namely, we claim that the conclusion of the corollary holds if there exists an integer e relatively prime to p such that A has semistable reduction at the places $u = 0$ and $u = \infty$ of $\mathbb{F}_q(u)$ where $u^e = t$. (The corollary is the case where $e = 1$.)

To check the claim, we first recall a result of Halle and Nicaise: Let A be an abelian variety over $\overline{\mathbb{F}_p}((t))$. For d prime to p , let c_d denote the order of the component group of the special fiber of the Néron model of A over $\overline{\mathbb{F}_p}((t^{1/d}))$. Then [Halle and Nicaise 2010, Theorem 6.5] states that if we assume that A acquires semistable reduction over $\overline{\mathbb{F}_p}((t^{1/e}))$ for some e prime to p , then the series

$$\sum_{(p,d)=1} c_d T^d$$

is a rational function in T and $1/(T^j - 1)$ for $j \geq 1$. This implies in particular that the c_d have at worst polynomial growth: $c_d = O(d^N)$ for some N .

Applying this result in the context of part (1) of the lemma for the places $t = 0$ and $t = \infty$ of $\mathbb{F}_q(t)$, we see that

$$\tau(A_d) \leq C_1^{d/\log d} d^{C_6} = O(H(A_d)^\epsilon)$$

for all $\epsilon > 0$.

2.7. Estimating $\deg(\omega_J)$. When $A = J$ is the Jacobian of a curve X over a function field, computing $\deg(\omega_J)$ typically involves knowledge of a regular model of X (or a mildly singular model), information which is sometimes difficult to obtain. The following lemma allows us to reduce to easy cases in two examples later in the paper.

Lemma 2.7.1. *Let $K = k(C)$ be the function field of a curve over a perfect field k . Let X be a smooth, projective curve of genus g over K . Let J be the Jacobian of X , let $\pi : \mathcal{X} \rightarrow \mathcal{C}$ be a regular minimal model of X over K , and let $\mathcal{J} \rightarrow \mathcal{C}$ be the Néron model of J with zero-section $z : \mathcal{C} \rightarrow \mathcal{J}$. Let*

$$\omega_J := \wedge^g(z^* \Omega_{\mathcal{J}/\mathcal{C}}^1)$$

be the Hodge bundle of J .

Let K' be a finite, separable, geometric extension of K , and let $\rho : \mathcal{C}' \rightarrow \mathcal{C}$ be the corresponding morphism of curves over k . Let $R = (2g_{\mathcal{C}'} - 2) - [K' : K](2g_{\mathcal{C}} - 2)$.

Let $X' = X \times_K K'$ with Jacobian J' , models \mathcal{X}' and \mathcal{J}' , and Hodge bundle $\omega_{J'}$. Then

$$[K' : K] \deg(\omega_J) \leq \deg(\omega_{J'}) + gR.$$

The point of the lemma is that we do not lose much information in passing to a finite extension.

Proof of Lemma 2.7.1. Since \mathcal{X} is regular and π has a section, we have that

$$\omega_J \cong \wedge^g (\pi_* \Omega_{\mathcal{X}/k}^2 \otimes (\Omega_{\mathcal{C}/k}^1)^{-1}) \cong (\wedge^g \pi_* \Omega_{\mathcal{X}/k}^2) \otimes (\Omega_{\mathcal{C}/k}^1)^{\otimes -g}$$

and similarly for $\omega_{J'}$. This argument, which uses results on Néron models and relative duality, is given in the proof of [Berger et al. 2015, Prop. 7.4].

There is a dominant rational map $\mathcal{X}' \dashrightarrow \mathcal{X}$ covering ρ , so pull back of 2-forms induces a nonzero morphism of sheaves

$$\rho^* \wedge^g (\pi_* \Omega_{\mathcal{X}/k}^2) \rightarrow \wedge^g (\pi'_* \Omega_{\mathcal{X}'/k}^2).$$

By Riemann–Hurwitz, we have

$$\rho^* (\Omega_{\mathcal{C}/K}^1) \cong \Omega_{\mathcal{C}'/k}^1 \otimes \mathcal{O}_{\mathcal{C}'}(D)$$

where D is a divisor on \mathcal{C}' of degree R .

Thus we get a nonzero morphism of sheaves

$$\rho^*(\omega_J) \rightarrow \omega_{J'} \otimes \mathcal{O}_{\mathcal{C}'}(gD).$$

Taking degrees, we conclude that

$$[K' : K] \deg(\omega_J) \leq \deg(\omega_{J'}) + gR$$

as desired. □

3. Integrality of the regulator and general lower bounds

In this section, we give a lower bound on the regulator $\text{Reg}(A)$ in terms of Tamagawa numbers. Combined with the bounds on $\tau(A)$ given in the preceding section, this yields a lower bound on the Brauer–Siegel ratio. A more general version of the same lower bound was proven in [Hindry and Pacheco 2016, Proposition 7.6], but our proof is arguably simpler and more uniform, and avoids a forward reference in [Hindry and Pacheco 2016].

3.1. Integrality of regulators. We continue with the standard notations introduced in Section 2. In particular, A is an abelian variety over the function field $K = k(\mathcal{C})$ with Néron model \mathcal{A} and dual abelian variety \hat{A} . We consider the height pairing $A(K) \times \hat{A}(K) \rightarrow \mathbb{Q}$ (which we recall is the canonical Néron–Tate height divided by $\log q$ and which takes values in \mathbb{Q}) and its determinant $\text{Reg}(A)$.

Our main goal in this section is to bound the denominator of the regulator in terms of the orders c_v of the component groups of \mathcal{A} at places v of K . Recall that $\tau(A) = \prod_v c_v$.

Proposition 3.1.1. *The rational number*

$$\tau(A) \operatorname{Reg}(A)$$

is an integer.

Proof. We refer to [Hindry and Silverman 2000] for general background on heights. Given an invertible sheaf \mathcal{L} on A and a point $x \in A(K)$, the general theory of heights on abelian varieties defines a rational number $h_{\mathcal{L}}(x)$. The canonical height pairing we are discussing is defined using this machine and the identification of \hat{A} with $\operatorname{Pic}^0(A)$, the group of invertible sheaves algebraically equivalent to zero. In other words, given $x \in A(K)$ and $y \in \hat{A}(K)$, we take \mathcal{L} to be the invertible sheaf associated to y and define

$$\langle x, y \rangle = h_{\mathcal{L}}(x).$$

Néron’s theory [1965] decomposes the height $h_{\mathcal{L}}(x)$ into a sum of local terms indexed by the places of K . In [Moret-Bailly 1985, III.1], Moret-Bailly proves that the contribution at a place v has denominator at most $2c_v$, and at most c_v if c_v is odd. Moreover, he gives an example which shows that this is in general best possible. The upper bound on the denominator comes from a property of “pointed maps of degree 2,” [Moret-Bailly 1985, I.5.6], namely that a pointed map of degree 2 from a group of exponent n has exponent at worst $2n$, or n if n is odd. (These terms will be defined just below.)

In our situation there is slightly more structure: Since \mathcal{L} is algebraically equivalent to zero, it is antisymmetric, i.e., if $[-1]$ is the inverse map on A , the $[-1]^*\mathcal{L} \cong \mathcal{L}^{-1}$. The functoriality in [Moret-Bailly 1985, III.1.1] then shows that the corresponding pointed map of degree 2 is also antisymmetric. In the next lemma, we define antisymmetric pointed maps of degree 2, and we prove that such a map from a group of exponent c has exponent dividing c .

Thus we see that $\langle x, y \rangle$ is a sum of local terms, and the term at a place v has denominator at worst c_v . It follows from the bilinearity of the local terms $\langle \cdot, \cdot \rangle_v$ that if x passes through the identity component at v , then $\langle x, y \rangle_v$ is an integer. We define a “reduced Mordell–Weil group”

$$A(K)^{\operatorname{red}} := \{x \in A(K) \mid x \text{ meets the identity component of } \mathcal{A} \text{ at every } v\},$$

and note that if $x \in A(K)^{\operatorname{red}}$, then $\langle x, y \rangle$ is an integer for every $y \in \hat{A}(K)$. Since the index of $A(K)^{\operatorname{red}}$ in $A(K)$ divides $\tau(A) = \prod_v c_v$, we see that

$$\operatorname{Reg}(A) \in \tau^{-1}\mathbb{Z}$$

as desired. The proposition thus follows from the next lemma. □

Lemma 3.1.2. *Let A and G be abelian groups and let $f : A \rightarrow G$ be a function such that:*

(1) *f is a “pointed map of degree 2,” namely,*

$$f(x_1 + x_2 + x_3) - f(x_1 + x_2) - f(x_1 + x_3) - f(x_2 + x_3) + f(x_1) + f(x_2) + f(x_3) = 0$$

for all $x_1, x_2, x_3 \in A$.

(2) f is “antisymmetric,” i.e., $f(-x) = -f(x)$ for all $x \in A$.

Then for all integers n and all $x \in A$, $f(nx) = nf(x)$. In particular, if A has exponent c , then $cf = 0$, i.e., $cf(x) = 0$ for all $x \in A$.

Proof. This follows from a simple inductive argument. Clearly it suffices to treat the case $n \geq 0$. Taking $x_1 = x_2 = x_3 = 0$ in the pointed map property shows that $f(0) = 0$. Taking $x_1 = x_2 = x$ and $x_3 = -x$ then shows that $f(2x) = 2f(x)$. Finally, for $n \geq 2$, taking $x_1 = (n - 1)x, x_2 = x_3 = x$, we have

$$\begin{aligned} f((n + 1)x) &= f((n - 1)x + x + x) \\ &= f(nx) + f(nx) + f(2x) - f((n - 1)x) - f(x) - f(x) \\ &= (n + n + 2 - (n - 1) - 1 - 1)f(x) \\ &= (n + 1)x, \end{aligned}$$

where we use induction to pass from the second displayed line to the third. This yields the lemma. \square

Without the antisymmetry hypothesis, we would have

$$f(nx) = \frac{n(n + 1)}{2} f(x) + \frac{n(n - 1)}{2} f(-x),$$

by the same argument leading from the theorem of the cube [Hindry and Silverman 2000, A.7.2.1] to Mumford’s formula [Hindry and Silverman 2000, A.7.2.5].

3.2. Further comments on integrality. Let $\mathcal{X} \rightarrow \mathcal{C}$ be a fibered surface with generic fiber X/K and assume X has a K -rational point. Let A be the Jacobian J_X . In [Berger et al. 2015, Proposition 7.2], we proved that the rational number

$$\frac{|\mathrm{NS}(\mathcal{X})_{\mathrm{tor}}|^2}{|A(K)_{\mathrm{tor}}|^2} \tau(A) \mathrm{Reg}(A) \tag{3.2.1}$$

is an integer. (By the factorization of birational maps into blow-ups and the blow-up formula, $\mathrm{NS}(\mathcal{X})_{\mathrm{tor}}$ is a birational invariant, so the displayed quantity depends only on X and K .)

Note that this bound on the denominator of $\mathrm{Reg}(A)$ is in general stronger than that of Proposition 3.1.1. For example, for the Jacobians studied in [Ulmer 2014b; Berger et al. 2015], (3.2.1) is stronger than Proposition 3.1.1.

When X has genus 1, it is known that $\mathrm{NS}(\mathcal{X})_{\mathrm{tor}}$ is trivial, so (3.2.1) says that

$$\frac{\tau(A)}{|A(K)_{\mathrm{tor}}|^2} \mathrm{Reg}(A) \in \mathbb{Z} \tag{3.2.2}$$

This bound (unlike (3.2.1)) makes sense for general abelian varieties, and it is reasonable to ask whether it holds in general. In the rest of this subsection, we sketch a proof that (3.2.2) does not hold in general, not even for Jacobians over $\mathbb{F}_q(t)$.

Let \mathcal{Y} be a classical Enriques surface over \mathbb{F}_p . It is known that

$$\text{NS}(\mathcal{Y})_{\text{tor}} \cong \mathbb{Z}/2\mathbb{Z}, \quad \text{NS}(\mathcal{Y})/\text{tor} \cong \mathbb{Z}^{10}, \quad \text{and} \quad \det(\text{NS}(\mathcal{Y})) = 1;$$

see [Cossec and Dolgachev 1989].

Next, embed \mathcal{Y} in some projective space and take a Lefschetz pencil, extending \mathbb{F}_p to \mathbb{F}_q if necessary. Let \mathcal{X} be the result of blowing up \mathcal{Y} at the base points of the pencil. Thus we have $\pi : \mathcal{X} \rightarrow \mathbb{P}^1$ over \mathbb{F}_q whose fibers are irreducible and either smooth or with single a node. Moreover π has a section. Choose such a section O and a fiber F . We have intersection pairings $O^2 = -1$, $F^2 = 0$, and $F \cdot O = 1$. Also, the Néron–Severi groups satisfy

$$\text{NS}(\mathcal{X}) = \text{NS}(\mathcal{Y}) \oplus \langle -1 \rangle^d$$

where the direct sum is orthogonal, $\langle -1 \rangle$ stands for a copy of \mathbb{Z} whose generator has self-intersection -1 , and d is the number of blow-ups. Thus $\det(\text{NS}(\mathcal{X})) = 1$.

Let $X/K = \mathbb{F}_q(t)$ be the generic fiber of π . This is a smooth curve with a K -rational point. Let $A = J_X$ be its Jacobian. We will see shortly that A is a counterexample to (3.2.2).

Since $\text{Pic}^0(\mathcal{X}) = \text{Pic}^0(\mathcal{Y}) = 0$, we have $\text{Tr}_{K/\mathbb{F}_q}(A) = 0$. The Shioda–Tate theorem gives an exact sequence

$$0 \rightarrow (\mathbb{Z}O + \mathbb{Z}F) \rightarrow \text{NS}(\mathcal{X}) \rightarrow A(K) \rightarrow 0.$$

Moreover, the fact that π has irreducible fibers implies that there is a splitting $A(K) \rightarrow \text{NS}(\mathcal{X})$ which sends the canonical height (divided by $\log q$) to the intersection pairing on $\text{NS}(\mathcal{X})$. It follows from the intersection formulas for O and F noted above that

$$\text{Reg}(A) := \det(A(K)/\text{tor}) = \det(\text{NS}(\mathcal{X})/\text{tor}) = 1.$$

Since π has irreducible fibers, $\tau(A) = 1$. The Shioda–Tate exact sequence above shows that $A(K)_{\text{tor}}$ has order at least 2 (in fact, exactly 2), so

$$\frac{\tau(A)}{|A(K)_{\text{tor}}|^2} \text{Reg}(A) = \frac{1}{4}.$$

Thus (3.2.2). fails for A .

3.3. Lower bounds on Brauer–Siegel ratio from integrality. We now state the main consequence for the Brauer–Siegel ratio of our Proposition 3.1.1.

Proposition 3.3.1. *Let A_d be a family of abelian varieties over K with $H(A_d) \rightarrow \infty$. Assume that $\tau(A_d) = O(H(A_d)^\epsilon)$ for all $\epsilon > 0$. Then $\liminf \text{BS}(A_d) \geq 0$.*

Proof. Noting that $|\text{III}(A_d)|$ is a positive integer and is therefore ≥ 1 , we have that

$$\log(|\text{III}(A_d)| \text{Reg}(A_d)) \geq \log(\text{Reg}(A_d)).$$

Proposition 3.1.1 implies that

$$\log(\text{Reg}(A_d)) \geq -\log(\tau(A_d)).$$

It follows from the hypothesis $\tau(A_d) = O(H(A_d)^\epsilon)$ that

$$\text{BS}(A_d) = \frac{\log(|\text{III}(A_d)| \text{Reg}(A_d))}{\log(H(A_d))} \geq \frac{-\log(\tau(A_d))}{\log(H(A_d))}$$

has $\liminf \geq 0$ as $d \rightarrow \infty$. □

Corollary 3.3.2. *If A_d is a family of abelian varieties over K such that $H(A_d) \rightarrow \infty$, then in any of the following situations $\liminf \text{BS}(A_d) \geq 0$:*

- (1) $\dim(A_d) = 1$ for all d .
- (2) A is an abelian variety over $K = \mathbb{F}_q(t)$, A_d is the family associated to A and the Kummer tower, and A has semistable reduction at $t = 0$ and $t = \infty$.
- (3) A is an abelian variety over $K = \mathbb{F}_q(t)$, A_d is the family associated to A and the Artin–Schreier tower, and A has semistable reduction at $t = \infty$.
- (4) A is an abelian variety over $K = \mathbb{F}_q(E)$, and A_d is the family associated to A and the division tower.
- (5) A is an abelian variety over $K = \mathbb{F}_q(t)$, A_d is the family associated to A and the PGL_2 tower, and A has semistable reduction at $t = 0$ and $t = 1$.

Proof. This is immediate from Lemma 2.2.1, Corollary 2.5.2, and Proposition 3.3.1. □

4. Lower bounds via the dimension of the Tate–Shafarevich functor

In this section, we assume that the conjecture of Birch and Swinnerton-Dyer (more precisely, the finiteness of $\text{III}(A)$) holds for all abelian varieties considered. Given an abelian variety A over $K = \mathbb{F}_q(C)$, we will consider the functor from finite extensions of \mathbb{F}_q to groups given by

$$\mathbb{F}_{q^n} \mapsto \text{III}(A \times_{\mathbb{F}_q(C)} \mathbb{F}_{q^n}(C))$$

and we will show that the dimension of this functor (to be defined below) gives information on the Brauer–Siegel ratio of A over K . This technical device will be extremely convenient as it allows us to bound the Brauer–Siegel ratio without considering the regulator.

Proposition/Definition 4.1. *For each positive integer n , let $K_n := \mathbb{F}_{q^n}(C)$. Given an abelian variety A over $K = K_1$, write A/K_n for $A \times_K K_n$. Then the limit*

$$\lim_{n \rightarrow \infty} \frac{\log |\text{III}(A/K_n)[p^\infty]|}{\log(q^n)}$$

exists and is an integer. We call it the dimension of $\text{III}(A)$, and denote it $\dim \text{III}(A)$.

The proof of the proposition will be given later in this section, after giving a formula for $\dim \text{III}(A)$ in terms of the L -function of A . We give a justification of the terminology “dimension” in Remarks 4.3 below.

In order to state a formula for $\dim \text{III}(A)$, we recall some well-known results on the L -function $L(A, s)$. Let $A_0 = \text{Tr}_{K/k}(A)$ be the K/k -trace of A , an abelian variety over k (where as usual $k = \mathbb{F}_q$). (See [Conrad 2006] for a modern account of the K/k -trace.) Then $L(A, s)$ has the form

$$L(A, s) = \frac{P(q^{-s})}{Q(q^{-s})Q(q^{1-s})}$$

where P and Q are polynomials with the following properties:

- (1) $P(T) = \prod_i (1 - \alpha_i T)$ where the α_i are Weil numbers of size q .
- (2) Q has degree $2 \dim(A_0)$ and $Q(T) = \prod_j (1 - \beta_j T)$ where the β_j are the Weil numbers of size $q^{1/2}$ associated to A_0 . (In other words, they are the eigenvalues of Frobenius on $H^1(A_0 \times \overline{\mathbb{F}_q}, \mathbb{Q}_\ell)$ for any $\ell \neq p$.)
- (3) $Q(1) = |A_0(\mathbb{F}_q)|$ and $Q(q^{-1}) = q^{-d_0} |A_0(\mathbb{F}_q)|$.
- (4) Replacing A with A/K_n has the effect of replacing the α_i and β_j with α_i^n and β_j^n .

Let F be the number field generated by the α_i , and choose a prime of F over p with associated valuation v normalized so that $v(q) = 1$. We define the *slopes* associated to A to be the rational numbers $\lambda_i = v(\alpha_i)$. It is known that the set of slopes (with multiplicities) is independent of the choice of v , that $0 \leq \lambda_i \leq 2$ for all i , and that the set of slopes is invariant under $\lambda_i \mapsto 2 - \lambda_i$.

We can now state a formula for the dimension of $\text{III}(A)$.

Proposition 4.2. $\dim \text{III}(A) = \deg(\omega) + \dim(A)(g_C - 1) + \dim(A_0) - \sum_{\lambda_i < 1} (1 - \lambda_i)$.

The last sum is over indices i such that $\lambda_i < 1$.

Before giving the proof of Propositions 4.1 and 4.2, we record an elementary lemma on p -adic numbers.

Lemma 4.2.1. *Let E be a finite extension of \mathbb{Q}_p , let \mathfrak{m} be the maximal ideal of E , and let $\text{ord} : E^\times \rightarrow \mathbb{Z}$ be the valuation of E . If $\gamma \in E^\times$ has $\text{ord}(\gamma) = 0$ and is not a root of unity, then*

$$\text{ord}(1 - \gamma^n) = O(\log n).$$

Proof. First we note that replacing γ with γ^a , we may assume without loss of generality that γ is a 1-unit, i.e., that $\text{ord}(1 - \gamma) > 0$. Next, if $n = p^e m$ with $p \nmid m$, then

$$\frac{1 - \gamma^n}{1 - \gamma^{p^e}} = 1 + \gamma^{p^e} + \dots + \gamma^{p^e(m-1)} \equiv m \not\equiv 0 \pmod{\mathfrak{m}},$$

so $\text{ord}(1 - \gamma^n) = \text{ord}(1 - \gamma^{p^e})$. Thus it suffices to treat the case where $n = p^e$.

We write \exp_p and \log_p for the p -adic exponential and logarithm respectively. (See, e.g., [Koblitz 1984, IV.1] for basic facts on these functions.) For y sufficiently close to 1 (namely for $|y - 1| < |p^{1/(p-1)}|$), we have $y = \exp_p(\log_p(y))$. Also, it follows from the power series definition of \exp_p , the ultrametric

property of E , and the estimate $v_p(n!) \leq n/(p - 1)$ that if $x \neq 0$ and $\text{ord}(x)$ is sufficiently large (e.g., $\text{ord}(x) > 2/(p - 1)$ suffices), then

$$\text{ord}(1 - \exp_p(x)) = \text{ord}(x).$$

Now if e is sufficiently large, then γ^{p^e} is close to 1, and $x = \log_p(\gamma^{p^e}) = p^e \log_p(\gamma)$ has large valuation and is not zero, so we may apply the estimate above to deduce that

$$\text{ord}(1 - \gamma^{p^e}) = \text{ord}(1 - \exp_p(\log_p(\gamma^{p^e}))) = \text{ord}(\log_p(\gamma^{p^e})) = \text{ord}(p^e) + \text{ord}(\log_p(\gamma)).$$

This last quantity is a linear function of e and thus a linear function of $\log(p^e)$, and this proves our claim. □

Proof of Propositions 4.1 and 4.2. We use the leading coefficient part of the BSD conjecture and consider the p -adic valuations of the elements of the formula. For simplicity, we first consider the case where $A_0 := \text{Tr}_{K/k}(A) = 0$ and then discuss the modifications needed to handle the general case at the end.

As a first step, we establish that several factors in the BSD formula do not contribute to the limit in Proposition/Definition 4.1. More precisely, as n varies, $\text{Reg}(A/K_n)$, $\tau(A/K_n)$, and $|A(K_n)_{\text{tor}}| \cdot |\hat{A}(K_n)_{\text{tor}}|$ are bounded. To see that $\text{Reg}(A/K_n)$ is bounded, we note that it is sensitive to the ground field \mathbb{F}_{q^n} only via the Mordell–Weil group $A(K_n)/\text{tor}$. In other words, if $A(K_n)/\text{tor} = A(K_m)/\text{tor}$, then $\text{Reg}(A/K_n) = \text{Reg}(A/K_m)$. This follows from the geometric nature of the definition of Reg (i.e., its definition in terms of intersection numbers). From the Lang–Néron theorem on the finite generation of $A(K\overline{\mathbb{F}}_q)$, it follows that there are only finitely many possibilities for $A(K_n)/\text{tor}$, so only finitely many possibilities for $\text{Reg}(A/K_n)$. It also follows that $|A(K_n)_{\text{tor}}|$ and $|\hat{A}(K_n)_{\text{tor}}|$ are bounded. (Our use of the Lang–Néron theorem here depends on the assumption that $A_0 = 0$.) Similarly, since the orders of the component groups of the fibers of the Néron model of A over $\overline{\mathbb{F}}_q(\mathbb{C})$ are bounded, there are only finitely possibilities for $\tau(A/K_n)$. Finally, we note that the geometric quantities $\text{deg}(\omega)$, $\dim(A)$, and g_C do not vary with n .

Write $L^*(A/K_n)_p$ for the p -part of the rational number $L^*(A/K_n)$. Then the BSD formula and the remarks above imply that

$$\begin{aligned} \lim_{n \rightarrow \infty} \frac{\log|\text{III}(A/K_n)[p^\infty]|}{\log(q^n)} &= \lim_{n \rightarrow \infty} \frac{\log(L^*(A/K_n)_p q^{n(\text{deg}(\omega) + \dim(A)(g_C - 1))})}{\log(q^n)} \\ &= \lim_{n \rightarrow \infty} \frac{\log(L^*(A/K_n)_p)}{\log(q^n)} + \text{deg}(\omega) + \dim(A)(g_C - 1). \end{aligned}$$

Thus to complete the proof of the existence of the limit in Proposition/Definition 4.1 and the formula of Proposition 4.2 in the case $A_0 = 0$, we need only check that

$$\lim_{n \rightarrow \infty} \frac{\log(L^*(A/K_n)_p)}{\log q^n} = \sum_{\lambda_i < 1} (\lambda_i - 1).$$

Again under the assumption that $A_0 = 0$, we have

$$L^*(A/K_n) = \prod_i' (1 - (\alpha_i/q)^n)$$

where \prod_i' is the product over indices i such that $(\alpha_i/q)^n \neq 1$. We view the right hand side as an element of the number field F introduced above to define the slopes, and we let E (as in Lemma 4.2.1) be the completion of F at the chosen prime of F over p . If $\lambda = v(\alpha_i) < 1$, then

$$v(1 - (\alpha_i/q)^n) = v((\alpha_i/q)^n) = n(\lambda_i - 1),$$

whereas if $\lambda_i > 1$, then

$$v(1 - (\alpha_i/q)^n) = v(1) = 0.$$

In the intermediate case where $\lambda_i = 1$, there are two cases: if α_i/q is not a root of unity, then Lemma 4.2.1 implies that

$$v(1 - (\alpha_i/q)^n) = O(\log n).$$

If α_i/q is a root of unity, then there are only finitely many possibilities for $v(1 - (\alpha_i/q)^n)$ with $(\alpha_i/q)^n \neq 1$, and if $(\alpha_i/q)^n = 1$, then it does not contribute to $L^*(A/K_n)$. Taking the product over i , we find that

$$\lim_{n \rightarrow \infty} \frac{\log(L^*(A/K_n)_p)}{\log q^n} = \sum_{\lambda_i < 1} (\lambda_i - 1).$$

This establishes the formula in Proposition 4.2.

Since the break points of a Newton polygon have integer coordinates, $\sum_{\lambda_i < 1} (\lambda_i - 1)$ is an integer. In the case $A_0 = 0$, we have thus established that the limit in Proposition/Definition 4.1 exists and is an integer, and we have established the formula in Proposition 4.2 for the limit, i.e., for $\dim \text{III}(A)$.

In case $A_0 = \text{Tr}_{K/k}(A)$ is nonzero, the L -function is more complicated, the torsion is not uniformly bounded, and we have to be slightly more careful with the regulator. Here are the details: The Lang–Néron theorem says that $A(K\overline{\mathbb{F}}_q)/A_0(\overline{\mathbb{F}}_q)$ is finitely generated. This implies that there are only finitely many possibilities for $A(K_n)/A_0(\mathbb{F}_{q^n})$ and for the regulator (since $A(K_n)/\text{tor}$ is a quotient of $A(K_n)/A_0(\mathbb{F}_{q^n})$). Moreover,

$$|A(K_n)_{\text{tor}}| = |(A(K_n)/A_0(\mathbb{F}_{q^n}))_{\text{tor}}| \cdot |A_0(\mathbb{F}_{q^n})_{\text{tor}}|$$

and similarly for \hat{A} . On the other hand, writing

$$L(A, s) = \frac{P(q^{-s})}{Q(q^{-s})Q(q^{1-s})} = \frac{\prod_i (1 - \alpha_i q^{-s})}{\prod_j (1 - \beta_j q^{-s})(1 - \beta_j q^{1-s})},$$

we have that

$$L^*(A/K_n) = \frac{\prod_{(\alpha_i/q)^n \neq 1} (1 - (\alpha_i/q)^n)}{\prod_j (1 - (\beta_j/q)^n)(1 - \beta_j^n)}.$$

The denominator is

$$q^{-n \dim(A_0)} |A_0(\mathbb{F}_{q^n})|^2 = q^{-n \dim(A_0)} |A_0(\mathbb{F}_{q^n})| \cdot |\hat{A}_0(\mathbb{F}_{q^n})|$$

so the ratio

$$\frac{|A(K_n)_{\text{tor}}| \cdot |\hat{A}(K_n)_{\text{tor}}|}{\prod_j (1 - (\beta_j/q)^n)(1 - \beta_j^n)} = q^{n \dim(A_0)} |(A(K_n)/A_0(\mathbb{F}_{q^n}))_{\text{tor}}| \cdot |(\hat{A}(K_n)/\hat{A}_0(\mathbb{F}_{q^n}))_{\text{tor}}|$$

is $q^{n \dim(A_0)}$ times a quantity which is bounded as n varies. It then follows that

$$\lim_{n \rightarrow \infty} \frac{\log(|A(K_n)_{\text{tor}}| \cdot |\hat{A}(K_n)_{\text{tor}}| \cdot L^*(A/K_n)_p)}{\log q^n} = \dim(A_0) + \sum_{\lambda_i < 1} (\lambda_i - 1).$$

Therefore,

$$\begin{aligned} \lim_{n \rightarrow \infty} \frac{\log |\text{III}(A/K_n)|}{\log(q^n)} &= \lim_{n \rightarrow \infty} \frac{\log(|A(K_n)_{\text{tor}}| \cdot |\hat{A}(K_n)_{\text{tor}}| \cdot L^*(A/K_n) q^{n(\deg(\omega) + \dim(A)(g_c - 1))})}{\log(q^n)} \\ &= \dim(A_0) + \sum_{\lambda_i < 1} (\lambda_i - 1) + \deg(\omega) + \dim(A)(g_c - 1). \end{aligned}$$

This completes the proof of Propositions 4.1 and 4.2. □

Remarks 4.3. (1) In our applications, we will compute $\dim \text{III}(A)$ directly from its definition using crystalline methods. Proposition 4.2 suggests that these methods will succeed exactly in those situations where one can compute the slopes λ_i , i.e., exactly in the cases where the methods of Hindry–Pacheco and Griffon succeed.

- (2) We explain why the terminology “dimension of $\text{III}(A)$ ” is reasonable. Assume that A is a Jacobian. If $\text{Sel}(A, p^m)$ denotes the Selmer group for multiplication by p^m on A , then it is known that the functor $\mathbb{F}_{q^n} \mapsto \text{Sel}(A \times_K K\mathbb{F}_{q^n}, p^m)$ from finite extensions of \mathbb{F}_q to groups is represented by a group scheme which is an extension of an étale group scheme by a unipotent connected quasisubalgebraic group $U[p^m]$, and the dimension of $U[p^m]$ is constant for large m [Artin 1974]. (One may even replace “finite extensions of \mathbb{F}_q ” with “affine perfect \mathbb{F}_q -schemes,” but unfortunately, not with “general affine schemes.”) Since the order of $A(K\mathbb{F}_{q^n})/p^m A(K\mathbb{F}_{q^n})$ is bounded for varying n , we may detect the dimension of $U[p^m]$ by computing the order of $\text{III}(A \times K\mathbb{F}_{q^n})[p^m]$ asymptotically as $n \rightarrow \infty$. Thus $\dim \text{III}(A)$ as we have defined it in this paper is equal to the dimension of the unipotent quasisubalgebraic group $U[p^m]$. (Note however that $\mathbb{F}_{q^n} \mapsto \text{III}(A \times_K K\mathbb{F}_{q^n})[p^\infty]$ is not in general represented by a group scheme.)
- (3) The formula in Proposition 4.2 for $\dim \text{III}(A)$ is proven using the BSD formula. Conversely, in case A is a Jacobian, Milne [1975, §7] computes the dimension of the group scheme mentioned in the previous remark, and this calculation is a key input into his proof of the leading coefficient formula of the BSD and Artin–Tate conjectures. Our approach is thus somewhat ahistorical, but it is elementary (modulo the BSD conjecture) and completely general.

- (4) In the case where A is a Jacobian, the formula of Proposition 4.2 is equivalent to the formula of Milne for the unipotent group scheme mentioned above, i.e., to the last displayed equation in [Milne 1975, §7].
- (5) The proof of Proposition 4.2 suggests that $\dim \text{III}(A)$ can be viewed as an analog of the Iwasawa μ -invariant.
- (6) If $K_n = \mathbb{F}_{q^n}(C)$, A is an abelian variety over K_1 with $\deg(\omega_A) > 0$, and we define the “ p -Brauer–Siegel ratio of A ” by

$$\text{BS}_p(A) := \frac{\log(R|\text{III}(A)|)_p}{\log H(A)}$$

where $(x)_p$ denotes the p -part of the rational number x , then we have

$$\lim_{n \rightarrow \infty} \text{BS}_p(A/K_n) = \frac{\dim \text{III}(A)}{\deg(\omega_A)}.$$

This gives an interpretation of $\dim \text{III}$ in terms of a modified Brauer–Siegel ratio.

In situations where we can control $\tau(A)$, the following proposition gives a tool to bound the Brauer–Siegel ratio of A from below.

Proposition 4.4. *We have*

$$\frac{\log(|\text{III}(A)| \text{Reg}(A)\tau(A))}{\log(q)} \geq \dim \text{III}(A).$$

Proof. We keep the notation of the proof of Proposition 4.2. In particular, A_0 denotes the K/k trace of A . Using the BSD formula and estimating the denominator of $L^*(A)$, we have

$$\begin{aligned} |\text{III}(A)| \text{Reg}(A)\tau(A) &\geq \frac{|\text{III}(A)| \text{Reg}(A)\tau(A)}{|(A(K_n)/A_0(\mathbb{F}_{q^n}))_{\text{tor}}| \cdot |(\hat{A}(K_n)/\hat{A}_0(\mathbb{F}_{q^n}))_{\text{tor}}|} \\ &= |A_0(\mathbb{F}_{q^n})| \cdot |\hat{A}_0(\mathbb{F}_{q^n})| L^*(A) q^{\deg(\omega) + \dim(A)(g_C - 1)} \\ &\geq q^{\deg(\omega) + \dim(A)(g_C - 1) + \dim(A_0) - \sum(1 - \lambda_i)} \\ &= q^{\dim \text{III}(A)} \end{aligned}$$

and this yields the proposition. □

Remark 4.5. The bound of the proposition is more subtle than it may seem at first: $\dim \text{III}(A)$ is defined in terms of the asymptotic growth of $\text{III}(A)$ as the ground field grows (i.e., replacing \mathbb{F}_q with \mathbb{F}_{q^n}), whereas the left-hand side of the inequality concerns invariants over the given ground field \mathbb{F}_q . In fact, a lower bound on the dimension of $\text{III}(A)$ is not sufficient to give nontrivial lower bounds on $\text{III}(A)$ itself. (This is related to the nonrepresentability of III mentioned above.) For example, if E denotes the Legendre curve studied in [Ulmer 2014b] over $K = \mathbb{F}_{p^{2f}}(t^{1/(p^f+1)})$, then [Ulmer 2014b, Corollary 10.2] shows that $\dim \text{III}(E) = (p^f - 1)/2$, whereas [Ulmer 2014c, Theorem 1.1] shows that when $f \leq 2$, $\text{III}(E)$ is trivial. This example also shows that the second inequality displayed above is sharp.

Next, we state the result which is our main motivation for considering $\dim \text{III}(A)$.

Proposition 4.6. *Let A_d be a family of abelian varieties over K with $H(A_d) \rightarrow \infty$. Assume that $\tau(A_d) = O(H(A_d)^\epsilon)$ for all $\epsilon > 0$. Then*

$$\liminf_{d \rightarrow \infty} \text{BS}(A_d) \geq \liminf_{d \rightarrow \infty} \frac{\dim \text{III}(A_d)}{\deg(\omega_{A_d})}.$$

Proof. The hypothesis $\tau(A_d) = O(H(A_d)^\epsilon)$ for all $\epsilon > 0$ implies that

$$\lim_{d \rightarrow \infty} \log(\tau(A_d)) / \log(H(A_d)) = 0,$$

so the proposition follows immediately from the estimate of Proposition 4.4. □

Corollary 4.7. *If A_d is a family of abelian varieties over K such that $H(A_d) \rightarrow \infty$, then*

$$\liminf_{d \rightarrow \infty} \text{BS}(A_d) \geq \liminf_{d \rightarrow \infty} \frac{\dim \text{III}(A_d)}{\deg(\omega_{A_d})}$$

in any of the following situations:

- (1) $\dim(A_d) = 1$ for all n
- (2) A is an abelian variety over $K = \mathbb{F}_q(t)$, A_d is the family associated to the Kummer tower, and A has semistable reduction at $t = 0$ and $t = \infty$.
- (3) A is an abelian variety over $K = \mathbb{F}_q(t)$, A_d is the family associated to the Artin–Schreier tower, and A has semistable reduction at $t = \infty$.
- (4) A is an abelian variety over $K = \mathbb{F}_q(E)$, and A_d is the family associated to the division tower.
- (5) A is an abelian variety over $K = \mathbb{F}_q(t)$, A_d is the family associated to the PGL_2 tower, and A has semistable reduction at $t = 0$ and $t = 1$.

Proof. This is immediate from Lemma 2.2.1, Corollary 2.5.2, and Proposition 4.6. □

5. Brauer–Siegel ratio and Frobenius

As a first application of our results on the dimension of III, we compute the Brauer–Siegel ratio for sequences of abelian varieties associated to the Frobenius isogeny.

More precisely, let E be an elliptic curve over the function field $K = \mathbb{F}_q(C)$, and for $n \geq 1$, let E_n be the Frobenius base change:

$$E_n := E^{(p^n)} = E \times_K K$$

where the right hand morphism $K \rightarrow K$ is the p^n -power Frobenius.

Our goal is the following result.

Theorem 5.1. *Assume that E is nonisotrivial. Then*

$$\lim_{n \rightarrow \infty} \text{BS}(E_n) = 1.$$

Proof. First we note that since E is nonisotrivial, $H(E_n) \rightarrow \infty$ as $n \rightarrow \infty$. Indeed, the j -invariant of E has a pole, say of order e , at some place of K , so the j invariant of E_n has a pole of order ep^n at the same place. This implies that the degrees of the divisors of one or both of $c_4(E_n)$ and $c_6(E_n)$ also tend to infinity, and this is possible only if $\deg(\omega_{E_n})$ also tends to infinity. Since $H(E_n) = q^{\deg(\omega_{E_n})}$, we have that $H(E_n) \rightarrow \infty$.

Next we note that Proposition 4.2 shows that $\dim \text{III}(E_n) - \deg(\omega_{E_n})$ depends only on the L -function of E_n , indeed only on the slopes of the L -function. Since E and E_n are isogenous, they have the same L -function, so we have

$$\dim \text{III}(E_n) - \deg \omega_{E_n} = \dim \text{III}(E) - \deg \omega_E$$

for all n .

Dividing the last displayed equation by $\deg \omega_{E_n}$ and taking the limit as $n \rightarrow \infty$, we get

$$\frac{\dim \text{III}(E_n)}{\deg \omega_{E_n}} \rightarrow 1$$

since $\deg(\omega_{E_n}) \rightarrow \infty$.

Applying part (1) of Corollary 4.7, we see that $\liminf_{n \rightarrow \infty} \text{BS}(E_n) \geq 1$. On the other hand, by [Hindry and Pacheco 2016, Corollary 1.13], $\limsup_{n \rightarrow \infty} \text{BS}(E_n) \leq 1$, so we find that $\lim_{n \rightarrow \infty} \text{BS}(E_n) = 1$, as desired. □

Remark 5.2. The same argument works for an abelian variety A as long as $\deg(\omega_{A^{(p^n)}}) \rightarrow \infty$ with n and $\tau(A^{(p^n)}) = o(H(A^{(p^n)}))$.

Remark 5.3. The theorem says that the product $|\text{III}(E_n)| \text{Reg}(E_n)$ grows with n . Our earlier results on p -descent [Ulmer 1991] can be used to show directly that $\text{III}(E_n)$ grows with n . Full details require an unilluminating consideration of many cases, so we limit ourselves to a sketch in the simplest situation. First, let $V : E^{(p)} \rightarrow E$ be the Verschiebung isogeny, and note that the Selmer group $\text{Sel}(E^{(p)}, p)$ contains $\text{Sel}(E, V)$. Also, let L be the (Galois) extension of K obtained by adjoining the $(p - 1)$ -st root of a Hasse invariant of E , and let $G = \text{Gal}(L/K)$. In [Ulmer 1991, Theorem 3.2 and Lemma 1.4], we computed that

$$\text{Sel}(E, V) \cong \text{Hom}(J_m / \langle \text{cusps} \rangle, \mathbb{Z}/p\mathbb{Z})^G$$

where J_m is the generalized Jacobian of the curve whose function field is L for a “modulus” m related to the places of bad and/or supersingular reduction of E . Rosenlicht showed that J_m is an extension of J by a linear group (see [Serre 1988]), and the unipotent part of this group contributes to the “dimension” of $\text{Sel}(E, V)$ and therefore to $\dim \text{III}(E^{(p)})$. The contribution is roughly the number of zeroes (with multiplicity) of the Hasse invariant, namely $(p - 1) \deg(\omega_E)$ which is approximately $\deg(\omega_{E^{(p)}}) - \deg(\omega)$. Thus we find

$$\dim \text{III}(E^{(p)}) \geq \deg(\omega_{E^{(p)}}) - \deg(\omega),$$

in agreement with what we deduced from Proposition 4.2.

6. Bounding III for a class of Jacobians

In this section, we review a general method for computing the p -part of the Tate–Shafarevich group of certain Jacobians, generalizing our previous work [Ulmer 2014c] on the Legendre elliptic curve. Although these methods suffice to compute the p -part of III on the nose, for simplicity we focus just on \dim III as this is what is needed to bound the Brauer–Siegel ratio from below.

6.1. Jacobians related to products of curves. Let k be the finite field \mathbb{F}_q of characteristic p with q elements. Let \mathcal{C} and \mathcal{D} be curves over k , and let $\mathcal{S} = \mathcal{C} \times_k \mathcal{D}$. Suppose that Δ is a group of k -automorphisms of \mathcal{S} with order prime to p and such that

$$\Delta \subset \text{Aut}_k(\mathcal{C}) \times \text{Aut}_k(\mathcal{D}) \subset \text{Aut}_k(\mathcal{S}).$$

Suppose that the quotient \mathcal{S}/Δ is birational to a smooth, projective surface \mathcal{X} over k and that \mathcal{X} is equipped with a surjective and generically smooth morphism $\pi : \mathcal{X} \rightarrow C$ where C is a smooth projective curve over k . Let $K = k(C)$ and let X be the generic fiber of π , a smooth projective curve over K . We assume that X has a K -rational point. (A vast supply of such data is given in [Berger 2008; Ulmer 2013].)

Let J be the Jacobian of X . We write $\text{Br}(\mathcal{X})$ for the cohomological Brauer group of \mathcal{X} : $\text{Br}(\mathcal{X}) = H^2(\mathcal{X}, \mathbf{G}_m)$.

Proposition 6.2. (1) $\text{III}(J_X)$ and $\text{Br}(\mathcal{X})$ are finite groups.

(2) There is a canonical isomorphism $\text{III}(J_X) \cong \text{Br}(\mathcal{X})$.

(3) There is a canonical isomorphism

$$\text{Br}(\mathcal{X})[p^\infty] \cong (\text{Br}(\mathcal{S})[p^\infty])^\Delta.$$

Proof. In substance, parts (2) and (3) are due to Grothendieck [1968] and part (1) is due to Tate [1966]. The details to deduce the statements here are given in [Ulmer 2014c, §4]. □

6.3. Brauer group of a product of curves. We keep the notation of the preceding subsection. In addition, let $W = W(k)$ be the ring of Witt vectors over k with Frobenius endomorphism σ . We write $H^1(\mathcal{C})$ for the crystalline cohomology $H^1_{\text{crys}}(\mathcal{C}/W)$ and similarly for $H^1(\mathcal{D})$. These are modules over the Dieudonné ring $A = W\{F, V\}$, which is the noncommutative polynomial ring generated over W by symbols F and V with relations $FV = VF = p$, $F\alpha = \sigma(\alpha)F$, and $\alpha V = V\sigma(\alpha)$ for all $\alpha \in W$.

The following crystalline calculation of the p part of the Brauer group of \mathcal{S} is originally due to Dummigan (with additional hypotheses) using results of Milne, and is proven in general in [Ulmer 2014c, §10].

Proposition 6.4. There is a canonical isomorphism

$$\text{Br}(\mathcal{S})[p^n] \cong \frac{\text{Hom}_A(H^1(\mathcal{C})/p^n, H^1(\mathcal{D})/p^n)}{\text{Hom}_A(H^1(\mathcal{C}), H^1(\mathcal{D}))/p^n}$$

which is compatible with the actions of Δ on both sides.

Here Hom_A denotes W -linear homomorphisms which commute with F and V .

Propositions 6.2 and 6.4 give us a powerful tool for bounding $\dim \text{III}(J)$ from below. Recall that this means bounding the growth of the order of $\text{III}(J)$ as we extend the ground field from \mathbb{F}_q to \mathbb{F}_{q^v} . The denominator on the right hand side of the displayed equation in Proposition 6.4 is known to be bounded as v varies (a fact we will see explicitly in Section 8 for the examples we consider), so we have:

Corollary 6.5. *For all sufficiently large n ,*

$$\dim \text{III}(J) = \dim \text{Hom}_A(H^1(\mathcal{C})/p^n, H^1(\mathcal{D})/p^n)^\Delta.$$

Here the \dim on the right-hand side is defined analogously to that on the left:

$$\dim \text{Hom}_A(H^1(\mathcal{C})/p^n, H^1(\mathcal{D})/p^n)^\Delta := \lim_{v \rightarrow \infty} \frac{\log |\text{Hom}_A(H^1(\mathcal{C} \times_k \mathbb{F}_{q^v})/p^n, H^1(\mathcal{D} \times_k \mathbb{F}_{q^v})/p^n)^\Delta|}{\log(q^v)}.$$

Computing the cardinality of the numerator on the right amounts to an interesting exercise in p -linear algebra, at least for certain curves \mathcal{C} and \mathcal{D} . We carry out these exercises in Section 8.

7. Cohomology of Fermat curves

We review some well-known result on the cohomology of Fermat curves.

As usual, let $k = \mathbb{F}_q$ be the finite field of cardinality q and characteristic p . We write \bar{k} for the algebraic closure of k . For a positive integer d relatively prime to p , let F_d be the smooth projective curve over k given by

$$x_0^d + x_1^d + x_2^d = 0.$$

We write μ_d for the group of d -th roots of unity in \bar{k} . There is an evident action of μ_d^3 on $F_d \times_k \bar{k}$ under which $(\zeta_i) \in \mu_d^3$ acts via $x_i \mapsto \zeta_i x_i$, and the diagonal $(\zeta_0 = \zeta_1 = \zeta_2)$ acts trivially, so we have $G := \mu_d^3 / \mu_d \subset \text{Aut } \bar{k}(F_d)$.

Let

$$A = \left\{ (a_0, a_1, a_2) \mid \sum a_i = 0 \right\} \subset (\mathbb{Z}/d\mathbb{Z})^3.$$

Abusively writing ζ both for a root of unity in \bar{k} and for its Teichmüller lift to the Witt vectors $W(\bar{k})$, we may identify A with the character group $\text{Hom}(G, W(\bar{k})^\times)$. Let

$$A' = \{(a_i) \in A \mid a_i \neq 0, i = 0, 1, 2\}.$$

Given $(a_0, a_1, a_2) \in A$, let $\langle a_i/d \rangle$ be the fractional part of \tilde{a}_i/d , where \tilde{a}_i is any representative in \mathbb{Z} of the class a_i . Define subsets A_0 and A_1 as follows:

$$A_0 = \left\{ (a_i) \in A' \mid \left\langle \frac{a_0}{d} \right\rangle + \left\langle \frac{a_1}{d} \right\rangle + \left\langle \frac{a_2}{d} \right\rangle = 2 \right\}$$

and

$$A_1 = \left\{ (a_i) \in A' \mid \left\langle \frac{a_0}{d} \right\rangle + \left\langle \frac{a_1}{d} \right\rangle + \left\langle \frac{a_2}{d} \right\rangle = 1 \right\}$$

It is a simple exercise to see that A' is the disjoint union of A_0 and A_1 . Let $\langle p \rangle$ be the subgroup of \mathbb{Q}^\times generated by p . Then $\langle p \rangle$ acts on A' coordinatewise: $p(a_0, a_1, a_2) = (pa_0, pa_1, pa_2)$.

Let $H = H_{crys}^1(F_d/W(k))$ be the crystalline cohomology of F_d equipped with its action of the p -power Frobenius F and Verschiebung V . Then $\bar{H} := H \otimes_{W(k)} W(\bar{k})$ inherits an action of G .

The following summarizes the main results on H . The argument in [Dummigan 1995, §6], stated in the special case where $d = q + 1$, works for general d prime to p .

Proposition 7.1. *There is W -basis $\{e_a\}$ of H indexed by $a \in A'$ with the following properties:*

(1) $F(e_a) = c_a e_{pa}$ where $c_a \in W(k)$ and

$$\text{ord}_p(c_a) = \begin{cases} 0 & \text{if } a \in A_0, \\ 1 & \text{if } a \in A_1. \end{cases}$$

(2) For $(\zeta_i) \in G$ and $a \in A'$,

$$(\zeta_i)e_a = a(\zeta_i)e_a = \zeta_0^{a_0} \zeta_1^{a_1} \zeta_2^{a_2} e_a$$

(an equality in \bar{H}).

7.2. A remark on twists. It is sometimes convenient to work with a different model of the Fermat curve, namely

$$F'_d : y_0^d + y_1^d = y_2^d.$$

This is a twist of F_d in the sense that they F_d and F'_d become isomorphic over \bar{k} via

$$(x_0, x_1, x_2) \mapsto (y_0, y_1, \epsilon y_2)$$

where ϵ is a d -th root of -1 . It follows that Proposition 7.1 holds for F'_d as well, with possibly different constants c_a which nevertheless continue to satisfy the valuation formula in part (1).

7.3. A remark on quotients. If \mathcal{C} is the quotient of F_d by a subgroup of $G' \subset G$, then the crystalline cohomology of \mathcal{C} can be identified with the W -submodule of H generated by the e_a whose indices a are trivial on G' .

For example, the hyperelliptic curve

$$C_{2,d} : y^2 = x^d + 1$$

is the quotient of F'_{2d} by a subgroup of G isomorphic to $\mu_d \times \mu_2$. (If d is even, it is also a quotient of F'_d , but it is more convenient to have a uniform statement.)

More generally, the superelliptic curve

$$C_{r,d} : y^r = x^d + 1$$

is the quotient of F'_{rd} by a subgroup of G isomorphic to $\mu_d \times \mu_r$.

The crystalline cohomology $H_{crys}^1(C_{r,d}/W(k))$ can then be identified with the W -submodule of $H_{crys}^1(F'_{rd}/W(k))$ generated by the e_a where a has the form

$$a = (a_0, a_1, a_2) = (ir, -ir - jd, jd) \quad 0 < i < d, \quad 0 < j < r, \quad ir + jd \not\equiv 0 \pmod{rd}.$$

The set I of such indices has cardinality $(r - 1)(d - 1) - \gcd(r, d) + 1$, and it is the disjoint union $I = I_0 \cup I_1$ where

$$I_0 = I \cap A_0 \cong \{(i, j) \mid 0 < i < d, \ 0 < j < r, \ ir + jd > rd\}$$

and

$$I_1 = I \cap A_1 \cong \{(i, j) \mid 0 < i < d, \ 0 < j < r, \ ir + jd < rd\}.$$

In the case where $r = 2$ we may further simplify this to

$$I_0 \cong \{i \mid \frac{d}{2} < i < d\} \quad \text{and} \quad I_1 \cong \{i \mid 0 < i < \frac{d}{2}\}.$$

These sets, with their action of $\langle p \rangle$, will play a key role in the p -adic exercises that compute $\dim III$ for the Jacobians introduced in Section 6.

8. p -adic exercises

In this section, we carry out the exercises in semilinear algebra needed to compute the dimension of III for several families of abelian varieties.

Let p be a prime and let \mathbb{F}_q be the field of cardinality q and characteristic p . Let $W = W(\mathbb{F}_q)$ be the Witt vectors over \mathbb{F}_q , and let $W_n = W/p^n$. Write σ for the p -power Witt-vector Frobenius. For a positive integer v , we write \mathbb{F}_{q^v} for the field of q^v elements, $W_v = W(\mathbb{F}_{q^v})$ for the corresponding Witt ring, and $W_{n,v}$ for W_v/p^n .

Let $A = W\{F, V\}$ be the Dieudonné ring of noncommutative polynomials in F and V with relations $FV = VF = p$, $F\alpha = \sigma(\alpha)F$, and $\alpha V = V\sigma(\alpha)$ for $\alpha \in W$. Also, let A_v be the ring $W_v\{F, V\}$ with analogous relations.

Let $\langle p \rangle$ be the cyclic subgroup of \mathbb{Q}^\times generated by p .

8.1. Data. Fix a finite set I equipped with an action of $\langle p \rangle$, which we write multiplicatively: $i \mapsto pi$. (In the applications below, I will typically be a subset of $\mathbb{Z}/d\mathbb{Z}$ for some d not divisible by p .) Let M be the free W -module with basis indexed by I :

$$M := \bigoplus_{i \in I} We_i.$$

Write I as a disjoint union $I = I_0 \cup I_1$ and choose elements $c_i \in W$ such that

$$\text{ord}(c_i) = \begin{cases} 0 & \text{if } i \in I_0, \\ 1 & \text{if } i \in I_1. \end{cases}$$

Define a σ -semilinear map $F : M \rightarrow M$ by setting

$$F(e_i) = c_i e_{pi}$$

and a σ^{-1} -semilinear map $V : M \rightarrow M$ by setting

$$V(e_i) = \frac{p}{\sigma^{-1}(c_{i/p})} e_{i/p}.$$

These definitions give M the structure of an A -module, and there is an induced A -module structure on $M_n := M \otimes_W W_n$. Parallel definitions make $M_v := M \otimes_W W_v$ and $M_{n,v} := M \otimes_W W_{n,v}$ into A_v -modules.

Fix another finite set J equipped with an action of $\langle p \rangle$, write J as a disjoint union $J = J_0 \cup J_1$, and choose elements $d_j \in W$ with

$$\text{ord}(d_j) = \begin{cases} 0 & \text{if } j \in J_0, \\ 1 & \text{if } j \in J_1. \end{cases}$$

Define

$$N := \bigoplus_{j \in J} W f_j,$$

with semilinear maps $F : N \rightarrow N$ and $V : N \rightarrow N$ defined by

$$F(f_j) = d_j f_{pj}$$

and

$$V(f_j) = \frac{p}{\sigma^{-1}(d_{j/p})} f_{j/p}.$$

Then N and $N_n := N \otimes_W W_n$ are A -modules, and parallel definitions make $N_v := N \otimes_W W_v$ and $N_{n,v} := N \otimes_W W_{n,v}$ into A_v -modules.

Let $\langle p \rangle$ act on $I \times J$ diagonally, and let O be the set of orbits of this action. For an orbit $o \in O$, define

$$d(o) := \min(|(I_0 \times J_1) \cap o|, |(I_1 \times J_0) \cap o|).$$

Consider $\text{Hom}_{W_v}(N_v, M_v)$, a free W_v -module with basis φ_{ij} defined by

$$\varphi_{ij}(f_{j'}) = \begin{cases} e_i & \text{if } j' = j, \\ 0 & \text{if } j' \neq j. \end{cases}$$

These elements induce elements of

$$\text{Hom}_{W_v}(N_{n,v}, M_{n,v}) = \text{Hom}_{W_v}(N_v, M_v)/p^n$$

which form a basis over $W_{n,v}$ and which we abusively also denote φ_{ij} .

8.2. Statement. Our main objects of study in this section are the subgroups

$$H_v := \text{Hom}_{A_v}(N_v, M_v) \subset \text{Hom}_{W_v}(N_v, M_v)$$

and

$$H_{n,v} := \text{Hom}_{A_v}(N_{n,v}, M_{n,v}) \subset \text{Hom}_{W_v}(N_{n,v}, M_{n,v})$$

consisting of A_v -module homomorphisms, i.e., homomorphisms φ such that $F \circ \varphi = \varphi \circ F$ and $V \circ \varphi = \varphi \circ V$.

To state the results, we first decompose the groups of interest into components indexed by the set of orbits O . For $o \in O$, let

$$\text{Hom}_{W_v}(N_v, M_v)^o := \left\{ \varphi = \sum_{i,j} \alpha_{i,j} \varphi_{i,j} \mid \alpha_{i,j} = 0 \text{ for all } (i,j) \notin o \right\}$$

and

$$\text{Hom}_{W_v}(N_{n,v}, M_{n,v})^o := \left\{ \varphi = \sum_{i,j} \alpha_{i,j} \varphi_{i,j} \mid \alpha_{i,j} = 0 \text{ for all } (i,j) \notin o \right\}.$$

We define

$$H_v^o := H_v \cap \text{Hom}_{W_v}(N_v, M_v)^o \quad \text{and} \quad H_{n,v}^o := H_{n,v} \cap \text{Hom}_{W_v}(N_{n,v}, M_{n,v})^o.$$

Here is the main result of this section:

Theorem 8.3. (1) $H_v = \bigoplus_{o \in O} H_v^o$ and $H_{n,v} = \bigoplus_{o \in O} H_{n,v}^o$.

(2) $|H_v^o/p^n|$ is at most $p^{n|o|}$ and in particular is bounded independently of v .

(3) For all sufficiently large n ,

$$\lim_{v \rightarrow \infty} \frac{\log |H_{n,v}^o|}{\log(q^v)} = d(o).$$

Proof. Let

$$\varphi = \sum_{(i,j) \in I \times J} \alpha_{i,j} \varphi_{i,j}$$

be a typical element of $\text{Hom}_{W_v}(N_v, M_v)$ (with $\alpha_{i,j} \in W_v$) or $\text{Hom}_{W_v}(N_{n,v}, M_{n,v})$ (with $\alpha_{i,j} \in W_{n,v}$). Then a straightforward calculation shows that $F \circ \varphi = \varphi \circ F$ if and only if

$$c_i \sigma(\alpha_{i,j}) = d_j \alpha_{p(i,j)} \quad \text{for all } (i,j) \in I \times J, \tag{8.3.1}$$

and $V \circ \varphi = \varphi \circ V$ if and only if

$$\left(\frac{p}{d_j}\right) \sigma(\alpha_{i,j}) = \left(\frac{p}{c_i}\right) \alpha_{p(i,j)} \quad \text{for all } (i,j) \in I \times J. \tag{8.3.2}$$

Defining

$$\varphi^o = \sum_{(i,j) \in o} \alpha_{i,j} \varphi_{i,j},$$

it is clear that $\varphi^o \in H_v^o$ or $H_{n,v}^o$ and that $\varphi = \sum_{o \in O} \varphi^o$. This shows that $H_v = \sum_{o \in O} H_v^o$ and $H_{n,v} = \sum_{o \in O} H_{n,v}^o$, and it is immediate that the sums are direct. This proves part (1) of the theorem.

For part (2), take a typical element $\varphi^o = \sum_{(i,j) \in o} \alpha_{i,j} \varphi_{i,j}$ of H_v^o . Since W_v is torsion-free, the conditions (8.3.1) and (8.3.2) are equivalent, so we focus on (8.3.1). Fix a base point $(i_0, j_0) \in o$ and note that α_{i_0, j_0} determines the other coefficients $\alpha_{i,j}$ with $(i, j) \in o$ by repeatedly using (8.3.1). Indeed, we have

$$\begin{aligned} c_{i_0} \sigma(\alpha_{i_0, j_0}) &= d_{j_0} \alpha_{p(i_0, j_0)} \\ c_{p i_0} \sigma(c_{i_0}) \sigma^2(\alpha_{i_0, j_0}) &= d_{p j_0} \sigma(d_{j_0}) \alpha_{p^2(i_0, j_0)} \\ &\vdots \\ c_{p^{|o|-1} i_0} \sigma(c_{p^{|o|-2} i_0}) \cdots \sigma^{|o|-1}(c_{i_0}) \sigma^{|o|}(\alpha_{i_0, j_0}) &= d_{p^{|o|-1} j_0} \sigma(d_{p^{|o|-2} j_0}) \cdots \sigma^{|o|-1}(d_{j_0}) \alpha_{i_0, j_0} \end{aligned}$$

Here $|o|$ is the cardinality of o and in the last line we use that $p^{|o|}(i_0, j_0) = (i_0, j_0)$. Moreover, α_{i_0, j_0} determines a solution to (8.3.1) only if the last displayed line holds. (There may be other integrality conditions, but they are not important for our argument.) If the valuations of

$$c_{p^{|o|-1} i_0} \sigma(c_{p^{|o|-2} i_0}) \cdots \sigma^{|o|-1}(c_{i_0}) \quad \text{and} \quad d_{p^{|o|-1} j_0} \sigma(d_{p^{|o|-2} j_0}) \cdots \sigma^{|o|-1}(d_{j_0})$$

are distinct, then it is clear that the only solution is $\alpha_{i_0, j_0} = 0$. On the other hand, if the valuations are the same, the last equation is equivalent to one of the form $\sigma^{|o|}(\alpha_{i_0, j_0}) = \gamma \alpha_{i_0, j_0}$ where $\gamma \in W_v$ is a unit. Written in terms of Witt vector components, this last equation is a polynomial of degree $p^{|o|}$ in each component of α_{i_0, j_0} (with coefficients given by γ and the lower Witt components of α_{i_0, j_0}). Therefore, taking α_{i_0, j_0} modulo p^n , there are at most $p^{n|o|}$ solutions, and this proves part (2) of the theorem.

We now turn to part (3) of the theorem, which follows from a somewhat more elaborate version of the calculation of [Ulmer 2014c, §7, §10]. Namely, we fix an orbit o and consider (8.3.1) and (8.3.2) with $(i, j) \in o$ and $\alpha_{i,j} \in W_{n,v}$. These are the equations defining $H_{n,v}^o$ as a subset of $\text{Hom}_{W_v}(N_{n,v}, M_{n,v})^o$, and analyzing them will allow us to estimate the size of $H_{n,v}^o$.

Fix an orbit $o \in O$ and a base point $(i_0, j_0) \in o$. We associate a word w on the alphabet $\{u, l, m\}$ to o as follows: $w = w_1 w_2 \cdots w_{|o|}$ where

$$w_\ell = \begin{cases} u & \text{if } p^{\ell-1}(i_0, j_0) \in I_1 \times J_0, \\ l & \text{if } p^{\ell-1}(i_0, j_0) \in I_0 \times J_1, \\ m & \text{if } p^{\ell-1}(i_0, j_0) \in (I_0 \times J_0) \cup (I_1 \times J_1). \end{cases}$$

Changing the base point changes w by a cyclic permutation. Note that $d(o)$ is the smaller of the number of appearances of l or u in w .

The motivation for these letters is as follows: If $w_\ell = u$, then in (8.3.1) and (8.3.2) for $(i, j) = p^{\ell-1}(i_0, j_0)$, d_j is a unit and p/c_j is a unit. It follows that the two equations are equivalent and either of them determines $\alpha_{p^\ell(i_0, j_0)}$ in terms of $\alpha_{p^{\ell-1}(i_0, j_0)}$. i.e., the “upper” $\alpha_{p^\ell(i_0, j_0)}$ is determined by the “lower” $\alpha_{p^{\ell-1}(i_0, j_0)}$. Similarly, if $w_\ell = l$, the “lower” $\alpha_{p^{\ell-1}(i_0, j_0)}$ is determined by the “upper” $\alpha_{p^\ell(i_0, j_0)}$. Finally, if $w_\ell = m$, then one of (8.3.1) and (8.3.2) implies other and shows that $\alpha_{p^{\ell-1}(i_0, j_0)}$ and $\alpha_{p^\ell(i_0, j_0)}$ determine

each other. We will use these observations to eliminate most of the variables in the systems (8.3.1) and (8.3.2), and use the simplified system to estimate the size of $H_{n,v}^o$ and prove part (3) of the theorem.

We first deal with three degenerate cases, namely those where w is a power of m , or has no letters l , or has no letters u . In all three cases, $d(o) = 0$, so it will suffice to prove that $|H_{n,v}^o|$ is bounded independently of v . If $w = m^{|\sigma|}$, then α_{i_0, j_0} determines all of the $\alpha_{p^\ell(i_0, j_0)}$, and the system ((8.3.1)–(8.3.2)) reduces to a single equation

$$\sigma^{|\sigma|} \alpha_{i_0, j_0} = \gamma \alpha_{i_0, j_0}$$

where $\gamma \in W$ is a unit. This is easily seen to have at most $p^{n|\sigma|}$ solutions for any v , as desired. If w contains no letters l , then again α_{i_0, j_0} determines all of the $\alpha_{p^\ell(i_0, j_0)}$, and the system ((8.3.1)–(8.3.2)) reduces to a single equation

$$p^e \sigma^{|\sigma|} \alpha_{i_0, j_0} = \gamma \alpha_{i_0, j_0}$$

where $e \geq 0$ and $\gamma \in W$ is a unit. (Here e is the number of appearances of u in w .) If $e = 0$, we are in the previous case, and the equation has at most $p^{n|\sigma|}$ solutions for any v , whereas if $e > 0$, then this equation is easily seen to have no solutions. Finally, if w has no letter u , then the system again reduces to a single equation of the form

$$\sigma^{|\sigma|} \alpha_{i_0, j_0} = \gamma p^e \alpha_{i_0, j_0}$$

which has at most $p^{n|\sigma|}$ solutions for any v if $e = 0$ and has no solutions if $e > 0$.

For the rest of the argument, we may assume w contains at least one u and at least one l . Define a function $a : \{0, 1, \dots, |\sigma|\} \rightarrow \mathbb{Z}$ by setting $a(0) = 0$ and

$$a(\ell) = a(\ell - 1) + \begin{cases} 1 & \text{if } w_\ell = u, \\ -1 & \text{if } w_\ell = l, \\ 0 & \text{if } w_\ell = m. \end{cases}$$

for $1 \leq \ell \leq |\sigma|$.

Define the *height* of o , denoted $ht(o)$, to be the maximum value of a minus the minimum value of a . Note that this is independent of the choice of a base point for o .

We divide into two cases depending on whether $a(|\sigma|) \geq 0$ or $a(|\sigma|) \leq 0$.

If $a(|\sigma|) \geq 0$, we may change base point so that $0 = a(0)$ is the minimum value of a (i.e., $a(\ell) \geq 0$ for $0 \leq \ell \leq |\sigma|$) and $a(|\sigma| - 1) > a(|\sigma|)$. Indeed, start with any base point (i_0, j_0) and let ℓ_0 be such that $a(\ell_0)$ is minimum among the $a(\ell)$. Then replacing (i_0, j_0) with $(i_1, j_1) = p^{\ell_0}(i_0, j_0)$ ensures that $a(\ell) \geq 0$ for all $0 \leq \ell \leq |\sigma|$. If the new word w ends with m or u , we may replace (i_1, j_1) with $p^{-1}(i_1, j_1)$ without affecting the inequality $a(\ell) \geq 0$. Iterate until the last letter is l , thus yielding the desired base point. We fix such as base point and denote it (i_0, j_0) .

Choose

$$0 = \ell_0 < \ell^0 < \ell_1 < \ell^1 \dots < \ell^{k-1} < \ell_k = |\sigma|$$

such that a is nondecreasing on $\{\ell_\lambda, \dots, \ell^\lambda\}$ and nonincreasing on $\{\ell^\lambda, \dots, \ell_{\lambda+1}\}$ for $0 \leq \lambda \leq k-1$. In particular, the ℓ_λ are the arguments of local minima of a . Now let

$$\beta_\lambda = \alpha_{p^{\ell_\lambda(i_0, j_0)}} \quad 0 \leq \lambda \leq k.$$

(Note that $\beta_k = \beta_0$.) Then the motivating remarks above about the letters u, l, m show that the β_λ determine all the $\alpha_{i,j}$ with $(i, j) \in o$. The equations (8.3.1) and (8.3.2) hold if and only if the β_λ satisfy the system:

$$\begin{aligned} p^{e_1} \sigma^{\ell_1 - \ell_0} \beta_0 &= \gamma_1 p^{e_2} \beta_1 \\ p^{e_3} \sigma^{\ell_2 - \ell_1} \beta_1 &= \gamma_2 p^{e_4} \beta_2 \\ &\vdots \\ p^{e_{2k-1}} \sigma^{\ell_k - \ell_{k-1}} \beta_{k-1} &= \gamma_k p^{e_{2k}} \beta_k \end{aligned} \tag{8.3.3}$$

where

$$e_{2\lambda-1} = \# \text{ of appearances of } u \text{ in the subword } w_{\ell_{\lambda-1}+1} \cdots w_{\ell_\lambda}$$

$$e_{2\lambda} = \# \text{ of appearances of } l \text{ in the subword } w_{\ell_{\lambda-1}+1} \cdots w_{\ell_\lambda}$$

and the units γ_λ are defined by

$$\gamma_\lambda = p^{e_{2\lambda-1} - e_{2\lambda}} \prod_{\ell=\ell_{\lambda-1}}^{\ell_\lambda-1} \sigma^{\ell_\lambda-1-\ell} \left(\frac{d p^\ell j_0}{c p^\ell i_0} \right).$$

To recap, the assignment $\varphi \mapsto (\beta_\lambda)$ gives an injection $H_{n,v}^o \hookrightarrow W_{n,v}^k$ whose image is the set of solutions to equations (8.3.3). We will finish the proof of part (3) of the theorem by estimating the number of such solutions.

Since the theorem is an assertion about $H_{n,v}^o$ for sufficiently large n , we will assume for the rest of the proof that $n \geq ht(o)$. Then we have an exact sequence

$$0 \rightarrow p^{n-ht(o)} H_{n,v}^o \rightarrow H_{n,v}^o \rightarrow W_{n-ht(o),v}$$

where the right hand map sends a tuple (β_λ) to the reduction modulo $p^{n-ht(o)}$ of β_0 . (Exactness in the middle follows from the fact that if $\mu \leq n - ht(o)$, then we may recover the Witt components $\beta_\lambda^{(\mu)}$ from β_0 modulo $p^{n-ht(o)}$ using the equations (8.3.3) and the fact that $a(\ell) \geq a(0)$ for all ℓ .) Moreover, we have

$$\beta_0 \equiv (\gamma_1 \cdots \gamma_k)^{-1} p^{a(|o|)} \sigma^{|o|} \beta_0 \pmod{p^{n-ht(o)}}.$$

It follows that the image of $H_{n,v}^o$ in $W_{n-ht(o),v}$ has order at most $p^{|o|(n-ht(o))}$ independently of v . (We may even conclude that it is 0 if $a(|o|) > 0$.) Thus this image does not contribute to the limit in the theorem, and it will suffice to bound $p^{n-ht(o)} H_{n,v}^o$.

Note also that if $n' > n \geq ht(o)$, then

$$p^{n-ht(o)} H_{n,v}^o \xrightarrow{\sim} p^{n'-ht(o)} H_{n',v}^o$$

via $(\beta_\lambda) \mapsto (p^{n'-n}\beta_\lambda)$. Thus we may assume that $n = ht(o)$ for the rest of the proof.

To finish the estimation, we ‘‘break’’ the circular system (8.3.3) into a triangular system, as in [Ulmer 2014c, §7.6]. To that end, choose λ so that $a(\ell^\lambda)$ is the maximum of a , and note that $ht(0) = a(\ell^\lambda) - a(0) = a(\ell^\lambda)$. Then we have

$$ht(o) = a(\ell^\lambda) = e_1 - e_2 + \cdots + e_{2\lambda+1}$$

and

$$0 = p^{ht(o)} \beta_0 = p^{e_1 - e_2 + \cdots + e_{2\lambda+1}} \beta_0 = p^{e_3 - e_4 + \cdots + e_{2\lambda+1}} \sigma^{-\ell_1}(\gamma_1 \beta_1) = \cdots = p^{e_{2\lambda+1}} \sigma^{-\ell_1}(\gamma_1) \cdots \sigma^{-\ell_\lambda}(\gamma_\lambda \beta_\lambda).$$

It follows that $p^{e_{2\lambda+1}} \beta_\lambda = 0$. Using this in (8.3.3) and reordering, we obtain a lower-triangular system

$$\begin{aligned} 0 &= \gamma_{\lambda+1} p^{e_{2\lambda+2}} \beta_{\lambda+1} \\ 0 &= -p^{e_{2\lambda+3}} \sigma^{\ell_{\lambda+2} - \ell_{\lambda+1}} \beta_{\lambda+1} + \gamma_{\lambda+2} p^{e_{2\lambda+4}} \beta_{\lambda+2} \\ &\vdots \\ 0 &= -p^{e_{2k-1}} \sigma^{\ell_k - \ell_{k-1}} \beta_{k-1} + \gamma_k p^{e_{2k}} \beta_k \\ 0 &= -p^{e_1} \sigma^{\ell_1 - \ell_0} \beta_0 + \gamma_1 p^{e_2} \beta_1 \\ &\vdots \\ 0 &= -p^{e_{2\lambda-1}} \sigma^{\ell_\lambda - \ell_{\lambda-1}} \beta_{\lambda-1} + \gamma_\lambda p^{e_{2\lambda}} \beta_\lambda. \end{aligned}$$

This system can be rewritten in the form

$$U_1 B U_2 \begin{pmatrix} \beta_{\lambda+1} \\ \vdots \\ \beta_k \\ \beta_1 \\ \vdots \\ \beta_\lambda \end{pmatrix} = 0$$

where U_1 and U_2 are diagonal with powers of σ and products of the units γ_i in the diagonal entries and where

$$B = \begin{pmatrix} p^{e_{2\lambda+2}} & & & & & \\ -p^{e_{2\lambda+3}} & p^{e_{2\lambda+4}} & & & & \\ & & \ddots & & & \\ & & & -p^{e_{2k-1}} & p^{e_{2k}} & \\ & & & -p^{e_1} & p^{e_2} & \\ & & & & & \ddots \\ & & & & & & -p^{e_{2\lambda-1}} & p^{e_{2\lambda}} \end{pmatrix}.$$

It follows that the number of solutions to this system is

$$q^{v(e_2+e_4+\dots+e_{2k})}.$$

On the other hand, $e_2 + e_4 + \dots + e_{2k}$ is the total number of appearances of l in the word w , and since $a(|o|) \geq 0$, w has at least as many appearances of u as of l , so this sum is equal to $d(o)$. It follows that $|H_{ht(o),v}^o| = q^{vd(o)}$ and that

$$\lim_{v \rightarrow \infty} \frac{\log |H_{n,v}^o|}{\log(q^v)} = d(o)$$

for any $n \geq ht(o)$. This completes the proof of part (3) of the theorem under the hypothesis that $a(|o|) \geq 0$.

The proof when $a(|o|) \leq 0$ is very similar. Roughly speaking, one proceeds as above, but with a base point so that $a(|o|)$ is the minimum of a and with β_k playing the role of β_0 . More precisely, assuming that w has at least one u and at least one l and that $a(|o|) \leq 0$, we may choose a base point for o such that $a(|o|)$ is the minimum value of a and $a(1) > a(0) = 0$. Fix such a base point, denoted (i_0, j_0) , for the rest of the argument.

As before, choose

$$0 = \ell_0 < \ell^0 < \ell_1 < \ell^1 \dots < \ell^{k-1} < \ell_k = |o|$$

such that a is nondecreasing on $\{\ell_\lambda, \dots, \ell^\lambda\}$ and nonincreasing on $\{\ell^\lambda, \dots, \ell_{\lambda+1}\}$ for $0 \leq \lambda \leq k - 1$. Let

$$\beta_\lambda = \alpha_{p^{\ell_\lambda(i_0, j_0)}} \quad 0 \leq \lambda \leq k.$$

Then as before, the coefficients $\alpha_{i,j}$ satisfy equations (8.3.1) and (8.3.2) if and only if the β_λ satisfy (8.3.3).

The same dévissage as before shows that it suffices to estimate the order of $H_{n,v}^o$ in the case where $n = ht(o)$. We make the circular system (8.3.3) triangular as follows: Choose λ so that $a(\ell^\lambda)$ is the maximum of a . Then

$$ht(o) = a(\ell^\lambda) - a(|o|) = e_{2k} - e_{2k-1} + \dots + e_{2\lambda+2}.$$

Therefore,

$$\begin{aligned} 0 &= p^{ht(o)} \beta_k = p^{e_{2k} - e_{2k-1} + \dots + e_{2\lambda+2}} \beta_k \\ &= p^{e_{2k-2} - e_{2k-3} + \dots + e_{2\lambda+2}} \gamma_k^{-1} \sigma^{\ell_k - \ell_{k-1}} (\beta_{k-1}) \\ &\quad \vdots \\ &= p^{e_{2\lambda+2}} \gamma_k^{-1} \sigma^{\ell_k - \ell_{k-1}} (\gamma_{k-1}^{-1}) \sigma^{\ell_k - \ell_{k-2}} (\gamma_{k-2}^{-1}) \dots \sigma^{\ell_k - \ell_{\lambda+1}} (\beta_{\lambda+1}). \end{aligned}$$

It follows that $p^{e_{2\lambda+2}} \beta_{\lambda+1} = 0$. Using this in (8.3.3) and reordering, we obtain (up to units and powers of σ) an upper-triangular system whose diagonal entries are $p^{e_1}, p^{e_3}, \dots, p^{e_{2k-1}}$.

It follows that the number of solutions to (8.3.3) with coefficients in $W_{n,v}$ (with $n = ht(o)$) is $q^{v(e_1 + \dots + e_{2k-1})}$. Observing that $a(|o|) \leq 0$ implies that $d(o) = e_1 + \dots + e_{2k-1}$, we find that $|H_{ht(o),v}^o| = q^{vd(o)}$

and that

$$\lim_{v \rightarrow \infty} \frac{\log |H_{n,v}^o|}{\log(q^v)} = d(o)$$

for any $n \geq ht(o)$. This completes the proof of part (3) of the theorem in the remaining case when $a(|o|) \leq 0$. □

9. Equidistribution

We record three equidistribution statements to be used to control the average behavior of the invariant $d(o)$ from the preceding section. The first is a consequence of what is proven in [Griffon 2018, Theorem 4.1]. The second is a straightforward “two-variable” generalization, and the third is a simple corollary of the first. We omit the proofs since they are orthogonal to our main concerns.

Proposition 9.1 (Helfgott, Hindry–Pacheco, Griffon). *Let $A \subset [0, 1]$ be an interval of length α . Let p be a prime number and let d run through positive integers prime to p . Let $\langle p \rangle$ act on $\mathbb{Z}/d\mathbb{Z}$ by multiplication, and let O be the set of orbits. Then*

$$\lim_{d \rightarrow \infty} \frac{1}{d} \sum_{o \in O} \left| \frac{|\{a \in o \mid \langle a/d \rangle \in A\}|}{|o|} - \alpha \right| = 0.$$

Proposition 9.2. *Let p be a prime number, let r be a fixed integer prime to p and let d run through integers prime to p . Let $\langle p \rangle$ act on $(\mathbb{Z}/r\mathbb{Z}) \times (\mathbb{Z}/d\mathbb{Z})$ diagonally, and let O be the set of orbits. Then*

$$\lim_{d \rightarrow \infty} \frac{1}{d} \sum_{o \in O} \left| \frac{|\{(a, b) \in o \mid \langle a/r \rangle + \langle b/d \rangle < 1\}|}{|o|} - \frac{1}{2} \right| = 0.$$

Proposition 9.3. *Let p be a prime number, let $I = \mathbb{Z}/d\mathbb{Z}$ with d prime to p equipped with the multiplication action of $\langle p \rangle$, and let $J = \{0, 1\}$ be a two-element set equipped with the nontrivial action of $\langle p \rangle$. Let $\langle p \rangle$ act on $I \times J$ diagonally, and let O be the set of orbits. Then*

$$\lim_{d \rightarrow \infty} \frac{1}{d} \sum_{o \in O} \left| \frac{|\{(a, b) \in o \mid \langle a/d \rangle < 1/2, b = 0\}| + |\{(a, b) \in o \mid \langle a/d \rangle > 1/2, b = 1\}|}{|o|} - \frac{1}{2} \right| = 0.$$

10. Calculations for curves defined by four monomials

In this section we compute the limit of Brauer–Siegel ratios for a family of elliptic curves related to the constructions in [Shioda 1986; Ulmer 2002]. We then explain how the same can be done for families of Jacobians of every genus in every positive characteristic.

Throughout, let $k = \mathbb{F}_q$, the finite field of cardinality q and characteristic p , and let $K = k(t)$, the rational function field over k .

10.1. The curve of [Ulmer 2002]. Let p be a prime number, let d be a positive integer prime to p , and let E_d be the elliptic curve over K defined by

$$y^2 + xy = x^3 - t^d \tag{10.1.1}$$

This family of curves was introduced in [Ulmer 2002] where it was shown that $\text{III}(E_d)$ is finite and the rank of $E_d(K)$ is unbounded as d varies. Hindry and Pacheco [2016] computed the Brauer–Siegel ratio of E_d as $d \rightarrow \infty$ by analytic means, i.e., by a careful study of the L -function of E_d . Here we compute it via algebraic means, more precisely, through a consideration of $\dim \text{III}(E_d)$.

Theorem 10.2. *We have*

$$\lim_{d \rightarrow \infty} \text{BS}(E_d) = 1.$$

Proof. Because $E_{pd} = E_d^{(p)}$, Theorem 5.1 implies that it will suffice to compute the limit as d runs through positive integers relatively prime to p and tending to infinity.

We are going to bound $\text{BS}(E_d)$ from below by estimating $\dim \text{III}(E_d)$. Since the latter is invariant under extension of the ground field, we are free to extend k as needed and will do so in the geometric argument below.

Let \mathcal{E}_d be the smooth projective surface equipped with a relatively minimal morphism $\pi : \mathcal{E}_d \rightarrow \mathbb{P}^1$ whose generic fiber is E_d . The procedure for constructing a model \mathcal{E}_d is explained in general in [Ulmer 2011, Lecture 3], and this particular example is carried out in detail in [Ulmer 2002, §3]. The important thing to know about \mathcal{E}_d is that it is birational to the hypersurface in $\mathbb{A}_{(x,y,t)}^3$ defined by (10.1.1).

Using the method of [Shioda 1986], it is proven in [Ulmer 2002, §4] that \mathcal{E}_d is birational to the quotient of the Fermat surface of degree d by a group of order d^2 . It is proven in [Shioda and Katsura 1979] that the Fermat surface of degree d is birational to the quotient of the product of two Fermat curves of degree d by a group of order d . (Here we may need to extend k so that it contains the $2d$ -th roots of unity.) Putting these together, we find that \mathcal{E}_d is birational to the quotient of $F_d \times F_d$ by the group

$$\Delta \subset (\mu_d^3 / \mu_d)^2 \subset \text{Aut}(F_d) \times \text{Aut}(F_d)$$

generated by

$$([\zeta^2, \zeta, 1], [1, 1, 1]), ([1, \zeta, 1], [\zeta^3, 1, 1]), \text{ and } ([1, 1, \zeta], [1, 1, \zeta])$$

where ζ is a primitive d -th root of unity in k .

It follows from Corollary 6.5 that

$$\dim \text{III}(E_d) = \dim \text{Hom}_A(H^1(F_d)/p^n, H^1(F_d)/p^n)^\Delta \tag{10.2.1}$$

for all sufficiently large n . Section 7 and Proposition 7.1 describe the cohomology group $H^1(F_d)$ with its action of Frobenius. They show in particular that the dimension in the last display can be computed by the methods of Section 8.

To spell this out, recall that the cohomology of F_d splits into lines indexed by

$$A' = \left\{ (a_0, a_1, a_2) \mid a_i \neq 0, \sum a_i = 0 \right\} \subset (\mathbb{Z}/d\mathbb{Z})^3$$

and that A' is the disjoint union of A_0 and A_1 as in Section 7. The curves F_d and their cohomology furnish data $M = N = H_{\text{crys}}^1(F_d/W(k))$, $I = J = A'$, and (c_i, d_j) as in Section 8.1.

A short calculation reveals that the basis elements φ_{ij} which contribute to the right hand side of (10.2.1) are those indexed by (i, j) of the form

$$(i, j) = (a_0, a_1, a_2, b_0, b_1, b_2) = b_1(-3, 6, -3, 2, 1, -3)$$

where $b_1 \in d\mathbb{Z}$ is such that $6b_1 \neq 0$. In other words, projection to the b_1 coordinate allows us to identify the orbits of $\langle p \rangle$ on $I \times J$ which contribute to (10.2.1) with the orbits of $\langle p \rangle$ on

$$B = \{b \in \mathbb{Z}/d\mathbb{Z} \mid 6b \neq 0\}.$$

Under this identification, $(i, j) \in I_0 \times J_1$ if and only if

$$0 < \left\langle \frac{b}{d} \right\rangle < \frac{1}{6}$$

and $(i, j) \in I_1 \times J_0$ if and only if

$$\frac{5}{6} < \left\langle \frac{b}{d} \right\rangle < 1$$

where $\langle \cdot \rangle$ denotes the fractional part. Thus, the invariant $d(o)$ of Section 8.1 becomes the following invariant of orbits of $\langle p \rangle$ on B : Setting

$$B_0 = \left\{ b \in \mathbb{Z}/d\mathbb{Z} \mid 0 < \left\langle \frac{b}{d} \right\rangle < \frac{1}{6} \right\} \quad \text{and} \quad B_1 = \left\{ b \in \mathbb{Z}/d\mathbb{Z} \mid \frac{5}{6} < \left\langle \frac{b}{d} \right\rangle < 1 \right\},$$

we have

$$d(o) = \min(|o \cap B_0|, |o \cap B_1|).$$

Finally, the equidistribution result Proposition 9.1 yields that

$$\sum_{o \in \mathcal{O}} d(o) = \frac{d}{6} + \epsilon$$

where $\epsilon/d \rightarrow 0$ as $d \rightarrow \infty$, and so

$$\dim \text{III}(E_d) = \frac{d}{6} + \epsilon.$$

It follows from [Ulmer 2002, §2] that $\deg \omega_{E_d} = \lceil \frac{d}{6} \rceil$, so by applying Corollary 4.7, we conclude that

$$\liminf_{d \rightarrow \infty} \text{BS}(E_d) \geq 1.$$

Taking into account the upper bound (1.1) of Hindry and Pacheco, we finally conclude that

$$\lim_{d \rightarrow \infty} \text{BS}(E_d) = 1. \quad \square$$

10.3. Other elliptic curves. The methods employed in the previous subsection can be used to compute the limiting Brauer–Siegel ratio for several other families of elliptic curves, namely those defined by equations involving 4 monomials. This includes the Hessian family studied in [Griffon 2016, Chapter 5] and a closely related family introduced by Davis and Occhipinti [2016] and studied in [Griffon 2016, Chapter 7]. We will not give the details here, since no fundamentally new phenomena arise.

10.4. Higher genus Jacobians. For every prime p and every $g > 0$, there is a sequence of curves of genus g over $\mathbb{F}_p(t)$ whose Jacobians are absolutely simple, satisfy the Birch and Swinnerton-Dyer conjecture, and have unbounded analytic and algebraic ranks; see [Ulmer 2007, §7]. Since these curves are defined by four monomials, the methods of this paper suffice to compute the limit of their Brauer–Siegel ratios. In the rest of this subsection, we explain the details for the main case, namely when g is a positive integer and p is a prime such that $p \nmid (2g + 2)(2g + 1)$. The other cases are similar and we omit them in the interest of brevity.

Fix a positive integer g , a prime p such that $p \nmid (2g + 2)(2g + 1)$, and a positive integer d . Let X_d be the smooth, proper curve of genus g over $K = \mathbb{F}_p(t)$ defined by

$$y^2 = x^{2g+2} + x^{2g+1} + t^d \tag{10.4.1}$$

and let J_d be its Jacobian.

Theorem 10.5. $\lim_{d \rightarrow \infty} \text{BS}(J_d) = 1.$

Proof. Once again, it suffices to restrict to d not divisible by p . We will bound $\text{BS}(J_d)$ from below by estimating $\dim \text{III}(J_d)$ using that X_d has a model which is dominated by a product of Fermat curves. As usual, we are free to expand the ground field \mathbb{F}_p and we do so as needed below.

Let \mathcal{X}_d be the smooth projective surface equipped with a relatively minimal morphism $\pi : \mathcal{X}_d \rightarrow \mathbb{P}^1$ with generic fiber X_d . Again, what is most important is that \mathcal{X}_d is birational to the hypersurface in \mathbb{A}^3 defined by (10.4.1).

Using the method of [Shioda 1986] (see also [Ulmer 2007]), one sees that \mathcal{X}_d is birational to the quotient of the Fermat surface of degree $2d$ by a group of order $(2d)^2$, and therefore birational to the quotient of $F_{2d} \times F_{2d}$ by a group of order $(2d)^3$. (Here we enlarge \mathbb{F}_p to a finite extension k that contains the $2d$ -th roots of unity.) More precisely, carrying out the procedure of [Ulmer 2007, §6] and using [Shioda and Katsura 1979], one finds that \mathcal{X}_d is birational to the quotient of $F_{2d} \times F_{2d}$ by the group

$$\Delta \subset (\mu_{2d}^3 / \mu_{2d})^2 \subset \text{Aut}(F_{2d}) \times \text{Aut}(F_{2d})$$

generated by

$$([\zeta^2, 1, 1], [1, 1, 1]), \quad ([1, 1, 1], [1, \zeta^d, 1]), \quad ([1, 1, 1], [\zeta, \zeta^{2g+2}, 1]), \quad \text{and} \quad ([1, 1, \zeta], [1, 1, \zeta])$$

where ζ is a primitive $2d$ -th root of unity in k .

It follows from Corollary 6.5 that

$$\dim \text{III}(E_d) = \dim \text{Hom}_A(H^1(F_d)/p^n, H^1(F_d)/p^n)^\Delta \tag{10.5.1}$$

for all sufficiently large n .

As in the previous subsections, the curves F_{2d} and their cohomology furnish data $M = N = H_{\text{crys}}^1(F_{2d}/W(k))$, $I = J = A'$, and (c_i, d_j) as in Section 8.1.

A short calculation reveals that the basis elements φ_{ij} which contribute to the right hand side of (10.5.1) are those indexed by (i, j) of the form

$$(i, j) = (a_0, a_1, a_2, b_0, b_1, b_2) = (-(4g+4)b, 2b, (4g+2)b, d, d - (4g+2)b, (4g+2)b)$$

where $b \in \mathbb{Z}/d\mathbb{Z}$ is such that none of the coordinates a_0, \dots, b_2 are zero in $\mathbb{Z}/2d\mathbb{Z}$. (Note that all of the coefficients of b above are even, so the display gives a well-defined element of $(\mathbb{Z}/2d\mathbb{Z})^6$ even though b lies in $\mathbb{Z}/d\mathbb{Z}$.) Thus the relevant orbits of $\langle p \rangle$ on $I \times J$ can be identified with the orbits of $\langle p \rangle$ on the subset B of $\mathbb{Z}/d\mathbb{Z}$ where none of the coordinates of (i, j) is 0.

Next we work out conditions on b for the corresponding (i, j) to lie in $I_0 \times J_1$ or $I_1 \times J_0$. One finds that

$$i = (a_0, a_1, a_2) = (-(4g+4)b, 2b, (4g+2)b)$$

lies in I_0 if and only if the fractional part $\langle b/d \rangle$ lies in one of the intervals

$$\left(\frac{k+1}{2g+2}, \frac{k+1}{2g+1} \right), \quad k = 0, \dots, 2g$$

and i lies in I_1 if and only if the fractional part $\langle b/d \rangle$ lies in one of the intervals

$$\left(\frac{k}{2g+1}, \frac{k+1}{2g+2} \right), \quad k = 0, \dots, 2g.$$

On the other hand,

$$j = (b_0, b_1, b_2) = (d, d - (4g+2)b, (4g+2)b)$$

lies in J_0 if and only if the fractional part $\langle b/d \rangle$ lies in one of the intervals

$$\left(\frac{2\ell+1}{4g+2}, \frac{2\ell+2}{4g+2} \right), \quad \ell = 0, \dots, 2g$$

and j lies in J_1 if and only if the fractional part $\langle b/d \rangle$ lies in one of the intervals

$$\left(\frac{2\ell}{4g+2}, \frac{2\ell+1}{4g+2} \right), \quad \ell = 0, \dots, 2g.$$

It follows that (i, j) lies in $I_0 \times J_1$ if and only if

$$\left\langle \frac{b}{d} \right\rangle \in \left(\frac{k+1}{2g+2}, \frac{2k+1}{4g+2} \right)$$

with $k = g+1, \dots, 2g$ and it lies in $I_1 \times J_0$ if and only if

$$\left\langle \frac{b}{d} \right\rangle \in \left(\frac{2k+1}{4g+2}, \frac{k+1}{2g+2} \right)$$

with $k = 0, \dots, g-1$.

The total length of the intervals corresponding to $I_0 \times J_1$ is

$$\sum_{k=g+1}^{2g} \left(\frac{2k+1}{4g+2} - \frac{k+1}{2g+2} \right) = \frac{g}{8g+4}$$

and the total length of the intervals corresponding to $I_1 \times J_0$ is

$$\sum_{k=0}^{g-1} \left(\frac{k+1}{2g+2} - \frac{2k+1}{4g+2} \right) = \frac{g}{8g+4}.$$

Transferring the definition of $d(o)$ to B and applying the equidistribution result Proposition 9.1, we find that

$$\dim \text{III}(J_d) = \sum_o d(o) = \frac{dg}{8g+4} + \epsilon$$

where $\epsilon/d \rightarrow 0$ as $d \rightarrow \infty$.

We pause briefly to consider the case $g = 1$. By [Weil 1954], the Jacobian of X_d is the elliptic curve

$$y^2 = x^3 - 4t^d x + t^d.$$

It is easy to see that the bundle ω_d attached to J_d has degree $\lceil \frac{d}{12} \rceil$. It then follows from our estimation of $\dim \text{III}(J_d)$ and Corollary 4.7 that $\liminf_{d \rightarrow \infty} \text{BS}(J_d) \geq 1$ and thus, by the Hindry–Pacheco upper bound (1.1), that $\lim_{d \rightarrow \infty} \text{BS}(J_d) = 1$.

To extend this to higher genus, we will give an upper bound on the degree of ω_d of the form $dg/(8g+4) + \epsilon$ where $\epsilon/d \rightarrow 0$ as $d \rightarrow \infty$. More precisely, we will show that $\deg(\omega_d) = dg/(8g+4)$ for all d divisible by $(2g+1)(2g+2)$. For a general d , we let

$$d' = \text{lcm}(d, (2g+1)(2g+2))$$

and apply Lemma 2.7.1 to conclude that

$$\deg(\omega_d) \leq \frac{dg}{(8g+4)} + 2g((2g+1)(2g+2) - 1)$$

which gives the desired estimate.

For $i = 1, \dots, g$, let ω_i be the 1-form $x^{i-1} dx/y$ on X_d over K . These 1-forms are regular and give a basis of $H^0(X, \Omega_{X/K}^1)$. We will consider their extensions to a suitable model $\pi : \mathcal{X} \rightarrow \mathbb{P}^1$ of X and use them to compute $\deg(\omega_d)$.

In [Ulmer 2007, §7.7], a model of X over $U = \mathbb{P}^1 \setminus \{0, \infty\}$ is constructed which is regular and a Lefschetz pencil, i.e., its singular fibers are irreducible with one ordinary node each. It is easy to see that the differentials ω_i extend to this model and

$$\sigma := \omega_1 \wedge \dots \wedge \omega_g$$

defines a nowhere vanishing section of ω_d over U . To compute $\deg(\omega_d)$ it will thus suffice to compute the order of vanishing of σ at $t = 0$ and $t = \infty$. This is where we use the hypothesis that d is a multiple of $(2g + 1)(2g + 2)$.

Indeed, if $d = 2(2g + 1)k$, then the change of coordinates $x \rightarrow t^{2k}x'$, $y \rightarrow t^{(2g+1)k}y'$ brings X into the form

$$y'^2 = t^{2k}x'^{2g+2} + x'^{2g+1} + 1$$

which has good reduction at $t = 0$. Moreover, we see that $\omega_i = t^{(2i-2g-1)k}\omega'_i$ where $\omega'_i = (x'^{i-1}dx')/y'$, and that the ω'_i have linearly independent reductions at $t = 0$. This shows that σ has a pole at $t = 0$ of order

$$\sum_{i=1}^g \frac{(2g + 1 - 2i)d}{2(2g + 1)}.$$

Similarly, when $d = (2g + 2)\ell$, the change of coordinates $x \rightarrow t^{2\ell}x$, $y \rightarrow t^{(2g+2)\ell}y$ brings X into the form

$$y^2 = x^{2g+2} + t^{-\ell}x^{2g+1} + 1,$$

which has good reduction at $t = \infty$. Moreover, we see that $\omega_i = t^{(i-g-1)\ell}\omega'_i$ where $\omega'_i = (x'^{i-1}dx')/y'$, and that the ω'_i have linearly independent reductions at $t = \infty$. This shows that σ has a zero at $t = \infty$ of order

$$\sum_{i=1}^g \frac{(g + 1 - i)d}{2g + 2}.$$

A short computation then shows that $\deg(\omega_d)$ is $dg/(8g + 4)$.

Note that these calculations also show that J_d has good reduction at $t = 0$ and $t = \infty$ when d is divisible by $(2g + 1)(2g + 2)$. Using Section 2.6, these reduction results imply that $\tau(J_d) = O(H(J_d)^\epsilon)$ for all $\epsilon > 0$. Then Proposition 4.6 shows that

$$\liminf_{d \rightarrow \infty} \text{BS}(J_d) \geq \liminf_{d \rightarrow \infty} \frac{\dim(\text{III}(J_d))}{\deg(\omega_{J_d})} \geq 1.$$

Taking into account the upper bound (1.1) of Hindry and Pacheco, we finally conclude that

$$\lim_{d \rightarrow \infty} \text{BS}(J_d) = 1. \quad \square$$

11. Calculations for Jacobians related to Berger's construction

In this section we compute the limiting Brauer–Siegel ratio for some families of curves related to the construction in [Berger 2008; Ulmer 2013].

Throughout, let $k = \mathbb{F}_q$, the finite field of cardinality q and characteristic p , and let $K = k(t)$, the rational function field over k .

11.1. The Legendre curve. Assume that $p > 2$, let d be a positive integer, and let E_d be the elliptic curve over K defined by

$$y^2 = x(x + 1)(x + t^d). \tag{11.1.1}$$

This family of curves has been studied extensively, in particular in [Ulmer 2014b; Conceição et al. 2014; Ulmer 2014c; Griffon 2016, Chapter 4]. In the latter, the limit of the Brauer–Siegel ratio of E_d as $d \rightarrow \infty$ was computed by analytic means, i.e., by a careful study of the L -function of E_d . Here we compute it via algebraic means, more precisely, through a consideration of $\dim \text{III}(E_d)$.

Theorem 11.2. *We have*

$$\lim_{d \rightarrow \infty} \text{BS}(E_d) = 1.$$

Proof. As usual, it suffices to consider values of d not divisible by p .

Let \mathcal{E}_d be the smooth projective surface equipped with a relatively minimal morphism $\pi : \mathcal{E}_d \rightarrow \mathbb{P}^1$ whose generic fiber is E_d . This is constructed in [Ulmer 2014b] (under the simplifying hypothesis that d is even, but the odd case is similar). The main thing we need to know about \mathcal{E}_d is that it is birational to the hypersurface in $\mathbb{A}_{(x,y,t)}^3$ defined by the (11.1.1).

Let \mathcal{C}_d be the curve with affine equation

$$x^2 = z^d + 1$$

and let \mathcal{D}_d be the curve with affine equation

$$y^2 = w^d + 1.$$

Both curves admit an evident action of $\Delta = \mu_2 \times \mu_d$ (over \bar{k}). Let Δ act “antidiagonally” on $\mathcal{C}_d \times \mathcal{D}_d$:

$$(\zeta_2, \zeta_d)(x, z, y, w) = (\zeta_2 x, \zeta_d z, \zeta_2^{-1} y, \zeta_d^{-1} w).$$

Our first main claim is that \mathcal{E}_d is birational to the quotient $\mathcal{C}_d \times \mathcal{D}_d / \Delta$ via the map

$$(x, z, y, w) \mapsto (x = z^d, y = z^d xy, t = wz).$$

Indeed, it is evident that this defines a dominant rational map from $\mathcal{C}_d \times \mathcal{D}_d$ to \mathcal{E}_d which factors through the quotient by Δ . Degree considerations then show that the induced map has degree 1, i.e., it is a birational isomorphism.

We are thus in position to apply the machinery of Section 6. In particular, it follows from Corollary 6.5 that

$$\dim \text{III}(E_d) = \dim \text{Hom}_A(H^1(\mathcal{C}_d)/p^n, H^1(\mathcal{D}_d)/p^n)^\Delta \tag{11.2.1}$$

for all sufficiently large n . Section 7.3 and Proposition 7.1 describe the cohomology groups $H^1(\mathcal{C}_d)$ and $H^1(\mathcal{D}_d)$ with their actions of Frobenius. They show in particular, that the dimension in the last display can be computed by the methods of Section 8.

To spell this out, let

$$I = J = \mathbb{Z}/d\mathbb{Z} \setminus \{0, d/2 \text{ (if } d \text{ is even)}\},$$

decomposed as $I_0 = J_0 = \{i \mid d/2 < i < d\}$ and $I_1 = J_1 = \{i \mid 0 < i < d/2\}$. Section 7 shows that the crystalline cohomology groups $H^1(\mathcal{C}_d)$ and $H^1(\mathcal{D}_d)$ with their action of Frobenius furnish data (M, N, I, J, c_i, d_j) as in Section 8.1, as well as the invariant $d(o)$ for each orbit o of $\langle p \rangle$ on $I \times J$.

Since Δ acts antidiagonally, the orbits that contribute to the right hand side of (11.2.1) are those whose elements (i, j) satisfy $j = -i$. Write O^Δ for the set of such orbits. Applying Theorem 8.3, we conclude that

$$\dim \text{III}(E_d) = \sum_{o \in O^\Delta} d(o). \quad (11.2.2)$$

We may identify the orbits in O^Δ with the orbits of $\langle p \rangle$ on I via the projection $\pi_I : I \times J \rightarrow I$. Also, since $(i, -i) \in I_0 \times J_1$ if and only if $i \in I_0$, and $(i, -i) \in I_1 \times J_0$ if and only if $i \in I_1$, we have

$$d(o) = \min(|\pi_I(o) \cap I_0|, |\pi_I(o) \cap I_1|).$$

Thus the sum on the right hand side of (11.2.2) becomes a sum over orbits of $\langle p \rangle$ on I , and the invariant $d(o)$ is described “on average” in Section 9. In particular, the equidistribution result Proposition 9.1 implies that

$$\dim \text{III}(E_d) = \sum_{o \in O^\Delta} d(o) = \frac{d}{2} + \epsilon_d$$

where $\epsilon_d/d \rightarrow 0$ as $d \rightarrow \infty$.

Since $\deg(\omega_{E_d}) = \lceil \frac{d}{2} \rceil$ (e.g., by [Ulmer 2014b, Lemma 7.1]), Corollary 4.7 implies that

$$\liminf_{d \rightarrow \infty} \text{BS}(E_d) \geq \liminf_{d \rightarrow \infty} \frac{\dim \text{III}(E_d)}{\deg(\omega_{E_d})} = 1.$$

Taking into account the upper bound (1.1) of Hindry and Pacheco, we conclude that

$$\lim_{d \rightarrow \infty} \text{BS}(J_d) = 1. \quad \square$$

11.3. Other elliptic curves. The methods employed in the previous subsection can be used to compute the limiting Brauer–Siegel ratio for several other families of elliptic curves, namely those coming from Berger’s construction where the dominating curves are related to Fermat curves. This is the case in particular for the universal curve over $X_1(4)$ studied in [Griffon 2016, Chapter 6] and the curve “ $B_{1/2,d}$ ” introduced in [Berger 2008, §4] and studied in [Griffon 2016, Chapter 8]. We will not give the details here, since no fundamentally new phenomena arise.

11.4. Higher dimensional Jacobians. Let p be a prime number, let q be a power of p , and let $k = \mathbb{F}_q$. Let r and d be integers relatively prime to p . Let $X = X_{r,d}$ be the smooth projective curve over $K = k(t)$ associated to the equation

$$y^r = x^{r-1}(x+1)(x+t^d). \quad (11.4.1)$$

This is a curve of genus $r - 1$, and the case $r = 2$ is the Legendre curve of Section 11.1. Let $J = J_{r,d}$ be the Jacobian of X . This family of Jacobians was studied in [Berger et al. 2015], where among other things it was proven that $\text{III}(J_{r,d})$ is finite for all p, q, r , and d as above. Here we will compute the limiting Brauer–Siegel ratio for fixed q and r as $d \rightarrow \infty$.

Theorem 11.5. *For all q and r as above,*

$$\lim_{\substack{d \rightarrow \infty \\ (p,d)=1}} \text{BS}(J_{r,d}) = 1.$$

Here the limit is through integers prime to p . It would be possible to include those d divisible by p using a straightforward generalization of the ideas in Section 5, but will not do that here.

Proof. Since r will be fixed throughout, we omit it from the notation. Let \mathcal{X}_d be the smooth projective surface equipped with a relatively minimal morphism $\pi : \mathcal{X}_d \rightarrow \mathbb{P}^1$ whose generic fiber is X_d . This is constructed in [Berger et al. 2015, §3.1]. The important thing to know about \mathcal{X}_d is that it is birational to the hypersurface in $\mathbb{A}_{(x,y,t)}^3$ defined by (11.4.1).

Let \mathcal{C}_d be the curve with affine equation

$$x^r = z^d + 1$$

and let \mathcal{D}_d be the curve with affine equation

$$y^r = w^d + 1.$$

Both curves admit an evident action of $\Delta = \mu_r \times \mu_d$ (over \bar{k}). Let Δ act “antidiagonally” on $\mathcal{C}_d \times \mathcal{D}_d$:

$$(\zeta_r, \zeta_d)(x, z, y, w) = (\zeta_r x, \zeta_d z, \zeta_r^{-1} y, \zeta_d^{-1} w).$$

It is proven in [Berger et al. 2015, §3.3] that \mathcal{X}_d is birational to the quotient $\mathcal{C}_d \times \mathcal{D}_d / \Delta$ via the map

$$(x, z, y, w) \mapsto (x = z^d, y = z^d xy, t = wz).$$

We are thus in position to apply the machinery of Section 6. In particular, it follows from Corollary 6.5 that

$$\dim \text{III}(J_d) = \dim \text{Hom}_A(H^1(\mathcal{C}_d)/p^n, H^1(\mathcal{D}_d)/p^n)^\Delta \tag{11.5.1}$$

for all sufficiently large n . Section 7.3 and Proposition 7.1 describe the cohomology groups $H^1(\mathcal{C}_d)$ and $H^1(\mathcal{D}_d)$ with their actions of Frobenius. They show in particular, that the dimension in the last display can be computed by the methods of Section 8.

To spell this out, let

$$\begin{aligned}
 I = J &= \left\{ (a, b) \in \mathbb{Z}/r\mathbb{Z} \times \mathbb{Z}/d\mathbb{Z} \mid a \neq 0, b \neq 0, \left\langle \frac{a}{r} \right\rangle + \left\langle \frac{b}{d} \right\rangle \neq 1 \right\}, \\
 I_0 = J_0 &= \left\{ (a, b) \in \mathbb{Z}/r\mathbb{Z} \times \mathbb{Z}/d\mathbb{Z} \mid a \neq 0, b \neq 0, \left\langle \frac{a}{r} \right\rangle + \left\langle \frac{b}{d} \right\rangle > 1 \right\}, \quad \text{and} \\
 I_1 = J_1 &= \left\{ (a, b) \in \mathbb{Z}/r\mathbb{Z} \times \mathbb{Z}/d\mathbb{Z} \mid a \neq 0, b \neq 0, \left\langle \frac{a}{r} \right\rangle + \left\langle \frac{b}{d} \right\rangle < 1 \right\}.
 \end{aligned}$$

Section 7 shows that the crystalline cohomology groups $H^1(\mathcal{C}_d)$ and $H^1(\mathcal{D}_d)$ with their action of Frobenius furnish data (M, N, I, J, c_i, d_j) as in Section 8.1, as well as the invariant $d(o)$ for each orbit o of $\langle p \rangle$ on $I \times J$.

Since Δ acts antidiagonally, the orbits that contribute to the right hand side of (11.5.1) are those whose elements $(i, j) = (a, b, a', b')$ satisfy $j = -i$, i.e., $a' = -a$ and $b' = -b$. Write O^Δ for the set of such orbits. Applying Theorem 8.3, we conclude that

$$\dim \text{III}(J_d) = \sum_{o \in O^\Delta} d(o). \tag{11.5.2}$$

We may identify the orbits in O^Δ with the orbits of $\langle p \rangle$ on I via the projection $\pi_I : I \times J \rightarrow I$. Also, since $(i, -i) \in I_0 \times J_1$ if and only if $i \in I_0$, and $(i, -i) \in I_1 \times J_0$ if and only if $i \in I_1$, we have

$$d(o) = \min(|\pi_I(o) \cap I_0|, |\pi_I(o) \cap I_1|).$$

We note that

$$|I_0| = |I_1| = \frac{1}{2}((r-1)(d-1) - (\gcd(r, d) - 1)),$$

which for fixed r is asymptotic to $d(r-1)/2$ as $d \rightarrow \infty$.

Thus the sum on the right hand side of (11.5.2) becomes a sum over orbits of $\langle p \rangle$ on I , and the invariant $d(o)$ is described “on average” in Section 9. In particular, the equidistribution result Proposition 9.2 implies that

$$\dim \text{III}(J_d) = \sum_{o \in O^\Delta} d(o) = \frac{1}{2}d(r-1) + \epsilon_d$$

where $\epsilon_d/d \rightarrow 0$ as $d \rightarrow \infty$.

To finish the proof, we will show that $\tau(J_d) = O(H(J_d)^\epsilon)$ for all $\epsilon > 0$ and that $\deg(\omega_{J_d}) \leq d(r-1)/2 + \epsilon_d$ where $\epsilon_d/d \rightarrow 0$ as $d \rightarrow \infty$. Once these claims are established, Proposition 4.6 implies that

$$\liminf_{d \rightarrow \infty} \text{BS}(J_d) \geq \liminf_{d \rightarrow \infty} \frac{\dim \text{III}(J_d)}{\deg(\omega_{J_d})} \geq 1.$$

Taking into account the upper bound (1.1) of Hindry and Pacheco, we conclude that

$$\lim_{d \rightarrow \infty} \text{BS}(J_d) = 1.$$

The assertion about $\tau(J_d)$ follows from the discussion of Section 2.6 and the fact (proven in [Berger et al. 2015, §3.1]) that X_d has semistable reduction at $t = 0$ and $t = \infty$ whenever r divides d .

It is proven in [Berger et al. 2015, Proof of Proposition 7.5] that when r divides d , we have $\deg(\omega_{J_d}) = d(r - 1)/2$. In general, if $d' = \text{lcm}(d, r)$, we have $\deg(\omega_{J_{d'}}) = d'(r - 1)/2$ and Lemma 2.7.1 shows that

$$\deg(\omega_{J_d}) \leq \frac{d(r - 1)}{2} + \frac{2(r - 1)^2}{d'/d} = \frac{d(r - 1)}{2} + \epsilon_d.$$

Since d'/d is an integer, ϵ_d is bounded independently of d , so $\epsilon_d/d \rightarrow 0$ as $d \rightarrow \infty$.

This completes the proof of the theorem. □

12. Quadratic twists of constant curves

We conclude the paper with a study of Brauer–Siegel ratios of quadratic twists of constant elliptic curves. Throughout we let p be an odd prime number, \mathbb{F}_q a finite field of characteristic p , and $K = \mathbb{F}_q(t)$.

12.1. Twists of a constant supersingular curve. Fix a supersingular elliptic curve E_0 over \mathbb{F}_q and let $E = E_0 \times_{\mathbb{F}_q} K$. For a positive integer d relatively prime to p , let E_d be the twist of E by the quadratic extension $\mathbb{F}_q(t, \sqrt{t^d + 1})$ of K . By results of Milne, the Tate–Shafarevich group of E_d is finite.

Theorem 12.2. *We have*

$$\lim_{\substack{d \rightarrow \infty \\ (p,d)=1}} \text{BS}(E_d) = 1.$$

Proof. Let $\mathcal{E}_d \rightarrow \mathbb{P}^1$ be the Néron model of E_d/K , and let \mathcal{C}_d be the smooth projective curve over \mathbb{F}_q defined by $y^2 = x^d + 1$ and equipped with the action of μ_2 given by the hyperelliptic involution. It is easy to see that \mathcal{E}_d is birational to the quotient of $\mathcal{C}_d \times_{\mathbb{F}_q} E_0$ by the (anti) diagonal action of μ_2 , i.e., by μ_2 acting via the hyperelliptic involution on both factors.

We are thus in position to apply the machinery of Section 6. In particular, it follows from Corollary 6.5 that

$$\dim \text{III}(E_d) = \dim \text{Hom}_A(H^1(\mathcal{C}_d)/p^n, H^1(E_0)/p^n)^{\mu_2} \tag{12.2.1}$$

for all sufficiently large n .

Section 7.3 and Proposition 7.1 describe the cohomology group $H^1(\mathcal{C}_d)$. We recall the well-known description of $H^1(E_0)$: It is a free W -module of rank 2 with a basis e_0, e_1 such that $F(e_0) = d_0 e_1$ and $F(e_1) = d_1 e_0$ where d_0 is a unit of W and d_1 is p times a unit. (See [Dummigan 1995, §5] for a detailed account.) To harmonize with earlier notation, let $J_0 = \{0\}$, $J_1 = \{1\}$, and $J = J_0 \cup J_1$, and equip J with the nontrivial action of $\langle p \rangle$.

Also, let

$$I = \mathbb{Z}/d\mathbb{Z} \setminus \left\{0, \frac{d}{2} \text{ (if } d \text{ is even)}\right\},$$

decomposed as $I_0 = \{i \mid \frac{d}{2} < i < d\}$ and $I_1 = \{i \mid 0 < i < \frac{d}{2}\}$. Section 7 and the preceding paragraph show that the crystalline cohomology groups $H^1(\mathcal{C}_d)$ and $H^1(E_0)$ with their actions of Frobenius furnish

data (M, N, I, J, c_i, d_j) as in Section 8.1, as well as the invariant $d(o)$ for each orbit o of $\langle p \rangle$ on $I \times J$. We may thus compute the dimension in the last display by the methods of Section 8.

Since \mathcal{C}_d and E_0 are hyperelliptic, the μ_2 -invariant part of their cohomology is trivial, so

$$\mathrm{Hom}_A(H^1(\mathcal{C}_d)/p^n, H^1(E_0)/p^n)^{\mu_2} = \mathrm{Hom}_A(H^1(\mathcal{C}_d)/p^n, H^1(E_0)/p^n).$$

Applying Theorem 8.3, we conclude that

$$\dim \mathrm{III}(E_d) = \sum_{o \in O} d(o) \tag{12.2.2}$$

where the sum is over all orbits of $\langle p \rangle$ on $I \times J$.

The equidistribution result Proposition 9.3 implies that

$$\sum_{o \in O} d(o) = \frac{d}{2} + \epsilon_d$$

where $\epsilon_d/d \rightarrow 0$ as $d \rightarrow \infty$.

Since $t^d + 1$ has distinct roots, it is easy to see that $\deg(\omega_{E_d}) = \lceil \frac{d}{2} \rceil$. Thus Corollary 4.7 implies that

$$\liminf_{d \rightarrow \infty} \mathrm{BS}(E_d) \geq \liminf_{d \rightarrow \infty} \frac{\dim \mathrm{III}(E_d)}{\deg(\omega_{E_d})} = 1.$$

Taking into account the upper bound (1.1) of Hindry and Pacheco, we conclude that

$$\lim_{d \rightarrow \infty} \mathrm{BS}(E_d) = 1. \quad \square$$

12.3. Twists of an constant ordinary curve. Now let E_0 be an ordinary elliptic curve over \mathbb{F}_q and set $E = E_0 \times_{\mathbb{F}_q} K$. One could use methods similar to those in the last section to compute $\dim \mathrm{III}(E_d)$ for the twist of E by $\mathbb{F}_q(t, \sqrt{t^d + 1})$, but much more is easily deduced from results of Katz in p -adic cohomology.

Theorem 12.4. *Let E' be any quadratic twist of E . Then*

$$\dim \mathrm{III}(E') = 0.$$

Proof. A variety X over a finite field is said to be *Hodge–Witt* if all of its deRham–Witt cohomology groups $H^i(X, W\Omega_X^j)$ are finitely generated. A curve is automatically Hodge–Witt, and a surface which satisfies the Tate conjecture is Hodge–Witt if and only if the dimension of its Brauer group (in the sense of Proposition/Definition 4.1) is 0 [Milne 1975, §1]. In other words, a surface X over \mathbb{F}_q satisfying the Tate conjecture is Hodge–Witt if and only if

$$\lim_{n \rightarrow \infty} \frac{\log |H^2(X \times_{\mathbb{F}_q} \mathbb{F}_{q^n}, G_m)[p^\infty]|}{\log(q^n)} = 0.$$

A theorem of Katz [1983] says that a product of varieties is Hodge–Witt if and only if one of the factors is ordinary and the other is Hodge–Witt.

Now let $\mathcal{C} \rightarrow \mathbb{P}^1$ be a double cover corresponding to a quadratic extension K'/K . Then the Néron model $\mathcal{E}' \rightarrow \mathbb{P}^1$ of E'/K is birational to the quotient of $\mathcal{C} \times_{\mathbb{F}_q} E_0$ by μ_2 acting diagonally by the hyperelliptic

involutions. Since $p > 2$, the Brauer group of the quotient is the μ_2 -invariant part of the Brauer group of $\mathcal{C} \times_{\mathbb{F}_q} E_0$, and the latter has dimension 0 since E_0 is ordinary. It follows that the Brauer group of \mathcal{E}' has dimension 0 and so $\text{III}(E')$ has dimension zero. \square

Thus for a quadratic twist of a constant, ordinary elliptic curve, our p -adic methods do not give a nontrivial lower bound on the Brauer–Siegel ratio. This is compatible with Conjecture 1.7 of [Hindry and Pacheco 2016], which predicts that the \liminf of $\text{BS}(E')$ as E' runs over all quadratic twists is 0.

We finish by remarking that Griffon [2015] has shown that if E_d is the twist of a constant ordinary E/K by the quadratic extension $\mathbb{F}_q(t, \sqrt{t^d + 1})$, then as d runs through “supersingular” integers, i.e., those that divide $p^f + 1$ for some f , the limit of $\text{BS}(E_d)$ is 1. In conjunction with Theorem 12.4, this shows that the Brauer–Siegel ratio of an elliptic curve E' may be large even when the dimension of $\text{III}(E')$ is zero.

References

- [Artin 1974] M. Artin, “Supersingular $K3$ surfaces”, *Ann. Sci. École Norm. Sup. (4)* **7** (1974), 543–567. MR Zbl
- [Berger 2008] L. Berger, “Towers of surfaces dominated by products of curves and elliptic curves of large rank over function fields”, *J. Number Theory* **128**:12 (2008), 3013–3030. MR Zbl
- [Berger et al. 2015] L. Berger, C. Hall, R. Pannekoek, J. Park, R. Pries, S. Sharif, A. Silverberg, and D. Ulmer, “Explicit arithmetic of Jacobians of generalized Legendre curves over global function fields”, 2015. arXiv
- [Bosch et al. 1990] S. Bosch, W. Lütkebohmert, and M. Raynaud, *Néron models*, *Ergebnisse der Mathematik und ihrer Grenzgebiete (3) [Results in Mathematics and Related Areas (3)]* **21**, Springer, 1990. MR Zbl
- [Brauer 1950] R. Brauer, “On the zeta-functions of algebraic number fields, II”, *Amer. J. Math.* **72** (1950), 739–746. MR Zbl
- [Conceição et al. 2014] R. P. Conceição, C. Hall, and D. Ulmer, “Explicit points on the Legendre curve II”, *Math. Res. Lett.* **21**:2 (2014), 261–280. MR Zbl
- [Conrad 2006] B. Conrad, “Chow’s K/k -image and K/k -trace, and the Lang–Néron theorem”, *Enseign. Math. (2)* **52**:1-2 (2006), 37–108. MR Zbl
- [Cossec and Dolgachev 1989] F. R. Cossec and I. V. Dolgachev, *Enriques surfaces, I*, *Progress in Mathematics* **76**, Birkhäuser, Boston, 1989. MR Zbl
- [Davis and Occhipinti 2016] C. Davis and T. Occhipinti, “Explicit points on $y^2 + xy - t^d y = x^3$ and related character sums”, *J. Number Theory* **168** (2016), 13–38. MR Zbl
- [Dummigan 1995] N. Dummigan, “The determinants of certain Mordell–Weil lattices”, *Amer. J. Math.* **117**:6 (1995), 1409–1429. MR Zbl
- [Griffon 2015] R. Griffon, “Analogue of the Brauer–Siegel theorem for some families of elliptic curves over function fields”, poster, 2015. Presented at the *Silvermania* conference at Brown University.
- [Griffon 2016] R. Griffon, *Analogues du théorème de Brauer–Siegel pour quelques familles de courbes elliptiques*, Ph.D. thesis, Université Paris Diderot, 2016, Available at http://math.richardgriffon.me/thesis/Griffon_thesis.pdf.
- [Griffon 2018] R. Griffon, “A Brauer–Siegel theorem for Fermat surfaces over finite fields”, *J. Lond. Math. Soc. (2)* **97**:3 (2018), 523–549. MR Zbl
- [Grothendieck 1968] A. Grothendieck, “Le groupe de Brauer, III: Exemples et compléments”, pp. 88–188 in *Dix exposés sur la cohomologie des schémas*, *Adv. Stud. Pure Math.* **3**, North-Holland, Amsterdam, 1968. MR Zbl
- [Halle and Nicaise 2010] L. H. Halle and J. Nicaise, “The Néron component series of an abelian variety”, *Math. Ann.* **348**:3 (2010), 749–778. MR Zbl
- [Hindry 2007] M. Hindry, “Why is it difficult to compute the Mordell–Weil group?”, pp. 197–219 in *Diophantine geometry*, edited by U. Zannier, CRM Series **4**, Ed. Norm., Pisa, 2007. MR Zbl

- [Hindry and Pacheco 2016] M. Hindry and A. Pacheco, “An analogue of the Brauer–Siegel theorem for abelian varieties in positive characteristic”, *Mosc. Math. J.* **16**:1 (2016), 45–93. MR Zbl
- [Hindry and Silverman 2000] M. Hindry and J. H. Silverman, *Diophantine geometry*, Graduate Texts in Mathematics **201**, Springer, 2000. MR Zbl
- [Kato and Trihan 2003] K. Kato and F. Trihan, “On the conjectures of Birch and Swinnerton-Dyer in characteristic $p > 0$ ”, *Invent. Math.* **153**:3 (2003), 537–592. MR Zbl
- [Katz 1983] N. M. Katz, “On the ubiquity of “pathology” in products”, pp. 139–153 in *Arithmetic and geometry, I*, edited by M. Artin and J. Tate, Progr. Math. **35**, Birkhäuser, Boston, 1983. MR Zbl
- [Koblitz 1984] N. Koblitz, *p -adic numbers, p -adic analysis, and zeta-functions*, 2nd ed., Graduate Texts in Mathematics **58**, Springer, 1984. MR Zbl
- [Lang and Néron 1959] S. Lang and A. Néron, “Rational points of abelian varieties over function fields”, *Amer. J. Math.* **81** (1959), 95–118. MR Zbl
- [Milne 1975] J. S. Milne, “On a conjecture of Artin and Tate”, *Ann. of Math. (2)* **102**:3 (1975), 517–533. MR Zbl
- [Milne 1980] J. S. Milne, *Étale cohomology*, Princeton Mathematical Series **33**, Princeton University Press, 1980. MR Zbl
- [Moret-Bailly 1985] L. Moret-Bailly, “Pinceaux de variétés abéliennes”, *Astérisque* **129** (1985), 266. MR Zbl
- [Néron 1965] A. Néron, “Quasi-fonctions et hauteurs sur les variétés abéliennes”, *Ann. of Math. (2)* **82** (1965), 249–331. MR Zbl
- [Ogg 1967] A. P. Ogg, “Elliptic curves and wild ramification”, *Amer. J. Math.* **89** (1967), 1–21. MR Zbl
- [Saito 1988] T. Saito, “Conductor, discriminant, and the Noether formula of arithmetic surfaces”, *Duke Math. J.* **57**:1 (1988), 151–173. MR Zbl
- [Serre 1988] J.-P. Serre, *Algebraic groups and class fields*, Graduate Texts in Mathematics **117**, Springer, 1988. MR Zbl
- [Shioda 1986] T. Shioda, “An explicit algorithm for computing the Picard number of certain algebraic surfaces”, *Amer. J. Math.* **108**:2 (1986), 415–432. MR Zbl
- [Shioda and Katsura 1979] T. Shioda and T. Katsura, “On Fermat varieties”, *Tôhoku Math. J. (2)* **31**:1 (1979), 97–115. MR Zbl
- [Tate 1966] J. Tate, “Endomorphisms of abelian varieties over finite fields”, *Invent. Math.* **2** (1966), 134–144. MR Zbl
- [Ulmer 1991] D. L. Ulmer, “ p -descent in characteristic p ”, *Duke Math. J.* **62**:2 (1991), 237–265. MR Zbl
- [Ulmer 2002] D. Ulmer, “Elliptic curves with large rank over function fields”, *Ann. of Math. (2)* **155**:1 (2002), 295–315. MR Zbl
- [Ulmer 2007] D. Ulmer, “ L -functions with large analytic rank and abelian varieties with large algebraic rank over function fields”, *Invent. Math.* **167**:2 (2007), 379–408. MR Zbl
- [Ulmer 2011] D. Ulmer, “Elliptic curves over function fields”, pp. 211–280 in *Arithmetic of L -functions*, edited by C. Popescu et al., IAS/Park City Math. Ser. **18**, Amer. Math. Soc., Providence, RI, 2011. MR Zbl
- [Ulmer 2013] D. Ulmer, “On Mordell–Weil groups of Jacobians over function fields”, *J. Inst. Math. Jussieu* **12**:1 (2013), 1–29. MR Zbl
- [Ulmer 2014a] D. Ulmer, “Curves and Jacobians over function fields”, pp. 283–337 in *Arithmetic geometry over global function fields*, edited by F. Bars et al., Springer, 2014. MR Zbl
- [Ulmer 2014b] D. Ulmer, “Explicit points on the Legendre curve”, *J. Number Theory* **136** (2014), 165–194. MR Zbl
- [Ulmer 2014c] D. Ulmer, “Explicit points on the Legendre curve III”, *Algebra Number Theory* **8**:10 (2014), 2471–2522. MR Zbl
- [Weil 1954] A. Weil, “Remarques sur un mémoire d’Hermite”, *Arch. Math. (Basel)* **5** (1954), 197–202. MR Zbl
- [Weil 1982] A. Weil, *Adeles and algebraic groups*, Progress in Mathematics **23**, Birkhäuser, Boston, 1982. MR Zbl

Communicated by Joseph H. Silverman

Received 2018-06-11 Revised 2019-02-27 Accepted 2019-04-02

ulmer@math.arizona.edu

Department of Mathematics, University of Arizona, Tucson, AZ, United States

A five-term exact sequence for Kac cohomology

César Galindo and Yiby Morales

Dedicated to Nicolás Andruskiewitsch on the occasion of his 60th birthday

We use relative group cohomologies to compute the Kac cohomology of matched pairs of finite groups. This cohomology naturally appears in the theory of abelian extensions of finite dimensional Hopf algebras. We prove that Kac cohomology can be computed using relative cohomology and relatively projective resolutions. This allows us to use other resolutions, besides the bar resolution, for computations. We compute, in terms of relative cohomology, the first two pages of a spectral sequence which converges to the Kac cohomology and its associated five-term exact sequence. Through several examples, we show the usefulness of the five-term exact sequence in computing groups of abelian extensions.

1. Introduction

Extension theory of groups plays a significant role in the construction and the classification of finite groups. In the same way, the extension theory of Hopf algebras has led to results on the still wide open problem of construction and classification of finite-dimensional semisimple Hopf algebras, [Kashina 2000; Masuoka 1995; Natale 1999; 2001; 2004]. The set of equivalence classes of extensions of a group G by a G -module M is an abelian group with the Baer product of extensions, which is isomorphic to the second cohomology group $H^2(G, M)$. A generalization of this theory to Hopf algebras is obtained for the so-called *abelian extensions*, that is, cleft Hopf algebra extensions of a commutative Hopf algebra K by a cocommutative Hopf algebra H , see [Hofstetter 1994; Kac 1969; Masuoka 1997a; 1999; 2000; Singer 1972].

In this paper, we deal with $H = kF$, a group algebra, and $K = k^G$, the dual of such an algebra, where F and G are finite groups. In this case, each abelian extension has an associated matched pair, that is, a larger group Σ such that G and F are subgroups of Σ satisfying $G \cap F = \{e\}$ and $\Sigma = GF$. The set of equivalence classes of abelian extensions associated to a fixed matched pair, denoted by $\text{Opext}(kF, k^G)$, is an abelian group that can be computed as the second total cohomology of a certain double complex whose cohomology is called *Kac cohomology*.

Obtaining a computation for the $\text{Opext}(kF, k^G)$ of a matched pair of groups can be quite difficult. In fact, there are few general computations in the literature [Masuoka 1997a]. One obstacle for the computation of $\text{Opext}(kF, k^G)$ comes from the fact that it is defined as the cohomology of a very specific

C.G. is partially supported by Faculty of Science of Universidad de los Andes, Convocatoria 2018–2019 para la Financiación de Programas de Investigación, programa "Simetría T (inversión temporal) en categorías de fusión y modulares".

MSC2010: 16T05.

Keywords: Hopf algebras, relative cohomology, abelian extensions of Hopf algebras.

total complex, and the unique “cocycle free” tool is the so-called *Kac exact sequence* (see [Masuoka 1997a] and Corollary 3.10). Perhaps one of the first results that provide a cocycle free description and interpretation of the Kac cohomology is given in [Baaj et al. 2005, Proposition 7.1], where the authors describe the Kac cohomology as the *singular cohomology* of the mapping cone $BG \sqcup BF \rightarrow B\Sigma$.

In this paper, we use two different kinds of relative cohomology groups to compute $\text{Opext}(kF, k^G)$: Auslander relative cohomology and Hochschild relative cohomology (see Section 3). We prove that $\text{Opext}(kF, k^G)$ can be computed using Auslander relative cohomology and that Auslander relative cohomology of a matched pair can be computed using relatively projective resolutions. This allows us to use other resolutions, besides the bar resolution, for computing $\text{Opext}(kF, k^G)$. In addition, we compute the first and second page of a spectral sequence which converges to the Kac cohomology. As a consequence, we compute the associated five-term exact sequence, whose second term is $\text{Opext}(kF, k^G)$. In the particular case of a semidirect product, the five-term exact sequence is described in terms of ordinary group cohomology. Finally, doing use of the five-term exact sequence and some nonstandard resolutions, we compute $\text{Opext}(kF, k^G)$ for several families of matched pairs.

The organization of the paper is as follows: In Section 2 we discuss preliminaries on group cohomology and abelian extensions of Hopf algebras. In Section 3 we recall the definitions of Auslander relative cohomology [Auslander and Solberg 1993] and Hochschild relative cohomology [Hochschild 1956]. We prove that $\text{Opext}(kF, k^G)$ can be computed using Auslander Relative cohomology and Auslander relative cohomology of matched pairs can be computed using relatively projective resolutions. In Section 4 we compute the first and second page of a spectral sequence which converges to the Kac cohomology. We also compute, in terms of Hochschild relative cohomology, the associated five-term exact sequence, whose second term is $\text{Opext}(kF, k^G)$. Finally, in Section 5 we compute $\text{Opext}(kF, k^G)$ for several families of matched pairs.

2. Preliminaries

Cohomology of groups. Let G be a group and let M be a G -module. The n -th cohomology group of G with coefficients in M is defined as

$$H^n(G, M) = \text{Ext}_G^n(\mathbb{Z}, M).$$

We will use occasionally the *normalized bar resolution* $(\mathbb{Z} \xleftarrow{\epsilon} P_i, \delta)$ of \mathbb{Z} as a trivial G -module. That is

$$P_i = \mathbb{Z}G[G]^i := \mathbb{Z}G[s_1 | \cdots | s_i],$$

where $s_i \in G$, $s_i \neq e$ for all i . The differentials are given by

$$\delta([s_1 | \cdots | s_{p+1}]) = s_1[s_2 | \cdots | s_{p+1}] + \sum_{i=1}^p (-1)^i [s_1 | \cdots | s_i s_{i+1} | \cdots | s_{p+1}] + (-1)^{p+1} [s_1 | \cdots | s_p].$$

For a finite cyclic group $C_n = \langle g \rangle$ of order n , we occasionally use a periodic resolution of \mathbb{Z} , defined as

$$\dots \rightarrow \mathbb{Z}C_n \xrightarrow{g-1} \mathbb{Z}C_n \xrightarrow{N} \mathbb{Z}C_n \xrightarrow{g-1} \mathbb{Z}C_n \xrightarrow{\epsilon} \mathbb{Z}, \tag{2-1}$$

where, $N = \sum_{i=0}^{n-1} g^i$. From this, we have that

$$H^m(C_n, M) = \begin{cases} M^{C_n}/\text{Im}(N) & m = 2k, \\ \text{Ker}(N)/\text{Im}(g-1) & m = 2k + 1. \end{cases} \tag{2-2}$$

Resolutions and cohomology for direct products. Let $G = G_1 \times G_2$ be a direct product of groups. Let $(\mathbb{Z} \xleftarrow{\epsilon} P_i, \delta_i)$ and $(\mathbb{Z} \xleftarrow{\epsilon} Q_i, \delta'_i)$ be projective resolutions of \mathbb{Z} as a G_1 -module and a G_2 -module, respectively. Then, the total complex $\text{Tot}(P_i \otimes Q_i)$ is a G -projective resolution of \mathbb{Z} . A useful description of the cohomology for direct products is the following. If M is a trivial G -module, then it holds that (see e.g., [Karpilovsky 1985])

$$H^2(G_1 \times G_2, M) \cong H^2(G_1, M) \oplus H^2(G_2, M) \oplus P(G_1, G_2; M), \tag{2-3}$$

where $P(G_1, G_2; M)$ is the abelian group of all pairings from $G_1 \times G_2$ to M . An isomorphism is given by $\bar{\alpha} \mapsto (\bar{\alpha}_1, \bar{\alpha}_2, \phi_\alpha)$, where α_i is the restriction of α to $Q_i \times Q_i$ and

$$\phi_\alpha(x, y) = \text{Alt}(\alpha) = \alpha(x, y) - \alpha(y, x).$$

the inverse isomorphism is defined by $(\bar{\alpha}_1, \bar{\alpha}_2, \phi) \mapsto \bar{\alpha}$ with

$$\alpha((x_1, y_1), (x_2, y_2)) = \alpha_1(x_1, x_2)\alpha_2(y_1, y_2)\phi(x_1, y_2).$$

Second cohomology group and skewsymmetric matrices. Another useful description of the second cohomology group in the case that V is a finite abelian group is provided by using the universal coefficient theorem. Let k be a field and let us consider the group k^\times of units of k as a trivial V -module. There is a short exact sequence

$$0 \rightarrow \text{Ext}(V, k^\times) \rightarrow H^2(V, k^\times) \xrightarrow{\text{Alt}} \text{Hom}(\wedge^2(V), k^\times) \rightarrow 0. \tag{2-4}$$

If V has exponent n and $(k^\times)^n = k^\times$, then $\text{Ext}(V, k^\times) = 0$. Therefore, the map

$$\text{Alt} : H^2(V, k^\times) \rightarrow \wedge^2 \hat{V}, \tag{2-5}$$

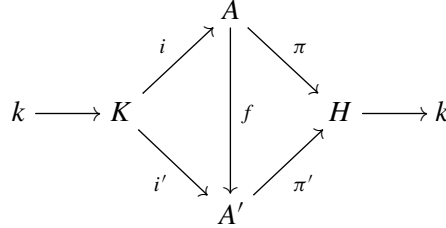
where $\hat{V} = \text{Hom}(V, k^\times)$, defines an isomorphism $H^2(V, k^\times) \cong \wedge^2 \hat{V}$.

Extensions of Hopf algebras. Let k be a field. A sequence of *finite dimensional* Hopf algebras and Hopf algebra maps

$$(A) : k \rightarrow K \xrightarrow{i} A \xrightarrow{\pi} H \rightarrow k$$

is called an *extension* of H by K if, i is injective, π is surjective, and $K = A^{\text{co}H}$ (see [Andruskiewitsch and Devoto 1995; Kac 1969; Majid 1990]).

Two extensions (A) and (A') of H by K are said to be *equivalent* if there is an homomorphism $f : A \rightarrow A'$ of Hopf algebras such that the following diagram commutes:



Matched pairs of groups. Let us recall (see e.g., [Takeuchi 1981]) that a *matched pair of groups* is a collection $(F, G, \triangleright, \triangleleft)$ where G, F are groups and $\triangleright, \triangleleft$ are permutation actions

$$G \xleftarrow{\triangleleft} G \times F \xrightarrow{\triangleright} F$$

such that

$$s \triangleright xy = (s \triangleright x)((s \triangleleft x) \triangleright y), \quad st \triangleleft x = (s \triangleleft (t \triangleright x))(t \triangleleft x),$$

for all $s, t \in G$ and $x, y \in F$.

Having groups G and F with a matched pair structure is equivalent to having a group Σ with an exact factorization; the actions \triangleright and \triangleleft are determined by the relations

$$sx = (s \triangleright x)(s \triangleleft x),$$

where $x \in F$ and $s \in G$.

The group Σ associated to a matched pair of groups will be denoted by $F \bowtie G$; it is $F \times G$ with product given by

$$(x, s)(y, t) = (x(s \triangleright y), (s \triangleleft y)t).$$

It is easy to see that the following conditions are equivalent:

- (i) The action \triangleright is trivial.
- (ii) The action $\triangleleft : G \times F \rightarrow G$ is by group automorphisms.

In this case, the associated group $\Sigma = F \bowtie G$ is a semidirect product of groups.

Abelian extensions. Let $(F, G, \triangleright, \triangleleft)$ be a matched pair of groups and let us consider 2-cocycles $\sigma \in Z^2(F, (k^G)^\times)$ and $\tau \in Z^2(G, (k^F)^\times)$. On the vector space

$$k^G \#_{\sigma, \tau} k^F := \text{Span}_k \{e_s \# x : s \in G, x \in F\},$$

we can define a unital associative algebra and counital coassociative coalgebra structure by

$$\begin{aligned}
 (e_s \# x)(e_t \# y) &= \delta_{s \triangleleft x, t} \sigma(s; x, y) e_s \# xy, \\
 \Delta(e_s \# x) &= \sum_{s=ab} \tau(a, b; x) e_a \# (b \triangleright x) \otimes e_b \# x.
 \end{aligned}$$

Here the 2-cocycles σ and τ are seen as functions

$$\begin{aligned} \sigma &: G \times F \times F \rightarrow k^\times, & (s, x, y) &\mapsto \sigma(s; x, y), \\ \tau &: G \times G \times F \rightarrow k^\times, & (s, t, x) &\mapsto \tau(s, t; x). \end{aligned}$$

The map Δ is an algebra map if and only if the 2-cocycles satisfy the following compatibility condition

$$\sigma(st; x, y)\tau(s, t; xy) = \sigma(s; t \triangleright x, (t \triangleleft x) \triangleright y)\sigma(t; x, y) \times \tau(s, t; x)\tau(s \triangleleft (t \triangleright x), t \triangleleft x; y),$$

for all $x, y \in G$ and $s, t \in F$. In the case that the 2-cocycles are compatible, the sequence

$$k \rightarrow k^G \xrightarrow{i} k^G \#_{\sigma, \tau} kF \xrightarrow{\pi} kF \rightarrow k,$$

is a Hopf algebra extension, where $i(e_s) = e_s \# e$ and $\pi(e_s \# x) = x$. These kinds of extensions are called *abelian extensions* of Hopf algebras.

The set of equivalence classes of abelian extensions associated to a fixed matched pair $(F, G, \triangleright, \triangleleft)$ is an abelian group with the Baer product of extensions and will be denoted by $\text{Opext}_{\triangleright, \triangleleft}(kF, k^G)$ (see [Masuoka 2002] for more details).

3. Kac cohomology and relative cohomology

In this section, we recall the definitions of two different kinds of relative group cohomology: the Auslander relative cohomology [Auslander and Solberg 1993] and the Hochschild relative cohomology [Hochschild 1956]. Our aim is to prove that $\text{Opext}_{\triangleright, \triangleleft}(kF, k^G)$ can be computed using Auslander relative cohomology which, in the case of matched pairs, can be computed using relatively projective resolutions. This allows us to use other resolutions besides the bar resolution for computing $\text{Opext}_{\triangleright, \triangleleft}(kF, k^G)$.

Auslander relative cohomology of groups. Let Σ be a group and X a Σ -set. We will denote by Λ_X the kernel of the augmentation map

$$\epsilon_X : \mathbb{Z}[X] \rightarrow \mathbb{Z}, \quad x \mapsto 1, \tag{3-1}$$

where $\mathbb{Z}[X]$ is the Σ -module associated to X .

Definition 3.1. Given a Σ -module A , the n -th cohomology group of Σ relative to X with coefficients in A is defined by

$$H_{\mathcal{A}}^k(\Sigma, X; A) := \text{Ext}_{\mathbb{Z}\Sigma}^{k-1}(\Lambda_X, A), \quad k \geq 1.$$

Let X be a Σ -set and \mathcal{R}_X a set of representatives of the Σ -orbits in X . Using Shapiro’s lemma, we have that

$$\text{Ext}_{\Sigma}^k(\mathbb{Z}[X], A) = \prod_{x \in \mathcal{R}_X} \text{Ext}_{\Sigma}^k(\mathbb{Z}[\mathcal{O}(x)], A) \cong \prod_{x \in \mathcal{R}_X} \text{Ext}_{\text{St}(x)}^k(\mathbb{Z}, A),$$

where $\text{St}(x)$ denotes the stabilizer of $x \in X$. Hence,

$$\text{Ext}_{\Sigma}^k(\mathbb{Z}[X], A) = \prod_{x \in \mathcal{R}_X} H^k(\text{St}(x), A).$$

If we apply the functor $\text{Ext}_\Sigma(-, A)$ to the exact sequence of Σ -modules

$$0 \rightarrow \Lambda_X \rightarrow \mathbb{Z}[X] \rightarrow \mathbb{Z} \rightarrow 0,$$

we obtain the well-known long exact sequence for relative cohomology

$$\dots \rightarrow H^k(\Sigma, A) \rightarrow \prod_{x \in \mathcal{R}_X} H^k(\text{St}(x), A) \rightarrow H_{\mathcal{S}}^{k+1}(\Sigma, X, A) \rightarrow H^{k+1}(\Sigma, A) \rightarrow \dots \quad (3-2)$$

Hochschild Relative cohomology of groups. Relative cohomology of groups was originally defined by Hochschild [1956] and Adamson [1954]. We follow the description given in [Alperin 1986].

Let U be a G -module and S a subgroup of G . We say that U is *relatively S -projective* if it satisfies the following equivalent properties (see [Alperin 1986, Proposition 1, page 65]):

- (i) If $\psi : U \rightarrow V$ is a surjective G -homomorphism and ψ splits as an S -homomorphism then ψ splits as a G -homomorphism.
- (ii) If $\psi : V \rightarrow W$ is a surjective G -homomorphism and $\phi : U \rightarrow W$ is a G -homomorphism, then there is a G -homomorphism $\lambda : U \rightarrow V$ with $\psi\lambda = \phi$, provided that there is an S -homomorphism $\lambda_0 : U \rightarrow V$ with the same property.
- (iii) U is a direct summand of $U \downarrow_S \uparrow_G$.

Here, \downarrow_S means the restriction to S , and \uparrow_G the induction to G .

A complex

$$\mathcal{R} : \dots \rightarrow R_3 \xrightarrow{\delta_3} R_2 \xrightarrow{\delta_2} R_1 \xrightarrow{\delta_1} R_0 \xrightarrow{\epsilon} M \rightarrow 0$$

of G -modules is called a *relatively S -projective resolution* if:

- (1) each G -module R_i is relatively S -projective,
- (2) the sequence has a contracting homotopy as S -modules.

Remarks 3.2. • Since the canonical map $M \downarrow_S \uparrow_G \rightarrow M$ splits as an S -homomorphism, if \mathcal{T} is a projective S -resolution of $M \downarrow_S$, then $\mathcal{T} \uparrow_G$ is a relatively S -projective resolution of M .

- If S is the trivial subgroup of G , the relatively S -projective resolutions of G -module are the same as projective resolutions of G -modules.

Definition 3.3. Given a relatively S -projective resolution \mathcal{R} of M , the n -th relative S -cohomology group of G is defined by

$$H^m(G, S; M) = H^n(\text{Hom}_G(\mathcal{R}, M), \delta^*).$$

As expected, this definition does not depend on the chosen relatively S -projective resolution of M , (see, e.g., [Hochschild 1956]).

From now on, all relatively projective resolutions are assumed to be free as \mathbb{Z} -modules.

Example 3.4 (Standard complex [Snapper 1964]). Let X be a transitive left G -set. Let $C_i = \mathbb{Z}X^{(i+1)}$ be the free \mathbb{Z} -module generated by all $(i + 1)$ -tuples of elements of X . The group G acts diagonally on C_i and the sequence

$$C_*^X := \dots \rightarrow C_3 \xrightarrow{\delta_3} C_2 \xrightarrow{\delta_2} C_1 \xrightarrow{\delta_1} C_0 \xrightarrow{\epsilon} \mathbb{Z} \rightarrow 0,$$

where

$$\delta_i(x_1, \dots, x_{r+1}) = \sum_{j=1}^{r+1} (-1)^{j+1} (x_1, \dots, \widehat{x}_j, \dots, x_{r+1}) \tag{3-3}$$

and $\epsilon(x) = 1$ for all $x \in X$, is a complex of G -modules.

If F denotes the stabilizer subgroup of $x_0 \in X$, the complex C_*^X is relatively F -projective resolution of \mathbb{Z} called the *standard complex of (G, X)* .

Proposition 3.5. *Let $\Sigma = F \bowtie G$ be a matched pair and $Q := (\mathbb{Z} \xleftarrow{\epsilon_Q} Q_i, \delta'_i)$ be the normalized right bar resolution of the trivial G -module \mathbb{Z} . Then the group Σ acts on Q_i by*

$$[s_i | \dots | s_2 | s_1] s_0 \cdot (x, s) = [s_i \triangleleft ((s_{i-1} \dots s_0) \triangleright x) | \dots | s_1 \triangleleft (s_0 \triangleright x)] (s_0 \triangleleft x) s. \tag{3-4}$$

and Q is a relatively F -projective resolution of right Σ -modules.

Proof. Since Σ is a matched pair, the right Σ -set of cosets $F \backslash \Sigma$ can be identified with the set G and Σ -action $s \cdot (f, g) = (s \triangleleft f)g$. This Σ -set will be denoted by X . Applying the construction of Example 3.4 to X , we obtain a relatively F -projective resolution $C := C_i \xrightarrow{\epsilon} \mathbb{Z}$, since F is the stabilizer of $e \in G$. The resolution C coincides with the standard G -free resolution of \mathbb{Z} as a trivial G -module. A G -basis of C_i (called bar basis) is given by

$$[s_i | \dots | s_2 | s_1] = (s_i \dots s_1, \dots, s_2 s_1, s_1, e).$$

The action of Σ in this basis is given by (3-4). The normalized bar resolution is a quotient of the bar resolution, and it is easy to see that this is also relatively F -projective. □

Remarks 3.6. • In Proposition 3.5, we may consider the normalized *left bar resolution* $(\mathbb{Z} \xleftarrow{\epsilon_P} P_i, \delta'_i)$ for the trivial G -module \mathbb{Z} : with action of Σ given by

$$(s, x) \cdot x_0 [x_1 | \dots | x_i] = x (s \triangleright x_0) [(s \triangleleft x_0) \triangleright x_1 | \dots | (s \triangleleft x_0 x_1 \dots x_{i-1}) \triangleright x_i]. \tag{3-5}$$

This is a relatively G -projective resolution.

- The formulas (3-4) and (3-5) appear in [Masuoka 1997a].

Theorem 3.7. *Let $\Sigma = F \bowtie G$ be a matched pair. Let $(\mathbb{Z} \xleftarrow{\epsilon_P} P_i, \delta'_i)$ and $(\mathbb{Z} \xleftarrow{\epsilon_Q} Q_i, \delta'_i)$ be a relatively F -projective and a relatively G -projective Σ -resolutions of \mathbb{Z} , respectively. Then the total complex of the tensor product double complex*

$$P_i \otimes Q_j \quad \text{for } i, j \geq 0.$$

is a projective Σ -resolution of \mathbb{Z} .

Proof. Since P_i is relatively F -projective, $P_i \downarrow_F \uparrow^\Sigma = P_i \oplus P'_i$ as Σ -modules. Analogously, $Q_j \downarrow_G \uparrow^\Sigma = Q_j \oplus Q'_j$.

Using the fact that (F, G) is an exact factorization of Σ and the Mackey's tensor product theorem (see e.g., [Curtis and Reiner 1990, Theorem 10.18]), we have that

$$\begin{aligned} (P_i \otimes Q_j) \downarrow_{\{e\}} \uparrow^\Sigma &\cong P_i \downarrow_F \uparrow^\Sigma \otimes Q_j \downarrow_G \uparrow^\Sigma, \\ &= (P_i \oplus P'_i) \otimes (Q_j \oplus Q'_j) \\ &= P_i \otimes Q_j \oplus (P_i \otimes Q'_j \oplus P'_i \otimes Q_j \oplus P'_i \otimes Q_j \oplus P'_i \otimes Q'_j). \end{aligned}$$

Hence $P_i \otimes Q_j$ is direct summand of the Σ -free module $(P_i \otimes Q_j) \downarrow_{\{e\}} \uparrow^\Sigma$, that is, $P_i \otimes Q_j$ is a projective Σ -module.

Finally, since each P_i and Q_j are flat \mathbb{Z} -modules, it follows from the Künneth formula that, for $n > 0$, $H_n(\text{Tot}(P_* \otimes Q_*)) = 0$. □

Theorem 3.7 generalizes the results of [Masuoka 1997b; 2003] about the construction of nonstandard free resolutions of \mathbb{Z} associated to a matched pairs of groups. In fact, taking the relatively projective resolutions of Proposition 3.5 and Remarks 3.6 we can obtain the resolutions in [Masuoka 1997b; 2003].

Kac cohomology as relative group cohomology. Using the bijective maps,

$$\begin{aligned} \Sigma/G &\rightarrow F, & (f, g)G &\mapsto f, \\ F \backslash \Sigma &\rightarrow G, & F(f, g) &\mapsto g \end{aligned}$$

we can endow the set F with a left Σ -action $(f, g)x = f(g \triangleright x)$ and G with a right Σ -action $s(f, g) = (s \triangleleft f)g$. From now on, X will denote the left Σ -set defined as the disjoint union $F \sqcup G$, where G is considered a left Σ -set using the inverse. We denote by Λ_X the kernel of the augmentation map (3-1).

Proposition 3.8. *Let $\Sigma = F \bowtie G$ be a matched pair. Let $(\mathbb{Z} \xleftarrow{\epsilon_P} P_i, \delta_i)$ and $(\mathbb{Z} \xleftarrow{\epsilon_Q} Q_j, \delta'_j)$ be a relatively F -projective and a relatively G -projective Σ -resolutions of \mathbb{Z} , respectively. Let $D_{*,*}$ be the truncated tensor product double complex*

$$D_{i,j} := P_{i+1} \otimes Q_{j+1}, \quad \text{for } i, j \geq 0. \tag{3-6}$$

Then, the total complex $(\text{Tot}(D_{,*}), d_i)$ completed with the map*

$$P_0 \otimes Q_0 \xleftarrow{\overline{\delta_1 \otimes \delta'_1}} \text{Tot}(D_{*,*}),$$

is a projective Σ -resolution of Λ_X .

Proof. Let Λ be the kernel of the map $\epsilon : P_0 \oplus Q_0 \rightarrow \mathbb{Z}$ defined by $\epsilon(x \oplus y) = \epsilon_P(x) + \epsilon_Q(y)$. Let us consider the total complex $(\text{Tot}(D_{*,*}), d_i)$ completed with the maps

$$0 \leftarrow \Lambda \xleftarrow{\theta} P_0 \otimes Q_0 \xleftarrow{\overline{\delta_1 \otimes \delta'_1}} \text{Tot}(D_{*,*}), \tag{3-7}$$

where $\theta(p \otimes q) = -p\epsilon_Q(q) \oplus \epsilon_P(p)q$. Let us see that $H_n(\text{Tot}(D_{*,*})) = 0$ for all $n \geq 1$. Let $B_{*,*}$ be the subcomplex of $P_* \otimes Q_*$ consisting of the first row and first column, that is,

$$\begin{aligned} B_{i,j} &= 0 && \text{for } i, j \geq 1, \\ B_{i,j} &= P_i \otimes Q_j && \text{for } i = 0 \vee j = 0. \end{aligned}$$

Let $S_{*,*} = P_* \otimes Q_*/B_{*,*}$. Note that $\text{Tot}_n(D_{*,*}) = \text{Tot}_{n+2}(S_{*,*})$ for all n .

Let us now see that (3-7) is exact in $\text{Tot}_0(D_{*,*}) = P_1 \otimes Q_1$. The map $d_1 : P_2 \otimes Q_1 \oplus P_1 \otimes Q_2 \rightarrow P_1 \otimes Q_1$ in $\text{Tot}(D_{*,*})$ is defined by $d_1(a \oplus b) = (\delta_2 \otimes \text{id})(a) - (\text{id} \otimes \delta'_2)(b)$, where $a \in P_2 \otimes Q_1$ and $b \in P_1 \otimes Q_2$. Hence, the composed map is given by

$$(-\delta_1 \otimes \delta'_1) \circ d_1(a, b) = (-\delta_1 \otimes \delta'_1)(\delta_2 \otimes \text{id})(a) + (\delta_1 \otimes \delta'_1)(\text{id} \otimes \delta'_2)(b) = 0.$$

Thus, $\text{Im}(d_1) \subseteq \text{Ker}(-\delta_1 \otimes \delta'_1)$. Now, since P_i and Q_i are free \mathbb{Z} -modules, then

$$\text{Ker}(-\delta_1 \otimes \delta'_1) = \text{Ker}(\delta_1) \otimes Q_1 + P_1 \otimes \text{Ker}(-\delta'_1) = \text{Im}(\delta_2) \otimes Q_1 - P_1 \otimes \text{Im}(\delta'_2).$$

so $\text{Ker}(-\delta_1 \otimes \delta'_1) \subseteq \text{Im}(d_1)$. To see the exactness in $P_0 \otimes Q_0$, note that

$$\theta \circ (\delta_1 \otimes \delta'_1)(p \otimes q) = \delta_1(p)\epsilon_Q(\delta'_1(q)) \oplus \epsilon_P(\delta_1(p))\delta_1(q) = 0.$$

Hence, $\text{Im}(\delta_1 \otimes \delta'_1) \subseteq \text{Ker}(\theta)$. To see that $\text{Ker}(\theta) \subseteq \text{Im}(\delta_1 \otimes \delta'_1)$, let us consider the tensor product of the two chain complexes $0 \leftarrow \mathbb{Z} \xleftarrow{\epsilon_P} P_i$ and $0 \leftarrow \mathbb{Z} \xleftarrow{\epsilon_Q} Q_i$. That is,

$$\begin{array}{ccccc} \mathbb{Z} \otimes Q_1 & \xleftarrow{\epsilon_P \otimes \text{id}} & P_0 \otimes Q_1 & \xleftarrow{\delta_1 \otimes \text{id}} & P_1 \otimes Q_1 & & (3-8) \\ \text{id} \otimes \delta'_1 \downarrow & & \downarrow \text{id} \otimes \delta'_1 & & \downarrow \text{id} \otimes \delta'_1 & & \\ \mathbb{Z} \otimes Q_0 & \xleftarrow{\epsilon_P \otimes \text{id}} & P_0 \otimes Q_0 & \xleftarrow{\delta_1 \otimes \text{id}} & P_1 \otimes Q_0 & & \\ \text{id} \otimes \epsilon_Q \downarrow & & \downarrow \text{id} \otimes \epsilon_Q & & \downarrow \text{id} \otimes \epsilon_Q & & \\ \mathbb{Z} & \xleftarrow{\epsilon_P \otimes \text{id}} & P_0 \otimes \mathbb{Z} & \xleftarrow{\delta_1 \otimes \text{id}} & P_1 \otimes \mathbb{Z} & & \end{array}$$

whose total complex is given by the exact sequence

$$0 \leftarrow \mathbb{Z} \otimes \mathbb{Z} \xleftarrow{\epsilon} P_0 \otimes \mathbb{Z} \oplus \mathbb{Z} \otimes Q_0 \xleftarrow{d_1} P_1 \otimes \mathbb{Z} \oplus P_0 \otimes Q_0 \oplus \mathbb{Z} \otimes Q_1 \leftarrow \dots \quad (3-9)$$

which can be written as

$$\mathbb{Z} \xleftarrow{\epsilon} P_0 \oplus Q_0 \xleftarrow{d_1} P_1 \oplus P_0 \otimes Q_0 \oplus Q_1 \leftarrow \dots \quad (3-10)$$

Note that θ is the restriction of d_1 to $P_0 \otimes Q_0$. Suppose that $c \in \text{Ker}(\theta)$. Then, $(\text{id} \otimes \epsilon_Q)(c) = (\epsilon_P \otimes \text{id})(c) = 0$. Let b_2, b_3 be the preimages of c under the vertical and horizontal differentials in (3-8) respectively. There is a tuple $b = (b_1, b_2, -b_3, -b_4)$ such that $d_2(b) = 0$. Since the total complex of (3-8) is acyclic, there is a tuple $a = (a_1, a_2, a_3, a_4, a_5)$ such that $d_3(a) = b$, and, it can be verified that $a_3 \in P_1 \otimes Q_1$ satisfies $\theta(a_3) = c$. Therefore, $\text{Ker}(\theta) \subseteq \text{Im}(\delta_1 \otimes \delta'_1)$.

To see that θ is surjective, note that $\epsilon = \epsilon_P + \epsilon_Q$ in (3-10), and

$$d_1(a \oplus b \oplus c) = (\delta_1(a) + (\text{id} \otimes \epsilon_Q)(b)) \oplus ((\epsilon_P \otimes \text{id})(b) - \delta'_1(c)).$$

Let $p_0 \in d_1(P_1) = \text{Ker}(\epsilon_P)$ and let $q \in Q_0$ such that $\epsilon_Q(q) = 1$. Then,

$$d_1(p_0 \otimes q) = (-\text{id} \otimes \epsilon_Q + \epsilon_P \otimes \text{id})(p_0 \otimes q) = p_0,$$

so $d_1(P_1) \subseteq d_1(P_0 \otimes Q_0)$. Similarly, $d_1(Q_1) \subseteq d_1(P_0 \otimes Q_0)$. Hence,

$$\Lambda = \text{Ker}(\epsilon) = \text{Im}(d_1) = \text{Span}\{d_1(P_1) \cup d_1(P_0 \otimes Q_0) \cup d_1(Q_1)\} = d_1(P_0 \otimes Q_0).$$

Also, the map d_1 restricted to $P_0 \otimes Q_0$ is given by $-\text{id} \otimes \epsilon_Q \oplus \epsilon_P \otimes \text{id} = \theta$, then $\Lambda = \text{Im}(\theta)$ and therefore the sequence (3-7) is exact.

Finally, we see that Λ and Λ_X are isomorphic as Σ -modules. Let us take $P'_* = C_*^G$ and $Q'_* = C_*^F$, the standard resolutions as in Example 3.4, where F and G are considered as Σ -sets. In this case,

$$C_0^G \oplus C_0^F = \mathbb{Z}[G] \oplus \mathbb{Z}[F] \cong \mathbb{Z}[G \sqcup F] = \mathbb{Z}[X],$$

and ϵ is the augmentation map (3-1). Then, for this resolution $\Lambda = \Lambda_X$.

If $(\mathbb{Z} \xleftarrow{\epsilon_P} P_i, \delta_i)$ and $(\mathbb{Z} \xleftarrow{\epsilon_Q} Q_i, \delta'_i)$ are relatively projective resolutions, there exists homotopy equivalences

$$s : P_* \rightarrow C_*^G, \quad l : Q_* \rightarrow C_*^F$$

This implies that $P_0 \otimes Q_0 \xleftarrow{\delta_1 \otimes \delta'_1} \text{Tot}(D_{*,*})$ is homotopically equivalent to $P'_0 \otimes Q'_0 \xleftarrow{\delta_1 \otimes \delta'_1} \text{Tot}(D_{*,*})$. Hence,

$$\Lambda_X \cong \text{Coker}(P'_1 \otimes Q'_1 \rightarrow P'_0 \otimes Q'_0) \cong \text{Coker}(P_1 \otimes Q_1 \rightarrow P_0 \otimes Q_0) \cong \Lambda. \quad \square$$

Theorem 3.9. *Let k be a field and $\Sigma = F \bowtie G$ a matched pair of groups. Then,*

$$\text{Opext}_{\triangleright, \triangleleft}(kF, k^G) \cong H^3_{\mathcal{A}}(\Sigma, X; k^\times),$$

where k^\times is considered as a trivial Σ -module.

Proof. Let $(\mathbb{Z} \xleftarrow{\epsilon_P} P_i, \delta_i)$ and $(\mathbb{Z} \xleftarrow{\epsilon_Q} Q_i, \delta'_i)$ be the resolutions in Proposition 3.5. These are the resolutions used in [Masuoka 1997b] to compute $\text{Opext}_{\triangleright, \triangleleft}(kF, k^G)$; they consider the truncated tensor product $D^{i,j}$ of the two resolutions to get

$$\text{Opext}_{\triangleright, \triangleleft}(kF, k^G) \cong H^1(\text{Tot}(\text{Hom}_\Sigma(D^{i,j}))).$$

If Λ is the kernel of the map $\epsilon : P_0 \oplus Q_0 \rightarrow \mathbb{Z}$ defined by $\epsilon(x \oplus y) = \epsilon_P(x) + \epsilon_Q(y)$, (here, $P_0 := \mathbb{Z}F = \mathbb{Z}X$ and $Q_0 := \mathbb{Z}Y$ for the Σ -sets $X = F$ and $Y = G$) then, by Proposition 3.8, the total complex of $E_{i,j} := P_{i+1} \otimes Q_{j+1}$ for $i, j \geq 0$ completed in the following way

$$0 \longleftarrow \Lambda \xleftarrow{\theta} P_0 \otimes Q_0 \xleftarrow{\delta_1 \otimes \delta'_1} \text{Tot}(D_{*,*}),$$

is a resolution of Λ , and $\Lambda = \Lambda_X$. Therefore, if we apply $\text{Hom}_\Sigma(-, k^\times)$ to this total complex, we get the relative cohomology groups $H_{\mathcal{A}}^n(\Sigma, X; A)$. That is

$$H^k(\text{Hom}_\Sigma(\text{Tot}(D_{*,*})) \cong H_{\mathcal{A}}^{k+2}(\Sigma, X; A). \tag{3-11}$$

In particular, since $\text{Hom}_\Sigma(\text{Tot}(D_{*,*})) \cong \text{Tot}(\text{Hom}_\Sigma(D_{*,*}))$, then,

$$\text{Opext}_{\triangleright, \triangleleft}(kF, k^G) \cong H^1(\text{Tot}(\text{Hom}_\Sigma(D_{*,*})) \cong H_{\mathcal{A}}^3(\Sigma, X; k^\times), \tag{3-12}$$

which completes the proof. □

As a consequence of Theorem 3.9 and the long exact sequence (3-2) we obtain Kac's exact sequence (see [Masuoka 1997a; Kac 1969]).

Corollary 3.10 (Kac's exact sequence). *For a fixed matched pair of groups $(F, G, \triangleright, \triangleleft)$, we have a long exact sequence*

$$\begin{aligned} 0 \rightarrow H^1(F \bowtie G, k^\times) &\rightarrow H^1(F, k^\times) \oplus H^1(G, k^\times) \rightarrow H_{\mathcal{A}}^3(\Sigma, X; k^\times) \\ &\rightarrow H^2(F \bowtie G, k^\times) \rightarrow H^2(F, k^\times) \oplus H^2(G, k^\times) \rightarrow \text{Opext}_{\triangleright, \triangleleft}(kF, k^G) \\ &\rightarrow H^3(F \bowtie G, k^\times) \rightarrow H^3(F, k^\times) \oplus H^3(G, k^\times) \rightarrow H_{\mathcal{A}}^4(\Sigma, X; k^\times). \end{aligned}$$

4. The five-term exact sequence for Kac double complex

The group $\text{Opext}_{\triangleright, \triangleleft}(kF, k^G)$ can be obtained, as described in [Masuoka 1997a], as the first cohomology group of a double cochain complex, which can be computed by means of a spectral sequence. We compute the first pages of the spectral sequence associated to the double cochain complex $D_{*,*}$ in (3-6), which is a particular case of the double cochain complex of Kac. The five-term exact sequence for this spectral sequence will be useful for computing the group $\text{Opext}_{\triangleright, \triangleleft}(kF, k^G)$ for different kinds of matched pairs.

Spectral sequence of a double cochain complex. Through this section we deal with a first quadrant double complex, that is, a double cochain complex $M^{p,q}$ such that $M^{p,q} = \{0\}$ when $p, q < 0$. There is a spectral sequence associated to a first quadrant double complex, whose first pages are obtained taking vertical and horizontal cohomology of the double complex.

Let $M^{*,*}$ be a first quadrant double complex with vertical and horizontal differentials given by δ_v, δ_h . Let $\text{Tot}(M^{*,*})$ be the total complex associated to $M^{*,*}$. There is a spectral sequence $(E_r^{*,*}, d_r)$ with differentials $d_r^{p,q} : E_r^{p,q} \rightarrow E_r^{p+r, q-r+1}$, which converges to $H^*(\text{Tot}(M^{*,*}))$, whose first pages are given by

$$E_0^{p,q} = M^{p,q}, \quad E_1^{p,q} = H^q(M^{p,*}, d_0), \quad E_2^{p,q} = H^p(E_1^{*,q}, d_1),$$

see [McCleary 2001] for more details. The differentials for each page are given by

$$d_0^{p,q} = d_v, \quad d_1^{p,q} = d'_h, \quad d_2^{p,q}(\bar{\alpha}) = \overline{d_h(\gamma)}, \tag{4-1}$$

where d'_h is the differential induced by d_h on $H^q(M^{p,*}, d_0)$ and $\gamma \in M^{p+1, q-1}$ is such that $d_h(\alpha) = d_v(\gamma)$. Associated to the spectral sequence $(E_r^{*,*}, d_r)$, there is a five-term exact sequence

$$0 \rightarrow E_2^{1,0} \xrightarrow{i} H^1(\text{Tot}(M^{*,*})) \xrightarrow{p} E_2^{0,1} \xrightarrow{d_2^{0,1}} E_2^{2,0} \xrightarrow{i} H^2(\text{Tot}(M^{*,*})), \quad (4-2)$$

where i is a restriction map, the map p is a projection map.

The five-term exact sequence. Given a matched pair $(F, G \triangleright, \triangleleft)$ we compute a five-term exact sequence to calculate $\text{Opext}_{\triangleright, \triangleleft}(kF, kG)$.

Theorem 4.1. *Let $\Sigma = F \bowtie G$ be a matched pair and A be a Σ -module. The first and second pages of the spectral sequence associated to the double complex $\text{Hom}_\Sigma(D_{*,*}, A)$, where $D^{i,j} := P_{i+1} \otimes Q_{j+1}$, for $i, j \geq 0$, is the double complex defined in Proposition 3.8, are given by*

$$E_1^{i,n} = H^n(\Sigma, G; \text{Hom}(P_i, A)), \quad E_2^{n,m} = H^m(H^n(\Sigma, G; \text{Hom}(P_*, A))),$$

for $m, n > 0$. The first page does not depend on the resolution Q_* and $E_2^{n,m}$ ($m, n > 0$) depends neither on the resolution P_* nor the resolution Q_* .

Proof. The double complex $\text{Hom}_\Sigma(D_{*,*}, A)$ with only the vertical differentials is the zeroth page of the spectral sequence

$$E_0^{i,j} := \text{Hom}_\Sigma(P_i \otimes Q_j, A) \cong \text{Hom}_\Sigma(Q_j, \text{Hom}_\mathbb{Z}(P_i, A)). \quad (4-3)$$

Since $(\mathbb{Z} \xleftarrow{\epsilon_Q} Q_i, \delta'_i)$ is a relatively projective resolution of \mathbb{Z} , the first page of the spectral sequence is

$$E_1^{i,n} = H^n(\Sigma, G, \text{Hom}(P_i, A)) = F_n(P_i), \quad i > 0.$$

where $F_n : \Sigma\text{-Mod} \rightarrow \text{Ab}$ is the functor given by $H_1^n(\Sigma, G; \text{Hom}(-, A))$. Hence it does not depend on the resolution Q_i . The second page is

$$E_2^{n,m} = H^m(F_n(P_*)), \quad m, n > 0.$$

To see that $E_2^{n,m} = H^m(F_n(P_*))$, $m, n > 0$ do not depend on the resolution, let P'_i be another relatively F -projective resolution of \mathbb{Z} . Then, there exists a homotopy equivalence $f : P_i \rightarrow P'_i$ as F -modules, that is, there exists $g : P'_i \rightarrow P_i$ and $h_i : P_i \rightarrow P_{i-1}$ such that

$$\delta_i h + h \delta_i = f g - \text{id}.$$

Since the functor F_n is additive, we get

$$F_n(\delta_i) F_n(h_i) + F_n(h_i) F_n(\delta_i) = F_n(f) F_n(f^{-1}) - F_n(\text{id})$$

for each n , so the map $F_n(f)$ is a homotopy equivalence between the resolutions $F_n(P_i)$ and $F_n(P'_i)$. This means that the second page, which consist on the cohomology groups of the resolutions $F_n(P_i)$ and $F_n(P'_i)$ with the respective induced differentials, is isomorphic to the first one. \square

Theorem 4.2. *Let $(F, G, \triangleright, \triangleleft)$ be a matched pair where the action \triangleright is trivial and let k be a field. The spectral sequence in Theorem 4.1 associated to the group $\Sigma = F \bowtie G$ has second page given by*

$$\begin{aligned} E_2^{p,q} &= H^{p+1}(F, H^{q+1}(G, k^\times)), & E_2^{p,0} &\cong H^{p+1}(F, \hat{G}), \\ E_2^{0,q} &\cong \text{Der}(F, H^{q+1}(G, k^\times)), & E_2^{0,0} &= \text{Der}(F, \hat{G}), \end{aligned}$$

for $p \geq 1, q \geq 1$. Therefore, we have the five-term exact sequence:

$$0 \rightarrow H^2(F, \hat{G}) \xrightarrow{i} \text{Opext}_{\triangleright, \triangleleft}(kF, k^G) \xrightarrow{\pi} \text{Der}(F, H^2(G, k^\times)) \xrightarrow{d_2} H^3(F, \hat{G}) \rightarrow H_{\mathcal{A}}^4(\Sigma, X, k^\times). \quad (4-4)$$

Proof. According to (4-3), the zeroth page is given by

$$E_0^{i,j} := \text{Hom}_\Sigma(P_i \otimes Q_j, k^\times) \cong \text{Hom}_\Sigma(Q_j, \text{Hom}_\mathbb{Z}(P_i, k^\times)).$$

If we take $(\mathbb{Z} \xleftarrow{\epsilon_p} P_i, \delta_i)$ and $(\mathbb{Z} \xleftarrow{\epsilon_q} Q_j, \delta'_j)$ to be the resolutions in Proposition 3.5, then we have the group isomorphism

$$\text{Hom}_\Sigma(P_i \otimes Q_j, k^\times) \cong \text{Map}_+(G^{q+1} \times F^{p+1}, k^\times),$$

and the vertical and horizontal differentials of the double complex of groups

$$\text{Map}_+(G^{q+1} \times F^{p+1}, k^\times),$$

are respectively given by

$$\begin{aligned} \delta_i f(s_{i+1}, \dots, s_1; x_1, \dots, x_p)^{(-1)^p} \\ = f(s_{i+1}, \dots, s_2; s_1 \triangleright x_1, (s_1 \triangleleft x_1) \triangleright x_2, \dots, (s_1 \triangleleft x_1 \cdots x_{p-1}) \triangleright x_p) \\ \times \prod_{k=1}^i f(s_{i+1}, \dots, s_{i+1} s_i, \dots, s_1; x_1, \dots, x_p)^{(-1)^k} \times f(s_i, \dots, s_1; x_1, \dots, x_p)^{(-1)^{q+1}}. \end{aligned}$$

and

$$\begin{aligned} \delta'_i f(s_q, \dots, s_1; x_1, \dots, x_{i+1}) \\ = f(s_q \triangleleft (s_{q-1} \cdots s_1 \triangleright x_1), \dots, s_2 \triangleleft (s_1 \triangleright x_1), s_1 \triangleleft x_1; x_2, \dots, x_{p+1}) \\ \times \prod_{k=1}^i f(s_q, \dots, s_1; x_1, \dots, x_i x_{i+1}, \dots, x_{i+1})^{(-1)^k} \times f(s_q, \dots, s_1; x_1, \dots, x_i)^{(-1)^{q+1}}. \end{aligned}$$

Since the action \triangleleft is trivial, the vertical differentials are given by

$$\begin{aligned} \delta(f)(s_{q+1}, \dots, s_1; x_1, \dots, x_p)^{(-1)^p} \\ = f(s_{q+1}, \dots, s_2; x_1, \dots, x_p) \\ \prod_{i=1}^q f(s_{q+1}, \dots, s_{i+1} s_i, \dots, s_1; x_1, \dots, x_p)^{(-1)^i} f(s_q, \dots, s_1; x_1, \dots, x_p)^{(-1)^{q+1}}. \end{aligned}$$

We have that $\text{Map}_+(G^q \times F^p, k^\times) \cong C^q(G, C^p(F, k^\times))$, where $C^q(G, C^p(F, k^\times))$ denotes the group of functions $f : G^q \rightarrow C^p(F, k^\times)$ (with a normalization property) and the group G acts trivially on the group $C^p(F, k^\times)$.

Taking vertical cohomology to the zeroth page, we get $H^q(G, C^p(F, k^\times))$. Therefore, the first page $E_1^{*,*}$ of the spectral sequence is given by

$$\begin{aligned} E_1^{p,q} &= H^{q+1}(G, C^{p+1}(F, k^\times)) \cong C^{p+1}(F, H^{q+1}(G, k^\times)) \quad \text{for } q \geq 1 \\ E_1^{p,0} &= \text{Der}(G, C^{p+1}(F, k^\times)) \cong C^{p+1}(F, \text{Der}(G, k^\times)), \end{aligned}$$

where the isomorphisms hold since the vertical differential leaves every element of F fixed. On the other hand, the horizontal differentials are given by

$$\begin{aligned} \delta'(f)(s_q, \dots, s_1; x_1, \dots, x_{p+1}) \\ &= f(s_q \triangleleft x_1, \dots, s_1 \triangleleft x_1; x_2, \dots, x_{p+1}) \\ &\quad \prod_{i=1}^p f(s_q, \dots, s_1; x_1, \dots, x_i x_{i+1}, \dots, x_{p+1})^{(-1)^i} f(s_q, \dots, s_1; x_1, \dots, x_p)^{(-1)^{p+1}}, \end{aligned}$$

by differentiating each row by the induced horizontal differentials,

$$\begin{aligned} E_2^{p,q} &= H^{p+1}(F, H^{q+1}(G, k^\times)), \quad E_2^{p,0} \cong H^{p+1}(F, \hat{G}) \\ E_2^{0,q} &\cong \text{Der}(F, H^{q+1}(G, k^\times)), \quad E_2^{0,0} = \text{Der}(F, \hat{G}) \end{aligned} \tag{4-5}$$

Since k^\times is a trivial G -module, the sequence (4-2) turns into

$$\begin{aligned} 0 \rightarrow H^2(F, \text{Der}(G, k^\times)) \xrightarrow{i} H^1(\text{Tot}(\text{Hom}_\Sigma(P_i \otimes Q_j, k^\times))) \xrightarrow{p} \text{Der}(F, H^2(G, k^\times)) \\ \xrightarrow{d_2^{0,1}} H^3(F, \text{Der}(G, k^\times)) \xrightarrow{i} H^2(\text{Tot}(\text{Hom}_\Sigma(P_i \otimes Q_j, k^\times))), \end{aligned}$$

From (3-11) and (3-12) we get the five-term exact sequence

$$0 \rightarrow H^2(F, \hat{G}) \rightarrow \text{Opext}_{\triangleright, \triangleleft}(kF, k^G) \rightarrow \text{Der}(F, H^2(G, k^\times)) \xrightarrow{d_2^{0,1}} H^3(F, \hat{G}) \rightarrow H^4(\Sigma, X, k^\times). \quad \square$$

Note that, in the case that \triangleright is a trivial action, the terms $E_2^{p,q}$ with $p \geq 1, q \geq 1$ of the second page of the spectral sequence associated to the semidirect product $\Sigma = F \ltimes G$ coincide with the second page of the Lyndon–Hochschild–Serre spectral sequence [Evens 1991].

Corollary 4.3. *Let $(F, G, \triangleright, \triangleleft)$ be a matched pair with trivial \triangleright action and let k be a field. Then:*

- (1) *If $H^2(G, k^\times) = 1$ then $\text{Opext}_{\triangleright, \triangleleft}(kF, k^G) \cong H^2(F, \hat{G})$.*
- (2) *If $(|F|, |\hat{G}|) = 1$, then $\text{Opext}_{\triangleright, \triangleleft}(kF, k^G) \cong \text{Der}(F, H^2(G, k^\times))$.*
- (3) *If G is a perfect group, then $\text{Opext}_{\triangleright, \triangleleft}(kF, k^G) = \text{Der}(F, H^2(G, k^\times))$*
- (4) *If $|F| = 2k + 1$ and $G = S_n$ with $n \geq 4$, then, $\text{Opext}_{\triangleright, \triangleleft}(\mathbb{C}F, \mathbb{C}^G) = 0$.*

Proof. Part (1) is straightforward.

(2) Since $(|F|, |\hat{G}|) = 1$, then $H^n(F, \hat{G}) = H^n(F, \hat{G}) = \{1\}$ and the result holds.

(3) Since the abelianization of G is trivial, then $\hat{G} \cong \{0\}$. Therefore, $H^2(F, \hat{G}) = H^2(F, \hat{G}) = \{0\}$, similarly $H^3(F, \hat{G}) = 0$, then, $\text{Opext}_{\triangleright, \triangleleft}(kF, k^G) = \text{Der}(F, H^2(G, k^\times))$.

(4) Under the given conditions $(|F|, |\hat{G}|) = 1$, as in (2). So $\text{Opext}_{\triangleright, \triangleleft}(\mathbb{C}F, \mathbb{C}^G) = \text{Der}(F, H^2(G, \mathbb{C}^\times))$. Now, $H^2(S_n, \mathbb{C}^\times) = \mathbb{Z}/2$, for $n \geq 4$ and $\text{Der}(F, H^2(G, \mathbb{C}^\times)) = \text{Hom}(F, \mathbb{Z}/2) = 0$, so the result holds. \square

5. Computations

We compute some examples of the group $\text{Opext}_{\triangleright, \triangleleft}(kF, k^G)$ for different semidirect products: the right action \triangleright is trivial, so we denote the group $\text{Opext}_{\triangleright, \triangleleft}(kF, k^G)$ by $\text{Opext}_{\triangleleft}(kF, k^G)$. The first calculation generalizes one from Masuoka [1997b].

Theorem 5.1. *Let k be a field. Let G be a group and $\mathbb{Z}/2 \rtimes (G \times G)$ be the semidirect product with $(a, b) \triangleleft 1 = (b, a)$. Then*

$$\text{Opext}_{\triangleleft}(k\mathbb{Z}/2, k^{G \times G}) \cong H^2(G, k^\times) \oplus P_{\text{Sym}}(G, G; k^\times),$$

where $P_{\text{Sym}}(G, G; k^\times)$ is the groups of all symmetric bicharacters of G .

Proof. It follows from (2-2) that $H^n(\mathbb{Z}/2, H^1(G \times G, k^\times)) = 0$ for $n \geq 1$. Then, the sequence (4-4) implies that

$$\text{Opext}_{\triangleleft}(k\mathbb{Z}/2, k^{G \times G}) \cong \text{Der}(\mathbb{Z}/2, H^2(G \times G, k^\times)).$$

According to (2-3), we have $H^2(G \times G, k^\times) \cong H^2(G, k^\times) \oplus P(G, G, k^\times) \oplus H^2(G, k^\times)$. Given $\alpha \in Z^2(G \times G, k^\times)$, we have $(\bar{\alpha}_1, \bar{\alpha}_2, \phi_{\bar{\alpha}}) = \psi(\bar{\alpha}) \in H^2(G, k^\times) \oplus H^2(G, k^\times) \oplus P(G, G, k^\times)$ given by

$$\alpha_1(x, y) = \alpha((x, e), (y, e)), \quad \alpha_2(x, y) = \alpha((e, x), (e, y)), \quad \phi_{\bar{\alpha}}(x, y) = \frac{\alpha((x, e), (e, y))}{\alpha((e, y), (x, e))}.$$

Hence the induced action of $\bar{1} \in \mathbb{Z}/2$ on $H^2(G, k^\times) \times H^2(G, k^\times) \times P(G, G; k^\times)$ is

$$\begin{aligned} \bar{1}(\alpha_1, \alpha_2, \phi_{\bar{\alpha}})(x, y) &= (\bar{1}\alpha_1(x, y), \bar{1}\alpha_2(x, y), \bar{1}\phi_{\bar{\alpha}}(x, y)) \\ &= (\alpha_2(x, y), \alpha_1(x, y), \phi_{\bar{\alpha}}^{-1}(y, x)) \\ &:= (\alpha_2, \alpha_1, (\phi_{\bar{\alpha}}^T)^{-1})(x, y). \end{aligned}$$

Then $\alpha \in \text{Der}(\mathbb{Z}/2, H^2(G \times G, k^\times))$ if and only if $\alpha := \alpha(\bar{1})$ satisfies $\alpha \bar{1} \alpha = 1$, that is,

$$(\alpha_1, \alpha_2, \phi_{\bar{\alpha}})(\alpha_2, \alpha_1, (\phi_{\bar{\alpha}}^T)^{-1}) = 1 \Leftrightarrow \alpha_1 = \alpha_2^{-1},$$

and $\phi_{\bar{\alpha}}$ is a symmetric bicharacter. \square

The following example includes the previous one in the case that $b = c = 1, a = 0$.

Theorem 5.2. Let $\mathbb{Z}/2 \ltimes (\mathbb{Z}/n \oplus \mathbb{Z}/n)$ be the semidirect product where the action of $\mathbb{Z}/2$ on $(\mathbb{Z}/n \oplus \mathbb{Z}/n)$ is defined by the matrix

$$A = \begin{pmatrix} a & b \\ c & -a \end{pmatrix}$$

with $\text{Det}(A) = -1$. Let k be a field such that $k^\times / (k^\times)^{2n} = 0$. Then,

$$\text{Opext}_{\triangleright}(k\mathbb{Z}/2, k^{\mathbb{Z}/n \oplus \mathbb{Z}/n}) \cong \frac{\text{Ker}(A - I)}{\text{Im}(A + I)} \oplus \mu_n(k),$$

where $\mu_n(k)$ is the group of n -th roots of unity in k .

Proof. In this case the sequence (4-4) is given by

$$\begin{aligned} 0 \rightarrow H^2(\mathbb{Z}/2, \text{Der}(\mathbb{Z}/n \oplus \mathbb{Z}/n, k^\times)) &\xrightarrow{i} \text{Opext}_{\triangleleft}(k\mathbb{Z}/2, k^{\mathbb{Z}/n \oplus \mathbb{Z}/n}) \\ &\xrightarrow{\pi} \text{Der}(\mathbb{Z}/2, H^2(\mathbb{Z}/n \oplus \mathbb{Z}/n, k^\times)) \xrightarrow{d_2^{0,1}} H^3(\mathbb{Z}/2, \text{Der}(\mathbb{Z}/n \oplus \mathbb{Z}/n, k^\times)) \rightarrow H^4(\Sigma, X, A). \end{aligned}$$

We will see that:

- (i) $H^2(\mathbb{Z}/2, H^1(\mathbb{Z}/n \oplus \mathbb{Z}/n, k^\times)) \cong \text{Ker}(A - I)/\text{Im}(A + I)$.
- (ii) $\text{Der}(\mathbb{Z}/2, H^2(\mathbb{Z}/n \oplus \mathbb{Z}/n, k^\times)) \cong \mu_n(k)$.
- (iii) $d_2^{0,1} = 0$.

Therefore, $\text{Opext}_{\triangleleft}(k\mathbb{Z}/2, k^{\mathbb{Z}/n \oplus \mathbb{Z}/n})$ fits in a short exact sequence

$$0 \rightarrow \frac{\text{Ker}(A - I)}{\text{Im}(A + I)} \xrightarrow{i} \text{Opext}_{\triangleleft}(k\mathbb{Z}/2, k^{\mathbb{Z}/n \oplus \mathbb{Z}/n}) \xrightarrow{\pi} \mu_n(k) \rightarrow 0, \quad (5-1)$$

moreover, we will see (5-1) is split, so

$$\text{Opext}_{\triangleleft}(k\mathbb{Z}/2, k^{\mathbb{Z}/n \oplus \mathbb{Z}/n}) \cong \frac{\text{Ker}(A - I)}{\text{Im}(A + I)} \oplus \mu_n(k).$$

- (i) It follows immediately from (2-2).
- (ii) We identify $\wedge^2(\mathbb{Z}/n \oplus \mathbb{Z}/n)$ with the abelian group of alternating 2×2 matrices over the ring $\mathbb{Z}/n\mathbb{Z}$. Therefore, $\wedge^2(\mathbb{Z}/n \oplus \mathbb{Z}/n) \cong \mathbb{Z}/n$ and

$$H^2(\mathbb{Z}/n \oplus \mathbb{Z}/n, k^\times) \cong \text{Hom}(\wedge^2(\mathbb{Z}/n \oplus \mathbb{Z}/n), k^\times) \cong \mu_n(k),$$

where $\mu_n(k)$ is the group of n -th roots unit. Since $A^T M A = -M$ for all $M \in \wedge^2(\mathbb{Z}/n \oplus \mathbb{Z}/n)$ we have that

$$\text{Der}(\mathbb{Z}/2, H^2(\mathbb{Z}/n \oplus \mathbb{Z}/n, k^\times)) \cong \mu_n(k).$$

- (iii) To compute

$$d_2^{0,1} : \mu_n(k) = \text{Der}(\mathbb{Z}/2, H^2(\mathbb{Z}/n \oplus \mathbb{Z}/n, k^\times)) \rightarrow H^3(\mathbb{Z}/2, \text{Der}(\mathbb{Z}/n \oplus \mathbb{Z}/n, k^\times)),$$

we follow (4-1). Given $\zeta \in \mu_n(k) = \text{Der}(\mathbb{Z}/2, H^2(\mathbb{Z}/n \oplus \mathbb{Z}/n, k^\times))$, we need to find

$$\gamma \in \text{Map}_+((\mathbb{Z}/n \oplus \mathbb{Z}/n) \times (\mathbb{Z}/2)^2, k^\times) \cong C^2(\mathbb{Z}/2, C^1(\mathbb{Z}/n \oplus \mathbb{Z}/n, k^\times))$$

such that

$$\delta_h(\alpha_\zeta) = \delta_v(\gamma), \tag{5-2}$$

where

$$\alpha_\zeta \in C^2(\mathbb{Z}/n \oplus \mathbb{Z}/n, k^\times), \quad \alpha_\zeta(x, y) = \zeta^{x_1 y_2} \tag{5-3}$$

and then compute the cohomology class of $\delta_h(\gamma)$ in $H^3(\mathbb{Z}/2, \text{Der}(\mathbb{Z}/n \oplus \mathbb{Z}/n, k^\times))$. We have that

$$\delta_h(\alpha_\zeta)(1, 1)(x, y) = \alpha_\zeta(Ax, Ay)\alpha_\zeta(x, y) = \zeta^{x^T \begin{pmatrix} ac & bc \\ bc & -ab \end{pmatrix} y}.$$

This is a bicharacter with associated quadratic form

$$\omega(x, y) = \zeta^{-acx^2 - 2bcxy + aby^2}.$$

Therefore, the cochain $\gamma \in C^2(\mathbb{Z}/2, C^1(\mathbb{Z}/n \oplus \mathbb{Z}/n, k^\times))$ defined by

$$\gamma(1, 1) = \zeta^{-(acx^2)/2 - bcxy + (aby^2)/2}, \quad (x, y) \in \mathbb{Z}/n \oplus \mathbb{Z}/n, \tag{5-4}$$

and $\gamma(0, 1) = \gamma(1, 0) = \gamma(0, 0) = 1$, satisfies (5-2).

Finally, the horizontal differential of γ is given by

$$\delta_h(\gamma)(1, 1, 1) = \gamma(1, 1)\gamma(1, 0)(\gamma(1, 1)\gamma(0, 1))^{-1} = \gamma(1, 1)(\gamma(1, 1))^{-1} = 1.$$

Hence, $d_2(\zeta) = 1$.

A section of π in the exact sequence (5-1) is given by cohomology of $s(\zeta) = (\alpha_\zeta, \gamma_\zeta)$, where α_ζ and γ_ζ are given by (5-3) and (5-4), respectively. It is clear from the definition that s is a group homomorphism, that is, (5-1) splits. \square

Theorem 5.3. *Let F be an arbitrary group acting on a finite abelian group V with odd order. Suppose that $(k^\times)^n = k^\times$, where n is the exponent of V . Then*

$$\text{Opext}_{\triangleleft}(kF, k^V) \cong H^2(F, \hat{V}) \oplus \text{Der}(F, \wedge^2 \hat{V}),$$

where $\hat{V} = \text{Hom}(V, k^\times)$.

Proof. By (2-4), we have $H^2(V, k^\times) \cong \wedge^2 \hat{V}$. The sequence (4-4) is given by

$$0 \rightarrow H^2(F, \hat{V}) \xrightarrow{i} \text{Opext}_{\triangleleft}(kF, k^V) \xrightarrow{\pi} \text{Der}(F, \wedge^2 \hat{V}) \xrightarrow{d_2} H^3(F, \hat{V}) \rightarrow H^4(\Sigma, X, A),$$

where $\Sigma = V \rtimes F$. We will see that $d_2 = 0$ and the resulting short exact sequence splits, hence we get the result.

Let $\alpha \in \text{Der}(F, \wedge^2 \hat{V})$, that is, $\alpha : F \rightarrow \wedge^2 \hat{V}$ such that $\alpha(gh) = {}^g\alpha(h)\alpha(g)$. By (2-5), α can be identified with a map $\alpha : F \rightarrow H^2(V, k^\times)$ which can be lifted to a map $\tilde{\alpha} : F \rightarrow Z^2(V, k^\times)$ considering that, since V has odd order, the map

$$\text{Alt} : \wedge^2 \hat{V} \rightarrow \wedge^2 \hat{V}$$

given by $\text{Alt}(\phi)(x, y) = \phi(x, y)/\phi(y, x) = \phi(x, y)^2$ is an isomorphism, so we can define the lifting map by $\tilde{\alpha} : F \rightarrow Z^2(V, k^\times)$ by

$$\tilde{\alpha}(g) = \alpha(g)^{1/2}.$$

In order to compute $d_2(\alpha)$, we must find a function $\gamma \in C^2(F, C^1(V, k^\times))$ such that

$$\delta(\gamma(g, h)) = \frac{{}^g\tilde{\alpha}(h)\tilde{\alpha}(g)}{\tilde{\alpha}(gh)} = \frac{{}^g\tilde{\alpha}(h)\tilde{\alpha}(g)}{\tilde{\alpha}(gh)} = \left(\frac{{}^g\alpha(h)\alpha(g)}{\alpha(gh)} \right)^{1/2} = 1.$$

Hence γ can be taken to be the constant cochain and, therefore, $d_2(\alpha) = 1$ for all $\alpha \in \text{Der}(F, \wedge^2 \hat{V})$. \square

Corollary 5.4. *Let $F = C_m = \langle \sigma \rangle$ be a cyclic group of order m acting on a finite abelian group V with odd order. Suppose that $(k^\times)^n = k^\times$, where n is the exponent of V . Then*

$$\text{Opext}_{\triangleleft}(kF, k^V) \cong \{\psi \in \hat{V} : \sigma\psi = \psi\} / \{N_\sigma\psi : \psi \in \hat{V}\} \oplus \{b \in \wedge^2 \hat{V} : N_\sigma b = 0\},$$

where $N_\sigma = 1 + \sigma + \dots + \sigma^{m-1}$.

An example with nontrivial differential d_2 . The next example illustrates the fact that the hypothesis in Theorem 5.3 stating that the order of the order of V must be odd, can not be avoided since otherwise the differential d_2 can be not trivial.

Remarks 5.5. (a) Let G be an elementary abelian p -group of rank n . Once a basis of G is fixed, using the isomorphism (2-5) we can identify $H^2(G, \mathbb{C}^\times)$ with alternating matrices over \mathbb{Z}/p . A representative 2-cocycle $\alpha_M \in H^2(G, \mathbb{C}^\times)$ corresponding to a matrix M is defined by

$$\alpha_M(\mathbf{x}, \mathbf{y}) = \exp\left(\frac{2\pi i}{n} \mathbf{x}^T \tilde{M} \mathbf{y}\right), \tag{5-5}$$

where \tilde{M} is the upper triangular part of M .

(b) Let $F = \langle t_1 \rangle \oplus \langle t_2 \rangle$ be a product of cyclic groups and let M be an left F -module. If $(\mathbb{Z} \xleftarrow{\epsilon} P_i)$ and $(\mathbb{Z} \xleftarrow{\epsilon} P'_i)$ are periodic resolutions as in (2-1) for the groups $\langle t_1 \rangle$ and $\langle t_2 \rangle$ respectively, then the total complex $\text{Tot}(P \otimes P')$ is a free F -resolution of \mathbb{Z} . Therefore, given a F -module M , we can compute

$H^*(F, M)$ as the cohomology of total complex of

$$\begin{array}{ccccccc}
 \vdots & & \vdots & & \vdots & & \\
 \uparrow & & \uparrow & & \uparrow & & \\
 t_{2-1} & & t_{2-1} & & t_{2-1} & & \\
 \uparrow & & \uparrow & & \uparrow & & \\
 M & \xrightarrow{t_{1-1}} & M & \xrightarrow{N_{t_1}} & M & \xrightarrow{t_{1-1}} & \dots \\
 \uparrow & & \uparrow & & \uparrow & & \\
 N_{t_2} & & N_{t_2} & & N_{t_2} & & \\
 \uparrow & & \uparrow & & \uparrow & & \\
 M & \xrightarrow{t_{1-1}} & M & \xrightarrow{N_{t_1}} & M & \xrightarrow{t_{1-1}} & \dots \\
 \uparrow & & \uparrow & & \uparrow & & \\
 t_{2-1} & & t_{2-1} & & t_{2-1} & & \\
 \uparrow & & \uparrow & & \uparrow & & \\
 M & \xrightarrow{t_{1-1}} & M & \xrightarrow{N_{t_1}} & M & \xrightarrow{t_{1-1}} & \dots
 \end{array}$$

Since we are mainly interested in $H^2(F, M)$, the second and third differentials $\delta_2 : M \oplus M \rightarrow M \oplus M \oplus M$ and $\delta_3 : M \oplus M \oplus M \rightarrow M \oplus M \oplus M \oplus M$ of the total complex are given by

$$\delta_2(A, B) = (A + {}^{s_1}A, A - {}^{s_2}A - (B - {}^{s_1}B), B + {}^{s_2}B), \tag{5-6}$$

$$\delta_3(A, B, C) = ({}^{s_1}A - A, {}^{s_2}A - A + B + {}^{s_1}B, B + {}^{s_2}B + {}^{s_1}C - C, {}^{s_2}C - C). \tag{5-7}$$

Lemma 5.6. *Let $F = \langle t_1, t_2 \rangle$, $G = \langle s_1, \dots, s_4 \rangle$ be elementary abelian 2-groups of rank 2 and 4, respectively. Consider the (right) action of F on G determined by the matrices*

$$F_1 = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 \\ 0 & 1 & 0 & 1 \end{pmatrix}, \quad F_2 = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 1 \end{pmatrix}.$$

and the induced left action of F on $H^2(G, \mathbb{C}^\times)$. Then:

- (1) The group $\text{Der}(F, H^2(G, \mathbb{C}^\times))$ is in correspondence with the set of pairs of matrices

$$A = \begin{pmatrix} 0 & 0 & b & c \\ 0 & 0 & d & e \\ b & d & 0 & f \\ c & e & f & 0 \end{pmatrix}, \quad B = \begin{pmatrix} 0 & 0 & b' & c' \\ 0 & 0 & d' & e' \\ b' & d' & 0 & f' \\ c' & e' & f' & 0 \end{pmatrix}$$

with entries in \mathbb{Z}/p , such that

$$\begin{aligned}
 c' + b' + d' &= 0 \\
 c + d + e &= 0 \\
 b + e + c' + d' + e &= 0.
 \end{aligned} \tag{5-8}$$

- (2) The group $H^2(F, H^1(G, \mathbb{C}^\times))$ is isomorphic to $\mathbb{Z}/2^2$.

Proof. (1) By (5-6), elements in $\text{Der}(F, H^2(G, \mathbb{C}^\times))$ are in correspondence with pairs (A, B) of alternating 4×4 matrices such that

$$\begin{aligned} A + F_1 A F_1^T &= 0 \\ B + F_2 B F_2^T &= 0 \\ A - F_2 A F_2^T - B + F_1 B F_1^T &= 0. \end{aligned} \tag{5-9}$$

The system (5-9) is equivalent to (5-8).

(2) In order to compute $H^2(F, H^1(G, \mathbb{C}^\times))$ we use the canonical identification

$$H^1(G, \mathbb{C}^\times) = \text{Hom}(G, \mathbb{C}^\times) \cong G$$

as left F -modules. By (5-7), we have that $\text{Ker}(\delta_3)$ is in correspondence with 4×3 matrices $S = [n_a, n_b, n_c]$ over $\mathbb{Z}/2$ such that

$$\begin{aligned} (F_1 - I)n_a &= 0, & (F_2 - I)n_c &= 0, \\ (I - F_1)n_c &= (F_2 + I)n_b, & (I + F_1)n_b &= (I - F_2)n_a. \end{aligned}$$

Thus, the space $\text{Ker}(\delta_3)$ corresponds with all 4×3 matrices over \mathbb{F}_2 such that $S_{ij} = 0$ for $1 \leq i \leq 2$ and $1 \leq j \leq 3$. On the other hand, by (5-6) we have

$$\text{Im}(\delta_2) = \{(l_a + F_1 l_a, l_a - l_b + F_1 l_b - F_2 l_a, l_b + F_2 l_b) : l_a, l_b \in N\},$$

that is, $\text{Im}(\delta_2)$ is in correspondence with all matrices of the form

$$\begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ x_1 + x_2 & x_1 + y_1 + y_2 & y_1 \\ x_2 & x_1 + x_2 + y_2 & y_1 + y_2 \end{pmatrix}$$

where $x_i, y_i \in \mathbb{Z}/2$. Hence $H^2(F, H^1(G, \mathbb{C}^\times)) \cong \mathbb{Z}/2^2$. □

Lemma 5.7. *Let $\Sigma = F \ltimes G$ be a semidirect product and let*

$$\cdots \rightarrow R_3 \rightarrow R_2 \rightarrow R_1 \rightarrow R_0, \tag{5-10}$$

be a free resolution of a right F -modules M . The action of F on R_i can be extended to an action of Σ by $r \cdot (f, g) = r \cdot f$. With this action, the sequence (5-10) turns out to be a relatively G -projective resolution of the right Σ -module M . □

Theorem 5.8. *Let $F = \langle t_1, t_2 \rangle$ and $G = \langle s_1, \dots, s_4 \rangle$ be elementary abelian 2-groups of rank 2 and 4, respectively. Consider the (right) action of F on G determined by the matrices*

$$F_1 = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 \\ 0 & 1 & 0 & 1 \end{pmatrix}, \quad F_2 = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 1 \end{pmatrix}.$$

Then $\text{Opext}_{\triangleleft}(\mathbb{C}F, \mathbb{C}G) \cong (\mathbb{Z}/2)^3 \oplus (\mathbb{Z}/4)^2$.

Proof. The five-term exact sequence (4-4) for this case is

$$0 \rightarrow H^2(F, \text{Der}(G, \mathbb{C}^\times)) \xrightarrow{i} \text{Opext}_{\triangleleft}(\mathbb{C}F, \mathbb{C}^G) \xrightarrow{\pi} \text{Der}(F, H^2(G, \mathbb{C}^\times)) \xrightarrow{d_2} H^3(F, \text{Der}(G, \mathbb{C}^\times)) \rightarrow H^2(\text{Tot}(M^{*,*})).$$

For this computation the relatively F -projective resolution used in Theorem 4.1 ($\mathbb{Z} \xleftarrow{\epsilon_Q} Q_i, \delta'_i$) will be as in Proposition 3.5, and the relatively G -projective resolution ($\mathbb{Z} \xleftarrow{\epsilon_P} P_i, \delta_i$) will be the total complex of the tensor product of the cyclic resolutions for $\langle t_1 \rangle$ and $\langle t_2 \rangle$, considered as a relatively G -projective Σ -resolution of \mathbb{Z} using Lemma 5.7.

The zeroth page of the spectral sequence in Theorem 4.1 is given by

$$E_0^{i,j} = \text{Hom}_{\mathbb{Z}\Sigma}(P_{i+1} \otimes Q_{j+1}, \mathbb{C}^\times) = \text{Hom}_{\mathbb{Z}\Sigma}\left(\bigoplus_{k=1}^{i+2} \mathbb{Z}F \otimes \mathbb{Z}G[G]^{j+1}, \mathbb{C}^\times\right)$$

The horizontal and vertical differentials d_h and d_v are induced by the differentials of the resolutions ($\mathbb{Z} \xleftarrow{\epsilon_P} P_i$) and ($\mathbb{Z} \xleftarrow{\epsilon_Q} Q_i$), respectively. Each Σ -module $\mathbb{Z}F \otimes \mathbb{Z}G[G]^{j+1}$ is free with basis

$$\{e \otimes [g_1 | \cdots | g_{j+1}] : e \neq g_1, \cdots, g_{j+1} \in N\}.$$

Therefore, an element in $E_0^{i,j}$ is defined by a tuple (h_1, \dots, h_{i+2}) with $h_k \in C^{j+1}(F, \mathbb{C}^\times)$, where $h_k(f_1, \dots, f_{j+1}) = 1$ if any of the entries is the identity of F . Similarly, the differentials $d_0^{k,0} : \text{Hom}_{\Sigma}(P^{k+1} \otimes Q^1, \mathbb{C}^\times)$ of the zeroth page, induced by the vertical differentials of the double complex are given by

$$d_0(f)(e \otimes [g_1 | g_2]) = f(e \otimes (g_1[g_2] - [g_1g_2] + [g_1])).$$

Since we are considering \mathbb{C}^\times as a trivial Σ -module, the elements in $E_1^{k,0} = \text{Ker}(d_0^{k,0})$ are in correspondence to tuples $(\chi_1, \dots, \chi_{i+2})$ with $\chi_i \in \hat{G}$.

First, we will compute $\text{Ker}(d_2)$. By Lemma 5.6, the group $E_2^{0,1} = \text{Der}(F, H^2(G, \mathbb{C}^\times))$ is in correspondence with pairs (A, B) of alternating 4×4 matrices satisfying the equations in (5-8). According to Remarks 5.5(a), a representative element for (A, B) in $E_0^{0,1} = \text{Hom}_{\Sigma}((\mathbb{Z}F \oplus \mathbb{Z}F) \otimes \mathbb{Z}G[G]^2, \mathbb{C}^\times)$ is defined by $\alpha = (\alpha_A, \alpha_B)$, where α_A, α_B are 2-cocycles defined in (5-5).

By (4-1), we have that $d_2(A, B) = \overline{d_h(\gamma)}$ were

$$\gamma \in E_0^{1,0} = \text{Hom}_{\Sigma}((\mathbb{Z}F \oplus \mathbb{Z}F \oplus \mathbb{Z}F) \otimes \mathbb{Z}G[G]),$$

satisfies $d_h(\alpha_A, \alpha_B) = d_v(\gamma)$.

By (5-6) we have that

$$d_h(\alpha_A, \alpha_B) = (b_{M_1}, b_{M_2}, b_{M_3}) \in E_0^{1,1} = \text{Hom}_{\Sigma}((\mathbb{Z}F \oplus \mathbb{Z}F \oplus \mathbb{Z}F) \otimes \mathbb{Z}G[G]^2, \mathbb{C}^\times), \tag{5-11}$$

where $b_{M_i}(x, y) = (-1)^{x^T M_i y}$ and

$$M_1 = \tilde{A} + F_1 \tilde{A} F_1^T = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & b+d & d \\ 0 & 0 & d & e \end{pmatrix}$$

$$M_2 = \tilde{A} - F_2 \tilde{A} F_2^T - (\tilde{B} - F_1 \tilde{B} F_1^T) = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & b+b'+d' & b+d+d' \\ 0 & b+d+d' & c+e+e' & \end{pmatrix}$$

$$M_3 = \tilde{B} + F_2 \tilde{B} F_2^T = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & b' & c' \\ 0 & 0 & c' & c'+e' \end{pmatrix}.$$

The cochain $\gamma = (\gamma_{M_1}, \gamma_{M_2}, \gamma_{M_3}) \in E_0^{1,0} = \text{Hom}_\Sigma((\mathbb{Z}F \oplus \mathbb{Z}F \oplus \mathbb{Z}F) \otimes \mathbb{Z}G[G]^2, \mathbb{C}^\times)$ defined by

$$\begin{aligned} \gamma_{M_1}(\mathbf{x}) &= \exp\left(-\frac{\pi}{2}((b+d)x_3^2 + 2dx_3x_4 + ex_4^2)\right), \\ \gamma_{M_2}(\mathbf{x}) &= \exp\left(-\frac{\pi}{2}((b+b'+d')x_3^2 + 2(b+d+d')x_3x_4 + (c+e+e')x_4^2)\right), \\ \gamma_{M_3}(\mathbf{x}) &= \exp\left(-\frac{\pi}{2}(b'x_3^2 + 2c'x_3x_4 + (c'+e')x_4^2)\right), \end{aligned} \tag{5-12}$$

satisfies (5-11). Therefore,

$$\begin{aligned} \delta_h(\gamma_{M_1}, \gamma_{M_2}, \gamma_{M_3}) &= \left(\frac{{}^{t_1}\gamma_{M_1}}{\gamma_{M_1}}, \frac{{}^{t_2}\gamma_{M_1}\gamma_{M_2}({}^{t_1}\gamma_{M_2})}{\gamma_{M_1}}, \frac{\gamma_{M_2}({}^{t_2}\gamma_{M_2}){}^{t_1}\gamma_{M_3}}{\gamma_{M_3}}, \frac{{}^{t_2}\gamma_{M_3}}{\gamma_{M_3}} \right) \\ &= (1, \gamma_{M_2}^2, \gamma_{M_2}^2, 1) \in \ker(d_1 : E_1^{2,0} \rightarrow E_1^{3,0}). \end{aligned}$$

Since G is an elementary abelian 2-group, we will use the canonical identification of \hat{G} with G . Under this identification we have that $\gamma_{M_2}^2 = (0, 0, b+b'+d', c+e+e')$.

The pair (A, B) belongs to $\text{Ker}(d_2)$ if and only if $(0, \gamma_{M_2}^2, \gamma_{M_2}^2, 0)$ belongs to the image of $d_1 : E_1^{1,0} \rightarrow E_1^{2,0}$ if and only if there exists $(\mu_A, \mu_B, \mu_C) \in E_1^{1,0} = F^{\times 3}$ such that $d_h(\mu_A, \mu_B, \mu_C) = (0, \gamma_{M_2}^2, \gamma_{M_2}^2, 0)$. This means that

$$(F_1 - I)\mu_A = 0, \quad (F_2 - I)\mu_C = 0, \quad (F_2 - I)\mu_A + (F_1 + I)\mu_B = \gamma_{M_2}^2, \quad (F_1 - I)\mu_C + (F_2 + I)\mu_B = \gamma_{M_2}^2.$$

From this equation we obtain $b+b'+d' = c+e+e' = 0$. Joining these two equations with (5-8) we get a system of equation with 5 free variables, hence $\text{Ker}(d_2) = (\mathbb{Z}/2)^5$.

Hence we have the exact sequence

$$0 \rightarrow H^2(F, G) \rightarrow \text{Opext}_{\triangleleft}(\mathbb{C}F, \mathbb{C}G) \xrightarrow{\pi} \text{Ker}(d_2) \rightarrow 0. \tag{5-13}$$

An element in $\text{Ker}(d_2)$ is represented by a pair of matrices A, B as in Lemma 5.6. Let us assign $c' = 1$ and consider the remaining variables zero, and let us call the respective pair of matrices (A'_c, B'_c) .

A section of the sequence (5-13) send (A_c, B_c) to the class of the extension

$$(\alpha_c, \gamma_c) \in H^1(\text{Hom}_\Sigma(\text{Tot}(D_{*,*}, \mathbb{C}^*)) \cong \text{Opext}_{\triangleright, \triangleleft}(\mathbb{C}F, \mathbb{C}^G))$$

where α_c is the 2-cocycle associated to (A'_c, B'_c) and γ_c is given according to (5-12), by

$$\gamma_{M_1}(\mathbf{x}) = \exp\left(-\frac{\pi}{2}(x_3^2 + x_4^2)\right), \quad \gamma_{M_2}(\mathbf{x}) = \exp\left(-\frac{\pi}{2}(x_3^2)\right), \quad \gamma_{M_3}(\mathbf{x}) = \exp\left(-\frac{\pi}{2}(x_4^2)\right).$$

It can be verified that the class of (α_c, γ_c) has order 4 in $\text{Opext}_{\triangleright, \triangleleft}(\mathbb{C}F, \mathbb{C}^G)$. In the same way, if we take the variable d' to be 1 and consider the remaining variables null we get an element (α_d, γ_d) of order 4. Any other element outside the subgroup $\langle (\alpha_c, \gamma_c), (\alpha_d, \gamma_d) \rangle \cong (\mathbb{Z}/4)^2$ has order 2, otherwise the order of $\text{Opext}_{\triangleright, \triangleleft}(\mathbb{C}F, \mathbb{C}^G)$ could not be 2^7 . That is why $\text{Opext}_{\triangleleft}(\mathbb{C}F, \mathbb{C}^G) \cong (\mathbb{Z}/2)^3 \oplus (\mathbb{Z}/4)^2$. Since $H_n(P) = H_n(Q) = 0$ for $n > 0$, then $H_n(P \otimes Q) = 0$ for $n > 1$ and $H_1(P \otimes Q) = \text{Tor}_1(H_0(P), H_0(Q))$. \square

Acknowledgements

We thank Paul Bressler and Bernardo Uribe for kindly answering some questions and for very useful conversations.

References

- [Adamson 1954] I. T. Adamson, "Cohomology theory for non-normal subgroups and non-normal fields", *Proc. Glasgow Math. Assoc.* **2** (1954), 66–76. MR Zbl
- [Alperin 1986] J. L. Alperin, *Local representation theory*, Cambridge Studies in Advanced Mathematics **11**, Cambridge University Press, 1986. MR Zbl
- [Andruskiewitsch and Devoto 1995] N. Andruskiewitsch and J. Devoto, "Extensions of Hopf algebras", *Algebra i Analiz* **7**:1 (1995), 22–61. In Russian; translated in *St. Petersburg Math. J.* **7**:1 (1996), 17–52. MR Zbl
- [Auslander and Solberg 1993] M. Auslander and O. . Solberg, "Relative homology and representation theory, I: Relative homology and homologically finite subcategories", *Comm. Algebra* **21**:9 (1993), 2995–3031. MR Zbl
- [Baaj et al. 2005] S. Baaj, G. Skandalis, and S. Vaes, "Measurable Kac cohomology for bicrossed products", *Trans. Amer. Math. Soc.* **357**:4 (2005), 1497–1524. MR Zbl
- [Curtis and Reiner 1990] C. W. Curtis and I. Reiner, *Methods of representation theory, I: With applications to finite groups and orders*, John Wiley & Sons, New York, 1990. MR Zbl
- [Evens 1991] L. Evens, *The cohomology of groups*, The Clarendon Press, Oxford University Press, New York, 1991. MR Zbl
- [Hochschild 1956] G. Hochschild, "Relative homological algebra", *Trans. Amer. Math. Soc.* **82** (1956), 246–269. MR Zbl
- [Hofstetter 1994] I. Hofstetter, "Extensions of Hopf algebras and their cohomological description", *J. Algebra* **164**:1 (1994), 264–298. MR Zbl
- [Kac 1969] G. I. Kac, "Extensions of groups to ring groups", *Math. USSR, Sb.* **5** (1969), 451–474. Zbl
- [Karpilovsky 1985] G. Karpilovsky, *Projective representations of finite groups*, Monographs and Textbooks in Pure and Applied Mathematics **94**, Marcel Dekker, New York, 1985. MR Zbl
- [Kashina 2000] Y. Kashina, "Classification of semisimple Hopf algebras of dimension 16", *J. Algebra* **232**:2 (2000), 617–663. MR Zbl
- [Majid 1990] S. Majid, "Physics for algebraists: noncommutative and noncocommutative Hopf algebras by a bicrossproduct construction", *J. Algebra* **130**:1 (1990), 17–64. MR Zbl
- [Masuoka 1995] A. Masuoka, "Self-dual Hopf algebras of dimension p^3 obtained by extension", *J. Algebra* **178**:3 (1995), 791–806. MR Zbl

- [Masuoka 1997a] A. Masuoka, “Calculations of some groups of Hopf algebra extensions”, *J. Algebra* **191**:2 (1997), 568–588. MR Zbl
- [Masuoka 1997b] A. Masuoka, “Corrigendum: “Calculations of some groups of Hopf algebra extensions” [J. Algebra **191** (1997), no. 2, 568–588; MR1448809 (98i:16042)]”, *J. Algebra* **197**:2 (1997), 656. MR
- [Masuoka 1999] A. Masuoka, “Extensions of Hopf algebras”, Trabajo de matemática 41/99, FaMAF, Universidad Nacional de Córdoba (Argentina), 1999. Zbl
- [Masuoka 2000] A. Masuoka, “Extensions of Hopf algebras and Lie bialgebras”, *Trans. Amer. Math. Soc.* **352**:8 (2000), 3837–3879. MR Zbl
- [Masuoka 2002] A. Masuoka, “Hopf algebra extensions and cohomology”, pp. 167–209 in *New directions in Hopf algebras*, edited by S. Montgomery and H.-J. Schneider, Math. Sci. Res. Inst. Publ. **43**, Cambridge Univ. Press, 2002. MR Zbl
- [Masuoka 2003] A. Masuoka, “Cohomology and coquasi-bialgebra extensions associated to a matched pair of bialgebras”, *Adv. Math.* **173**:2 (2003), 262–315. MR Zbl
- [McCleary 2001] J. McCleary, *A user’s guide to spectral sequences*, 2nd ed., Cambridge Studies in Advanced Mathematics **58**, Cambridge University Press, 2001. MR Zbl
- [Natale 1999] S. Natale, “On semisimple Hopf algebras of dimension pq^2 ”, *J. Algebra* **221**:1 (1999), 242–278. MR Zbl
- [Natale 2001] S. Natale, “On semisimple Hopf algebras of dimension pq^2 , II”, *Algebr. Represent. Theory* **4**:3 (2001), 277–291. MR Zbl
- [Natale 2004] S. Natale, “On semisimple Hopf algebras of dimension pq^r ”, *Algebr. Represent. Theory* **7**:2 (2004), 173–188. MR Zbl
- [Singer 1972] W. M. Singer, “Extension theory for connected Hopf algebras”, *J. Algebra* **21** (1972), 1–16. MR Zbl
- [Snapper 1964] E. Snapper, “Cohomology of permutation representations, I: Spectral sequences”, *J. Math. Mech.* **13** (1964), 133–161. MR Zbl
- [Takeuchi 1981] M. Takeuchi, “Matched pairs of groups and bismash products of Hopf algebras”, *Comm. Algebra* **9**:8 (1981), 841–882. MR Zbl

Communicated by Susan Montgomery

Received 2018-06-15 Revised 2018-09-18 Accepted 2019-02-22

cn.galindo1116@uniandes.edu.co

Departamento de Matemáticas, Universidad de Los Andes, Bogota, Colombia

yk.morales964@uniandes.edu.co

Departamento de Matemáticas, Universidad de los Andes, Bogota, Colombia

On the paramodularity of typical abelian surfaces

Armand Brumer, Ariel Pacetti, Cris Poor, Gonzalo Tornara, John Voight and David S. Yuen

Generalizing the method of Faltings–Serre, we rigorously verify that certain abelian surfaces without extra endomorphisms are paramodular. To compute the required Hecke eigenvalues, we develop a method of specialization of Siegel paramodular forms to modular curves.

An errata, with an appendix by J.-P. Serre, was submitted on 9 Nov 2020 and posted online on 11 Nov 2020.

1. Introduction	1145
2. A general Faltings–Serre method	1149
3. Core-free subextensions	1156
4. Galois representations	1163
5. Group theory and Galois theory for $\mathrm{GSp}_4(\mathbb{F}_2)$	1172
6. Computing Hecke eigenvalues by specialization	1177
7. Verifying paramodularity	1187
Acknowledgements	1192
References	1192

1. Introduction

1.1. Paramodularity. The Langlands program predicts deep connections between geometry and automorphic forms, encoded in associated L -functions and Galois representations. The celebrated modularity of elliptic curves E over \mathbb{Q} [Wiles 1995; Taylor and Wiles 1995; Breuil et al. 2001] provides an important instance of this program: to the isogeny class of E of conductor N , we associate a classical cuspidal newform $f \in S_2(\Gamma_0(N))$ of weight 2 and level N with rational Hecke eigenvalues such that $L(E, s) = L(f, s)$, and conversely. In particular, $L(E, s)$ shares the good analytic properties of $L(f, s)$ including analytic continuation and functional equation, and the ℓ -adic Galois representations of E and of f are equivalent. More generally, by work of Ribet [1992] and the proof of Serre’s conjecture by Khare and Wintenberger [2009a; 2009b], isogeny classes of abelian varieties A of dimension d , of GL_2 -type over \mathbb{Q} , and of conductor N^d are in bijection with Galois orbits of classical cuspidal newforms $f \in S_2(\Gamma_1(N))$, with matching (imprimitive) L -functions and ℓ -adic Galois representations.

Continuing this program, let A be an abelian surface over \mathbb{Q} ; for instance, we may take $A = \mathrm{Jac}(X)$ the Jacobian of a curve of genus 2 over \mathbb{Q} . We suppose that $\mathrm{End}(A) = \mathbb{Z}$, i.e., A has minimal endomorphisms defined over \mathbb{Q} , and in particular A is *not* of GL_2 -type over \mathbb{Q} . For example, if A has prime conductor, then $\mathrm{End}(A) = \mathbb{Z}$ by a theorem of Ribet (see Lemma 4.1.2). A conjecture of H. Yoshida [1980; 2007]

MSC2010: primary 11F46; secondary 11Y40.

Keywords: abelian surfaces, Siegel modular forms, computation.

compatible with the Langlands program is made precise by a conjecture of Brumer and Kramer [2014, Conjecture 1.1], restricted here for simplicity.

Conjecture 1.1.1 (Brumer and Kramer). *To every abelian surface A over \mathbb{Q} of conductor N with $\text{End}(A) = \mathbb{Z}$, there exists a cuspidal Siegel paramodular newform f of degree 2, weight 2, and level N with rational Hecke eigenvalues that is not a Gritsenko lift, such that*

$$L(A, s) = L(f, s, \text{spin}). \quad (1.1.2)$$

Moreover, f is unique up to (nonzero) scaling and depends only on the isogeny class of A ; and if N is squarefree, then this association is bijective.

Conjecture 1.1.1 is often referred to as the *paramodular conjecture*; in what follows, we say *nonlift* for not a Gritsenko lift. As pointed out by Frank Calegari, in general it is necessary to include abelian fourfolds with quaternionic multiplication for the converse assertion: for a precise statement for arbitrary N and further discussion, see [Brumer and Kramer 2019, Section 8].

Extensive experimental evidence supports Conjecture 1.1.1 [Brumer and Kramer 2014; Poor and Yuen 2015]. There is also theoretical evidence for this conjecture when the abelian surface A is potentially of GL_2 -type, acquiring extra endomorphisms over a quadratic field: see Johnson-Leung and Roberts [2012] for real quadratic fields, Berger, Dembele, Pacetti, and ˘engun [2015] for imaginary quadratic fields, and Dembele and Kumar [2016] for explicit examples. For a complete treatment of the many possibilities for the association of modular forms to abelian surfaces with potentially extra endomorphisms, see work of Booker, Sijssling, Sutherland, Voight, and Yasaki [2016]. What remains is the case where $\text{End}(A_{\mathbb{Q}^{\text{al}}}) = \mathbb{Z}$, which is to say that A has minimal endomorphisms defined over the algebraic closure \mathbb{Q}^{al} ; we say then that A is *typical*. (We do not say *generic*, since it is not a Zariski open condition on the moduli space.)

Recently, there has been dramatic progress in modularity lifting theorems for nonlift Siegel modular forms (i.e., forms not of *endoscopic type*): see Pilloni [2012] for p -adic overconvergent modularity lifting, as well as recent work by Calegari and Geraghty [2016, §1.2], Berger and Klosin with Poor, Shurman and Yuen [Berger and Klosin 2017] establishing modularity in the reducible case when certain congruences are provided, and a recent manuscript by Boxer, Calegari, Gee, and Pilloni [2018] establishing potential modularity over totally real fields.

1.2. Main result. For all *prime* levels $N < 277$, the paramodular conjecture is known: there are no paramodular forms of the specified type by work of Poor and Yuen [2015, Theorem 1.2], and correspondingly there are no abelian surfaces by work of Brumer and Kramer [2014, Proposition 1.5]. At level $N = 277$, there exists a cuspidal, nonlift Siegel paramodular cusp form, unique up to scalar multiple, by work of Poor and Yuen [2015, Theorem 1.3]: this form is given explicitly as a rational function in Gritsenko lifts of ten weight 2 theta blocks — see (6.2.2).

Our main result is as follows.

Theorem 1.2.1. *Let X be the curve over \mathbb{Q} defined by*

$$y^2 + (x^3 + x^2 + x + 1)y = -x^2 - x;$$

let $A = \text{Jac}(X)$ be its Jacobian, a typical abelian surface over \mathbb{Q} of conductor 277. Let f be the cuspidal, nonlift Siegel paramodular form of genus 2, weight 2, and conductor 277, unique up to scalar multiple. Then

$$L(A, s) = L(f, s, \text{spin}).$$

Theorem 1.2.1 is not implied by any of the published or announced results on paramodularity, and its announcement in October 2015 makes it the first established typical case of the paramodular conjecture. More recently, Berger and Klosin with Poor, Shurman, and Yuen [Berger and Klosin 2017] recently established the paramodularity of an abelian surface of conductor 731 using a congruence with a Siegel Saito–Kurokawa lift.

Returning to the paramodular conjecture, by work of Brumer and Kramer [2018, Theorem 1.2] there is a unique isogeny class of abelian surfaces (LMFDB label 277.a) of conductor 277. Therefore, the proof of Conjecture 1.1.1 for $N = 277$ is completed by Theorem 1.2.1. (More generally, Brumer and Kramer [2014] also consider odd semistable conductors at most 1000.)

The theorem implies, and we prove directly, the equality of polynomials $L_p(A, T) = Q_p(f, T)$ for all primes p arising in the Euler product for the corresponding L -series. These equalities are useful in two ways. On the one hand, the Euler factors $L_p(A, T)$ can be computed much more efficiently than for $Q_p(f, T)$: without modularity, to compute the eigenvalues of a Siegel modular form f is difficult and sensitive to the manner in which f was constructed, whereas computing $L_p(A, T)$ can be done in average polynomial time [Harvey 2014] and also efficiently in practice [Harvey and Sutherland 2016]. On the other hand, the L -series $L(A, s)$ is endowed with the good analytic properties of $L(f, s, \text{spin})$: without (potential) modularity, one knows little about $L(A, s)$ beyond convergence in a right half-plane.

By work of Johnson-Leung and Roberts [2014, Main Theorem] there are infinitely many quadratic characters χ such that the twist f_χ of the paramodular cusp form by χ is nonzero. By a local calculation [Johnson-Leung and Roberts 2017, Theorem 3.1], we have $Q_p(f_\chi, T) = Q_p(f, \chi(p)T)$ and similarly $L_p(A_\chi, T) = L_p(A, \chi(p)T)$ for good primes p . Consequently, we have $L(A_\chi, s) = L(f_\chi, s, \text{spin})$ for infinitely many characters χ , and in this way we also establish the paramodularity of infinitely many twists.

We also establish paramodularity for two other isogeny classes in this article of conductors $N = 353$ and $N = 587$, and our method is general enough to establish paramodularity in a wide variety of cases.

1.3. The method of Faltings–Serre. We now briefly discuss the method of proof and a few relevant details. Let $\text{Gal}_{\mathbb{Q}} := \text{Gal}(\mathbb{Q}^{\text{al}} | \mathbb{Q})$ be the absolute Galois group of \mathbb{Q} . To establish paramodularity, we associate 2-adic Galois representations $\rho_A, \rho_f : \text{Gal}_{\mathbb{Q}} \rightarrow \text{GSp}_4(\mathbb{Q}_2^{\text{al}})$ to A and f , and then we prove by an extension of the Faltings–Serre method that these Galois representations are equivalent. The Galois representation for A arises via its Tate module. By contrast, the construction of the Galois representation for the Siegel paramodular form — for which the archimedean component of the associated automorphic representation is a holomorphic limit of discrete series — is much deeper: see Theorem 4.3.4 for a precise statement, attribution, and further discussion.

The first step in carrying out the Faltings–Serre method is to prove equivalence modulo 2, which can be done using information on $\bar{\rho}_f$ obtained by computing $Q_p(f, T)$ modulo 2 for a few small primes p . For example, $p = 3, 5$ are enough for $N = 277$ (see Lemma 7.1.4) and in this case the mod 2 residual Galois representations

$$\bar{\rho}_A, \bar{\rho}_f : \text{Gal}_{\mathbb{Q}} \rightarrow \text{GSp}_4(\mathbb{F}_2) \simeq S_6$$

have common image $S_5(b)$ up to conjugation. (There are two nonconjugate subgroups of S_6 isomorphic to S_5 , interchanged by an outer automorphism of S_6 : see (5.1.8).)

The second step is to show that the traces of the two representations agree for an effectively computable set of primes p . For example, to finish the proof of Theorem 1.2.1 in level $N = 277$, it suffices to show equality of traces for primes $p \leq 43$.

We also carry out this strategy to prove paramodularity for two other isogeny classes of abelian surfaces. For $N = 353$, we have the isogeny class with LMFDB label 353.a; we again represent the paramodular form as a rational function in Gritsenko lifts; and the common mod 2 image is instead the wreath product $S_3 \wr S_2$ of order 72. For $N = 587$, we have the class with label 587.a; instead, we represent the form as a Borcherds product; and in this case the mod 2 image is the full group S_6 .

1.4. Contributions and organization. Our contributions in this article are threefold. First, we show how to extend the Faltings–Serre method from GL_2 to a general algebraic group when the residual mod ℓ representations are absolutely irreducible. We then discuss making this practical by consideration of core-free subgroups in a general context, and we hope this will be useful in future investigations. We then make these extensions explicit for GSp_4 and $\ell = 2$. Whereas for GL_2 , Serre’s original “quartic method” considers extensions whose Galois groups are no larger than S_4 , for GSp_4 we must contemplate large polycyclic extensions of S_6 –extensions — accordingly, the Galois theory and class field theory required to make the method explicit and to work in practice are much more involved. It would be much more difficult (perhaps hopeless) to work with GL_4 instead of GSp_4 , so our formulation is crucial for practical implementation.

By other known means, the task of calculating the required traces for ρ_f would be extremely difficult. Our second contribution in this article is to devise and implement a method of *specialization* of the Siegel modular form to a classical modular form, making this calculation a manageable task.

Our third contribution is to carry out the required computations. There are nine absolutely irreducible subgroups of $\text{GSp}_4(\mathbb{F}_2)$. The three examples we present cover each of the three possibilities for the residual image when it is absolutely irreducible and the level is squarefree (see Lemma 5.2.1). Our methods work for any abelian surface whose mod 2 image is absolutely irreducible, as well as situations for paramodular forms of higher weight. Our implementations are suitable for further investigations along these lines.

The paper is organized as follows. In Section 2, we explain the extension of the method of Faltings–Serre in a general (theoretical) algorithmic context; we continue in Section 3 by noting a practical extension of this method using some explicit Galois theory. We then consider abelian surfaces, paramodular forms, and their associated Galois representations tailored to our setting in Section 4. Coming to our intended application, we provide in Section 5 the group theory and Galois theory needed for the Faltings–Serre

method for $\mathrm{GSp}_4(\mathbb{Z}_2)$. In Section 6, we explain a method to compute Hecke eigenvalues of Siegel paramodular forms using restriction to a modular curve. Finally, in Section 7, we combine these to complete our task and verify paramodularity.

2. A general Faltings–Serre method

In this section, from the point of view of general algorithmic theory, we formulate the Faltings–Serre method to show that two ℓ -adic Galois representations are equivalent, under the hypothesis that the residual representations are absolutely irreducible. A practical method for the group $\mathrm{GSp}_4(\mathbb{Z}_2)$ is given in Section 5. For further reading on the Faltings–Serre method, see the original criterion given by Serre [1985] for elliptic curves over \mathbb{Q} , an extension for residually reducible representations by Livné [1987, §4], the general overview for GL_2 over number fields by Dieulefait, Guerberoff, and Pacetti [2010, §4], and the description for GL_n by Schütt [2006, §5]. For an algorithmic approach in the pro- p setting, see [Grenié 2007].

2.1. Trace computable representations. Let F be a number field with ring of integers \mathbb{Z}_F . Let F^{al} be an algebraic closure of F ; we take all algebraic extensions of F inside F^{al} . Let $\mathrm{Gal}_F := \mathrm{Gal}(F^{\mathrm{al}} | F)$ be the absolute Galois group of F . Let S be a finite set of places of F , let $\mathrm{Gal}_{F,S}$ be the Galois group of the maximal subextension of $F^{\mathrm{al}} \supseteq F$ unramified away from S . By a *prime* of F we mean a nonzero prime ideal $\mathfrak{p} \subset \mathbb{Z}_F$, or equivalently, a finite place of F .

Let $G \subseteq \mathrm{GL}_n$ be an embedded algebraic group over \mathbb{Q} . Let ℓ be a prime of good reduction for the inclusion $G \subseteq \mathrm{GL}_n$. A *representation* $\mathrm{Gal}_{F,S} \rightarrow G(\mathbb{Z}_\ell)$ is a continuous homomorphism.

Definition 2.1.1. Let $\rho_1, \rho_2 : \mathrm{Gal}_{F,S} \rightarrow G(\mathbb{Z}_\ell)$ be two representations. We say ρ_1 and ρ_2 are (GL_n) -*equivalent*, and we write $\rho_1 \simeq \rho_2$, if there exists $g \in \mathrm{GL}_n(\mathbb{Z}_\ell)$ such that

$$\rho_1(\sigma) = g\rho_2(\sigma)g^{-1}, \quad \text{for all } \sigma \in \mathrm{Gal}_{F,S}.$$

Definition 2.1.2. A representation $\rho : \mathrm{Gal}_{F,S} \rightarrow G(\mathbb{Z}_\ell)$ is *trace computable* if $\mathrm{tr} \rho$ takes values in a computable subring of \mathbb{Z}_ℓ and there exists a deterministic algorithm to compute $\mathrm{tr}(\mathrm{Frob}_{\mathfrak{p}})$ for $\mathfrak{p} \notin S$, where $\mathrm{Frob}_{\mathfrak{p}}$ denotes the conjugacy class of the Frobenius automorphism at \mathfrak{p} .

For precise definitions and a thorough survey of the subject of computable rings, see [Stoltenberg-Hansen and Tucker 1999]. See [Cohen 1993] for background on algorithmic number theory.

Remark 2.1.3. Galois representations arising in arithmetic geometry are often trace computable. For example, by counting points over finite fields, we may access the trace of Frobenius acting on Galois representations arising from the étale cohomology of a nice variety: then the trace takes values in $\mathbb{Z} \subseteq \mathbb{Z}_\ell$ (independent of ℓ). Similarly, algorithms to compute modular forms give as output Hecke eigenvalues, which can then be interpreted in terms of the trace of Frobenius on the associated Galois representation.

Looking only at the trace of a representation is justified in certain cases by the following theorem, a cousin to the Brauer–Nesbitt theorem. For $r \geq 1$, write

$$\rho \bmod \ell^r : \text{Gal}_{F,S} \rightarrow \text{G}(\mathbb{Z}/\ell^r\mathbb{Z})$$

for the reduction of ρ modulo ℓ^r , and as a shorthand write

$$\bar{\rho} : \text{Gal}_{F,S} \rightarrow \text{G}(\mathbb{F}_\ell)$$

for the residual representation $\bar{\rho} = \rho \bmod \ell$. Given two representations $\rho_1, \rho_2 : \text{Gal}_{F,S} \rightarrow \text{G}(\mathbb{Z}_\ell)$, we write $\rho_1 \simeq \rho_2 \pmod{\ell^r}$ to mean that $(\rho_1 \bmod \ell^r) \simeq (\rho_2 \bmod \ell^r)$ are equivalent as in Definition 2.1.1 but over $\mathbb{Z}/\ell^r\mathbb{Z}$; we write $\rho_1 \equiv \rho_2 \pmod{\ell^r}$ to mean that $(\rho_1 \bmod \ell^r) = (\rho_2 \bmod \ell^r)$; and we write $\text{tr } \rho_1 \equiv \text{tr } \rho_2 \pmod{\ell^r}$ if $\text{tr } \rho_1(\sigma) \equiv \text{tr } \rho_2(\sigma) \pmod{\ell^r}$ for all $\sigma \in \text{Gal}_{F,S}$. Finally, we say that $\bar{\rho}$ is *absolutely irreducible* if the representation $\text{Gal}_{F,S} \rightarrow \text{G}(\mathbb{F}_\ell) \hookrightarrow \text{GL}_n(\mathbb{F}_\ell)$ is absolutely irreducible.

Theorem 2.1.4 (Carayol). *Let $\rho_1, \rho_2 : \text{Gal}_{F,S} \rightarrow \text{G}(\mathbb{Z}_\ell)$ be two representations such that $\bar{\rho}_1$ is absolutely irreducible and let $r \geq 1$. Then $\rho_1 \simeq \rho_2 \bmod \ell^r$ if and only if $\text{tr } \rho_1 \equiv \text{tr } \rho_2 \pmod{\ell^r}$.*

Proof. See [Carayol 1994, Theoreme 1]. □

We now state the main result of this section. We say that a prime \mathfrak{p} of F is a *witness* to the fact that $\rho_1 \not\simeq \rho_2$ if $\text{tr } \rho_1(\text{Frob}_\mathfrak{p}) \neq \text{tr } \rho_2(\text{Frob}_\mathfrak{p})$.

Theorem 2.1.5. *There is a deterministic algorithm that takes as input*

$$\begin{aligned} & \text{an algebraic group } G \text{ over } \mathbb{Q}, \text{ a number field } F, \\ & \text{a finite set } S \text{ of primes of } F, \text{ a prime } \ell, \text{ and} \\ & \rho_1, \rho_2 : \text{Gal}_{F,S} \rightarrow \text{G}(\mathbb{Z}_\ell) \text{ trace computable representations} \\ & \text{with } \bar{\rho}_1, \bar{\rho}_2 \text{ absolutely irreducible,} \end{aligned} \tag{2.1.6}$$

and gives as output

$$\begin{cases} \text{true} & \text{if } \rho_1 \simeq \rho_2; \\ \text{false and a witness prime } \mathfrak{p} \notin S & \text{if } \rho_1 \not\simeq \rho_2. \end{cases}$$

The algorithm does not operate on the representations ρ_1, ρ_2 themselves, only their traces. The proof of Theorem 2.1.5 will occupy us throughout this section.

2.2. Testing equivalence of residual representations. We first prove a variant of our theorem for the residual representations. For a finite extension $K_0 \supseteq F$ of fields with $[K_0 : F] = n$ and with Galois closure K , we write $\text{Gal}(K_0 | F) \leq S_n$ for the Galois group $\text{Gal}(K | F)$ as a permutation group on the roots of a minimal polynomial of a primitive element for K_0 .

Lemma 2.2.1. *There exists a deterministic algorithm that takes as input*

$$\begin{aligned} & \text{a number field } F, \\ & \text{a finite set } S \text{ of places of } F, \text{ and} \\ & \text{a transitive group } G \leq S_n, \end{aligned}$$

and gives as output

all extensions $K_0 \supseteq F$ (up to isomorphism) of degree n unramified at all places $v \notin S$
such that $\text{Gal}(K_0 | F) \simeq G$ as permutation groups.

Moreover, every Galois extension $K \supseteq F$ unramified outside S such that $\text{Gal}(K | F) \simeq G$ as groups appears as the Galois closure of at least one such $K_0 \supseteq F$.

Proof. The extensions K_0 have degree n and are unramified away from S , so they have effectively bounded discriminant by Krasner’s lemma. Therefore, there are finitely many such fields up to isomorphism, by a classical theorem of Hermite. The enumeration can be accomplished algorithmically by a *Hunter search*: see [Cohen 2000, §9.3]. The computation and verification of Galois groups can also be accomplished effectively.

The second statement follows from basic Galois theory. □

Remark 2.2.2. For theoretical purposes, it is enough to consider $G \hookrightarrow S_n$ in its regular representation ($n = \#G$), for which the algorithm yields Galois extensions $K = K_0 \supseteq F$. For practical purposes, it is crucial to work with small permutation representations.

Algorithm 2.2.3. The following algorithm takes as input the data (2.1.6) and gives as output

$$\begin{cases} \text{true} & \text{if } \bar{\rho}_1 \simeq \bar{\rho}_2; \\ \text{false and a witness prime } \mathfrak{p} \notin S & \text{if } \bar{\rho}_1 \not\simeq \bar{\rho}_2. \end{cases}$$

1. Using the algorithm of Lemma 2.2.1, enumerate all Galois extensions $K \supseteq F$ up to isomorphism that are unramified away from S and such that $\text{Gal}(K | F)$ is isomorphic to a subgroup of $G(\mathbb{F}_\ell)$.
2. For each of these finitely many fields, enumerate all injective group homomorphisms $\theta : \text{Gal}(K | F) \hookrightarrow G(\mathbb{F}_\ell)$ up to conjugation by $\text{GL}_n(\mathbb{F}_\ell)$.
3. Looping over primes $\mathfrak{p} \notin S$ of F , rule out pairs (K, θ) such that

$$\text{tr } \rho_1(\text{Frob}_\mathfrak{p}) \not\equiv \text{tr } \theta(\text{Frob}_\mathfrak{p}) \pmod{\ell}$$

for some \mathfrak{p} until only one possibility (K_1, θ_1) remains.

4. Let \mathcal{P} be the set of primes used in Step 3. If

$$\text{tr } \rho_2(\text{Frob}_\mathfrak{p}) \equiv \text{tr } \theta_1(\text{Frob}_\mathfrak{p}) \pmod{\ell}$$

for all $\mathfrak{p} \in \mathcal{P}$, return `true`; otherwise, return `false` and a prime $\mathfrak{p} \in \mathcal{P}$ such that $\text{tr } \rho_2(\text{Frob}_\mathfrak{p}) \not\equiv \text{tr } \theta_1(\text{Frob}_\mathfrak{p})$.

Proof of correctness. Let K_1 be the fixed field under $\ker \bar{\rho}_1$; then K_1 is unramified away from S , and we have an injective homomorphism $\bar{\rho}_1 : \text{Gal}(K_1 | F) \hookrightarrow G(\mathbb{F}_\ell)$. Thus $(K_1, \bar{\rho}_1)$ is among the finite list of pairs (K, θ) computed in Step 2.

Combining Theorem 2.1.4 (for $r = 1$) and the Chebotarev density theorem, we can effectively determine if $\bar{\rho}_1 \not\equiv \theta$ by finding a prime \mathfrak{p} such that $\text{tr } \rho_1(\text{Frob}_\mathfrak{p}) \not\equiv \text{tr } \theta(\text{Frob}_\mathfrak{p}) \pmod{\ell}$. So by looping over the

primes $\mathfrak{p} \notin S$ of F in Step 3, we will eventually rule out all of the finitely many candidates except one (K'_1, θ'_1) and, in the style of Sherlock Holmes, we must have $K_1 = K'_1$ and $\bar{\rho}_1 \simeq \theta_1$.

For the same reason, if $\text{tr } \rho_2(\text{Frob}_{\mathfrak{p}}) \equiv \text{tr } \theta_1(\text{Frob}_{\mathfrak{p}}) \pmod{\ell}$ for all $\mathfrak{p} \in \mathcal{P}$ we must have $\bar{\rho}_2 \simeq \theta_1 \simeq \bar{\rho}_1$. Otherwise, we find a witness prime $\mathfrak{p} \in \mathcal{P}$. □

Remark 2.2.4. In practice, we may also use the characteristic polynomial of $\bar{\rho}_i(\text{Frob}_{\mathfrak{p}})$ when it is computable, since it gives more information about the residual image and thereby limits the possible subgroups of $G(\mathbb{F}_{\ell})$ we need to consider in Step 1. This allows for a smaller list of pairs (K, θ) and a smaller list of primes: see Lemma 7.1.4 for an example.

2.3. Faltings–Serre and deformation. With the residual representations identified, we now explain the key idea of the Faltings–Serre method: we exhibit another representation that measures the failure of two representations to be equivalent. This construction is quite natural when viewed in the language of deformation theory: see [Gouvea 2001, Lecture 4] for background.

For the remainder of this section, let $\rho_1, \rho_2 : \text{Gal}_{F,S} \rightarrow G(\mathbb{Z}_{\ell})$ be representations such that $\rho_1 \simeq \rho_2 \pmod{\ell^r}$ for some $r \geq 1$. Conjugating ρ_2 , we may assume $\rho_1 \equiv \rho_2 \pmod{\ell^r}$, and we write $\bar{\rho} := \bar{\rho}_1 = \bar{\rho}_2$ for the common residual representation modulo ℓ . We suppose throughout that $\bar{\rho}$ is absolutely irreducible.

Let $\text{Lie}(G) \leq M_n$ be the Lie algebra of G over \mathbb{Q} as a commutative algebraic group. Attached to $\bar{\rho}$ is the *adjoint residual representation*

$$\begin{aligned} \text{ad } \bar{\rho} : \text{Gal}_{F,S} &\rightarrow \text{Aut}_{\mathbb{F}_{\ell}}(M_n(\mathbb{F}_{\ell})) \\ \sigma &\mapsto \sigma_{\text{ad}} \end{aligned} \tag{2.3.1}$$

defined by $\sigma_{\text{ad}}(a) := \bar{\rho}(\sigma)a\bar{\rho}(\sigma)^{-1}$ for $a \in M_n(\mathbb{F}_{\ell})$. The adjoint residual representation $\text{ad } \bar{\rho}$ also restricts to take values in $\text{Aut}_{\mathbb{F}_{\ell}}(\text{Lie}(G)(\mathbb{F}_{\ell}))$, but we will not need to introduce new notation for this restriction.

Because we consider representations with values in G up to equivalence in GL_n , it is natural that our deformations will take values in $\text{Lie}(G)$ up to equivalence in M_n . With this in mind, we define the group of *cocycles*

$$\begin{aligned} Z^1(F, \text{ad } \bar{\rho}; \text{Lie}(G)(\mathbb{F}_{\ell})) \\ := \{(\mu : \text{Gal}_{F,S} \rightarrow \text{Lie}(G)(\mathbb{F}_{\ell})) : \mu(\sigma\tau) = \mu(\sigma) + \sigma_{\text{ad}}(\mu(\tau)), \forall \sigma, \tau \in \text{Gal}_{F,S}\} \end{aligned} \tag{2.3.2}$$

and the subgroup of *coboundaries*

$$\begin{aligned} B^1(F, \text{ad } \bar{\rho}; M_n(\mathbb{F}_{\ell})) \\ := \{\mu \in Z^1(F, \text{ad } \bar{\rho}; \text{Lie}(G)(\mathbb{F}_{\ell})) : \exists a \in M_n(\mathbb{F}_{\ell}) \text{ such that } \mu(\sigma) = a - \sigma_{\text{ad}}(a), \forall \sigma \in \text{Gal}_{F,S}\}. \end{aligned} \tag{2.3.3}$$

From the exact sequence

$$1 \rightarrow 1 + \ell^r \text{Lie}(G)(\mathbb{F}_{\ell}) \rightarrow G(\mathbb{Z}/\ell^{r+1}\mathbb{Z}) \rightarrow G(\mathbb{Z}/\ell^r\mathbb{Z}) \rightarrow 1, \tag{2.3.4}$$

we conclude that for all $\sigma \in \text{Gal}_{F,S}$ there exists $\mu(\sigma) \in \text{Lie}(G)(\mathbb{F}_{\ell})$ such that

$$\rho_1(\sigma) \equiv (1 + \ell^r \mu(\sigma))\rho_2(\sigma) \pmod{\ell^{r+1}}. \tag{2.3.5}$$

Lemma 2.3.6. *The following statements hold:*

- (a) *The map $\sigma \mapsto \mu(\sigma)$ defined by (2.3.5) is a cocycle $\mu \in Z^1(F, \text{ad } \bar{\rho}; \text{Lie}(\mathbf{G})(\mathbb{F}_\ell))$.*
- (b) *We have $\rho_1 \simeq \rho_2 \pmod{\ell^{r+1}}$ if and only if $\mu \in B^1(F, \text{ad } \bar{\rho}; \mathbf{M}_n(\mathbb{F}_\ell))$.*

Proof. We verify the cocycle condition as follows:

$$\begin{aligned} \rho_1(\sigma\tau) &= \rho_1(\sigma)\rho_1(\tau) \equiv (1 + \ell^r \mu(\sigma))\rho_2(\sigma)(1 + \ell^r \mu(\tau))\rho_2(\tau) \\ &\equiv (1 + \ell^r (\mu(\sigma) + \rho_2(\sigma)\mu(\tau)\rho_2(\sigma)^{-1}))\rho_2(\sigma)\rho_2(\tau) \\ &\equiv (1 + \ell^r \mu(\sigma\tau))\rho_2(\sigma\tau) \pmod{\ell^{r+1}}, \end{aligned}$$

so $\mu(\sigma\tau) = \mu(\sigma) + \sigma_{\text{ad}}(\mu(\tau))$ as claimed. For the second statement, by definition $\rho_1 \simeq \rho_2 \pmod{\ell^{r+1}}$ if and only if there exists $a_r \in \text{GL}_n(\mathbb{Z}/\ell^{r+1}\mathbb{Z})$ such that for all $\sigma \in \text{Gal}_{F,S}$ we have

$$\rho_1(\sigma) \equiv a_r \rho_2(\sigma) a_r^{-1} \pmod{\ell^{r+1}}. \tag{2.3.7}$$

Since $\rho_1(\sigma) \equiv \rho_2(\sigma) \pmod{\ell^r}$, the image of a_r in $\text{GL}_n(\mathbb{Z}/\ell^r\mathbb{Z})$ centralizes the image of $\rho \pmod{\ell^r}$. Since the image is irreducible, by Schur’s lemma we have $a_r \pmod{\ell^r}$ is scalar, so without loss of generality we may suppose $a_r \equiv 1 \pmod{\ell^r}$, so that $a_r = 1 + \ell^r a$ for some $a \in \mathbf{M}_n(\mathbb{F}_\ell)$. Expanding (2.3.7) then yields

$$\begin{aligned} \rho_1(\sigma) &\equiv (1 + \ell^r a)\rho_2(\sigma)(1 + \ell^r a)^{-1} \equiv (1 + \ell^r a)\rho_2(\sigma)(1 - \ell^r a) \\ &\equiv (1 + \ell^r a - \ell^r \rho_2(\sigma)a\rho_2(\sigma)^{-1})\rho_2(\sigma) \\ &\equiv (1 + \ell^r (a - \sigma_{\text{ad}}(a)))\rho_2(\sigma) \pmod{\ell^{r+1}} \end{aligned}$$

so $\mu(\sigma) = a - \sigma_{\text{ad}}(a)$ by definition (2.3.5). □

Our task now turns to finding an effective way to detect when μ is a coboundary. For this purpose, we work with extensions of our representations using explicit parabolic groups. The adjoint action of GL_n on \mathbf{M}_n gives an exact sequence

$$0 \rightarrow \mathbf{M}_n \rightarrow \mathbf{M}_n \rtimes \text{GL}_n \rightarrow \text{GL}_n \rightarrow 1 \tag{2.3.8}$$

which extends to a linear representation via the *parabolic subgroup*, as follows. We embed

$$\begin{aligned} \mathbf{M}_n \rtimes \text{GL}_n &\hookrightarrow \text{GL}_{2n} \\ (a, g) &\mapsto \begin{pmatrix} 1 & a \\ 0 & 1 \end{pmatrix} \begin{pmatrix} g & 0 \\ 0 & g \end{pmatrix} = \begin{pmatrix} g & ag \\ 0 & g \end{pmatrix} \end{aligned} \tag{2.3.9}$$

(on points, realizing $\mathbf{M}_n \rtimes \text{GL}_n$ as an algebraic matrix group). The embedding (2.3.9) is compatible with the exact sequence (2.3.8): the natural projection map

$$\pi : \mathbf{M}_n \rtimes \text{GL}_n \rightarrow \text{GL}_n \tag{2.3.10}$$

corresponds to the projection onto the top left entry, it is split by the diagonal embedding $\text{GL}_n \hookrightarrow \text{GL}_{2n}$, and it has kernel isomorphic to \mathbf{M}_n in the upper-right entry. We will identify $\mathbf{M}_n \rtimes \text{GL}_n$ and its subgroups with their image in GL_{2n} .

Let $\text{utr} : (\mathbf{M}_n \rtimes \text{GL}_n)(\mathbb{F}_\ell) \rightarrow \mathbb{F}_\ell$ denote the trace of the upper right $n \times n$ -block.

Lemma 2.3.11. *The map utr is well-defined on conjugacy classes in $(\mathbf{M}_n \rtimes \text{GL}_n)(\mathbb{F}_\ell)$.*

Proof. For all $g, h \in \text{GL}_n(\mathbb{F}_\ell)$ and $a, b \in \mathbf{M}_n(\mathbb{F}_\ell)$ we have

$$\begin{pmatrix} h & bh \\ 0 & h \end{pmatrix} \begin{pmatrix} g & ag \\ 0 & g \end{pmatrix} \begin{pmatrix} h^{-1} & -h^{-1}b \\ 0 & h^{-1} \end{pmatrix} = \begin{pmatrix} hgh^{-1} & hagh^{-1} + bhgh^{-1} - hgh^{-1}b \\ 0 & hgh^{-1} \end{pmatrix} \quad (2.3.12)$$

so the upper trace is $\text{tr}(hagh^{-1} + bhgh^{-1} - hgh^{-1}b) = \text{tr}(ag)$. \square

For $\mu \in Z^1(F, \text{ad } \bar{\rho}; \text{Lie}(\mathbf{G})(\mathbb{F}_\ell))$ we define

$$\begin{aligned} \varphi_\mu : \text{Gal}_{F,S} &\rightarrow (\text{Lie}(\mathbf{G}) \rtimes \mathbf{G})(\mathbb{F}_\ell) \leq \text{GL}_{2n}(\mathbb{F}_\ell) \\ \sigma &\mapsto (\mu(\sigma), \bar{\rho}(\sigma)) = \begin{pmatrix} \bar{\rho}(\sigma) & \mu(\sigma)\bar{\rho}(\sigma) \\ 0 & \bar{\rho}(\sigma) \end{pmatrix}. \end{aligned} \quad (2.3.13)$$

Proposition 2.3.14. *Let $\mu \in Z^1(F, \text{ad } \bar{\rho}; \text{Lie}(\mathbf{G})(\mathbb{F}_\ell))$. Then the following statements hold:*

- (a) *The map φ_μ defined by (2.3.13) is a group homomorphism, and $\pi \circ \varphi_\mu = \bar{\rho}$.*
- (b) *We have $\mu \in \mathbf{B}^1(F, \text{ad } \bar{\rho}; \mathbf{M}_n(\mathbb{F}_\ell))$ if and only if φ_μ is conjugate to $\varphi_0 = \begin{pmatrix} \bar{\rho} & 0 \\ 0 & \bar{\rho} \end{pmatrix}$ by an element of $\mathbf{M}_n(\mathbb{F}_\ell) \leq (\mathbf{M}_n \rtimes \text{GL}_n)(\mathbb{F}_\ell)$.*
- (c) *Suppose μ is defined by (2.3.5). Then for all $\sigma \in \text{Gal}_{F,S}$,*

$$\text{utr } \varphi_\mu(\sigma) = \text{tr}(\mu(\sigma)\bar{\rho}(\sigma)) \equiv \frac{\text{tr } \rho_1(\sigma) - \text{tr } \rho_2(\sigma)}{\ell^r} \pmod{\ell}. \quad (2.3.15)$$

Proof. For (a), the cocycle condition implies that φ_μ is a group homomorphism: the upper right entry of $\varphi_\mu(\sigma\tau)$ is

$$\mu(\sigma\tau)\bar{\rho}(\sigma\tau) = (\mu(\sigma) + \bar{\rho}(\sigma)\mu(\tau)\bar{\rho}(\sigma)^{-1})\bar{\rho}(\sigma)\bar{\rho}(\tau) = \mu(\sigma)\bar{\rho}(\sigma)\bar{\rho}(\tau) + \bar{\rho}(\sigma)\mu(\tau)\bar{\rho}(\tau)$$

which is equal to the upper right entry of $\varphi_\mu(\sigma)\varphi_\mu(\tau)$ obtained by matrix multiplication.

For (b), the calculation

$$\begin{pmatrix} 1 & a \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \bar{\rho}(\sigma) & 0 \\ 0 & \bar{\rho}(\sigma) \end{pmatrix} \begin{pmatrix} 1 & -a \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} \bar{\rho}(\sigma) & a\bar{\rho}(\sigma) - \bar{\rho}(\sigma)a \\ 0 & \bar{\rho}(\sigma) \end{pmatrix} \quad (2.3.16)$$

shows that $\varphi_\mu = a\varphi_0a^{-1}$ for $a \in \mathbf{M}_n(\mathbb{F}_\ell)$ if and only if $\mu(\sigma)\bar{\rho}(\sigma) = a\bar{\rho}(\sigma) - \bar{\rho}(\sigma)a$ for all $\sigma \in \text{Gal}_{F,S}$. Multiplying on the right by $\bar{\rho}(\sigma)^{-1}$, we see this is equivalent to $\mu(\sigma) = a - \sigma_{\text{ad}}(a)$ for all $\sigma \in \text{Gal}_{F,S}$.

Finally, (c) follows directly from (2.3.5). \square

Definition 2.3.17. Let K be the fixed field under $\bar{\rho}$. We say a pair (L, φ) extends $(K, \bar{\rho})$ if

$$\varphi : \text{Gal}_{F,S} \rightarrow (\text{Lie}(\mathbf{G}) \rtimes \mathbf{G})(\mathbb{F}_\ell) \leq \text{GL}_{2n}(\mathbb{F}_\ell)$$

is a representation with fixed field L such that $\pi \circ \varphi = \bar{\rho}$.

If (L, φ) extends $(K, \bar{\rho})$, then $L \supseteq K$ is an ℓ -elementary abelian extension unramified outside S , since φ induces an injective group homomorphism $\text{Gal}(L | K) \hookrightarrow \text{Lie}(\mathbf{G})(\mathbb{F}_\ell)$.

Definition 2.3.18. A pair (L, φ) extending $(K, \bar{\rho})$ is *obstructing* if $\text{utr } \varphi \not\equiv 0 \pmod{\ell}$, and we call the group homomorphism φ an *obstructing extension* of $\bar{\rho}$. An element $\sigma \in \text{Gal}(L | F)$ such that $\text{utr } \varphi(\sigma) \not\equiv 0 \pmod{\ell}$ is called *obstructing* for φ .

We note the following corollary of Proposition 2.3.14.

Corollary 2.3.19. *Let μ be defined by (2.3.5) and φ_μ by (2.3.13). Then φ_μ extends $\bar{\rho}$, and φ_μ is obstructing if and only if $\mu \notin B^1(F, \text{ad } \bar{\rho}; M_n(\mathbb{F}_\ell))$.*

Proof. The map φ_μ extends $\bar{\rho}$ by Proposition 2.3.14(a). We prove the contrapositive of the second statement: $\mu \in B^1(F, \text{ad } \bar{\rho}; M_n(\mathbb{F}_\ell))$ if and only if $\text{utr } \varphi_\mu \equiv 0 \pmod{\ell}$. The implication (\Rightarrow) is immediate from Proposition 2.3.14(b) and the invariance of utr by conjugation (Lemma 2.3.11). For (\Leftarrow) , if $\text{utr } \varphi_\mu \equiv 0 \pmod{\ell}$ then $\text{tr } \rho_1 \equiv \text{tr } \rho_2 \pmod{\ell^{r+1}}$ by Proposition 2.3.14(c). Now Theorem 2.1.4 implies $\rho_1 \simeq \rho_2 \pmod{\ell^{r+1}}$, hence $\mu \in B^1(F, \text{ad } \bar{\rho}; M_n(\mathbb{F}_\ell))$ by Lemma 2.3.6(b). \square

Before we conclude this section, we note the following important improvement. Let $\text{Lie}^0(G) \leq \text{Lie}(G)$ be the subgroup of trace zero matrices, and note that $\text{Lie}^0(G)(\mathbb{F}_\ell)$ is invariant by the adjoint residual representation.

Lemma 2.3.20. *If $\det \rho_1 = \det \rho_2$, then μ takes values in $\text{Lie}^0(G)(\mathbb{F}_\ell)$.*

Proof. By (2.3.5), we have $1 = \det(\rho_1 \rho_2^{-1}) = \det(1 + \ell^r \mu) \equiv 1 + \ell^r \text{tr } \mu \pmod{\ell^{2r}}$ so accordingly $\text{tr } \mu(\sigma) \equiv 0 \pmod{\ell}$ and $\mu(\sigma) \in \text{Lie}^0(G)(\mathbb{F}_\ell)$ for all $\sigma \in \text{Gal}_{F,S}$. \square

In view of Lemma 2.3.20, we note that Proposition 2.3.14 and Corollary 2.3.19 hold when replacing $\text{Lie}(G)$ by $\text{Lie}^0(G)$.

2.4. Testing equivalence of representations. We now use Corollary 2.3.19 to prove Theorem 2.1.5.

Algorithm 2.4.1. The following algorithm takes as input the data (2.1.6) and gives as output

$$\begin{cases} \text{true} & \text{if } \rho_1 \simeq \rho_2; \\ \text{false and a witness prime } \mathfrak{p} & \text{if } \rho_1 \not\simeq \rho_2. \end{cases}$$

1. Apply Algorithm 2.2.3; if $\bar{\rho}_1 \not\cong \bar{\rho}_2$, return `false` and the witness prime \mathfrak{p} . Otherwise, let K be the fixed field under the common residual representation $\bar{\rho}$.
2. Using the algorithm of Lemma 2.2.1, enumerate all ℓ -elementary abelian extensions $L \supseteq K$ unramified away from S and such that $\text{Gal}(L | F)$ is isomorphic to a subgroup of $(\text{Lie}(G) \rtimes G)(\mathbb{F}_\ell)$.
3. For each of these finitely many fields L , by enumeration of injective group homomorphisms $\text{Gal}(L | F) \hookrightarrow (\text{Lie}(G) \rtimes G)(\mathbb{F}_\ell)$, find all obstructing pairs (L, φ) extending $(K, \bar{\rho})$ up to conjugation by $(M_n \rtimes \text{GL}_n)(\mathbb{F}_\ell)$.
4. For each such pair (L, φ) , find a prime $\mathfrak{p} \notin S$ such that $\text{utr } \varphi(\text{Frob}_{\mathfrak{p}}) \not\equiv 0 \pmod{\ell}$.
5. Check if $\text{tr } \rho_1(\text{Frob}_{\mathfrak{p}}) = \text{tr } \rho_2(\text{Frob}_{\mathfrak{p}})$ for the primes in Step 4. If equality holds for all primes, return `true`; if equality fails for \mathfrak{p} , return `false` and the prime \mathfrak{p} .

Remark 2.4.2. In Step 2, we may instead use algorithmic class field theory (and we will do so in practice). Moreover, if we know that $\det \rho_1 = \det \rho_2$, then we can replace $\text{Lie}(G)$ by $\text{Lie}^0(G)$ by Lemma 2.3.20.

Proof of correctness. By the Chebotarev density theorem, in Step 4 we will eventually find a prime $\mathfrak{p} \notin S$, since utr is well-defined on conjugacy classes by Lemma 2.3.11. In the final step, if equality does not hold for some prime \mathfrak{p} , we have found a witness, and we correctly return `false`.

Otherwise, we return `true` and we claim that $\rho_1 \simeq \rho_2$ so the output is correct. Indeed, assume for purposes of contradiction that $\rho_1 \not\simeq \rho_2$. Then there exists $r \geq 1$ such that $\rho_1 \simeq \rho_2 \pmod{\ell^r}$ but $\rho_1 \not\simeq \rho_2 \pmod{\ell^{r+1}}$. We can assume as before that $\rho_1 \equiv \rho_2 \pmod{\ell^r}$. We define μ by (2.3.5) and φ_μ by (2.3.13). Let L_μ be the fixed field of φ_μ . By Lemma 2.3.6 we have $\mu \notin B^1(F, \text{Lie}(G)(\mathbb{F}_\ell); M_n(\mathbb{F}_\ell))$, hence by Corollary 2.3.19 φ_μ extends $\bar{\rho}$ and is obstructing. It follows that the pair (L_μ, φ_μ) is, up to conjugation by $(M_n \rtimes \text{GL}_n)(\mathbb{F}_\ell)$, among the pairs computed in Step 3. In particular there is a prime \mathfrak{p} in Step 4 such that $\text{utr} \varphi_\mu(\text{Frob}_\mathfrak{p}) \not\equiv 0 \pmod{\ell}$. But then by (2.3.15) we would have $\text{tr} \rho_1(\text{Frob}_\mathfrak{p}) \neq \text{tr} \rho_2(\text{Frob}_\mathfrak{p})$, contradicting the verification carried out in Step 5. \square

The correctness of Algorithm 2.4.1 then proves Theorem 2.1.5.

Remark 2.4.3. In the case $G = \text{GSp}_{2g}$, using an effective version of the Chebotarev density theorem, Achter [2005, Lemma 1.2] has given an effective upper bound in terms of the conductor and genus to detect when two abelian surfaces are isogenous. This upper bound is of theoretical interest, but much too large to be useful in practice. In a similar way, following the above strategy one could give theoretical (but practically useless) upper bounds to detect when two Galois representations are equivalent.

3. Core-free subextensions

The matrix groups arising in the previous section are much too large to work with in practice. In this section, we find comparatively small extensions whose Galois closure give rise to the desired representations.

3.1. Core-free subgroups. We begin with a condition that arises naturally in group theory and Galois theory.

Definition 3.1.1. Let G be a finite group. A subgroup $H \leq G$ is *core-free* if G acts faithfully on the cosets G/H .

Equivalently, $H \leq G$ is core-free if and only if $\bigcap_{g \in G} gHg^{-1} = \{1\}$. For example, the subgroup $\{1\}$ is core-free.

Definition 3.1.2. Let $K \supseteq F$ be a finite Galois extension of fields with $G = \text{Gal}(K | F)$. A subextension $K \supseteq K_0 \supseteq F$ is *core-free* if $\text{Gal}(K | K_0) \leq G$ is a core-free subgroup.

Lemma 3.1.3. *The subextension $K \supseteq K_0 \supseteq F$ is core-free if and only if K is the Galois closure of K_0 over F .*

Proof. Immediate. \square

If $K \supseteq K_0 \supseteq F$ is a core-free subextension of $K \supseteq F$ with $K_0 = F(\alpha)$, then by definition the action of $\text{Gal}(K | F)$ on the conjugates of α defines a faithful permutation representation, equivalent to its action on the left cosets of $\text{Gal}(K | K_0)$.

We slightly augment the notion of core-free subextension for two-step extensions of fields, as follows.

Definition 3.1.4. Let

$$1 \rightarrow V \rightarrow E \xrightarrow{\pi} G \rightarrow 1 \tag{3.1.5}$$

be an exact sequence of finite groups. A core-free subgroup $D \leq E$ is *exact (relative to (3.1.5))* if $\pi(D)$ is a core-free subgroup of G .

If $D \leq E$ is an exact core-free subgroup we let $H := \pi(D)$ and $W := V \cap D = \ker \pi|_D$, so there is an exact subsequence

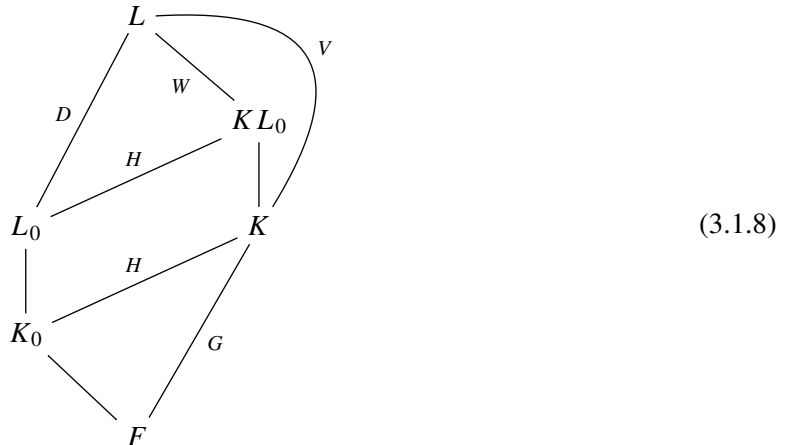
$$1 \rightarrow W \rightarrow D \xrightarrow{\pi} H \rightarrow 1 \tag{3.1.6}$$

with both $D \leq E$ and $H \leq G$ core-free. (We do not assume that $W \leq V$ is core-free.)

Now let $L \supseteq K \supseteq F$ be a two-step Galois extension with $V := \text{Gal}(L | K)$, $E := \text{Gal}(L | F)$, $G := \text{Gal}(K | F)$ and $\pi : E \rightarrow G$ the restriction, so we have an exact sequence as in (3.1.5).

Definition 3.1.7. We say $L_0 \supseteq K_0 \supseteq F$ is an *exact core-free subextension* of $L \supseteq K \supseteq F$ if $L_0 = L^D$ and $K_0 = K^{\pi(D)}$ where $D \leq E$ is an exact core-free subgroup.

Let $L_0 \supseteq K_0 \supseteq F$ be an exact core-free subextension of $L \supseteq K \supseteq F$, so that $\text{Gal}(L | L_0) = D$. As above we let $H := \pi(D) = \text{Gal}(K | K_0)$ and $W := V \cap D = \text{Gal}(L | KL_0)$. By (3.1.6) we have $H \simeq D/W = \text{Gal}(KL_0 | L_0)$, and we have the following field diagram:



By Lemma 3.1.3, L is the Galois closure of L_0 over F , and K is the Galois closure of K_0 over F . We read the diagram (3.1.8) as giving us a way to reduce the Galois theory of the extension $L \supseteq K \supseteq F$ to $L_0 \supseteq K_0 \supseteq F$: the larger we can make D , the smaller the extension $L_0 \supseteq K_0 \supseteq F$, and the better for working explicitly with the corresponding Galois groups.

3.2. Application to Faltings–Serre. We now specialize the preceding discussion to our case of interest; although working with core-free extensions does not improve the theoretical understanding, it is a crucial simplification in practice.

In Steps 2–3 of Algorithm 2.4.1, we are asked to enumerate obstructing pairs (L, φ) extending $(K, \bar{\rho})$, with $\varphi : \text{Gal}(L | F) \hookrightarrow (\text{Lie}(G) \rtimes G)(\mathbb{F}_\ell)$.

Let $G := \text{img } \bar{\rho} \leq G(\mathbb{F}_\ell)$. Given (L, φ) , the image of φ is a subgroup $E \leq \text{Lie}(G)(\mathbb{F}_\ell) \rtimes G$ with $\pi(E) = G$; letting $V := \text{Lie}(G)(\mathbb{F}_\ell) \cap E$ we have an exact sequence

$$1 \rightarrow V \rightarrow E \xrightarrow{\pi} G \rightarrow 1 \tag{3.2.1}$$

arising from (2.3.8).

So we enumerate the subgroups $E \leq \text{Lie}(G)(\mathbb{F}_\ell) \rtimes G$ with $\pi(E) = G$, up to conjugation by $M_n(\mathbb{F}_\ell) \rtimes G$. The enumeration of these subgroups depends only on G , so it may be done as a precomputation step, independent of the representations.

For each such E , let D be an exact core-free subgroup relative to (3.2.1). We let $L_0 = L^D$ and $K_0 = K^{\pi(D)}$, hence $L_0 \supseteq K_0 \supseteq F$ is an exact core-free subextension of $L \supseteq K \supseteq F$ and we have the field diagram (3.1.8) where $H = \pi(D)$ and $W = V \cap D$ as before. Since V is abelian, $KL_0 \supseteq K$ is Galois and hence $L_0 \supseteq K_0$ is also Galois, with common abelian Galois group $\text{Gal}(L_0 | K_0) \simeq \text{Gal}(KL_0 | K) \simeq V/W$. So better than a Hunter search as in Lemma 2.2.1, we can use algorithmic class field theory (see [Cohen 2000, Chapter 4]) to enumerate the possible fields $L_0 \supseteq K_0$.

Accordingly, we modify Steps 2–3 of Algorithm 2.4.1 then as follows.

- 2'. Enumerate the subgroups $E \leq \text{Lie}(G)(\mathbb{F}_\ell) \rtimes G$ with $\pi(E) = G$, up to conjugation by $M_n(\mathbb{F}_\ell) \rtimes G$, such that $\text{utr}(E) \not\equiv 0 \pmod{\ell}$. For each such subgroup E , perform the following steps:
 - a. Compute a set of representatives ξ of (outer) automorphisms of E such that ξ acts by an inner automorphism on G , modulo inner automorphisms by elements of $M_n(\mathbb{F}_\ell) \rtimes G$.
 - b. Find an exact core-free subgroup $D \leq E$ and let W, H be as in (3.1.6).
 - c. Let $K_0 = K^H$ and use algorithmic class field theory to enumerate all possible extensions $L_0 \supseteq K_0$ unramified away from S such that $\text{Gal}(L_0 | K_0) \simeq V/W$.
- 3'. For each extension L_0 from Step 2'c and for each E , perform the following steps:
 - a. Compute an isomorphism of groups $\varphi_0 : \text{Gal}(L | F) \xrightarrow{\sim} E$ extending $\bar{\rho}$; if no such isomorphism exists, proceed to the next group E .
 - b. Looping over ξ computed in Step 2'a, let $\varphi := \xi \circ \varphi_0$, and record the pair (L, φ) .

Proof of equivalence with Steps 2–3. We show that these steps enumerate all obstructing pairs (L, φ) up to equivalence.

Let L be an obstructing extension. For an obstructing extension φ of $\bar{\rho}$, the image $E = \text{img } \varphi$ arises up to conjugation in the list computed in Step 2'; such conjugation gives an equivalent representation. So we may restrict our attention to the set Φ of obstructing extensions φ whose image is *equal* to E .

With respect to the core-free subgroup D , the field L arises as the Galois closure of the field $L_0 = L^D$, and so L_0 will appear in the list computed in Step 2'c. An exact core-free subgroup always exists as we can always take D the trivial group.

In Step 3'a, we compute one obstructing extension $\varphi_0 \in \Phi$. Any other obstructing extension $\varphi \in \Phi$ is of the form $\varphi = \xi \circ \varphi_0$ where ξ is an automorphism of E that induces an inner automorphism on G ; when ξ arises from conjugation by an element of $\text{Lie}(G)(\mathbb{F}_\ell) \rtimes G$, we obtain a representation equivalent to φ_0 , so the representatives ξ computed in Step 2'a cover all possible extensions φ up to equivalence. \square

We now explain in a bit more detail Steps 2'a and 3'a—in these steps, we need to understand how $\text{Gal}(L | F)$ restricts to $\text{Gal}(K | F)$ via its permutation representation. The simplest thing to do is just to ignore the conditions on ξ , i.e., in Step 2'a allow all outer automorphisms and in Step 3'a take any isomorphism of groups; *a fortiori*, we will still encounter every one satisfying the extra constraint. To nail it down precisely, we compute the group $\text{Aut}(L_0 | F)$ of F -automorphisms of the field L_0 , for each automorphism τ of order 2 compute the fixed field, until we find a field isomorphic to K_0 ; then $\text{Gal}(K | F)$ is the stabilizer of $\{\beta, \tau(\beta)\}$, and so we can look up the indices of these roots in the permutation representation of $\text{Gal}(L | F)$.

In the above, we may also use $\text{Lie}^0(G)$ in place of $\text{Lie}(G)$ if we are also given $\det \rho_1 = \det \rho_2$, by the discussion at the end of Section 2.3.

3.3. Computing conjugacy classes, in stages. We now discuss Step 4 of Algorithm 2.4.1, where we are given (L, φ) and we are asked to find a witness prime. In theory, to accomplish this task we compute the conjugacy class of Frob_p in $\text{Gal}(L | K)$ using an algorithm of Dokchitser and Dokchitser [2013] and then calculate $\text{utr } \varphi(\sigma)$ for any σ in this conjugacy class.

In practice, because of the enormity of the computation, we may not want to spend time computing the conjugacy class if we can get away with less. In particular, we would like to minimize the amount of work done per field. So we now describe in stages ways to find obstructing primes; each stage gives correct output, but in refining the previous stage we may be able to find smaller primes. Each of these stages involves a precomputation step that only depends of the group-theoretic data.

In Step 2' above, we enumerate subgroups E and identify an exact core-free subgroup D . We identify E with the permutation representation on the cosets E/D .

In Step 3' above, we see the extension $L \supseteq K \supseteq F$ via a core-free extension $L_0 \supseteq K_0 \supseteq F$, and these fields are encoded by minimal polynomials of primitive elements. We may compute $\text{Gal}(L | F)$ as a permutation group with respect to some numbering of the roots, and then insist that the isomorphism $\varphi_0 : \text{Gal}(L | F) \xrightarrow{\sim} E$ computed in Step 3'a is an isomorphism of permutation representations.

For $p \notin S$, for the conjugacy class Frob_p , the cycle type $c(\text{Frob}_p, L_0)$ can be computed very quickly by factoring the minimal polynomial of L_0 modulo a power p^k where it is separable (often but not always $k = 1$ suffices). This cycle type may not uniquely identify the conjugacy class, but we can try to find a cycle type which is *guaranteed* to be obstructing as follows.

4'. Perform the following steps:

- a. For each group E computed in Step 2' with core-free subgroup D , identify E with the permutation representation on the cosets E/D . For each ξ computed in Step 2'a for E , compute the set of cycle types

$$\text{Obc}(E, \xi) := \{c(\xi(\gamma)) : \gamma \in E \text{ and } \text{utr } \gamma \not\equiv 0 \pmod{\ell}\} \setminus \{c(\xi(\gamma)) : \gamma \in E \text{ and } \text{utr } \gamma \equiv 0 \pmod{\ell}\}.$$

- b. For each field (L, φ) , with L encoded by the core-free subfield L_0 and $\varphi \leftrightarrow \xi$ as computed in Step 3'b find a prime \mathfrak{p} such that $c(\text{Frob}_{\mathfrak{p}}, L_0) \in \text{Obc}(E, \xi)$.

In computing $\text{Obc}(E, \xi)$, of course it suffices to restrict to γ in a set of conjugacy classes for E .

Step 4' gives correct output because the set of cycle types in $\text{Obc}(E, \xi)$ are precisely those for which every conjugacy class in E with the given cycle type is obstructing. It is the simplest version, and it is the quickest to compute provided that $\text{Obc}(E, \xi)$ is nonempty.

Remark 3.3.1. In Step 4'a, there may be a cycle type which arises in two ways, from $\gamma, \gamma' \in E$, with $\text{utr } \gamma \not\equiv 0 \pmod{\ell}$ and $\text{utr } \gamma' \equiv 0 \pmod{\ell}$; such a cycle type is not guaranteed to be obstructing.

Remark 3.3.2. In a situation where there are many outer automorphisms ξ to consider, it may be more efficient (but give potentially larger primes and possibly fail more often) to work with the set

$$\text{Obc}(E) := \bigcap_{\xi} \text{Obc}(E, \xi) \tag{3.3.3}$$

consisting of cycle types with the property that every conjugacy class in E under every outer automorphism ξ is obstructing. In this setting, in Step 4'b, we can loop over just the fields L and look for \mathfrak{p} with $c(\text{Frob}_{\mathfrak{p}}) \in \text{Obc}(E)$.

In the next stage, we seek to combine also cycle type information from $\text{Gal}(K | F)$, arising as a permutation group from the field K_0 . Via the isomorphism $\varphi : \text{Gal}(L | F) \xrightarrow{\sim} E$ and the construction of the core-free extension, as a permutation group $\text{Gal}(L | F)$ is isomorphic to the permutation representation of E on the cosets of D . (The numbering might be different, but there is a renumbering for which the representations are equal.) In the same way, the group $\text{Gal}(K | F)$ is isomorphic as a permutation group to the permutation representation of $\pi(E) = G$ on the cosets of the subgroup $\pi(D) = H$, where $\pi : E \rightarrow G$ is the projection. So we have the following second stage.

4''. Perform the following steps:

- a. For each group E computed in Step 2' and each ξ computed in Step 2'a for E , compute the set of pairs of cycle types

$$\text{Obc}(E, G, \xi) := \{(c(\xi(\gamma)), c(\pi(\gamma))) : \gamma \in E \text{ and } \text{utr } \gamma \not\equiv 0 \pmod{\ell}\} \setminus \{(c(\xi(\gamma)), c(\pi(\gamma))) : \gamma \in E \text{ and } \text{utr } \gamma \equiv 0 \pmod{\ell}\}.$$

- b. For each field (L, φ) , with L encoded by L_0 and $\varphi \leftrightarrow \xi$, find a prime \mathfrak{p} such that

$$(c(\text{Frob}_{\mathfrak{p}}, L_0), c(\text{Frob}_{\mathfrak{p}}, K_0)) \in \text{Obc}(E, G, \xi).$$

Step 4'' works for the same reason as in Step 4': the cycle type pairs in $\text{Obc}(E, G, \xi)$ are precisely those for which every conjugacy class in E with the given pair of cycle types is obstructing. The precomputation is a bit more involved in this case, but the check for each field is still extremely fast.

Remark 3.3.4. Instead of the cycle type, a weaker alternative to Step 4'' would be to record the order of $\text{Frob}_{\mathfrak{p}} \in \text{Gal}(K | F)$.

Remark 3.3.5. Assuming that $\text{tr } \bar{\rho}(\text{Frob}_{\mathfrak{p}})$ can be computed efficiently, one additional piece of data that may be appended to the pair of cycle types is $\text{tr } \bar{\rho}(\gamma)$.

Remark 3.3.6. If L arises from several different choices of core-free subgroup, then these subgroups give different (but conjugate) fields L_0 . Because we are not directly accessing the conjugacy class above, but only cycle type information, it is possible that replacing L_0 by a conjugate field will give smaller witnesses. In other words, in Step 4'b or 4''b above, we could loop over the core-free subgroups D and take the smallest witness among them.

Finally, we may go all the way and compute conjugacy classes. Write $[\gamma]_E$ for the conjugacy class of a group element $\gamma \in E$.

4'''. Perform the following steps:

- a. For each group E computed in Step 2' and each ξ computed in Step 2'a for E , compute the set of obstructing conjugacy classes

$$\text{Ob}(E, \xi) := \{[\gamma]_E : \gamma \in E \text{ and } \text{utr } \gamma \not\equiv 0 \pmod{\ell}\}$$

- b. For each field (L, φ) , with L encoded by L_0 and $\varphi \leftrightarrow \xi$, find a prime \mathfrak{p} such that $\text{Frob}_{\mathfrak{p}} \in \text{Ob}(E, G, \xi)$.

We now explain some examples in detail which show the difference between these stages.

Example 3.3.7. Anticipating one of our three core cases, we consider $G = \text{GSp}_4$ and $\ell = 2$ over $F = \mathbb{Q}$. (The reader may wish to skip ahead and read Sections 4–5 to read the details of the setup, but this example is still reasonably self-contained.) We consider the case of a residual representation with image $G = S_5(b) \leq \text{GSp}_4(\mathbb{F}_2)$ (see (5.1.8)), and then a subgroup $E \leq \text{sp}_4 \rtimes G$ with $\dim_{\mathbb{F}_2} V = 10$. We find a core-free subgroup D where $\#H = 10$ and $[V : W] = 2$.

We compute in Step 2'a that we need to consider 8 automorphisms ξ , giving rise to 8 homomorphisms φ . With respect to one such ξ , we find that there are 48 conjugacy classes that are obstructing. Among these, computing as in Step 4'a, we find that 17 are recognized by their L_0 -cycle type:

$$\text{Obc}(E; \xi) = \{3^6 2^1, 4^1 2^4 1^8, 4^1 2^5 1^6, 4^3 1^8, 4^3 2^1 1^6, 6^1 3^4 2^1, 8^1 4^2 2^2, 8^1 4^3, 10^2, 12^1 3^2 2^1, 12^1 6^1 2^1\}. \quad (3.3.8)$$

If instead we call Step 4''a, we find that $35 = \#\text{Obc}(E, G, \xi)$ are recognized by the pair of L_0, K_0 -cycle types (and 22 recognized by L_0 -cycle type and K_0 -order). This leaves 13 conjugacy classes that cannot be recognized purely by cycle type considerations, for which Step 4''' would be required.

For the other choices of ξ , we obtain similar numbers but different cycle types. If we restrict to just L_0 -cycle types that work for *all* such as in Remark 3.3.2, we are reduced to a set of 8:

$$\text{Obc}(E) = \{4^1 2^4 1^8, 4^1 2^5 1^6, 4^3 1^8, 4^3 2^1 1^6, 6^1 3^4 2^1, 8^1 4^2 2^2, 8^1 4^3, 10^2\}. \tag{3.3.9}$$

To see how this plays out with respect to the sizes of primes, we work with the field K arising as the Galois closure of $K_0 = K^H$ defined by a root of the polynomial

$$x^{10} + 3x^9 + x^8 - 10x^7 - 17x^6 - 7x^5 + 11x^4 + 18x^3 + 13x^2 + 5x + 1$$

and similarly $L_0 = L^D$ by a root of

$$x^{20} + 3x^{18} + 5x^{16} + 2x^{14} - 5x^{12} - 13x^{10} - 13x^8 - 6x^6 + x^4 + x^2 - 1.$$

If we restrict to the cycle types in (3.3.8) (or (3.3.9)), we obtain the multiset of witnesses $\{5, 5, 5, 5, 23, 23, 29, 29\}$. If we work with $\text{Obc}(E, G, \xi)$, we find $\{5, 5, 5, 5, 19, 19, 23, 23\}$ instead; the difference is two cases where the witness $p = 29$ is replaced by $p = 19$, so we dig a bit deeper into one of these two cases.

In L_0 , the factorization pattern of 19 is $6^2 3^2 2^1$. But apparently we cannot be guaranteed to have $\text{utr}(\text{Frob}_p) \equiv 1 \pmod{2}$ just looking at cycle type. Indeed, there are three conjugacy classes with this cycle type: one of order 1280 and two of order 2560, represented by the permutations

$$\begin{aligned} &(1\ 9\ 18)(2\ 15\ 6\ 12\ 5\ 16)(3\ 20\ 7\ 13\ 10\ 17)(4\ 14)(8\ 11\ 19), \\ &(1\ 19\ 8\ 11\ 9\ 18)(2\ 15\ 6\ 12\ 5\ 16)(3\ 20\ 17)(4\ 14)(7\ 13\ 10), \\ &(1\ 10\ 2\ 3\ 8\ 4)(5\ 9\ 6)(7\ 17)(11\ 20\ 12\ 13\ 18\ 14)(15\ 19\ 16) \end{aligned}$$

in S_{20} mapping respectively to the matrices

$$\begin{pmatrix} 1 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 0 & 1 & 0 & 1 & 0 & 1 & 1 \\ 1 & 0 & 0 & 0 & 1 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \end{pmatrix}.$$

So precisely the first two conjugacy classes have upper trace 1 and are obstructing, whereas the third has upper trace 0 and is not obstructing. So by cycle types in L_0 alone, indeed, we cannot proceed.

But we recover using the K_0 -cycle type. For the obstructing classes, the cycle type in the permutation representation of G is $3^3 1^1$, whereas for the nonobstructing class the cycle type is $6^1 3^1 1^1$. We compute that the factorization pattern for 19 in K_0 is type $3^3 1^1$, which means 19 belongs to an obstructing class. If we go all the way to the end, we can compute that the conjugacy class of Frob_{19} in fact belongs to the second case.

4. Abelian surfaces, paramodular forms, and Galois representations

We pause now to set up notation and input from the theory of abelian surfaces, paramodular forms, and Galois representations in our case of interest.

4.1. Galois representations from abelian surfaces. Let A be a polarized abelian variety over \mathbb{Q} . For example, if X is a nice (smooth, projective, geometrically integral) genus g curve over \mathbb{Q} , then its Jacobian $\text{Jac } X$ with its canonical principal polarization is a principally polarized abelian variety over \mathbb{Q} of dimension g . Let $N := \text{cond}(A)$ be the conductor of A . We say A is *typical* if $\text{End}(A^{\text{al}}) = \mathbb{Z}$, where $A^{\text{al}} := A_{\mathbb{Q}^{\text{al}}}$ is the base change of A to \mathbb{Q}^{al} .

Lemma 4.1.1. *Let A be a simple, semistable abelian surface over \mathbb{Q} with nonsquare conductor. Then A is typical.*

Proof. By Albert’s classification, either $\text{End}(A) = \mathbb{Z}$ or $\text{End}(A)$ is an order in a quadratic field. In the latter case, $\text{cond}(A)$ is a square by the conductor formula (see [Brumer and Kramer 2014, Lemma 3.2.9]), a contradiction. Therefore $\text{End}(A) = \mathbb{Z}$. Since A is semistable, all endomorphisms of A^{al} are defined over \mathbb{Q} by a result of Ribet [1975, Corollary 1.4]. Thus $\text{End}(A^{\text{al}}) = \text{End}(A) = \mathbb{Z}$, and A is typical. \square

Lemma 4.1.2. *An abelian surface over \mathbb{Q} of prime conductor is typical.*

Proof. If A is not simple over \mathbb{Q} , then we have any isogeny $A \sim A_1 \times A_2$ over \mathbb{Q} to the product of abelian varieties A_1, A_2 over \mathbb{Q} , and $\text{cond}(A) = \text{cond}(A_1) \text{cond}(A_2)$. But since A is prime, without loss of generality $\text{cond}(A_1) = 1$, contradicting the result of Fontaine [1985] that there is no abelian variety over \mathbb{Q} with everywhere good reduction. Therefore A is simple over \mathbb{Q} . Since $N = \text{cond}(A)$ is prime, A is semistable at N , and the result then follows from Lemma 4.1.1. \square

From now on, suppose that $g = 2$ and A is a polarized abelian surface over \mathbb{Q} . Let ℓ be a prime with $\ell \nmid N$ and ℓ coprime to the degree of the polarization on A . Let S be a finite set of places of \mathbb{Q} containing ℓ, ∞ and the primes of bad reduction of A . Let

$$\chi_\ell : \text{Gal}_{\mathbb{Q}, S} \rightarrow \mathbb{Z}_\ell^\times$$

denote the ℓ -adic cyclotomic character, so that $\chi_\ell(\text{Frob}_p) = p$. Then the action of $\text{Gal}_{\mathbb{Q}}$ on the ℓ -adic Tate module

$$T_\ell(A) := \varprojlim_n A[\ell^n] \simeq H_{\text{ét}}^1(A, \mathbb{Z}_\ell) \simeq \mathbb{Z}_\ell^4$$

(where $A[\ell^n]$ denotes the ℓ^n -torsion of A) provides a continuous Galois representation

$$\rho_{A, \ell} : \text{Gal}_{\mathbb{Q}, S} \rightarrow \text{GSp}_4(\mathbb{Z}_\ell) \tag{4.1.3}$$

with determinant χ_ℓ^2 and similitude character χ_ℓ that is unramified outside ℓN . We may reduce the representation (4.1.3) modulo ℓ to obtain a residual representation

$$\bar{\rho}_{A, \ell} : \text{Gal}_{\mathbb{Q}, S} \rightarrow \text{GSp}_4(\mathbb{F}_\ell),$$

which can be concretely understood via the Galois action on the field $\mathbb{Q}(A[\ell])$.

For a prime $p \neq \ell$, slightly more generally we define

$$L_p(A, T) := \det(1 - T \text{Frob}_p^* | H_{\text{et}}^1(A^{\text{al}}, \mathbb{Q}_\ell)^{I_p}) \tag{4.1.4}$$

where Frob_p^* is the geometric Frobenius automorphism, $I_p \leq \text{Gal}_{\mathbb{Q}, S}$ is an inertia group at p , and the definition is independent of the auxiliary prime $\ell \neq p$ (by the semistable reduction theorem of Grothendieck [SGA 7₁ 1972, Expose IX, Theoreme 4.3(b)]). In particular, when $p \nmid \ell N$, we have

$$\det(1 - \rho_{A, \ell}(\text{Frob}_p)T) = L_p(A, T) = 1 - a_p T + b_{p^2} T^2 - p a_p T^3 + p^2 T^4 \in 1 + T\mathbb{Z}[T]. \tag{4.1.5}$$

Moreover, if $A = \text{Jac } X$ and p does not divide the minimal discriminant Δ of X , then

$$Z(X \bmod p, T) := \exp\left(\sum_{r=1}^{\infty} \#X(\mathbb{F}_{p^r}) \frac{T^r}{r}\right) = \frac{L_p(A, T)}{(1-T)(1-pT)}$$

so the polynomials $L_p(A, T)$ may be efficiently computed by counting points on X over finite fields. We define

$$L(A, s) := \prod_p L_p(A, p^{-s})^{-1}; \tag{4.1.6}$$

this series converges for $s \in \mathbb{C}$ in a right half-plane.

4.2. Paramodular forms. We follow Freitag [1983] for the theory of Siegel modular forms. Let $\mathcal{H}_2 \subset M_2(\mathbb{C})$ be the Siegel upper half-space. For $M = \begin{pmatrix} A & B \\ C & D \end{pmatrix} \in \text{GSp}_4^+(\mathbb{R})$, $J = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$ as usual, and T the transpose, we have $M^T J M = \mu J$ with $\mu = \det(M)^{1/2} > 0$ the similitude factor.

For a holomorphic function $f : \mathcal{H}_2 \rightarrow \mathbb{C}$ and $M \in \text{GSp}_4^+(\mathbb{R})$ and $k \in \mathbb{Z}_{\geq 0}$, we define the classical slash

$$(f|_k M)(Z) := \mu^{2k-3} \det(CZ + D)^{-k} f((AZ + B)(CZ + D)^{-1}). \tag{4.2.1}$$

Let $\Gamma \leq \text{Sp}_4(\mathbb{R})$ be a subgroup commensurable with $\text{Sp}_4(\mathbb{Z})$. We denote by

$$M_k(\Gamma) := \{f : \mathcal{H}_2 \rightarrow \mathbb{C} : (f|_k \gamma)(Z) = f(Z) \text{ for all } \gamma \in \Gamma\}$$

the \mathbb{C} -vector space of Siegel modular forms with respect to Γ , and $S_k(\Gamma) \subseteq M_k(\Gamma)$ the subspace of forms vanishing at the cusps of Γ , called the space of cuspforms.

To each double coset $\Gamma M \Gamma$ with $M \in \text{GSp}_4^+(\mathbb{Q})$, we define the *Hecke operator*

$$T(\Gamma M \Gamma) : M_k(\Gamma) \rightarrow M_k(\Gamma) \tag{4.2.2}$$

as follows: from a decomposition $\Gamma M \Gamma = \bigsqcup_j \Gamma M_j$ of the double coset into disjoint single cosets, we define $f|_k T(\Gamma M \Gamma) = \sum_j f|_k M_j$. The action is well-defined, depending only on the double coset, and $T(\Gamma M \Gamma)$ maps $S_k(\Gamma)$ to $S_k(\Gamma)$.

Let $N \in \mathbb{Z}_{\geq 1}$. The *paramodular group* $K(N)$ of level N in degree two is defined by

$$K(N) := \begin{pmatrix} \mathbb{Z} & N\mathbb{Z} & \mathbb{Z} & \mathbb{Z} \\ \mathbb{Z} & \mathbb{Z} & \mathbb{Z} & N^{-1}\mathbb{Z} \\ \mathbb{Z} & N\mathbb{Z} & \mathbb{Z} & \mathbb{Z} \\ N\mathbb{Z} & N\mathbb{Z} & N\mathbb{Z} & \mathbb{Z} \end{pmatrix} \cap \mathrm{Sp}_4(\mathbb{Q}). \tag{4.2.3}$$

The paramodular group $K(N)$ has a normalizing *paramodular Fricke involution*, $\mu_N \in \mathrm{Sp}_4(\mathbb{R})$, given by

$$\mu_N = \begin{pmatrix} (F_N^{-1})^\top & 0 \\ 0 & F_N \end{pmatrix}$$

where $F_N = \frac{1}{\sqrt{N}} \begin{pmatrix} 0 & -N \\ 1 & 0 \end{pmatrix}$ is the Fricke involution for $\Gamma_0(N)$. Consequently, for all k we may decompose

$$M_k(K(N)) = M_k(K(N))^+ \oplus M_k(K(N))^- \tag{4.2.4}$$

into plus and minus μ_N -eigenspaces.

Write $e(z) = \exp(2\pi\sqrt{-1}z)$ for $z \in \mathbb{C}$. The Fourier expansion of $f \in M_k(K(N))$ is

$$f(Z) = \sum_{T \geq 0} a(T; f) e(\mathrm{tr}(TZ)) \tag{4.2.5}$$

for $Z \in \mathcal{H}_2$ and the sum over semidefinite matrices

$$T = \begin{pmatrix} n & r/2 \\ r/2 & Nm \end{pmatrix} \in \mathbf{M}_2^{\mathrm{sym}}(\mathbb{Q})_{\geq 0} \quad \text{with } n, r, m \in \mathbb{Z}.$$

For a subring $R \subseteq \mathbb{C}$, we denote by

$$M_k(K(N), R) := \{f \in M_k(K(N)) : a(T; f) \in R \text{ for all } T \geq 0\} \tag{4.2.6}$$

the R -module of paramodular forms whose Fourier coefficients all lie in R , and similarly we write $S_k(K(N), R)$ for cusp forms and $S_k(K(N), R)^\pm$ for the eigenspaces under the Fricke involution. The ring of paramodular forms with coefficients in R

$$M(K(N), R) := \bigoplus_{k=0}^{\infty} M_k(K(N), R)$$

is a graded R -algebra.

For a prime $p \nmid N$, the first (more familiar) Hecke operator we will use is

$$T(p) := T(K(N) \mathrm{diag}(1, 1, p, p) K(N)) \tag{4.2.7}$$

whose decomposition into left cosets is given by

$$\begin{aligned}
 &K(N) \operatorname{diag}(1, 1, p, p)K(N) \\
 &= K(N) \begin{pmatrix} p & 0 & 0 & 0 \\ 0 & p & 0 & 0 \\ & & 1 & 0 \\ & & & 0 & 1 \end{pmatrix} + \sum_{i \bmod p} K(N) \begin{pmatrix} 1 & 0 & i & 0 \\ 0 & p & 0 & 0 \\ & & p & 0 \\ & & & 0 & 1 \end{pmatrix} \\
 &\quad + \sum_{i, j \bmod p} K(N) \begin{pmatrix} p & 0 & 0 & 0 \\ i & 1 & 0 & j \\ & & 1 & -i \\ & & 0 & p \end{pmatrix} + \sum_{i, j, k \bmod p} K(N) \begin{pmatrix} 1 & 0 & i & j \\ 0 & 1 & j & k \\ & & p & 0 \\ & & & 0 & p \end{pmatrix} \tag{4.2.8}
 \end{aligned}$$

with indices taken over residue classes modulo p . Writing $T[u] = u^T T u$ for $T, u \in M_2(\mathbb{Q})$, the action of $T(p)$ on Fourier coefficients $a(T; f)$ is given by

$$a(T; f|_k T(p)) = a(pT; f) + p^{k-2} \sum_{j \bmod p} a\left(\frac{1}{p} T \begin{bmatrix} 1 & 0 \\ j & p \end{bmatrix}; f\right) + p^{k-2} a\left(\frac{1}{p} T \begin{bmatrix} p & 0 \\ 0 & 1 \end{bmatrix}; f\right) + p^{2k-3} a\left(\frac{1}{p} T; f\right). \tag{4.2.9}$$

Hence for $k \geq 2$, the Hecke operator $T(p)$ stabilizes $S_k(K(N), R)$. In particular, taking $R = \mathbb{Z}$ we see that if f has integral Fourier coefficients, then $f|_k T(p)$ has integral Fourier coefficients for $k \geq 2$.

We will also make use of another, perhaps less familiar, Hecke operator. For $K(N)$ and a prime $p \nmid N$, we define

$$T_1(p^2) = T(K(N) \operatorname{diag}(1, p, p^2, p)K(N)). \tag{4.2.10}$$

Lemma 4.2.11. *The coset decomposition for $T_1(p^2)$ is given by*

$$\begin{aligned}
 &K(N) \operatorname{diag}(1, p, p^2, p)K(N) \\
 &= K(N) \begin{pmatrix} p & 0 & 0 & 0 \\ 0 & p^2 & 0 & 0 \\ & & p & 0 \\ & & & 0 & 1 \end{pmatrix} + \sum_{i \bmod p} K(N) \begin{pmatrix} p^2 & 0 & 0 & 0 \\ pi & p & 0 & 0 \\ & & 1 & -i \\ & & & 0 & p \end{pmatrix} \\
 &\quad + \sum_{i \not\equiv 0 \pmod p} K(N) \begin{pmatrix} p & 0 & i & 0 \\ 0 & p & 0 & 0 \\ & & p & 0 \\ & & & 0 & p \end{pmatrix} + \sum_{\substack{i \bmod p, \\ j \not\equiv 0 \pmod p}} K(N) \begin{pmatrix} p & 0 & i^2 j & ij \\ 0 & p & ij & j \\ & & p & 0 \\ & & & 0 & p \end{pmatrix} \\
 &\quad + \sum_{\substack{i \bmod p, \\ j \bmod p^2}} K(N) \begin{pmatrix} 1 & 0 & j & i \\ 0 & p & pi & 0 \\ & & p^2 & 0 \\ & & & 0 & p \end{pmatrix} + \sum_{\substack{i, j \bmod p, \\ k \bmod p^2}} K(N) \begin{pmatrix} p & 0 & 0 & pj \\ i & 1 & j & k \\ & & p & -pi \\ & & & 0 & p^2 \end{pmatrix} \tag{4.2.12}
 \end{aligned}$$

Proof. The cosets are from [Roberts and Schmidt 2007, (6.6)] after swapping rows one and two and columns one and two, applying an inverse, and multiplying by the similitude p^2 . □

Define the indicator function $\mathbf{1}(p \mid y)$ by 1 if $p \mid y$ and by 0 if $p \nmid y$. Then the action of $T_1(p^2)$ on the Fourier coefficients is:

$$\begin{aligned} a(T; f|_k T_1(p^2)) &= p^{k-3} \sum_{x \bmod p} a(T \begin{bmatrix} 1 & 0 \\ x & p \end{bmatrix}; f) + p^{k-3} a(T \begin{bmatrix} p & 0 \\ 0 & 1 \end{bmatrix}; f) \\ &+ p^{3k-6} \sum_{j \bmod p} a(\frac{1}{p^2} T \begin{bmatrix} 1 & 0 \\ j & p \end{bmatrix}; f) + p^{3k-6} a(\frac{1}{p^2} T \begin{bmatrix} p & 0 \\ 0 & 1 \end{bmatrix}; f) \\ &+ p^{2k-6} (p \mathbf{1}(p \mid T \begin{bmatrix} 1 & \\ & 0 \end{bmatrix}) - 1) a(T; f) \\ &+ p^{2k-6} \sum_{\lambda \bmod p} (p \mathbf{1}(p \mid T \begin{bmatrix} \lambda & \\ & 1 \end{bmatrix}) - 1) a(T; f). \end{aligned} \tag{4.2.13}$$

Hence for $k \geq 3$, the Hecke operator $T_1(p^2)$ stabilizes $S_k(K(N), R)$. In particular, if f has integral Fourier coefficients, then $f|_k T_1(p^2)$ has integral Fourier coefficients for $k \geq 3$. However, for $k = 2$, we only know that $p^2 f|_k T_1(p^2)$ is integral when f is (and there are examples where $f|_2 T_1(p^2)$ has p^2 in the denominator of some Fourier coefficients).

Summarizing the above, we have:

$$\begin{aligned} T(p) &= T(K(N) \operatorname{diag}(1, 1, p, p) K(N)); & \deg T(p) &= (1+p)(1+p^2); \\ T_1(p^2) &= T(K(N) \operatorname{diag}(1, p, p^2, p) K(N)); & \deg T_1(p^2) &= (1+p)(1+p^2). \end{aligned} \tag{4.2.14}$$

We define two new operators:

$$\begin{aligned} T_2(p^2) &:= T(K(N) \operatorname{diag}(p, p, p, p) K(N)) = p^{2k-6} \operatorname{id} \\ B(p^2) &:= p(T_1(p^2) + (1+p^2)T_2(p^2)) \end{aligned} \tag{4.2.15}$$

If f is an eigenform of weight k for the operators $T(p)$ and $T_1(p^2)$, with corresponding eigenvalues $a_p(f), a_{1,p^2}(f) \in \mathbb{C}$, then f is an eigenform for the operator $B(p^2)$ with eigenvalue

$$b_{p^2}(f) := p a_{1,p^2}(f) + p^{2k-5} (1+p^2). \tag{4.2.16}$$

Lemma 4.2.17. *If $k = 2$ and f has integral Fourier coefficients, then $b_{p^2}(f) \in \mathbb{Z}$.*

Proof. We have observed that $p^2 a_{1,p^2}(f) \in \mathbb{Z}$. From (4.2.13), we observe the congruence

$$p^2 (f|_2 T_1(p^2)) = p^2 a_{1,p^2}(f) f \equiv -f \pmod{p}.$$

so $p \mid (p^2 a_{1,p^2}(f) + 1)$. Therefore

$$b_{p^2}(f) = p a_{1,p^2}(f) + (1+p^2)/p = (p^2 a_{1,p^2}(f) + 1)/p + p \in \mathbb{Z}. \quad \square$$

Following Roberts and Schmidt [2006; 2007], to f we then assign the *spinor Euler factor* at $p \nmid N$ in the arithmetic normalization by

$$Q_p(f, T) := 1 - a_p(f)T + b_{p^2}(f)T^2 - p^{2k-3} a_p(f)T^3 + p^{4k-6} T^4 \in 1 + T\mathbb{C}[T]. \tag{4.2.18}$$

We will also call $Q_p(f, T)$ the *spinor Hecke polynomial* at p . If f has integral Fourier coefficients, then by Lemma 4.2.17 we have $Q_p(f, T) \in 1 + T\mathbb{Z}[T]$.

4.3. Galois representations from Siegel modular forms. We now seek to match the Galois representation coming from an abelian surface with one coming from an automorphic form. In this section, we explain the provenance of the latter.

We follow the presentation of Schmidt [2018] for the association of an automorphic representation to a paramodular eigenform. Let $\Gamma \leq \mathrm{GSp}_4(\mathbb{Q})^+$ be a subgroup commensurable with $\mathrm{Sp}_4(\mathbb{Z})$ and let $f \in S_k(\Gamma)$ be a cuspidal eigenform at all but finitely many places. In general, the representation π_f generated by the adelization of f may be reducible and hence not an automorphic representation at all. It is still possible however, to associate a global Arthur parameter for $\mathrm{GSp}_4(\mathbb{A})$ to f as follows. Because f is cuspidal, the representation π_f decomposes as the direct sum of a finite number of automorphic representations, and each summand has the same global Arthur parameter among one of six types: the general type (**G**), the Yoshida type (**Y**), the finite type (**F**), or types (**P**), (**Q**) or (**B**) named after parabolic subgroups. Thus we may associate a global Arthur parameter directly to a paramodular eigenform f . The only type of global Arthur parameter that concerns us here is type (**G**) given by the formal tensor $\mu \boxtimes 1$, where μ is a cuspidal, self-dual, symplectic, unitary, automorphic representation of $\mathrm{GL}_4(\mathbb{A})$ and 1 is the trivial representation of $\mathrm{SU}_2(\mathbb{A})$.

Remark 4.3.1. One can consider the eigenforms of type (**G**) to be those that *genuinely* belong on GSp_4 .

Second, when f is of type (**G**) or (**Y**), the associated representation π_f is irreducible and f is necessarily an eigenform at all good primes. Third, the type of f may be determined by checking *one* Euler factor at a good prime. We state the paramodular case $\Gamma = K(N)$.

Proposition 4.3.2 (Schmidt). *Let $f \in S_k(K(N))$ be a cuspidal eigenform for all primes $p \nmid N$. Let $p \nmid N$ be prime and let $Q_p(f, T)$ be the Hecke polynomial of f at p defined in (4.2.18) in the arithmetic normalization. Then f is of type (**G**) if and only if all reciprocal roots of $Q_p(f, T)$ have complex absolute value $p^{k-3/2}$.*

Proof. Converting from analytic to arithmetic normalization, by [Schmidt 2018, Proposition 2.1] the stated local factor condition implies that f is of type (**G**) or (**Y**), but paramodular cusp forms cannot be type (**Y**) also by [Schmidt 2018, Lemma 2.5]. \square

Fourth, continuing in the paramodular case $\Gamma = K(N)$, the global conductor of π_f divides N , and is equal to N if and only if f is a newform. Finally, if f is a newform — see [Roberts and Schmidt 2006] for the global newform theory of paramodular forms — then f is a Hecke eigenform at all primes and for all paramodular Atkin–Lehner involutions.

We need one final bit of notation, concerning archimedean L -parameters. The real Weil group is $W(\mathbb{R}) = \mathbb{C}^\times \cup \mathbb{C}^\times j$, with $j^2 = -1$ and $jzj^{-1} = \bar{z}$ for $z \in \mathbb{C}^\times$. For $w, m_1, m_2 \in \mathbb{Z}$ with $m_1 > m_2 \geq 0$ and $w + 1 \equiv m_1 + m_2 \pmod{2}$, we define the *archimedean L -parameter* $\phi(w, m_1, m_2) : W(\mathbb{R}) \rightarrow \mathrm{GSp}_4(\mathbb{R})$

by sending $z \in \mathbb{C}^\times$ to the diagonal matrix

$$|z|^{-w} \operatorname{diag} \left(\left(\frac{z}{\bar{z}} \right)^{(m_1+m_2)/2}, \left(\frac{z}{\bar{z}} \right)^{(m_1-m_2)/2}, \left(\frac{z}{\bar{z}} \right)^{(m_2-m_1)/2}, \left(\frac{z}{\bar{z}} \right)^{-(m_1+m_2)/2} \right) \tag{4.3.3}$$

and j to the antidiagonal matrix $\operatorname{antidiag}((-1)^{w+1}, (-1)^{w+1}, 1, 1)$. The archimedean L -packet of $\operatorname{GSp}_4(\mathbb{R})$ corresponding to $\phi(w, m_1, m_2)$ has two elements, one holomorphic and one generic: for $m_2 > 0$ these are both discrete series representations, whereas for $m_2 = 0$ they are limits of discrete series.

We are now ready to associate a Galois representation to a paramodular eigenform of type **(G)**.

Theorem 4.3.4 (Taylor, Laumon, Weissauer, Schmidt, and Mok). *Let $f \in S_k(K(N))$ be a Siegel paramodular newform of weight $k \geq 2$ and level N . Suppose that f is of type **(G)**. Then for any prime $\ell \nmid N$, there exists a continuous, semisimple Galois representation*

$$\rho_{f,\ell} : \operatorname{Gal}_{\mathbb{Q}} \rightarrow \operatorname{GSp}_4(\mathbb{Q}_\ell^{\text{al}})$$

with the following properties:

- (i) $\det(\rho_{f,\ell}) = \chi_\ell^{4k-6}$.
- (ii) The similitude character of $\rho_{f,\ell}$ is χ_ℓ^{2k-3} .
- (iii) $\rho_{f,\ell}$ is unramified outside ℓN .
- (iv) $\det(1 - \rho_{f,\ell}(\operatorname{Frob}_p)T) = Q_p(f, T)$ for all $p \nmid \ell N$.
- (v) The local Langlands correspondence holds for all primes $p \neq \ell$, up to semisimplification.

By (v), we mean that the Weil–Deligne representations associated to the restriction of the Galois representation $\rho_{f,\ell}$ to $\operatorname{Gal}(\mathbb{Q}_p^{\text{al}} | \mathbb{Q}_p)$ agrees with that associated to the $\operatorname{GL}_n(\mathbb{Q}_p)$ -representation π_p attached by the local Langlands correspondence up to semisimplification *without* information about the nilpotent operator N : in the notation of Taylor and Yoshida [2007, p. 468] we mean $(V, r, N)^{\text{ss}} = (V, r^{\text{ss}}, 0)$.

Proof. The existence and properties (i)–(ii) follow from the construction and an argument of Taylor [1991, Example 1, §1.3]. Properties (iii) and (iv) are provided by Berger and Klosin [2017, Theorem 8.2] (they claim in the subsequent Remark 8.3 that the result is “well-known”).

We now sketch the construction, and we use the argument of Mok to conclude also property (v). By the discussion above, following Schmidt [2018], we may attach to f a cuspidal automorphic representation Π_f of $\operatorname{GSp}_4(\mathbb{A})$ of type **(G)**. The hypothesis that f is of type **(G)** assures that the automorphic representation Π_f is irreducible. If $k \geq 3$, then the automorphic representation is of cohomological type, and from a geometric construction we obtain a Galois representation $\rho_{f,\ell} : \operatorname{Gal}_{\mathbb{Q}} \rightarrow \operatorname{GSp}_4(E)$ by work of Laumon [2005] and Weissauer [2005, Theorems I and IV], where E is the finite extension of \mathbb{Q}_ℓ containing the Hecke eigenvalues of f (choosing an isomorphism between the algebraic closure of \mathbb{Q} in \mathbb{C} and in $\mathbb{Q}_\ell^{\text{al}}$): one shows that the representation takes values in $\operatorname{GL}_4(E)$ and that it preserves a nondegenerate symplectic bilinear form invariant under $\rho_{f,\ell}(\operatorname{Gal}_{\mathbb{Q}})$ so lands in $\operatorname{GSp}_4(E)$. Thereby, properties (i)–(iv) are verified.

For all $k \geq 2$, with the above conventions (including archimedean L -parameters) we verify that Π_f satisfies the hypotheses of a theorem of Mok [2014, Theorem 4.14]: from this theorem we obtain a unique, continuous semisimple representation $\rho_{f,\ell} : \text{Gal}_{\mathbb{Q}} \rightarrow \text{GL}_4(\mathbb{Q}_{\ell}^{\text{al}})$ where $\mathbb{Q}_{\ell}^{\text{al}}$ is an algebraic closure of \mathbb{Q}_{ℓ} . For $k = 2$, Mok constructs the representation by ℓ -adic deformation using Hida theory from those of Laumon and Weissauer, and so properties (i)–(iv) and the fact that the representation is symplectic continue to hold in the limit; and property (v) is a conclusion of his theorem.

To illustrate this convergence argument, we show that the representation is symplectic. Let $\{f_n\}_n$ be a sequence of Siegel paramodular forms of weights $k_n > 2$ such that f_n converge p -adically to f (for example, multiplying by powers of the Hasse invariant). By the previous paragraph, each f_n is symplectic with representation ρ_n so

$$\bigwedge^2 \rho_n(3 - 2k_n) \simeq \rho_{\text{triv}} \oplus \psi_n \tag{4.3.5}$$

is equivalent to the direct sum of the trivial representation ρ_{triv} of degree 1 and the representation ψ of degree 5 with values in $\text{SO}_5(\mathbb{Q}_{\ell}^{\text{al}})$. The sequence $\text{Tr } \psi_n$ of pseudorepresentations converges to a pseudorepresentation by (4.3.5) and continuity of the trace, and this limit is the trace of a representation ψ . From this identity of traces, we conclude

$$\bigwedge^2 \rho(-1) \simeq \rho_{\text{triv}} \oplus \psi$$

and thus ρ is symplectic with cyclotomic similitude character.

Mok’s theorem relies on work of Arthur in a crucial way. For further attribution and discussion, see [Mok 2014, About the proof, pp. 524ff] and the overview of the method by Jorza [2012, §§1–3]. \square

Let f be as in Theorem 4.3.4, with Galois representation $\rho_{f,\ell} : \text{Gal}_{\mathbb{Q},S} \rightarrow \text{GSp}_4(\mathbb{Q}_{\ell}^{\text{al}})$ where $S := \{p : p \mid N\} \cup \{\ell, \infty\}$. By the Baire category theorem, we may descend the representation to a finite extension $E' \subseteq \mathbb{Q}_{\ell}^{\text{al}}$ of \mathbb{Q}_{ℓ} . Let ℓ' be the prime above ℓ in the valuation ring R' of E' and let k' be the residue field of R' . Choose a stable R' -lattice in the representation space $V' := (E')^4$ and reduce modulo ℓ' ; the semisimplification yields a semisimple residual representation $\bar{\rho}_{f,\ell}^{\text{ss}} : \text{Gal}_{\mathbb{Q},S} \rightarrow \text{GL}_4(k')$, unique up to equivalence.

Applying a recent result of Serre, we now show that the residual representation is symplectic.

Lemma 4.3.6. *The semisimplification $\bar{\rho}_{f,\ell}^{\text{ss}} : \text{Gal}_{\mathbb{Q},S} \rightarrow \text{GL}_4(k')$ is compatible with a nondegenerate alternating form with similitude character $\bar{\chi}_{\ell}^{2k-3}$; in particular, up to equivalence its image lies in $\text{GSp}_4(k')$.*

Proof. We refer to Serre [2018]. Let $\langle \cdot, \cdot \rangle$ be the alternating form on V' with similitude character $\epsilon := \chi_{\ell}^{2k-3}$ provided by Theorem 4.3.4. Then V' is a module over $A' := R'[\text{Gal}_{\mathbb{Q},S}]$ via $\rho_{f,\ell}$ (we suppress this from the notation for convenience); moreover, the map $\sigma^* := \epsilon(\sigma)\sigma^{-1}$ for $\sigma \in \text{Gal}_{\mathbb{Q},S}$ extends by R' -linearity to an involution of A' . Therefore, for all $\sigma \in \text{Gal}_{\mathbb{Q},S}$ and all $x, y \in V'$ we have

$$\langle \sigma x, y \rangle = \langle \sigma x, \sigma \sigma^{-1} y \rangle = \epsilon(\sigma) \langle x, \sigma^{-1} y \rangle = \langle x, \epsilon(\sigma) \sigma^{-1} y \rangle = \langle x, \sigma^* y \rangle. \tag{4.3.7}$$

Extending by R' -linearity, we conclude that (\cdot, \cdot) is compatible with the action of A' [Serre 2018, (5.1.1)].

Let $V_{k'}$ be the k' -vector space underlying the semisimplification $\bar{\rho}_{f,\ell}^{\text{ss}}$. Then Serre proves [2018, Theorem 5.1.4] that there exists a nondegenerate k' -bilinear alternating form on $V_{k'}$ that is compatible with $A'_k := k'[\text{Gal}_{\mathbb{Q},S}]$. Running the equality of (4.3.7) again, we conclude that $V_{k'}$ has similitude character $\bar{\epsilon}$, as claimed.

The final statement holds because up to equivalence by $\text{GL}_4(k')$ we may assume the alternating form is the standard form, so now the image lands in $\text{GSp}_4(k')$, as claimed. \square

Next, we seek descent preserving the symplectic form. Let E be the extension of \mathbb{Q}_ℓ generated by the Hecke eigenvalues of f (with respect to a choice of isomorphism between the algebraic closure of \mathbb{Q} in \mathbb{C} and in $\mathbb{Q}_\ell^{\text{al}}$); then E also contains all coefficients of the Hecke polynomials $Q_p(f, T)$. Let R be the valuation ring of E and let k be its residue field. We have $E \subseteq E'$, and we would like to be able to descend the representation to take values in $\text{GSp}_4(E)$. However, there is a possible obstruction coming from the Brauer group of \mathbb{Q}_ℓ ; such an obstruction arises for example in the Galois representation afforded by a QM abelian fourfold at a prime ℓ dividing the discriminant of the quaternion algebra B , which has image in $\text{GL}_2(B \otimes \mathbb{Q}_\ell)$ and not $\text{GSp}_4(\mathbb{Q}_\ell)$. Under an additional hypothesis, we may ensure descent following Carayol and Serre as follows.

Lemma 4.3.8. *With hypotheses as in Theorem 4.3.4, the following statements hold:*

- (a) *The semisimplified residual representation $\bar{\rho}_{f,\ell}^{\text{ss}}$ descends to*

$$\bar{\rho}_{f,\ell}^{\text{ss}} : \text{Gal}_{\mathbb{Q},S} \rightarrow \text{GSp}_4(k)$$

up to equivalence.

- (b) *If $\bar{\rho}_{f,\ell}^{\text{ss}} = \bar{\rho}_{f,\ell}$ is absolutely irreducible, then $\rho_{f,\ell}$ descends to*

$$\rho_{f,\ell} : \text{Gal}_{\mathbb{Q},S} \rightarrow \text{GSp}_4(E)$$

up to equivalence, where E is the extension of \mathbb{Q}_ℓ generated by the Hecke eigenvalues of f as above.

Proof. We begin with (a). First, a semisimple representation into $\text{GL}_4(k')$ is determined by its traces, and so up to equivalence we may descend $\bar{\rho}_{f,\ell}^{\text{ss}}$ to take values in $\text{GL}_4(k) \subseteq \text{GL}_4(k')$ (for a complete proof, see e.g., [Taylor 1991, Lemma 2, part 2]). The semisimplification $\bar{\rho}_{f,\ell}^{\text{ss}}$ was only well-defined up to equivalence (in $\text{GL}_4(k')$) anyway, so Lemma 4.3.6 still applies and the underlying space $V_k = k^4$ of $\bar{\rho}_{f,\ell}^{\text{ss}}$ has the property that its extension $V_{k'} = (k')^4$ to k' carries an alternating form with k -valued similitude character $\bar{\chi}_\ell^{2k-3}$. The set of such alternating forms with fixed similitude character is defined by linear conditions over k since the image of $\bar{\rho}_{f,\ell}$ belongs to $\text{GL}_4(k)$; therefore, the existence of a form defined on $V_{k'}$ implies the existence of such a form on V_k with the same similitude character. Again up to equivalence, the image of $\bar{\rho}_{f,\ell}^{\text{ss}}$ may be taken to lie in $\text{GSp}_4(k)$.

For statement (b), by a theorem of Carayol [1994, Théorème 2] under the hypothesis that the *residual representation is absolutely irreducible*, the representation $\rho_{f,\ell}$ takes values in $\text{GL}_4(E)$. Again we have a

nondegenerate alternating form compatible with $\text{Gal}_{\mathbb{Q},S}$, and repeating the first part of the argument in the previous paragraph we may assume it takes values in E ; conjugating, we conclude that the image is in $\text{GSp}_4(E)$. \square

Remark 4.3.9. The statement of Theorem 4.3.4 is not the most general statement that could be proven (in several respects), but it is sufficient for our purposes.

Berger and Klosin [2017, Theorem 8.2] attach to any paramodular newform f a Galois representation into $\text{GL}_4(\mathbb{Q}_\ell^{\text{al}})$, not just those of type **(G)**. The remaining types are related to constructions of automorphic representations from those in $\text{GL}_2(\mathbb{A})$, where the local Langlands correspondence is known. We do not know a reference for a complete argument for these remaining cases. In this article, we are only concerned with forms of type **(G)**.

A consequence of Mok’s proof of Theorem 4.3.4(v) is encoded in the following result.

Lemma 4.3.10. *Let K be the fixed field of $\ker \bar{\rho}_{f,\ell}$ and let $\text{cond}(\bar{\rho}_{f,\ell})$ be the Artin conductor of the representation $\bar{\rho}_{f,\ell}$ of $\text{Gal}(K | \mathbb{Q})$. If $p \parallel N$ is odd, then $\text{ord}_p(\text{cond}(\bar{\rho}_{f,\ell})) \leq 1$.*

Proof. The proof of Theorem 4.3.4(v) is only up to semisimplification, so we do not know the complete statement of local Langlands under the patching argument that is employed. However, in specializing the family to the accumulation point f in the family, there is nevertheless an *upper bound* on the level: the representation is necessarily either unramified or is Steinberg with level p , and accordingly the conductor has p -valuation 0 or 1. \square

5. Group theory and Galois theory for $\text{GSp}_4(\mathbb{F}_2)$

In this section, we carry out the needed Galois theory for the group $\text{GSp}_4(\mathbb{F}_2)$. Specifically, we carry out the task outlined in Section 3.2: given $G = \text{img } \bar{\rho} \leq \text{GSp}_4(\mathbb{F}_2)$, and for each obstructing extension φ extending $\bar{\rho}$, we compute an exact core-free subgroup $D \leq E$ (as large as possible) and the list of E -conjugacy classes of elements whose upper trace is nonzero. The arguments provided in this section are done once and for all for the group $\text{GSp}_4(\mathbb{F}_2)$; we apply these to our examples in Section 7.

5.1. Symplectic group as permutation group. We pause for some basic group theory. We have an isomorphism $\iota : S_6 \xrightarrow{\sim} \text{Sp}_4(\mathbb{F}_2)$, where S_6 is the symmetric group on 6 letters, which we make explicit in the following manner. Let $U := \mathbb{F}_2^6$, and equip U with the coordinate action of S_6 and the standard nondegenerate alternating (equivalently, symmetric) bilinear form $\langle x, y \rangle = \sum_{i=1}^6 x_i y_i$ visibly compatible with the S_6 -action. Let $U^0 \subset U$ be the trace 0 hyperplane, let L be the \mathbb{F}_2 -span of $(1, \dots, 1)$, and let $Z := U^0/L$ be the quotient, so $\dim_{\mathbb{F}_2} Z = 4$. Then Z inherits both an action of S_6 and a symplectic pairing, which remains nondegenerate: specifically, the images

$$e_1 := (1, 1, 0, 0, 0, 0), \quad e_2 := (0, 0, 1, 1, 0, 0), \quad e_3 := (0, 0, 0, 1, 1, 0), \quad e_4 := (0, 1, 0, 0, 0, 1) \in Z$$

are a basis for Z in which the Gram matrix of the induced pairing is the antiidentity matrix, so e.g., $\langle e_1, e_4 \rangle = \langle e_2, e_3 \rangle = 1$. (An alternating pairing over \mathbb{F}_2 is symmetric, and we have chosen the standard such form.) We compute that

$$\iota : S_6 \rightarrow \mathrm{Sp}_4(\mathbb{F}_2)$$

$$(1\ 2\ 3\ 4\ 5), (1\ 6) \mapsto \begin{pmatrix} 1 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 1 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix}. \tag{5.1.1}$$

We have

$$\mathrm{Lie}^0(\mathrm{GSp}_4)(\mathbb{F}_2) = \mathrm{sp}_4(\mathbb{F}_2) = \{A \in \mathrm{M}_4(\mathbb{F}_2) : A^\top J + JA = 0\} \simeq \mathbb{F}_2^{10}, \tag{5.1.2}$$

where $J \in \mathrm{M}_4(\mathbb{F}_2)$ is the antiidentity matrix (with 1 along the antidiagonal), and we have an exact sequence

$$1 \rightarrow \mathrm{sp}_4(\mathbb{F}_2) \rightarrow \mathrm{sp}_4(\mathbb{F}_2) \rtimes G \xrightarrow{\pi} G \rightarrow 1 \tag{5.1.3}$$

with $\pi : \mathrm{sp}_4(\mathbb{F}_2) \rtimes G \rightarrow G$ the natural projection map. As in (2.3.9) we identify

$$\mathrm{sp}_4(\mathbb{F}_2) \rtimes G \leq \mathrm{M}_4(\mathbb{F}_2) \rtimes G \hookrightarrow \mathrm{GL}_8(\mathbb{F}_2)$$

$$(a, g) \mapsto \begin{pmatrix} 1 & a \\ 0 & 1 \end{pmatrix} \begin{pmatrix} g & 0 \\ 0 & g \end{pmatrix} = \begin{pmatrix} g & ag \\ 0 & g \end{pmatrix} \tag{5.1.4}$$

The following lemmas follow from straightforward computation.

Lemma 5.1.5. *The group $\mathrm{Sp}_4(\mathbb{F}_2)$ has elements of orders $1, \dots, 6$ with the following possibilities for their characteristic polynomials:*

order	characteristic polynomial	
1, 2, 4	$x^4 + 1$	
3, 6	$x^4 + x^2 + 1$ or $x^4 + x^3 + x + 1$	
5	$x^4 + x^3 + x^2 + x + 1$	

(5.1.6)

There is a unique outer automorphism of S_6 up to inner automorphisms [Howard et al. 2008]; it sends transpositions to products of three transpositions, and interchanges the trace of some order 3 and order 6 elements.

Lemma 5.1.7. *There are, up to inner automorphism, exactly 9 subgroups of $\mathrm{Sp}_4(\mathbb{F}_2) \simeq S_6$ with absolutely irreducible image. They are listed in the following table with a property that determines them uniquely (where “–” indicates there is a unique conjugacy class of subgroup with that order):*

subgroup	order	element orders	distinguishing property
S_6	720	1, . . . , 6	–
A_6	360	1, . . . , 5	–
$S_5(a)$	120	1, . . . , 6	elements of order 3, 6 have trace 0
$S_5(b)$	120	1, . . . , 6	elements of order 3, 6 have trace 1
$S_3 \wr S_2$	72	1, 2, 3, 4, 6	–
$A_5(b)$	60	1, 2, 3, 5	elements of order 3 have trace 1
$C_3^2 \rtimes C_4$	36	1, 2, 3, 4	no elements of order 6
$S_3(a)^2$	36	1, 2, 3, 6	elements of order 6 have trace 0
$C_5 \rtimes C_4$	20	1, 2, 4, 5	–

(5.1.8)

Example 5.1.9. The conjugacy classes of subgroups $S_5(a), S_5(b) \leq S_6$ are exchanged by the outer automorphism of S_6 . For example, under the restriction of (5.1.1), we have

$$\begin{aligned} \iota : S_5(b) &\rightarrow \mathrm{Sp}_4(\mathbb{F}_2) \\ (1\ 2\ 3\ 4\ 5), (1\ 2), (1\ 2\ 3) &\mapsto \begin{pmatrix} 1 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 1 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \end{pmatrix}, \begin{pmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 \end{pmatrix}. \end{aligned} \tag{5.1.10}$$

Another way to distinguish $S_5(a)$ from $S_5(b)$ is that $\iota(S_5(b))$ has transvections while $\iota(S_5(a))$ does not.

Example 5.1.11. There is a subgroup $A_5(a) \leq S_6$ that is similarly exchanged with $A_5(b)$ but that is not absolutely irreducible.

5.2. Images and discriminants. For the purposes of establishing the first typical cases of the paramodular conjecture, we observe the following.

Lemma 5.2.1. *Suppose N is odd and squarefree and let A be an abelian surface over \mathbb{Q} of conductor N equipped with a polarization of odd degree. Then the residual representation*

$$\bar{\rho}_{A,2} : \mathrm{Gal}_{\mathbb{Q},S} \rightarrow \mathrm{GSp}_4(\mathbb{F}_2)$$

(where $S = \{p : p \mid N\} \cup \{\ell, \infty\}$) is absolutely irreducible if and only if its image is isomorphic to $S_5(b)$, S_6 , or $S_3 \wr S_2$.

Proof. By work of Brumer and Kramer [2014, §7.3], whenever N is not a square, the image is either S_5 , S_6 , or $S_3 \wr S_2$. To force $S_5(b)$, it suffices that there is a prime $p \mid N$ such that A_p has toroidal dimension one (i.e., $p \parallel N$) and that p be ramified in $\mathbb{Q}(A[2])$. If A is semistable and the Galois group is $S_5(a)$, then the toroidal dimension at the bad primes is 2 since there are no transvections. □

Remark 5.2.2. In general, if $A[2]$ is absolutely irreducible, then the degree of any minimal polarization on A is odd.

Next, we convert the upper bound from Lemma 4.3.10 on the conductor into an upper bound on the discriminant. We first recall the following standard result.

Lemma 5.2.3. *Let $a(x) \in \mathbb{Q}[x]$ be irreducible and let Ω be the set of roots of $a(x)$ in \mathbb{Q}^{al} . Let $\alpha \in \Omega$, let $K_0 = \mathbb{Q}(\alpha)$, and let K be the normal closure of K_0 . Let \mathfrak{p} be a prime of K that is tamely ramified in the extension $K \supseteq \mathbb{Q}$, and let $p \in \mathbb{Z}$ be the prime lying below \mathfrak{p} . Finally, let $I_{\mathfrak{p}} \leq \text{Gal}(K | \mathbb{Q})$ denote the inertia group at \mathfrak{p} . Then*

$$\text{ord}_p(d_{K_0}) = \deg a(x) - \#\Omega/I_{\mathfrak{p}}$$

where $\#\Omega/I_{\mathfrak{p}}$ denotes the number of orbits of $I_{\mathfrak{p}}$ acting on Ω .

We now specialize to our case of interest.

Proposition 5.2.4. *Let $p \parallel N$ be odd. Let K be the fixed field of $\ker \bar{\rho}_{f,2}$.*

- (a) *If $\text{Gal}(K | \mathbb{Q}) \simeq S_3 \wr S_2$, or S_m with $m = 5, 6$, then K is the normal closure of a field K_0 of degree 6, or respectively m , with $\text{ord}_p d_{K_0} \leq 1$.*
- (b) *If $\text{Gal}(K | \mathbb{Q}) \simeq A_m$, with $m = 5, 6$, then K is the normal closure of a field K_0 of degree m with p unramified in K_0 (i.e., $\text{ord}_p d_{K_0} = 0$).*

Proof. Decomposing the Weil–Deligne representation at p , we see by Lemma 4.3.10 that the image of inertia is either trivial or a 2×2 -Jordan block. If trivial, the extension is unramified and the result holds, so suppose we are in the latter case. Under the isomorphism $\text{GSp}_4(\mathbb{F}_2) \simeq S_6$ above (5.1.1), nontrivial elements of this Jordan block correspond to cycle decomposition $2 + 2 + 2$ or $2 + 1 + 1 + 1 + 1$, and these are exchanged by an outer automorphism.

For (a), by a faithful permutation representation on the cosets of a core-free subgroup, a field K_0 of the given degree exists. If the residual image inside S_6 is invariant under such an automorphism (which holds for S_6 and $S_3 \wr S_2$), then we can choose our subfield K_0 corresponding to the latter case, and conclude $\text{ord}_p d_{K_0} \leq 1$ by Lemma 5.2.3. If $\text{Gal}(K | \mathbb{Q}) \simeq S_5$, we have only the possibility $2 + 1 + 1 + 1$ again giving $\text{ord}_p d_{K_0} \leq 1$.

Finally, for (b) and the groups A_5, A_6 , we find no possibilities and reach a contradiction, so we conclude that K_0 is unramified at p . □

5.3. Core-free extensions and obstructing elements. We will compute all obstructing extensions $\varphi : \text{Gal}(L|F) \hookrightarrow E$ extending $\bar{\rho}$ (Definition 2.3.17); we represent $L \supseteq K \supseteq F$ by an exact core-free subextension $L_0 \supseteq K_0 \supseteq F$ (Definition 3.1.7) arising from an exact core-free subgroup $D \leq E$ which is as large as possible, to make the degree of the subextension as small as possible.

For each G in (5.1.8), we therefore first seek subgroups $\varphi : E \hookrightarrow \text{sp}_4(\mathbb{F}_2) \rtimes G$ such that $\pi(E) = G$; such extensions are obstructing (Definition 2.3.18) if they have nonzero upper trace in the matrix realization (5.1.4). Consider first the case $G = S_5(b)$.

Theorem 5.3.1. *For $G = S_5(b)$, there are exactly 10 extension groups E up to conjugacy in $\text{M}_4(\mathbb{F}_2) \rtimes G$, with $\#V = [E : G] = 2^k$ where $k = 0, 0, 1, 4, 4, 5, 5, 6, 9, 10$, respectively.*

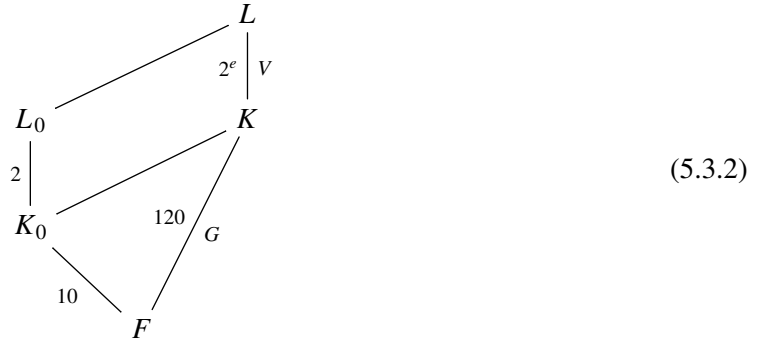
Furthermore, let

$$H = D_6(b) := \langle (1\ 2), (1\ 3), (4\ 5) \rangle \leq G;$$

then for all $E \not\cong G$, there is an exact core-free subgroup $D \leq E$ of index 2 such that $\pi(D) = H$ as in (3.1.6).

Proof. This theorem is proven by explicit computation in Magma [Bosma et al. 1997]; the code is available online [Tornara 2018] together with the verbose output. There are exactly 18 conjugacy classes of subgroups $\varphi : E \hookrightarrow \text{sp}_4(\mathbb{F}_2) \rtimes G$ with $\pi(E) = G$; these subgroups fall into 10 conjugacy classes in $\text{M}_4(\mathbb{F}_2) \rtimes G$. Let $H = D_6(b) := \langle (1\ 2), (1\ 3), (4\ 5) \rangle \leq G$ be as in the statement. Then H is dihedral of order $\#H = 12$ and index $[G : H] = 10$ and it can be verified that for each such $E \not\cong G$, there is at least one subgroup $W \leq V$ of index 2 such that $D \leq E$ is an exact core-free subgroup. \square

The somewhat complicated field diagram (3.1.8) in our case simplifies to:



We understand the large extension $L \supseteq K \supseteq F$ as the Galois closure of the exact core-free subextension $L_0 \supseteq K_0 \supseteq F$, with $L_0 \supseteq K_0$ quadratic. The extension K_0 is realized explicitly as follows: if $K \supseteq F$ is the splitting field of a quintic polynomial $f(x)$ with roots $\alpha_1, \dots, \alpha_5$ permuted by S_5 , then $K_0 = K^H = F(\alpha_4 + \alpha_5)$.

In a similar way, we have the result for the remaining two groups.

Theorem 5.3.3. (a) For $G = S_3 \wr S_2 \leq \text{GSp}_4(\mathbb{F}_2)$, there are exactly 20 extension groups E up to conjugacy in $\text{M}_4(\mathbb{F}_2) \rtimes G$, with $\#V = [E : G] = 2^k$ and

$$k = 0, 0, 1, 1, 2, 4, 4, 4, 4, 5, 5, 5, 5, 6, 6, 8, 8, 9, 9, 10.$$

Let $H = C_2^2 \leq G$ with $[G : H] = 18$. Then for each such E , there is an exact core-free subgroup $D \leq E$ such that $\pi(D) = H$.

(b) For $G = S_6 \simeq \text{GSp}_4(\mathbb{F}_2)$, the analogous statement to (a) holds, with 7 groups having $k = 0, 0, 1, 5, 5, 6, 10$ and $H = S_3(b)^2$.

Remark 5.3.4. With reference to computing conjugacy classes in stages as in Section 3.3, we note that the index 2 subgroups of the 18 subgroups C_2^2 of $S_3 \wr S_2$ are not sufficient to find obstructing classes for all 20 extension groups if one applies the more limited strategy exhibited in Remark 3.3.2.

Remark 5.3.5. The remaining cases of subgroups $G \leq \mathrm{GSp}_4(\mathbb{F}_2)$ may be computed with the same method and the same code.

6. Computing Hecke eigenvalues by specialization

Having set up the required Galois theory, we now compute Hecke eigenvalues of particular Siegel paramodular newforms. In this section, we use the technique of restriction to a modular curve to accomplish these eigenvalue computations. We continue the notation from Section 4.2.

6.1. Jacobi forms and Borcherds products. We construct our paramodular forms using Gritsenko lifts of Jacobi forms and Borcherds products. In this section, we quickly review what we need from these theories.

We begin with Jacobi forms; we refer to [Eichler and Zagier 1985] for further reference. Each Jacobi form $\phi \in J_{k,N}$ of weight k and index N has a Fourier expansion

$$\phi(\tau, z) = \sum_{n,r \in \mathbb{Z}} c(n, r; \phi) q^n \zeta^r, \tag{6.1.1}$$

where $q = e(\tau)$ and $\zeta = e(z)$. We write $\phi \in J_{k,N}(R)$ if all the Fourier coefficients of ϕ lie in a ring $R \subseteq \mathbb{C}$. We will need the level-raising operators $V_m : J_{k,N} \rightarrow J_{k,mN}$ (see [Eichler and Zagier 1985, p. 41]) that act on $\phi \in J_{k,N}$ via

$$c(n, r; \phi | V_m) = \sum_{\delta | \gcd(n,r,m)} \delta^{k-1} c\left(\frac{mn}{\delta^2}, \frac{r}{\delta}; \phi\right). \tag{6.1.2}$$

The Gritsenko lift [1995]

$$\mathrm{Grit} : J_{k,N} \mathrm{ cusp} \rightarrow S_k(K(N))$$

lifts a Jacobi cusp form ϕ to a paramodular form f by the rule

$$a\left(\begin{matrix} n & r/2 \\ r/2 & Nm \end{matrix}; \mathrm{Grit}(\phi)\right) = c(n, r; \phi | V_m).$$

We also have $\mathrm{Grit}(\phi)|_k \mu_N = (-1)^k \mathrm{Grit}(\phi)$, so that a Gritsenko lift has paramodular Fricke sign $(-1)^k$.

One convenient way to construct Jacobi forms is to use the theta blocks created by Gritsenko, Skoruppa and Zagier [2018]. Recall the Dedekind η -function and the Jacobi ϑ -function

$$\begin{aligned} \eta(\tau) &= q^{1/24} \prod_{n=1}^{\infty} (1 - q^n) = \sum_{n=1}^{\infty} \left(\frac{12}{n}\right) q^{n^2/24}, \\ \vartheta(\tau, z) &= \sum_{n=-\infty}^{\infty} (-1)^n q^{(2n+1)^2/8} \zeta^{(2n+1)/2} = q^{1/8} (\zeta^{1/2} - \zeta^{-1/2}) \sum_{n=1}^{\infty} (-1)^{n+1} q^{\binom{n}{2}} \sum_{j=-(n-1)}^{n-1} \zeta^j. \end{aligned}$$

For $d \in \mathbb{Z}_{>0}$ let $\vartheta_d(\tau, z) = \vartheta(\tau, dz)$. For $d_1, \dots, d_\ell \in \mathbb{Z}_{>0}$ and $k \in \mathbb{Z}$, define the *theta block*

$$\mathrm{TB}_k[\mathbf{d}] = \mathrm{TB}_k[d_1, d_2, \dots, d_\ell] = \eta^{2k} \prod_{j=1}^{\ell} \frac{\vartheta_{d_j}}{\eta}. \tag{6.1.3}$$

The theta block $TB_k[\mathbf{d}]$ defines a meromorphic Jacobi form (with multiplier) of weight k and index $m = \frac{1}{2}(d_1^2 + \dots + d_\ell^2)$. Moreover, by [Eichler and Zagier 1985] (compare [Poor and Yuen 2015, Theorem 4.3]), the theta block $TB_k[\mathbf{d}]$ is a Jacobi cusp form if

$$12 \mid (k + \ell) \quad \text{and} \quad \frac{k}{12} + \frac{1}{2} \sum_{j=1}^{\ell} \bar{B}_2(d_j x) > 0, \tag{6.1.4}$$

where $B_2(x) := x^2 - x + \frac{1}{6}$ and $\bar{B}_2(x) := B_2(x - \lfloor x \rfloor)$.

Second, we use Borcherds products in the construction of paramodular forms. Let ψ be a weakly holomorphic Jacobi form of weight 0 and index N with integral Fourier coefficients on singular indices with Fourier expansion (6.1.1). Define

$$A(\psi) := \frac{1}{24} \sum_{r \in \mathbb{Z}} c(0, r; \psi), \quad B(\psi) := \frac{1}{2} \sum_{r \geq 1} r c(0, r; \psi), \quad C(\psi) := \frac{1}{4} \sum_{r \in \mathbb{Z}} r^2 c(0, r; \psi).$$

Then $A(\psi), B(\psi), C(\psi) \in \mathbb{Q}$. The Borcherds product of ψ is a meromorphic paramodular form $\text{Borch}(\psi)$, perhaps with nontrivial character on $K(N)$, with

$$\text{Borch}(\psi) = q^{A(\psi)} \zeta^{B(\psi)} \xi^{C(\psi)} \prod_{n,r,m} (1 - q^n \zeta^r \xi^{mN})^{c(mn,r;\psi)}, \tag{6.1.5}$$

where the product is over $n, r, m \in \mathbb{Z}$ such that: (i) $m \geq 0$; (ii) if $m = 0$, then $n \geq 0$; and (iii) if $m = n = 0$, then $r < 0$. Borcherds products are not always holomorphic and, when holomorphic, not always cuspidal.

6.2. Construction of newforms. In this section, we define the nonlift paramodular newforms of interest to this article, with levels 277, 353, 587. We will see later that this way of writing paramodular forms makes the computation of Hecke eigenvalues feasible.

We refer to Section 4.2 for notation. We now define the nonlift paramodular form $f_{277} \in S_2(K(277), \mathbb{Z})^+$ following Poor and Yuen [2015, Theorem 7.1]. Define the following ten theta blocks:

$$\begin{aligned} \Xi_1 &:= TB_2(2, 4, 4, 4, 5, 6, 8, 9, 10, 14) & \Xi_6 &:= TB_2(2, 3, 3, 5, 5, 7, 8, 10, 10, 13) \\ \Xi_2 &:= TB_2(2, 3, 4, 5, 5, 7, 7, 9, 10, 14) & \Xi_7 &:= TB_2(2, 3, 3, 4, 5, 6, 7, 9, 10, 15) \\ \Xi_3 &:= TB_2(2, 3, 4, 4, 5, 7, 8, 9, 11, 13) & \Xi_8 &:= TB_2(2, 2, 4, 5, 6, 7, 7, 9, 11, 13) \\ \Xi_4 &:= TB_2(2, 3, 3, 5, 6, 6, 8, 9, 11, 13) & \Xi_9 &:= TB_2(2, 2, 4, 4, 6, 7, 8, 10, 11, 12) \\ \Xi_5 &:= TB_2(2, 3, 3, 5, 5, 8, 8, 8, 11, 13) & \Xi_{10} &:= TB_2(2, 2, 3, 5, 6, 7, 9, 9, 11, 12). \end{aligned} \tag{6.2.1}$$

We have, for $i = 1, \dots, 10$,

$$\Xi_i \in J_{2,277}^{\text{cusp}}(\mathbb{Z}) \quad \text{and} \quad G_i := \text{Grit}(\Xi_i) \in S_2(K(277), \mathbb{Z}).$$

Let f_{277} be the (a priori) meromorphic function on \mathcal{H}_2 defined by

$$\begin{aligned}
 f_{277} := & (-14G_1^2 - 20G_8G_2 + 11G_9G_2 + 6G_2^2 - 30G_7G_{10} + 15G_9G_{10} + 15G_{10}G_1 - 30G_{10}G_2 \\
 & - 30G_{10}G_3 + 5G_4G_5 + 6G_4G_6 + 17G_4G_7 - 3G_4G_8 - 5G_4G_9 - 5G_5G_6 + 20G_5G_7 \\
 & - 5G_5G_8 - 10G_5G_9 - 3G_6^2 + 13G_6G_7 + 3G_6G_8 - 10G_6G_9 - 22G_7^2 \\
 & + G_7G_8 + 15G_7G_9 + 6G_8^2 - 4G_8G_9 - 2G_9^2 + 20G_1G_2 - 28G_3G_2 + 23G_4G_2 \quad (6.2.2) \\
 & + 7G_6G_2 - 31G_7G_2 + 15G_5G_2 + 45G_1G_3 - 10G_1G_5 - 2G_1G_4 - 13G_1G_6 \\
 & - 7G_1G_8 + 39G_1G_7 - 16G_1G_9 - 34G_3^2 + 8G_3G_4 + 20G_3G_5 + 22G_3G_6 + 10G_3G_8 \\
 & + 21G_3G_9 - 56G_3G_7 - 3G_4^2) / (-G_4 + G_6 + 2G_7 + G_8 - G_9 + 2G_3 - 3G_2 - G_1).
 \end{aligned}$$

A main result of Poor and Yuen [2015, Theorem 7.1] is that f_{277} is actually *holomorphic*: in fact, $f_{277} \in S_2(K(277), \mathbb{Z})^+$ is a cuspidal, nonlift, paramodular form of weight 2 that is an eigenform for all Hecke operators and has integral Fourier coefficients whose greatest common divisor is 1. There are no nontrivial weight 2 paramodular cusp forms of level 1, so since 277 is prime, f_{277} is a newform. Equation (4.2.9) and Lemma 4.2.17 imply that the Euler factors $Q_p(f_{277}, t)$ are integral.

The first few eigenvalues for f_{277} were computed [Poor and Yuen 2015] as

$$a_p(f_{277}) = -2, -1, -1, 1, -2 \quad \text{for } p = 2, 3, 5, 7, 11 \quad (6.2.3)$$

and the first three Hecke polynomials, identifying f_{277} as type **(G)**, are:

$$\begin{aligned}
 Q_2(f_{277}, t) &= 1 + 2t + 4t^2 + 4t^3 + 4t^4, \\
 Q_3(f_{277}, t) &= 1 + t + t^2 + 3t^3 + 9t^4, \\
 Q_5(f_{277}, t) &= 1 + t - 2t^2 + 5t^3 + 25t^4.
 \end{aligned} \quad (6.2.4)$$

Remark 6.2.5. The form f_{277} can also be realized as the sum of a Borchers product and a Gritsenko lift, giving a second, independent construction by Poor, Shurman, and Yuen [2018].

In a similar way, we construct a second form

$$f_{353} := Q(G_1, \dots, G_{11}) \in S_2(K(353), \mathbb{Z})^+ \quad (6.2.6)$$

(plus eigenspace for the Fricke involution, as in (4.2.4)) a quotient of a quadratic polynomial by a linear polynomial of 11 Gritsenko lifts of theta blocks: see [Poor and Yuen 2015, Theorem 7.4] for the specific formula for Q and the forms G_i . This construction was contingent upon assuming the existence of some nonlift in $S_2(K(353))$; however, the dimension $\dim S_2(K(353)) = 12$ is now known [Poor et al. 2018] via the construction of a nonlift Borchers product in $S_2(K(353))$.

The first two Euler factors, each showing that f_{353} is of type **(G)**, are

$$Q_2(f_{353}, t) = 1 + t + 3t^2 + 2t^3 + 4t^4, \quad Q_3(f_{353}, t) = 1 + 2t + 4t^2 + 6t^3 + 9t^4. \quad (6.2.7)$$

Finally, we construct a form of level 587 as a Borchers product. An antisymmetric nonlift Borchers product $f_{587}^- \in S_2(K(587), \mathbb{Z})^-$ was recently constructed by Gritsenko, Poor, and Yuen [2019]. The

form f_{587}^- is necessarily an eigenform because $\dim S_2(K(587))^- = 1$. The Fourier expansion is given by formally expanding

$$f_{587}^- = \text{Borch}(\psi) = \xi^{587} \phi \exp(-\text{Grit}(\psi)) \quad \text{for} \quad \psi = (\phi \mid V_2 - \Xi)/\phi, \tag{6.2.8}$$

where

$$\begin{aligned} \phi &= \text{TB}_2(1, 1, 2, 2, 2, 3, 3, 4, 4, 5, 5, 6, 6, 7, 8, 8, 9, 10, 11, 12, 13, 14) \in J_{2,587}^{\text{cusp}}, \\ \Xi &= \text{TB}_2(1, 10, 2, 2, 18, 3, 3, 4, 4, 15, 5, 6, 6, 7, 8, 16, 9, 10, 22, 12, 13, 14) \in J_{2,1174}^{\text{cusp}}. \end{aligned} \tag{6.2.9}$$

For the Borcherds product that appears in the formula for f_{587}^- , we have $\text{Borch}(\psi) \in S_k(K(587))$ with $k = \frac{1}{2}c(0, 0; \psi) = 2$ [Gritsenko et al. 2019]. The first two Euler factors, verifying type **(G)**, are computed to be

$$Q_2(f_{587}^-, t) = 1 + 3t + 5t^2 + 6t^3 + 4t^4, \quad Q_3(f_{587}^-, t) = 1 + 4t + 9t^2 + 12t^3 + 9t^4. \tag{6.2.10}$$

6.3. Specialization. To compute the action of the Hecke operators directly on a Fourier expansion of a Siegel paramodular form would require manipulations with series in three variables. To avoid this, we specialize our form. Possibilities for this specialization include restriction to Humbert surfaces (typically producing Hilbert modular forms), restriction to modular curves (producing classical modular forms), or evaluation at CM points (producing a numerical result, see Colman, Ghitza, and Ryan [2019]). Each of these methods has certain advantages and disadvantages — we choose to restrict to modular curves and work with one-variable q -series to avoid rigorous analysis of the upper bounds on the tails of convergent numerical series. The biggest advantage of our choice, however, is that Proposition 6.3.8 allows us to sum over only $O(p^2)$ cosets instead of $O(p^3)$ cosets, a significant savings; it is not clear whether such a speedup is available to a method that numerically evaluates at CM points.

Remark 6.3.1. Specialization of Siegel modular forms is not a new idea, but here we take a different approach. In previous work of Poor and Yuen [2015], only three Euler factors were computed for f_{277} because the computation relied on multiplying initial expansions of multivariable Fourier series. Instead, below we will write the action of the Hecke operator $T(p)$ on a paramodular form f as a sum of slashes $f|_k T(p) = \sum_j f|_k M_j$, and the main innovation is to specialize each part of $f|M_j$ to a one variable q -series prior to any addition, multiplication, or division. Specialization was also used by Poor and Yuen [2007] to compute upper bounds on dimensions and some Fourier coefficients by taking advantage of the known structure of the target space of elliptic modular forms, whereas here we only use the one variable nature of the target space.

Let $s \in M_2^{\text{sym}}(\mathbb{Q})_{>0}$ be a symmetric, positive definite matrix with rational coefficients. Let \mathcal{H}_g be the Siegel upper half space of dimension g , so \mathcal{H}_1 is the upper half-plane. Define the holomorphic map

$$\begin{aligned} \phi_s : \mathcal{H}_1 &\rightarrow \mathcal{H}_2 \\ \tau &\mapsto s\tau. \end{aligned} \tag{6.3.2}$$

Lemma 6.3.3. *Let $R \subseteq \mathbb{C}$ be a subring. Let $s = \begin{pmatrix} a & b \\ b & c/N \end{pmatrix} \in M_2^{\text{sym}}(\mathbb{Q})_{>0}$ with $a, b, c \in \mathbb{Z}$. Then the pullback under ϕ_s defines a ring homomorphism*

$$\phi_s^* : M(K(N), R) \rightarrow M(\Gamma_0(\det(s)N), R) \tag{6.3.4}$$

from the graded ring of Siegel paramodular forms of level N with coefficients in R to the graded ring of classical modular forms of level $\det(s)N$ with coefficients in R . The map ϕ_s^ multiplies weights by 2 and maps cusp forms to cusp forms.*

Proof. The proof follows from a straightforward modification of a result of Poor and Yuen [2007, Proposition 5.4]. □

Let $f \in M_k(K(N), R)$ be a paramodular form with Fourier expansion (4.2.5), the Fourier expansion of the specialization $\phi_s^* f \in M_{2k}(\Gamma_0(\det(s)N), R)$ is

$$(\phi_s^* f)(\tau) = f(s\tau) = \sum_{n=0}^{\infty} \left(\sum_{T: \text{Tr}(sT)=n} a(T; f) \right) q^n. \tag{6.3.5}$$

Furthermore, the specialization of f after slashing with a block upper-triangular matrix $\begin{pmatrix} A & B \\ 0 & D \end{pmatrix} \in \text{GSp}_4^+(\mathbb{Q})$ with similitude $\mu = \det(AD)^{1/2}$ is given by

$$\begin{aligned} \phi_s^*(f|_k \begin{pmatrix} A & B \\ 0 & D \end{pmatrix})(\tau) &= (f|_k \begin{pmatrix} A & B \\ 0 & D \end{pmatrix})(s\tau) = \det(AD)^{k-3/2} \det(D)^{-k} f(AsD^{-1}\tau + BD^{-1}) \\ &= \det(A)^k \det(AD)^{-3/2} \sum_{n \in \mathbb{Q}_{\geq 0}} \left(\sum_{T: \text{Tr}(AsD^{-1}T)=n} e(\text{Tr}(BD^{-1}T)) a(T; f) \right) q^n. \end{aligned} \tag{6.3.6}$$

Let $s = \begin{pmatrix} a & b \\ b & c/N \end{pmatrix} \in M_2^{\text{sym}}(\mathbb{Q})_{>0}$ with $a, b, c \in \mathbb{Z}$. Using (4.2.8), the specialization of $f|_k T(p)$ may be written

$$\begin{aligned} \phi_s^*(f|_k T(p))(\tau) &= p^{2k-3} f(ps\tau) + p^{k-3} \sum_{i \bmod p} f\left(\begin{pmatrix} a/p & b \\ b & pc/N \end{pmatrix} \tau + \begin{pmatrix} i/p & 0 \\ 0 & 0 \end{pmatrix}\right) \\ &\quad + p^{k-3} \sum_{i \bmod p} \left(\sum_{j \bmod p} f\left(\begin{pmatrix} pa & b+ia \\ b+ia & (c/N+2ib+i^2a)/p \end{pmatrix} \tau + \begin{pmatrix} 0 & 0 \\ 0 & j/p \end{pmatrix}\right) \right) \\ &\quad + p^{-3} \sum_{i,j,k \bmod p} f\left(s\tau/p + \begin{pmatrix} i/p & j/p \\ j/p & k/p \end{pmatrix}\right). \end{aligned} \tag{6.3.7}$$

Upon expanding in Puiseux q -series, there is cancellation among these sums of specializations. The following proposition shows that partial summation gives new specializations whose sum over smaller index sets equals the original sum for integral powers of q . For a Puiseux series $f \in \mathbb{C}[[q^{1/\infty}]]$ and $e \in \mathbb{Q}_{\geq 0}$, we denote by $\text{coeff}_e f \in \mathbb{C}$ the coefficient of q^e in f .

Proposition 6.3.8. *Let $s = \begin{pmatrix} a & b \\ b & c/N \end{pmatrix} \in M_2^{\text{sym}}(\mathbb{Q})_{>0}$ with $a, b, c \in \mathbb{Z}$. Let p be prime, and let $f \in M_k(K(N))$. Then the following statements hold for all $e \in \mathbb{Z}_{\geq 0}$:*

(a) If $p \nmid a$, then

$$\begin{aligned} \text{coeff}_e \sum_{i \bmod p} f\left(\begin{pmatrix} a/p & b \\ b & pc/N \end{pmatrix} \tau + \begin{pmatrix} i/p & 0 \\ 0 & 0 \end{pmatrix}\right) &= p \text{coeff}_e f\left(\begin{pmatrix} a/p & b \\ b & pc/N \end{pmatrix} \tau\right) \\ \text{coeff}_e \sum_{i,j,k \bmod p} f\left(s\tau/p + \begin{pmatrix} i/p & j/p \\ j/p & k/p \end{pmatrix}\right) &= p \text{coeff}_e \sum_{j,k \bmod p} f\left(s\tau/p + \begin{pmatrix} 0 & j/p \\ j/p & k/p \end{pmatrix}\right). \end{aligned}$$

(b) If $p \nmid b$, then

$$\text{coeff}_e \sum_{i,j,k \bmod p} f\left(s\tau/p + \begin{pmatrix} i/p & j/p \\ j/p & k/p \end{pmatrix}\right) = p \text{coeff}_e \sum_{i,k \bmod p} f\left(s\tau/p + \begin{pmatrix} i/p & 0 \\ 0 & k/p \end{pmatrix}\right).$$

(c) If $p \nmid c$, then

$$\text{coeff}_e \sum_{i,j,k \bmod p} f\left(s\tau/p + \begin{pmatrix} i/p & j/p \\ j/p & k/p \end{pmatrix}\right) = p \text{coeff}_e \sum_{i,j \bmod p} f\left(s\tau/p + \begin{pmatrix} i/p & j/p \\ j/p & 0 \end{pmatrix}\right).$$

(d) For $i \in \mathbb{Z}$, if $p \nmid (c + 2ibN + i^2aN)$, then

$$\text{coeff}_e \sum_{j \bmod p} f\left(\begin{pmatrix} pa & b+ia \\ b+ia & (c/N+2ib+i^2a)/p \end{pmatrix} \tau + \begin{pmatrix} 0 & 0 \\ 0 & j/p \end{pmatrix}\right) = p \text{coeff}_e f\left(\begin{pmatrix} pa & b+ia \\ b+ia & (c/N+2ib+i^2a)/p \end{pmatrix} \tau\right).$$

Proof. We prove (c); the other proofs are similar. Suppose $p \nmid c$. Let $e \in \mathbb{Z}_{\geq 0}$. Then the coefficient of q^e in the left-hand side is equal to

$$\sum_{\substack{i,j,k \bmod p \\ n,r,m:an+br+cm=pe}} e((in + jr + km)/p)a(T; f) \tag{6.3.9}$$

where $T = \begin{pmatrix} n & r/2 \\ r/2 & mN \end{pmatrix}$. If any of n, r, m is not a multiple of p , then summing over i, j, k modulo p in (6.3.9) would yield a contribution of zero. Hence we may restrict the sum to the terms where $p \mid n, p \mid r$, and $p \mid m$. But since $p \nmid c$ and given $an + br + cm = pe$, the conditions $p \mid n$ and $p \mid r$ imply $p \mid m$. Thus (6.3.9) becomes simply

$$\begin{aligned} \sum_{\substack{i,j,k \bmod p \\ n,r,m:an+br+cm=pe \\ p \mid n, p \mid r}} e((in + jr + 0)/p)a(T; f) &= p \sum_{\substack{i,j \bmod p \\ n,r,m:an+br+cm=pe \\ p \mid n, p \mid r}} e((in + jr)/p)a(T; f) \\ &= p \sum_{\substack{i,j \bmod p \\ n,r,m:an+br+cm=pe}} e((in + jr)/p)a(T; f) \\ &= p \text{coeff}_e \sum_{i,j \bmod p} f\left(s\tau/p + \begin{pmatrix} i/p & j/p \\ j/p & 0 \end{pmatrix}\right). \quad \square \end{aligned}$$

Remark 6.3.10. Proposition 6.3.8 provides a certain subtle speedup because the coefficients at integral powers are equal, even though the series themselves are not necessarily equal. Further simplifying the above sums to

$$p^3 \sum_{\substack{n,r,m:an+br+cm=pe \\ p|n,p|r,p|m}} a(T; f).$$

does not help: we want to leave the sums in terms of coefficients of specializations.

In a similar way, we can compute the specialization $\phi_s^*(f|_k T_1(p^2))$ and there are similar cancellations in the character sums as in Proposition 6.3.8.

6.4. Algorithmic detail. In this section, we provide three further bits of algorithmic detail.

First, we describe the choice of s . Suppose f has a nonzero coefficient $a(t_0; f)$ where t_0 has small determinant and small entries. If we choose s to be the adjoint of $2t_0$, then the restriction $\phi_s^*(f)$ likely begins with $a(t_0; f)q^{\det(s)}$. In particular if t_0 has minimal determinant, then this is forced. In practice, we can just check the initial expansion to see that

$$\phi_s^*(f)(\tau) = a(t_0; f)q^{\det(s)} + \text{higher powers of } q.$$

For each $T(p)$, we want to expand $\phi_s^*(f|T(p))$ to at least q^e where $e = \det(s)$ is the target exponent of q . For a polynomial combination of Gritsenko lifts and Borcherds products, the target exponent of each part $g(G\tau + H)$ would also be e . But for a rational function of Gritsenko lifts and Borcherds products, we have to be slightly more careful. If the denominator of this rational functional restricted to $(G\tau + H)$ has leading term q^μ , then we must expand both the numerator and denominator to a higher target term $q^{e+\mu}$. Therefore, we may end up evaluating the restriction of the denominator twice, with the initial execution used to get the leading exponent μ .

Second, we provide our algorithm for finding all T such that $\langle G, T \rangle \leq u$. Let G and H be two rational, symmetric 2×2 matrices with G positive definite. We explain how to effectively compute specializations of the form $f(G\tau + H)$, as in (6.3.7) or Proposition 6.3.8. We adapt our index sets \mathcal{S} to the type used in (6.1.5) for Borcherds products but they can be used in all the cases we need to program. For any $u, \delta \in \mathbb{R}$, let

$$\mathcal{S}(N, G, u, \delta) = \left\{ (n, r, m) \in \mathbb{Z}^3 : \text{tr}\left(\begin{pmatrix} n & r/2 \\ r/2 & mN \end{pmatrix} G\right) \leq u, m \geq 0, 4mnN - r^2 \geq \delta, \right. \\ \left. \text{if } m = 0 \text{ then } n \geq 0 \text{ and if } m = n = 0 \text{ then } r < 0 \right\}.$$

Proposition 6.4.1. *Let $G = \begin{pmatrix} \alpha & \beta \\ \beta & \gamma \end{pmatrix} \in M_2(\mathbb{R})$ be positive definite. Let $u, \delta \in \mathbb{R}$. Let $\Delta = \det G = \alpha\gamma - \beta^2 > 0$. Let $X = 4\alpha umN - \alpha^2\delta - 4\Delta(mN)^2$. Then the elements $(n, r, m) \in \mathcal{S}(N, G, u, \delta)$ satisfy the following bounds:*

(a) *If $m \geq 1$, then*

$$1 \leq m \leq \frac{\alpha(u + \sqrt{u^2 - \delta\Delta})}{2\Delta N}, \quad \frac{-2\beta mN - \sqrt{X}}{\alpha} \leq r \leq \frac{-2\beta mN + \sqrt{X}}{\alpha}, \quad \text{and} \quad \frac{r^2 + \delta}{4mN} \leq n \leq \frac{u - \beta r - \gamma mN}{\alpha}.$$

(b) If $m = 0$ and $n > 0$, then

$$r^2 \leq -\delta \quad \text{and} \quad 1 \leq n \leq \frac{u - \beta r}{\alpha}.$$

(c) If $m = n = 0$, then

$$r^2 \leq -\delta \quad \text{and} \quad r < 0.$$

Proof. The main two conditions that need to be satisfied are $\alpha n + \beta r + \gamma m N \leq u$ and $4mnN - r^2 \geq \delta$. The case $m = 0$ is straightforward, so we only deal with the case $m \geq 1$ here. These two inequalities lead immediately to the third inequality as stated in the proposition. From this third inequality, we work with terms on the left and right of n ; multiply through by $4mN\alpha$ and put the terms on one side:

$$\alpha r^2 + \alpha \delta - 4mNu + 4mN\beta r + 4\gamma(mN)^2 \leq 0.$$

Solving this quadratic inequality for r yields the second inequality stated in the proposition. A condition for there to be a solution in r is that the inside X of the square root must be nonnegative. Solving the resulting quadratic inequality yields the first inequality in the proposition. \square

We conclude with a final speedup. Suppose we wish to calculate the coefficient of q^e in $f(G\tau + H)$. If there are no $(n, r, m) \in \mathcal{S}(N, G, u, \delta)$ such that $\text{tr}\left(\binom{n}{r/2} \binom{r/2}{mN} G\right) = e$, then we may skip the term involving G . This simple observation is especially useful for terms in the second summand in (6.3.7): for well chosen s , there are typically at most 2 choices of i for which such (n, r, m) exist. It often happens that, for these surviving i , Proposition 6.3.8(d) applies.

6.5. Example of restricting f_{277} . Now suppose that f is represented as a rational function in Gritsenko lifts G_i with coefficients in a commutative ring R by $f = Q(G_1, \dots, G_r)$. Both the slash by M and the specialization by ϕ_s^* may be applied directly to each Gritsenko lift, so that we obtain

$$\phi_s^*(f | M) = Q(\phi_s^*(G_1 | M), \dots, \phi_s^*(G_r | M)). \tag{6.5.1}$$

If the Fourier coefficients of f satisfy $a(T; f) \in R \subseteq \mathbb{C}$, then for the representative matrices M_j appearing in the coset decomposition (4.2.8) for the Hecke operator $T(p)$, the sum in (6.3.6) can be taken over $n \in \frac{1}{p}\mathbb{Z}_{\geq 0}$ and the coefficients of $\phi_s^*(f | M_j)$ belong to the ring $R[\frac{1}{p}, \zeta_p]$ where $\zeta_p = e(\frac{1}{p})$ is a primitive p -th root of unity. From specializing $f | T(p) = \sum_j f | M_j = a_p(f)f$, the eigenvalue $a_p(f)$ for $T(p)$ can be computed by performing field operations on Laurent–Puiseux series in q via

$$a_p(f) = \frac{1}{\phi_s^*(f)} \sum_j \phi_s^*(f | M_j) \in R[\frac{1}{p}, \zeta_p][[q^{1/p}]] \tag{6.5.2}$$

whenever the specializing curve ϕ_s is chosen so that $\phi_s^*(f)$ is not identically zero. In practice, we choose a target exponent e such that $\text{coeff}_e \phi_s^* f \neq 0$ and then

$$a_p(f) = \frac{\text{coeff}_e(\sum_j \phi_s^*(f | M_j))}{\text{coeff}_e(\phi_s^*(f))}. \tag{6.5.3}$$

Remark 6.5.4. One practical advantage of this technique of restricting to modular curves is that when more than one coefficient in the q -expansion of (6.5.2) is computed, it constitutes a double check on the value of $a_p(f)$.

Example 6.5.5. We consider the core example of the form f_{277} of level $N = 277$ constructed above (6.2.2). A Fourier coefficient of f_{277} whose matrix index has the smallest determinant is $a(t_0; f_{277}) = -3$, where $t_0 = \begin{pmatrix} 49 & -233/2 \\ -233/2 & 277 \end{pmatrix}$ and $\det(2t_0) = 3$. Accordingly we select $s = \begin{pmatrix} 544 & 233 \\ 233 & 98 \end{pmatrix}$, which is the adjoint of $2t_0$. Working over $R = \mathbb{Z}$, we find

$$\phi_s^*(f_{277}) = -3q^3 + 6q^6 + 6q^9 + 3q^{12} + 3q^{15} - 12q^{18} + 3q^{21} + O(q^{24}). \tag{6.5.6}$$

As a sanity check, we recognized $\phi_s^*(f_{277})$ using modular symbols as a classical modular form of weight 4 and level $3 \cdot 277$ to order $O(q^{400})$. We then compute

$$\phi_s^*(f_{277} | T_2) = 6q^3 - 12q^6 - 12q^9 - 6q^{12} - 6q^{15} + 24q^{18} - 6q^{21} + O(q^{24}) \tag{6.5.7}$$

so quite convincingly, $a_2(f_{277}) = -2$, in agreement with (6.2.3).

To compute the action of Hecke operators on the specialized expansion (6.5.2), we work (to a finite degree of q -adic precision) with coefficients over \mathbb{C} or over $\mathbb{Z}/m\mathbb{Z}$ with m suitably large — we consider these two approaches in turn in the next two sections.

6.6. Over floating point complex numbers. We may also compute $a_p(f)$ via (6.5.2) over the complex numbers using interval arithmetic.

Example 6.6.1. We perform our Hecke computation with in-house C++ code. Continuing with $f = f_{277}$ as in Example 6.5.5, for $p = 2$ we work with 512 bits of precision: the upper size encountered was $3.40282 \cdot 10^{38}$ and the lower size was $2.9387 \cdot 10^{-39}$, giving

$$a_2(f) = \frac{\phi_s^*(f | T_2)}{\phi_s^*(f)} \equiv \frac{6q^3 + O(q^5)}{-3q^3 + O(q^4)} = -2 + O(q)$$

up to an error 10^{-75} under a second on a standard desktop CPU. The largest computation required for this f was $a_{43}(f) = 4$; with the same bit precision and maximum error smaller than 10^{-40} , it took less than 90 minutes.

Remark 6.6.2. Given the first few Dirichlet coefficients of an L -function in the Selberg class with specified conductor and Γ -factors, Farmer, Koutsoliotas, and Lemurell [≥ 2019] can (in principle) rigorously compute complex approximations to the next few Dirichlet coefficients using just the approximate functional equation. This method is practical for small examples — and it is especially useful when the L -function is of unknown, speculative, or otherwise complicated origin. Prolonging an initial L -series is a possible avenue for extending the range of examples of modularity proven in this article.

6.7. Expansion over a finite field. As an alternative to complex expansion, we may also work in a finite ring. To do so, we need the following archimedean information about the Hecke eigenvalue.

Proposition 6.7.1. *Let $f \in S_k(K(N))$ be an eigenform for the Hecke operators $T(p), T_1(p^2)$ with eigenvalues $a_p(f), a_{1,p^2}(f) \in \mathbb{C}$ where $p \nmid N$. Then*

$$|a_p(f)| \leq p^{k-3}(1+p)(1+p^2), \quad |a_{1,p^2}(f)| \leq p^{2k-6}(1+p)(1+p^2)p. \tag{6.7.2}$$

Proof. By an elementary estimate, there exists a $B > 0$ such that $|a(T; f)| \leq B \det(T)^{k/2}$ for all T . Clearly $B = \sup_{T>0} |a(T; f)| \det(T)^{-k/2}$ is optimal. By (4.2.9), we have

$$\begin{aligned} |a_p(f)||a(T; f)| &= |a(T; f | T(p))| \\ &\leq |a(pT; f)| + p^{k-2} \sum_{j \bmod p} |a(\frac{1}{p}T \begin{bmatrix} 1 & 0 \\ j & p \end{bmatrix}; f)| + p^{k-2} |a(\frac{1}{p}T \begin{bmatrix} p & 0 \\ 0 & 1 \end{bmatrix}; f)| + p^{2k-3} |a(\frac{1}{p}T; f)| \\ &\leq Bp^k \det(T)^{k/2} + Bp^{k-1} \det(T)^{k/2} + Bp^{k-2} \det(T)^{k/2} + Bp^{k-3} \det(T)^{k/2}. \end{aligned}$$

From the equation $|a_p(f)||a(T; f)| \det(T)^{-k/2} \leq B(p^k + p^{k-1} + p^{k-2} + p^{k-3})$, we obtain the desired result by taking the supremum over $T > 0$.

A similar argument shows the inequality for $a_{1,p^2}(f)$. □

If $a \in \mathbb{Z}$ and $|a| < C$, then we can recover $a \in \mathbb{Z}$ from its congruence class modulo m whenever $m > 2C$. For our purposes, we might as well work with a *prime* modulus m , and indeed, because of the needed p -th roots of unity, we choose a large prime m such that $m \equiv 1 \pmod{p}$ and work in $R = \mathbb{Z}[\zeta_p]/\mathfrak{m}$ where \mathfrak{m} is a fixed choice of split prime above m , and we compute the expansion (6.5.2) in $R[[q]]$ as

$$a_p(f) \equiv \frac{1}{\phi_s^*(f)} \sum_j \phi_s^*(f | M_j) \pmod{\mathfrak{m}}$$

and then lift the result to $\mathbb{Z} \subseteq \mathbb{Z}[\zeta_p]$. The computational benefit is that we may replace ζ_p by an integer and compute modulo m .

Example 6.7.3. Let $f_{587}^- \in S_2(K(587))^-$ be the Borchers product defined in (6.2.8). We choose $t_0 = \begin{pmatrix} 4 & -137/2 \\ -137/2 & 1174 \end{pmatrix}$ and have $a(t_0, f) = -1$. We used $s = \begin{pmatrix} 2348 & 137 \\ 137 & 8 \end{pmatrix}$ and target exponent $e = \text{tr}(st_0) = 15$. We used the finite field method in our computations, which required a choice of a prime modulus m and an integer γ such that $\gamma \not\equiv 1 \pmod{m}$ and $\gamma^p \equiv 1 \pmod{m}$. The modulus m must be chosen large enough so that $m > \lfloor 2C \rfloor$ where $C = p^2(1 + \frac{1}{p})(1 + \frac{1}{p^2})$ from Proposition 6.7.1. The code was written in C++ using FLINT for operations of polynomials in one variable modulo an integer, and the computation of the restriction method to compute $a_{41}(f_{587}^-)$ took less than 2 hours on a typical CPU. The computation of $a_{1,p^2}(f)$ for $p \leq 11$ took just a few minutes.

7. Verifying paramodularity

In this section, we carry out the Faltings–Serre method for our case of interest $G = \mathrm{GSp}_4$ and $\ell = 2$, proving our main Theorem 1.2.1 as well as the other two advertised cases. We employ the conventions and notation of Section 4, in particular for Galois representations and L -functions.

7.1. The case $N = 277$. Let $X = X_{277}$ be the smooth projective curve over \mathbb{Q} given by the equation

$$X : y^2 + (x^3 + x^2 + x + 1)y = -x^2 - x \tag{7.1.1}$$

with LMFDB label 277.a.277.1, or equivalently by

$$y^2 + y = x^5 - 2x^3 + 2x^2 - x. \tag{7.1.2}$$

Both models are minimal with discriminant $\Delta = 277$. Let $A = A_{277} = \mathrm{Jac} X_{277}$ be the Jacobian of X_{277} , a principally polarized abelian surface over \mathbb{Q} of conductor 277. Let $f = f_{277} \in S_2(K(277))$ be the Siegel modular form of weight 2 constructed in (6.2.2).

Our main result (implying Theorem 1.2.1) is as follows.

Theorem 7.1.3. *For all primes p , we have $L_p(A_{277}, T) = Q_p(f_{277}, T)$. In particular, we have $L(A_{277}, s) = L(f_{277}, s, \mathrm{spin})$ and the abelian surface A_{277} is paramodular.*

To ease notation, we now dispense with subscripts. To prove this theorem, we use the strategy described in Section 3.2, with the further practical improvements from Section 3.3. Attached to A by (4.1.3) and to f by Theorem 4.3.4 and by the remarks afterward are 2-adic Galois representations

$$\rho_A, \rho_f : \mathrm{Gal}_{\mathbb{Q}, S} \rightarrow \mathrm{GSp}_4(\mathbb{Q}_2^{\mathrm{al}})$$

where $S = \{2, 277, \infty\}$ such that $\det \rho_A = \det \rho_f = \chi_2^2$ the square of the 2-adic cyclotomic character. Our first task is to verify equivalence of residual representations. We start with Lemma 4.3.8(a), which allows us to conclude that the residual representations $\bar{\rho}_A^{\mathrm{ss}}, \bar{\rho}_f^{\mathrm{ss}} : \mathrm{Gal}_{\mathbb{Q}, S} \rightarrow \mathrm{GSp}_4(\mathbb{F}_2)$ take values in \mathbb{F}_2 .

Lemma 7.1.4. *The residual representations $\bar{\rho}_A, \bar{\rho}_f : \mathrm{Gal}_{\mathbb{Q}, S} \rightarrow \mathrm{GSp}_4(\mathbb{F}_2)$ are equivalent and have absolutely irreducible image $S_5(b)$.*

Proof. We apply Algorithm 2.2.3. The representation $\bar{\rho}_A$ is given by the action on $A[2]$; completing the square in (7.1.2) to obtain the model $y^2 = g(x) = 4x^5 - 8x^3 + 8x^2 - 4x + 1$ we obtain $\bar{\rho}_A$ via the action on the roots of $g(x)$, which we verify is isomorphic to $G = S_5(b)$ as the elements of order 3 have trace 1 by (5.1.8). As implied by the general theory, the field $\mathbb{Q}(A[2])$ is ramified only at 2, 277.

For $\bar{\rho}_f$, we only have indirect access to the Galois representation. By (6.2.4), we have

$$\det(1 - \bar{\rho}_f(\mathrm{Frob}_3)T) = 1 + T + T^2 + T^3 + T^4 \in \mathbb{F}_2[T],$$

so $\mathrm{img} \bar{\rho}_f$ contains an element of order 5. Similarly Frob_5 has order divisible by 3, so $\mathrm{img} \bar{\rho}_f$ is isomorphic to one of A_5, S_5, A_6, S_6 . Therefore the fixed field under $\ker \bar{\rho}_f$ is the splitting field of an irreducible, separable polynomial $g(x)$ of degree 5 or 6. Let $F := \mathbb{Q}[x]/(g(x))$; then F is unramified away from 2, 277.

But we know a bit more: by Lemma 4.3.10, the 277-valuation of the Artin conductor of $\bar{\rho}_f$ is at most 1, so $\text{ord}_{277}(d_F) \leq 1$. A Hunter search, or looking up the possible fields in the database of Jones and Roberts [2014], shows that there are no such degree 6 polynomials, and exactly two polynomials of degree 5, namely $x^5 - x^4 + 2x^2 - x + 1$ and $x^5 - x^4 + 4x^3 + 5x - 1$. Both polynomials have the same Galois closure, with Galois group S_5 ; we need to distinguish the representations afforded by the inclusion $S_5 \subseteq S_6$ and the fixed representation (5.1.1). We refer to (5.1.8): for the second one Frob_3 does not have order 5, so we must have a match with the representation afforded by the first one. \square

With Lemma 7.1.4 in hand, we apply Lemma 4.3.8(b) to conclude that our 2-adic representations descend to $\rho_A, \rho_f : \text{Gal}_{\mathbb{Q},S} \rightarrow \text{GSp}_4(\mathbb{Z}_2)$. We now finish the proof of the theorem.

Proof of Theorem 7.1.3. We apply Algorithm 2.4.1. Step 1 was done in Lemma 7.1.4, and the residual representations have a common image

$$G := \text{img } \bar{\rho} \leq \text{GSp}_4(\mathbb{F}_2) = \text{Sp}_4(\mathbb{F}_2)$$

with $G \simeq S_5(b)$. Let K be the fixed field under $\ker \bar{\rho}$, so $\text{Gal}(K | \mathbb{Q}) \simeq G$ under $\bar{\rho}$.

Using Theorem 5.3.3, we now find all obstructing extension groups E , an exact core-free subgroup $D \leq E$, and a list of conjugacy classes of obstructing elements. We refer to the field diagram (5.3.2). The extension $K_0 = K^H$ has degree 10, explicitly it is given by adjoining a root of the polynomial

$$x^{10} + 3x^9 + x^8 - 10x^7 - 17x^6 - 7x^5 + 11x^4 + 18x^3 + 13x^2 + 5x + 1.$$

The possible obstructing extensions $\varphi : \text{Gal}(L | \mathbb{Q}) \hookrightarrow E$ are obtained as the Galois closure of the quadratic extension $L_0 \supseteq K_0$, still unramified away from S so they may be constructed using class field theory: we find there are 4095 quadratic extensions $L_0 \supseteq K_0$ unramified away from S . To write down polynomials (not necessarily small) that represent these fields takes about 5 minutes; as we developed the algorithm, we found it convenient to optimize these polynomials (using polredabs), which took about 6 hours. In the course of the algorithm we consider 24062 obstructing pairs (L, φ) .

For each such obstructing pair (L, φ) , we compute a small prime $p \neq 2, 277$ such that the conjugacy class of Frob_p is obstructing, according to the stages of Section 3.3. Computing obstructing primes by their L_0 -cycle type as in Step 4', we obtain the list of primes $\{3, 5, 7, 11, 13, 17, 19, 23, 29, 31, 37, 41, 43, 53\}$; going a bit further, considering obstructing primes by the pair of L_0, K_0 -cycle type as in Step 4'', we manage only to remove the prime $p = 53$ from the list (but reduce the sizes of primes in many cases), so we refine the list of primes to those with $p \leq 43$. The total running time for this step was about 90 minutes on a standard CPU.

There are 8 pairs (L, φ) that require $p = 53$. The field L_0 generated by a root of

$$\begin{aligned} x^{20} + 121x^{18} + 7459x^{16} + 286418x^{14} + 7324711x^{12} + 126372663x^{10} + 1387797423x^8 \\ + 7013797890x^6 - 30031807329x^4 - 582846604659x^2 - 1630793025157 \end{aligned}$$

has Galois closure L with $\text{Gal}(L | \mathbb{Q}) \simeq E \leq \text{sp}_4(\mathbb{F}_2) \rtimes G$ with $\#E = 2^{10}5!$. There are four outer automorphisms ξ , and with respect to one of these, we find that Frob_5 is an obstructing conjugacy class based on the L_0, K_0 -cycle type pair $6^3 1^2, 6^1 3^1 1^1$ but Frob_{53} is the first obstructing prime based *only* on the L_0 -cycle type $8^1 4^2 2^2$ (and this cycle type works for all four ξ).

We are now in Step 5 of the algorithm, and to conclude we will show that $\text{tr } \rho_A(\text{Frob}_p) = \text{tr } \rho_f(\text{Frob}_p)$ for all $p \leq 43$. The former traces can be done by counting points, the latter traces were computed using the method in Example 6.6.1, and we check that they are equal, completing the proof. (In fact, we went further than necessary and checked the equality of traces for all $p \leq 97$.) \square

7.2. The case $N = 353$. We now turn to a case with residual image $S_3 \wr C_2$. Let $X = X_{353}$ be the genus 2 curve with LMFDB label 353.a.353.1 defined by

$$X : y^2 + (x^3 + x + 1)y = x^2$$

and $A = A_{353} = \text{Jac } X$, a typical abelian surface of conductor 353. Let $f = f_{353} \in S_2(K(353))$ be the paramodular form constructed in (6.2.6).

Theorem 7.2.1. *For all primes p , we have $L_p(A_{353}, T) = Q_p(f_{353}, T)$. In particular, $L(A, s) = L(f_{353}, s, \text{spin})$ and the abelian surface A_{353} is paramodular.*

Proof. The proof is similar to that of Theorem 7.1.3, but with some slightly different arguments. To supplement the data (6.2.7), we compute $a_p(f), a_{1,p^2}(f)$ for $p \leq 11$, and counting points yields equality of the additional Euler factors

$$\begin{aligned} L_5(A, T) &= Q_5(f, T) = 1 - T + 2T^2 - 5T^3 + 25T^4, \\ L_7(A, T) &= Q_7(f, T) = 1 - 6T^2 + 49T^4, \\ L_{11}(A, T) &= Q_{11}(f, T) = 1 - 2T + T^2 - 22T^3 + 121T^4. \end{aligned} \tag{7.2.2}$$

Our first task is to verify that the mod 2 representations $\bar{\rho}_A$ and $\bar{\rho}_f$ are equivalent and absolutely irreducible. For A , we find the 2-torsion field generated by the splitting field of the polynomial $x^6 + 2x^4 + 2x^3 + 5x^2 + 2x + 1$ and Galois group $S_3 \wr C_2$.

Let K be the fixed field of $\ker \bar{\rho}_f$ and $G := \text{Gal}(K | \mathbb{Q})$. Since $L_3(A, T) \equiv 1 + T + T^3 + T^4 \pmod{2}$ we see that G has an element of order 3 or 6 with trace 0. Since $L_{11}(A, T) \equiv 1 + T^2 + T^4$, we see G has an element of order 3 or 6 with trace 1. Squaring such elements preserves their trace, so G contains elements of order 3 with either trace. Thus $G \leq S_6$ has an element with cycle decomposition 3^1 and one with cycle decomposition 3^2 . Listing all subgroups of S_6 with this property, we see that G must be isomorphic to one of the permutation groups

$$C_3^2, \quad C_3 : S_3, \quad C_3 \times S_3 \text{ (twice)}, \quad C_3 : S_3 \cdot C_2, \quad S_3^2 \text{ (twice)}, \quad S_3 \wr C_2, \quad A_6, \quad S_6.$$

The subgroups in this list that are intransitive are $C_3^2, C_3 : S_3, C_3 \times S_3, S_3^2$. The groups $C_3^2, C_3 \times S_3$ have C_3 as a quotient, and by the Kronecker–Weber theorem there are no C_3 -extensions unramified outside 2 and 353 since $353 \equiv 2 \pmod{3}$. The groups $C_3 : S_3$ and S_3^2 have as quotient S_3 , but there is a

unique S_3 extension ramified only at 2 and 353 (verified by a class field calculation and the Jones and Roberts database [2014]) defined by $x^3 - x^2 - 6x + 14$, and we compute that there are no cyclic cubic extensions of this field unramified away from primes dividing 2, 353. This leaves the transitive groups $C_3 : S_3 \cdot C_2$, $S_3 \wr C_2$, A_6 , S_6 arising as the normal closure of a degree 6 subfield K' . If $G = C_3 : S_3 \cdot C_2$, then as in the proof of Proposition 5.2.4, we have $\text{ord}_{353} d_{K'} = 0, 1, 3$ but if $\text{ord}_{353} d_{K'} = 3$ then G contains an element with cycle structure 2^3 , a contradiction. Combined with Proposition 5.2.4 in the remaining cases, we have $\text{ord}_{353} d_{K'} \leq 1$. Again by consulting the Jones and Roberts database [2014], we find exactly two candidates, the extensions defined by $x^6 - 2x^5 + 2x^4 - x^2 + 1$ and $x^6 - 2x^5 - 3x^4 + 4x^3 + x^2 - 6x + 1$. In the first extension, Frob_3 has order 6 contradicting $Q_3(f, T) \equiv 1 + T^4 \pmod{2}$, so we have the latter, and G is isomorphic to $S_3 \wr C_2$. Finally, since the trace of $\bar{\rho}_f(\text{Frob}_3)$ equals that of A , we see that the two residual images are isomorphic and absolutely irreducible (recall that there are two embeddings of $S_3 \wr C_2$ into $\text{GSp}_4(\mathbb{F}_2)$ up to inner automorphisms, and they differ in the trace of order 3 and 6 elements).

Next, using Theorem 5.3.3 we compute the extension K_0 corresponding to the core-free subgroup C_2^2 , defined by

$$x^{18} - 10x^{14} + 3x^{12} + 25x^{10} - 5x^8 - 19x^6 + 5x^2 + 1. \quad (7.2.3)$$

Using computational class field theory, we list all quadratic extensions $L_0 \supseteq K_0$ unramified away from primes above 2, 353. We find that there are 65535 such extensions. For each extension, we find an obstructing element; after computing for just over 5 hours on a standard CPU (about 0.2 seconds per field) we find the list of primes

$$\{3, 5, 7, 11, 13, 19, 23, 29, 31, 37, 41, 43, 53, 97, 137\}. \quad (7.2.4)$$

(The prime $p = 181$ arose from 2 extensions L_0 and 4 maps φ each looking only at cycle types, but by identifying the precise conjugacy classes we find obstructing classes for $p = 5, 137$.)

To conclude, using the floating point algorithm we compute $\text{tr } \rho_f(\text{Frob}_p)$ for all primes $p \leq 109$ as well as the primes $p = 137, 139, 251$ (for robustness) in 29 hours on a standard CPU, and we see they agree with the traces obtained from point counts on X , completing the proof. \square

Example 7.2.5. We pause to consider an extreme example where the refinement in Section 3.3 provides a significant improvement. Consider the extension defined by adjoining a square root of the element

$$-430a^{16} + 302a^{14} + 3956a^{12} - 3904a^{10} - 6944a^8 + 5348a^6 + 3628a^4 - 1454a^2 - 510$$

where a is a root of (7.2.3), the defining polynomial for K_0 .

There are 4 outer automorphisms giving rise to possible maps φ : but in fact, we will see below that only 2 of these maps extend $\bar{\rho}$, which is to say the other 2 do not preserve the residual representation. If we only consider cycle types that obstruct all 4 possible maps φ as in Step 4', we have the types $8^4 2^2$, $4^6 2^2 1^8$, $4^2 2^{10} 1^8$. For one of these 4 extensions, the smallest prime p with this cycle type is $p = 251$. If we push further in this extension, and look at the L_0 -cycle type and the order in K_0 , we compute that $p = 101$ works. Going even further and using L_0 , K_0 -cycle type, we find that $p = 11$ works!

7.3. The case $N = 587$. We conclude with one final case. Let $X = X_{587}$ be the genus 2 curve with LMFDB label 587.a.587.1 defined by

$$X : y^2 + (x^3 + x + 1)y = -x^2 - x$$

and $A = A_{587} = \text{Jac } X_{587}$, a typical abelian surface of conductor 587 and rank 1. Let $f = f_{587}^- \in S_2(K(587))$ be the paramodular form constructed using (6.2.9).

Theorem 7.3.1. *For all primes p we have $L_p(A_{587}, T) = Q_p(f_{587}^-, T)$, and A_{587} is paramodular.*

Proof. We first verify that the mod 2 representations $\bar{\rho}_A$ and $\bar{\rho}_f$ are equivalent and absolutely irreducible. For A , we find the 2-torsion field generated by the splitting field of the polynomial $x^6 - 2x^5 + 2x^4 - x^2 + 2x - 1$ with Galois group $G = S_6$. For f , we have

$$Q_3(f, T) = 1 + 4T + 9T^2 + 12T^3 + 9T^4 \equiv 1 + T^2 + T^4 \pmod{2}$$

and

$$Q_{11}(f, T) = 1 + T - T^2 + 11T^3 + 121T^4 \equiv 1 + T + T^2 + T^3 + T^4 \pmod{2}$$

by [Poor and Yuen 2007, Table 5] and Example 6.7.3. In particular, the residual image has order divisible by 3 and 5.

The subgroups of S_6 (up to isomorphisms) of order divisible by 15 are

$$A_5, S_5, A_6, S_6.$$

In all cases, there exists a polynomial of degree 5 or 6 unramified outside $\{2, 587\}$ and we can choose them such that the discriminant valuation is at most 1 at 587 by Proposition 5.2.4. By [Jones and Roberts 2014] there are only two degree 5 polynomials with field discriminant having valuation 1 at 587, namely: $x^5 - x^3 - x - 2$ and $x^5 + 2x^3 - 8x^2 - 13x - 8$ and two degree 6 polynomials with field discriminant having valuation 1 at 587: $x^6 - 2x^5 + 2x^4 - x^2 + 2x - 1$ and $x^6 - 2x^5 + 3x^4 + 4x^3 - 2x^2 - 4x + 2$. For the degree 5 polynomials, the first field has Frob_3 of order 4 (then it would have even trace) while Frob_{11} has order 2 in the second field. Regarding the degree six ones, in the second extension Frob_{11} has order 2, but odd trace in A . We deduce that the residual representation of f_{587}^- corresponds then to the same extension as A , and since both representations have the same trace at Frob_3 , we deduce that they are indeed equivalent and absolutely irreducible.

By Theorem 5.3.3 we are led to compute all quadratic extensions of the degree 20 extension

$$x^{20} + x^{18} - 4x^{17} - 3x^{16} - 2x^{15} + 7x^{14} - 6x^{13} - 18x^{12} - 8x^{11} + 8x^{10} + 8x^9 - 18x^8 + 6x^7 + 7x^6 + 2x^5 - 3x^4 + 4x^3 + x^2 + 1. \quad (7.3.2)$$

We find that there are $2^{19} - 1 = 524287$ such extensions. Writing down minimal polynomials (not necessarily small) that represent these fields takes about 10 minutes; for convenience, we also computed optimized representatives, which took many CPU weeks.

Finding an obstructing element for each of them, we find the list of primes to verify:

$$\{3, 5, 7, 11, 13, 17, 19, 23, 29, 37, 41\}. \quad (7.3.3)$$

The total CPU time to compute this list of primes was about 2.5 hours (about 0.2 seconds per field). Finally, we computed the corresponding traces above and they match, completing the proof. \square

Acknowledgements

The authors would like to thank several people for helpful conversations: Frank Calegari, Jennifer Johnson-Leung, Kenneth Kramer, Chung Pang Mok, David P. Roberts (in particular for Proposition 5.2.4), Drew Sutherland, and Eric Urban (in particular for help with showing that the representation is symplectic in Theorem 4.3.4). We also thank Fordham University’s Academic Computing Environment for the use of its servers. Thanks also to the anonymous referees for their feedback. Pacetti was partially supported by PIP 2014–2016 11220130100073 and Voight was supported by an NSF CAREER Award (DMS-1151047) and a Simons Collaboration Grant (550029).

This large collaborative project was made possible by the generous support of several host institutes, to which we express our thanks: the Institute for Computational and Experimental Research in Mathematics (ICERM), the International Centre for Theoretical Physics (ICTP), and the Hausdorff Institute of Mathematics (HIM).

References

- [Achter 2005] J. D. Achter, “Detecting complex multiplication”, pp. 38–50 in *Computational aspects of algebraic curves* (Moscow, ID, 2005), edited by T. Shaska, Lecture Notes Ser. Comput. **13**, World Sci., Hackensack, NJ, 2005. MR Zbl
- [Berger and Klosin 2017] T. Berger and K. Klosin, “Deformations of Saito–Kurokawa type and the paramodular conjecture”, preprint, 2017. With an appendix by C. Poor, J. Shurman and D. S. Yuen. To appear in *Amer. J. Math.* arXiv
- [Berger et al. 2015] T. Berger, L. Dembele, A. Pacetti, and M. H. Sengun, “Theta lifts of Bianchi modular forms and applications to paramodularity”, *J. Lond. Math. Soc.* (2) **92**:2 (2015), 353–370. MR Zbl
- [Booker et al. 2016] A. R. Booker, J. Sijsling, A. V. Sutherland, J. Voight, and D. Yasaki, “Sato–Tate groups and modularity for atypical genus 2 curves”, preprint, 2016.
- [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
- [Boxer et al. 2018] G. Boxer, F. Calegari, T. Gee, and V. Pilloni, “Abelian surfaces over totally real fields are potentially modular”, preprint, 2018. arXiv
- [Breuil et al. 2001] C. Breuil, B. Conrad, F. Diamond, and R. Taylor, “On the modularity of elliptic curves over \mathbb{Q} : wild 3-adic exercises”, *J. Amer. Math. Soc.* **14**:4 (2001), 843–939. MR Zbl
- [Brumer and Kramer 2014] A. Brumer and K. Kramer, “Paramodular abelian varieties of odd conductor”, *Trans. Amer. Math. Soc.* **366**:5 (2014), 2463–2516. MR Zbl
- [Brumer and Kramer 2018] A. Brumer and K. Kramer, “Certain abelian varieties bad at only one prime”, *Algebra Number Theory* **12**:5 (2018), 1027–1071. MR Zbl
- [Brumer and Kramer 2019] A. Brumer and K. Kramer, “Corrigendum to “Paramodular abelian varieties of odd conductor””, *Trans. Amer. Math. Soc.* (online publication May 2019).
- [Calegari and Geraghty 2016] F. Calegari and D. Geraghty, “Minimal modularity lifting for non-regular symplectic representations”, submitted, 2016, Available at <https://tinyurl.com/siegcleg>.

- [Carayol 1994] H. Carayol, “Formes modulaires et représentations galoisiennes à valeurs dans un anneau local complet”, pp. 213–237 in *p-adic monodromy and the Birch and Swinnerton–Dyer conjecture* (Boston, 1991), edited by B. Mazur and G. Stevens, Contemp. Math. **165**, Amer. Math. Soc., Providence, RI, 1994. MR Zbl
- [Cohen 1993] H. Cohen, *A course in computational algebraic number theory*, Graduate Texts in Math. **138**, Springer, 1993. MR Zbl
- [Cohen 2000] H. Cohen, *Advanced topics in computational number theory*, Graduate Texts in Math. **193**, Springer, 2000. MR Zbl
- [Colman et al. 2019] O. Colman, A. Ghitza, and N. C. Ryan, “Analytic evaluation of Hecke eigenvalues for Siegel modular forms of degree two”, pp. 207–220 in *Proceedings of the thirteenth Algorithmic Number Theory Symposium* (Madison, WI, 2018), edited by R. Scheidler and J. Sorenson, Open Book Series **2**, MSP, Berkeley, CA, 2019.
- [Dembélé and Kumar 2016] L. Dembélé and A. Kumar, “Examples of abelian surfaces with everywhere good reduction”, *Math. Ann.* **364**:3-4 (2016), 1365–1392. MR Zbl
- [Dieulefait et al. 2010] L. Dieulefait, L. Guerberoff, and A. Pacetti, “Proving modularity for a given elliptic curve over an imaginary quadratic field”, *Math. Comp.* **79**:270 (2010), 1145–1170. MR Zbl
- [Dokchitser and Dokchitser 2013] T. Dokchitser and V. Dokchitser, “Identifying Frobenius elements in Galois groups”, *Algebra Number Theory* **7**:6 (2013), 1325–1352. MR Zbl
- [Eichler and Zagier 1985] M. Eichler and D. Zagier, *The theory of Jacobi forms*, Progr. Math. **55**, Birkhäuser, Boston, 1985. MR Zbl
- [Farmer et al. ≥ 2019] D. Farmer, S. Koutsoliotas, and S. Lemurell, “ L -functions with rational integer coefficients, I: Degree 4 and weight 0”, preprint.
- [Fontaine 1985] J.-M. Fontaine, “Il n’y a pas de variété abélienne sur \mathbb{Z} ”, *Invent. Math.* **81**:3 (1985), 515–538. MR Zbl
- [Freitag 1983] E. Freitag, *Siegelsche Modulfunktionen*, Grundlehren der Math. Wissenschaften **254**, Springer, 1983. MR Zbl
- [Gouvêa 2001] F. Q. Gouvêa, “Deformations of Galois representations”, pp. 233–406 in *Arithmetic algebraic geometry* (Park City, UT, 1989), edited by B. Conrad and K. Rubin, IAS/Park City Math. Ser. **9**, Amer. Math. Soc., Providence, RI, 2001. MR Zbl
- [Grenié 2007] L. Grenié, “Comparison of semi-simplifications of Galois representations”, *J. Algebra* **316**:2 (2007), 608–618. MR Zbl
- [Gritsenko 1995] V. Gritsenko, “Arithmetical lifting and its applications”, pp. 103–126 in *Number theory* (Paris, 1992–1993), edited by S. David, London Math. Soc. Lecture Note Ser. **215**, Cambridge Univ. Press, 1995. MR Zbl
- [Gritsenko et al. 2018] V. Gritsenko, N.-P. Skoruppa, and D. Zagier, “Theta blocks”, preprint, 2018, Available at <https://tinyurl.com/thetablo>.
- [Gritsenko et al. 2019] V. Gritsenko, C. Poor, and D. S. Yuen, “Antisymmetric paramodular forms of weights 2 and 3”, *Int. Math. Res. Not.* (online publication February 2019).
- [Harvey 2014] D. Harvey, “Counting points on hyperelliptic curves in average polynomial time”, *Ann. of Math. (2)* **179**:2 (2014), 783–803. MR Zbl
- [Harvey and Sutherland 2016] D. Harvey and A. V. Sutherland, “Computing Hasse–Witt matrices of hyperelliptic curves in average polynomial time, II”, pp. 127–147 in *Frobenius distributions: Lang–Trotter and Sato–Tate conjectures*, edited by D. Kohel and I. Shparlinski, Contemp. Math. **663**, Amer. Math. Soc., Providence, RI, 2016. MR Zbl
- [Howard et al. 2008] B. Howard, J. Millson, A. Snowden, and R. Vakil, “A description of the outer automorphism of S_6 , and the invariants of six points in projective space”, *J. Combin. Theory Ser. A* **115**:7 (2008), 1296–1303. MR Zbl
- [Johnson-Leung and Roberts 2012] J. Johnson-Leung and B. Roberts, “Siegel modular forms of degree two attached to Hilbert modular forms”, *J. Number Theory* **132**:4 (2012), 543–564. MR Zbl
- [Johnson-Leung and Roberts 2014] J. Johnson-Leung and B. Roberts, “Twisting of paramodular vectors”, *Int. J. Number Theory* **10**:4 (2014), 1043–1065. MR Zbl
- [Johnson-Leung and Roberts 2017] J. Johnson-Leung and B. Roberts, “Twisting of Siegel paramodular forms”, *Int. J. Number Theory* **13**:7 (2017), 1755–1854. MR Zbl

- [Jones and Roberts 2014] J. W. Jones and D. P. Roberts, “A database of number fields”, *LMS J. Comput. Math.* **17**:1 (2014), 595–618. MR Zbl
- [Jorza 2012] A. Jorza, “ p -adic families and Galois representations for $\mathrm{GSp}(4)$ and $\mathrm{GL}(2)$ ”, *Math. Res. Lett.* **19**:5 (2012), 987–996. MR Zbl
- [Khare and Wintenberger 2009a] C. Khare and J.-P. Wintenberger, “Serre’s modularity conjecture, I”, *Invent. Math.* **178**:3 (2009), 485–504. MR Zbl
- [Khare and Wintenberger 2009b] C. Khare and J.-P. Wintenberger, “Serre’s modularity conjecture, II”, *Invent. Math.* **178**:3 (2009), 505–586. MR Zbl
- [Laumon 2005] G. Laumon, “Fonctions zetas des varietes de Siegel de dimension trois”, pp. 1–66 in *Formes automorphes, II: Le cas du groupe $\mathrm{GSp}(4)$* , edited by J. Tilouine et al., Asterisque **302**, Soc. Math. France, Paris, 2005. MR Zbl
- [Livne 1987] R. Livne, “Cubic exponential sums and Galois representations”, pp. 247–261 in *Current trends in arithmetical algebraic geometry* (Arcata, CA, 1985), edited by K. A. Ribet, Contemp. Math. **67**, Amer. Math. Soc., Providence, RI, 1987. MR Zbl
- [Mok 2014] C. P. Mok, “Galois representations attached to automorphic forms on GL_2 over CM fields”, *Compos. Math.* **150**:4 (2014), 523–567. MR Zbl
- [Pilloni 2012] V. Pilloni, “Modularite, formes de Siegel et surfaces abeliennes”, *J. Reine Angew. Math.* **666** (2012), 35–82. MR Zbl
- [Poor and Yuen 2007] C. Poor and D. S. Yuen, “Computations of spaces of Siegel modular cusp forms”, *J. Math. Soc. Japan* **59**:1 (2007), 185–222. MR Zbl
- [Poor and Yuen 2015] C. Poor and D. S. Yuen, “Paramodular cusp forms”, *Math. Comp.* **84**:293 (2015), 1401–1438. MR Zbl
- [Poor et al. 2018] C. Poor, J. Shurman, and D. S. Yuen, “Nonlift weight two paramodular eigenform constructions”, preprint, 2018. arXiv
- [Ribet 1975] K. A. Ribet, “Endomorphisms of semi-stable abelian varieties over number fields”, *Ann. Math. (2)* **101** (1975), 555–562. MR Zbl
- [Ribet 1992] K. A. Ribet, “Abelian varieties over \mathbb{Q} and modular forms”, pp. 53–79 in *Algebra and topology* (Taejon, 1992), edited by S. G. Hahn and D. Y. Suh, Korea Adv. Inst. Sci. Tech., Taejon, South Korea, 1992. MR Zbl
- [Roberts and Schmidt 2006] B. Roberts and R. Schmidt, “On modular forms for the paramodular groups”, pp. 334–364 in *Automorphic forms and zeta functions* (Tokyo, 2004), edited by S. Bocherer et al., World Sci., Hackensack, NJ, 2006. MR Zbl
- [Roberts and Schmidt 2007] B. Roberts and R. Schmidt, *Local newforms for $\mathrm{GSp}(4)$* , Lecture Notes in Math. **1918**, Springer, 2007. MR Zbl
- [Schmidt 2018] R. Schmidt, “Packet structure and paramodular forms”, *Trans. Amer. Math. Soc.* **370**:5 (2018), 3085–3112. MR Zbl
- [Schutt 2006] M. Schutt, “On the modularity of three Calabi–Yau threefolds with bad reduction at 11”, *Canad. Math. Bull.* **49**:2 (2006), 296–312. MR Zbl
- [Serre 1985] J.-P. Serre, “Resume des cours de 1984–1985”, pp. 85–90 in *Annuaire du College de France 1984–1985*, Imprimerie Nationale, Paris, 1985.
- [Serre 2018] J.-P. Serre, “On the mod p reduction of orthogonal representations”, pp. 527–540 in *Lie groups, geometry, and representation theory*, edited by V. G. Kac and V. L. Popov, Progr. Math. **326**, Birkhuser, Cham, 2018. MR
- [SGA 7_I 1972] A. Grothendieck, *Groupes de monodromie en geometrie algebrique, I: Exposes I–II, VI–IX* (Seminaire de Geometrie Algebrique du Bois Marie 1967–1969), Lecture Notes in Math. **288**, Springer, 1972. MR Zbl
- [Stoltenberg-Hansen and Tucker 1999] V. Stoltenberg-Hansen and J. V. Tucker, “Computable rings and fields”, pp. 363–447 in *Handbook of computability theory*, edited by E. R. Griffor, Stud. Logic Found. Math. **140**, North-Holland, Amsterdam, 1999. MR Zbl
- [Taylor 1991] R. Taylor, “Galois representations associated to Siegel modular forms of low weight”, *Duke Math. J.* **63**:2 (1991), 281–332. MR Zbl
- [Taylor and Wiles 1995] R. Taylor and A. Wiles, “Ring-theoretic properties of certain Hecke algebras”, *Ann. of Math. (2)* **141**:3 (1995), 553–572. MR Zbl

- [Taylor and Yoshida 2007] R. Taylor and T. Yoshida, “Compatibility of local and global Langlands correspondences”, *J. Amer. Math. Soc.* **20**:2 (2007), 467–493. MR Zbl
- [Tornaria 2018] G. Tornaria, Paramodularity code repository, 2018, Available at <https://gitlab.fing.edu.uy/tornaria/modularity>.
- [Weissauer 2005] R. Weissauer, “Four dimensional Galois representations”, pp. 67–150 in *Formes automorphes, II: Le cas du groupe $GS\mathrm{p}(4)$* , edited by J. Tilouine et al., Astérisque **302**, Soc. Math. France, Paris, 2005. MR Zbl
- [Wiles 1995] A. Wiles, “Modular elliptic curves and Fermat’s last theorem”, *Ann. of Math. (2)* **141**:3 (1995), 443–551. MR Zbl
- [Yoshida 1980] H. Yoshida, “Siegel’s modular forms and the arithmetic of quadratic forms”, *Invent. Math.* **60**:3 (1980), 193–248. MR Zbl
- [Yoshida 2007] H. Yoshida, “On generalization of the Shimura–Taniyama conjecture, I and II”, pp. 1–26 in *Siegel modular forms and abelian varieties* (Hamamatsu, Japan, 2007), edited by T. Ibukiyama, Ryushido, Kobe, 2007.

Communicated by Bjorn Poonen

Received 2018-07-06 Revised 2019-01-24 Accepted 2019-04-02

brumer@fordham.edu	<i>Department of Mathematics, Fordham University, Bronx, NY, United States</i>
apacetti@famaf.unc.edu.ar	<i>FAMAF-CIEM, Universidad Nacional de Córdoba, Argentina</i>
poor@fordham.edu	<i>Department of Mathematics, Fordham University, Bronx, NY, United States</i>
tornaria@cmat.edu.uy	<i>Centro de Matemática, Universidad de la República, Montevideo, Uruguay</i>
jvoight@gmail.com	<i>Department of Mathematics, Dartmouth College, Hanover, NH, United States</i>
yuen888@hawaii.edu	<i>Department of Mathematics, University of Hawaii, Honolulu, HI, United States</i>

Contragredient representations over local fields of positive characteristic

Wen-Wei Li

It was conjectured by Adams, Vogan and Prasad that under the local Langlands correspondence, the L -parameter of the contragredient representation equals that of the original representation composed with the Chevalley involution of the L -group. We verify a variant of their prediction for all connected reductive groups over local fields of positive characteristic, in terms of the local Langlands parametrization of A. Genestier and V. Lafforgue. We deduce this from a global result for cuspidal automorphic representations over function fields, which is in turn based on a description of the transposes of Lafforgue's excursion operators.

1. Introduction	1197
2. Review of representation theory	1203
3. Statement of a variant of the conjecture	1210
4. Overview of the global Langlands parametrization	1219
5. The transposes of excursion operators	1231
Acknowledgements	1240
References	1240

1. Introduction

Let F be a local field. Choose a separable closure $\bar{F}|F$ and let W_F be the Weil group of F . For a connected reductive F -group G , the *local Langlands conjecture* asserts the existence of a map

$$\text{LLC} : \Pi(G) \rightarrow \Phi(G)$$

where $\Pi(G)$ is the set of isomorphism classes of irreducible smooth representations π of $G(F)$ (or Harish-Chandra modules when F is archimedean), and $\Phi(G)$ is the set of \hat{G} -conjugacy classes of L -parameters $W_F \xrightarrow{\phi} {}^L G$. Here the representations and the L -groups are taken over \mathbb{C} , but we will soon switch to the setting of nonarchimedean F and ℓ -adic coefficients.

It is expected that the L -packets $\Pi_\phi := \text{LLC}^{-1}(\phi)$ are finite sets; if $\pi \in \Pi_\phi$, we say ϕ is the L -parameter of π . The local Langlands correspondence also predicates on the internal structure of Π_ϕ when ϕ is a *tempered* parameter; this requires additional structures as follows:

MSC2010: primary 11F70; secondary 11R58, 22E55.

Keywords: contragredient representation, function field, local Langlands conjecture.

- When G is quasisplit, choose a Whittaker datum $\mathfrak{w} = (U, \chi)$ of G , taken up to $G(F)$ -conjugacy, where $U \subset G$ is a maximal unipotent subgroup and χ is a generic smooth character of $U(F)$. The individual members of Π_ϕ are described in terms of

$$\mathcal{S}_\phi := Z_{\hat{G}}(\text{im}(\phi)), \quad \mathcal{S}_\phi := \pi_0(\mathcal{S}_\phi).$$

Specifically, to each $\pi \in \Pi_\phi$ one should be able to attach an irreducible representation ρ of the finite group \mathcal{S}_ϕ (up to isomorphism), such that a \mathfrak{w} -generic $\pi \in \Pi_\phi$ maps to $\rho = \mathbf{1}$.

- For nonsplit G , one needs to connect G to a quasisplit group by means of a pure inner twist, or more generally a rigid inner twist [Kaletha 2016b]; in parallel, the L -packets will extend across various inner forms of G . We refer to [loc. cit., §5.4] for a discussion in this generality.

One natural question is to describe various operations on $\Pi(G)$ in terms of L -parameters. Among them, we consider the *contragredient* $\check{\pi}$ of π . The question is thus:

$$\text{How is } \pi \mapsto \check{\pi} \text{ in } \Pi(G) \text{ reflected on } \Phi(G)?$$

Despite its immediate appearance, this question has not been considered in this generality until the independent work of Adams and Vogan [2016, Conjecture 1.1] and D. Prasad [2018, §4]. The answer hinges on the *Chevalley involution* ${}^L\theta$ on ${}^L G$ to be reviewed in Section 3.1.

Conjecture 1.1 (Adams and Vogan; Prasad). Let π be an irreducible smooth representation of $G(F)$.

- (1) If π has L -parameter ϕ , then $\check{\pi}$ has L -parameter ${}^L\theta \circ \phi$.
- (2) Assume for simplicity that G is quasisplit and fix a Whittaker datum \mathfrak{w} . If a tempered representation $\pi \in \Pi_\phi$ corresponds to an irreducible representation ρ of \mathcal{S}_ϕ , then $\check{\pi}$ corresponds to $(\rho \circ {}^L\theta)^\vee$ tensored with a character ξ of \mathcal{S}_ϕ .

To define ξ , we use the general recipe [Kaletha 2013, Lemma 4.1]:

$$\begin{array}{ccc} \pi_0(\mathcal{S}_\phi / Z_{\hat{G}}^{\text{Gal}(\bar{F}|F)}) & \longrightarrow & \ker[\mathrm{H}^1(\mathbf{W}_F, Z_{\hat{G}^{\text{sc}}}) \rightarrow \mathrm{H}^1(\mathbf{W}_F, Z_{\hat{G}})] \\ \uparrow & & \downarrow \\ \mathcal{S}_\phi & & \left(\frac{G^{\text{ad}}(F)}{\text{im}[G(F) \rightarrow G^{\text{ad}}(F)]} \right)^{\text{Pontryagin dual}} \end{array}$$

Let B be the Borel subgroup of G included in the Whittaker datum, and choose a maximal torus $T \subset B$. Take the $\kappa \in T^{\text{ad}}(F)$ acting as -1 on each \mathfrak{g}_α where α is any B -simple root. This furnishes the character ξ of \mathcal{S}_ϕ . When G is not quasisplit, we have to endow it with a pure or rigid inner twist alluded to above.

Conjecture 1.1 comprises two layers: the second one is due to Prasad [2018]. In this article, we will focus exclusively on the first layer.

A precondition of the Adams–Vogan–Prasad conjecture is the existence of a map $\Pi(G) \rightarrow \Phi(G)$, baptized the *Langlands parametrization*, which has been constructed for many groups in various ways:

- When F is archimedean, the local Langlands correspondence is Langlands’ paraphrase of Harish-Chandra’s works. The “first layer” of the Adams–Vogan–Prasad conjecture is established by Adams and Vogan [2016], and Kaletha [2013, Theorem 5.9] obtained the necessary refinement for the “second layer”.
- When F is nonarchimedean of characteristic zero and G is a symplectic or quasisplit orthogonal group, Kaletha [2013, Theorem 5.9, Corollary 5.10] verified the Adams–Vogan–Prasad conjecture in terms of Arthur’s *endoscopic classification* of representations, which offers the local Langlands correspondence for these groups.
- For nonarchimedean F and general G , Kaletha [2013, §6] also verified the conjecture for the depth-zero and epipelagic supercuspidal L -packets, constructed by DeBacker, Reeder and Kaletha using induction from open compact subgroups.

The aim of this article is to address the first layer of Conjecture 1.1 when F is a nonarchimedean local field of characteristic $p > 0$ and G is arbitrary, in terms of the Langlands parametrization $\Pi(G) \rightarrow \Phi(G)$ of A. Genestier and V. Lafforgue [2017]. Their method is based on the geometry of the moduli stack of *restricted chtoucas*, intimately related to the global Langlands parametrization of cuspidal automorphic representations by Lafforgue [2018]. Accordingly, our representations π will be realized on $\overline{\mathbb{Q}}_\ell$ -vector spaces, where ℓ is a prime number not equal to p , and the L -group ${}^L G$ is viewed as a $\overline{\mathbb{Q}}_\ell$ -group. As $\mathbb{C} \simeq \overline{\mathbb{Q}}_\ell$ as abstract fields, passing to $\overline{\mathbb{Q}}_\ell$ does not alter the smooth representation theory of $G(F)$. On the other hand, there are subtle issues such as the independence of ℓ in the Langlands parametrization, which we refer to [Lafforgue 2018, §12.2.4] for further discussions.

Our main local result is.

Theorem 1.2 (Theorem 3.2.2). *Let F be a nonarchimedean local field of characteristic $p > 0$ and G be a connected reductive F -group. Fix $\ell \neq p$ as above. If an irreducible smooth representation π of $G(F)$ has parameter $\phi \in \Phi(G)$ under the Langlands parametrization of Genestier and Lafforgue, then $\check{\pi}$ has parameter ${}^L\theta \circ \phi$.*

Remark 1.3. The prefix L for local parameters and local packets is dropped for the following reason. The parameters of Genestier and Lafforgue are always *semisimple* or completely reducible in the sense of Serre [2005]; in other words, the monodromy part of the Weil–Deligne parameter is trivial; see Lemma 2.4.4. As mentioned in [Genestier and Lafforgue 2017], one expects that their parameter is the semisimplification of the “true” L -parameter of π . Hence the packets Π_ϕ in question are larger than expected, and the Langlands parametrization we adopt is coarser, unless when ϕ does not factorize through any Levi ${}^L M \hookrightarrow {}^L G$, i.e., ϕ is *semisimple and elliptic*.

Our strategy is to reduce it into a global statement. Let \mathring{F} be a global field of characteristic $p > 0$, say $\mathring{F} = \mathbb{F}_q(X)$ for a geometrically irreducible smooth proper \mathbb{F}_q -curve X , and set $\mathbb{A} = \mathbb{A}_{\mathring{F}}$. Let G be a connected reductive \mathring{F} -group. Fix a level $N \subset X$, whence the corresponding congruence subgroup $K_N \subset G(\mathbb{A})$ and the Hecke algebra $C_c(K_N \backslash G(\mathbb{A}) / K_N; \overline{\mathbb{Q}}_\ell)$. Also fix a cocompact lattice Ξ in $A_G(\mathring{F}) \backslash A_G(\mathbb{A})$

where $A_G \subset G$ is the maximal central split torus. *Grosso modo*, the global Langlands parametrization in [Lafforgue 2018] is deduced from a commutative $\overline{\mathbb{Q}}_\ell$ -algebra \mathcal{B} acting on the Hecke module

$$H_{\{0\},1} := C_c^{\text{cusp}}(\text{Bun}_{G,N}(\mathbb{F}_q)/\Xi; \overline{\mathbb{Q}}_\ell) = \bigoplus_{\alpha \in \ker^1(\mathring{F}, G)} C_c^{\text{cusp}}(G_\alpha(\mathring{F}) \backslash G_\alpha(\mathbb{A})/K_N \Xi; \overline{\mathbb{Q}}_\ell)$$

of $\overline{\mathbb{Q}}_\ell$ -valued cusp forms, extended across pure inner forms indexed by $\ker^1(\mathring{F}, G)$ (finite in number). The algebra \mathcal{B} is generated by the *excursion operators* $S_{I,f,\vec{\gamma}}$. For any character $\nu : \mathcal{B} \rightarrow \overline{\mathbb{Q}}_\ell$ of algebras, denote by \mathfrak{H}_ν the generalized ν -eigenspace of $H_{\{0\},1}$. Then $H_{\{0\},1} = \bigoplus_\nu \mathfrak{H}_\nu$ as Hecke modules. Moreover, Lafforgue’s machinery of ${}^L G$ -pseudocharacters associates a semisimple L -parameter $\sigma : \text{Gal}(\overline{F}|F) \rightarrow {}^L G(\overline{\mathbb{Q}}_\ell)$ to ν . In fact ν is determined by σ , so that we may write $\mathfrak{H}_\sigma = \mathfrak{H}_\nu$.

There is an evident Hecke-invariant bilinear form on $H_{\{0\},1}$, namely the *integration pairing*

$$\langle h, h' \rangle := \sum_{\alpha \in \ker^1(\mathring{F}, G)} \int_{G_\alpha(\mathring{F}) \backslash G_\alpha(\mathbb{A})/\Xi} hh', \quad h, h' \in H_{\{0\},1},$$

with respect to some Haar measure on $G(\mathbb{A}) = G_\alpha(\mathbb{A})$ which is \mathbb{Q} -valued on compact open subgroups. It is nondegenerate as easily seen by passing to $\overline{\mathbb{Q}}_\ell \simeq \mathbb{C}$. Now comes our global theorem.

Theorem 1.4 (Theorem 3.3.2). *If σ, σ' are two semisimple L -parameters for G such that $\langle \cdot, \cdot \rangle$ is nontrivial on $\mathfrak{H}_\sigma \otimes \mathfrak{H}_{\sigma'}$, then $\sigma' = {}^L\theta \circ \sigma$ up to $\hat{G}(\overline{\mathbb{Q}}_\ell)$ -conjugacy.*

Our local-global argument runs by first reducing Theorem 1.2 to the case that π is integral supercuspidal such that ω_π has finite order when restricted to A_G ; this step makes use of the compatibility of Langlands parametrization with parabolic induction, as established in [Genestier and Lafforgue 2017]. The second step is to globalize π into a cuspidal automorphic representation $\hat{\pi}$ with a suitable global model of G and Ξ , satisfying $\hat{\pi}^{K_N} \neq \{0\}$. The subspaces \mathfrak{H}_σ of $H_{\{0\},1}$ might have isomorphic irreducible constituents in common, but upon modifying the automorphic realization, one can always assume that $\hat{\pi}^{K_N}$ lands in some \mathfrak{H}_σ . An application of Theorem 1.4 and the local-global compatibility of Langlands parametrization [Genestier and Lafforgue 2017] will conclude the proof.

The proof of Theorem 1.4 relies upon the determination of the transpose $S \mapsto S^*$ of excursion operators with respect to $\langle \cdot, \cdot \rangle$, namely the Lemma 5.3.3:

$$S_{I,f,\vec{\gamma}}^* = S_{I,f^\dagger,\vec{\gamma}^{-1}}$$

where $f \in \mathcal{O}(\hat{G} \backslash ({}^L G)^I // \hat{G})$, the finite set I and $\vec{\gamma} \in \text{Gal}(\overline{F}|F)^I$ are the data defining excursion operators, and $f^\dagger(\vec{g}) = f({}^L\theta(\vec{g})^{-1})$ for $\vec{g} \in ({}^L G)^I$. This property entails that if $\nu : \mathcal{B} \rightarrow \overline{\mathbb{Q}}_\ell$ corresponds to σ , then $\nu^* : S \mapsto \nu(S^*)$ corresponds to ${}^L\theta \circ \sigma$ (Proposition 5.3.4).

The starting point of the computation of the transpose is the fact that $\langle \cdot, \cdot \rangle$ is of geometric origin: it stems from the Verdier duality on the moduli stack $\text{Cht}_{N,I}^{(I_1, \dots, I_k)} / \Xi$ of chtoucas. The Chevalley involution intervenes ultimately in the effect of Verdier duality in geometric Satake equivalence, which is in turn connected to $\text{Cht}_{N,I}^{(I_1, \dots, I_k)} / \Xi$ via certain canonical smooth morphisms.

These geometric ingredients are already implicit in [Lafforgue 2018]. We just recast the relevant parts into our needs and supply some more details. In fact, the pairing $\langle \cdot, \cdot \rangle$ and its geometrization were used in a crucial way in older versions of [Lafforgue 2018]; that usage is now deprecated, and this article finds another application thereof.

Our third main result concerns the *duality involution* proposed by Prasad [2018, §3]. Assume that G is quasisplit. Fix an additive character ψ of G , an F -pinning \mathcal{P} of G and the corresponding Whittaker datum \mathfrak{w} ; replacing ψ by ψ^{-1} yields another Whittaker datum \mathfrak{w}' . Prasad defined an involution $\iota_{G,\mathcal{P}}$ as the commuting product of the Chevalley involution $\theta = \theta_{\mathcal{P}}$ of G and some inner automorphism ι_- which calibrates the Whittaker datum. Up to $G(F)$ -conjugation, this recovers the MVW involutions on classical groups [Mœglin et al. 1987, Chapitre 4] as well as the transpose-inverse on $\mathrm{GL}(n)$, whose relation with contragredient is well known.

Theorem 1.5 (Theorem 3.5.4). *Let $\phi \in \Phi(G)$ be a semisimple parameter such that Π_{ϕ} contains a unique \mathfrak{w} -generic member π . Then $\Pi_{\iota_{\theta \circ \phi}}$ satisfies the same property with respect to \mathfrak{w}' , and $\tilde{\pi} \simeq \pi \circ \iota_{G,\mathcal{P}} \in \Pi_{\iota_{\theta \circ \phi}}$.*

Besides the crucial assumption which is expected to hold for tempered parameters if one works over \mathbb{C} with true L -packets (called Shahidi’s tempered L -packet conjecture [1990]), the main inputs are Theorem 1.2 and the local “trivial functoriality” applied to $\iota_{G,\mathcal{P}}$ (see [Genestier and Lafforgue 2017, Théorèmes 0.1 and 8.1]). Due to these assumptions and the coarseness of our LLC, one should regard this result merely as some heuristic for Prasad’s conjectures [2018].

To conclude this introduction, let us mention two important issues that are left unanswered in this article:

- As in [Lafforgue 2018; Genestier and Lafforgue 2017], these techniques can be generalized to some *metaplectic coverings*, i.e., central extensions of locally compact groups

$$1 \rightarrow \mu_m(F) \rightarrow \tilde{G} \rightarrow G(F) \rightarrow 1$$

where $\mu_m(R) = \{z \in R^{\times} : z^m = 1\}$ as usual; it is customary to assume $\mu_m(F) = \mu_m(\bar{F})$ here. Fix a character $\zeta : \mu_m \hookrightarrow \overline{\mathbb{Q}}_{\ell}^{\times}$. One studies the irreducible smooth representations π of \tilde{G} that are ζ -genuine, i.e., $\pi(\varepsilon) = \zeta(\varepsilon) \cdot \mathrm{id}$ for all $\varepsilon \in \mu_m(F)$. The most satisfactory setting for metaplectic coverings is due to Brylinski and Deligne [2001] that classifies the central extensions of G by \mathbf{K}_2 as sheaves over $(\mathrm{Spec} F)_{\mathrm{Zar}}$. Taking F -points and pushing-out from $\mathbf{K}_2(F)$ by norm-residue symbols yields a central extension above.

The L -group ${}^L\tilde{G}_{\zeta}$ associated to a Brylinski–Deligne \mathbf{K}_2 -central extension, m and ζ has been constructed in many situations; see the references in [Lafforgue 2018, §14]. Now consider the metaplectic variant of Conjecture 1.1. If π is ζ -genuine, $\tilde{\pi}$ will be ζ^{-1} -genuine so one needs a canonical L -isomorphism ${}^L\theta : {}^L\tilde{G}_{\zeta} \rightarrow {}^L\tilde{G}_{\zeta^{-1}}$; this is further complicated by the fact that ${}^L G_{\zeta}$ is not necessarily a split extension of groups. Although some results seem within reach when G is split, it seems more reasonable to work in the broader \mathbf{K}_2 -setting and incorporate the framework of Gaitsgory and Lysenko [2018] for the geometric part. Nonetheless, this goes beyond the scope of the present article.

- With powerful tools from p -adic Hodge theory, Fargues and Scholze proposed a program to obtain a local Langlands parametrization in characteristic zero, akin to that of Genestier and Lafforgue; see

[Fargues 2016] for an overview. It would certainly be interesting to try to adopt our techniques to characteristic zero. However, our key tools are global adélic in nature, whilst the setting of Fargues and Scholze is global in a different sense (over the Fargues–Fontaine curve). This hinders a direct translation into the characteristic zero setting.

Organization of this article. In Section 2, we collect the basic backgrounds on cusp forms, the integration pairing $\langle \cdot, \cdot \rangle$, contragredient representations and L -parameters, all in the ℓ -adic setting.

In Section 3, we begin by defining the Chevalley involution with respect to a chosen pinning and its extension to the L -group. Then we state the main Theorems 3.2.2 and 3.3.2 in the local and global cases, respectively. The local-global argument and the heuristic on duality involutions (Theorem 3.5.4) are also given there.

We give a brief overview of some basic vocabulary of [Lafforgue 2018] in Section 4. The only purpose of this section is to fix notation and serve as a preparation of the next section. As in [Lafforgue 2018; Genestier and Lafforgue 2017], we allow nonsplit groups as well.

The transposes of excursion operators are described in Section 5. It boils down to explicating the interplay between Verdier duality and partial Frobenius morphisms on the moduli stack of chtoucas. As mentioned before, a substantial part of this section can be viewed as annotations to [Lafforgue 2018], together with a few new computations. The original approach in Section 5 in an earlier manuscript has been substantially simplified following suggestions of Lafforgue.

In Sections 4 and 5, we will work exclusively in the global setting.

Conventions. Throughout this article, we fix a prime number ℓ distinct from the characteristic $p > 0$ of the fields under consideration. We also fix an algebraic closure $\overline{\mathbb{Q}}_\ell$ of the field \mathbb{Q}_ℓ of ℓ -adic numbers.

The six operations on ℓ -adic complexes are those defined in [Laszlo and Olsson 2008a; 2008b], for algebraic stacks locally of finite type over a reasonable base scheme, for example over $\text{Spec } \mathbb{F}_q$ where q is some power of p . Given a morphism f of finite type between such stacks, the symbols $f_!$, f_* , etc. will always stand for the functors between derived categories $D_c^b(\dots, E)$ unless otherwise specified, where the field of coefficients E is some algebraic extension of \mathbb{Q}_ℓ . The perverse t -structure on such stacks is defined in [Laszlo and Olsson 2009]; further normalizations will be recalled in Section 4.1. The constant sheaf associated to E on such a stack \mathcal{X} is denoted by $E_{\mathcal{X}}$.

We use the notation $C_c(X; E)$ to indicate the space of compactly supported smooth E -valued functions on a topological space X , where E is any ring. Since we work exclusively over totally disconnected locally compact spaces, smoothness here means locally constant.

For a local or global field F , we denote by W_F the Weil group F with respect to a choice of separable closure $\overline{F}|F$, and by $I_F \subset \text{Gal}(\overline{F}|F)$ the inertia subgroup. The arithmetic Frobenius automorphism is denoted by Frob . If F is local nonarchimedean, \mathfrak{o}_F will stand for its ring of integers.

If \mathring{F} is a global field, we write $\mathbb{A} = \mathbb{A}_{\mathring{F}} := \prod'_v \mathring{F}_v$ for its ring of adèles, where v ranges over the places of \mathring{F} . We also write $\mathfrak{o}_v = \mathfrak{o}_{\mathring{F}_v}$ in this setting.

For a scheme T , we write:

- $\Delta : T \hookrightarrow T^I$ for the diagonal morphism, where I is any set.
- $\pi_1(T, \bar{t})$ for the étale fundamental group with respect to a geometric point $\bar{t} \rightarrow T$, when T is connected, normal and locally Noetherian.
- $\mathcal{O}(T)$ for the ring of regular functions on T .
- $T_B := T \times_{\text{Spec } A} \text{Spec } B$ if T is a scheme over $\text{Spec } A$, and B is a commutative A -algebra.
- $E(T) := \text{Frac } \mathcal{O}(T)$ for the function field, when T is an irreducible variety over a field E .

Suppose that T is a variety over a field. The geometric invariant-theoretic quotient of T under the right action of some group variety Q , if it exists, is written as $T//Q$. Similar notation pertains to left or bilateral actions.

Let G be a connected reductive group over a field F . For any F -algebra A , denote the group of A -points of G by $G(A)$, endowed with a topology whenever A is. Denote by $Z_G, G^{\text{der}}, G^{\text{ad}}$ for the center, derived subgroup and the adjoint group of G , respectively. Normalizers and centralizers in G are written as $N_G(\cdot)$ and $Z_G(\cdot)$. If $T \subset G$ is a maximal torus, we write T^{ad} , etc. for the corresponding subgroups in G^{ad} , etc. The character and cocharacter groups of a torus T are denoted by $X^*(T)$ and $X_*(T)$ as \mathbb{Z} -modules, respectively.

The L -group and Langlands dual group of G are denoted by ${}^L G$ and \hat{G} , respectively. We use the Galois form of L -groups: details will be given in Section 2.4.

For an affine algebraic group H over some field E , the additive category of finite-dimensional algebraic representations of H will be denoted as $\text{Rep}_E(H)$. The trivial representation is denoted by $\mathbf{1}$. For any object $W \in \text{Rep}_E(H)$, we write \check{W} or W^\vee for its contragredient representation on $\text{Hom}_E(W, E)$. For any automorphism θ of H , write W^θ for the representation on W such that every $h \in H$ acts by $w \mapsto \theta(h) \cdot w$.

The same notation $\check{\pi}$ applies to the contragredient of a smooth representation π of a locally compact totally disconnected group. This will be the topic of Section 2.3. We denote the central character of an irreducible smooth representation π as ω_π .

2. Review of representation theory

2.1. Cusp forms. Let \mathring{F} be a global field of characteristic $p > 0$. We may write $\mathring{F} = \mathbb{F}_q(X)$ where q is some power of p , and X is a smooth, geometrically irreducible proper curve over \mathbb{F}_q . Denote $\mathbb{A} = \mathbb{A}_{\mathring{F}}$. Fix a closed subscheme $N \subset X$ which is finite over \mathbb{F}_q , known as the level.

Let G be a connected reductive group over \mathring{F} . We associate to N a compact open subgroup

$$K_N := \ker \left[G \left(\prod_{v \in |X|} \mathfrak{o}_v \right) \rightarrow G(\mathcal{O}(N)) \right] \subset G(\mathbb{A}).$$

Denote the maximal split central torus in G by A_G . It is also known that there is a cocompact lattice

$$\Xi \subset A_G(\mathring{F}) \backslash A_G(\mathbb{A}),$$

which we fix once and for all. The space $G(\mathring{F}) \backslash G(\mathbb{A}) / \mathfrak{E}$ is known to have finite volume with respect to any Haar measure on $G(\mathbb{A})$.

In what follows, we use a Haar measure on $G(\mathbb{A})$ such that $\text{mes}(K) \in \mathbb{Q}$ for any compact open subgroup K . The existence of such measures is established in [Vignéras 1996, Théorème 2.4]. The same convention pertains to the subgroups of G .

For all subextension $E | \mathbb{Q}_\ell$ of $\overline{\mathbb{Q}_\ell} | \mathbb{Q}_\ell$, we have the space

$$C_c(G(\mathring{F}) \backslash G(\mathbb{A}) / \mathfrak{E}; E) = \bigcup_{N:\text{levels}} C_c(G(\mathring{F}) \backslash G(\mathbb{A}) / K_N \mathfrak{E}; E)$$

of smooth E -valued functions of compact support on $G(\mathring{F}) \backslash G(\mathbb{A}) / \mathfrak{E}$. Then $G(\mathbb{A})$ acts on the left of $C_c(G(\mathring{F}) \backslash G(\mathbb{A}) / \mathfrak{E}; E)$ by $(gf)(x) = f(xg)$. Accordingly, $C_c(G(\mathring{F}) \backslash G(\mathbb{A}) / K_N \mathfrak{E}; E)$, the space of K_N -invariants, is a left module under the unital E -algebra $C_c(K_N \backslash G(\mathbb{A}) / K_N; E)$, the Hecke algebra under convolution \star .

Our convention on Haar measures means that we can integrate E -valued smooth functions on $G(\mathring{F}) \backslash G(\mathbb{A}) / \mathfrak{E}$, etc.

The subspace of $C_c(G(\mathring{F}) \backslash G(\mathbb{A}) / \mathfrak{E}; E)$ of cuspidal functions

$$C_c^{\text{cusp}}(G(\mathring{F}) \backslash G(\mathbb{A}) / \mathfrak{E}; E) = \bigcup_{N:\text{levels}} C_c^{\text{cusp}}(G(\mathring{F}) \backslash G(\mathbb{A}) / K_N \mathfrak{E}; E)$$

is defined by either

- requiring that the constant terms $f_P(x) = \int_{U(\mathring{F}) \backslash U(\mathbb{A})} f(ux) \, du$ are zero whenever $P = MU \subsetneq G$ is a parabolic subgroup, or
- using the criterion in terms of Hecke-finiteness in [Lafforgue 2018, Proposition 8.23].

We record two more basic facts:

- The E -vector space $C_c^{\text{cusp}}(G(\mathring{F}) \backslash G(\mathbb{A}) / K_N \mathfrak{E}; E)$ is finite-dimensional. This result is originally due to Harder, and can be deduced from the uniform bound on supports of such functions in [Mœglin and Waldspurger 1994, I.2.9].
- As a smooth $G(\mathbb{A})$ -representation, $C_c^{\text{cusp}}(G(\mathring{F}) \backslash G(\mathbb{A}) / \mathfrak{E}; E)$ is absolutely semisimple, i.e., it is semisimple after $- \otimes_E \overline{\mathbb{Q}_\ell}$; see [Bourbaki 2012, VIII.226]. Indeed, the semisimplicity in the case $E = \overline{\mathbb{Q}_\ell} \simeq \mathbb{C}$ is well known.

In parallel, $C_c^{\text{cusp}}(G(\mathring{F}) \backslash G(\mathbb{A}) / K_N \mathfrak{E}; E)$ is also absolutely semisimple as a $C_c(K_N \backslash G(\mathbb{A}) / K_N; E)$ -module. Recall the module structure: $f \in C_c(K_N \backslash G(\mathbb{A}) / K_N; E)$ acts on h as

$$(f \cdot h)(x) := \int_{K_N \backslash G(\mathbb{A}) / K_N} h(xg) f(g) \, dg = (h \star \check{f})(x), \quad x \in G(\mathbb{A}) \tag{2-1}$$

where $\check{f}(g) = f(g^{-1})$ and the convolution \star is defined in the usual manner.

We record the following standard result for later use.

Proposition 2.1.1. *For every $G(\mathbb{A})$ -representation $\dot{\pi}$, assumed to be smooth, let $\dot{\pi}^{K_N}$ be the space of K_N -invariant vectors. It is a left module under $C_c(K_N \backslash G(\mathbb{A})/K_N; \overline{\mathbb{Q}}_\ell)$.*

- (i) *For all irreducible $G(\mathbb{A})$ -representations $\dot{\pi}_1, \dot{\pi}_2$ generated by K_N -invariants, we have $\dot{\pi}_1 \simeq \dot{\pi}_2 \iff \dot{\pi}_1^{K_N} \simeq \dot{\pi}_2^{K_N}$ as simple $C_c(K_N \backslash G(\mathbb{A})/K_N; \overline{\mathbb{Q}}_\ell)$ -modules.*
- (ii) *Given any decomposition $C_c^{\text{cusp}}(G(\dot{F}) \backslash G(\mathbb{A})/\mathfrak{E}; \overline{\mathbb{Q}}_\ell) = \bigoplus_{\dot{\pi} \in \Pi} \dot{\pi}$ into irreducibles, where Π is a set (with multiplicities) of irreducible subrepresentations, we have*

$$C_c^{\text{cusp}}(G(\dot{F}) \backslash G(\mathbb{A})/K_N \mathfrak{E}; \overline{\mathbb{Q}}_\ell) = \bigoplus_{\dot{\pi} \in \Pi, \dot{\pi}^{K_N} \neq 0} \dot{\pi}^{K_N}$$

in which each $\dot{\pi}^{K_N}$ is simple.

- (iii) *For every irreducible $G(\mathbb{A})$ -representation $\dot{\pi}$ generated by K_N -invariants, we have a natural isomorphism of multiplicity spaces*

$$\begin{aligned} \text{Hom}_{G(\mathbb{A})\text{-Rep}}(\dot{\pi}, C_c^{\text{cusp}}(G(\dot{F}) \backslash G(\mathbb{A})/\mathfrak{E}; \overline{\mathbb{Q}}_\ell)) \\ \xrightarrow{\sim} \text{Hom}_{C_c(K_N \backslash G(\mathbb{A})/K_N; \overline{\mathbb{Q}}_\ell)\text{-Mod}}(\dot{\pi}^{K_N}, C_c^{\text{cusp}}(G(\dot{F}) \backslash G(\mathbb{A})/K_N \mathfrak{E}; \overline{\mathbb{Q}}_\ell)). \end{aligned}$$

Property (i) actually holds for representations of $G(\dot{F}_v)$ and of its Hecke algebras, for any place v of \dot{F} .

The C_c^{cusp} in (ii) and (iii) can be replaced by $\bigoplus_{\alpha \in \ker^1(\dot{F}, G)} C_c^{\text{cusp}}(G_\alpha(\dot{F}) \backslash G(\mathbb{A})/\mathfrak{E}; \overline{\mathbb{Q}}_\ell)$; see (2-2).

Proof. By semisimplicity, $C_c^{\text{cusp}}(G(\dot{F}) \backslash G(\mathbb{A})/\mathfrak{E}; \overline{\mathbb{Q}}_\ell)$ (or the \bigoplus_α version) decomposes uniquely into $W \oplus W'$ such that

- W is a subrepresentation isomorphic to a direct sum of irreducibles, each summand is generated by K_N -invariants;
- W' is a subrepresentation satisfying $(W')^{K_N} = \{0\}$.

For (ii)–(iii), it suffices to look at the $G(\mathbb{A})$ -representation W and the $C_c(K_N \backslash G(\mathbb{A})/K_N; \overline{\mathbb{Q}}_\ell)$ -module W^{K_N} ; both are semisimple. The required assertions follow from the standard equivalences between categories in [Renard 2010, I.3 and III.1.5] and Schur’s lemma [Renard 2010, III.1.8 and B.II]. \square

Next, we introduce the moduli stack $\text{Bun}_{G,N}$ over \mathbb{F}_q of G -torsors on X with level N structures: it maps any \mathbb{F}_q -scheme S to the groupoid

$$\text{Bun}_{G,N}(S) = \left\{ (\mathcal{G}, \psi) \left| \begin{array}{l} \mathcal{G} \text{ a } G\text{-torsor over } X \times S \text{ and} \\ \psi: \mathcal{G}|_{N \times S} \xrightarrow{\sim} G|_{N \times S} \text{ a trivialization over } N \end{array} \right. \right\}, \quad \text{Bun}_G := \text{Bun}_{G,\emptyset}.$$

For this purpose, we need suitable models of G over X . Let $U \subset X$ be the maximal open subscheme such that G extends to a connected reductive U -group scheme. We follow [Lafforgue 2018, §12.1] to take parahoric models at the formal neighborhoods of all points of $X \setminus U$. Glue these parahoric models with the smooth model over U , à la Beauville–Laszlo, to yield a smooth affine X -group scheme with geometrically connected fibers, known as a *Bruhat–Tits group scheme* over X ; see also [Heinloth 2010, §1].

Regard $\text{Bun}_{G,N}(\mathbb{F}_q)$ as a set, on which Ξ acts naturally. As explained in [Lafforgue 2018], we have

$$\text{Bun}_{G,N}(\mathbb{F}_q) = \bigsqcup_{\alpha \in \ker^1(\mathring{F}, G)} G_\alpha(\mathring{F}) \backslash G_\alpha(\mathbb{A}) / K_N \tag{2-2}$$

where

- $\ker^1(\mathring{F}, G)$ is the kernel of $H^1(\mathring{F}, G) \rightarrow \prod_{v \in |X|} H^1(\mathring{F}_v, G)$;
- to each $\alpha \in \ker^1(\mathring{F}, G)$ is attached a locally trivial pure inner twist G_α of G , and we fix an identification $G_\alpha(\mathbb{A}) \simeq G(\mathbb{A})$.

The decomposition is compatible with Ξ -actions. The pointed set $\ker^1(\mathring{F}, G)$ is finite; it is actually trivial when G is split. As before, we have the spaces

$$C_c^{\text{cusp}}(\text{Bun}_{G,N}(\mathbb{F}_q) / \Xi; E) = \bigoplus_{\alpha \in \ker^1(\mathring{F}, G)} C_c^{\text{cusp}}(G_\alpha(\mathring{F}) \backslash G_\alpha(\mathbb{A}) / K_N \Xi; E).$$

The cuspidality on the left-hand side can be defined in terms of Hecke-finiteness as before. We shall also use compatible Haar measures on various $G_\alpha(\mathbb{A})$.

From the viewpoint of harmonic analysis, the mere effect of working with $\text{Bun}_{G,N}(\mathbb{F}_q)$ is to consider all the inner twists from $\ker^1(\mathring{F}, G)$ at once. See also [Lafforgue 2018, §12.2.5].

2.2. Integration pairing. Let E be a subextension of $\overline{\mathbb{Q}_\ell} | \mathbb{Q}_\ell$.

Definition 2.2.1. With the Haar measures as in Section 2.1, we define the *integration pairing*

$$\langle \cdot, \cdot \rangle : C_c^{\text{cusp}}(G(\mathring{F}) \backslash G(\mathbb{A}) / \Xi; E) \otimes_E C_c^{\text{cusp}}(G(\mathring{F}) \backslash G(\mathbb{A}) / \Xi; E) \rightarrow E$$

$$h \otimes h' \mapsto \langle h, h' \rangle := \int_{G(\mathring{F}) \backslash G(\mathbb{A}) / \Xi} hh'.$$

The pairing is clearly E -bilinear, symmetric and $G(\mathbb{A})$ -invariant. There is an obvious variant for not necessarily cuspidal functions.

Lemma 2.2.2. *The pairing $\langle \cdot, \cdot \rangle$ above is absolutely nondegenerate, i.e., its radical equals $\{0\}$ after $- \otimes_E \overline{\mathbb{Q}_\ell}$.*

Proof. It is legitimate to assume $E = \overline{\mathbb{Q}_\ell}$, and there exists an isomorphism of fields $\overline{\mathbb{Q}_\ell} \simeq \mathbb{C}$. The nondegeneracy over \mathbb{C} is well known: we have $\int h \bar{h} \geq 0$, and equality holds if and only if $h = 0$. \square

Remark 2.2.3. For a chosen level $N \subset X$, we have an analogous pairing

$$\langle \cdot, \cdot \rangle : C_c^{\text{cusp}}(G(\mathring{F}) \backslash G(\mathbb{A}) / K_N \Xi; E) \otimes_E C_c^{\text{cusp}}(G(\mathring{F}) \backslash G(\mathbb{A}) / K_N \Xi; E) \rightarrow E$$

$$h \otimes h' \mapsto \langle h, h' \rangle := \int_{G(\mathring{F}) \backslash G(\mathbb{A}) / K_N \Xi} hh'.$$

The integration here is actually a “stacky” sum over $G(\mathring{F}) \backslash G(\mathbb{A}) / K_N \Xi$, i.e., $\langle h, h' \rangle$ equals that of Definition 2.2.1 if one starts with a Haar measure on $G(\mathbb{A})$ with $\text{mes}(K_N) = 1$. It is also E -bilinear, symmetric, absolutely nondegenerate and invariant in the sense that

$$\langle f \cdot h, h' \rangle = \langle h, \check{f} \cdot h' \rangle, \quad f \in C_c(K_N \backslash G(\mathbb{A}) / K_N; E);$$

see (2-1). There is an obvious variant for not necessarily cuspidal functions.

The spaces in question being finite-dimensional, it makes sense to talk about the *transpose* of a linear operator. For example, the transpose of the left multiplication by f is given by that of \check{f} .

As in Section 2.1, the integration pairing extends to

$$\langle \cdot, \cdot \rangle : C_c^{\text{cusp}}(\text{Bun}_{G,N}(\mathbb{F}_q)/\Xi; E) \otimes_E C_c^{\text{cusp}}(\text{Bun}_{G,N}(\mathbb{F}_q)/\Xi; E) \rightarrow E$$

$$h \otimes h' \mapsto \int_{\text{Bun}_{G,N}(\mathbb{F}_q)/\Xi} hh'.$$

This is the orthogonal sum of the integrations pairings on various $G_\alpha(\mathbb{A})$.

2.3. Representations. In this subsection, we let F be a local field of characteristic $p > 0$. Denote the cardinality of the residue field of F as q . Let G be a connected reductive F -group. The smooth representations of $G(F)$ will always be realized on $\overline{\mathbb{Q}_\ell}$ -vector spaces. Irreducible smooth representation of $G(F)$ are admissible; see [Renard 2010, VI.2.2].

The smooth characters of $G(F)$ are homomorphisms $G(F) \rightarrow \overline{\mathbb{Q}_\ell}^\times$ with open kernel. We will need to look into a class of particularly simple characters, namely those trivial on the open subgroup

$$G(F)^1 := \bigcap_{\chi \in X^*(G)} \ker |\chi|_F \tag{2-3}$$

of $G(F)$, where $X^*(G) := \text{Hom}_{\text{alg. grp}}(G, \mathbb{G}_m)$ and

$$|\cdot|_F : F^\times \rightarrow q^{\mathbb{Z}} \subset \mathbb{Q}^\times$$

is the normalized absolute value on F . Note that $G(F)/G(F)^1 \simeq \mathbb{Z}^r$ with $r := \text{rk}_{\mathbb{Z}} X^*(G)$. Moreover, $G(F)^1 \supset G^{\text{der}}(F) = G^{\text{der}}(F)^1$.

For any smooth character ω of $Z_G(F)$, denote by $C_c(G(F), \omega)$ the space of functions $f : G(F) \rightarrow \overline{\mathbb{Q}_\ell}$ such that $f(zg) = \omega(z)f(g)$ for all $z \in Z_G(F)$ and $\text{Supp}(f)$ is compact modulo $Z_G(F)$.

Let $\text{Ind}_P^G(\cdot)$ denote the unnormalized parabolic induction from the Levi quotient of $P \subset G$. Let δ_P denote the modulus character of $P(F)$ taking values in $q^{\mathbb{Z}}$. Upon choosing $q^{1/2} \in \overline{\mathbb{Q}_\ell}$, we can also form the *normalized parabolic induction* $I_P^G(\cdot) := \text{Ind}_P^G(\cdot \otimes \delta_P^{1/2})$.

We need the notion [Renard 2010, VI.7.1] of the *cuspidal support* (M, τ) of an irreducible smooth representation π . Here $M \subset G$ is a Levi subgroup and τ is a supercuspidal irreducible representation of $M(F)$, such that π is a subquotient of $I_P^G(\tau)$ for any parabolic subgroup $P \subset G$ with Levi component M . The cuspidal support is unique up to $G(F)$ -conjugacy. It is known that one can choose P with Levi component M such that $\pi \hookrightarrow I_P^G(\tau)$. See [Renard 2010, VI.5.4].

We collect below a few properties of an irreducible smooth representation π of $G(F)$:

(1) Suppose that π is supercuspidal. There exists a finite extension E of \mathbb{Q}_ℓ such that π is defined over E . Indeed, since the central character ω_π can be defined over some finite extension of \mathbb{Q}_ℓ , so is $\pi \hookrightarrow C_c(G(F), \omega_\pi)$.

From this and the discussion on cuspidal supports, it follows that every π can be defined over some finite extension E of \mathbb{Q}_ℓ .

(2) We say π is *integral* if it admits an \mathfrak{o}_E -model of finite type, where E is a finite extension of \mathbb{Q}_ℓ . See [Vignéras 2001, §1.4] for details. Then an irreducible supercuspidal π is integral if and only if ω_π has ℓ -adically bounded image in $\overline{\mathbb{Q}_\ell}^\times$.

Again, this is a consequence of $\pi \hookrightarrow C_c(G(F), \omega_\pi)$. It also implies the notion of integrality stated in the beginning of [Genestier and Lafforgue 2017].

(3) Let V be the underlying vector space of π . The *contragredient representation* $\check{\pi}$ of a smooth representation π is realized on the space V^\vee of the smooth vectors in $\text{Hom}_{\overline{\mathbb{Q}_\ell}}(V, \overline{\mathbb{Q}_\ell})$. It satisfies $\langle \check{\rho}(g)\check{v}, v \rangle = \langle \check{v}, \rho(g^{-1})v \rangle$. If π is defined over E , so is $\check{\pi}$. Taking contragredient preserves irreducibility and supercuspidality. It is clear that $(\pi \otimes \chi)^\vee = \check{\pi} \otimes \chi^{-1}$ for any smooth character $\chi : G(F) \rightarrow \overline{\mathbb{Q}_\ell}^\times$.

(4) Moreover, $(\pi)^{\vee\vee} \simeq \pi$ for all smooth irreducible π ; see [Renard 2010, III.1.7]. Also, $\omega_{\check{\pi}} = \omega_\pi^{-1}$.

Proposition 2.3.1. *If π is an irreducible smooth representation of $G(F)$ with cuspidal support (M, τ) , then $\check{\pi}$ has cuspidal support $(M, \check{\tau})$.*

Proof. Choose a parabolic subgroup $P \subset G$ with M as Levi component such that $\pi \hookrightarrow I_P^G(\tau)$. Once the Haar measures are chosen, we have $I_P^G(\tau)^\vee \simeq I_P^G(\check{\tau})$ canonically; see [Bushnell and Henniart 2006, §3.5]. Dualizing, we deduce $I_P^G(\check{\tau}) \twoheadrightarrow \check{\pi}$. Thus $\check{\pi}$ is a subquotient of $I_P^G(\check{\tau})$. \square

2.4. L-parameters. Let F be a local or global field of characteristic $p > 0$. For a connected reductive F -group G , we denote by $\tilde{F}|F$ the splitting field of G , which is a finite Galois extension inside a chosen separable closure \bar{F} .

Denote by W_F the absolute Weil group of F . It comes with canonical continuous homomorphisms (i) $W_F \rightarrow \text{Gal}(\bar{F}|F)$, and (ii) $W_{F_v} \rightarrow W_F$ if F is global and v is a place of F . For (ii) we choose an embedding $\bar{F} \hookrightarrow \bar{F}_v$ of separable closures.

Definition 2.4.1. The *Langlands dual group* \hat{G} of G is a pinned connected reductive \mathbb{Q}_ℓ -group (in fact, definable over \mathbb{Z}), on which $\text{Gal}(\tilde{F}|F)$ operates by pinned automorphisms. Throughout this article, we use the *finite Galois forms* of the L -group of G , namely

$${}^L G := \hat{G} \rtimes \text{Gal}(\tilde{F}|F)$$

viewed as an affine algebraic group.

If $M \hookrightarrow G$ is a Levi subgroup, we obtain a the corresponding embedding ${}^L M \rightarrow {}^L G$ of standard Levi subgroup.

Definition 2.4.2. An L -parameter for G is a homomorphism $\sigma : W_F \rightarrow {}^L G(\overline{\mathbb{Q}_\ell})$ such that:

- The following diagram commutes:

$$\begin{array}{ccc}
 W_F & \xrightarrow{\sigma} & {}^L G \\
 & \searrow & \swarrow \\
 & & \text{Gal}(\tilde{F}|F)
 \end{array}$$

- σ is continuous with respect to the ℓ -adic topology on ${}^L G(\overline{\mathbb{Q}}_\ell)$.
- σ is *relevant* in the sense of [Borel 1979, §8.2], which matters only when G is not quasisplit.
- (The local case) σ is Frobenius semisimple: $\rho(\sigma(\text{Frob}))$ is semisimple for every algebraic representation $\rho : {}^L G(\overline{\mathbb{Q}}_\ell) \rightarrow \text{GL}(N, \overline{\mathbb{Q}}_\ell)$, where Frob stands for any Frobenius element in W_F (see [Bushnell and Henniart 2006, 32.7 Proposition] for more discussions on Frobenius-semisimplicity).
- (The global case) σ is semisimple in the sense of [Serre 2005], to be described below. We do not require Frobenius-semisimplicity here because for ℓ -adic representations of geometric origin, that property is a long-standing conjecture in étale cohomology.

The set of $\hat{G}(\overline{\mathbb{Q}}_\ell)$ -conjugacy classes of L -parameters is denoted as $\Phi(G)$. By [Borel 1979, §3.4], there is a natural map $\Phi(M) \rightarrow \Phi(G)$ for any Levi subgroup M .

Remark 2.4.3. Since ${}^L G(\overline{\mathbb{Q}}_\ell)$ carries the ℓ -adic topology and σ is required to be continuous, when F is local we get rid of the Weil–Deligne group in the usual formulation in terms of ${}^L G(\mathbb{C})$. Besides, we do not consider Arthur parameters in this article.

As recalled earlier, the structure of Weil groups allows us to

- localize a global L -parameter at a place v ;
- talk about L -parameters of the form $\text{Gal}(\overline{F}|F) \rightarrow {}^L G(\overline{\mathbb{Q}}_\ell)$ and their localizations when F is global.

Next, we recall the *semisimplicity* of L -parameters following [Lafforgue 2018; Serre 2005]: a continuous homomorphism $\sigma : W_F \rightarrow {}^L G(\overline{\mathbb{Q}}_\ell)$ is called semisimple if the Zariski closure of $\text{im}(\sigma)$ is reductive in ${}^L G(\overline{\mathbb{Q}}_\ell)$, in the sense that its identity connected component is reductive. When G is split, this is exactly the definition of complete reducibility in [Serre 2005, 3.2.1], say by applying [loc. cit., Proposition 4.2].

Lemma 2.4.4. *Assume F is local. The following are equivalent for any L -parameter σ for G :*

- (i) σ is semisimple.
- (ii) The Weil–Deligne parameter associated to σ has trivial nilpotent part.

Proof. By composing σ with any faithful algebraic representation $\rho : {}^L G(\overline{\mathbb{Q}}_\ell) \hookrightarrow \text{GL}(N, \overline{\mathbb{Q}}_\ell)$, we may assume that σ is an ℓ -adic representation $W_F \rightarrow \text{GL}(N, \overline{\mathbb{Q}}_\ell)$. To σ is associated the Weil–Deligne representation $\text{WD}(\sigma)$: it comes with a nilpotent operator \mathfrak{n} . For details, see [Bushnell and Henniart 2006, 32.5].

(i) \implies (ii): The line $\overline{\mathbb{Q}}_\ell \mathfrak{n}$ is preserved by $\text{im}(\sigma)$ -conjugation. Since $\exp(t\mathfrak{n}) \in \text{im}(\sigma)$ for $t \in \mathbb{Z}_\ell$ with $|t| \ll 1$, the semisimplicity of σ forces $\mathfrak{n} = 0$.

(ii) \implies (i): As $\mathfrak{n} = 0$, the smooth representation underlying $\text{WD}(\sigma)$ is just σ , hence σ is semisimple as a smooth representation of W_F by [Bushnell and Henniart 2006, 32.7 Theorem]. The reductivity (or complete reducibility) of the Zariski closure of $\text{im}(\sigma)$ then follows from the theory in [Serre 2005]. \square

Finally, we define parabolic subgroups of ${}^L G$ as in [Borel 1979, 3.2]. They are subgroups of the form $N_{L_G}(\hat{P})$ where $\hat{P} \subset \hat{G}$ is a parabolic subgroups, and whose projection to $\text{Gal}(\tilde{F}|F)$ has full image. Define the unipotent radical of such a parabolic subgroup to be that of \hat{P} . We still have the notion of Levi decomposition in this setting; see [Borel 1979, 3.4].

Following [Lafforgue 2018, §13], the *semisimplification* σ^{ss} of an L -parameter σ is defined as follows:

- First, take the smallest parabolic subgroup ${}^L P \subset {}^L G$ containing $\text{im}(\sigma)$.
- Project to the Levi quotient.
- Then embed back into ${}^L G$ using some Levi decomposition.

The resulting parameter is well-defined up to $\hat{G}(\overline{\mathbb{Q}_\ell})$ -conjugacy.

By definition, an L -homomorphism ${}^L H \rightarrow {}^L G$ between L -groups is an algebraic homomorphism respecting the projections to $\text{Gal}(\tilde{F}|F)$.

Lemma 2.4.5. *Up to $\hat{G}(\overline{\mathbb{Q}_\ell})$ -conjugacy, semisimplification commutes with L -automorphisms of ${}^L G$.*

Proof. Indeed, an L -automorphism permutes the parabolic subgroups of ${}^L G$ together with their Levi decompositions. □

3. Statement of a variant of the conjecture

3.1. Chevalley involutions. To begin with, we consider a split connected reductive group H over a field, equipped with a pinning $\mathcal{P} = (B, T, (X_\alpha)_{\alpha \in \Delta_0})$, where

- (B, T) is a Borel pair of H , and
- X_α is a nonzero vector in the root subspace \mathfrak{h}_α , where α ranges over the set Δ_0 of B -simple roots.

Definition 3.1.1. The *Chevalley involution* $\theta = \theta_{\mathcal{P}}$ is the unique pinned automorphism of H acting as $t \mapsto w_0(t^{-1})$ on T , where w_0 stands for the longest element in the Weyl group associated to T .

This is the definition in [Prasad 2018, §4], and it is clear that $\theta^2 = \text{id}_H$.

The Chevalley involution will be considered in the following settings. Let F be a field with separable closure \bar{F} .

- (1) Let $H = \hat{G}$ be the dual group of G , which is connected reductive over F . The dual group is endowed with a pinning and we obtain $\theta : \hat{G} \rightarrow \hat{G}$. Since $\text{Gal}(\tilde{F}|F)$ operates by pinned automorphisms on \hat{G} , the Chevalley involution extends to

$${}^L \theta : {}^L G \rightarrow {}^L G, \quad g \rtimes \sigma \mapsto \theta(g) \rtimes \sigma,$$

which is still an involution.

- (2) Let G be a quasisplit connected reductive group over F . Then G admits an F -pinning \mathcal{P} , i.e., a Galois-invariant pinning of $H := G_{\bar{F}}$. Therefore the Chevalley involution $\theta = \theta_{\mathcal{P}}$ for $G_{\bar{F}}$ descends to G .

Furthermore, observe that if $H = \prod_{i=1}^r H_i$ and \mathcal{P} decomposes into $(\mathcal{P}_1, \dots, \mathcal{P}_r)$ accordingly, the corresponding Chevalley involution $\theta_{\mathcal{P}}$ equals $\prod_{i=1}^r \theta_{\mathcal{P}_i}$.

3.2. The local statement. Let F be a local field of characteristic $p > 0$. Let G be a connected reductive F -group. The set of isomorphism classes of irreducible smooth representations over $\overline{\mathbb{Q}}_\ell$ of $G(F)$ will be denoted by $\Pi(G)$. The local statement to follow presumes a given *Langlands parametrization* of representations, namely an arrow

$$\begin{aligned} \Pi(G) &\rightarrow \Phi(G) \\ \pi &\mapsto \phi. \end{aligned}$$

This is the “automorphic to Galois” direction of the local Langlands correspondence for G . We say that ϕ is the parameter of π , and denote by $\Pi_\phi \subset \Pi(G)$ the fiber over ϕ , called the *packet* associated to ϕ .

For the local statement, we employ the Langlands parametrization furnished by Genestier and Lafforgue [2017]. It is actually an arrow

$$\Pi(G) \rightarrow \{\text{semisimple } L\text{-parameters}\} / \hat{G}(\overline{\mathbb{Q}}_\ell)\text{-conj.} \subset \Phi(G).$$

Remark 3.2.1. The Genestier–Lafforgue parameters are expected to be the semisimplifications of authentic (yet hypothetical) Langlands parameters. As a consequence, the packets Π_ϕ for general Genestier–Lafforgue parameters are expected to be a disjoint union of authentic L -packets, unless when ϕ is an elliptic parameter (see Lemma 2.4.4), i.e., $\text{im}(\phi)$ is ${}^L G$ -ir in the sense of [Serre 2005, 3.2.1].

Further descriptions and properties of the Genestier–Lafforgue parametrization will be reviewed in due course. Let us move directly to the main local statement.

Theorem 3.2.2. *Let $\phi \in \Phi(G)$ be a semisimple L -parameter. In terms of the Langlands parametrization of Genestier–Lafforgue, we have*

$$\{\check{\pi} : \pi \in \Pi_\phi\} = \Pi_{{}^L\theta \circ \phi},$$

where ${}^L\theta : {}^L G \rightarrow {}^L G$ is the Chevalley involution in Section 3.1.

If the Genestier–Lafforgue parametrization is replaced by an authentic Langlands parametrization, the statement above becomes [Adams and Vogan 2016, Conjecture 1.1]; it is also a part of [Prasad 2018, §4, Conjecture 2], but Prasad’s conjecture also predicates on the internal structure of L -packets. The conjecture of Adams, Vogan, and Prasad applies to any local field F ; known cases in this generality include:

- The case $F = \mathbb{R}$ in [Adams and Vogan 2016, Theorem 7.1(a)], with admissible representations of $G(\mathbb{R})$ over \mathbb{C} .
- The tempered L -packets for symplectic groups $\text{Sp}(2n)$ and quasisplit SO groups over nonarchimedean local fields F of characteristic zero in terms of Arthur’s endoscopic classification, see [Kaletha 2013, Corollary 5.10].
- The depth-zero and epipelagic L -packets for many p -adic groups [Kaletha 2013, §6].

Each case above requires a different construction of L -packets, applicable to different groups or parameters, whereas the Theorem 3.2.2 furnishes a uniform statement. On the other hand, Theorem 3.2.2 is weaker since the Langlands parametrization here is coarser, in view of the Remark 3.2.1.

The proof of Theorem 3.2.2 will occupy Section 3.4.

3.3. The global statement. Theorem 3.2.2 will be connected to the global result below.

Let $\mathring{F} = \mathbb{F}_q(X)$ and fix the level $N \subset X$ as in Section 2.1. Let G, K_N and Ξ be as in Section 2.1, so that $\text{Bun}_{G,N}$ is defined. Note that we need to choose a model of G over X which is a Bruhat–Tits group scheme, still denoted as G . Let $U \subset X$ denote the (open) locus of good reduction of G , and set

$$\hat{N} := N \cup (X \setminus U). \tag{3-1}$$

This is a finite closed \mathbb{F}_q -subscheme of X , the “unramified locus”. Let $\eta \rightarrow X$ be the generic point of X ; fix a geometric generic point $\bar{\eta} \rightarrow \eta$ of X .

The main global result of Lafforgue [2018, Théorème 12.3] gives a canonical decomposition of $C_c(K_N \backslash G(\mathbb{A})/K_N; \overline{\mathbb{Q}}_\ell)$ -modules

$$C_c^{\text{cusp}}(\text{Bun}_{G,N}(\mathbb{F}_q)/\Xi; \overline{\mathbb{Q}}_\ell) = \bigoplus_{\sigma} \mathfrak{H}_{\sigma} \tag{3-2}$$

indexed by L -parameters $\sigma : \text{Gal}(\bar{F}|\mathring{F}) \rightarrow {}^L G(\overline{\mathbb{Q}}_\ell)$ up to $\hat{G}(\overline{\mathbb{Q}}_\ell)$ -conjugacy that

- are semisimple, and
- factor continuously through $\text{Gal}(\bar{F}|\mathring{F}) \rightarrow \pi_1(X \setminus \hat{N}, \bar{\eta})$.

Remark 3.3.1. Since the left-hand side of (3-2) is a semisimple module, of finite dimension over $\overline{\mathbb{Q}}_\ell$, so are its submodules \mathfrak{H}_{σ} . To each σ we may associate a set (with multiplicities) of simple submodules C_{σ} , such that

$$\mathfrak{H}_{\sigma} = \bigoplus_{\mathcal{L} \in C_{\sigma}} \mathcal{L}, \quad \text{hence} \quad C_c^{\text{cusp}}(\text{Bun}_{G,N}(\mathbb{F}_q)/\Xi; \overline{\mathbb{Q}}_\ell) = \bigoplus_{\sigma} \bigoplus_{\mathcal{L} \in C_{\sigma}} \mathcal{L}$$

as $C_c(K_N \backslash G(\mathbb{A})/K_N; \overline{\mathbb{Q}}_\ell)$ -modules.

The decomposition (3-2) is built on two pillars: the theories of excursion operators and pseudocharacters for ${}^L G$. As in the local case, we defer the necessary details of [Lafforgue 2018] to Section 4.

Theorem 3.3.2. *Suppose that $\mathfrak{H}_{\sigma}, \mathfrak{H}_{\sigma'}$ are two nonzero summands in (3-2) such that the restriction*

$$\langle \cdot, \cdot \rangle_{\sigma, \sigma'} : \mathfrak{H}_{\sigma} \otimes_{\overline{\mathbb{Q}}_\ell} \mathfrak{H}_{\sigma'} \rightarrow \overline{\mathbb{Q}}_\ell$$

of the integration pairing $\langle \cdot, \cdot \rangle$ of Remark 2.2.3 (extended to $\text{Bun}_{G,N}(\mathbb{F}_q)/\Xi$) is not identically zero. Then we have

$$\sigma' = {}^L \theta \circ \sigma \quad \text{in } \Phi(G);$$

here ${}^L \theta$ is the Chevalley involution of ${}^L G$.

The proof of Theorem 3.3.2 will be accomplished at the end of Section 5.3.

3.4. Local-global argument. Consider a connected reductive group G over a local field F of characteristic p as in the local setting Section 3.2. As usual, A_G stands for the maximal central split torus in G , and $\tilde{F}|F$ stands for the splitting field of G . Take a maximal torus $T \subset G$ with splitting field equal to \tilde{F} . Let $H^1(W_F, Z_{\hat{G}})$ denote the continuous cohomology with values in $Z_{\hat{G}}(\overline{\mathbb{Q}_\ell})$ with discrete topology.

The first lemma concerns the Langlands parametrization of smooth characters of $G(F)$. The general case turns out to be delicate: by the discussion in [Lapid and Mao 2015, Appendix A], the usual cohomological construction actually yields an arrow in the opposite direction:

$$\begin{array}{ccc} H^1(W_F, Z_{\hat{G}}) & \longrightarrow & \{\eta : G(F) \rightarrow \overline{\mathbb{Q}_\ell}^\times, \text{ smooth character}\} \\ \downarrow & & \\ \Phi(G) & & \end{array}$$

It is injective but not necessarily surjective. However, we only need the invert it when $\eta|_{G(F)^1}$ is trivial. This is well known to experts, and below is a sketch.

Lemma 3.4.1. *For G as above, there is a canonical homomorphism of groups*

$$\{\eta : G(F)/G(F)^1 \rightarrow \overline{\mathbb{Q}_\ell}^\times, \text{ a smooth character}\} \rightarrow H^1(W_F, Z_{\hat{G}}).$$

Here we do not assume $\text{char}(F) > 0$.

Proof. Fix η . First, one can take a z -extension of G as in [Lapid and Mao 2015, Proof of Lemma A.1], i.e., a central extension

$$1 \rightarrow C \rightarrow G_1 \xrightarrow{p} G \rightarrow 1, \quad C \text{ is an induced torus, } G_1^{\text{der}} \text{ simply connected.}$$

Then $\eta_1 := \eta \circ p$ is trivial on $G_1(F)^1$. We know that $H^1(F, G_1^{\text{der}})$ is trivial. Put $S := G_1/G_1^{\text{der}}$ so that $G_1(F)/G_1^{\text{der}}(F) \xrightarrow{\sim} S(F)$ and $\hat{S} \simeq Z_{\hat{G}_1} = Z_{\hat{G}_1}^\circ$. Then $G_1^{\text{der}}(F) \subset G_1(F)^1$ implies that η_1 factors through $S(F)$. The local classfield theory affords an element $a \in H^1(W_F, Z_{\hat{G}_1})$. Since $\eta_1|_C = 1$, we infer that a has trivial image in $H^1(W_F, \hat{C})$.

Furthermore, using the fact that C is induced, in [loc. cit.] the following natural isomorphism is constructed:

$$H^1(W_F, Z_{\hat{G}}) \simeq \ker[H^1(W_F, Z_{\hat{G}_1}) \rightarrow H^1(W_F, \hat{C})].$$

All in all, we obtain $a \in H^1(W_F, Z_{\hat{G}})$. It is routine to check that $\eta \mapsto a$ is independent of the choice of z -extensions, see [loc. cit.]. \square

In fact, η corresponds to some class in $H^1(W_F/I_F, Z_{\hat{G}}^{I_F})$. To see this, one readily reduces to the case of a torus S as above. Since $S(F)^1$ contains the parahoric subgroup, one can infer, for example by the Satake isomorphism [Haines and Rostami 2010, Proposition 1.0.2] for S , that we obtain a parameter in $H^1(W_F/I_F, \hat{S}^{I_F})$.

The second lemma concerns the globalization of groups.

Lemma 3.4.2. *Given G and F as above, one can choose*

- \mathring{F} : a global field of characteristic p ;
- \mathring{G} : a connected reductive \mathring{F} -group with maximal \mathring{F} -torus \mathring{T} , sharing the same splitting field $\mathring{F}|\mathring{F}$;
- v : a place of \mathring{F} , and w is the unique place of $\tilde{\mathring{F}}$ lying over v , in particular $\text{Gal}(\tilde{\mathring{F}}|\mathring{F})$ equals the decomposition group $\Gamma_w := \text{Gal}(\tilde{\mathring{F}}_w|\mathring{F}_v)$;

such that

- there exist isomorphisms $\mathring{F}_v \simeq F, \tilde{\mathring{F}}_w \simeq \tilde{F}$, which identify $\Gamma := \text{Gal}(\tilde{F}|F)$ with Γ_w ;
- under the identifications above, there is an isomorphism

$$\begin{array}{ccc} \mathring{G}_{\mathring{F}_v} & \xrightarrow{\sim} & G \\ \cup & & \cup \\ \mathring{T}_{\mathring{F}_v} & \xrightarrow{\sim} & T \end{array},$$

i.e., $\mathring{G} \supset \mathring{T}$ is an \mathring{F} -model of $G \supset T$;

- \mathring{G} and G share the same root datum endowed with actions of $\Gamma \simeq \Gamma_w$, relative to \mathring{T} and T respectively.

Proof. Standard. See for instance [Arthur 1988, p.526] or [Vignéras 2001, 3.12]. □

Remark 3.4.3. The matching of root data in Lemma 3.4.2 also implies that $A_{\mathring{G}}$ is “the same” as A_G . Hereafter, we shall drop the clumsy notation $\mathring{G}, \mathring{T}$ or $A_{\mathring{G}}$, and denote them abusively as G, T or A_G instead.

For any closed discrete subgroup $\Xi \subset A_G(F)$ isomorphic to $\mathbb{Z}^{\dim A_G}$, its isomorphic image in $A_G(\mathring{F}) \backslash A_G(\mathbb{A})$ will also be denoted by Ξ . Another consequence of Lemma 3.4.2 is that Ξ is a cocompact lattice in $A_G(\mathring{F}) \backslash A_G(\mathbb{A})$ satisfying the requirements in Section 2.1.

Proof of Theorem 3.2.2 from Theorem 3.3.2. In what follows, we write $\pi \rightsquigarrow \phi$ if $\pi \in \Pi(G)$ has Genestier–Lafforgue parameter $\phi \in \Phi(G)$. It suffices to show that for every $\pi \in \Pi(G)$,

$$(\pi \rightsquigarrow \phi) \implies (\check{\pi} \rightsquigarrow {}^L\theta \circ \phi). \tag{3-3}$$

Indeed, this assertion amounts to $\{\check{\pi} : \pi \in \Pi_\phi\} \subset \Pi_{L\theta \circ \phi}$. The reverse inclusion will follow by applying (3-3) to any $\pi_1 \in \Pi(G)$ with $\pi_1 \rightsquigarrow {}^L\theta \circ \phi$, which in turn yields $\pi := \check{\pi}_1 \rightsquigarrow L\theta \circ {}^L\theta \circ \phi = \phi$ whilst $\pi_1 = \check{\pi}$.

The assertion (3-3) will be established in steps:

Step 1. We reduce (3-3) to the case π supercuspidal. Indeed, let (M, τ) be the cuspidal support of π reviewed in Section 2.3. By Proposition 2.3.1, $\check{\pi}$ has cuspidal support $(M, \check{\tau})$.

On the dual side, choose an L -embedding $\iota: {}^L M \hookrightarrow {}^L G$ as reviewed in Section 2.4. Suppose that $\tau \rightsquigarrow \phi_\tau$ in M . By [Genestier and Lafforgue 2017, Théorème 0.1], ϕ equals to the composite of $W_F \xrightarrow{\phi_\tau} {}^L M \hookrightarrow {}^L G$ up to $\hat{G}(\overline{\mathbb{Q}_\ell})$ -conjugacy. The same relation holds for the parameters for $\check{\pi}$ and $\check{\tau}$. Denoting ${}^L\theta_M$ the

Chevalley involution on ${}^L M$, the diagram

$$\begin{array}{ccc} {}^L M & \xrightarrow{\iota} & {}^L G \\ {}^L \theta_M \downarrow & & \downarrow {}^L \theta \\ {}^L M & \xrightarrow{\iota} & {}^L G \end{array}$$

is commutative up to an explicit $\hat{G}(\overline{\mathbb{Q}}_\ell)$ -conjugacy, by [Prasad 2018, §5, Lemma 4]. Upon replacing (G, π) by (M, τ) , we have reduced (3-3) to the supercuspidal case.

Step 2. Consider the smooth character $\omega := \omega_\pi|_{A_G(F)}$. We reduce (3-3) to the case that ω is of finite order as follows (see also Remark 3.4.4). First, recalling (2-3), there exists a character

$$\eta_0 : A_G(F)/A_G(F)^1 \rightarrow \overline{\mathbb{Q}}_\ell^\times$$

such that $\eta_0 \otimes \omega$ is of finite order. Indeed, this is easily reduced to the case $A_G \simeq \mathbb{G}_m$, and it suffices to take $\eta_0(\varpi) = \omega(\varpi)^{-1}$ where $\varpi \in F^\times$ is some uniformizer.

Secondly, the inclusion of discrete free commutative groups of finite type

$$A_G(F)/A_G(F)^1 = A_G(F)/A_G(F) \cap G(F)^1 \hookrightarrow G(F)/G(F)^1$$

has finite cokernel, whereas $\overline{\mathbb{Q}}_\ell^\times$ is divisible. Therefore η_0 extends to a smooth character

$$\eta : G(F)/G(F)^1 \rightarrow \overline{\mathbb{Q}}_\ell^\times.$$

The central character of $\pi \otimes \eta$ has finite order when restricted to $A_G(F)$.

Attach $a \in H^1(W_F, Z_{\hat{G}})$ to η by Lemma 3.4.1; it can be used to twist elements of $\Phi(G)$ by the homomorphism

$$W_F \times (Z_{\hat{G}} \times \hat{G}) \rightarrow W_F \times \hat{G}, \quad w \times (z, g) \mapsto w \times (zg)$$

by choosing any cocycle representative of a ; see [Genestier and Lafforgue 2017, Remarque 0.2].

In the construction above, $-\otimes \eta^{-1}$ corresponds to twisting a parameter by a^{-1} . We have $(\pi \otimes \eta)^\vee \simeq \check{\pi} \otimes \eta^{-1}$. Concurrently, ${}^L \theta \circ (\phi \cdot a) = ({}^L \theta \circ \phi) \cdot a^{-1}$ since Chevalley involution acts as $z \mapsto z^{-1}$ on the center. Therefore, by replacing π by $\pi \otimes \eta$, it suffices to prove (3-3) when ω has finite order.

Step 3. Now we can assume π to be integral supercuspidal (see Section 2.3) with $\omega := \omega_\pi|_{A_G(F)}$ of finite order. By [Genestier and Lafforgue 2017], we know that the parameter ϕ of π factors through $\text{Gal}(\bar{F}|F)$. Take a global \mathring{F} -model of $G \supset A_G$ as in Lemma 3.4.2 with $\mathring{F}_v \simeq F$. As A_G is split over \mathring{F} , by reducing to \mathbb{G}_m and applying [Artin and Tate 1968, Chapter X, §2, Theorem 5], there exists an automorphic character

$$\hat{\omega} = \bigotimes_u \hat{\omega}_u : A_G(\mathring{F}) \backslash A_G(\mathbb{A}) \rightarrow \overline{\mathbb{Q}}_\ell^\times$$

of finite order, such that $\hat{\omega}_v = \omega$.

Since ω is smooth, there exists a closed discrete subgroup $\Xi \subset A_G(F)$ such that $\omega|_\Xi = 1$ and $\Xi \simeq \mathbb{Z}^{\dim A_G}$. In view of Remark 3.4.3, Ξ also affords the cocompact lattice in $A_G(\mathring{F}) \backslash A_G(\mathbb{A})$ required in Section 2.1.

Claim: there exists a cuspidal automorphic representation $\mathring{\pi} = \bigotimes_u \mathring{\pi}_u$ of $G(\mathbb{A})$ (in the extended sense that we consider all G_α simultaneously, $\alpha \in \ker^1(\mathring{F}, G)$) such that:

- The central character of $\mathring{\pi}$ equals $\mathring{\omega}$ on $A_G(\mathbb{A})$.
- We have $\mathring{\pi}_v \simeq \pi$.
- Relative to the chosen lattice Ξ and a sufficiently deep level N , the $C_c(K_N \backslash G(\mathbb{A})/K_N; \overline{\mathbb{Q}}_\ell)$ -module $\mathring{\pi}^{K_N}$ can be embedded in some summand \mathfrak{H}_σ in (3-2).

This can be achieved by the following variant of the argument in [Henniart 1983, Appendice 1] (which works over \mathbb{C}) via Poincaré series; see also the proof of [Genestier and Lafforgue 2017, Lemme 1.4]. For each place u of \mathring{F} , choose a smooth function $f_u \in C_c(G(\mathring{F}_u), \mathring{\omega}_u)$ such that:

- There exists a finite set S of places of \mathring{F} containing v and the ramification locus of G , such that when $u \notin S$, the function f_u is right $G(\mathfrak{o}_u)$ -invariant, supported on $A_G(\mathring{F}_u)G(\mathfrak{o}_u)$ and $f_u(1) = 1$, where $G(\mathfrak{o}_u)$ is the hyperspecial subgroup arising from some reductive model of G over the ring of S -integers in \mathring{F} .
- We require f_v to a matrix coefficient of π and assume $f_v(1) \neq 0$.
- For every $u \in S \setminus \{v\}$, we require that $C_u := \text{Supp}|f_u|$ is a sufficiently small neighborhood of 1 modulo $A_G(F_u)$, so that the image of

$$\text{Supp}(f_v) \times \prod_{u \in S \setminus \{v\}} C_u \times \prod_{u \notin S} G(\mathfrak{o}_u)A_G(\mathring{F}_u)$$

in $A_G(\mathbb{A}) \backslash G(\mathbb{A}) = (A_G \backslash G)(\mathbb{A})$ intersects $A_G(\mathring{F}) \backslash G(\mathring{F}) = (A_G \backslash G)(\mathring{F})$ only at 1. To see why this can be achieved, embed $A_G \backslash G$ into some affine space over F .

Take $f := \prod_u f_u : G(\mathbb{A}) \rightarrow \overline{\mathbb{Q}}_\ell$ and form

$$P_f(g) = \sum_{\gamma \in (A_G \backslash G)(\mathring{F})} f(\gamma g), \quad g \in G(\mathbb{A}).$$

The sum is finite when g is constrained in any compact subset modulo $A_G(\mathbb{A})$. By choosing N sufficiently deep, it furnishes an element of $C_c(G(\mathring{F}) \backslash G(\mathbb{A})/K_N \Xi; \overline{\mathbb{Q}}_\ell)$. Moreover, $P_f(1) = f(1) \neq 0$ by the condition on supports. By looking at f_v , we see that P_f is a cusp form.

Decompose $C_c^{\text{cusp}}(\text{Bun}_{G,N}(\mathbb{F}_q)/\Xi; \overline{\mathbb{Q}}_\ell)$ into simple submodules as in Remark 3.3.1. There exists a summand \mathcal{L} contained in some \mathfrak{H}_σ such that P_f has nonzero component in \mathcal{L} . Let $\mathring{\pi}$ be the cuspidal automorphic representation corresponding to \mathcal{L} via Proposition 2.1.1 (realized in $\bigoplus_\alpha C_c^{\text{cusp}}(G_\alpha(\mathring{F}) \backslash G(\mathbb{A}) \cdots)$ where $\alpha \in \ker^1(\mathring{F}, G)$) so that $\mathring{\pi}^{K_N} = \mathcal{L} \hookrightarrow \mathfrak{H}_\sigma$. Then $\mathring{\pi}$ has central character $\mathring{\omega}$ on $A_G(\mathbb{A})$ and $\mathring{\pi}_v \simeq \pi$, since P_f and \mathcal{L} have similar properties under $C_c(K_N \backslash G(\mathbb{A})/K_N; \overline{\mathbb{Q}}_\ell)$.

Step 4. Since the integration pairings $\langle \cdot, \cdot \rangle$ of Remark 2.2.3 are nondegenerate, $\mathring{\pi}^{K_N} \subset \mathfrak{H}_\sigma$ must pair nontrivially with some simple $C_c(K_N \backslash G(\mathbb{A})/K_N; \overline{\mathbb{Q}}_\ell)$ -submodule of some $\mathfrak{H}_{\sigma'}$. Proposition 2.1.1 implies that the simple submodule takes the form $(\mathring{\pi}')^{K_N} \subset \mathfrak{H}_{\sigma'}$ for some cuspidal automorphic representation $\mathring{\pi}'$.

Theorem 3.3.2 then asserts $\sigma' = {}^L\theta \circ \sigma$ in $\Phi(G)$ (global version). On the other hand, $\check{\pi}'$ pairs nontrivially with $\check{\pi}$ under the integration pairing $\langle \cdot, \cdot \rangle$ of Definition 2.2.1. The invariance of $\langle \cdot, \cdot \rangle$ therefore implies that, as $G(\mathbb{A})$ -representations,

$$\bigotimes_u (\check{\pi}_u)^\vee = \check{\pi}^\vee \simeq \check{\pi}'.$$

The local-global compatibility in [Genestier and Lafforgue 2017, Théorème 0.1(b)] says that

$$\begin{aligned} \pi &\simeq \check{\pi}_v \rightsquigarrow (\sigma|_{\text{Gal}(\bar{F}|F)})^{\text{ss}}, \\ \check{\pi} &\simeq (\check{\pi}_v)^\vee \simeq \check{\pi}'_v \rightsquigarrow (\sigma'|_{\text{Gal}(\bar{F}|F)})^{\text{ss}} = ({}^L\theta \circ \sigma|_{\text{Gal}(\bar{F}|F)})^{\text{ss}}. \end{aligned}$$

Here we choose an embedding of the separable closure of \mathring{F} into \bar{F} , and the semisimplification is defined as in Section 2.4. In particular, $\phi = (\sigma|_{\text{Gal}(\bar{F}|F)})^{\text{ss}}$ in $\Phi(G)$ (local version).

By Lemma 2.4.5 we have $({}^L\theta \circ \sigma|_{\text{Gal}(\bar{F}|F)})^{\text{ss}} = {}^L\theta \circ (\sigma|_{\text{Gal}(\bar{F}|F)})^{\text{ss}}$. Summarizing,

$$\check{\pi} \rightsquigarrow {}^L\theta \circ (\sigma|_{\text{Gal}(\bar{F}|F)})^{\text{ss}} = {}^L\theta \circ \phi$$

holds in $\Phi(G)$ (local version). This establishes (3-3) and the Theorem 3.2.2 follows. □

Remark 3.4.4. As pointed out by a referee, Lemma 3.4.1 can be avoided in Step 2 by the following arguments. Restrict quot $: G \rightarrow T := G/G_{\text{der}}$ to an isogeny $A_G \rightarrow T$. The same arguments show that some smooth character $\eta : T(F) \rightarrow \overline{\mathbb{Q}_\ell}^\times$ pulls back to our given $\eta_0 : A_G(F) \rightarrow \overline{\mathbb{Q}_\ell}^\times$. To complete Step 2, it remains to compare (a) the parameters of η and η^{-1} and (b) the parameters of π and $\pi \otimes \eta$. For (a), apply local trivial functoriality [Genestier and Lafforgue 2017, Théorème 8.1] to the automorphism $t \mapsto t^{-1}$ of T . For (b), apply it to the homomorphism $G \xrightarrow{(\text{id}, \text{quot})} G \times T$ with normal image, as performed in [Genestier and Lafforgue 2017, Remarque 0.3].

From Section 4 onwards, we will focus exclusively on Theorem 3.3.2 and the underlying geometric considerations.

3.5. Remarks on the duality involution. Conserve the assumptions for the local statement in Section 3.2 and assume G is quasisplit. Fix an F -pinning $\mathcal{P} = (B, T, (X_\alpha)_\alpha)$ of G . Choose the unique $\kappa \in T^{\text{ad}}(F)$ such that $\kappa X_\alpha \kappa^{-1} = -X_\alpha$, for all simple root α with respect to (B, T) . Observe that $\kappa^2 = 1$ in G^{ad} .

Let $\theta = \theta_{\mathcal{P}}$ be the Chevalley involution of G , and ι_- be the inner involution $g \mapsto \kappa g \kappa^{-1}$. Observe that $\iota_- \theta = \theta \iota_-$. Indeed, $\iota_- \theta \iota_-$ is seen to preserve \mathcal{P} and coincides with θ on T , hence $\iota_- \theta \iota_- = \theta$ by the characterization of the Chevalley involution.

Definition 3.5.1 [Prasad 2018, §3]. Relative to the F -pinning \mathcal{P} , set $\iota_{G, \mathcal{P}} := \iota_- \theta = \theta \iota_-$. It is called the *duality involution* of G .

Recall that $\iota_{G, \mathcal{P}}$ induces a pinned automorphism of \hat{G} , called the *dual* automorphism of $\iota_{G, \mathcal{P}}$, which depends only on $\iota_{G, \mathcal{P}}$ modulo $G^{\text{ad}}(F)$; see [Borel 1979, §2.5] for the general set-up. This recipe applies to any base field F .

Lemma 3.5.2. *The Chevalley involution on \hat{G} is the dual of $\iota_{G,\mathcal{P}}$ in the sense above. This result holds over any field F .*

Proof. Since ι_- comes from $G^{\text{ad}}(F)$ -action, $\iota_{G,\mathcal{P}}$ and θ have the same dual. It suffices to show that the Chevalley involution of \hat{G} is dual to that of G . Since both automorphisms are pinned, it suffices to show that the induced automorphisms on $X_*(T_{\bar{F}})$ and $X^*(T_{\bar{F}})$ are mutually dual. Recall that the Chevalley involution of G and \hat{G} act on $X_*(T_{\bar{F}})$ and $X^*(T_{\bar{F}})$, respectively, as $x \mapsto -w_0(x)$, where w_0 is the longest element in the Weyl group. Since $w_0^2 = 1$, these two automorphisms are indeed mutually dual. \square

Fix a nontrivial smooth character $\psi : F \rightarrow \overline{\mathbb{Q}_\ell}^\times$. From the F -pinning $\mathcal{P} = (B, T, (X_\alpha)_\alpha)$ we produce a Whittaker datum $\mathfrak{w} := (U, \chi)$ for G taken up to $G(F)$ -conjugacy, that is,

- U is the unipotent radical of B ,
- $\chi : U(F) \rightarrow \overline{\mathbb{Q}_\ell}^\times$ is the composition of ψ with the algebraic character $U \rightarrow \mathbb{G}_a$ mapping each X_α to 1.

The automorphisms of G act on F -pinnings, thereby act on Whittaker data. Put

$$\mathfrak{w}' := (U, \chi^{-1}) = \iota_- \mathfrak{w}.$$

Fix ψ, \mathcal{P} and the associated Whittaker datum \mathfrak{w} for G . Let $\phi \in \Phi(G)$ be a semisimple parameter. Define the Genestier–Lafforgue packet Π_ϕ as in Section 3.2. We say that *Shahidi’s property* holds for Π_ϕ and \mathfrak{w} , if

$$\exists! \pi \in \Pi_\phi \quad \text{such that } \pi \text{ is } \mathfrak{w}\text{-generic.} \tag{3-4}$$

Further discussions about this property will be given in Remark 3.5.5.

Lemma 3.5.3. *The following are equivalent for an irreducible smooth representation π of $G(F)$:*

- (i) π is \mathfrak{w} -generic.
- (ii) $\pi \circ \theta$ is \mathfrak{w} -generic.
- (iii) $\check{\pi}$ is \mathfrak{w}' -generic.
- (iv) $\pi \circ \iota_-$ is \mathfrak{w}' -generic.

Proof. (i) \iff (ii) since θ preserves \mathcal{P} . (i) \iff (iii) is [Prasad 2018, §4, Lemma 2]. (i) \iff (iv) follows from transport of structure by the involution ι_- . \square

The following result serves as a partial heuristic for [Prasad 2018, §3, Conjecture 1].

Theorem 3.5.4. *Define the Whittaker data \mathfrak{w} and \mathfrak{w}' as above. Let $\phi \in \Phi(G)$ be a semisimple parameter such that Π_ϕ satisfies Shahidi’s property (3-4) with respect to \mathfrak{w} . Then the following hold:*

- (i) *The packet $\Pi_{L_{\theta \circ \phi}}$ satisfies Shahidi’s property (3-4) with respect to \mathfrak{w}' .*
- (ii) *Let π be the unique \mathfrak{w} -generic member of Π_ϕ , then $\check{\pi}$ is the unique \mathfrak{w}' -generic member of $\Pi_{L_{\theta \circ \phi}}$.*
- (iii) *If $\pi \in \Pi_\phi$ is \mathfrak{w} -generic, then $\check{\pi} \simeq \pi \circ \iota_{G,\mathcal{P}}$.*

Proof. Parts (i), (ii) follow immediately from Lemma 3.5.3 and Theorem 3.2.2, which says that $\Pi_{L_{\theta \circ \phi}} = \{\tilde{\pi} : \pi \in \Pi_{\phi}\}$.

Now consider (iii). We claim that $\pi \circ \iota_{G, \mathcal{P}} \in \Pi_{L_{\theta \circ \phi}}$. In view of Lemma 3.5.2, the corresponding statement for global Langlands parametrization of cuspidal automorphic representations follows from the “trivial functoriality” (under the dual of $\iota_{G, \mathcal{P}}$) in [Genestier and Lafforgue 2017, Théorème 0.1 and 8.1].

Using Lemma 3.5.3, we see that $\pi \circ \iota_{G, \mathcal{P}} = (\pi \circ \theta) \circ \iota_{-}$ is also a \mathfrak{w}' -generic member of $\Pi_{L_{\theta \circ \phi}}$. It follows from (ii) that $\tilde{\pi} \simeq \pi \circ \iota_{G, \mathcal{P}}$. □

Remark 3.5.5. Choose an isomorphism $\overline{\mathbb{Q}}_{\ell} \xrightarrow{\sim} \mathbb{C}$ and let $\phi \in \Phi(G)$ be semisimple. By a conjecture of Shahidi [1990, Conjecture 9.4], one expects that when ϕ is a tempered L -parameter, (3-4) will hold for the *authentic L -packet* associated to Π_{ϕ} and for any \mathfrak{w} .

On the other hand, [Gross and Prasad 1992, Conjecture 2.6] proposes a characterization of L -parameters satisfying (3-4). It is stated in terms of adjoint L -factors, thus applies directly to the ℓ -adic case. The author is grateful to Yeansu Kim for this comment.

Because of the semisimplified nature of our packet Π_{ϕ} , see Remark 3.2.1, we expect (3-4) to hold only when ϕ is not the semisimplification of any other L -parameter. This occurs when ϕ is elliptic, in which case every $\pi \in \Pi_{\phi}$ is supercuspidal: otherwise the compatibility of the parametrization $\pi \rightsquigarrow \phi$ with cuspidal supports will force ϕ to factor through some proper Levi. It is believed that the authentic L -packets for elliptic ϕ have the same property. Many constructions of such L -packets have been proposed, such as in [Kaletha 2016a]. Nonetheless, the precise relation of these packets to the Langlands parametrization of Genestier and Lafforgue [2017] remains to be settled.

Remark 3.5.6. As shown in [Prasad 2018], up to $G(F)$ -conjugacy, $\iota_{G, \mathcal{P}}$ reduces to the well-known MVW involution when G is classical; it reduces to $g \mapsto {}^t g^{-1}$ when $G = \mathrm{GL}(n)$. According to [Prasad 2018, §3, Corollary 1], when Z_G is an elementary 2-group, $\iota_{G, \mathcal{P}}$ is independent of \mathcal{P} up to $G(F)$ -conjugacy.

4. Overview of the global Langlands parametrization

4.1. Geometric setup. Fix some power q of a prime number p . Take $E \subset \overline{\mathbb{Q}}_{\ell}$ to be a finite extension of \mathbb{Q}_{ℓ} containing a square root $q^{1/2}$ of q , which we fix once and for all. The sheaves and complexes under consideration will be E -linear.

Suppose that S is a smooth \mathbb{F}_q -scheme of finite type and of pure dimension d . For any reasonable algebraic stack \mathcal{X} equipped with a morphism $\mathfrak{p} : \mathcal{X} \rightarrow S$, define the *normalized perverse sheaves* on \mathcal{X} with respect to S to be of the form

$$\mathcal{F}[-d]\left(-\frac{d}{2}\right), \quad \mathcal{F} \text{ a nonnormalized perverse sheaf.}$$

The usual operations on constructible complexes continue to hold in the normalized setting, with the proviso that the dualizing complex in [Laszlo and Olsson 2008b, §7.3] becomes

$$\Omega_{\mathcal{X}} := (\text{the nonnormalized one})[-2d](-d) \simeq \mathfrak{p}^!(E_S)$$

and the duality operator becomes $\mathbb{D} = R\mathcal{H}om(-, \Omega_X)$ accordingly. This formalism extends to ind-stacks, etc. with a morphism to S . When $S = \text{Spec } \mathbb{F}_q$, we revert to the usual definitions.

Next, assume $\mathring{F} = \mathbb{F}_q(X)$ is a global field and G is a connected reductive \mathring{F} -group with a chosen Bruhat–Tits model over X , as in Section 2.1. Fix a maximal \mathring{F} -torus $T \subset G$. Also recall that \hat{G} carries a Galois-stable pinning $(\hat{B}, \hat{T}, (X_\alpha)_\alpha)$. Enlarging E if necessary, we can assume that:

All irreducible $\overline{\mathbb{Q}}_\ell$ -representations of ${}^L G$ are realized over E .

Fix a partition of a finite set

$$I = I_1 \sqcup \cdots \sqcup I_k$$

used to label points on X and a level $N \subset X$. Set $\hat{N} = |N| \cup (X \setminus U)$ as in (3-1). Define the *Hecke stack* $\text{Hecke}_{N,I}^{(I_1, \dots, I_k)}$ that maps each \mathbb{F}_q -scheme S to the groupoid

$$\text{Hecke}_{N,I}^{(I_1, \dots, I_k)}(S) = \left\{ \begin{array}{l} (x_i)_{i \in (X \setminus \hat{N})(S)}^I, \\ ((\mathcal{G}_j, \psi_j) \in \text{Bun}_{G,N}(S))_{j=0}^k, \\ \phi_j : \mathcal{G}_{j-1} \dashrightarrow \mathcal{G}_j \end{array} \left| \begin{array}{l} \phi_j \text{ defined off } \bigcup_{i \in I_j} \Gamma_{x_i}, \\ \psi_j \phi_j|_{N \times S} = \psi_{j-1} \\ \forall j = 1, \dots, k \end{array} \right. \right\} \quad (4-1)$$

where Γ_{x_i} stands for the graph of $x_i : S \rightarrow X$. The points $(x_i)_{i \in I}$ are known as the “paws”.

The reason for partitioning I into I_1, \dots, I_k is to define *partial Frobenius morphisms*, see Section 4.3.

The ind-scheme $\text{Gr}_I^{(I_1, \dots, I_k)}$, the factorization version of affine Grassmannian of Beilinson–Drinfeld, is the space classifying the same data (4-1) as $\text{Hecke}_{I, \emptyset}^{(I_1, \dots, I_k)}$ together with a trivialization θ of \mathcal{G}_k . It also admits a morphism of “paws” to X^I . In fact $\text{Gr}_I^{(I_1, \dots, I_k)}$ is ind-projective; we refer to [Lafforgue 2018, §1] for further details. When I is a singleton and $k = 1$, the usual Beilinson–Drinfeld Grassmannian over X is recovered.

The *factorization structure* here means that given a surjection $\zeta : I \rightarrow J$, we have, for $U_\zeta := \{(x_i)_{i \in I} : \zeta(a) \neq \zeta(b) \implies x_a \neq x_b\} \subset X^I$ and $I'_a := I_a \cap \zeta^{-1}(j), \forall j$, the canonical isomorphism

$$\text{Gr}_I^{(I_1, \dots, I_k)} \times_{X^I} U_\zeta \xrightarrow{\sim} \prod_{j \in J} \text{Gr}_{\zeta^{-1}(j)}^{(I'_1, \dots, I'_k)}$$

over U_ζ see [Lafforgue 2018, Remarque 1.9]. The factorization structure is mainly to be employed together with the complexes that are *universally locally acyclic*, hereafter abbreviated as *ULA*, with respect to the base (say X^I). This property (see [Richarz 2014, §3.2] or [Braverman and Gaitsgory 2002, §5.1]) is immensely useful for “spreading out” certain properties of complexes from some open subset in the base, see e.g., [Richarz 2014, Theorem 3.16].

Given $(n_i)_i \in \mathbb{Z}_{\geq 0}^I$. Define $\Gamma_{\sum_i n_i x_i} \subset X \times X^I$ as the closed subscheme Zariski-locally defined by $\prod_{i \in I} t_i^{n_i}$, with t_i being a local equation for x_i in X , where $(x_i)_{i \in I}$ are the aforementioned “paws”. Then define

$$G_{\sum_{i \in I} n_i x_i} := \text{the Weil restriction of } G \text{ with respect to } \Gamma_{\sum_i n_i x_i} \rightarrow X^I.$$

One interprets $G_{\sum_i \infty x_i}$ in the same manner by considering formal neighborhoods, but we won't go into the details.

As in the discussion preceding [Lafforgue 2018, Proposition 1.10], there is a notion of $G_{\sum_i \infty x_i}$ -action on $\mathrm{Gr}_I^{(I_1, \dots, I_k)}$, namely by altering the trivialization θ of \mathcal{G}_k at $\Gamma_{\sum_i \infty x_i}$.

Let $\mathrm{Perv}_{G_{\sum_i \infty x_i}}(\mathrm{Gr}_I^{(I_1, \dots, I_k)})$ denote the category of $G_{\sum_i \infty x_i}$ -equivariant normalized perverse sheaves on the ind-scheme $\mathrm{Gr}_I^{(I_1, \dots, I_k)}$ relative to X^I ; for nonsplit G , we confine ourselves to $(X \setminus \hat{N})^I$ as in [Lafforgue 2018, §12.3.1]. The factorization version of *geometric Satake equivalence* [Lafforgue 2018, Théorèmes 1.17 and 12.16] gives an additive functor

$$\begin{aligned} \mathrm{Rep}_E({}^L G)^I &\rightarrow \mathrm{Perv}_{G_{\sum_i \infty x_i}}(\mathrm{Gr}_I^{(I_1, \dots, I_k)}) \\ W &\mapsto \mathcal{S}_{I, W, E}^{(I_1, \dots, I_k)}. \end{aligned}$$

For later reference, we record some of the basic properties of this functor, all of which can be found in [loc. cit.]:

- (1) The normalized perverse sheaves $\mathcal{S}_{I, W, E}^{(I_1, \dots, I_k)}$ are ULA relative to the morphism to X^I (or $(X \setminus \hat{N})^I$).
- (2) When $|I| = 1$, the geometric Satake equivalence [Richarz 2014; Zhu 2015] yields $W \mapsto \mathcal{S}_{I, W, E}^{(I)}$. This extends to general I and “factorizable” W using the factorization structure on affine Grassmannians, see [Lafforgue 2018]. Namely, for any family $(W_i)_{i \in I}$ of objects in $\mathrm{Rep}_E({}^L G)$, one can associate $\mathcal{S}_{I, \boxtimes_i W_i, E}^{(I_1, \dots, I_k)}$ in $\mathrm{Perv}_{G_{\sum_i \infty x_i}}(\mathrm{Gr}_I^{(I_1, \dots, I_k)})$.
- (3) Write ${}^L G^I$ for $({}^L G)^I$. In order to obtain a functorial construction in all $W \in \mathrm{Rep}_E({}^L G^I)$, we take the ${}^L G^I \times {}^L G^I$ -representation $\mathcal{R} := \boxtimes_{i \in I} \theta({}^L G)$ over E . This becomes an ind-object of $\mathrm{Rep}_E({}^L G^I)$ using the ${}^L G^I$ -action on the first slot, and this ind-object carries a ${}^L G^I$ -action from the second slot. Take a system of representatives of irreducible objects $V \in \mathrm{Rep}_E({}^L G)$. As ${}^L G^I \times {}^L G^I$ -representations, we have

$$\bigoplus_{V: \text{irred}} V \otimes_E V^\vee \xrightarrow{\sim} \mathcal{R} \quad \text{by taking matrix coefficients.}$$

The decomposition above and the available $\mathcal{S}_{I, V, E}^{(I_1, \dots, I_k)}$ define a normalized ind-perverse sheaf $\mathcal{S}_{I, \mathcal{R}, E}^{(I_1, \dots, I_k)}$, with the \hat{G}^I -action inherited from the second slot of \mathcal{R} . Now we can define, for each $W \in \mathrm{Rep}_E({}^L G^I)$,

$$\mathcal{S}_{I, W, E}^{(I_1, \dots, I_k)} := (\mathcal{S}_{I, \mathcal{R}, E}^{(I_1, \dots, I_k)} \otimes_E W)^{{}^L G^I} \simeq \bigoplus_{V: \text{irred}} \mathcal{S}_{I, V, E}^{(I_1, \dots, I_k)} \otimes_E (V^\vee \otimes_E W)^{{}^L G^I} \simeq \bigoplus_{V: \text{irred}} \mathcal{S}_{I, V, E}^{(I_1, \dots, I_k)} \otimes_E \mathfrak{W}_V, \quad (4-2)$$

where: (a) W is viewed as a constant sheaf on $\mathrm{Gr}_I^{(I_1, \dots, I_k)}$. (b) ${}^L G^I$ acts diagonally. (c) \mathfrak{W}_V stands for the multiplicity space of V in W . Functoriality in W is clear, and it is readily seen to agree with the previous step if $W = V$, up to isomorphism.

Given $W \in \mathrm{Rep}_E({}^L G^I)$, we define the reduced closed subscheme

$$\mathrm{Gr}_{I, W}^{(I_1, \dots, I_k)} := \mathrm{Supp} \mathcal{S}_{I, W, E}^{(I_1, \dots, I_k)} \subset \mathrm{Gr}_I^{(I_1, \dots, I_k)}.$$

In this manner, the objects of $\mathrm{Rep}_E({}^L G^I)$ will serve as truncation parameters for $\mathrm{Gr}_{I, W}^{(I_1, \dots, I_k)}$. For the traditional definition in terms of weights and relative positions, see [Lafforgue 2018, Définition 1.12].

When $|I| = 1$ and W is irreducible, $\mathcal{S}_{I,W,E}^{(I)}$ is well known to be isomorphic to the normalized IC-complex of the stratum $\text{Gr}_{I,W}^{(I)}$.

We move to the *moduli stack of chtoucas* with level structures, whose the details can be found in [Lafforgue 2018, §2, §12.3.2]. For $I = I_1 \sqcup \dots \sqcup I_k$ and N as before, $\text{Cht}_{N,I}^{(I_1, \dots, I_k)}$ is defined by a pull-back diagram

$$\begin{array}{ccc} \text{Cht}_{N,I}^{(I_1, \dots, I_k)} & \longrightarrow & \text{Hecke}_{N,I}^{(I_1, \dots, I_k)} \\ \downarrow & \square & \downarrow (\mathcal{G}_0, \mathcal{G}_k) \\ \text{Bun}_{G,N} & \xrightarrow{\text{id} \times \text{Frob}} & \text{Bun}_{G,N} \times \text{Bun}_{G,N} \end{array}$$

of ind-stacks over \mathbb{F}_q . It classifies the chains

$$(\mathcal{G}_0, \psi_0) \xrightarrow{\phi_1} \dots \xrightarrow{\phi_{k-1}} (\mathcal{G}_k, \psi_k) \xrightarrow{\phi_k} (\tau \mathcal{G}_0, \tau \psi_0)$$

of G -torsors with N -level structures (see (4-1)). Here, for every \mathbb{F}_q -scheme S and $(\mathcal{G}, \psi) \in \text{Bun}_{G,N}(S)$ we set

$$(\tau \mathcal{G}, \tau \psi) := (\text{id}_X \times \text{Frob}_S)^*(\mathcal{G}, \psi)$$

and similarly for the morphisms in $\text{Bun}_{G,N}(S)$. Note that $\text{Cht}_{N,I}^{(I_1, \dots, I_k)}$ is an ind-stack of ind-finite type over \mathbb{F}_q endowed with a morphism of ‘‘paws’’

$$\mathfrak{p}_{N,I}^{(I_1, \dots, I_k)} : \text{Cht}_{N,I}^{(I_1, \dots, I_k)} \rightarrow (X \setminus \hat{N})^I$$

coming from that of $\text{Hecke}_{N,I}^{(I_1, \dots, I_k)}$. Stability conditions of Harder–Narasimhan type attached to dominant coweights $\mu \in X_*(T^{\text{ad}})$ of G^{ad} on the datum \mathcal{G}_0 gives rise to the truncated piece $\text{Cht}_{N,I}^{(I_1, \dots, I_k), \leq \mu}$. Choose any Borel subgroup (over the separable closure) of G containing T . For coweights μ, μ' , write

$$\mu' \geq \mu \iff \mu' - \mu \in \sum_{\check{\alpha}: \text{simple coroot}} \mathbb{Q}_{\geq 0} \cdot \check{\alpha}.$$

As μ grows with respect to \geq , we have the filtered limit

$$\text{Cht}_{N,I}^{(I_1, \dots, I_k)} = \varinjlim_{\mu} \text{Cht}_{N,I}^{(I_1, \dots, I_k), \leq \mu}.$$

Exactly as in the case of affine Grassmannians, there is another truncation indexed by $W \in \text{Rep}_E(({}^L G)^I)$; see [Lafforgue 2018, §2] for details. They give rise to

$$\text{Cht}_{N,I,W}^{(I_1, \dots, I_k)} \stackrel{\text{open}}{\supseteq} \text{Cht}_{N,I,W}^{(I_1, \dots, I_k), \leq \mu}.$$

By [Lafforgue 2018, Proposition 2.6], $\text{Cht}_{N,I,W}^{(I_1, \dots, I_k)}$ is a reduced Deligne–Mumford stack locally of finite type over $(X \setminus \hat{N})^I$, for any W . The connected components of an open substack of the form $\text{Cht}_{N,I,W}^{(I_1, \dots, I_k), \leq \mu}$ are quotients of quasiprojective $(X \setminus \hat{N})^I$ -schemes by finite groups; when N is large relative to μ and to the highest weights of W , those connected components are even quasiprojective $(X \setminus \hat{N})^I$ -schemes. The

last property can serve to justify some geometric reasoning over such stacks, by reducing them to the usual scheme-theoretic setting.

We have $Z_G(\mathring{F}) \backslash Z_G(\mathbb{A}) \hookrightarrow \text{Bun}_{Z_G, N}(\mathbb{F}_q)$, and the latter acts on $\text{Cht}_{N, I}^{(I_1, \dots, I_k)}$ by twisting G -torsors by Z_G -torsors. This action leaves each truncated piece invariant. In particular, for a lattice $\Xi \subset Z_G(\mathring{F}) \backslash Z_G(\mathbb{A})$ chosen as in Section 2.1, we have Ξ -action on $\text{Cht}_{N, I, W}^{(I_1, \dots, I_k), \leq \mu}$, etc. One can shrink Ξ to make it act freely, and consider the quotients $\text{Cht}_{N, I, W}^{(I_1, \dots, I_k), \leq \mu} / \Xi$, etc.

By the discussions before [Lafforgue 2018, Définition 2.14], $\text{Cht}_{N, I, W}^{(I_1, \dots, I_k), \leq \mu} / \Xi$ is a Deligne–Mumford stack of finite type.

4.2. Cohomologies. We keep the notation from Section 4.1. In what follows, normalization of perverse sheaves will always be with respect to the base $(X \setminus \hat{N})^I$.

The first ingredient [Lafforgue 2018, Proposition 2.8] is a canonical smooth morphism

$$\epsilon_{(I), W, \underline{n}}^{(I_1, \dots, I_k)} : \text{Cht}_{N, I, W}^{(I_1, \dots, I_k)} \rightarrow \text{Gr}_{I, W}^{(I_1, \dots, I_k)} / G_{\sum_{i \in I} n_i x_i}$$

where $\underline{n} = (n_i)_{i \in I} \in \mathbb{Z}_{\geq 0}^I$ is sufficiently positive with respect to $W \in \text{Rep}_E(({}^L G)^I)$, so that the $G_{\sum_i \infty x_i}$ -action factors through $G_{\sum_i n_i x_i}$. Assume furthermore that $W = \boxtimes_{j=1}^k W_j$ where each $W_j \in \text{Rep}_E(({}^L G)^{I_j})$ is irreducible. In [Lafforgue 2018, (2.5)] the canonical smooth morphism

$$\epsilon_{(I_1, \dots, I_k), W, \underline{n}}^{(I_1, \dots, I_k)} : \text{Cht}_{N, I, W}^{(I_1, \dots, I_k)} \rightarrow \prod_{j=1}^k \text{Gr}_{I_j, W_j}^{(I_j)} / G_{\sum_{i \in I_j} n_i x_i}$$

is constructed. These two are related by the canonical smooth morphism [Lafforgue 2018, (1.12)]

$$\kappa_{I, W}^{(I_1, \dots, I_k)} : \text{Gr}_{I, W}^{(I_1, \dots, I_k)} \rightarrow \prod_{j=1}^k \text{Gr}_{I_j, W_j}^{(I_j)} / G_{\sum_{i \in I_j} n_i x_i}$$

that chops a chain $\mathcal{G}_0 \dashrightarrow \mathcal{G}_1 \dashrightarrow \dots \dashrightarrow \mathcal{G}_k$ (the trivialization forgotten) classified by $\text{Gr}_I^{(I_1, \dots, I_k)}$ into segments indexed by I_j . By [Lafforgue 2018, (1.13)], when $m_i \gg n_i$ it factorizes through a smooth

$$\tilde{\kappa}_{I, W}^{(I_1, \dots, I_k)} : \text{Gr}_{I, W}^{(I_1, \dots, I_k)} / G_{\sum_{i \in I} m_i x_i} \rightarrow \prod_{j=1}^k \text{Gr}_{I_j, W_j}^{(I_j)} / G_{\sum_{i \in I_j} n_i x_i}.$$

For an interesting result on *local models* of $\text{Cht}_{N, I, W}^{(I_1, \dots, I_k)}$ based on these morphisms, see [Lafforgue 2018, Proposition 2.11]. However, we do not need that result in this article.

As an application, for each $W \in \text{Rep}_E(({}^L G)^I)$ we take the normalized perverse sheaf $\mathcal{S}_{I, W, E}^{(I_1, \dots, I_k)}$ on $\text{Gr}_{I, W}^{(I_1, \dots, I_k)}$. Descend this complex to $\text{Gr}_{I, W}^{(I_1, \dots, I_k)} / G_{\sum_{i \in I} n_i x_i}$ by its equivariance given by geometric Satake. Hence on can form the complex $(\epsilon_{I, W, \underline{n}}^{(I_1, \dots, I_k)})_* \mathcal{S}_{I, W, E}^{(I_1, \dots, I_k)}$.

Since $\mathcal{S}_{I, W, E}^{(I_1, \dots, I_k)}$ is ULA with respect to $(X \setminus \hat{N})^I$, so is its inverse image via the smooth morphism $\epsilon_{I, W, \underline{n}}^{(I_1, \dots, I_k)}$; see [Braverman and Gaitsgory 2002, 5.1.2, item 2]. We claim that the complex $(\epsilon_{I, W, \underline{n}}^{(I_1, \dots, I_k)})_* \mathcal{S}_{I, W, E}^{(I_1, \dots, I_k)}$ is moreover normalized perverse on $\text{Cht}_{N, I, W}^{(I_1, \dots, I_k)}$ for irreducible $W = \boxtimes_{j=1}^k W_j$.

Indeed, the claim is a routine consequence of the factorization structure on $\mathrm{Gr}_{I,W}^{(I_1, \dots, I_k)}$ and the ULA property, smoothness, etc.

This completes our construction when G is semisimple. In general, one has to consider a lattice Ξ as in Section 2.1. According to [Lafforgue 2018, Remarque 1.20], $\mathcal{S}_{I,W,E}^{(I_1, \dots, I_k)}$ descends to $\mathrm{Gr}_{I,W}^{(I_1, \dots, I_k)} / G_{\sum_i n_i x_i}^{\mathrm{ad}}$. By the discussions after [loc. cit., Définition 2.14], $\epsilon_{N,I,W,\underline{n}}^{(I_1, \dots, I_k)}$ induces

$$\epsilon_{N,I,W,\underline{n}}^{(I_1, \dots, I_k), \Xi} : \mathrm{Cht}_{N,I,W}^{(I_1, \dots, I_k)} / \Xi \rightarrow \mathrm{Gr}_{I,W}^{(I_1, \dots, I_k)} / G_{\sum_i n_i x_i}^{\mathrm{ad}}$$

which is smooth of relative dimension equal to $\dim G_{\sum_i n_i x_i}^{\mathrm{ad}}$. We define accordingly

$$\mathcal{F}_{N,I,W,\Xi,E}^{(I_1, \dots, I_k)} := (\epsilon_{N,I,W,\underline{n}}^{(I_1, \dots, I_k), \Xi})^* \mathcal{S}_{I,W,E}^{(I_1, \dots, I_k)}. \tag{4-3}$$

This is still a normalized perverse sheaf on $\mathrm{Cht}_{N,I,W}^{(I_1, \dots, I_k)} / \Xi$. In [loc. cit.], one actually deduces that $\mathcal{F}_{N,I,W,\Xi,E}^{(I_1, \dots, I_k)}$ is isomorphic to the normalized IC-sheaf on $\mathrm{Cht}_{N,I,W}^{(I_1, \dots, I_k)} / \Xi$.

Thus far we have assumed $W = \boxtimes_{j=1}^k W_j$. A general definition, functorial in arbitrary $W \in \mathrm{Rep}_E(({}^L G)^I)$, can be crafted by repeating the construction for $W \mapsto \mathcal{S}_{I,W,E}^{(I_1, \dots, I_k)}$ reviewed in Section 4.1. The result still takes the form (4-3), except that the right-hand side is now constructed functorially in $W \in \mathrm{Rep}_E(({}^L G)^I)$; see [loc. cit., §4.5].

Next, introduce the other truncation parameter μ from Section 4.1. The morphism of paws induces

$$\mathfrak{p}_{N,I}^{(I_1, \dots, I_k), \leq \mu} : \mathrm{Cht}_{N,I,W}^{(I_1, \dots, I_k), \leq \mu} / \Xi \rightarrow (X \setminus \hat{N})^I.$$

Recall that $\mathrm{Cht}_{N,I,W}^{(I_1, \dots, I_k), \leq \mu}$ is open and Ξ -invariant in $\mathrm{Cht}_{N,I,W}^{(I_1, \dots, I_k)}$. Define

$$\begin{aligned} \mathcal{H}_{N,I,W}^{\leq \mu, E} &:= (\mathfrak{p}_{N,I}^{(I_1, \dots, I_k), \leq \mu})^! \mathcal{F}_{N,I,W,\Xi,E}^{(I_1, \dots, I_k)} \Big|_{\mathrm{Cht}_{N,I,W}^{(I_1, \dots, I_k), \leq \mu} / \Xi}, \\ \mathcal{H}_{N,I,W}^{i, \leq \mu, E} &:= \mathbf{H}^i \mathcal{H}_{N,I,W}^{\leq \mu, E}, \end{aligned} \tag{4-4}$$

$i \in \mathbb{Z}$, here \mathbf{H}^i is taken with respect to the ordinary t -structure on $\mathrm{D}_c((X \setminus \hat{N})^I, E)$.

- By using the forgetful morphisms as in [Lafforgue 2018, Construction 2.7 and Corollaire 2.18], these complexes are seen to be independent of the partition (I_1, \dots, I_k) . The notation in (4-4) is thus justified.
- For $\mu \leq \mu'$, the open immersion $j : \mathrm{Cht}_{N,I,W}^{(I_1, \dots, I_k), \leq \mu} / \Xi \rightarrow \mathrm{Cht}_{N,I,W}^{(I_1, \dots, I_k), \leq \mu'} / \Xi$ induces a canonical arrow $\mathcal{H}_{N,I,W}^{\leq \mu, E} \rightarrow \mathcal{H}_{N,I,W}^{\leq \mu', E}$. This is a standard consequence of the formalism of six operations as $j^* = j^!$.
- They also respect the *coalescence* of paws with respect to any map $\zeta : I \rightarrow J$. We refer to [Lafforgue 2018, Proposition 4.12] for further explanations.

Let I be a finite set and $W \in \mathrm{Rep}_E(({}^L G)^I)$ arbitrary. Denote the generic point of X and X^I by η and η^I , respectively, and choose geometric points over them

$$\bar{\eta} \rightarrow \eta, \quad \bar{\eta}^I \rightarrow \eta^I.$$

Let $\Delta : X \rightarrow X^I$ be the diagonal embedding. Following [Lafforgue 2018, §8] or [Varshavsky 2007, §1.3], we choose an *arrow of specialization*

$$\mathfrak{sp} : \bar{\eta}^I \rightarrow \Delta(\bar{\eta}),$$

i.e., a morphism $(X^I)_{(\bar{\eta}^I)} \rightarrow (X^I)_{(\Delta(\bar{\eta}))}$ or equivalently $\bar{\eta}^I \rightarrow (X^I)_{(\Delta(\bar{\eta}))}$, where the subscripts indicate strict Henselizations at the corresponding geometric points. By [Lafforgue 2018, Proposition 8.24], the induced pull-back morphism

$$\mathfrak{sp}^* : \varinjlim_{\mu} \mathcal{H}_{N,I,W}^{0,\leq\mu,E} \Big|_{\Delta(\bar{\eta})} \rightarrow \varinjlim_{\mu} \mathcal{H}_{N,I,W}^{0,\leq\mu,E} \Big|_{\bar{\eta}^I}$$

between E -vector spaces is injective.

Now comes the Hecke action. Let $f \in C_c(K_N \backslash G(\mathbb{A})/K_N; E)$. According to [Lafforgue 2018, Corollaire 6.5], taking a coweight $\kappa \gg 0$ with respect to f , there is an induced morphism

$$T(f) : \mathcal{H}_{N,I,W}^{\leq\mu,E} \rightarrow \mathcal{H}_{N,I,W}^{\leq\mu+\kappa,E} \tag{4-5}$$

in $D_c^b((X \backslash \hat{N})^I, E)$, with various compatibilities. It is E -linear in f and satisfies $T(ff') = T(f)T(f')$. After passing to \varinjlim_{μ} , we are led to the left $C_c(K_N \backslash G(\mathbb{A})/K_N; E)$ -module

$$H_{I,W} := \left(\varinjlim_{\mu} \mathcal{H}_{N,I,W}^{0,\leq\mu,E} \Big|_{\Delta(\bar{\eta})} \right)^{\text{Hf}} \tag{4-6}$$

where ‘‘Hf’’ means Hecke-finite with respect to the action (4-5). This definition is clearly functorial in W . The following properties are established in [Lafforgue 2018, §§8–9]:

- Compatibility with coalescence of paws. Namely, every map $\zeta : J \rightarrow I$ induces a canonical isomorphism $\chi_{\zeta} : H_{I,W} \xrightarrow{\sim} H_{J,W^{\zeta}}$, where $W^{\zeta} \in \text{Rep}_E({}^L G^J)$ denotes the pull-back of W via ζ .
- The arrow \mathfrak{sp}^* commutes with Hecke action since the latter is defined on the level of $D_c^b((X \backslash \hat{N})^I, E)$. Moreover, it induces an isomorphism

$$\mathfrak{sp}^* : H_{I,W} \xrightarrow{\sim} \left(\varinjlim_{\mu} \mathcal{H}_{N,I,W}^{0,\leq\mu,E} \Big|_{\bar{\eta}^I} \right)^{\text{Hf}}.$$

- We have ${}^L G^{\emptyset} = \{1\}$, $\eta^{\emptyset} = \text{Spec } \mathbb{F}_q$ when $I = \emptyset$. There are natural isomorphisms

$$H_{\{0\},1} \xleftarrow{\chi} H_{\emptyset,1} \xrightarrow{\sim} C_c^{\text{cusp}}(\text{Bun}_{G,N}(\mathbb{F}_q)/\mathfrak{E}; E). \tag{4-7}$$

The arrow χ is induced by coalescence via the unique map $\emptyset \rightarrow \{0\}$. The rightward arrow stems from the fact [Varshavsky 2004, Proposition 2.16(c)] that $\text{Cht}_{N,\emptyset,1}/\mathfrak{E}$ is the constant stack $\text{Bun}_{G,N}(\mathbb{F}_q)/\mathfrak{E}$ over $\text{Spec } \mathbb{F}_q$, which implies a canonical isomorphism

$$\varinjlim_{\mu} \mathcal{H}_{N,\emptyset,1}^{0,\leq\mu,E} \Big|_{\Delta(\bar{\eta})} \xrightarrow{\sim} C_c(\text{Bun}_{G,N}(\mathbb{F}_q)/\mathfrak{E}; E) \tag{4-8}$$

of $C_c(K_N \backslash G(\mathbb{A})/K_N; E)$ -modules.

- For $I = \{1\}$, $W = \mathbf{1}$, coalescence induces $\text{Cht}_{N, \{1\}, \mathbf{1}}^{(\{0\})} / \mathfrak{E} \xrightarrow{\sim} (\text{Cht}_{N, \emptyset, \mathbf{1}} / \mathfrak{E}) \times_{\text{Spec } \mathbb{F}_q} (X \setminus \hat{N})$ by [Lafforgue 2018, (8.4)]. In this case, $\mathcal{H}_{N, \{0\}, \mathbf{1}}^{0, \leq \mu, E}$ is a constant sheaf and the \varinjlim_{μ} of its stalk at $\bar{\eta}$ is still $C_c(\text{Bun}_{G, N}(\mathbb{F}_q) / \mathfrak{E}; E)$.
- Via these isomorphisms, the $C_c(K_N \backslash G(\mathbb{A}) / K_N; E)$ -module structures on $H_{\emptyset, \mathbf{1}}$ and $H_{\{0\}, \mathbf{1}}$ match the one on $C_c^{\text{cusp}}(\text{Bun}_{G, N}(\mathbb{F}_q) / \mathfrak{E}; E)$ recorded in Section 2.1. See [Lafforgue 2018, §8].

The last item above is how harmonic analysis enters the geometric picture.

4.3. Partial Frobenius morphisms and Galois actions. We conserve the previous conventions and review the partial Frobenius morphisms. Let $J \subset I$ be finite sets. Choose a partition $I = I_1 \sqcup \cdots \sqcup I_k$ with $I_1 = J$, together with a specialization arrow $\mathfrak{sp} : \bar{\eta}^I \rightarrow \Delta(\bar{\eta})$. The choice of partition intervenes in the constructions, but will disappear in the final results.

Let $\text{Frob}_J = \text{Frob}_{I_1} : (X \setminus \hat{N})^I \rightarrow (X \setminus \hat{N})^I$ be the morphism that equals Frob on the coordinates indexed by I_1 , and id elsewhere.

Take $W \in \text{Rep}_E({}^L G)^I$ as well the lattice \mathfrak{E} as in Section 4.2. In [Lafforgue 2018, §3] is defined the partial Frobenius morphism

$$\text{Frob}_{I_1, N}^{(I_1, \dots, I_k)} : \text{Cht}_{N, I, W}^{(I_1, \dots, I_k)} \rightarrow \text{Cht}_{N, I, W}^{(I_2, \dots, I_k, I_1)} \tag{4-9}$$

covering Frob_{I_1} , that respects \mathfrak{E} -actions. In terms of the notations in Section 4.1, it sends the chain

$$(\mathcal{G}_0, \psi_0) \xrightarrow{\phi_1} \cdots \rightarrow (\mathcal{G}_k, \psi_k) \rightarrow (\tau \mathcal{G}_0, \tau \psi_0)$$

into

$$(\mathcal{G}_1, \psi_1) \xrightarrow{\phi_2} \cdots \rightarrow (\mathcal{G}_k, \psi_k) \rightarrow (\tau \mathcal{G}_0, \tau \psi_0) \xrightarrow{\tau \phi_1} (\tau \mathcal{G}_1, \tau \psi_1)$$

whereas the paws are transformed accordingly by Frob_{I_1} . The cyclic composition of k partial Frobenius morphism equals the total Frobenius endomorphism of $\text{Cht}_{N, I, W}^{(I_1, \dots, I_k)}$. An easy consequence is that Frob_{I_1} is a *universal homeomorphism*; see [Stacks 2005–, Tag 04DC].

The induced morphism between the quotients by \mathfrak{E} is also named $\text{Frob}_{I_1, N}^{(I_1, \dots, I_k)}$. Now introduce the dominant coweight μ of G^{ad} in Section 4.2 as truncation parameter. A basic fact is that whenever $\mu' \gg \mu$ with respect to W ,

$$(\text{Frob}_{I_1, N}^{(I_1, \dots, I_k)})^{-1} \text{Cht}_{N, I, W}^{(I_2, \dots, I_k, I_1), \leq \mu} \subset \text{Cht}_{N, I, W}^{(I_1, \dots, I_k), \leq \mu'}. \tag{4-10}$$

When $k = 1$, we have the usual Frobenius correspondence $\Phi : \text{Frob}^* \mathcal{S}_{I, W, E}^{(I)} \xrightarrow{\sim} \mathcal{S}_{I, W, E}^{(I)}$ between normalized perverse sheaves on $\text{Gr}_{I, W}^{(I)} / G_{\sum_i n_i x_i}^{\text{ad}}$. In general, by writing

$$\text{Frob}_{I_1}(x_i)_{i \in I} = (x'_i)_{i \in I}, \quad (x_i)_{i \in I} \in (X \setminus \hat{N})^I(S) \quad \forall S : \mathbb{F}_q\text{-scheme}$$

and supposing $W = \boxtimes_{j=1}^k W_j$ is irreducible, there is a commutative diagram:

$$\begin{array}{ccc}
 \text{Cht}_{N,I,W}^{(I_1, \dots, I_k)} / \mathfrak{E} & \xrightarrow{\text{Frob}_{I_1, N}^{(I_1, \dots, I_k)}} & \text{Cht}_{N,I,W}^{(I_2, \dots, I_k, I_1)} / \mathfrak{E} \\
 \downarrow \epsilon_{N, (I_1, \dots, I_k), \mathfrak{E}}^{(I_1, \dots, I_k)} & & \downarrow \epsilon_{N, (I_2, \dots, I_k, I_1), \mathfrak{E}}^{(I_2, \dots, I_k, I_1)} \\
 \prod_{j=1}^k \text{Gr}_{I_j, W_j}^{(I_j)} / G_{\sum_{i \in I_j} n_i x_i}^{\text{ad}} & \xrightarrow{\text{Frob} \times \text{id} \times \dots \times \text{id}} & \prod_{j=1}^k \text{Gr}_{I_j, W_j}^{(I_j)} / G_{\sum_{i \in I_j} n_i x_i'}^{\text{ad}}
 \end{array}$$

In view of the constructions in Section 4.2 using the smooth morphisms ϵ_{\dots} , the ULA property, etc., we obtain a canonical isomorphism in $D_c^b(\text{Cht}_{N,I,W}^{(I_1, \dots, I_k)} / \mathfrak{E}, E)$

$$F_{I_1, N, W}^{(I_1, \dots, I_k)} : (\text{Frob}_{I_1, N}^{(I_1, \dots, I_k)})^* \mathcal{F}_{N, I, W, \mathfrak{E}, E}^{(I_2, \dots, I_1)} \xrightarrow{\sim} \mathcal{F}_{N, I, W, \mathfrak{E}, E}^{(I_1, \dots, I_k)} \tag{4-11}$$

extending the previous case $k = 1$. See [Lafforgue 2018, Proposition 3.4]. This isomorphism can be extended functorially to arbitrary $W \in \text{Rep}_E(({}^L G)^I)$ by repeating the construction for $W \mapsto \mathcal{S}_{I, W, E}^{(I_1, \dots, I_k)}$.

Abbreviate the $\text{Frob}_{I_1, N}^{(I_1, \dots, I_k)}$ on $\text{Cht}_{N,I,W}^{(I_1, \dots, I_k)} / \mathfrak{E}$ as a_1 . It fits into the commutative diagram

$$\begin{array}{ccccc}
 \text{Cht}_{N,I,W}^{(I_2, \dots, I_1), \leq \mu} / \mathfrak{E} & \xleftarrow{a_1} & a_1^{-1} (\text{Cht}_{N,I,W}^{(I_2, \dots, I_1), \leq \mu} / \mathfrak{E}) & \xrightarrow{a_2} & \text{Cht}_{N,I,W}^{(I_1, \dots, I_k), \leq \mu'} / \mathfrak{E} \\
 \downarrow \mathfrak{p} & & \downarrow \mathfrak{p} & & \downarrow \mathfrak{p} \\
 (X \setminus \hat{N})^I & \xleftarrow{\text{Frob}_{I_1}} & (X \setminus \hat{N})^I & \xlongequal{\quad} & (X \setminus \hat{N})^I
 \end{array} \tag{4-12}$$

where \mathfrak{p} denotes the self-evident morphisms of paws, a_1 is a universal homeomorphism and a_2 is an open immersion. Hence (4-11) affords a *cohomological correspondence* between bounded constructible complexes in the sense of [Varshavsky 2007, §1]: for a_1, a_2 in (4-12),

$$\text{Frob}_{I_1, N, W}^{(I_1, \dots, I_k)} : a_1^* \underbrace{\mathcal{F}_{N, I, W, \mathfrak{E}, E}^{(I_2, \dots, I_1)}}_{\text{on the } \leq \mu \text{ part}} \rightarrow a_2^! \underbrace{\mathcal{F}_{N, I, W, \mathfrak{E}, E}^{(I_1, \dots, I_k)}}_{\text{on the } \leq \mu' \text{ part}}, \quad a_2^* = a_2^! \tag{4-13}$$

The left square of (4-12) is not Cartesian; however, in the commutative diagram defined with Cartesian square

$$\begin{array}{ccccc}
 & & a_1 & & a_1^{-1} (\text{Cht}_{N,I,W}^{(I_2, \dots, I_1), \leq \mu} / \mathfrak{E}) \\
 & & \swarrow & & \swarrow \exists! \varphi \\
 \text{Cht}_{N,I,W}^{(I_2, \dots, I_1), \leq \mu} / \mathfrak{E} & \xleftarrow{\tilde{a}_1} & \text{Frob}_{I_1}^* (\text{Cht}_{N,I,W}^{(I_2, \dots, I_1), \leq \mu} / \mathfrak{E}) & & \\
 \downarrow \mathfrak{p} & \square & \downarrow \tilde{\mathfrak{p}} & & \downarrow \mathfrak{p} \\
 (X \setminus \hat{N})^I & \xleftarrow{\text{Frob}_{I_1}} & (X \setminus \hat{N})^I & \xleftarrow{\quad} & (X \setminus \hat{N})^I
 \end{array}$$

the arrow φ is a universal homeomorphism since both Frob_{I_1} and a_1 are. Therefore we obtain

$$\text{BC} : \text{Frob}_{I_1}^* \mathfrak{p}_! \xrightarrow{\sim}_{\text{bc}} \tilde{\mathfrak{p}}_! \tilde{a}_1^* \xleftarrow{\sim} \tilde{\mathfrak{p}}_! \varphi_! \varphi^* \tilde{a}_1^* \simeq \mathfrak{p}_! a_1^*. \tag{4-14}$$

Indeed, bc is the isomorphism of base change by the universal homeomorphism Frob_{I_1} [Laszlo and Olsson 2008b, 12.2]; the second isomorphism is induced by $\varphi_! \varphi^* \xrightarrow{\sim} \text{id}$, which is in turn due to the topological invariance of the étale topos (see [SGA 4₁ 1972; SGA 4₂ 1972; SGA 4₃ 1973, Exp VIII, Théorème 1.1] or [Stacks 2005–, Tag 04DY]) under the universal homeomorphism φ .

In view of (4-4), we can now define

$$F_J = F_{I_1} : \text{Frob}_{I_1}^* \mathcal{H}_{N,I,W}^{\leq \mu, E} \rightarrow \mathcal{H}_{N,I,W}^{\leq \mu', E} \tag{4-15}$$

as the composite in $D_c^b((X \setminus \hat{N})^I, E)$

$$\text{Frob}_{I_1}^* \mathfrak{p}_! \underbrace{\mathcal{F}_{N,I,W,\Xi,E}^{(I_2, \dots, I_1)}}_{\text{on } \leq \mu} \xrightarrow{\sim}_{\text{BC}} \mathfrak{p}_! a_1^* \underbrace{\mathcal{F}_{N,I,W,\Xi,E}^{(I_2, \dots, I_1)}}_{\text{on } \leq \mu} \xrightarrow{\mathfrak{p}_! \text{Frob}_{I_1, N, W}^{(I_1, \dots, I_k)}} \mathfrak{p}_! a_2^! \underbrace{\mathcal{F}_{N,I,W,\Xi,E}^{(I_1, \dots, I_k)}}_{\text{on } \leq \mu'} \rightarrow \underbrace{\mathcal{F}_{N,I,W,\Xi,E}^{(I_1, \dots, I_k)}}_{\text{on } \leq \mu'}$$

the last arrow arising from $\mathfrak{p}_! a_2^! = \mathfrak{p}_! (a_2)_! a_2^! \xrightarrow{(a_2)_! a_2^! \rightarrow \text{id}} \mathfrak{p}_!$. It is functorial in $W \in \text{Rep}_E(({}^L G)^I)$ and is shown to be compatible with the coalescence of paws in [Lafforgue 2018, §§3–4]. Hence the dependence is only on $J \subset I$.

Consequently, the morphism F_J also acts E -linearly on $\varinjlim_{\mu} \mathcal{H}_{N,I,W}^{0, \leq \mu, E} |_{\bar{\eta}^I}$. Given any partition $I = I_1 \sqcup \dots \sqcup I_k$, the actions of F_{I_1}, \dots, F_{I_k} form a commuting family whose cyclic composition equals the total Frobenius action on $\varinjlim_{\mu} \mathcal{H}_{N,I,W}^{0, \leq \mu, E} |_{\bar{\eta}^I}$.

On the other hand, the standard theory [Stacks 2005–, Tag 03QW] yields a continuous representation of $\pi_1(\eta^I, \bar{\eta}^I)$ on $\mathcal{H}_{N,I,W}^{0, \leq \mu, E} |_{\bar{\eta}^I}$ which passes to \varinjlim_{μ} .

To conclude this subsection, we recall briefly the following extension of groups

$$1 \rightarrow \ker[\pi_1(\eta^I, \bar{\eta}^I) \rightarrow \hat{\mathbb{Z}}] \rightarrow \text{FWeil}(\eta^I, \bar{\eta}^I) \rightarrow \mathbb{Z}^I \rightarrow 0.$$

We refer to [Lafforgue 2018, Remarque 8.18] and the subsequent discussions for all further details. When $|I| = 1$, it becomes the Weil group $W_{\bar{F}}$ of $\bar{F} = \mathbb{F}_q(X)$; in general there is a surjection $\text{FWeil}(\eta^I, \bar{\eta}^I) \twoheadrightarrow W_{\bar{F}}^I$ depending on the choice of \mathfrak{sp} . The surjection induces an isomorphism from the profinite completion $\text{FWeil}(\eta^I, \bar{\eta}^I)$ to that of $W_{\bar{F}}^I$, i.e., $\pi_1(\eta, \bar{\eta})^I$. As mentioned in [loc. cit.], the action on $\varinjlim_{\mu} \mathcal{H}_{N,I,W}^{0, \leq \mu, E} |_{\bar{\eta}^I}$ of

- the partial Frobenius morphisms F_J for various $J \subset I$, and
- that of $\pi_1(\eta^I, \bar{\eta}^I)$

meld into an action of $\text{FWeil}(\eta^I, \bar{\eta}^I)$. The upshot of [loc. cit., §8] is to produce a continuous E -linear $\pi_1(\eta, \bar{\eta})^I$ -action on $(\varinjlim_{\mu} \mathcal{H}_{N,I,W}^{0, \leq \mu, E} |_{\bar{\eta}^I})^{\text{Hf}}$ therefrom. In other words, one wants to factorize the $\text{FWeil}(\eta^I, \bar{\eta}^I)$ -action through its profinite completion continuously.

The key for the passage to $\pi_1(\eta, \bar{\eta})^I$ -action is *Drinfeld’s Lemma*. This method requires some finiteness conditions which in turn involve the Eichler–Shimura relations. These important issues are addressed at length in [loc. cit., §8], but they are not needed in this article.

The aforementioned continuous representation transports to $H_{I,W}$, namely

$$\vec{\gamma} \cdot f := (\mathfrak{sp}^*)^{-1}(\vec{\gamma} \cdot (\mathfrak{sp}^* f)), \quad \vec{\gamma} \in \pi_1(\eta, \bar{\eta})^I, f \in H_{I,W}.$$

This action turns out to be independent of the choice of $\bar{\eta}^I$ and \mathfrak{sp} , by [loc. cit., Lemme 9.4].

4.4. Excursion operators and pseudocharacters. Consider finite sets I, J and $W \in \text{Rep}_E(({}^L G)^I)$ and $U \in \text{Rep}_E(({}^L G)^J)$. Let ζ_I, ζ_J be the unique maps from I, J into the singleton $\{0\}$. The diagonal action on W gives $W^{\zeta_I} \in \text{Rep}_E({}^L G)$; the space of \hat{G} -invariants $(W^{\zeta_I})^{\hat{G}}$ is therefore a representation of $\text{Gal}(\tilde{F}|F)$. Denote by $(W^{\zeta_I})^{\hat{G}}|_{X \setminus \hat{N}}$ the E -lisse sheaf on $X \setminus \hat{N}$ obtained by descent. Likewise, we have $(W^{\zeta_I})^{\hat{G}}|_{X \setminus \hat{N}}$ by taking the maximal quotient of W^{ζ_I} on which \hat{G} acts trivially. A pair of morphisms

$$\begin{aligned} \mathcal{H}_{N,J,U}^{\leq \mu, E} \boxtimes (W^{\zeta_I})^{\hat{G}}|_{X \setminus \hat{N}} &\rightarrow \mathcal{H}_{N,J \sqcup I, U \boxtimes W}^{\leq \mu, E}|_{(X \setminus \hat{N})^J \times \Delta(X \setminus \hat{N})} \\ \mathcal{H}_{N,J,U}^{\leq \mu, E} \boxtimes (W^{\zeta_I})^{\hat{G}}|_{X \setminus \hat{N}} &\leftarrow \mathcal{H}_{N,J \sqcup I, U \boxtimes W}^{\leq \mu, E}|_{(X \setminus \hat{N})^J \times \Delta(X \setminus \hat{N})} \end{aligned}$$

in $D_c^b((X \setminus \hat{N})^{J \sqcup \{0\}}, E)$ are constructed in [loc. cit., (12.18), (12.19)]. Roughly speaking, they are defined via coalescence and the functoriality of \mathcal{H} with respect to $(W^{\zeta_I})^{\hat{G}} \hookrightarrow W^{\zeta_I} \twoheadrightarrow (W^{\zeta_I})_{\hat{G}}$.

Now take $J = \emptyset$ and $U = \mathbf{1}$. Let $x \in W$ and $\xi \in W^\vee$ be \hat{G} -invariant under the diagonal action, viewed as maps $E \rightarrow (W^{\zeta_I})^{\hat{G}}$ and $(W^{\zeta_I})_{\hat{G}} \rightarrow E$, respectively. Taking $\varinjlim_\mu H^0(\dots|_{\bar{\eta}})$ yields the creation and annihilation operators (see [loc. cit., Définitions 5.1, 5.2 and 12.3.4])

$$\varinjlim_\mu \mathcal{H}_{N,\emptyset,\mathbf{1}}^{0, \leq \mu, E} \begin{array}{c} \xrightarrow{C_x^\sharp} \\ \xleftarrow{C_\xi^b} \end{array} \varinjlim_\mu \mathcal{H}_{N,I,W}^{0, \leq \mu, E}|_{\Delta(\bar{\eta})}$$

between E -vector spaces. Restriction to Hecke-finite parts yields arrows

$$H_{\emptyset,\mathbf{1}} \simeq H_{\{0\},\mathbf{1}} \begin{array}{c} \xrightarrow{C_x^\sharp} \\ \xleftarrow{C_\xi^b} \end{array} H_{I,W}, \quad \text{see (4-6).}$$

Given I, W, x, ξ as above and $\vec{\gamma} = (\gamma_i)_{i \in I} \in \pi_1(\eta, \bar{\eta})^I$, the *excursion operator* $S_{I,W,x,\xi,\vec{\gamma}}$ is the composite

$$\begin{array}{ccc} H_{\{0\},\mathbf{1}} & & H_{\{0\},\mathbf{1}} \\ C_x^\sharp \downarrow & & \uparrow C_\xi^b \\ H_{I,W} & \xrightarrow[\sim]{\mathfrak{sp}^*} (\varinjlim_\mu \mathcal{H}_{N,I,W}^{0, \leq \mu, E}|_{\bar{\eta}^I})^{\text{Hf}} & \xrightarrow[\sim]{\vec{\gamma}} (\varinjlim_\mu \mathcal{H}_{N,I,W}^{0, \leq \mu, E}|_{\bar{\eta}^I})^{\text{Hf}} & \xrightarrow[\sim]{(\mathfrak{sp}^*)^{-1}} H_{I,W} \end{array}$$

Here $\vec{\gamma}$ acts in the manner reviewed in Section 4.3. Upon recalling (4-7), we obtain

$$S_{I,W,x,\xi,\vec{\gamma}} \in \text{End}_E(H_{\{0\},\mathbf{1}}) \simeq \text{End}_E(C_c^{\text{cusp}}(\text{Bun}_{G,N}(\mathbb{F}_q)/\Xi; E)).$$

Moreover, by [loc. cit., Définition-Proposition 9.1]

- we have $S_{I,W,x,\xi,\vec{\gamma}} \in \text{End}_{C_c(K_N \backslash G(\mathbb{A})/K_N; E)}(H_{\{0,1\}})$;
- the formation of $S_{I,W,x,\xi,\vec{\gamma}}$ is E -bilinear in x, ξ and continuous in $\vec{\gamma}$ for the topology on the finite-dimensional space $\text{End}_E(H_{\{0,1\}})$ induced by E ;
- let \mathcal{B}_E be the E -subalgebra of $\text{End}_{C_c(K_N \backslash G(\mathbb{A})/K_N; E)}(H_{\{0,1\}})$ generated by $S_{I,W,x,\xi,\vec{\gamma}}$ for all quintuples $(I, W, x, \xi, \vec{\gamma})$. Then \mathcal{B}_E is a finite-dimensional commutative E -algebra by [loc. cit., (10.2)].

The foregoing constructions behave well under finite extensions of the field E of coefficients. Define the $\overline{\mathbb{Q}_\ell}$ -algebra

$$\mathcal{B} := \mathcal{B}_E \otimes_E \overline{\mathbb{Q}_\ell} \subset \text{End}_{\overline{\mathbb{Q}_\ell}}(H_{\{0,1\}} \otimes_E \overline{\mathbb{Q}_\ell}).$$

Upon enlarging E , we may assume that all homomorphisms $\nu : \mathcal{B} \rightarrow \overline{\mathbb{Q}_\ell}$ (finitely many) of $\overline{\mathbb{Q}_\ell}$ -algebras are defined over E . There is a decomposition of $C_c(K_N \backslash G(\mathbb{A})/K_N; E)$ -modules into generalized eigenspaces

$$H_{\{0,1\}} = \bigoplus_{\nu} \mathfrak{H}_{\nu}, \quad \mathfrak{H}_{\nu} := \{f \in H_{\{0,1\}} : \forall T \in \mathcal{B}_E, \exists d \geq 1 \mid (T - \nu(T))^d f = 0\}. \tag{4-16}$$

Here ν ranges over the characters of \mathcal{B} , and one may take $d = \dim_E H_{\{0,1\}}$. The same holds after passing to $\overline{\mathbb{Q}_\ell}$. All in all,

$$C_c^{\text{cusp}}(\text{Bun}_{G,N}(\mathbb{F}_q)/\Xi; \overline{\mathbb{Q}_\ell}) = \bigoplus_{\nu: \mathcal{B} \rightarrow \overline{\mathbb{Q}_\ell}} \mathfrak{H}_{\nu} \quad \text{in } C_c(K_N \backslash G(\mathbb{A})/K_N; \overline{\mathbb{Q}_\ell})\text{-Mod.}$$

It is conjectured that \mathcal{B} is reduced, which will imply that $d = 1$ suffices.

The next step is to reencode the excursion operators $S_{I,W,x,\xi,\vec{\gamma}}$. Let $f(\vec{g}) = \langle \xi, \vec{g} \cdot x \rangle_{W^\vee \otimes W}$ where $\vec{g} \in ({}^L G)^I$ and $\langle \cdot, \cdot \rangle_{W^\vee \otimes W}$ is the duality pairing $W^\vee \otimes_E W \rightarrow E$. Then $f \in \mathcal{O}(\hat{G} \backslash ({}^L G)^I // \hat{G})$, where \hat{G} acts by bilateral translations through diagonal embedding. By [loc. cit., Lemme 10.6], $S_{I,W,x,\xi,\vec{\gamma}}$ depends only on $(I, f, \vec{\gamma})$. Using some algebraic version of the Peter–Weyl theorem, one can uniquely define the operators

$$S_{I,f,\vec{\gamma}} \in \mathcal{B}_E, \quad f \in \mathcal{O}(\hat{G} \backslash ({}^L G)^I // \hat{G}),$$

in a manner compatible with the original $S_{I,W,x,\xi,\vec{\gamma}}$, such that if f comes from a function $\text{Gal}(\tilde{F}|\mathring{F})^I \rightarrow E$, then $S_{I,f,\vec{\gamma}} = f(\vec{\gamma}) \cdot \text{id}$. See [loc. cit., Remarque 12.20] for further explanations.

Take $n \in \mathbb{Z}_{\geq 1}$ and $I := \{0, \dots, n\}$. Then \hat{G} acts on $({}^L G)^n$ by simultaneous conjugation. There is a natural map

$$\begin{aligned} \mathcal{O}({}^L G)^n // \hat{G} &\rightarrow \mathcal{O}({}^L G \backslash ({}^L G)^{\{0, \dots, n\}} // \hat{G}) \subset \mathcal{O}(\hat{G} \backslash ({}^L G)^{\{0, \dots, n\}} // \hat{G}) \\ f &\mapsto [\tilde{f} : (g_0, \dots, g_n) \mapsto f(g_0^{-1}g_1, \dots, g_0^{-1}g_n)]. \end{aligned} \tag{4-17}$$

When n is fixed, the operators

$$\Theta_n(f)(\vec{\gamma}) := S_{\{0, \dots, n\}, \tilde{f}, (1, \vec{\gamma})}, \quad n \in \mathbb{Z}_{\geq 1}, \vec{\gamma} \in \pi_1(\eta, \bar{\eta})^n, f \in \mathcal{O}({}^L G)^n // \hat{G} \tag{4-18}$$

in \mathcal{B}_E afford a homomorphism $\mathcal{O}(({}^L G)^n // \hat{G}) \rightarrow C(\pi_1(X \setminus \hat{N}, \bar{\eta})^n, \mathcal{B}_E)$ between E -algebras, where $C(\dots)$ denotes the algebra of continuous functions under pointwise operations. See [loc. cit., Proposition 10.10] for the passage to $\pi_1(X \setminus \hat{N}, \bar{\eta})$.

Since $\mathcal{O}({}^L G // ({}^L G)^{\{0, \dots, n\}} // \hat{G}) \subsetneq \mathcal{O}(\hat{G} // ({}^L G)^{\{0, \dots, n\}} // \hat{G})$ in general, the map (4-17) is not always surjective. Nonetheless, the operators $S_{\{0, \dots, n\}, \tilde{f}, (1, \vec{\gamma})}$ still generate \mathcal{B}_E as $n, f, \vec{\gamma}$ vary; see [loc. cit., Remarque 12.20].

Finally, the machinery of ${}^L G$ -pseudocharacters associates a semisimple L -parameter $\sigma \in \Phi(G)$ to any character $\nu : \mathcal{B} \rightarrow \overline{\mathbb{Q}}_\ell$, characterized as follows:

- Version 1: for all $n \in \mathbb{Z}_{\geq 1}$, $\vec{\gamma} = (\gamma_1, \dots, \gamma_n)$ and $f \in \mathcal{O}(({}^L G)^n // \hat{G})$, we have (see [loc. cit., Proposition 11.7])

$$f(\sigma(\gamma_1), \dots, \sigma(\gamma_n)) = \nu \circ \Theta_n(f)(\vec{\gamma}).$$

- Version 2: for all $n \in \mathbb{Z}_{\geq 1}$, $\vec{\gamma} = (\gamma_1, \dots, \gamma_n)$ and $\tilde{f} \in \mathcal{O}(\hat{G} // ({}^L G)^{\{0, \dots, n\}} // \hat{G})$, we have

$$\tilde{f}(\sigma(1), \sigma(\gamma_1), \dots, \sigma(\gamma_n)) = \nu(S_{\{0, \dots, n\}, \tilde{f}, (1, \vec{\gamma})}). \tag{4-19}$$

The version 2 above is *a priori* stronger, but they are actually equivalent by the preceding remarks on generators.

By the discussions preceding [loc. cit., Remarque 12.21], the map $\text{Hom}_{\overline{\mathbb{Q}}_\ell\text{-Alg}}(\mathcal{B}, \overline{\mathbb{Q}}_\ell) \rightarrow \Phi(G)$ above is injective. Hence we may write $\mathfrak{H}_\sigma = \mathfrak{H}_\nu$ if $\nu \mapsto \sigma \in \Phi(G)$, and set $\mathfrak{H}_\sigma = \{0\}$ if σ does not match any ν . This leads to the desired decomposition (3-2).

5. The transposes of excursion operators

5.1. On Verdier duality. Retain the notation of Section 4.2. Among them, we recall only two points:

- (i) The duality operator \mathbb{D} is normalized with respect to $(X \setminus \hat{N})^I$.
- (ii) $W^{\vee, \theta}$ denotes the contragredient of $W \in \text{Rep}_E(({}^L G)^I)$ twisted by the Chevalley involution of $({}^L G)^I$.

The following results are recorded in [loc. cit., Remarque 5.4]. For the benefit of the readers, we will give some more details below.

Proposition 5.1.1. *There is a canonical isomorphism*

$$\mathbb{D} \mathcal{S}_{I, W, E}^{(I_1, \dots, I_k)} \xrightarrow{\sim} \mathcal{S}_{I, W^{\vee, \theta}, E}^{(I_1, \dots, I_k)}$$

between functors from $W \in \text{Rep}_E(({}^L G)^I)^{\text{op}}$ to $\text{Perv}_{\hat{G}_{\Sigma_i} \infty x_i}(\text{Gr}_I^{(I_1, \dots, I_k)})$.

Proof. As noted in [Braverman and Gaitsgory 2002, §B.6], \mathbb{D} preserves the ULA property with respect to $\text{Gr}_I^{(I_1, \dots, I_k)} \rightarrow (X \setminus \hat{N})^I$. Since $\mathcal{S}_{I, W, E}^{(I_1, \dots, I_k)}$ is ULA, the factorization structure on $\text{Gr}_I^{(I_1, \dots, I_k)}$ reduces the affairs to the case $|I| = 1$, i.e., the Beilinson–Drinfeld affine Grassmannian used in [Mirković and Vilonen 2007; Richarz 2014; Zhu 2015]. Consider its fiber Gr_x over some point $x \in |X \setminus \hat{N}|$. In the notation from Section 4.1, there is a left $G_{\infty x}$ -action on Gr_x .

In the local setting above, denote the usual duality operator $\text{Perv}_{G_\infty}(\text{Gr}_x)$ by \mathbb{D} ; normalization is not an issue here. The main ingredients are

- the Satake functor $\text{Rep}_E(({}^L G)^I) \rightarrow \text{Perv}_{G_\infty}(\text{Gr}_x)$, written as $W \mapsto \mathcal{S}_{W,E}$;
- a canonical isomorphism between functors in W :

$$\mathbb{D}\mathcal{S}_{W,E} \xrightarrow{\sim} \mathcal{S}_{W^{\vee,\theta},E}.$$

Granting these ingredients, for general $|I|$ we obtain canonical isomorphisms $\mathbb{D}\mathcal{S}_{I,W,E}^{(I_1,\dots,I_k)} \xrightarrow{\sim} \mathcal{S}_{I,W^{\vee,\theta},E}^{(I_1,\dots,I_k)}$

Let us explain the two ingredients in the local setting. The functor $W \mapsto \mathcal{S}_{W,E}$ is obtained in [Richarz 2014; Zhu 2015, Theorem A.12], which are based on the case over separably closed fields in [Mirković and Vilonen 2007]. In order to explain the effect of ${}^L\theta$, we shall review the case over the separable closure \mathbb{k} of \mathbb{F}_x first. The canonical isomorphism $\mathbb{D}\mathcal{S}_{W,E} \xrightarrow{\sim} \mathcal{S}_{W^{\vee,\theta},E}$ over \mathbb{k} can be found in [Bezrukavnikov and Finkelberg 2008, Lemma 14], for example, where a stronger equivariant version is established; they work over \mathbb{C} , but the argument is largely formal.

Next, apply Galois descent as explicated in [Richarz 2014, §6; Zhu 2015, Appendix]. Let $\mathcal{C} := \text{Perv}_{G_\infty}(\text{Gr}_x)$ and set \mathcal{C}' to be its avatar over \mathbb{k} . The absolute Galois group Υ of \mathbb{F}_x operates on \mathcal{C}' via \otimes -equivalences, in a manner compatible with the fiber functor (total cohomology), thus Υ acts on the Tannakian group \hat{G} as well. By [Richarz 2014, p.237], \mathcal{C} is equivalent as an abelian category to $(\mathcal{C}' + \text{continuous descent data under } \Upsilon)$. The Satake equivalence over \mathbb{k} and the machinery from [loc. cit.] furnish an equivalence of \otimes -categories

$$\text{Perv}_{G_\infty}(\text{Gr}_x) \rightarrow \text{Rep}_E(\hat{G} \rtimes_{\text{geom}} \Upsilon),$$

where c means continuity, and “geom” means the Tannakian or “geometric” Υ -action on \hat{G} . See [Lafforgue 2018, Remarque 1.19] for the choice of commutativity constraints.

By [Zhu 2015, Proposition A.6; Richarz 2014, Corollary 6.8], the geometric Υ -action on \hat{G} differs from the familiar “algebraic” one by the adjoint action via $\rho_{\hat{B}} \circ \chi_{\text{cycl}} : \Upsilon \rightarrow \mathbb{Z}_\ell^\times \rightarrow \hat{G}^{\text{ad}}(\mathbb{Q}_\ell)$, where χ_{cycl} is the ℓ -adic cyclotomic character and $\rho_{\hat{B}}$ is the half-sum of positive roots in \hat{B} ; in particular, $\hat{G} \rtimes_{\text{geom}} \Upsilon \simeq \hat{G} \rtimes_{\text{alg}} \Upsilon$ (= absolute Galois form of the L -group) continuously. Since $\theta \in \text{Aut}(\hat{G})$ stabilizes $\rho_{\hat{B}}$, the isomorphism matches $\theta \rtimes_{\text{geom}} \text{id}$ with the Chevalley involution $\theta \rtimes_{\text{alg}} \text{id} =: {}^L\theta$.

All in all, we obtain the Satake functor $W \mapsto \mathcal{S}_{W,E}$ as well as the canonical isomorphisms $\mathbb{D}\mathcal{S}_{W,E} \xrightarrow{\sim} \mathcal{S}_{W^{\vee,\theta},E}$. This completes the proof. □

Note that the equivariance can be upgraded to $G_{\sum_i \infty x_i}^{\text{ad}}$ or $G_{\sum_i n_i x_i}^{\text{ad}}$ where $n_i \gg 0$ relative to W , see [Lafforgue 2018, Remarque 1.20].

Proposition 5.1.2. *There is a canonical isomorphism*

$$\mathbb{D}\mathcal{F}_{N,I,W,\Xi,E}^{(I_1,\dots,I_k)} \xrightarrow{\sim} \mathcal{F}_{N,I,W^{\vee,\theta},\Xi,E}^{(I_1,\dots,I_k)}$$

between functors from $W \in \text{Rep}_E(({}^L G)^I)^{\text{op}}$ to $\text{Perv}(\text{Cht}_{N,I,W}^{(I_1,\dots,I_k)} / \Xi)$ that is compatible with coalescence of paws.

Proof. First, by combining [Lafforgue 2018, Proposition 2.8] and the explanations before Corollaire 2.15 of [loc. cit.], the smooth morphisms

$$\mathrm{Cht}_{N,I,W}^{(I_1,\dots,I_k)} / \Xi \xrightarrow{\epsilon_{N,I,W,n}^{(I_1,\dots,I_k),\Xi}} \mathrm{Gr}_{I,W}^{(I_1,\dots,I_k)} / G_{\sum_i n_i x_i}^{\mathrm{ad}}, \quad \mathrm{Gr}_{I,W}^{(I_1,\dots,I_k)} \rightarrow \mathrm{Gr}_{I,W}^{(I_1,\dots,I_k)} / G_{\sum_i n_i x_i}^{\mathrm{ad}}$$

have the same relative dimension; denote it by d .

Proposition 5.1.1 gives a functorial isomorphism between descent data of shifted perverse sheaves from $\mathrm{Gr}_{I,W}^{(I_1,\dots,I_k)}$ to $\mathrm{Gr}_{I,W}^{(I_1,\dots,I_k)} / G_{\sum_i n_i x_i}^{\mathrm{ad}}$, abbreviated as $\mathbb{D}\mathcal{S}_1 \xrightarrow{\sim} \mathcal{S}_2$. Denote the corresponding shifted perverse sheaves on $\mathrm{Gr}_{I,W}^{(I_1,\dots,I_k)} / G_{\sum_i n_i x_i}^{\mathrm{ad}}$ as \mathcal{S}_1^b and \mathcal{S}_2^b . By standard results, see [Laszlo and Olsson 2008b, 9.1.2], the isomorphism above descends to

$$(\mathbb{D}\mathcal{S}_1^b)[2d](d) \xrightarrow{\sim} \mathcal{S}_2^b.$$

Since $\mathcal{F}_{N,I,W,\Xi,E}^{(I_1,\dots,I_k)}$ and $\mathcal{F}_{N,I,W^{\vee,\theta},\Xi,E}^{(I_1,\dots,I_k)}$ are defined in (4-3) as $(\epsilon_{N,I,W,n}^{(I_1,\dots,I_k),\Xi})^* \mathcal{S}_1^b$ and $(\epsilon_{N,I,W^{\vee,\theta},n}^{(I_1,\dots,I_k),\Xi})^* \mathcal{S}_2^b$, respectively, the assertion follows immediately by the same standard result. \square

Take any partition $I = I_1 \sqcup \dots \sqcup I_k$, truncation parameter μ and $W \in \mathrm{Rep}_E(({}^L G)^I)$. As a consequence of Propositions 5.1.1 and 5.1.2, we deduce that $\mathrm{Gr}_{I,W}^{(I_1,\dots,I_k)} = \mathrm{Gr}_{I,W^{\vee,\theta}}^{(I_1,\dots,I_k)}$ and $\mathrm{Cht}_{N,I,W}^{(I_1,\dots,I_k),\leq\mu} = \mathrm{Cht}_{N,I,W^{\vee,\theta}}^{(I_1,\dots,I_k),\leq\mu}$.

Remark 5.1.3. Below is a review of the cup product of $!$ -pushforward. Let S be a regular scheme and let $p : \mathcal{X} \rightarrow S$ be an algebraic stack of finite type over S . Let $\mathcal{L}, \mathcal{L}'$ be in $D_c^-(\mathcal{X}, E)$. Our goal is to define a canonical arrow

$$p_! \mathcal{L} \otimes^L p_! \mathcal{L}' \rightarrow p_!(\mathcal{L} \otimes^L \mathcal{L}').$$

Denote by Δ and $p \times p$ the diagonal morphisms $\mathcal{X} \rightarrow \mathcal{X} \times_S \mathcal{X}$ and $\mathcal{X} \times_S \mathcal{X} \rightarrow S$, respectively. The Künneth formula [Laszlo and Olsson 2008b, 11.0.14 Theorem] yields a canonical isomorphism in $D_c^-(S)$

$$p_! \mathcal{L} \otimes^L p_! \mathcal{L}' \simeq (p \times p)_!(\mathcal{L} \boxtimes \mathcal{L}'),$$

where \boxtimes denotes the external tensor product. Since $\mathcal{L} \otimes^L \mathcal{L}' = \Delta^*(\mathcal{L} \boxtimes \mathcal{L}')$, to obtain the desired arrow, it remains to use the

$$(p \times p)_! \rightarrow (p \times p)_! \Delta_! \Delta^* = p_! \Delta^*$$

arising from $\mathrm{id} \rightarrow \Delta_* \Delta^* = \Delta_! \Delta^*$, as Δ is a closed immersion.

Consider the normalized dualizing complex Ω on $\mathrm{Cht}_{N,I,W}^{(I_1,\dots,I_k),\leq\mu} / \Xi$. The *trace map*

$$\mathrm{Tr} : (p_{N,I}^{(I_1,\dots,I_k),\leq\mu})_! \Omega \xrightarrow{\mathrm{Tr}} E_{(X \setminus \hat{N})^I}$$

in Verdier duality is obtained by adjunction from $\Omega \xrightarrow{\sim} (p_{N,I}^{(I_1,\dots,I_k),\leq\mu})^! E_{(X \setminus \hat{N})^I}$.

On the other hand, Proposition 5.1.2 affords a canonical arrow $\mathcal{F}_{N,I,W^{\vee,\theta},\Xi,E}^{(I_1,\dots,I_k)} \overset{L}{\otimes} \mathcal{F}_{N,I,W,\Xi,E}^{(I_1,\dots,I_k)} \rightarrow \Omega$. Apply the cup-product construction to the stack $\text{Cht}_{N,I,W}^{(I_1,\dots,I_k),\leq\mu} / \Xi$ over $(X \setminus \hat{N})^I$ to obtain canonical arrows in $D_c^b((X \setminus \hat{N})^I, E)$:

$$\begin{array}{ccc}
 \mathcal{H}_{N,I,W^{\vee,\theta}}^{\leq\mu,E} \overset{L}{\otimes} \mathcal{H}_{N,I,W}^{\leq\mu,E} & & E_{(X \setminus \hat{N})^I} \\
 \parallel & & \uparrow \\
 (\mathfrak{p}_{N,I}^{(I_1,\dots,I_k),\leq\mu})_!(\mathcal{F}_{N,I,W^{\vee,\theta},\Xi,E}^{(I_1,\dots,I_k)}) \overset{L}{\otimes} (\mathfrak{p}_{N,I}^{(I_1,\dots,I_k),\leq\mu})_!(\mathcal{F}_{N,I,W,\Xi,E}^{(I_1,\dots,I_k)}) & & \text{Tr} \\
 \downarrow & & \uparrow \\
 (\mathfrak{p}_{N,I}^{(I_1,\dots,I_k),\leq\mu})_!(\mathcal{F}_{N,I,W^{\vee,\theta},\Xi,E}^{(I_1,\dots,I_k)}) \overset{L}{\otimes} \mathcal{F}_{N,I,W,\Xi,E}^{(I_1,\dots,I_k)} & \longrightarrow & (\mathfrak{p}_{N,I}^{(I_1,\dots,I_k),\leq\mu})_!\Omega
 \end{array} \tag{5-1}$$

By homological common sense (see [Kashiwara and Schapira 1990, Example I.24(ii)] for example), taking H^\bullet in (5-1) with respect to the ordinary t -structure on $(X \setminus \hat{N})^I$ yield natural arrows between E -sheaves over $(X \setminus \hat{N})^I$

$$\mathfrak{B}_{N,I,W}^{\Xi,E} : \mathcal{H}_{N,I,W^{\vee,\theta}}^{i,\leq\mu,E} \otimes_E \mathcal{H}_{N,I,W}^{-i,\leq\mu,E} \rightarrow E_{(X \setminus \hat{N})^I}, \quad i \in \mathbb{Z};$$

we will only use the case $i = 0$ in this article.

Following [Lafforgue 2018, Remarque 9.2], we may even pass to \varinjlim_μ and look at the stalk at $\xi \in \{\eta^I, \Delta(\bar{\eta})\}$, thereby obtain from $\mathfrak{B}_{N,I,W}^{\Xi,E}$ the E -bilinear pairings

$$\langle \cdot, \cdot \rangle_\xi : \varinjlim_\mu \mathcal{H}_{N,I,W^{\vee,\theta}}^{0,\leq\mu,E} \Big|_\xi \otimes_E \varinjlim_\mu \mathcal{H}_{N,I,W}^{0,\leq\mu,E} \Big|_\xi \longrightarrow E.$$

Their relation with the arrow \mathfrak{sp} of specialization is given by [loc. cit., (9.6)]

$$\langle \mathfrak{sp}^* h, \mathfrak{sp}^* h' \rangle_{\eta^I} = \langle h, h' \rangle_{\Delta(\bar{\eta})}. \tag{5-2}$$

In the discussions surrounding (4-8), we have seen that $\text{Cht}_{N,\emptyset,1} / \Xi$ is the constant stack $\text{Bun}_{G,N}(\mathbb{F}_q) / \Xi$ over $\text{Spec } \mathbb{F}_q$. The upshot is that, as in [loc. cit., Remarque 9.2], the pairing $\langle \cdot, \cdot \rangle_{\Delta(\bar{\eta})}$ for $I = \emptyset$ reduces to the integration pairing on $C_c(G(\hat{F}) \backslash G(\mathbb{A}) / K_N \Xi; E)$, assuming $\text{mes}(K_N) = 1$. Upon restriction to Hecke-finite part, we get the pairing $\langle \cdot, \cdot \rangle$ for $H_{\emptyset,1}$ in Remark 2.2.3. The same holds for $H_{\{0\},1}$ by coalescence (4-7).

5.2. Frobenius invariance. For every morphism f between reasonable schemes or stacks, we will denote by “can” the canonical isomorphisms exchanging $f^* \leftrightarrow f^!$ and $f_* \leftrightarrow f_!$ under \mathbb{D} . When f is a universal homeomorphism or open immersion, we have $f_* = f_!$ and $f^* = f^!$ or only $f^* = f^!$, respectively.

Merge the conventions from Sections 5.1 and 4.3. Let $J \subset I$ be finite sets, $I = I_1 \sqcup \dots \sqcup I_k$ with $J = I_1$. We are going to explicate the compatibility between $\mathfrak{B}_{N,I,W}^{\Xi,E}$ and $\text{Frob}_J^* \mathcal{H}_{N,I,W}^{\leq\mu,E} \xrightarrow{F_J} \mathcal{H}_{N,I,W}^{\leq\mu',E}$, i.e., (4-15).

Lemma 5.2.1. *In $D_c^b(\text{Cht}_{N,I,W}^{(I_1,\dots,I_k)} / \Xi, E)$, there is a commutative diagram whose arrows are all invertible:*

$$\begin{array}{ccc}
 (\text{Frob}_{I_1,N,W^{\vee,\theta}}^{(I_1,\dots,I_k)})^* \mathcal{F}_{N,I,W^{\vee,\theta},E}^{(I_2,\dots,I_1)} & \longrightarrow & (\text{Frob}_{I_1,N,W}^{(I_1,\dots,I_k)})^* \mathbb{D} \mathcal{F}_{N,I,W,\Xi,E}^{(I_2,\dots,I_1)} \xrightarrow[\sim]{\text{can}} \mathbb{D}(\text{Frob}_{I_1,N,I}^{(I_1,\dots,I_k)})^* \mathcal{F}_{N,I,W,\Xi,E}^{(I_2,\dots,I_1)} \\
 \downarrow F_{I_1,N,W^{\vee,\theta}}^{(I_1,\dots,I_k)} & & \nearrow \mathbb{D} F_{I_1,N,W}^{(I_1,\dots,I_k)} \\
 \mathcal{F}_{N,I,W^{\vee,\theta},\Xi,E}^{(I_1,\dots,I_k)} & \longrightarrow & \mathbb{D} \mathcal{F}_{N,I,W,\Xi,E}^{(I_1,\dots,I_k)}
 \end{array}$$

where $F_{I_1,N,\dots}^{(I_1,\dots,I_k)}$ is from (4-11), and the horizontal arrows except can are induced by Proposition 5.1.2.

Proof. We may assume $W = \boxtimes_{j=1}^k W_j$ is irreducible. Using the definition (4-3), the smoothness of $\epsilon_{N,I,W,\underline{n}}^{(I_1,\dots,I_k),\Xi}$ as well as the ULA properties of \mathcal{F} and \mathcal{S} , the desired commutativity eventually reduces to that of

$$\begin{array}{ccc}
 \text{Frob}^* \mathcal{S}_{I_j,W_j^{\vee,\theta},E}^{(I_j)} & \longrightarrow & \text{Frob}^* \mathbb{D} \mathcal{S}_{I_j,W_j,E}^{(I_j)} \xrightarrow[\sim]{\text{can}} \mathbb{D} \text{Frob}^* \mathcal{S}_{I_j,W_j,E}^{(I_j)} \\
 \downarrow \Phi & & \nearrow \mathbb{D} \Phi \\
 \mathcal{S}_{I_j,W_j^{\vee,\theta},E}^{(I_j)} & \longrightarrow & \mathbb{D} \mathcal{S}_{I_j,W_j,E}^{(I_j)}
 \end{array}$$

in $D_c^b(\text{Gr}_{I_j,W_j}^{(I_j)} / G_{\sum_{i \in I_j} n_i x_i}^{\text{ad}}, E)$, for each $1 \leq j \leq k$. Here Φ stands for the usual Frobenius correspondences, and the horizontal arrows except can are from Proposition 5.1.1. The square commutes by the functoriality of Φ .

for the triangular part, [Laszlo and Olsson 2008a, 4.8.2 Corollary] says that $\text{can} : \text{Frob}^* \mathbb{D} \xrightarrow{\sim} \mathbb{D} \text{Frob}^*$ equals

$$\text{Frob}^* \text{R}\mathcal{H}om(-, \Omega) \xrightarrow{\text{natural}} \text{R}\mathcal{H}om(\text{Frob}^*(-), \text{Frob}^* \Omega) \xrightarrow{f_*} \text{R}\mathcal{H}om(\text{Frob}^*(-), \Omega),$$

where Ω is the dualizing complex and $f : \text{Frob}^* \Omega \xrightarrow{\sim} \Omega$ is the canonical isomorphism furnished by [loc. cit.]. Both f and “can” reflect the fact that universal homeomorphisms conserve duality. In our case, that fact is also realized by transport of structure via Frobenius, i.e., we have $f = \Phi_\Omega$, the Frobenius correspondence for Ω . The desired commutativity thus reduces to that of

$$\begin{array}{ccc}
 \text{Frob}^* \text{R}\mathcal{H}om(\mathcal{S}, \Omega) & \xrightarrow{\text{natural}} & \text{R}\mathcal{H}om(\text{Frob}^* \mathcal{S}, \text{Frob}^* \Omega) \\
 \downarrow \Phi_{\text{R}\mathcal{H}om(\mathcal{S},\Omega)} & & \swarrow (\Phi_S^{-1})^* \circ (\Phi_\Omega)_* \\
 \text{R}\mathcal{H}om(\mathcal{S}, \Omega) & &
 \end{array}$$

for all $\mathcal{S} \in D_c^b(\text{Gr}_{I_j,W_j}^{(I_j)} / G_{\sum_{i \in I_j} n_i x_i}^{\text{ad}}, E)$. This is by now standard. □

Reintroduce the truncation parameters $\mu' \gg \mu$ so that (4-10) holds with respect to both W and $W^{\vee,\theta}$. Let a_1 (universal homeomorphism) and a_2 (open immersion) be as in (4-12). As μ increases, $\text{Cht}_{N,I,W}^{\dots \leq \mu}$ and $\text{Cht}_{N,I,W}^{\dots \leq \mu'}$ form open coverings of $\text{Cht}_{N,I,W}^{\dots}$.

To state the next result, we write $\Omega^{\leq \mu}$ and Ω for the normalized dualizing complex on $\text{Cht}_{N,I,W}^{\leq \mu} / \Xi$ and $a_1^{-1} \text{Cht}_{N,I,W}^{(I_2, \dots, I_1), \leq \mu} / \Xi$, respectively. Recall that dualizing complexes are unique up to unique isomorphisms [Laszlo and Olsson 2008a, 3.4.5]. There are canonical isomorphisms $a_1^* \Omega^{\leq \mu} \xrightarrow{\sim} \Omega \xleftarrow{\sim} a_2^! \Omega^{\leq \mu'}$, since $a_1^* = a_1^!$.

Lemma 5.2.2. *In $D_c^b(a_1^{-1} \text{Cht}_{N,I,W}^{(I_2, \dots, I_1), \leq \mu} / \Xi, E)$, there is a commutative diagram*

$$\begin{array}{ccc}
 a_1^* \underbrace{\mathcal{F}_{N,I,W^{\vee,\theta},\Xi,E}^{(I_2, \dots, I_1)}}_{\text{on } \leq \mu} \otimes^{\mathbb{L}} a_1^* \underbrace{\mathcal{F}_{N,I,W,\Xi,E}^{(I_2, \dots, I_1)}}_{\text{on } \leq \mu} & \longrightarrow & a_1^* \Omega^{\leq \mu} \\
 \downarrow \text{Frob}_{I_1,N,W^{\vee,\theta}}^{(I_1, \dots, I_k)} \otimes^{\mathbb{L}} \text{Frob}_{I_1,N,W}^{(I_1, \dots, I_k)} & & \downarrow \cong \\
 a_2^! \underbrace{\mathcal{F}_{N,I,W^{\vee,\theta},\Xi,E}^{(I_1, \dots, I_k)}}_{\text{on } \leq \mu'} \otimes^{\mathbb{L}} a_2^! \underbrace{\mathcal{F}_{N,I,W,\Xi,E}^{(I_1, \dots, I_k)}}_{\text{on } \leq \mu'} & \longrightarrow & a_2^! \Omega^{\leq \mu'}
 \end{array}$$

Ω
 \cong
 \cong

where the arrows from $\dots \otimes^{\mathbb{L}} \dots$ to Ω are induced from Lemma 5.2.1.

Proof. It suffices to show the commutativity of the outer pentagon, since the triangle is defined to be commutative. Recall the passage from (4-11) to (4-13): $\text{Frob}_{I_1,N,\dots}^{(I_1, \dots, I_k)}$ is obtained by restricting $F_{I_1,N,\dots}^{(I_1, \dots, I_k)}$ to the open substacks cut out by the conditions $\leq \mu$ and $\leq \mu'$. It remains to apply Lemma 5.2.1; note that the effect of arrows $a_1^* \Omega^{\leq \mu} \xrightarrow{\sim} \Omega \xleftarrow{\sim} a_2^! \Omega^{\leq \mu'}$ match the morphism “can” in Lemma 5.2.1. \square

Proposition 5.2.3. *Write $J := I_1$. There is a commutative diagram in $D_c^b((X \setminus \hat{N}), E)$*

$$\begin{array}{ccc}
 \text{Frob}_J^* \mathcal{H}_{N,I,W^{\vee,\theta}}^{\leq \mu, E} \otimes^{\mathbb{L}} \text{Frob}_J^* \mathcal{H}_{N,I,W}^{\leq \mu, E} & \xrightarrow{\text{Frob}_J^*(5-1)} & \text{Frob}_J^* E_{(X \setminus \hat{N})^I} \\
 \downarrow \text{F}_J \otimes^{\mathbb{L}} \text{F}_J & & \downarrow \text{F}_J \\
 \mathcal{H}_{N,I,W^{\vee,\theta}}^{\leq \mu', E} \otimes^{\mathbb{L}} \mathcal{H}_{N,I,W}^{\leq \mu', E} & \xrightarrow{(5-1)} & E_{(X \setminus \hat{N})^I}
 \end{array}$$

where

- $\text{F}_J \otimes^{\mathbb{L}} \text{F}_J$ is induced from the F_J in (4-15),
- the F_J on the right is the evident partial Frobenius morphism for $E_{(X \setminus \hat{N})^I}$.

Proof. Retain the notation for Lemma 5.2.2 and let $\mathfrak{p} := \mathfrak{p}_{N,I}^{(I_1, \dots, I_k)}$. Upon recalling the formalism of $\text{K}\tilde{\mathbb{A}}_4$ -neth formula, cup products (Remark 5.1.3) and the trace maps Tr (see (5-1)), Lemma 5.2.2 produce a diagram in $D_c^b((X \setminus \hat{N})^I, E)$:

$$\begin{array}{ccc}
 \text{Frob}_J^* \mathfrak{p}_! \underbrace{\mathcal{F}_{N,I,W^{\vee,\theta},\Xi,E}^{(I_2,\dots,I_1)}}_{\text{on } \leq \mu} \overset{\mathbb{L}}{\otimes} \text{Frob}_J^* \mathfrak{p}_! \underbrace{\mathcal{F}_{N,I,W,\Xi,E}^{(I_2,\dots,I_1)}}_{\text{on } \leq \mu} & \longrightarrow & \text{Frob}_J^* \mathfrak{p}_! \Omega^{\leq \mu} \xrightarrow{\text{Frob}_J^* \text{Tr}} \text{Frob}_J^* E_{(X \setminus \hat{N})^I} \\
 \downarrow \text{BC} \overset{\mathbb{L}}{\otimes} \text{BC} \simeq & & \downarrow \simeq \text{BC} \\
 \mathfrak{p}_! a_1^* \mathcal{F}_{N,I,W^{\vee,\theta},\Xi,E}^{(I_2,\dots,I_1)} \overset{\mathbb{L}}{\otimes} \mathfrak{p}_! a_1^* \mathcal{F}_{N,I,W,\Xi,E}^{(I_2,\dots,I_1)} & \longrightarrow & \mathfrak{p}_! a_1^* \Omega^{\leq \mu} \simeq \mathfrak{p}_! \Omega \xrightarrow{\text{Tr}} E_{(X \setminus \hat{N})^I} \\
 \downarrow \mathfrak{p}_! \text{Frob}_{I_1,\dots,I_k}^{(I_1,\dots,I_k)} \overset{\mathbb{L}}{\otimes} \mathfrak{p}_! \text{Frob}_{I_1,\dots,I_k}^{(I_1,\dots,I_k)} & & \parallel \\
 \mathfrak{p}_! a_2^! \mathcal{F}_{N,I,W^{\vee,\theta},\Xi,E}^{(I_1,\dots,I_k)} \overset{\mathbb{L}}{\otimes} \mathfrak{p}_! a_2^! \mathcal{F}_{N,I,W,\Xi,E}^{(I_1,\dots,I_k)} & \longrightarrow & \mathfrak{p}_! a_2^! \Omega^{\leq \mu'} \simeq \mathfrak{p}_! \Omega \\
 \downarrow & & \downarrow \\
 \mathfrak{p}_! \underbrace{\mathcal{F}_{N,I,W^{\vee,\theta},\Xi,E}^{(I_1,\dots,I_k)}}_{\text{on } \leq \mu'} \overset{\mathbb{L}}{\otimes} \mathfrak{p}_! \underbrace{\mathcal{F}_{N,I,W,\Xi,E}^{(I_1,\dots,I_k)}}_{\text{on } \leq \mu'} & \longrightarrow & \mathfrak{p}_! \Omega^{\leq \mu'} \xrightarrow{\text{Tr}} E_{(X \setminus \hat{N})^I}
 \end{array}$$

where BC and $\mathfrak{p}_! a_2^! \rightarrow \mathfrak{p}_!$ are the arrows in (4-14) and explained after (4-15), respectively. The diagram commutes, indeed:

- The first two rows form a commutative diagram by the naturality of BC, which is ultimately based on the topological invariance of the étale topos together with the fact that universal homeomorphisms respect duality [Laszlo and Olsson 2008b, 9.1.5 Proposition and 12.2].
- The commutativity of the middle square comes from Lemma 5.2.2, by applying $\mathfrak{p}_!$.
- The remaining pieces commute by the naturality of $\mathfrak{p}_! a_2^! \rightarrow \mathfrak{p}_!$ and of Tr.

The composite of the last row is (5-1), and that of the first row is its Frob_J^* -image (now for the $\leq \mu$ part). The composite of the leftmost column yields $F_J \overset{\mathbb{L}}{\otimes} F_J : \text{Frob}_J^* \mathcal{H}_{N,I,W^{\vee,\theta}}^{\leq \mu,E} \overset{\mathbb{L}}{\otimes} \text{Frob}_J^* \mathcal{H}_{N,I,W}^{\leq \mu,E} \rightarrow \mathcal{H}_{N,I,W^{\vee,\theta}}^{\leq \mu',E} \overset{\mathbb{L}}{\otimes} \mathcal{H}_{N,I,W}^{\leq \mu',E}$ by the very definition of F_J . This completes the proof. \square

The case $k = 1$, i.e., when F_J is the total Frobenius morphism, is relatively straightforward; see the proof of Lemma 5.2.1.

Recall from Section 4.3 that F_J furnishes an E -linear endomorphism of $\varinjlim_{\mu} \mathcal{H}_{N,I,W}^{0,\leq \mu,E} |_{\eta^I}$, still denoted as F_J .

Corollary 5.2.4. *The pairing $\langle \cdot, \cdot \rangle_{\eta^I}$ in (5-2) is invariant under F_J for all $J \subset I$.*

Proof. After taking H^0 and \varinjlim_{μ} , Proposition 5.2.3 implies that

$$\langle h_1, h_2 \rangle_{\eta^I} = \langle F_J(h_1), F_J(h_2) \rangle_{\eta^I}$$

for all h_1, h_2 in $\varinjlim_{\mu} \mathcal{H}_{N,I,W^{\vee,\theta}}^{0,\leq \mu,E} |_{\eta^I}$ and $\varinjlim_{\mu} \mathcal{H}_{N,I,W}^{0,\leq \mu,E} |_{\eta^I}$, respectively. \square

The cautious reader might worry about a missing power of p in Corollary 5.2.4 due to Tate twists. It does not occur here by our normalizations of \mathcal{S} , \mathcal{F} and \mathbb{D} .

5.3. Computation of the transpose. We adopt the notation of Section 4.4. The integration pairing $\langle \cdot, \cdot \rangle$ of Remark 2.2.3 is nondegenerate symmetric on the finite-dimensional E -vector space $H_{\{0\},1} \simeq H_{\emptyset,1}$. The transpose S^* of any $S \in \text{End}_E(H_{\{0\},1})$, characterized by $\langle h', Sh \rangle = \langle S^*h', h \rangle$ for all $h, h' \in H_{\{0\},1}$. Identify $\langle \cdot, \cdot \rangle$ with the pairing $\langle \cdot, \cdot \rangle_{\Delta(\bar{\eta})}$ in (5-2).

Lemma 5.3.1. *For all data I, W, ξ, x and $\vec{\gamma} = (\gamma_i)_i \in \pi_1(\eta, \bar{\eta})^I$ for excursion operators, we have*

$$S_{I,W,\xi,x,\vec{\gamma}}^* = S_{I,W^{\vee,\theta},x,\xi,\vec{\gamma}^{-1}}.$$

In particular, the E -algebra \mathcal{B}_E is closed under transpose $S \mapsto S^$.*

Note that the roles of x, ξ are switched when one passes from W to $W^{\vee,\theta}$. The transpose-invariance of \mathcal{B}_E has already been sketched in [Lafforgue 2018, Remarque 12.15].

Proof. Recall from Section 4.3 that by choosing $\bar{\eta}^I$ and \mathfrak{sp} , there is a homomorphism $\text{FWeil}(\eta^I, \bar{\eta}^I) \rightarrow \mathbb{W}_{\mathbb{F}}^I$ inducing an isomorphism between profinite completions. As $S_{I,W,\xi,x,\vec{\gamma}}$ and $S_{I,W^{\vee,\theta},x,\xi,\vec{\gamma}^{-1}}$ are both continuous in $\vec{\gamma}$, it suffices to consider that case when $\vec{\gamma}$ comes from $\text{FWeil}(\eta^I, \bar{\eta}^I)$.

By [Lafforgue 2018, Remarque 5.4, (9.8)], \mathcal{C}_ξ^b and \mathcal{C}_ξ^\sharp are already transposes of each other on the sheaf level with respect to $\mathfrak{B}_{N,I,W}^{\Xi,E}$; in particular they commute with \mathfrak{sp}^* . Ditto for \mathcal{C}_x^\sharp and \mathcal{C}_x^b . Therefore, for all $h, h' \in H_{\{0\},1}$, we infer by using (5-2) that

$$\begin{aligned} \langle h', S_{I,W,x,\xi,\vec{\gamma}}(h) \rangle_{\Delta(\bar{\eta})} &= \langle h', \mathcal{C}_\xi^b(\mathfrak{sp}^*)^{-1}(\vec{\gamma} \cdot \mathfrak{sp}^* \mathcal{C}_x^\sharp h) \rangle_{\Delta(\bar{\eta})} \\ &= \langle \mathfrak{sp}^*(h'), \mathfrak{sp}^*(\mathcal{C}_\xi^b(\mathfrak{sp}^*)^{-1}(\vec{\gamma} \cdot \mathfrak{sp}^* \mathcal{C}_x^\sharp h)) \rangle_{\bar{\eta}^I} \\ &= \langle \mathfrak{sp}^* \mathcal{C}_\xi^\sharp(h'), \vec{\gamma} \cdot \mathfrak{sp}^* \mathcal{C}_x^\sharp(h) \rangle_{\bar{\eta}}, \\ \langle S_{I,W^{\vee,\theta},\xi,x,\vec{\gamma}^{-1}}(h'), h \rangle_{\Delta(\bar{\eta})} &= \langle \mathcal{C}_x^b(\mathfrak{sp}^*)^{-1}(\vec{\gamma}^{-1} \cdot \mathfrak{sp}^* \mathcal{C}_\xi^\sharp h'), h \rangle_{\Delta(\bar{\eta})} \\ &= \langle \mathfrak{sp}^*(\mathcal{C}_x^b(\mathfrak{sp}^*)^{-1}(\vec{\gamma}^{-1} \cdot \mathfrak{sp}^* \mathcal{C}_\xi^\sharp h')), \mathfrak{sp}^*(h) \rangle_{\bar{\eta}^I} \\ &= \langle \vec{\gamma}^{-1} \cdot \mathfrak{sp}^* \mathcal{C}_\xi^\sharp(h'), \mathfrak{sp}^* \mathcal{C}_x^\sharp(h) \rangle_{\bar{\eta}^I}. \end{aligned}$$

It remains to show that $\langle \cdot, \cdot \rangle_{\bar{\eta}^I}$ is $\text{FWeil}(\eta^I, \bar{\eta}^I)$ -invariant. Recall that the $\text{FWeil}(\eta^I, \bar{\eta}^I)$ -action unites those from $\pi_1(\eta^I, \bar{\eta}^I)$ and partial Frobenius morphisms F_J . The $\pi_1(\eta^I, \bar{\eta}^I)$ -action leaves $\langle \cdot, \cdot \rangle_{\bar{\eta}^I}$ invariant since the latter comes from the sheaf-level pairing $\mathfrak{B}_{N,I,W}^{\Xi,E}$ over $(X \setminus \hat{N})^I$. The F_J -invariance of $\langle \cdot, \cdot \rangle$ for all $J \subset I$ is assured by Corollary 5.2.4. □

We are now able to describe the transpose of excursion operators.

Definition 5.3.2. For every $f \in \mathcal{O}(\hat{G} \backslash \langle \langle \mathbb{L}G \rangle \rangle / \hat{G})$, set $f^\dagger(\vec{g}) := f(\mathbb{L}\theta(\vec{g}^{-1}))$ where $\vec{g} \in (\mathbb{L}G)^I$ and $\mathbb{L}\theta$ stands for the Chevalley involution of $(\mathbb{L}G)^I$. Then $f \mapsto f^\dagger$ defines an involution of $\mathcal{O}(\hat{G} \backslash \langle \langle \mathbb{L}G \rangle \rangle / \hat{G})$.

Lemma 5.3.3. *For all $I, f \in \mathcal{O}(\hat{G} \backslash \langle \langle \mathbb{L}G \rangle \rangle / \hat{G})$ and $\vec{\gamma} = (\gamma_i)_i \in \pi_1(\eta, \bar{\eta})^I$, we have $S_{I,f,\vec{\gamma}}^* = S_{I,f^\dagger,\vec{\gamma}^{-1}}$.*

Proof. It suffices to consider the case $f(\vec{g}) = \langle \xi, \vec{g} \cdot x \rangle_{W^\vee \otimes W}$, where $\xi \in W^\vee, x \in W$ are as in Lemma 5.3.1, and $\langle \cdot, \cdot \rangle_{W^\vee \otimes W}$ is the evident duality pairing. As before, denote by $W^\theta, W^{\vee,\theta}$ be the $\mathbb{L}\theta$ -twists of the

representations W, W^\vee etc., and the preceding convention on pairing still applies. Note that $(W^{\vee,\theta})^\vee \simeq W^\theta$ canonically in $\text{Rep}_E(({}^L G)^I)$.

For every $\vec{g} \in ({}^L G)^I$, we have

$$\begin{aligned} f^\dagger(\vec{g}) &= \langle \xi, \underbrace{{}^L\theta(g)^{-1} \cdot x}_{\text{original action}} \rangle_{W^\vee \otimes W} \\ &= \langle \xi, \underbrace{\vec{g}^{-1} \cdot x}_{{}^L\theta\text{-twisted}} \rangle_{W^{\theta,\vee} \otimes W^\theta} \\ &= \underbrace{\langle \vec{g} \cdot \xi \rangle}_{{}^L\theta\text{-twisted}}, x_{W^{\theta,\vee} \otimes W^\theta} \\ &= \langle x, \vec{g} \cdot \xi \rangle_{(W^{\vee,\theta})^\vee \otimes W^{\vee,\theta}}. \end{aligned}$$

In view of Lemma 5.3.1, we deduce that $S_{I,f,\vec{\gamma}}^* = S_{I,W,\xi,x,\vec{\gamma}}^*$ equals $S_{I,W^{\vee,\theta},x,\xi,\vec{\gamma}^{-1}} = S_{I,f^\dagger,\vec{\gamma}^{-1}}$, as asserted. \square

Consider any homomorphism $\nu : \mathcal{B} := \mathcal{B}_E \otimes_E \overline{\mathbb{Q}_\ell} \rightarrow \overline{\mathbb{Q}_\ell}$ of $\overline{\mathbb{Q}_\ell}$ -algebras. As \mathcal{B} is commutative and closed under transpose, $\nu^* : S \mapsto \nu(S^*)$ is also a homomorphism of $\overline{\mathbb{Q}_\ell}$ -algebras.

Proposition 5.3.4. *If $\sigma \in \Phi(G)$ is attached to $\nu : \mathcal{B} \rightarrow \overline{\mathbb{Q}_\ell}$, then ${}^L\theta \circ \sigma$ is attached to ν^* .*

Proof. Fix $n \in \mathbb{Z}_{\geq 0}$ and let $I := \{0, \dots, n\}$. Given the characterization (4-19) of the L -parameters attached to ν, ν^* , it boils down to the observation that for all $\vec{\gamma} = (\gamma_0, \dots, \gamma_n) \in \pi_1(\eta, \bar{\eta})^I$ and $f \in \mathcal{O}(\hat{G} \parallel {}^L G^I \parallel \hat{G})$,

$$\nu^*(S_{I,f,\vec{\gamma}}) = \nu(S_{I,f,\vec{\gamma}}^*) = \nu(S_{I,f^\dagger,\vec{\gamma}^{-1}}) = f^\dagger(\sigma(\gamma_0)^{-1}, \dots, \sigma(\gamma_n)^{-1}) = f({}^L\theta\sigma(\gamma_0), \dots, {}^L\theta\sigma(\gamma_n)),$$

in which the second equality stems from Lemma 5.3.3. \square

Write $\mathfrak{H}_\sigma := \mathfrak{H}_\nu$ if $\sigma \in \Phi(G)$ is attached to ν , and write $\langle \cdot, \cdot \rangle_{\sigma,\sigma'} := \langle \cdot, \cdot \rangle|_{\mathfrak{H}_\sigma \otimes \mathfrak{H}_{\sigma'}}$, for all $\sigma, \sigma' \in \Phi(G)$.

Proof of Theorem 3.3.2. Enlarge E so that all homomorphisms $\nu : \mathcal{B} \rightarrow \overline{\mathbb{Q}_\ell}$ are defined over E . Fix a ν such that $\mathfrak{H}_\nu \neq \{0\}$. Since \mathcal{B}_E is closed under transpose, the subspace $\mathfrak{H}_\nu^\perp \subset H_{\{0\},1}$ defined relative to $\langle \cdot, \cdot \rangle$ is \mathcal{B}_E -stable as well. Since $\langle \cdot, \cdot \rangle$ is nondegenerate, $\mathfrak{H}_\nu^\perp \neq H_{\{0\},1}$. Use the \mathcal{B}_E -invariance to decompose \mathcal{B}_E -modules as follows

$$\mathfrak{H}_\nu^\perp = \bigoplus_\mu \mathfrak{H}_\nu^\perp \cap \mathfrak{H}_\mu, \quad \frac{H_{\{0\},1}}{\mathfrak{H}_\nu^\perp} = \bigoplus_\mu \frac{\mathfrak{H}_\mu}{\mathfrak{H}_\nu^\perp \cap \mathfrak{H}_\mu} \neq \{0\}.$$

We contend that $\mathfrak{H}_\mu \not\subset \mathfrak{H}_\nu^\perp$ only if $\mu = \nu^*$, or equivalently $\mu^* = \nu$.

Indeed, $\langle \cdot, \cdot \rangle$ induces a nondegenerate pairing

$$\langle \cdot, \cdot \rangle_\nu : \mathfrak{H}_\nu \otimes_E \bigoplus_\mu \frac{\mathfrak{H}_\mu}{\mathfrak{H}_\nu \cap \mathfrak{H}_\mu^\perp} \rightarrow E.$$

Let $d := \dim H_{\{0\},1}$. For every $S \in \mathcal{B}_E$, write $S_v := S|_{\mathfrak{H}_v}$. Then $(S_v - \nu(S))^d = 0$. Taking transpose with respect to $\langle \cdot, \cdot \rangle_v$ yields $(S_v^* - \nu(S))^d = 0$ in $\text{End}_E(\mathfrak{H}_\mu / (\mathfrak{H}_\mu \cap \mathfrak{H}_v^\perp))$, for each μ .

On the other hand, the transpose $S^* \in \mathcal{B}_E$ with respect to $\langle \cdot, \cdot \rangle$ satisfies $(S^* - \mu(S^*))^d = 0$ on \mathfrak{H}_μ , and $S^*|_{\mathfrak{H}_\mu}$ induces $S^{*,\mu} \in \text{End}_E(\mathfrak{H}_\mu / (\mathfrak{H}_\mu \cap \mathfrak{H}_v^\perp))$ satisfying $(S^{*,\mu} - \mu(S^*))^d = 0$. Clearly $S^{*,\mu} = S_v^*$. All in all, we deduce that $\mu^*(S) := \mu(S^*) = \nu(S)$ whenever $\mathfrak{H}_\mu \neq \mathfrak{H}_v^\perp \cap \mathfrak{H}_\mu$.

It follows from the claim that if $\langle \cdot, \cdot \rangle_{\sigma, \sigma'}$ is not identically zero, then the corresponding $\nu, \nu' : \mathcal{B} \rightarrow \overline{\mathbb{Q}_\ell}$ satisfy $\nu^* = \nu'$. Now Proposition 5.3.4 implies ${}^L\theta \circ \sigma = \sigma'$. \square

Acknowledgements

The author is deeply grateful to Alain Genestier, Vincent Lafforgue, Dipendra Prasad and Changjian Su for their helpful comments, answers and corrections. Thanks also goes to the referees for pertinent suggestions.

References

- [Adams and Vogan 2016] J. Adams and D. A. Vogan, Jr., “Contragredient representations and characterizing the local Langlands correspondence”, *Amer. J. Math.* **138**:3 (2016), 657–682. MR Zbl
- [Arthur 1988] J. Arthur, “The invariant trace formula, II: Global theory”, *J. Amer. Math. Soc.* **1**:3 (1988), 501–554. MR Zbl
- [Artin and Tate 1968] E. Artin and J. Tate, *Class field theory*, W. A. Benjamin, New York, 1968. MR Zbl
- [Bezrukavnikov and Finkelberg 2008] R. Bezrukavnikov and M. Finkelberg, “Equivariant Satake category and Kostant–Whittaker reduction”, *Mosc. Math. J.* **8**:1 (2008), 39–72. MR Zbl
- [Borel 1979] A. Borel, “Automorphic L -functions”, pp. 27–61 in *Automorphic forms, representations and L -functions, II*, edited by A. Borel and W. Casselman, Proc. Sympos. Pure Math. **33**, Amer. Math. Soc., Providence, RI, 1979. MR Zbl
- [Bourbaki 2012] N. Bourbaki, *Algèbre, Chapitre VIII*, 2nd revised ed., Springer, 2012. MR Zbl
- [Braverman and Gaitsgory 2002] A. Braverman and D. Gaitsgory, “Geometric Eisenstein series”, *Invent. Math.* **150**:2 (2002), 287–384. MR Zbl
- [Brylinski and Deligne 2001] J.-L. Brylinski and P. Deligne, “Central extensions of reductive groups by \mathbb{K}_2 ”, *Publ. Math. Inst. Hautes Études Sci.* **94** (2001), 5–85. MR Zbl
- [Bushnell and Henniart 2006] C. J. Bushnell and G. Henniart, *The local Langlands conjecture for $GL(2)$* , Grundlehren der Math. Wissenschaften **335**, Springer, 2006. MR Zbl
- [Fargues 2016] L. Fargues, “Geometrization of the local Langlands correspondence: an overview”, preprint, 2016. arXiv
- [Gaitsgory and Lysenko 2018] D. Gaitsgory and S. Lysenko, “Parameters and duality for the metaplectic geometric Langlands theory”, *Selecta Math. (N.S.)* **24**:1 (2018), 227–301. MR Zbl
- [Genestier and Lafforgue 2017] A. Genestier and V. Lafforgue, “Chtoucas restreints pour les groupes réductifs et paramétrisation de Langlands locale”, preprint, 2017. arXiv
- [Gross and Prasad 1992] B. H. Gross and D. Prasad, “On the decomposition of a representation of SO_n when restricted to SO_{n-1} ”, *Canad. J. Math.* **44**:5 (1992), 974–1002. MR Zbl
- [Haines and Rostami 2010] T. J. Haines and S. Rostami, “The Satake isomorphism for special maximal parahoric Hecke algebras”, *Represent. Theory* **14** (2010), 264–284. MR Zbl
- [Heinloth 2010] J. Heinloth, “Uniformization of \mathcal{G} -bundles”, *Math. Ann.* **347**:3 (2010), 499–528. MR Zbl
- [Henniart 1983] G. Henniart, *La conjecture de Langlands locale pour $GL(3)$* , Mém. Soc. Math. France **11-12**, Soc. Math. France, Paris, 1983. MR Zbl
- [Kaletha 2013] T. Kaletha, “Genericity and contragredience in the local Langlands correspondence”, *Algebra Number Theory* **7**:10 (2013), 2447–2474. MR Zbl

- [Kaletha 2016a] T. Kaletha, “Regular supercuspidal representations”, preprint, 2016. arXiv
- [Kaletha 2016b] T. Kaletha, “Rigid inner forms of real and p -adic groups”, *Ann. of Math. (2)* **184**:2 (2016), 559–632. MR Zbl
- [Kashiwara and Schapira 1990] M. Kashiwara and P. Schapira, *Sheaves on manifolds*, Grundlehren der Math. Wissenschaften **292**, Springer, 1990. MR Zbl
- [Lafforgue 2018] V. Lafforgue, “Chtoucas pour les groupes réductifs et paramétrisation de Langlands globale”, *J. Amer. Math. Soc.* **31**:3 (2018), 719–891. MR Zbl
- [Lapid and Mao 2015] E. Lapid and Z. Mao, “A conjecture on Whittaker–Fourier coefficients of cusp forms”, *J. Number Theory* **146** (2015), 448–505. MR Zbl
- [Laszlo and Olsson 2008a] Y. Laszlo and M. Olsson, “The six operations for sheaves on Artin stacks, I: Finite coefficients”, *Publ. Math. Inst. Hautes Études Sci.* **107** (2008), 109–168. MR Zbl
- [Laszlo and Olsson 2008b] Y. Laszlo and M. Olsson, “The six operations for sheaves on Artin stacks, II: Adic coefficients”, *Publ. Math. Inst. Hautes Études Sci.* **107** (2008), 169–210. MR Zbl
- [Laszlo and Olsson 2009] Y. Laszlo and M. Olsson, “Perverse t -structure on Artin stacks”, *Math. Z.* **261**:4 (2009), 737–748. MR Zbl
- [Mirković and Vilonen 2007] I. Mirković and K. Vilonen, “Geometric Langlands duality and representations of algebraic groups over commutative rings”, *Ann. of Math. (2)* **166**:1 (2007), 95–143. MR Zbl
- [Mœglin and Waldspurger 1994] C. Mœglin and J.-L. Waldspurger, *Décomposition spectrale et séries d’Eisenstein*, Progress in Math. **113**, Birkhäuser, Basel, 1994. MR Zbl
- [Mœglin et al. 1987] C. Mœglin, M.-F. Vignéras, and J.-L. Waldspurger, *Correspondances de Howe sur un corps p -adique*, Lecture Notes in Math. **1291**, Springer, 1987. MR Zbl
- [Prasad 2018] D. Prasad, “Generalizing the MVW involution, and the contragredient”, *Trans. Amer. Math. Soc.* (online publication November 2018).
- [Renard 2010] D. Renard, *Représentations des groupes réductifs p -adiques*, Cours Spécialisés **17**, Soc. Math. France, Paris, 2010. MR Zbl
- [Richarz 2014] T. Richarz, “A new approach to the geometric Satake equivalence”, *Doc. Math.* **19** (2014), 209–246. MR Zbl
- [Serre 2005] J.-P. Serre, “Complète réductibilité”, exposé 932, pp. 195–217 in *Séminaire Bourbaki*, 2003/2004, Astérisque **299**, 2005. MR Zbl
- [SGA 4₁ 1972] M. Artin, A. Grothendieck, and J. L. Verdier, *Théorie des topos et cohomologie étale des schémas, Tome 1: Théorie des topos, Exposés I–IV* (Séminaire de Géométrie Algébrique du Bois Marie 1963–1964), Lecture Notes in Math. **269**, Springer, 1972. MR Zbl
- [SGA 4₂ 1972] M. Artin, A. Grothendieck, and J. L. Verdier, *Théorie des topos et cohomologie étale des schémas, Tome 2: Exposés V–VIII* (Séminaire de Géométrie Algébrique du Bois Marie 1963–1964), Lecture Notes in Math. **270**, Springer, 1972. MR Zbl
- [SGA 4₃ 1973] M. Artin, A. Grothendieck, and J. L. Verdier, *Théorie des topos et cohomologie étale des schémas, Tome 3: Exposés IX–XIX* (Séminaire de Géométrie Algébrique du Bois Marie 1963–1964), Lecture Notes in Math. **305**, Springer, 1973. MR Zbl
- [Shahidi 1990] F. Shahidi, “A proof of Langlands’ conjecture on Plancherel measures; complementary series for p -adic groups”, *Ann. of Math. (2)* **132**:2 (1990), 273–330. MR Zbl
- [Stacks 2005–] P. Belmans, A. J. de Jong, et al., “The Stacks project”, electronic reference, 2005–, Available at <http://stacks.math.columbia.edu>.
- [Varshavsky 2004] Y. Varshavsky, “Moduli spaces of principal F -bundles”, *Selecta Math. (N.S.)* **10**:1 (2004), 131–166. MR Zbl
- [Varshavsky 2007] Y. Varshavsky, “Lefschetz–Verdier trace formula and a generalization of a theorem of Fujiwara”, *Geom. Funct. Anal.* **17**:1 (2007), 271–319. MR Zbl
- [Vignéras 1996] M.-F. Vignéras, *Représentations l -modulaires d’un groupe réductif p -adique avec $l \neq p$* , Progress in Math. **137**, Birkhäuser, Boston, 1996. MR Zbl

[Vignéras 2001] M.-F. Vignéras, “Correspondance de Langlands semi-simple pour $GL(n, F)$ modulo $l \neq p$ ”, *Invent. Math.* **144**:1 (2001), 177–223. MR Zbl

[Zhu 2015] X. Zhu, “The geometric Satake correspondence for ramified groups”, *Ann. Sci. Éc. Norm. Supér. (4)* **48**:2 (2015), 409–451. MR Zbl

Communicated by Marie-France Vignéras

Received 2018-10-16 Revised 2019-01-09 Accepted 2019-03-10

wwli@bicmr.pku.edu.cn

*Beijing International Center for Mathematical Research, Peking University,
Beijing, China*

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 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 L^AT_EX but submissions in other varieties of T_EX, 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 BibT_EX 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 13 No. 5 2019

Surjectivity of Galois representations in rational families of abelian varieties AARON LANDESMAN, ASHVIN A. SWAMINATHAN, JAMES TAO and YUJIE XU	995
A unified and improved Chebotarev density theorem JESSE THORNER and ASIF ZAMAN	1039
On the Brauer–Siegel ratio for abelian varieties over function fields DOUGLAS ULMER	1069
A five-term exact sequence for Kac cohomology CÉSAR GALINDO and YIBY MORALES	1121
On the paramodularity of typical abelian surfaces ARMAND BRUMER, ARIEL PACETTI, CRIS POOR, GONZALO TORNARÍA, JOHN VOIGHT and DAVID S. YUEN	1145
Contragredient representations over local fields of positive characteristic WEN-WEI LI	1197



1937-0652(2019)13:5;1-8