THE SHAFAREVICH CONJECTURE FOR HYPERSURFACES IN ABELIAN VARIETIES

BRIAN LAWRENCE AND WILL SAWIN

Abstract. Faltings proved that there are finitely many abelian varieties of genus g over a number field K, with good reduction outside a finite set of primes S. Fixing one of these abelian varieties A, we prove that there are finitely many smooth hypersurfaces in A, with good reduction outside S, representing a given ample class in the Neron-Severi group of A, up to translation, as long as the dimension of A is at least 4. Our approach builds on the approach of [42] which studies p-adic variations of Hodge structure to turn finiteness results for p-adic Galois representations into geometric finiteness statements. A key new ingredient is an approach to proving big monodromy for the variations of Hodge structure arising from the middle cohomology of these hypersurfaces using the Tannakian theory of sheaf convolution on abelian varieties.

1. Introduction

Fix a number field K with ring of integers $\mathcal{O}_K$, and let S be a finite set of primes of $\mathcal{O}_K$. Fix an abelian variety A, defined over K, with good reduction at all primes outside S. We say a hypersurface $H \subseteq A$ has good reduction at $p \notin S$ if the closure of H in the unique smooth projective model of A over $\mathcal{O}_K[1/S]$ is smooth at p. Our main result is the following.

Theorem 1.1 (Theorem 11.4)**.** *Suppose* $\dim A \geq 4$*. Fix an ample class* ϕ *in the Neron-Severi group of* A*. There are only finitely many smooth hypersurfaces* $H \subseteq A$ *representing* ϕ*, with good reduction outside* S*, up to translation.*

If we fix a Picard class ψ, rather than a Neron-Severi class, this theorem becomes a finiteness result for a Diophantine equation, in principle concrete. The theorem is equivalent to the statement that there are only finitely many H representing a given Picard class ψ, because only finitely many translates of a given H will represent ψ. The hypersurfaces in a given Picard class form a projective space, and the singular ones form an irreducible divisor as soon as ψ is very ample, by a classical result (e.g. [60, Theorem 1.18]) which uses the fact that A is not ruled by projective spaces. Thus, the singular hypersurfaces are the vanishing locus of some discriminant polynomial $\Delta(x_1, \ldots, x_N)$ in the homogenous coordinates of that projective space. Theorem 11.4 is equivalent to the statement that, for u any S-unit in $\mathcal{O}_K$, there are only finitely many solutions of the equation $\Delta(x_1, ..., x_N) = u$ with all $x_i \in \mathcal{O}_K[1/S]$.

For $\dim A = 3$ there are additional combinatorial difficulties, leading to a more complicated result. Let $a(i)$ be the sequence

$$1, 5, 20, 76, 285, 1065, \ldots$$

satisfying

$$a(1) = 1, a(2) = 5, a(i + 2) = 4a(i + 1) + 1 - a(i)$$

Let $d(i)$ be the sequence

$$d(i) = \binom{a(i) + a(i+1)}{a(i)}$$

so that

$$d(1) = 6, d(2) = 53130, d(3) = 216182590635135019896, d(4) = 2.5 \ldots \times 10^{79}, \cdots$$

Theorem 1.2 (Theorem 11.5)**.** *Suppose* $\dim A = 3$*. Fix an ample class* ϕ *in the Neron-Severi group of* A*. Assume that the intersection number* $\phi \cdot \phi \cdot \phi$ *is not divisible by* $d(i)$ *for any* $i \geq 2$*. There are only finitely many smooth hypersurfaces* $H \subseteq A$ *representing* ϕ*, with good reduction outside* S*, up to translation.*

Since $a(i)$ increases exponentially, $d(i)$ increases superexponentially. Because of this rapid rate of increase, and because $d(2)$ is already large, a very small proportion of possible intersection numbers are ruled out by this.

If $\dim A = 2$, then hypersurfaces in A are curves, and the analogue of Theorems 11.4 and 11.5 follows from the Shafarevich conjecture for curves; see Theorem 11.6.

Our result is analogous to the Shafarevich conjecture for curves, now a theorem of Faltings [20], but (except in dimension 2) it doesn't seem to follow from Faltings's work; we'll say more about the relationship below. Instead, the proof uses a study of variation of Galois representations based on the work of one of the authors (B.L.) and Venkatesh [42], and the sheaf convolution formalism of Krämer and Weissauer [40].

The original Shafarevich conjecture (proved by Faltings) says that there are only finitely many isomorphism classes of curve of fixed genus g, defined over K, and having good reduction outside S. Similar results are now known for various families of varieties: abelian varieties ([20]), K3 surfaces ([2] and [58]), del Pezzo surfaces ([54]), flag varieties ([30]), complete intersections of Hodge level at most 1 ([29]), surfaces fibered smoothly over a curve ([27]), Fano threefolds ([31]), and some general type surfaces ([28]).

Javanpeykar and Loughran have suggested that the Shafarevich conjecture should hold in broad generality (see for example [29, Conj. 1.4]); the present result is further evidence in this direction. They show that the Lang–Vojta conjecture implies the Shafarevich conjecture for certain families of complete intersections [29, Thm. 1.5]; their argument uses the hyperbolicity of a certain moduli space. One expects the implication to hold for still more general families of varieties: for any family that gives rise to a locally injective period map, Griffiths transversality and the geometry of period domains imply that the base must be hyperbolic, and the argument of Javanpeykar–Loughran applies. Indeed, in our proof we use a big monodromy statement (Corollary 5.9) that may be seen as a strong form of injectivity of the period map. In fact, we show that this big monodromy statement implies the quasi-finiteness of a certain period map in Proposition 5.10 below.

To understand the relationship between our work and previous work, it is helpful to compare and contrast with two previous finiteness theorems, both due to Faltings, involving abelian varieties. The first is the Shafarevich conjecture for abelian varieties [20], i.e. the result that there are only finitely many isomorphism classes of abelian varieties of dimension n over K with good reduction outside S. The second is the result, in [21], that the K-rational points in any closed subvariety Z of an abelian variety defined over K lie in finitely many translates of abelian subvarieties contained in Z.

Both of these have been very useful for proving further arithmetic finiteness theorems. The result of [20] was applied, using the natural maps from the moduli space of curves, certain moduli spaces of K3 surfaces, and moduli spaces of complete intersections of Hodge level 1 to the moduli space of abelian varieties, to prove most of the Shafarevich-type statements discussed above. Similarly, finiteness results for points on curves over number fields of fixed degree are proven using [21] and the maps from symmetric powers of a curve to the Jacobian abelian variety.

There does not seem to be any logical relation between our work and these two finiteness theorems. There is no reason to believe that there exists a nonconstant map from the moduli space of smooth hypersurfaces $H \subseteq A$ to any moduli space of abelian varieties for general d and A (except when $\dim A = 2$). Thus, our result does not seem to follow from [20]. There does exist a map from the moduli space of hypersurfaces to an abelian variety - in fact A - by sending each hypersurface to its Picard class, but this is surjective so [21] is not helpful. Instead, this map can be used to reduce the finiteness problem to the moduli space of smooth hypersurfaces in a given Picard class, which is an open subset of projective space. Because an open subset of projective space does not have a nonconstant map to any abelian variety, [21] cannot be applied at this point.

Indeed, our main result seems to be synergistic with prior finiteness results in abelian varieties. Faltings proved that there only finitely many abelian varieties A of a given dimension with good reduction outside S. One can check that each of these abelian varieties has only finitely many ample Neron-Severi classes of a given intersection number, up to automorphism. We have proven that each of these ample classes contains only finitely many smooth hypersurfaces with good reduction outside S, up to translation. Finally Faltings proved that each of these hypersurfaces contains only finitely many K-rational points, outside of finitely many translates of abelian subvarieties.

The present work uses general machinery introduced by B.L. and Venkatesh in [42] to study of period maps and Galois representations applicable to cohomology in arbitrary degree. Significant work is required to apply this machinery in our setting. We develop a version of the sheaf convolution Tannakian category, and use it to prove a uniform big monodromy statement. We extend the methods of [42] to non-connected reductive groups. Finally, we need to do some difficult combinatorial calculations related to Hodge numbers. All of this will be explained in more detail after we recall some general ideas from [42].

The paper [42] introduces a method to bound integral points on a variety X, assuming one can find a family over X whose cohomology has big monodromy. Suppose $Y \to X$ is a smooth proper family of varieties, extending to a smooth proper S-integral model $\mathcal{Y} \to \mathcal{X}$ over $\mathcal{O}_{K,S}$. Then for every integral point $x \in \mathcal{X}(\mathcal{O}_{K,S})$, the étale cohomology of the fiber $\mathcal{Y}_x$ gives rise to a global Galois representation

$$\rho_x \colon \mathrm{Gal}_K \to \mathrm{Aut}(H^i_{et}(\mathcal{Y}_x, \mathbb{Q}_p)).$$

A lemma of Faltings shows that there are only finitely many possibilities for ρ_x, up to semisimplification. In various settings, it is possible to show that the representation ρ_x varies p-adically in x, and deduce that the S-integral points $\mathcal{X}(\mathcal{O}_{K,S})$ are not Zariski dense in X.

A key input to the methods of [42] is control on the image of the monodromy representation

$$\pi_1(X, x_0) \to \mathrm{Aut}(H^i_{sing}(\mathcal{Y}_{x_0})).$$

(The idea that big monodromy statements might have interesting Diophantine consequences goes back at least to Deligne's proof of the Weil conjectures [14].) In order to show that a certain period map has big image, we need to know that the Zariski closure of the image of monodromy is "big" in a certain sense. In particular, in the case studied in this paper the image of monodromy is sufficiently big if its Zariski closure includes one of the classical groups SL_N, Sp_N, SO_N. Because this is the sufficient condition we use in our argument, one can think of "bigness" in terms of classical groups, but the precise condition in [42] is substantially more flexible, which might prove useful elsewhere. For example, Theorem 8.21 requires that the monodromy group be "strongly c-balanced" in the sense of Definition 7.5, as well as two numerical conditions that are more easily satisfied when the monodromy group is larger.

A major technical difficulty in this present work is the need to prove a big monodromy statement that applies uniformly to all positive-dimensional subvarieties of the universal family of hypersurfaces in A. For the monodromy groups of the universal family itself, there are multiple geometric and topological arguments that could demonstrate that the monodromy contains a classical group. This would be sufficient to prove Zariski nondensity of the integral points. Then, one hopes to pass from Zariski nondensity of $\mathcal{X}(\mathcal{O}_{K,S})$ to finiteness by an inductive argument. (This idea was suggested in [42, Sec. 10.2].) Specifically, if $\mathcal{X}(\mathcal{O}_{K,S})$ is infinite, we may take Z the Zariski closure of $\mathcal{X}(\mathcal{O}_{K,S})$ in X, and repeat the argument with Z in place of X. However, this requires us to prove large monodromy, not just over X, but for *every* positive-dimensional subvariety $Z \subseteq X$. In our case, we may assume only that Z carries some smooth family of hypersurfaces in A which is not a translate of the constant family.

For most families of varieties, such as hypersurfaces in projective space, this problem would seem totally insurmountable. Either we know almost nothing about the monodromy groups of arbitrary subvarieties, or, as in the universal family of abelian varieties, we can construct explicit subfamilies with too-small monodromy, e.g. low-dimensional Shimura subvarieties. However, working with a family of subvarieties of a fixed abelian variety A provides us with a way out. The inverse image of H under the multiplication-by-ℓ^n map $A \to A$ has good reduction everywhere H does, except possibly at ℓ, and we can run the argument with its étale cohomology. The ℓ^n-torsion points $A[\ell^n]$ act on this inverse image, and thus on its cohomology; this action splits the cohomology into a sum of eigenspaces, each with its own monodromy representation. It suffices for our purposes to show that one of these representations has big monodromy.

This additional freedom allows a new type of argument, based on the Tannakian theory of sheaf convolution developed by Krämer and Weissauer [40]. They defined a group, the "Tannakian monodromy group", associated to a subvariety in an abelian variety (and in fact to much more general objects). Its definition is subtler than the definition of the usual monodromy group, but it is a better tool to work with because it depends only on a single hypersurface in the family, whose geometry can be controlled, rather than an arbitrary family of hypersurfaces, whose geometry is far murkier. We prove a group-theoretic relationship between the usual monodromy group of a typical $A[\ell^n]$-eigenspace in the cohomology of a family of hypersurfaces and the Tannakian monodromy group of a typical member of the family of hypersurfaces. One can think of this as analogous to the relationship between the monodromy groups of the generic horizontal and vertical fibers of a family of varieties over (an open subset of) a product $X_1 \times X_2$. Using purely geometric arguments, we show that the Tannakian monodromy group contains a classical group, and then using this relationship, we show that the usual monodromy group does as well.

We believe the problem of proving big monodromy for the restriction of a local system to an arbitrary subvariety to be very difficult without this Tannakian method, but owing to its arithmetic applications, it would be very interesting to look for new examples where this can be established by a different method. Roughly speaking, one looks for a variety X of dimension at least two and a smooth family $Y \to X$ such that for all pointed curves $(Z, z) \subseteq X$, the Zariski closure of the image of monodromy

$$\pi_1(Z, z) \to \operatorname{Aut} H^1(Y_z)$$

is large. Here "large" means, ideally, that the Zariski closure contains a classical group, but should at least mean that the dimension of the Zariski closure is greater than some constant. Ideally one wants an example where nothing is known already about the integral points on X - for instance X should not map to the moduli space of abelian varieties or to a hyperbolic curve. At the very least, one should avoid trivial examples, like X a product of curves and Y a fiber product of families over those curves, where the monodromy over any curve is necessarily at least as large as the monodromy over one of the original families. If the variety X has multiple natural smooth families $Y_i \to X$, it suffices to prove that for each Z, z there exists one Y_i with big monodromy.

The methods of this paper can likely be applied to many different classes of subvarieties of abelian varieties, beyond hypersurfaces. To make this generalization, the additional inputs needed are a result giving some control on the Tannakian monodromy group associated to the subvariety and the verification of a certain inequality involving the Hodge numbers and this group (see Lemma 10.1).

1.1. **Outline of the proof.** The argument of [42] derives bounds on $X(\mathcal{O}_{K,S})$ from a family $f \colon Y \to X$, through a study of various cohomology objects $R^i f_*(-)$ on X. The étale local system $R^i f_*(\mathbb{Q}_p)$ gives rise to the global Galois representations to which Faltings's lemma is applied; a filtered F-isocrystal coming from crystalline cohomology is used to study the p-adic variation of these Galois representations; and a variation of Hodge structure allows one to relate a p-adic period map to topological monodromy. The method allows one to conclude that $X(\mathcal{O}_{K,S})$ is not Zariski dense in X.

In the present setting, we will need to apply these results to $R^i f_*(\mathsf{L})$, where L is a nontrivial local system on Y. To this end, we introduce the notion of *Hodge–Deligne system* on X (Definition 6.2), a variant of Deligne's definition of a motive as a system of realizations [15]. A Hodge–Deligne system on X consists of a singular local system, an étale local system, a vector bundle with filtration and integrable connection, and a filtered F-isocrystal, equipped with the usual comparison isomorphisms.

For technical reasons (to be explained below), we will assume $K = \mathbb{Q}$. There is no harm in this: if X is a variety over K, then its Weil restriction $\mathrm{Res}_{\mathbb{Q}}^{K} X$ is a variety over $\mathbb{Q}$; we can also arrange that integral points of X correspond to integral points of $\mathrm{Res}_{\mathbb{Q}}^{K} X$.

We will work with Hodge–Deligne systems that are pure of some fixed weight, with integral Frobenius eigenvalues; by Faltings's lemma (Lemma 6.43), there are only finitely many possibilities for the semisimplification of the Galois representation

$$\rho_x \colon \mathrm{Gal}_K \to \mathrm{Aut}\, H^k(Y_x, \mathbb{Q}_p),$$

as x varies over $X(\mathcal{O}_{K,S})$.

As in [42], we want to show that the fibers of the map

$$x \mapsto \rho_x^{ss}$$

are not Zariski dense. To do this, we consider the map that takes a p-adic point $x \in X(\mathbb{Q}_p)$ to a *local* Galois representation. By p-adic Hodge theory, the local Galois representation

$$\rho_{x,p} \colon \mathrm{Gal}_{\mathbb{Q}_p} \to \mathrm{Aut}\, H^k(Y_x, \mathbb{Q}_p)$$

determines the filtered ϕ-module

$$(V_x, \phi_x, F_x) = H^k_{cris}(Y_x).$$

The global semisimplification $\rho_x \mapsto \rho_x^{ss}$ causes substantial technical difficulties; so much so that, in addition to our main argument in Section 8, we include an alternative, simpler version under the additional assumption that every relevant representation ρ_x is semisimple, in Section 9. For this sketch, to illustrate ideas, let us make the same assumption; that is, let us imagine that every global representation ρ_x is semisimple. Then there are literally only finitely many possibilities for the isomorphism class ρ_x, so (restricting to the local Galois representation and applying the crystalline Dieudonné functor) there are only finitely many possibilities (up to isomorphism) for the filtered ϕ-module (V_x, ϕ_x, F_x), as x ranges over all integral points. In this simplified setting, we need only show that

$$\{x \in X(\mathbb{Q}_p) \mid (V_x, \phi_x, F_x) \cong (V_0, \phi_0, F_0)\}$$

is contained in a positive-codimension algebraic subset of X.

We recall from [42, §3] some facts about the variation of (V_x, ϕ_x, F_x) with x. For x in a fixed mod-p residue disk Ω, the pair (V_x, ϕ_x) is constant: these spaces are canonically identified with the crystalline cohomology of the mod-p reduction of Y_x. The filtration F_x varies with x. The assignment $x \mapsto F_x$ defines a p-adic period map

$$\Phi_p \colon \Omega \to \mathcal{H}$$

to a certain flag variety. Isomorphism classes of triple (V_x, ϕ_x, F_x) correspond to orbits of the Frobenius centralizer $Z(\phi)$ on $\mathcal{H}$.

Thus we want to control $\Phi_p^{-1}(Z)$, where Z is an orbit of the Frobenius centralizer. We'll have the result we want if we can prove precise versions of the following two conditions.

(a) The Frobenius centralizer is small.
(b) The image of Φ_p is not contained in a small algebraic set.

In fact, since we don't know that the global Galois representations are semisimple, we need a stronger form of (a).

(a') (See Lemma 8.20.) Fix a ϕ-module (V, ϕ) and a semisimple global Galois representation ρ^{ss}. Consider all global Galois representations ρ whose semisimplification is ρ^{ss}, and such that $D_{cris}(\rho|_{\mathbb{Q}_p}) \cong (V, \phi, F)$, for some filtration F on V.

The F that arise in this way all lie in a subvariety $Z \subseteq \mathcal{H}$ of low dimension.

Once we have items (a') and (b), we know that $X(\mathbb{Z}[1/S])$ is contained in $\Phi_p^{-1}(Z)$. A p-adic version of the Bakker–Tsimerman transcendence theorem (Theorem 7.3) will imply that $X(\mathbb{Z}[1/S])$ is not Zariski dense.

Condition (a') comes from two ingredients. First, the semilinearity of Frobenius gives an upper bound on its centralizer (Lemma 6.32). Second, the possible subrepresentations of a global Galois representation are constrained by purity (Lemma 8.2), which restricts the structure of *local* Galois representations coming from *global* ρ having a given semisimplification. It is this latter result that requires us to work over $\mathbb{Q}$ (or at least a number field that has no CM subfield). As mentioned above, we can always pass to this situation by restriction of scalars.

Condition (b) is a question about the monodromy of the variation of Hodge structure given by V. As mentioned above, we only need a very weak lower bound on the Zariski closure of the monodromy group. We will show that this group is strongly c-balanced (Definition 7.5; see Corollary 5.9 and Lemma 7.6 for precise statements). The technical difficulty in Corollary 5.9 is that it applies uniformly to any family of hypersurfaces in an abelian variety; this will allow us to apply Noetherian induction to smaller and smaller subvarieties of X.

It is now crucial that, in our case, Y is a subvariety of $A \times X$, with the map f the restriction of the projection map $A \times X$ to X. The abelian variety A has many rank-one local systems L, each of which we can pull back to Y, push forward to X, and apply this machinery to. These local systems are associated to characters of the fundamental group $\pi_1(A)$.

For the Noetherian induction to work, it suffices to have, for each subvariety of X, a local system L such that $R^{n-1}f_*(\mathsf{L})$ has big monodromy in our sense. (There are some additional technical conditions that we suppress here to focus on the main difficulty.) In fact we will show big monodromy for almost all rank one local systems L, in a precise sense. To do this, it is necessary to have a framework in which the vector spaces $R^{n-1}f_*(\mathsf{L})_x$ for different local systems L can be studied all at once. This is accomplished by the Tannakian theory of sheaf convolution [40].

The fundamental objects of the Tannakian theory of sheaf convolution are perverse sheaves. The fundamental perverse sheaf for us is the constant sheaf on Y_x, pushed forward to A, and placed in degree $1-n$. The vector space $R^{n-1}f_*(\mathsf{L})_x$ can be recovered from this by applying the functor $K \mapsto H^0(A, K \otimes \mathsf{L})$. The theory of [40] views (a slight modification of) the category of perverse sheaves on A as the category of representations of a certain group. where the functors $K \mapsto H^0(A, K \otimes \mathsf{L})$ are all isomorphic to the functor taking a representation to the underlying vector space. The image of this group on the representation associated to a perverse sheaf K is the Tannakian monodromy group.

We show that, for Z a subvariety of X, if the Tannakian monodromy group of the constant sheaf on Y_x contains a classical group for some $x \in Z$, and if the family Y over Z is not equal to a translate of the constant family, then for almost all L, the monodromy group of $R^{n-1}f_*(\mathsf{L})_x$ contains a classical group. Because translations of the constant family produce only finitely many hypersurfaces up to translation, we can stop the induction there. Hence to make the argument work, it suffices to show for all x that the Tannakian monodromy group of the constant sheaf on Y_x contains a classical group. In other words, we must show this for all smooth hypersurfaces H.

To show that the Tannakian monodromy groups contains a classical group, we use recent results of Krämer [37, 38], to reduce to a small number of cases—essentially, the simple algebraic groups acting by their minuscule representations—and then some intricate but elementary combinatorics involving Hodge numbers to eliminate the non-classical cases.

1.2. **Sheaf convolution and uniform big monodromy.** Given an abelian variety A over an algebraically closed field, Krämer and Weissauer [39] construct a Tannakian category as a quotient of the category of perverse sheaves on A. A perverse sheaf N on A is said to be *negligible* if its Euler characteristic is zero; the negligible sheaves form a thick subcategory, and the sheaf convolution category is defined as the quotient of the category of all perverse sheaves by the negligible sheaves. The convolution of two perverse sheaves has negligible perverse homology in nonzero degrees; in other words, it is "perverse up to negligible sheaves," and convolution defines a tensor structure on this quotient category.

One original motivation for this construction was the Schottky problem [40]. Given a principally polarized abelian variety A (say of dimension g) with theta divisor Θ, one wants to know whether A is isomorphic to a Jacobian, say $\operatorname{Jac} C$. In this case, Θ would be the $(g-1)$-st convolution power of C. Informally, the role of the Tannakian formalism here is to determine whether Θ is "a $(g-1)$-st convolution power of something."

An alternate motivation for the sheaf convolution theory comes from work of Katz. This time, one works with an abelian variety A over a finite field $\mathbb{F}_q$. A perverse sheaf K on A has a trace function f_K on $A(\mathbb{F}_q)$. Associated to a character χ of $A(\mathbb{F}_q)$ is the character sum $\sum_{x \in A(\mathbb{F}_q)} f_K(x)\chi(x)$. Katz showed (in unpublished work analogous to [33]) that the distribution of $\sum_{x \in A(\mathbb{F}_q)} f_K(x)\chi(x)$, viewed as a random variable for uniformly random χ, converges to a distribution determined by the Tannakian group, in the limit as q goes to ∞. More precisely, the distribution is like the trace of a random element in the maximal compact subgroup of the Tannakian group. To gain some intuition for this, note that given representations V_1, V_2, we have $\operatorname{tr}(g, V_1 \otimes V_2) = \operatorname{tr}(g, V_1)\operatorname{tr}(g, V_2)$; that is, taking the tensor product of representations has the effect of multiplying the traces. For the character sums $\sum_{x \in A(\mathbb{F}_q)} f_K(x)\chi(x)$, convolution has the same effect:

$$\sum_{x \in A(\mathbb{F}_q)} (f_{K_1} * f_{K_2})(x)\chi(x) = \left(\sum_{x \in A(\mathbb{F}_q)} f_{K_1}(x)\chi(x) \right) \left(\sum_{x \in A(\mathbb{F}_q)} f_{K_2}(x)\chi(x) \right).$$

In other words, convolution of these functions f_K has a similar effect on this sum as tensor product of the representations V has on the trace. It stands to reason that a framework where perverse sheaves correspond to representations, and convolution of sheaves correspond to tensor product of representations, would have relevance to the distribution of the trace. In particular, this should be plausible if one is familiar with Deligne's equidistribution theorem [14, Theorem 3.5.3], whose proof is similar to the argument Katz uses to establish the relationship between the distribution and the Tannakian group [33, Corollary 7.4].

For non-algebraically closed fields, such as finite fields, we can construct a Tannakian category in almost the same way as Krämer and Weissauer did—again defining negligible sheaves as those with zero Euler characteristic. The key facts (for example, that the convolution of two perverse sheaves has negligible perverse homology in nonzero degrees) hold over the base field once checked over its algebraic closure.

To relate these two categories, it is convenient to restrict attention to geometrically semisimple perverse sheaves on A_k, and to perverse sheaves on $A_{\overline{k}}$ which are summands of the pullback from A_k to $A_{\overline{k}}$. Having done this, we obtain an exact sequence of pro-algebraic groups

$$1 \to G_{\overline{k}} \to G_k \to \mathrm{Gal}_k \to 1$$

where G_k is the Tannakian group of a suitable category of perverse sheaves on A_k, $G_{\overline{k}}$ is the Tannakian group of a suitable category of perverse sheaves on $A_{\overline{k}}$, and Gal_k is the Tannakian group of the category of ℓ-adic $\mathrm{Gal}(\overline{k}/k)$-representations. We think of this as a close analogue of the usual exact sequence

$$1 \to \pi_1^{geom}(X) \to \pi_1^{arith}(X) \to \mathrm{Gal}(\overline{k}/k) \to 1$$

for a variety X over a field k.

Just like this usual exact sequence, it often has splittings. In our case, splittings arise from local systems L on A defined over k, as their cohomology is then a Galois representation, on which Gal_k acts, and we can check that this action factors through the Tannakian group G_k, giving the splitting.

Fix now a subvariety Z of the moduli space of smooth hypersurfaces in an abelian variety A. Let k be the field of functions on the generic point of Z. Let H be the universal hypersurface in A, defined over k. Let K be the constant sheaf on H, pushed forward to A, placed in degree $1-n$; this is our perverse sheaf of interest. Associated to k is a representation of G_k. The action of $G_{\overline{k}}$ on this representation is a purely geometric object. By geometric methods, we will show that the image of $G_{\overline{k}}$ acting on this representation contains SL_N, SO_N, or Sp_N as a normal subgroup. So the image of G_k on the representation associated to k contains the same classical group as a normal subgroup. Because the action of Gal_k in this setting matches the action of the fundamental group, it will suffice for our induction step to show that the action of Gal_k also contains (as a normal subgroup) the same classical group.

To do this, we construct a battery of tests, each consisting of representations of the normalizer of the classical group, such that any subgroup of the normalizer contains the classical group if and only if it has no invariants on any of these representations. Associated to each of these representations is a perverse sheaf on A_k. We prove a lemma showing that the action of Gal_k on the cohomology of a perverse sheaf, defined using a generic local system L, has invariants if and only if the perverse sheaf has a very special form. Then we check that the relevant perverse sheaves do not have this special form unless the family of hypersurfaces over Z is constant, up to translation by a section of A.

Next we describe how we check that the image of the $G_{\overline{k}}$-action on the representation associated to a smooth hypersurface in A contains a classical group acting by the standard representation as a normal subgroup. This proceeds in two steps. The first step uses very general geometric arguments using deep machinery, and shows that the commutator of the identity component of this image group is a simple algebraic group acting by a minuscule representation. (Recall that a *minuscule representation* is one where the eigenvalues of the maximal torus action are conjugate under the Weyl group.) The second step eliminates all such pairs of a group and a representation except the standard representations of the classical groups. The first step uses sophisticated machinery (beyond the perverse sheaves and sheaf convolution needed to formulate the statement) but is relatively general and conceptual, while the second uses no additional machinery (except a bit of Hodge theory) but involves an intricate combinatorial argument.

For the first step, we apply results of Krämer that study the characteristic cycle of a perverse sheaf. This is a fundamental invariant of any perverse sheaf on a smooth variety, defined as an algebraic cycle on the cotangent bundle of that variety. For abelian varieties, the cotangent bundle is a trivial vector bundle, which gives it some interesting additional structure—for instance, it naturally maps to a vector space the same dimension as the variety. By examining how the characteristic cycle of a perverse sheaf changes when it is convolved with another perverse sheaf, Krämer was able to relate the Tannakian group to the characteristic cycle. In particular, he gave criteria for the commutator subgroup of the identity component to be a simple group, and for the representation of it to be minuscule. The fact that our hypersurface is smooth makes its characteristic cycle relatively simple—it is simply the conormal bundle to the hypersurface. This makes Krämer's minisculeness criterion straightforward to check, but to check simplicity we must make a modification of Krämer's argument. The reason for this is that Krämer, motivated by the theta divisor and the Schottky problem, assumed that a hypersurface in A was invariant under the inversion map, while we do not wish to assume this.

For the second step, the exceptional groups and spin groups are not too hard to eliminate, as they only occur for representations of very specific dimensions. The Tannakian dimension in our setting is the topological Euler characteristic of the hypersurface, which we have an explicit formula for. Comparing these, we can see that the problematic cases only occur for curves in an abelian surface, which we have assumed away by taking the dimension of A at least 4. The only remaining case, except for the good classical cases, is the case of a special linear group acting by a wedge power representation. For this representation, the Euler characteristic formula is not sufficient, but we are eventually able to rule this case out using a more sophisticated numerical invariant, the Hodge numbers. If the Tannakian group acts on the representation associated to H by the k'th wedge power of an m-dimensional representation, we might expect that the Hodge structure on the cohomology of H, or the cohomology of H twisted by a rank one local system, is the k'th wedge power of an m-dimensional Hodge structure. This would place some restrictions on the Hodge numbers. We don't prove this, but instead prove a p-adic Hodge-theoretic analogue, using the Gal_k action discussed earlier. On the other hand, we can calculate the Hodge numbers of the cohomology of H twisted by a rank one local system using the Hirzebruch-Riemann-Roch formula. Working this out gives a complicated set of combinatorial relations between the Hodge numbers of the original m-dimensional Hodge structure. By a lengthy combinatorial argument, we find all solutions of these relations, noting in particular that they occur only for abelian varieties of dimension less than four. This completes the proof.

1.3. **Outline of the paper.** In Section 2 we compute Euler characteristics of hypersurfaces in an abelian variety, as well as of vector bundles on those hypersurfaces, and use them to calculate the Hodge-Tate weights of the arithmetic local systems we will study in the rest of the paper. After this preamble, the argument proceeds in three parts.

First, we use the sheaf convolution formalism to prove a big monodromy result for families of hypersurfaces. In Section 3 we introduce the sheaf convolution category, a Tannakian category of perverse sheaves on an abelian variety. In Section 4 we investigate the Tannakian group of a hypersurface; we show in many cases that this group must be as big as possible. In Section 5 we relate the Tannakian group to the geometric monodromy group, which gives the big monodromy statement we need.

Sections 6–9 explain how to deduce non-density of integral points, following the strategy in [42]. Section 6 contains some technical preliminaries. We introduce the notion of Hodge–Deligne system, which is closely related to Deligne's "system of realizations" of a motive, although we include only the realizations that are relevant for our argument. We discuss "H^0-algebras", roughly, algebra objects in the category of Artin motives with rational coefficients, which we need to express the semilinearity of Frobenius. We also recall some facts from the theory of not-necessarily-connected reductive groups. Section 7 relates the big monodromy statement from Section 5 to the p-adic period map, via the theorem of Bakker and Tsimerman ([3]). In Section 8, we deduce the non-density of integral points. The argument here involves combinatorics on reductive groups, analogous to [42, §11]. We conclude with Theorem 8.21, which is analogous to Lemma 4.2, Prop. 5.3, and Thm. 10.1 in [42]. The main technical difficulty comes from semisimplification of the global representation; in Section 9 we present a much simpler argument, assuming all the global representations that arise are semisimple.

Finally, we wrap up the proof. In Section 10, we verify the two numerical conditions in the hypotheses of Theorem 8.21. The proof of our main theorem occupies Section 11.

Appendices A and B contains some purely combinatorial calculations involving Eulerian numbers. We prove Prop. 4.11, which is used to show that the representation of the Tannakian group associated to a smooth hypersurface is not the wedge power of a smaller-dimensional representation—the last remaining case where the Tannakian group could be too small.

1.4. **Acknowledgements.** We would like to thank Johan de Jong, Matthew Emerton, Sergey Gorchinskiy, Ariyan Javanpeykar, Shizhang Li, Benjamin Martin, and Akshay Venkatesh for interesting discussions related to this project.

This work was conducted while Will Sawin served as a Clay Research Fellow. Brian Lawrence would like to acknowledge support from the National Science Foundation. We met to work on this project at the Oberwolfach Research Institute for Mathematics, Columbia University, and the University of Chicago; we would like to thank these institutions for their hospitality.

Contents

2. Euler characteristics and Hodge numbers

In this section, we fix an abelian variety A and a hypersurface H, and compute various Euler characteristics of H - its arithmetic Euler characteristic, its topological Euler characteristic, and the Euler characteristics of the wedge powers of its cotangent sheaf. Using these, we will calculate the dimension, and Hodge numbers, of the cohomology of H with coefficients in a rank one lisse sheaf. These Hodge numbers will be used at multiple points in our later arguments.

Lemma 2.1. *Let L be a line bundle on A.*

$$\chi(A, \mathcal{L}) = c_1(\mathcal{L})^n/n!$$

Proof. By Hirzebruch-Riemann-Roch, the Euler characteristic of the coherent sheaf $\mathcal{L}$ is the integral of its Chern character against the Todd class. By definition, the Chern character of L is $e^{c_1[L]} = \sum_{k=0}^{n} c_1(L)^k/k!$. Because the tangent bundle of A is trivial, its Todd class is 1. Integrating is equivalent to taking the degree n term, which is $c_1(\mathcal{L})^n/n!$. □

Lemma 2.2. *Let H be a smooth hypersurface in an abelian variety A of dimension n. The arithmetic Euler characteristic $\chi(H, \mathcal{O}_H)$ of H is $(-1)^{n-1}[H]^n/n!$.*

Proof. Using the exact sequence $0 \to \mathcal{O}_A(-H) \to \mathcal{O}_A \to \mathcal{O}_H \to 0$, we observe that

$$\chi(H, \mathcal{O}_H) = \chi(A, \mathcal{O}_A) - \chi(A, \mathcal{O}_A(-H)) = 0 - (-[H])^n/n! = (-1)^{n-1}[H]^n/n!$$

by Lemma 2.1. □

Lemma 2.3. *Let H be a smooth hypersurface in an abelian variety A of dimension n. The topological Euler characteristic of H is $(-1)^{n-1}[H]^n$.*

Proof. The topological Euler characteristic of H is the top Chern class of the tangent bundle of H. Using the exact sequence $0 \to \mathcal{O}(-[H]) \to \Omega^1_A \to \Omega^1_H \to 0$, and the fact that all Chern classes of Ω^1_A vanish, we see that the top Chern class of Ω^1_H is $[H]^n$, □

Motivated by Lemma 2.2, we define the *degree* d of H to be $\frac{[H]^n}{n!}$, which is always a positive integer for H an ample hypersurface.

The Hodge numbers of H can be computed in terms of Eulerian numbers. For a general reference on Eulerian numbers in combinatorics, see [47, Chap. 1].

Lemma 2.4. *Let H be a smooth hypersurface in an abelian variety A of dimension n. We have*

$$\chi(H, \Omega^i_H) = (-1)^{n-1-i} d A(n, i) \tag{1}$$

where $A(n, i)$ is the Eulerian number.

Proof. Let $\mathcal{L} = \mathcal{O}_A(H)$. We have the exact sequence

$$0 \to \mathcal{L}^{-1} \to \mathcal{O}_A \to \mathcal{O}_H \to 0.$$

There is also an exact sequence for the cotangent bundle:

$$0 \to N^*_{H/A} \to \Omega^1_A|_H \to \Omega^1_H \to 0,$$

which because A is an abelian variety, so Ω^1_A is trivial, reduces (noncanonically) to

$$0 \to \mathcal{L}^{-1}|_H \to \mathcal{O}_H^n \to \Omega^1_H \to 0.$$

Passing to wedge powers gives

$$0 \to \mathcal{L}^{-1} \otimes \Omega_H^{i-1} \to \Omega_A^i|_H \to \Omega_H^i \to 0,$$

or, noncanonically,

$$0 \to \mathcal{L}^{-1} \otimes \Omega_H^{i-1} \to \mathcal{O}_H^{\binom{n}{i}} \to \Omega_H^i \to 0. \tag{2}$$

In what follows we use freely

$$\chi(\mathcal{L}^r) = r^n \chi(\mathcal{L}).$$

For every line bundle $\mathcal{E}$ on A, we have

$$\chi(\mathcal{E}|_H) = \chi(\mathcal{E}) - \chi(\mathcal{E} \otimes \mathcal{L}^{-1}).$$

In particular, we have

$$\chi(\mathcal{L}^r|_H) = (r^n - (r-1)^n)\chi(\mathcal{L})$$

and (combining this with (2))

$$\chi(\mathcal{L}^r \otimes \Omega_H^i) = \binom{n}{i}(r^n - (r-1)^n)\chi(\mathcal{L}) - \chi(\mathcal{L}^{r-1} \otimes \Omega_H^{i-1}).$$

We deduce that

$$\chi(\mathcal{L}^r \otimes \Omega_H^i) = \sum_{j=0}^{i} \binom{n}{i-j} (-1)^j ((r-j)^g - (r-j-1)^g)\chi(\mathcal{L}).$$

In particular, setting $r = 0$, we get

$$\chi(H, \Omega_H^i) = (-1)^n \sum_{j=0}^{i} \binom{n}{i-j} (-1)^j (j^n - (j+1)^n)\chi(\mathcal{L})$$

$$= (-1)^n \sum_{j=0}^{i+1} (-1)^j \left(\binom{n}{i-j} + \binom{n}{i-j+1} \right) j^n \chi(\mathcal{L}) = (-1)^n \sum_{j=0}^{i+1} (-1)^j \binom{n+1}{i+1-j} j^n \chi(\mathcal{L})$$

We now use the combinatorial identity ([47, Cor. 1.3])

$$(-1)^n \sum_{j=0}^{i+1} (-1)^j \binom{n+1}{i+1-j} j^n = (-1)^{n-1-i} A(n,i)$$

and Lemma 2.1

$$\chi(\mathcal{L}) = d$$

to derive (1). □

We will work with lisse rank one sheaves on an abelian variety. It will be convenient to parametrize them by representations of the fundamental group.

Definition 2.5. Let A be an abelian variety over a field k. Fix a character $\chi : \pi_1^{et}(A_{\overline{k}}) \to \overline{\mathbb{Q}}_\ell^\times$. We define the *character sheaf* $\mathcal{L}_\chi$ to be the unique rank one sheaf on $A_{\overline{k}}$ whose monodromy representation is χ.

We also have a canonical way to descend these sheaves to A_k:

Definition 2.6. Let A be an abelian variety over a field k. Let χ be a character of $\pi_1^{et}(A_{\overline{k}})$ that is $\mathrm{Gal}(\overline{k}|k)$-invariant. We define the *character sheaf* $\mathcal{L}_\chi$ to be the unique lisse rank one sheaf on A_k whose monodromy restricts to χ on $\pi_1^{et}(A_{\overline{k}})$ and whose stalk at the identity has trivial Galois action.

Lemma 2.7. *Let H be a smooth hypersurface defined over a field k. Let $\chi : \pi_1(A_{\overline{k}}) \to \overline{\mathbb{Q}_p}^\times$ be a finite-order character such that $H^i(X_{\overline{k}}, \mathcal{L}_\chi) = 0$ for $i \neq n-1$.*

Then the Hodge-Tate weights of the $\mathrm{Gal}(\overline{k}'|k')$ action on $H^{n-1}(X_{\overline{k}'}, \mathcal{L}_\chi)$ are $0, \dots, n-1$ where the multiplicity of the weight q is $dA(n,q)$.

Proof. As a finite order character of $\pi_1(A)$, χ factors through $A[m]$ for some m. Let X' be the inverse image of X under the multiplication-by-m map of A. Then we can express $H^*(H_{\overline{k}'}, \mathcal{L}_\chi)$ as the part of the étale cohomology of $H^*(H'_{\overline{k}'}, \mathbb{Q}_p)$ where $A[m]$ acts by the character χ. Applying p-adic Hodge theory, we see that the dimension of the Hodge Tate weight q subspace of $H^{p+q}(H_{\overline{k}'}, \mathcal{L}_\chi)$ is equal to the dimension of the subspace of $H^p(H', \Omega^q_{X'})$ on which $A[m]$ acts by the character χ. By descent, this is the dimension of $H^p(H, \Omega^q_X \otimes L)$ for a torsion line bundle L on X. Because $H^{p+q}(H_{\overline{k}'}, \mathcal{L}_\chi)$ vanishes for $p \neq n-1-q$, the dimension of $H^p(X, \Omega^q_X \otimes L)$ for $p = n-1-q$ is equal to $(-1)^{n-1-q}$ times the Euler characteristic $\chi(X, \Omega^q_X \otimes L) = \chi(X, \Omega^q_X)$ because, as L is torsion, its Chern class vanishes. By Lemma 2.4, this Euler characteristic is $(-1)^{n-1-q} dA(n,q)$, so the dimension and Hodge-Tate multiplicity are both $dA(n,q)$. □

3. Sheaf convolution over a field

A *Tannakian category* over $\mathbb{Q}_\ell$ is a rigid symmetric monoidal $\mathbb{Q}_\ell$-linear abelian category with an exact tensor functor to the category of vector spaces over $\mathbb{Q}_\ell$. The point of these conditions is that Tannakian categories are necessarily equivalent to the category of representations of some pro-algebraic group (the group of automorphisms of the functor), together with the underlying vector space functor. Thus, associated to each object is some representation of this pro-algebraic group. For such a representation V, we refer to the image of the Tannakian group inside $GL(V)$ as the *Tannakian monodromy group*.

Krämer and Weissauer [39] constructed a Tannakian category as a quotient of the category of perverse sheaves on an abelian variety over an algebraically closed field (initially of characteristic zero, but Weissauer [63] later extended it to characteristic p), where the tensor operation is *sheaf convolution*. We will use the Tannakian monodromy groups from their theory to control usual monodromy groups.

In this section, we check that these Tannakian monodromy groups behave similarly to the usual monodromy groups with respect to the distinction between the geometric and arithmetic monodromy groups. In the setting of the étale fundamental group, we can define both geometric and arithmetic monodromy groups, with the geometric a normal subgroup of the arithmetic. We will check that the same works here. The Tannakian group of the category defined by Krämer and Weissauer will function as the geometric group, and we will define a Tannakian category of perverse sheaves over a non-algebraically closed field whose Tannakian monodromy groups will function as the arithmetic groups. We will verify that the geometric groups are normal subgroups of the arithmetic groups.

Our construction of the Tannakian category over a non-algebraically closed field will follow a version of the strategy of Krämer and Weissauer, and thus will also serve as a very brief review of our construction.

Let A be an abelian variety over a field k of characteristic zero. Let $D_c^b(A,\mathbb{Q}_\ell)$ be the derived category of bounded complexes of ℓ-adic sheaves on A with constructible cohomology. Define a *sheaf convolution* functor $* : D_c^b(A,\mathbb{Q}_\ell) \times D_c^b(A,\mathbb{Q}_\ell) \to D_c^b(A,\mathbb{Q}_\ell)$ that sends complexes K_1, K_2 to

$$K_1 * K_2 = a_*(K_1 \boxtimes K_2)$$

for $a : A \times A \to A$ the group law.

Lemma 3.1. $(D_c^b(A,\mathbb{Q}_\ell), *)$ *is a rigid symmetric monoidal category, where the unit object is the skyscraper sheaf at* 0*, and the dual of a complex* K *is*

$$K^\vee = [-1]^* DK$$

where D *is Verdier duality and* $[-1] : A \to A$ *is the inversion map.*

Proof. These were proved in [61, §2.1] (the symmetric monoidality and unit) and [62, Proposition] (the rigidity and description of the dual). These results are stated in the case where k is an algebraically closed field, but they proceed without modification in the general case. □

Let $\mathcal{P}$ be the category of perverse sheaves on A with $\mathbb{Q}_\ell$-coefficients. Let $\mathcal{N}$ be the subcategory of perverse sheaves with Euler characteristic zero. Let $D^b(\mathcal{N})$ be the category of complexes in $D_c^b(A,\mathbb{Q}_\ell)$ whose perverse homology objects lie in $\mathcal{N}$.

The Tannakian category will be constructed by combining this rigid symmetric monoidal structure with the abelian structure on the category of perverse sheaves. This requires modifying the category of perverse sheaves slightly because it is not quite stable under convolution. Instead one verifies that it is stable under convolution "up to" $\mathcal{N}$, i.e. that the convolution of two perverse sheaves has all perverse homology objects in nonzero degrees lying in $\mathcal{N}$. This lets us give $\mathcal{P}/\mathcal{N}$ the structure of a rigid symmetric monoidal $\mathbb{Q}_\ell$-linear abelian category.

Lemma 3.2. *(1) Perverse sheaves on A have nonnegative Euler characters.*

(2) $\mathcal{N}$ is a thick subcategory of $\mathcal{P}$ (i.e. it is stable under taking subobjects, quotients, and extensions).

(3) $D^b(\mathcal{N})$ is a thick subcategory of $D^b_c(A,\mathbb{Q}_\ell)$ (i.e. for any distinguished triangle with two objects in $D^b(\mathcal{N})$, the third one is in $D^b(\mathcal{N})$ as well.)

(4) For $K_1, K_2 \in D^b_c(A,\mathbb{Q}_\ell)$, if K_1 or K_2 lies in $D^b(\mathcal{N})$, then $K_1 * K_2$ lies in $D^b(\mathcal{N})$.

(5) For $K_1, K_2 \in \mathcal{P}$, ${}^p\mathcal{H}^i(K_1 * K_2) \in \mathcal{N}$ if $i \neq 0$.

Proof. It suffices to check all these statements after passing to $A_{\overline{k}}$, where they were checked by Krämer and Weissauer.

Indeed, for (1) it suffices to pass to $A_{\overline{k}}$ because Euler characteristics are preserved by base change, and the result then follows from [23, Corollary 1.4]. (2) follows immediately from (1) because Euler characteristic is additive in exact sequences, so if the middle of a short exact sequence has Euler characteristic zero, then the sides must have Euler characteristic zero, and vice versa. (3) follows from (2) because each perverse homology object of the third member of the distinguished triangle is an extension of a subobject of a perverse homology object of the first member by a quotient of a perverse homology object of a second member.

For (4) and (5), it suffices to pass to $A_{\overline{k}}$ because sheaf convolution is preserved by base change. Furthermore it suffices to handle the case where K_1 and K_2 are irreducible perverse sheaves, and thus in particular are semisimple. Then (4) follows from [39, Prop 10.1 and preceding paragraph]. For (5), this follows from [39, Lemma 13.1]. ☐

Lemma 3.3. *(1) Convolution descends to a functor*

$$D^b_c(A,\mathbb{Q}_\ell)/D^b(\mathcal{N}) \times D^b_c(A,\mathbb{Q}_\ell)/D^b(\mathcal{N}) \to D^b_c(A,\mathbb{Q}_\ell)/D^b(\mathcal{N}).$$

(2) The essential image of $\mathcal{P}$ in $D^b_c(A,\mathbb{Q}_\ell)/D^b(\mathcal{N})$ is equivalent to $\mathcal{P}/N$.

(3) The essential image of $\mathcal{P}$ in $D^b_c(A,\mathbb{Q}_\ell)/D^b(\mathcal{N})$ is stable under convolution.

(4) $(\mathcal{P}/\mathcal{N}, *)$ is a rigid symmetric monoidal $\mathbb{Q}_\ell$-linear abelian category.

Proof. (1) follows from Lemma 3.2(4).

(2) follows from Lemma 3.2(2) and [25, Proposition 3.6.1].

(3) follows from Lemma 3.2(5) because using the t-structure we see that any object $K_1 * K_2$ in $D^b_c(A,\mathbb{Q}_\ell)$ satisfying ${}^p\mathcal{H}^i(K_1 * K_2) \in \mathcal{N}$ if $i \neq 0$ can be reduced to an object of $\mathcal{P}$ by taking two mapping cones with objects of $D^b(\mathcal{N})$.

For (4), we first check that the rigid symmetric monoidal structure of Lemma 3.1 descends from $D^b_c(A,\mathbb{Q}_\ell)$ to $D^b_c(A,\mathbb{Q}_\ell)/D^b(\mathcal{N})$ (because isomorphisms and commutative diagrams are preserved by taking quotients) and that it restricts to the full subcategory $\mathcal{P}/\mathcal{N}$. Again, isomorphisms and diagrams restrict to full subcategories, so the only things to check are that the identity object lies in $\mathcal{P}/\mathcal{N}$, the dual of any object in $\mathcal{P}/\mathcal{N}$ lies in $\mathcal{P}/\mathcal{N}$, and the convolution of any two objects in

$\mathcal{P}/\mathcal{N}$ lies in $\mathcal{P}/\mathcal{N}$. The first is because the skyscraper sheaf is perverse, the second because Verdier duality preserves perverse sheaves, and the last is part (3). $\square$

For χ a character of $\pi_1^{et}(A_{\overline{K}})$, let $\mathcal{P}^\chi$ be the subcategory of perverse sheaves K with $H^i(A_{\overline{k}}, Q \otimes \mathcal{L}_\chi) = 0$ for all Q a subquotient of K defined over $\overline{k}$ and all $i \neq 0$, and let $\mathcal{N}^\chi = \mathcal{P}^\chi \cap \mathcal{N}$.

Lemma 3.4. *(1) The essential image of $\mathcal{P}^\chi$ in $\mathcal{P}/\mathcal{N}$ is equivalent to $\mathcal{P}^\chi/\mathcal{N}^\chi$.*
(2) $\mathcal{P}^\chi/\mathcal{N}^\chi$ contains the unit and is stable under convolution and duality.
(3) $K \mapsto H^0(A_{\overline{k}}, K \otimes \mathcal{L}_\chi)$ is an exact tensor functor from $\mathcal{P}^\chi/\mathcal{N}^\chi$ to $\mathbb{Q}_\ell$-vector spaces.
(4) The category $\mathcal{P}^\chi/\mathcal{N}^\chi$, convolution, and the functor $H^0(A_{\overline{k}}, K \otimes \mathcal{L}_\chi)$ are a rigid symmetric monoidal $\mathbb{Q}_\ell$-linear abelian category with an exact tensor functor to $\mathbb{Q}_\ell$-vector spaces.

Proof. (1) follows from [39, Lemma 12.3] and the fact that $\mathcal{P}^\chi$, by construction, is a thick subcategory.

The claims in (2) may be checked after passing to an algebraically closed field. To check that it contains the unit, we must check that the skyscraper sheaf at zero has cohomology only in degree zero, which is obvious. To check that it is closed under duality, it suffices to observe that

$$\begin{aligned} H^i(A_{\overline{k}}, [-1]^* DQ \otimes \mathcal{L}_\chi) &= H^i(A_{\overline{k}}, DQ \otimes [-1]_* \mathcal{L}_\chi) = H^i(A_{\overline{k}}, DQ \otimes \mathcal{L}_\chi^{-1}) \\ &= H^i(A_{\overline{k}}, D(Q \otimes \mathcal{L}_\chi)) = H^{-i}(A_{\overline{k}}, Q \otimes \mathcal{L}_\chi)^\vee \end{aligned}$$

so if one vanishes for all $i \neq 0$ the other does. That it is closed under convolution is checked in [39, Theorem 13.2].

The claims in (3) may be checked after passing to an algebraically closed field, where they are proved in [39, Theorem 13.2]. Specifically, [39, Proposition 4.1] reduces this to the case where χ is trivial. In this case, exactness follows from the long exact sequence of cohomology, which reduces to a short exact sequence because higher and lower cohomology groups vanish, and tensorness follows from the Künneth formula, which gives

$$H^0(A_{\overline{k}}, a_*(K_1 \boxtimes K_2)) = H^0(A_{\overline{k}} \times A_{\overline{k}}, K_1 \boxtimes K_2) = H^0(A_{\overline{k}}, K_1) \otimes H^0(A_{\overline{k}}, K_2)$$

again using the vanishing of higher and lower cohomology.

(4) It is rigid symmetric monoidal by part (2) and Lemma 3.3(4). It is $\mathbb{Q}_\ell$-linear abelian because it is the quotient of a $\mathbb{Q}_\ell$-linear abelian category by a thick subcategory. The functor is an exact tensor functor by part (3). $\square$

Any two fiber functors of the same Tannakian category are equivalent, and give rise to equivalent Tannnakian groups, so it will not matter which χ we pick. Moreover, by [39, Theorem 1.1], such χ exists for any K. So we can refer to the "Tannakian group" of an object without fixing a choice of χ, as we can always make a choice and the answer is independent of our choice.

For A an abelian variety over a field k with algebraic closure $\overline{k}$, we say that a perverse sheaf K on A is geometrically semisimple if its pullback to $A_{\overline{k}}$ is a sum of irreducible perverse sheaves. If $k = \overline{k}$ then semisimple and geometrically semisimple are equivalent.

Lemma 3.5. *Let K_1 and K_2 be geometrically semisimple perverse sheaves on A. Then $\mathrm{Hom}_{\mathcal{P}/\mathcal{N}}(K_1, K_2)$ is the quotient of the space of homorphisms $K_1 \to K_2$ by the subspace of homomorphisms factoring through an element of $\mathcal{N}$.*

Proof. Without loss of generality, we may assume that K_1 and K_2 are indecomposable. Because K_1 is semisimple when pulled back to $\overline{k}$, the set of isomorphism classes of irreducible components of this pullback must form a single $\mathrm{Gal}(\overline{k}|k)$-orbit, as otherwise the splitting into orbits would be defined over k and make them fail to be indecomposable. Thus, either all irreducible components of K_1 are in $\mathcal{N}$ or none of them are. The same is true for K_2.

If all irreducible components of K_1 or K_2 are in $\mathcal{N}$, then K_1 or K_2 is in $\mathcal{N}$, so maps in the quotient category are zero and all maps factor through elements of $\mathcal{N}$, and the statement holds.

Thus, we may assume that no irreducible components of K_1 and K_2 are in N. By definition, $\mathrm{Hom}_{\mathcal{P}/\mathcal{N}}(K_1, K_2)$ is the limit of $\mathrm{Hom}(K_1', K_2')$ where K_1' is a subobject of K_1 whose quotient lies in $\mathcal{N}$ and K_2' is a quotient of K_2 by an object in $\mathcal{N}$. By assumption, we must have $K_1' = K_1$ and $K_2' = K_2$, and similarly no nonzero map from K_1 to K_2 factors through an object in $\mathcal{N}$, so the statement holds in this case as well. □

Lemma 3.6.
1. *The full subcategory of $\mathcal{P}^\chi/\mathcal{N}^\chi$ consisting of geometrically semisimple perverse sheaves is a Tannakian subcategory of $\mathcal{P}^\chi/\mathcal{N}^\chi$.*
2. *The full subcategory of $\mathcal{P}^\chi_{\overline{k}}/\mathcal{N}^\chi_{\overline{k}}$ consisting of summands of the pullbacks to $A_{\overline{k}}$ of geometrically semisimple elements of $\mathcal{P}^\chi/\mathcal{N}^\chi$ on A_k is a Tannakian subcategory of $\mathcal{P}^\chi_{\overline{k}}/\mathcal{N}^\chi_{\overline{k}}$.*

Proof. To prove part (1), we must check that this subcategory contains the unit, and is closed under kernels, cokernels, direct sums, convolution, duals. For the unit, this is clear because the skyscraper sheaf is geometrically semisimple. For kernels and cokernels, by Lemma 3.5 it suffices to check that kernels and cokernels of maps between geometrically semisimple sheaves are geometrically semisimple, which is clear. That sums of geometrically semisimple sheaves are semisimple is obvious, and that duals of geometrically semisimple sheaves are geometrically semisimple follows from double-duality. For convolution, this follows from Kashiwara's conjecture, proven in [17] and [26].

To prove part (2), the arguments are similar. Because the pullbacks from A_k to $A_{\overline{k}}$ of geometrically semisimple perverse sheaves are (geometrically) semisimple, their summands are as well. Hence any kernel or cokernel of a map in $\mathcal{P}^\chi_{\overline{k}}/\mathcal{N}^\chi_{\overline{k}}$ between them is a kernel or cokernel of an honest map, hence a subobject or quotient, thus (by semisimplicity) a summand, and hence is also a summand of the pullback of the original geometrically semisimple perverse sheaf. A direct sum of summands is a summand of the direct sum, a convolution of summands is a summand of the convolution, and a dual of a summand is a summand of the dual, so the remaining three properties proceed by the same argument. □

Fix an abelian variety A over k and a character χ of $\pi_1^{geom}(A)$.

Let G_k be the Tannakian fundamental group of the full subcategory of $\mathcal{P}^\chi/\mathcal{N}^\chi$ consisting of geometrically semisimple perverse sheaves.

Let $G_{k'}$ be the Tannakian fundamental group of the full subcategory of $\mathcal{P}^{\chi}_{\overline{k}}/\mathcal{N}^{\chi}_{\overline{k}}$ consisting of summands of the pullbacks to $A_{\overline{k}}$ of geometrically semisimple perverse sheaves on A_k.

Let Gal_k be the Tannakian group of the category of ℓ-adic Galois representations over k.

Lemma 3.7. *The group $G_{k'}$ is a normal subgroup of G_k, with quotient Gal_k.*

Proof. There is a functor from the Tannakian category of Galois representations over k to the geometrically semisimple objects of $\mathcal{P}^{\chi}/\mathcal{N}^{\chi}$ that sends a Galois representation to the corresponding skyscraper sheaf at the identity.

There is a functor from geometrically semisimple perverse sheaves on A_k to summands of pullbacks of geometrically semisimple perverse sheaves to $A_{\overline{k}}$, given by pullback to $A_{\overline{k}}$.

Because these functors are both exact tensor functors, they define maps $G_{\overline{k}} \to G_k \to \mathrm{Gal}_k$. We wish to show that this is an exact sequence of groups, i.e. $G_{\overline{k}}$ is a normal subgroup of G_k whose quotient is Gal_k. To do this, we check the criteria of [18, Theorem A.1].

First, to check that $G_k \to \mathrm{Gal}_k$ is surjective, it suffices to check that the functor from Galois representations to skyscraper sheaves at the origin is full, and that a subquotient of a skyscraper sheaf at the origin is a skyscraper sheaf at the origin. These are both easy to check.

Second, to check that $G_{\overline{k}} \to G_k$ is a closed immersion, we must check that every representation of $G_{\overline{k}}$ is a subquotient of a pullback from G_k to $G_{\overline{k}}$ representation of G_k. This is automatic, as the Tannakian category of representations of G_k is defined to consist of perverse sheaves that are summands of pullbacks of perverse, geometrically semisimple sheaves on A_k that lie in P^{χ}, which by definition are representations of G_k, and because summands are an example of subquotients.

Third we must check that a perverse sheaf on A_k is a skyscraper sheaf at the origin if and only if is trivial when pulled back to $A_{\overline{k}}$. This is obvious.

Fourth, we must check that given a geometrically semisimple perverse sheaf on A_k, its maximal trivial subobject over $A_{\overline{k}}$ (i.e. the maximal sub-perverse sheaf that is a skyscraper sheaf at the origin) is a subobject over A_k. By duality, it is equivalent to check this with quotient objects, where the maximal trivial quotient is simply the stalk at zero of the zeroth homology and hence is certainly defined over k.

The fifth condition is simply the second condition with "subquotient" replaced with "subobject". This follows again because summands are subobjects. $\square$

It is likely possible to prove the analogous theorem, without the "geometrically semisimple" conditions in the definitions of the key Tannakian categories, by a similar but more complicated argument. However, this additional level of generality is not needed for our paper, and so we did not pursue this.

Using the fact that $G_{k'}$ is a normal subgroup of G_k, we will show that the Galois action on our fiber functor normalizes the Tannakian monodromy. First, we need a definition to make this Galois action well-defined:

Lemma 3.8. *Let A be an abelian variety over a field k. Let χ be a character of $\pi_1^{et}(A_{\overline{k}})$ that is $\mathrm{Gal}(\overline{k}|k)$-invariant. Let $\mathcal{L}_\chi$ be the associated character sheaf.*

Let K be a geometrically semisimple perverse sheaf on A such that $H^i_c(A_{\overline{k}}, A \otimes \mathcal{L}_\chi)$ vanishes for $i \neq 0$. Then

(1) *The action of* $\mathrm{Gal}(\overline{k}|k)$ *on* $H^0_c(A_{\overline{k}}, A \otimes \mathcal{L}_\chi)$ *normalizes the geometric Tannakian monodromy group of* K*.*
(2) *The action of* $\mathrm{Gal}(\overline{k}|k)$ *on* $H^0_c(A_{\overline{k}}, A \otimes \mathcal{L}_\chi)$ *normalizes the commutator subgroup of the identity component of the geometric Tannakian monodromy group of* K*.*

Proof. For (1), note that the Galois group acts by automorphisms of the fiber functor $H^i_c(A_{\overline{k}}, A \otimes \mathcal{L}_\chi)$ of the arithmetic Tannakian category, giving a map $\mathrm{Gal}_k \to G_k$. A trivial calculation involving skyscraper sheaves at the origin shows that this map splits the exact sequence of Lemma 3.7; hence Gal_k normalizes the geometric Tannakian group. Here it is crucial that, because the category of representations of $G_{k'}$ is a full Tannakian subcategory of $\mathcal{P}^\chi_{\overline{k}}/\mathcal{N}^\chi_{\overline{k}}$, Tannakian groups measured by this category are equal to the usual Tannakian groups.

(2) follows from (1) because the commutator subgroup of the identity component is a characteristic subgroup. $\square$

4. The Tannakian group of a hypersurface

Let A be an abelian variety of dimension $n \geq 2$ over the complex numbers and let $H \subset A$ be a smooth hypersurface. Let $i : H \to A$ be the inclusion. Let G_H be the Tannakian group of the perverse sheaf $i_* \mathbb{Q}[n-1]$. Let $d = [H]^n/n!$, which by Lemma 2.2 is an integer, equal to $(-1)^{n-1}$ times the arithmetic Euler characteristic of H.

Let $N = (n!)d = [H]^n$.

Recall from the introduction that $a(i)$ is the sequence

$$1, 5, 20, 76, 285, 1065, \dots$$

satisfying

$$a(1) = 1, a(2) = 5, a(i+2) = 4a(i+1) + 1 - a(i)$$

(OEIS A061278).

Theorem 4.1. *Assume that H is not equal to the translate of H by any nontrivial point of A. Assume also that none of the following hold:*

(1) $n = 2$ and $d = \binom{2k-1}{k}$ for some $k > 2$.
(2) $n = 2$ and $d = 2^r$ for some $r > 1$.
(3) $n = 2$ and $d = 28$.
(4) $n = 3$ and $d = \binom{a(i)+a(i+1)}{a(i)}/6$ for some $i \geq 2$.

Then G_H contains as a normal subgroup either SL_N, Sp_N, or SO_N. If H is not equal to any translate of $[-1]^*H$ then the SL_N case holds. If H is equal to such a translate, then if n is even Sp_N holds and if n is odd SO_N holds.

Remark 4.2. In the exceptional cases, there are only a few other possibilities for G_H. Suppose G_H does not contain SL_N, Sp_N, or SO_N. Then in case (1), G_H contains as a normal subgroup SL_{2r} acting by the representation $\wedge^r$. In case (2), G contains as a normal subgroup some SO acting by a 2^{r+1} dimensional spin representation. In case (3) it contains E_7 acting by its 56-dimensional representation. In (4) it contains as a normal subgroup SL_{a+b} acting by the representation $\wedge^b$.

The proof occupies the remainder of this section.

Lemma 4.3. *The dimension of the standard representation of G_H is N.*

Proof. By construction, the dimension of the representation associated to any object in the Tannakian category is the Euler characteristic of the corresponding perverse sheaf, which is $(-1)^{n-1}$ times the topological Euler characteristic of H. This now follows from Lemma 2.3. □

Lemma 4.4. *Assume that H is not equal to the translate of H by any nontrivial point of A. Then the Lie algebra of G_H acts irreducibly on the distinguished representation of G_H.*

Proof. This is [37, Corollary 1.6]. □

Lemma 4.5. *Assume that H is not equal to the translate of H by any nontrivial point of A. Then the distinguished representation of G_H is a minuscule representation of the Lie algebra of G_H.*

Proof. Because H is smooth, the characteristic cycle of $i_*\mathbb{Q}[n]$ is simply the conormal bundle of H with multiplicity 1, hence has a single irreducible component, with multiplicity one. The result then follows from [37, Corollary 1.9] □

Lemma 4.6. *Assume H is not translation-invariant by any nontrivial point of A. Then the identity component of G is simple modulo center.*

Proof. We follow the argument of [38, Theorem 6.2.1], with minor modifications. That theorem is not directly applicable because it assumes that H is symmetric (i.e. stable under inversion), which is not necessarily true here. Thus we restate the proof, which we can also simplify somewhat because our assumption that H is smooth is stronger than the analogous assumption in [38]. First, we review some notation and terminology from [38].

The proof relies on the notion of the characteristic cycle of a perverse sheaf. Classically, the characteristic cycle of a perverse sheaf on a variety X of dimension n is an effective Lagrangian cycle in the cotangent bundle T^*X of X. In other words, it is a nonnegative-integer-weighted sum of irreducible n-dimensional subvarieties of T^*X whose tangent space at a generic point is isotropic for the natural symplectic form on T^*X. All such subvarieties are automatically the conormal bundle to an irreducible subvariety of X, of arbitrary dimension, their support.

For A an abelian variety, because T^*A is a trivial bundle, we can express it as a product $A \times H^0(A, \Omega^1_A)$. The projection onto the second factor is called the Gauss map. Krämer considers an irreducible component *negligible* if its image under the Gauss map is not dense, and a cycle *clean* if none of its components are negligible [38, Definition 1.2.2]. He defines $cc(K)$ for a perverse sheaf K to be the usual characteristic cycle but ignoring any negligible components, making it automatically clean [38, Definition 2.1.1].

The degree of a cycle is the degree of the Gauss map restricted to that cycle. It is manifestly a sum over components of the degree of the Gauss map on that component, which vanishes if and only if the component is negligible.

Clean cycles are determined by their restriction to any open set in $H^0(A, \Omega^1_A)$. In particular, given two such cycles Λ_1, Λ_2, because

$$\dim \Lambda_1 = \dim \Lambda_2 = n = \dim H^0(A, \Omega^1_A)$$

one can find an open set over which both Λ_1 and Λ_2 are finite. The fiber product $\Lambda_1 \times_U \Lambda_2$ then maps to U by the obvious projection and to A by composing the two projections $\Lambda_1 \to A, \Lambda_2 \to A$ with the multiplication map $A \times A \to A$. Hence $\Lambda_1 \times_U \Lambda_2$ maps to $A \times U$. Its image has a unique clean extension to $A \times H^0(A, \Omega^1_A)$. Krämer defines this $\Lambda_1 \circ \Lambda_2$ to be this extension [38, Example 1.3.2].

A key property of this convolution product is that $\deg(\Lambda_1 \circ \Lambda_2) = \deg(\Lambda_1)\deg(\Lambda_2)$; as a consequence, if Λ_1, Λ_2 are clean and nonzero, then $\Lambda_1 \circ \Lambda_2$ is nonzero as well.

We are now ready to begin the argument.

Assume for contradiction that the identity component of G is not simple modulo its center. Then its Lie algebra is not simple modulo its center. By [38, Proposition 6.1.1], it follows from this that there exists $m \in \mathbb{N}$ and effective clean cycles Λ_1, Λ_2 on A, with $\deg(\Lambda_i) > 1$, such that

$$[m]_* cc(i_* \mathbb{Q}[n]) = \Lambda_1 \circ \Lambda_2$$

where $[m]$ is the multiplication-by-n map. Because H is smooth , the characteristic cycle of $i_*\mathbb{Q}[n]$ is simply the conormal bundle Λ_H of H, which is irreducible. Its degree is $d \cdot n!$, because the degree of the Gauss map of the conormal bundle to H is the sum of the multiplicities of vanishing of a general 1-form on H, which is the Euler characteristic of H, which is $d \cdot n!$. In particular, this degree is nonzero, so $cc(i_*\mathbb{Q}[n]) = \Lambda_H$.

Because H is not translation-invariant, the map from H to its image under $[m]$ is generically one-to-one, and so $[m]_*\Lambda_H$ is an irreducible cycle with multiplicity one in the cotangent bundle of A. This implies Λ_1 and Λ_2 are irreducible: If not, say if $\Lambda_1 = \Lambda_1^a + \Lambda_1^b$, we would have

$$\Lambda_H = \Lambda_1 \circ \Lambda_2 = \Lambda_1^a \circ \Lambda_2 + \Lambda_1^b \circ \Lambda_2$$

with both $\Lambda_1^a \circ \Lambda_2$ and $\Lambda_1^b \circ \Lambda_2$ nonzero, contradicting irreducibility.

It follows that Λ_1 and Λ_2 must be the conormal bundles Λ_{Z_1} and Λ_{Z_2} of varieties Z_1 and Z_2. Because $\deg(\Lambda_i) > 1$, neither Z_1 nor Z_2 can be a point, as the conormal bundle to a point is simply an affine space, and its Gauss map is an isomorphism, and thus has degree 1.

Let Y be the image of H under $[m]$. By [38, Lemma 5.2.2] there is a dominant rational maps from Y to Z_1 (say), and thus a dominant rational map from H to Z_1. Because H is smooth and Z_1 is a subvariety of an abelian variety, this dominant rational map automatically extends to a surjective morphism [11, Theorem 4.4.1]. Moreover, by the Lefschetz hyperplane theorem, A is the Albanese of H, so the surjective morphism $f_1 : H \to Z_1 \subseteq A$ extends to a homomorphism $g_1 : A \to A$, giving a commutative diagram

$$\begin{tikzcd} H \arrow[r, two heads, "f_1"] \arrow[d, hook, "i"] & Z_1 \arrow[d, hook] \\ A \arrow[r, "g_1"] & A \end{tikzcd}$$

Let B_1 be the image of g_1. Because f_1 is surjective, $Z_1 \subseteq B_1$.

If $Z_1 = B_1$ then Z_1 is an abelian variety, and the conormal bundle to any nontrivial abelian subvariety of A has Gauss map degree 0, contradicting $\deg(\Lambda_i) > 1$. Otherwise, by commutativity of the diagram, $H \subseteq g_1^* Z_1$. Because H is a hypersurface, it is a maximal proper subvariety of A, so $H = g_1^* Z_1$. This contradicts ampleness of H unless g_1 is finite, and contradicts H not being translation-invariant unless g_i is an isomorphism. This means Z_1 and H are isomorphic as subvarieties of an abelian variety. Thus $\deg(\Lambda_{Z_1}) = \deg(T^*_{Z_1} Z_1) = \deg(T^*_H H)$ and so because $\deg(\Lambda_{Z_1}) \deg(\Lambda_{Z_2}) = \deg(T^*_H H)$, we have $\deg(\Lambda_{Z_2} = 1)$, contradicting $\deg(\Lambda_i) > 1$.

Because we have a contradiction in every case, we have shown that G is simple modulo its center. □

Lemma 4.7. *Assume that H is not equal to the translate of H by any nontrivial point of A. Then the commutator subgroup of the identity component of G_H is one of the following:*

(1) SL_N, Sp_N or SO_N acting by their standard representation.
(2) SO_m acting by one of its spin representations, or E_6 or E_7 acting by their lowest-dimensional representation.
(3) SL_m acting by the representation $\wedge^k$ for some $2 \leq k \leq m/2$.

Proof. It follows from the Lemma 4.6 that the commutator subgroup of the identity component of G_H is a simple Lie group. Furthermore, from Lemmas 4.4 and 4.5, its standard representation must be irreducible and minuscule. But the above is an exhaustive list of minuscule representations of simple Lie groups (see e.g. [37, p. 7]). □

Lemma 4.8. *Assume that G_H contains as a normal subgroup one of SL_N, Sp_N, or SO_N. Then it contains SL_N only if H is not equal to any translate of $[-1]^*H$, it contains Sp_N only if n is even, and it contains SO_N only if n is oddt.*

Proof. This is essentially the argument of [40, Lemma 2.1] in a different setting.

Note first that the standard representation of any subgroup of GL_N which contains Sp_N or SO_N as a normal subgroup is equal to the tensor product of its dual representation with a one-dimensional representation, since it is contained in the normalizer GSp_N or GO_N respectively. Conversely, if $N > 2$ then the standard representation of any subgroup of GL_N which contains SL_N as a normal subgroup is not equal to the tensor product of its dual representation with any one-dimensional representation.

Translating into the language of the Tannakian category, we see that under this assumption, G_H contains SL_N as a normal subgroup if, and only if, the perverse sheaf $i_*\mathbb{Q}[n-1]$ is not isomorphic, up to negligible factors, to the convolution of its dual $[-1]^*Di_*\mathbb{Q}[n-1] = [-1]^*i_*\mathbb{Q}[n-1]$ with any perverse sheaf corresponding to a one-dimensional representation. Now perverse sheaves corresponding to a one-dimensional representation are always skyscraper sheaves [39, Proposition 10.1], and convolution with a skyscraper sheaf is equivalent to translation, so it is equivalent to say that $i_*\mathbb{Q}[n-1]$ is not isomorphic, up to negligible factors, to any translate of $[-1]^*i_*\mathbb{Q}[n-1]$. Because $i_*\mathbb{Q}[n-1]$ and $[-1]^*i_*\mathbb{Q}[n-1]$ are both irreducible perverse sheaves, there can be no negligible factors, and so this happens if and only if they are isomorphism. Because $i_*\mathbb{Q}[n-1]$ and $[-1]^*i_*\mathbb{Q}[n-1]$ are each constant sheaves on their support, they are isomorphic if and only if their supports are equal, which happens exactly when H is equal to a translate of $[-1]^*H$. This handles the SL_N case.

To distinguish Sp_N and SO_N, we observe that the second tensor power of the standard representations of both GSP_N and GO_N contain a one-dimensional summand, but for GSP_N it is contained in the $\wedge^2$ part while in GO_N it is contained in the Sym^2 part. More specifically this one-dimensional summand is the same one-dimensional representation from the previous argument, so it is equal to the skyscraper sheaf δ_a supported at a point a such that $H = a - H$. The second tensor power of the standard representation corresponds to the convolution

$$(i_*\mathbb{Q}[n-1]) * (i_*\mathbb{Q}[n-1]) = (mult \circ (i \times i))_*\mathbb{Q}[2n-2].$$

The relevant skyscraper sheaf must equal the degree 0 part of the stalk of $(mult \circ (i \times i))_*\mathbb{Q}[2n-2]$ at a, which is the shift by $2n-2$ of the cohomology of the fiber over a of the map $mult \circ (i \times i) : H \times H \to A$. Because $H = a - H$, this fiber is a diagonally embedded copy of H, of dimension $n-1$, so the degree 0 part of the shift by $2n-2$ of its cohomology is its top degree cohomology. The involution switching the two copies of $(i_*\mathbb{Q}[n-1])$ acts on this cohomology group by its action on the top cohomology of this fiber, which is trivial, twisted by the standard $(-1)^{n-1}$ from the Koszul sign rule for the tensor product of two complexes. Hence if n is odd, that involution fixes this skyscraper sheaf, so the skyscraper sheaf lies in Sym^n, and the normal subgroup must be SO_N, and if n is even, the involution acts by (-1), so the skyscraper sheaf lies in $\wedge^2$, and the normal subgroup must be Sp_N.

∎

To prove the main theorem, it remains to give a complete list of n, d for which G_H can contain as a normal subgroup one of the groups in Lemma 4.7, cases (2) or (3).

Case (2) is relatively easy as for these groups the dimensions have a special form.

Lemma 4.9. *Let A, H, and G_H be as in the setting of Theorem 4.1, so that the commutator subgroup of the identity component of G_H is one of the groups listed in Lemma 4.7. Suppose it is one of the groups in case (2): SO_m acting by a spin representation, or E_6 or E_7 acting by the standard representation. Then $n = 2$ and d is equal to 28 or equal to 2^r for some $r > 1$. .*

Proof. The spin representations have dimension a power of 2, and have exceptional isomorphisms to SL_2, Sp_4, or SL_4 as long as their dimensions are ≤ 4. For a power of two greater than four to equal $(n!)d$, we must have $n = 2$ and d a power of 2 greater than two. The standard representations of E_6 and E_7 have dimension 27 and 56 respectively. We can never have $(n!)d = 27$ and we can only have $(n!)d = 56$ if $n = 2$ and $d = 28$. □

The remainder of the section is devoted to restricting the case of wedge powers. We will obtain further numerical obstructions by introducing a Hodge torus into our Tannakian group. The action of the Hodge torus is obtained using the Galois action on $H^{n-1}(X_{\overline{k}'}, \mathcal{L}_\chi)$ and p-adic Hodge theory, and thus relies on Lemma 3.8 and our earlier construction of a Tannakian category of perverse sheaves over a non-algebraically closed field. An alternate approach should be possible, using classical Hodge theory and a Tannakian category of mixed Hodge modules on $A_{\mathbb{C}}$, as suggested by [39, Example 5.2], but we did not take this approach as we found the Galois action useful elsewhere in the argument.

Lemma 4.10. *Let H be a hypersurface of "degree" d on on abelian variety A of dimension n. Suppose that the commutator subgroup of the identity component of G_H is $\wedge^k SL_m$. Then there exists a function m_H from the integers to the natural numbers and an integer s such that $\sum_i m_H(i) = m$ and such that, for all $q \in \mathbb{Z}$*

$$\sum_{\substack{m_S:\mathbb{Z}\to\mathbb{Z} \\ 0\leq m_S(i)\leq m_H(i) \\ \sum_i m_S(i)=k \\ \sum_i i m_S(i)=s+q}} \prod_i \binom{m_H(i)}{m_S(i)} = dA(n,q). \tag{3}$$

Here we use the convention that $A(n,q) = 0$ unless $0 \leq q < n$.

Proof. By our assumption on the Tannakian group G_H and Lemma 3.8(2), it follows that the $\mathrm{Gal}(\overline{k}'|k')$ action on $H^{n-1}(X_{\overline{k}'}, \mathcal{L}_\chi)$ lies in the normalizer of SL_m.

This normalizer is GL_m/μ_k if $k \neq m/2$ and $GL_m/\mu_k \ltimes (\mathbb{Z}/2)$ if $k = m/2$.

The category of Galois representations generated by $H^{n-1}(X_{\overline{k}'}, \mathcal{L}_\chi)$ under tensor products, direct sums, subobjects, and quotients is a Tannakian category isomorphic to the category of representations of the Zariski closure of the image of $\mathrm{Gal}(\overline{k}'|k')$ acting on $H^{n-1}(X_{\overline{k}'}, \mathcal{L}_\chi)$. Because $H^{n-1}(X_{\overline{k}'}, \mathcal{L}_\chi)$ is Hodge-Tate and Hodge-Tate representations are closed under all these operations, all representations in this category are Hodge-Tate, and so the functor to a graded vector space provided by p-adic Hodge theory is an exact tensor functor. Because the category of graded vector spaces is isomorphic to the category of representations of $\mathbb{G}_m$, by

Tannakian duality there is a map from $\mathbb{G}_m$ to the Zariski closure of the image of $\mathrm{Gal}(\overline{k}'|k')$, whose weights are the Hodge-Tate weights of $H^{n-1}(X_{\overline{k}'}, \mathcal{L}_\chi)$.

Thus we have a map from $\mathbb{G}_m$ to GL_m/μ_k (or its semidirect product with $\mathbb{Z}/2$). Because $\mathbb{G}_m$ is connected, this defines a map $\mathbb{G}_m \to GL_m/\mu_k$. Its weight mulitplicities on the representation $\wedge^k$ must equal the Hodge-Tate multiplicities of $H^{n-1}(X_{\overline{k}'}, \mathcal{L}_\chi)$.

Homomorphisms $\mathbb{G}_m \to GL_m$ are parameterized by their weights, an m-tuple of integers with $w_1 \leq \cdots \leq w_m$. We can parameterize homomorphisms $\mathbb{G}_m \to GL_m/\mu_k$ by tuples $w_1, \dots, w_m, s$ of integers $w_1 \leq \cdots \leq w_m$ and $0 \leq s \leq m-1$, where the "weights" of the standard representation are $w_1 - \frac{s}{k}, \dots, w_m - \frac{s}{k}$.

Then the weights of the representation $\wedge^k$ are exactly $(\sum_{i\in S} w_i) - s$ for all $S \subseteq \{1, \dots m\}$ with $|S| = k$.

It follows that the multiplicity of the weight q inside $\wedge^k$ of the standard representation is the number of subsets $S \subseteq \{1, \dots, m\}$ with $|S| = k$ and $\sum_{i\in S} w_i = s + q$.

To calculate this, let $m_H(j)$ be the number of i such that $w_i = j$. For $S \subseteq \{1, \dots, m\}$, let $m_S(j)$ be the number of $i \in S$ such that $w_i = j$. Then $|S| = \sum_j m_S(j)$ and $\sum_{i\in S} w_i = \sum_j j m_S(j)$. Furthermore the number of sets S attaining a given function m_S is $\prod_i \binom{m_H(i)}{m_S(i)}$.

The stated identities then follow from Lemma 2.7.

□

To complete the argument, we have the following purely combinatorial proposition, which is proven in Appendix A. Recall from the introduction

Proposition 4.11 (Appendix A)**.** *Suppose that there exists a natural number k, function m_H from the integers to the natural numbers and an integer s such that $1 < k < -1 + \sum_i m_H(i)$ and satisfying Equation* (3) *for all $q \in \mathbb{Z}$. Then we have one of the cases*

(1) $m = 4$ and $k = 2$
(2) $n = 2$ and $d = \binom{2k-1}{k}$ for some $k > 2$.
(3) $n = 3$ and $d = \binom{a(i)+a(i+1)}{a(i)}/6$ for some $i \geq 2$. (Here $a(i)$ is defined by

$$a(1) = 1, a(2) = 5, a(i+2) = 4a(i+1) + 1 - a(i).)$$

We are now ready to prove the main theorem of this section:

Proof of Theorem 4.1. The commutator subgroup of the identity component of G_H must be one of the groups listed in the three cases of Lemma 4.7. In case (1), we know immediately that G_H contains as a normal subgroup one of SL_N, Sp_N or SO_N, and we conclude by Lemma 4.8.

We rule out case (2) by Lemma 4.9 and case (3) by Proposition 4.11. Note that Proposition 4.11 does not rule out the case $m = 4, k = 2$, but $\wedge^2 SL_4$ is SO_6 so G_H contains SO_6 as a normal subgroup in this case. □

5. Big monodromy from big Tannakian monodromy

Let X be a variety, A an abelian variety, and $Y \subseteq X \times A$ a family over X of smooth hypersurfaces in A, with X, A, Y all defined over $\mathbb{C}$. Let $n = \dim A$, so that $n-1$ is the relative dimension of Y over X. Let η be the generic point of X and $\overline{\eta}$ the geometric generic point. Let $f : Y \to X$ and $g : Y \to A$ be the projections.

Let $i : Y_{\overline{\eta}} \to A_{\overline{\eta}}$ be the inclusion, and let K be the perverse sheaf $K = i_* \mathbb{Q}_\ell[n-1]$ on $A_{\overline{\eta}}$. Let G be the Tannakian group of K. Let G^* be the commutator subgroup of the identity component of G.

Lemma 5.1. *Assume that G^* is a simple algebraic group acting by an irreducible representation, and that Y is not equal to a constant family of hypersurfaces translated by a section of A.*

Let K' be an irreducible perverse sheaf on $A_{\overline{\eta}}$ in the Tannakian category generated by K. Assume that K' is a pullback from $A_{\mathbb{C}}$ to $A_{\overline{\eta}}$ of a perverse sheaf on $A_{\mathbb{C}}$. Then G^* acts trivially on the irreducible representation of G corresponding to K'.

Proof. Let $pr_1, pr_2 : A_{\overline{\eta \times \eta}} \to A_{\overline{\eta}}$ be the two natural projection maps. The Tannakian monodromy groups of $pr_1^* K$ and $pr_2^* K$ are both G, so the Tannakian monodromy group of $pr_1^* K + pr_2^* K$ is a subgroup of $G \times G$ whose projection onto each factor is surjective. By Goursat's lemma, there are normal subgroups H_1, H_2 in G and an isomorphism $a : (G/H_1) \to (G/H_2)$ such that the Tannakian group of $pr_1^* K + pr_2^* K$ is

$$\{(g_1, g_2) \in G \times G | a(g_1) = g_2 \mod H_2\}.$$

Note that G has a unique factor in its Jordan-Hölder decomposition which is a nonabelian connected simple group. Hence this factor appears either in H_1 and H_2 or in G/H_1 and G/H_2.

In the first case, we must have $G^* \subseteq H_1$ and $G^* \subseteq H_2$. This is because, if the nonabelian connected simple factor appears in H_i, then it must appear in $H_i \cap G^*$ as $H_i/(H_i \cap G^*) \subseteq G/G^*$ which, modulo scalars, is contained in the outer automorphism group of G^* and thus is virtually abelian and cannot contain a nonabelian connected simple factor. Furthermore $H_i \cap G^*$ is a normal subgroup of G^*, and since it cannot be a finite group, it must be G^*.

Now, using the fact that $G^* \subseteq H_1$ and $G^* \subseteq H_2$, we will show that G^* acts trivially on the irreducible representation corresponding to K'. To do this, observe that $pr_1^* K'$ and $pr_2^* K'$ are isomorphic. These correspond to two representations of the Tannakian group of $pr_1^* K + pr_2^* K$ that factor through the projection onto the first and second factors respectively. Because G^* lies in H_1 and H_2, the Tannakian group of $pr_1^* K + pr_2^* K$ contains two copies of G^*. The first copy of G^* acts trivially on $pr_2^* K'$, so it must act trivially on $pr_1^* K'$, so G^* acts trivially on K', as desired.

In the second case, H_1 and H_2 must both be contained in the scalars. To see this, because the scalars are the centralizers of G^*, it suffices to show that the image of H_i in the automorphisms of G^* vanishes. Equivalently, we must show that the image of H_i in the automorphisms of the Lie algebra G^* vanishes. This automorphism group is an extension of the finite outer automorphism group of G^* by G^* mod its center. Because the image of H_i in the automorphism group is normalized by G^*, it either contains G^* or is finite, and it cannot contain G^*, so it is finite. Because it is finite and normalized by G^*, it commutes with G^*. Because the Lie algebra of G^* is an irreducible representation of G^*, this forces the image of H_i to act as scalars. But there are no nontrivial scalar automorphisms of a nonabelian Lie algebra, as they would never preserve any equation $[x,y]=z$, and so the image of H_i is trivial, as desired.

Now, using the fact that H_1 and H_2 are both contained in the scalars, we will derive a contradiction. Thus the Tannakian group $pr_1^*K + pr_2^*K$ is contained in the set $\{(g_1,g_2)\in G^2 | a(g_1)=g_2\}$ for some automorphism a of G mod scalars. Let us first check that the automorphism a is inner. To do this, let $\overline{\eta}_m$ be the geometric generic point of X^m. Let $pr_1,\dots,pr_m$ be the projections to η. For each i we have a homomorphism ρ_i from the Tannakian group of $\bigoplus_{i=1}^m pr_i^*K$ to the Tannakian group of pr_i^*K modulo scalars. Then because for any pair i,j there is a projection $X^m \to X^2$ onto the ith and jth copies of X, $pr_i^*K + pr_j^*K$ is isomorphic to $pr_1^*K + pr_2^*K$, so its Tannakian group is equal to the Tannakian group of $pr_i^*K + pr_j^*K$. Hence for any i and j, ρ_i and ρ_j are equal up to an automorphism of G mod scalars. Hence if m is greater than the order of the outer automorphism group of G mod scalars, there are i and j such that ρ_i equals ρ_j up to an inner automorphism. So in fact the maps from the Tannakian group of $pr_1^*K + pr_2^*K$ to the Tannakian groups of pr_1^*K and pr_2^*K are equal, mod scalars up to an inner automorphism, which we can ignore because the Tannakian group is well-defined only up to inner automorphisms in the first place.

Now the representation associated to $pr_1^*K * [-1]^*Dpr_2^*K$ is the standard representation of the first G tensored with the dual of the standard representation of the second G. Because the Tannakian group consists of pairs (g_1,g_2) where $g_2=\lambda g_1$ for a scalar λ, this tensor product contains a one-dimensional subrepresentation χ. (Viewing this tensor product as the space of endomorphisms of the standard representation, the one-dimensional subrepresentation consists of scalar endomorphisms). Because χ admits a nontrivial map to $\mathrm{std}_1 \otimes \mathrm{std}_2^\vee$, we have a natural map $\chi \otimes \mathrm{std}_2 \to \mathrm{std}_1$, which must be an isomorphism because both sides are irreducible. Any one-dimensional representation of the Tannakian group must be a skyscraper sheaf δ_x [39, Proposition 10.1], so we have an isomorphism $K_2 * \delta_x = K_1$ for some $x \in A(\overline{\eta \times \eta})$. Considering the support, we see that the translation of Y_{η_2} by x is Y_{η_1}. Spreading out this identity and then specializing η_2 to a sufficiently general point, we see that Y is generically the translation of a constant variety by a section x of A. We can extend this section to some open set, and then Y over some open set will be the translation of a constant variety by a section, and then because Y is a smooth proper family this will be true globally, contradicting the assumption. ☐

Lemma 5.2. *Assume that G^* is a simple algebraic group acting by an irreducible representation, and that Y is not equal to a constant family of hypersurfaces translated by a section of A.*

Let c be a positive integer, and let $i_1,\dots,i_c$ be the inclusions of A into A^c that send A to one of the c coordinate axes.

Then the Tannakian monodromy group of $\bigoplus_{j=1}^c i_{j*}K$ contains $(G^*)^c$, acting by the sum of c copies of the standard representation of G^*, as a normal subgroup, and this normal subgroup acts trivially on any representation of the Tannakian monodromy group corresponding to a perverse sheaf that is pulled back from $A^c_{\mathbb{C}}$ to $A^c_{\overline{\eta}}$.

Proof. Because i_{j*c} has Tannakian group G for all j, the Tannakian group of $\bigoplus_{j=1}^c i_{j*}K$ is contained in G^c. Let us check that its Tannakian group is actually equal to G^c. To do this, it suffices to show that each irreducible representation of G^c remains irreducible when restricted to the Tannakian group, and that nonisomorphic irreducible representations remain nonisomorphic. (This implication follows from [16, Proposition 2.21(a)] where the first hypothesis is automatic and for the second hypothesis, because G is reductive, we may restrict to irreducible representations.) Each irreducible representation of G^c is the tensor product of an ordered tuple of c irreducible representations $V_1, \dots, V_c$ of G. Each of these irreducible representations V_j of G corresponds to an irreducible perverse sheaf K_j on one of the copies of A, and the associated perverse sheaf on A^c is

$$i_{1*}K_1 * i_{2*}K_2 * \cdots * i_{c*}K_c = K_1 \boxtimes K_2 \boxtimes \cdots \boxtimes K_c.$$

Because the external tensor product of irreducible perverse sheaves is irreducible and perverse, the representation remains irreducible. Similarly, the external tensor product of two tuples of irreducible perverse sheaves, not elementwise isomorphic to each other, will remain nonisomorphic.

Because G^c is the Tannakian group of $\bigoplus_{j=1}^c i_{j*}K$, it contains $(G^*)^c$ as a normal subgroup.

Now let us check that any perverse sheaf corresponding to an irreducible representation which is a pullback of a perverse sheaf from $A^c_{\mathbb{C}}$ to $A^c_{\overline{\eta}}$ is trivial on $(G^*)^c$. Again, such a perverse sheaf is an external tensor product $K_1 \boxtimes K_2 \boxtimes \cdots \boxtimes K_c$ of perverse sheaves K_j on $A_{\overline{\eta}}$ corresponding to irreducible representations V_j of G, and the only way it can be a pullback is if all these perverse sheaves K_j are individually pullbacks, since we can recover the individual perverse sheaves by restricting to horizontal or vertical fibers and taking irreducible components. But then by Lemma 5.1, G^* acts trivially on all the V_j , so $(G^*)^c$ acts trivially on $V_1 \otimes \cdots \otimes V_c$. $\square$

We say that *the dual torus of* $\pi_1(A)$ is the set of characters $\pi_1(A) \to \mathbb{C}^\times$. Fixing an isomorphism $\overline{\mathbb{Q}}_\ell^\times$ and $\mathbb{C}$, the dual torus is equal to the set of characters $\pi_1(A) \to \overline{\mathbb{Q}}_\ell^\times$, and thus contains the set of continuous characters $\pi_1^{et}(A) \to \overline{\mathbb{Q}}_\ell^\times$ as a subgroup.

We call the set of characters of $\pi_1(A)$ trivial on the fundamental group of a nontrivial abelian subvariety of A a *proper subtorus* of the dual torus of $\pi_1(A)$.

Note that if a continuous character of $\pi_1^{et}(A)$, when viewed as a character of $\pi_1(A)$, is trivial on the topological fundamental group of an abelian subvariety of A, then it is trivial on the étale fundamental group of that abelian subvariety, because the topological fundamental group is dense in the étale fundamental group.

Lemma 5.3. *Let $K \in D_c^b(A_{\mathbb{C}(\eta)}, \mathbb{Q}_\ell)$ be a perverse sheaf of geometric origin. If no irreducible component of K is a pullback from $A_{\mathbb{C}}$, then for all characters $\chi : \pi_1^{et}(A) \to \overline{\mathbb{Q}}_\ell^\times$ outside a finite set of torsion translates of proper subtori of the dual torus to $\pi_1(A)$, we have $H^i(A_{\overline{\mathbb{C}(\eta)}}, K \otimes \mathcal{L}_\chi) = 0$ for $i \neq 0$ and $\left(H^0(A_{\overline{\mathbb{C}(\eta)}}, K \otimes \mathcal{L}_\chi)\right)^{\mathrm{Gal}(\overline{\mathbb{C}(\eta)}|\mathbb{C}(\eta))} = 0$.*

Proof. The first claim is a statement of the generic vanishing theorem of Kramer and Weissauer [39, Lemma 11.2]. (This is stated for varieties over $\mathbb{C}$ and singular cohomology, but we may embed $\mathbb{C}(\eta)$ into $\mathbb{C}$ and then base change from the étale to the analytic site).

The second claim follows from the same theorem, but indirectly. By restricting to an open subset of X, we may assume X is smooth. Let m be the dimension of X. We may spread K out (using the fact that it is of geometric origin) to a sheaf

K' over $A \times X$ such that $K'[m]$ is perverse. Let $\pi : A \times X \to A$ and $\rho : A \times X \to X$ be the projections.

We will prove the second claim by contradiction. We will first assume that $\left(H^0(A_{\overline{\mathbb{C}(\eta)}}, K \otimes \mathcal{L}_\chi)\right)^{\mathrm{Gal}(\overline{\mathbb{C}(\eta)}|\mathbb{C}(\eta))} \neq 0$ for a particular χ such that $H^i(A_{\overline{\mathbb{C}(\eta)}}, K \otimes \mathcal{L}_\chi) = 0$ for $i \neq 0$, and derive some conclusions from this. We will then define a finite set of torsion translates of proper subtori of the dual torus, assume that this nonvanishing holds for some χ outside their union, and derive a contradiction from that.

Let us first see how to interpret the nonvanishing of monodromy invariants in terms of the perverse sheaf K'. This will essentially be the usual observation that sheafs with monodromy invariants have global sections, and thus have nonzero H^0. Additional care must be taken because $\rho_*(K' \otimes \mathcal{L}_\chi)$ is a complex of perverse sheaves, but the decomposition theorem will give us exactly what is needed.

The stalk of $\rho_*(K' \otimes \mathcal{L}_\chi)$ at the generic point is the complex $H^*(A_{\overline{\mathbb{C}(\eta)}}, K \otimes \mathcal{L}_\chi)$. By the assumption that this cohomology group vanishes in degree $\neq 0$, the stalk of $\rho_*(K' \otimes \mathcal{L}_\chi)$ at the generic point is supported in degree 0. There is some open subset of Y over which $\rho_*(K' \otimes \mathcal{L}_\chi)$ remains a lisse sheaf in degree 0, and the Galois action matches the monodromy action of that open subset. By the decomposition theorem, $\rho_*(K' \otimes \mathcal{L}_\chi)$ is a sum of shifts of irreducible perverse sheaves. In particular, this monodromy action is semisimple.

Now we assume that the Galois invariants are nonzero. It follows that the monodromy invariants are nonzero, and thus, by semisimplicity, there is a rank one monodromy-invariant summand. Equivalently, there is a summand of $\rho_*(K' \otimes \mathcal{L}_\chi)$, restricted to this open set, that is isomorphic to the constant sheaf $\mathbb{Q}_\ell$. Because $\rho_*(K' \otimes \mathcal{L}_\chi)$ is a sum of shifts of irreducible perverse sheaves, this irreducible summand $\mathbb{Q}_\ell$ on an open set must extend to a shift of an irreducible perverse sheaf on the whole space. Because X is smooth, the unique irreducible extension of the constant sheaf $\mathbb{Q}_\ell$ from an open subset to all of X is $\mathbb{Q}_\ell$, which is a perverse sheaf shifted by m. (In general, it would be the IC sheaf of X.)

Because $H^0(X, \mathbb{Q}_\ell) \neq 0$, and $\mathbb{Q}_\ell$ is a summand of $\rho_*(K' \otimes \mathcal{L}_\chi)$, it follows that $H^0(X, \rho_*(K' \otimes \mathcal{L}_\chi)) \neq 0$.

Now that we have interpreted the existence of nontrivial monodromy invariants cohomologically, we can re-express the cohomology group in terms of shaves on A, which will enable us to understand its dependence on χ using the generic vanishing theorem. It follows from the Leray spectral sequence that

$$0 \neq H^0(X, R\rho_*(K' \otimes \mathcal{L}_\chi)) = H^0(A \times X, K' \otimes \mathcal{L}_\chi) = H^0(A, \pi_* K' \otimes \mathcal{L}_\chi).$$

For the last equality, we need to know $\pi_*(K' \otimes \mathcal{L}_\chi) = (\pi_* K') \otimes \mathcal{L}_\chi$. This is a local question, and working étale-locally, $\mathcal{L}_\chi$ is trivial, so it is automatic.

Now we choose our finite union of torsion translates of subtori. We apply the generic vanishing theorem to every perverse sheaf in sight. For all $j \in \mathbb{Z}$ with ${}^p\mathcal{H}^j(\pi_* K') \neq 0$, we take from [39, Lemma 11.2] a finite set of torsion translates of subtori such that $H^i(A, {}^p\mathcal{H}^j(\pi_* K') \otimes \mathcal{L}_\chi) = 0$ for all χ not in this set and all $i \neq 0$. Because there are only finitely many j where ${}^p\mathcal{H}^j(\pi_* K') \neq 0$, the union of all of these is again a finite set.

Assume that $\left(H^0(A_{\overline{\mathbb{C}(\eta)}}, K \otimes \mathcal{L}_\chi)\right)^{\mathrm{Gal}(\overline{\mathbb{C}(\eta)}|\mathbb{C}(\eta))} \neq 0$ for a particular χ not in this set. The vanishing of $H^i(A, {}^p\mathcal{H}^j(\pi_* K') \otimes \mathcal{L}_\chi)$ for all j and all $i \neq 0$ forces the spectral sequence

$$H^i(A, {}^p\mathcal{H}^j(\pi_*K') \otimes \mathcal{L}_\chi) \mapsto H^{i+j}(A, \pi_*K' \otimes \mathcal{L}_\chi)$$

to degenerate, giving

$$0 \neq H^0(A, \pi_*K' \otimes \mathcal{L}_\chi) = H^0(A, {}^p\mathcal{H}^0(\pi_*K')\mathcal{L}_\chi).$$

In particular, we can conclude that

$${}^p\mathcal{H}^0(\pi_*K') \neq 0.$$

We will now derive a contradiction from this simpler statement, which notably is independent of χ.

Note first that because $K'[m]$ is perverse, and π has fibers of dimension at most m, we have ${}^p\mathcal{H}^i(\pi_*K') = 0$ for $i < 0$ by [8, 4.2.4]. Hence there is a natural map

$${}^p\mathcal{H}^0(\pi_*K') \to \pi_*K'$$

arising from the perverse t-structure. Because ${}^p\mathcal{H}^{2m}(\pi_*K')$ is nonzero, this map must be nonzero. Applying adjunction, we obtain a nonzero map $\pi^{*p}\mathcal{H}^0(\pi_*K')[m] \to K'$, and and shifting by m, a nonzero map

$$\pi^{*p}\mathcal{H}^0(\pi_*K')[m] \to K'[m].$$

Because this is a nonzero map between perverse sheaves, some irreducible component of the source is equal to some irreducible component of the target. This cannot happen because by assumption no irreducible component of K is a pullback from $A_{\mathbb{C}}$, giving a contradiction. □

Lemma 5.4. *Let G^* be a simple algebraic group with a standard representation, and let $N(G^*)$ be its normalizer inside the group of automorphisms of this representation. Then there is a finite list of irreducible representations of $N(G^*)^c$ such that a reductive subgroup of $N(G^*)^c$ contains $(G^*)^c$ if and only if it has no invariants on any of these representations.*

Proof. Let us first handle the case where $c = 1$. Let H be a reductive subgroup of $N(G^*)$. By Larsen's alternative, if H acts irreducibly on the Lie algebra of G^* then H is either finite mod scalars or contains G^*. (To prove this, note that the Lie algebra of $N(G^*)$ splits into the Lie algebra of G^* and scalars. The Lie algebra of H is an H-invariant subspace of this, so if the Lie algebra of G^* is irreducible as a representation of H, the Lie algebra of H either contains the Lie algebra of G^* or is contained in the scalars. In the first case, H contains G^*, and in the second case, H is finite modulo scalars. This is a slight variant of a statement due to Larsen in unpublished preprint work, and later published by Katz [35, Theorem 2.2.2].)

If H is finite mod scalars, then by the Jordan-Schur theorem H contains a normal abelian subgroup of index $\leq f(N)$ for some explicit function f. Thus all irreducible representations of H have dimension $\leq f(N)$. Let V be some fixed irreducible representation of $N(G^*)$ of dimension $> f(N)$ that remains irreducible when restricted to G^*. If H acts irreducibly on the Lie algebra of G^* and on V then H is finite. To check that H acts irreducibly on V, it is sufficient to check that it acts nontrivially on every irreducible subrepresentation of $V \otimes V^\vee/\mathbb{C}$, where we mod out by the trivial subrepresentation $\mathbb{C}$ arising from scalars. The same works for the adjoint representation.

Based on this, we take our finite list to be all irreducible subrepresentations of $(V \otimes V^\vee)/\mathbb{C}$ and $(\operatorname{ad} G^* \otimes \operatorname{ad} G^{*\vee})/\mathbb{C}$. We have shown that if these representations have no H-invariants then H contains G^*. Conversely, because V is irreducible upon restriction to G^*, $V \otimes V^\vee/\mathbb{C}$ has no G^*-invariants, and thus neither does any irreducible subrepresentation, and the same argument works for $\operatorname{ad} G^*$, so these have no invariants if H contains G^*. This proves the statement in the case $c = 1$.

For greater c, we take our finite list to consist of, for each $1 \leq i \leq c$, the list in the $c = 1$ case composed with the projection onto the ith copy of $N(G^*)$, combined with, for each $1 \leq i < j \leq c$, the Lie algebra of the ith copy of G^* tensored with the dual of the Lie algebra of the jth copy of G^*. These extra representations manifestly have no $(G^*)^c$-invariants. Conversely, if H acts without invariants on these representations, then by the previous discussion we know that the projection of H onto the ith copy of $N(G^*)$ contains G^*. Hence the projection of the commutator subgroup of the identity component of H onto the ith copy of G^* is G^*. By [36, Theorem on p. 1152], if the commutator subgroup of the identity component of H is not $(G^*)^c$, then there is i, j with $1 \leq i < j \leq c$ such that the projections of this subgroup to the ith and jth copy of G^* are isomorphic, mod scalars. This isomorphism is unique and thus is defined over H. Then the Lie algebras of the ith and jth copy of G^* must be isomorphic as representations of H, so the tensor product of one with the dual of the other has invariants, contradicting our assumption. □

Remark 5.5. Lists of representations satisfying the condition of Lemma 5.4 with smaller dimension here follow from Larsen's conjecture [41, Theorem 1.4] in the case that G^* is a classical group, but these depend on the classification of finite simple groups. For some applications of results of this type, optimizing the dimension of the representations is relevant, but not here.

The following theorem is the analogue of Pink's specialization theorem [34, Theorem 8.18.2], which shows, given any sheaf on the total space family of schemes, for "most" schemes in the family, the monodromy of the restricted sheaf is equal to a generic monodromy group. Our analogue shows that for "most" characters χ, the monodromy of $\bigoplus_{i=1}^c R^{n-1} f_*(g^*\mathcal{L}_{\chi_i})$ is (roughly) equal to a generic Tannakian monodromy group. This will connect our earlier investigations of the Tannakian group to our later arguments, which require control on monodromy groups.

Theorem 5.6. *Assume that $Y_{\overline{\eta}}$ is not translation-invariant by any nonzero element of A, that G^* is a simple algebraic group acting by an irreducible representation, and that Y is not equal to a constant family of hypersurfaces translated by a section of A.*

Then for $\chi_1, \dots, \chi_c$ characters of $\pi_1^{et}(A)$, with $(\chi_1, \dots, \chi_c)$ avoiding some finite set of torsion translates of proper subtori of the dual torus to $\pi_1(A^c)$, the following conditions are satisfied:

- *$R^k f_*(g^*\mathcal{L}_{\chi_i}) = 0$ for $k \neq n-1$, and*
- *the geometric monodromy group of $\bigoplus_{i=1}^c R^{n-1} f_*(g^*\mathcal{L}_{\chi_i})$ contains as a normal subgroup $(G^*)^c$ acting on the sum of c copies of the fixed irreducible representation.*

Proof. First, we use the generic vanishing theorem to find a finite set of torsion translates of proper subtori of the dual torus to $\pi_1(A)$ such that, for χ avoiding them, $R^k f_*(g^*\mathcal{L}_\chi) = 0$ for $k \neq n-1$. We will then take the inverse images of these subtori under the duals of the c inclusions $\pi_1(A) \to \pi_1(A^c)$ to be in our finite set of torsion translates of subtori of the dual torus to $\pi_1(A^c)$.

Next, to calculate the monodromy, we will use the fact that the monodromy group is equal to the Zariski closure of the image of $\mathrm{Gal}(\overline{\mathbb{C}(\eta)}|\mathbb{C}(\eta))$ acting on the stalk at the geometric generic point.

Let us first check that the geometric monodromy group of $\bigoplus_{i=1}^{c} R^{n-1}f_*(g^*\mathcal{L}_{\chi_i})$ is contained in $N(G^*)^c$. It suffices to show for each i that the geometric monodromy group of $R^{n-1}f_*(g^*\mathcal{L}_{\chi_i})$ is contained in $N(G^*)$. This follows from Lemma 3.8(2).

Now by Lemma 5.4 we can find an explicit list of representations of $N(G^*)^c$ such that any reductive subgroup of $N(G^*)^c$ contains $(G^*)^c$ if and only if its action on all these representations has no invariants. By Deligne's theorem, the monodromy group of $\bigoplus_{i=1}^{c} R^{n-1}f_*(g^*\mathcal{L}_{\chi_i})$ is reductive, and so we can apply this lemma.

By Lemma 3.7, each representation from the list of Lemma 5.4 corresponds to a perverse sheaf on A^c_η in the arithmetic Tannakian category generated by $\bigoplus_{j=1}^{c} i_{j*}i_*\mathbb{Q}_\ell[n-1]$. (We have to check that $\bigoplus_{j=1}^{c} i_{j*}i_*\mathbb{Q}_\ell[n-1]$ is geometrically semisimple, but this is clear as the constant sheaf on any closed subvariety is semisimple. It follows that the action of G_k on the representation associated to this complex factors through the normalizer of the action of $G_{k'}$, and thus factors through $N(G^*)^c$, and so all representations of $N(G^*)^c$ corerspond to geometrically semisimple perverse sheaves.)

Because $(G^*)^c$ acts nontrivially on all representations from Lemma 5.4, by Lemma 5.2 none of these perverse sheaves is a pullback from $A^c_{\mathbb{C}}$, so by Lemma 5.3, outside some finite set of torsion translates of subtori of the dual torus of $\pi_1(A^c)$, the Galois group has no invariants for these representations. Thus, outside some finite set of torsion translates, the Galois group contains $(G^*)^c$. □

Lemma 5.7. *Suppose A is defined over a number field K, and let p be a prime at which A has good reduction. For any natural number c, for any finite set S of torsion translates of proper subtori of the dual torus $\Pi(A)^c$ to $\pi_1(A^c)$, there exist $\sigma_1, \dots, \sigma_c \in \mathrm{Gal}(\overline{\mathbb{Q}}_p|K_p)$ and some finite set S' of torsion translates of proper subtori of the dual torus to A, such that for any χ outside the union of S', the tuple $(\sigma_1(\chi), \dots, \sigma_c(\chi))$ does not lie in the union of S.*

Proof. By enlarging K_p if necessary, we may assume that all the tori in S are defined over K_p. Let σ be a Frobenius element of $\mathrm{Gal}(\overline{\mathbb{Q}}_p|K_p)$, acting on $\pi_1(A^c)$. We will prove this lemma as a consequence of the fact that this action of σ is invertible, with no roots of unity as eigenvalues.

To illustrate the argument, let's first consider the case where $S = \{T + \xi\}$ consists of only a single torsion translate of a torus. Let Z be the quotient of $\mathrm{Hom}(\pi_1^{et}(A^c), \overline{\mathbb{Q}}_\ell^\times)$ by $\mathrm{Hom}(\pi_1^{et}(A^c), \overline{\mathbb{Q}}_\ell^\times) \cap T$. Then Z is an abelian group, and there is a natural homomorphism $f : \mathrm{Hom}(\pi_1^{et}(A^c), \overline{\mathbb{Q}}_\ell^\times) \to Z$. We can describe $T + \xi$ as the inverse image of $f(\xi)$. Because by assumption T is defined over K_p, there is an action of $\mathrm{Gal}(\overline{\mathbb{Q}}_p|K_p)$ on Z making f into $\mathrm{Gal}(\overline{\mathbb{Q}}_p|K_p)$-equivariant homomorphism.

By the definition of subtori, there exists an abelian subvariety B of A such that T consists of characters vanishing on B, and thus $Z = \mathrm{Hom}(\pi_1^{et}(B), \overline{\mathbb{Q}}_\ell^\times)$ and f is dual to the inclusion $B \to A^c$.

We'll take our Galois elements to be powers of Frobenius: $\sigma_i = \sigma^{e_i}$, for integers $e_1, \dots, e_c$. Consider the linear map

$$g = g_{e_1,\dots,e_c} \colon \mathrm{Hom}(\pi_1^{et}(A), \overline{\mathbb{Q}}_\ell^\times) \to Z$$

given by
$$g_{e_1,\dots,e_c}(x) = f(\sigma_1(x), \sigma_2(x), \dots, \sigma_c(x)).$$
If g is not uniformly zero, then $g^{-1}(f(\xi))$ is a finite union of torsion translates of proper subtori, and we are done. (This is because g is dual to the homomorphism of abelian varieties $B \to A^c \to A$ by composing the inclusion with $(x_1, \dots, x_c) \mapsto \sigma_1(x_1)+\cdots+\sigma_c(x_c)$. If g is not uniformly zero then the image of this homomorphism includes some nontrivial abelian subvariety. Thus, characters pulled back to zero under this homomorphism vanish on some fixed finite étale cover of that nontrivial abelian subvariety. Hence, characters pulled back to zero under this homomorphism lie in a finite union of torsion translates of characters that vanish on that nontrivial abelian subvariety. The same is true for characters that pull back to $f(\xi)$.) Hence, we just need to find some $e_1, \dots, e_c$ for which g is not uniformly zero. We'll say that the tuple $(e_1, \dots, e_c)$ is *bad* for T if g is uniformly zero.

For $1 \le r \le c$, let $\Pi_r(A^c)$ be the set of characters of $\pi_1(A^c) = \pi_1(A)^c$ trivial on each direct summand $\pi_1(A)$ except possibly the r-th, so we have the direct sum decomposition $\Pi(A^c) = \bigoplus_{r=1}^c \Pi_r(A^c)$. Since T is a proper subtorus of the dual torus, there exists some index r such that T does not contain $\Pi(A_r)$; in other words,
$$f(0, \dots, 0, x, 0, \dots, 0)(\text{with } x \text{ in the } r\text{th place})$$
is not uniformly zero. We'll call such an r nondegenerate for T.

Now if $e_r' \neq e_r$, the endomorphism
$$x \mapsto \sigma^{e_r'}(x) - \sigma^{e_r}(x)$$
of $\Pi(A)$ is surjective. If $(e_1, \dots, e_r, \dots, e_c)$ is bad for T, then $(e_1, \dots, e_r', \dots, e_c)$ is not bad for T.

Now return to the general case, where S is a finite set of torsion translates of proper subtori T of $\Pi(A)$. We need to show that there exists $(e_1, \dots, e_r, \dots, e_c)$ that is not bad for any of these finitely many T. For each T, let $r(T)$ be the least index that is nondegenerate for T. We'll work by reverse induction on r (from c to 1).

Suppose we have a tuple $(e_1, \dots, e_r, \dots, e_c)$ that is not bad for any T with $r(T) > r$. Any such T contains $\Pi_r(A^c)$, so changing e_r does not affect this property. Now consider all T such that $r(T) = r$. For each such T, only one value of e_r can give a bad tuple $(e_1, \dots, e_r, \dots, e_c)$. Thus, we can choose e_r' so that the tuple $(e_1, \dots, e_r', \dots, e_c)$ is not bad for any T with $r(T) \ge r$. Repeating this process for $r(T) = c, c-1, \dots, 1$, we produce a tuple $(e_1, \dots, e_r', \dots, e_c)$ for which no T is bad. ☐

Lemma 5.8. *Suppose A is defined over a number field K. For any finite set of torsion translates of proper subtori of the dual torus to $\pi_1(A)$, there exists a torsion character χ of $\pi_1(A)$ such that no $\mathrm{Gal}(\overline{\mathbb{Q}}|K)$-conjugate of χ lies in the union of that set. Furthermore, we can arrange that χ is of order a power of ℓ, for any given prime ℓ.*

Proof. For χ to have no Galois conjugate in any of these translates of subtori, it suffices that χ not lie in any Galois conjugate of these translates of subtori.

By definition, each proper subtorus of the dual torus corresponds to some abelian subvariety, which must be defined over a number field. Since every torsion point is defined over a number field, every torsion translate of a proper subtorus is defined over a number field. Hence they have finitely many Galois conjugates, and the union of all their Galois conjugates is a finite union of torsion translates of proper subtori.

Each proper subtorus can contain at most $\ell^{2k(n-1)}$ ℓ^k-torsion characters, and so any translate of a proper subtorus can contain at most $\ell^{2k(n-1)}$ ℓ^k-torsion characters, while there are ℓ^{2kn} ℓ^{2k}-torsion characters in total, so as soon as ℓ^{2k} is greater than this finite number of tori, there will be an ℓ^k-torsion character not in any of them. □

For the remainder of this section, we'll suppose A is an abelian variety over a number field K, X is a smooth scheme over $\mathbb{Q}$, and

$$Y \subseteq X \times_{\mathbb{Q}} A = X_K \times_K A$$

is a family of hypersurfaces in A, smooth, proper and flat over X_K. For every embedding $\iota\colon K \to \mathbb{C}$, we can form schemes

$$A_\iota = A \otimes_{\operatorname{Spec} K, \iota} \operatorname{Spec} \mathbb{C}$$

and

$$Y_\iota = Y \otimes_{\operatorname{Spec} K, \iota} \operatorname{Spec} \mathbb{C};$$

these are both schemes over $\mathbb{C}$, and Y_ι has a projection to

$$X_{\mathbb{C}} = X \otimes_{\operatorname{Spec} \mathbb{Q}} \operatorname{Spec} \mathbb{C}.$$

Let $f\colon Y_\iota \to X_{\mathbb{C}}$ and $g\colon Y_\iota \to A_\iota$ be the projections; for every torsion character χ of $\pi_1^{et}(A_\iota)$, let $\mathcal{L}_{(\iota,\chi)}$ be the corresponding character sheaf on A_ι.

Corollary 5.9. *Assume that $Y_{\overline{\eta}}$ is not translation-invariant by any nonzero element of A, that G^* is a simple algebraic group acting by an irreducible representation, and that Y is not equal to a constant family of hypersurfaces translated by a section of A.*

Then for any prime ℓ, there exist an embedding $\iota\colon K \to \mathbb{C}$ and a torsion character χ of $\pi_1^{et}(A_\iota)$, of order a power of ℓ, such that for every conjugate (ι', χ') of (ι, χ) by an element of $\operatorname{Gal}(\overline{\mathbb{Q}}|K)$:

- *for $k \neq n-1$, we have $R^k f_{\iota' *}(g_{\iota'}^* \mathcal{L}_{(\iota',\chi')}) = 0$, and*
- *there exist $\sigma_1, \dots, \sigma_c \in \operatorname{Gal}(\overline{\mathbb{Q}}_p|\mathbb{Q}_p)$ such that the monodromy group of $\bigoplus_{i=1}^c R^{n-1} f_{\iota' *}(g_{\iota'}^* \mathcal{L}_{\sigma_i(\iota',\chi')})$ contains as a normal subgroup $(G^*)^c$ acting on the sum of c copies of the fixed irreducible representation.*

Proof. This follows from the previous 3 results, applying Theorem 5.6 to each of the finitely many pairs (Y_ι, A_ι). □

Using Theorem 5.6, we can also prove a result on the period maps of certain variations of Hodge structures associated to families of hypersurfaces in an abelian variety. This is not used anywhere in this paper. Instead, it provides a different perspective on our main result, showing that it is compatible with the philosophy of Javenpeykar and Loughran that varieties with a quasi-finite period map should have finitely many $\mathcal{O}_K[1/S]$-points for any number field K and set S of prime ideals.

Proposition 5.10. *Let A be an abelian variety of dimension $n \geq 2$ over $\mathbb{C}$. Let ϕ be an ample class in the Picard group of A. For a natural number m, let $[m] : A \to A$ be the multiplication-by-m map.*

There exists a natural number m such that the natural period m from the moduli space of smooth hypersurfaces H in A of class ϕ to a period domain, which sends H to the Hodge structure on $H^{n-1}([m]^{-1}H,\mathbb{Q})$, is quasi-finite.

Proof. Let $\mathcal{M}$ be this moduli space. Suppose that the period map is not quasi-finite for some m. Then its fiber over some point must contain a positive dimensional analytic subvariety. Consider the variation of Hodge structures over $\mathcal{M}\times\mathcal{M}$ whose fiber over a pair of hypersurfaces H_1, H_2 is $H^{n-1}([m]^{-1}H_1,\mathbb{Q}) \otimes H^{n-1}([m]^{-1}H_2,\mathbb{Q})^\vee$. Over the diagonal in $\mathcal{M}\times\mathcal{M}$, this variation of Hodge structures has a Hodge class representing the identity isomorphism between the two Hodge structures. By [12, Corollary 1.3], the projection from the universal cover of $\mathcal{M}\times\mathcal{M}$ of the locus where this cohomology class is Hodge is a Zariski closed subset $Z_m \subseteq \mathcal{M}\times\mathcal{M}$. Because Z_m certainly contains the square of the positive-dimensional analytic subvariety discussed earlier, and hence the first projection $Z_m \to \mathcal{M}_1$ is not quasifinite.

If m_1 divides m_2 then we have $Z_{m_2} \subseteq Z_{m_1}$. Because $\mathcal{M}\times\mathcal{M}$ is Noetherian, it follows that there exists m such that $Z_{m'} = Z_m$ whenever m' is a multiple of m.

We claim now that the period map is quasi-finite for this m. Suppose not. Then the projection $Z_m \to \mathcal{M}_1$ is not quasi-finite. We will derive a contradiction from this. Fix $x \in \mathcal{M}$ such that the fiber of Z_m over x has positive dimension. and let X be the smooth locus of some irreducible component of the fiber of Z_m over x. Let $f : Y \to X$ be the universal family of hypersurfaces H_2 over X and $g : Y \to A$ the projection map. By assumption, for all multiples m' of m, for all $y \in X$, we have an isomorphism of Hodge structures

$$H^{n-1}([m']^{-1}H_{1,x},\mathbb{Q}) = H^{n-1}([m']^{-1}H_{1,y},\mathbb{Q}) \cong H^{n-1}([m]^{-1}H_{2,y},\mathbb{Q}).$$

Hence the variation of Hodge structures $H^{n-1}([m']^{-1}H_{2,y},\mathbb{Q})$ is constant, and thus has finite monodromy. This is the sum, over characters χ of $\pi_1(A)$ of order dividing m', of $R^{n-1}f^*(g^*\mathcal{L}_\chi)$, and so all these individual summands have finite monodromy.

The family $Y \to X$ is a family of smooth hypersurfaces. Because we have fixed an ample class in the Picard group, and there are only finitely many translates of a given hypersurface up to translation, and because X is a positive-dimensional subvariety of the moduli space $\mathcal{M}$, it is not the constant family up to translation. From this fact, and the finiteness of the mondromy of $R^{n-1}f^*(g^*\mathcal{L}_\chi)$ for all torsion characters χ, we will derive a contradiction.

Before proceeding, we consider the case where $Y_{\overline{\eta}}$ is translation-invariant by a nonzero element of A, for $\overline{\eta}$ the generic point of X. It follows that the whole family is invariant under the same element. this case, we consider the subgroup of all such elements and quotient A by it. The family Y is then a pullback from a family Y' of hypersurfaces in this quotient A' of A, and the pushforward from Y' of $g^{'*}\mathcal{L}_{\chi'}$ is a summand of the pushforward from Y of $g^*\mathcal{L}_\chi$, where χ is the composition of χ' with the map $A \to A'$, so our finite monodromy assumption remains for Y'. Hence we may assume that $Y_{\overline{\eta}}$ is not translation-invariant by a nonzero element of A.

Let G be the Tannakian group of $Y_{\overline{\eta}}$ and let G^* be the commutator subgroup of the identity component of G. By Lemmas 4.4 and 4.6, the G^* is a simple algebraic group acting by an irreducible representation. We have thus verified all the assumptions of Theorem 5.6. It follows that for all χ outside some finite set of proper subtori of the dual torus to $\pi_1^{et}(A)$, which necessarily includes at least one torsion character, the geometric monodromy group or $R^{n-1}f_*(g^*\mathcal{L}_\chi)$ contains G^*, contradicting our assumption that it is finite, as desired. $\square$

6. Hodge–Deligne systems

The goal of the next few sections is to prove Theorem 8.21, which is analogous to Lemma 4.2, Prop. 5.3, and Thm. 10.1 in [42]. Roughly, the theorem says that, if a smooth variety X over $\mathbb{Q}$ admits a Hodge–Deligne system that has big monodromy and satisfies two numerical conditions, then the integral points of X are not Zariski dense. We follow the same strategy as [42], but we'll need to work in greater generality. First, [42] works only with the primitive cohomology of a family of varieties, but we'll need to work with the cohomology with coefficients. Second, we'll need to work with Galois representations valued in a disconnected reductive group. Finally, we are unable to precisely identify the Zariski closure of the image of monodromy; we only know that it is a c-balanced subgroup of $\mathbf{G}$.

We'll begin by defining the notion of "Hodge–Deligne system", which will figure in our statement of Theorem 8.21. Let X be a variety over a number field K (which will eventually be taken to be $\mathbb{Q}$). A smooth, projective family of varieties over X gives rise to various cohomology objects on X. The argument of [42] relies on the interplay among several of these objects: a complex period map, a p-adic period map, and a family of p-adic global Galois representations on X. Deligne has called the collection of these cohomology objects a "system of realizations" for a motive [15]; our notion of "Hodge–Deligne system" will be closely related to Deligne's systems of realizations.

Definition 6.1. Let k be an integer, and q a prime power. A *rational q-Weil number* of weight k is an algebraic number, all of whose conjugates have complex absolute value $q^{k/2}$. [1]

An *integral q-Weil number* is a rational q-Weil number that is an algebraic integer.

When ℓ is a prime of $\mathcal{O}_K$, we write q_ℓ for the cardinality of the residue field at ℓ.

Definition 6.2. Suppose given a number field K with a chosen embedding $\iota_{\mathbb{C}}\colon K \to \mathbb{C}$. Let X be a smooth variety over K. Let S be a finite set of primes of K, and let $\mathcal{X}$ be a smooth model of X over $\mathcal{O}_K\left[\frac{1}{S}\right]$. Let p be a prime of $\mathbb{Q}$ not lying below any place of S and let v be a place of K lying over p.

A *Hodge–Deligne system* [2] on $\mathcal{X}$ at v consists of the following structures:

- A singular local system V_{Sing} of $\mathbb{Q}$-vector spaces on $X_{\mathbb{C}}$.
- An étale local system V_{et} of $\mathbb{Q}_p$-vector spaces on $\mathcal{X}_{\text{et}} \times_{\operatorname{Spec}\mathbb{Z}} \operatorname{Spec}\mathbb{Z}[1/p]$.
- A vector bundle V_{dR} on X, an integrable connection ∇ on V_{dR}, and a descending filtration $\operatorname{Fil}^i \mathsf{V}_{dR}$ of V_{dR} by subbundles
$$\mathsf{V}_{dR} = \operatorname{Fil}^{-M} \mathsf{V}_{dR} \supseteq \operatorname{Fil}^{-M+1} \mathsf{V}_{dR} \supseteq \cdots \supseteq \operatorname{Fil}^{M} \mathsf{V}_{dR} = 0$$
(not necessarily ∇-stable), each of which is locally a direct summand of V_{dR}.
- A filtered F-isocrystal V_{cris} on X_{K_v} (see, for example, [59, end of §3.1]),

with the following isomorphisms:

[1] It is important that we allow Weil numbers that are not algebraic integers, since we are going to work in a Tannakian setting.

[2] The name is meant to evoke variations of Hodge structure and Deligne's systems of realizations.

(1) An isomorphism on $X_{\mathbb{C},an}$ between $\mathsf{V}_{Sing} \otimes_{\mathbb{Q}} \mathbb{Q}_p$ and the pullback of V_{et} to $X_{\mathbb{C},an}$. [3]
(2) An isomorphism on $X_{\mathbb{C},an}$ between $\mathsf{V}_{Sing} \otimes_{\mathbb{Q}} O_{X_{\mathbb{C},an}}$ and $\mathsf{V}_{dR} \otimes_{\mathcal{O}_X} \mathcal{O}_{X_{\mathbb{C},an}}$.
(3) An isomorphism on an open neighborhood of the rigid analytic generic fiber of X_{K_v} between the underlying vector bundle to V_{cris} and the pullback of V_{dR}.
(4) An isomorphism on $X_{K_v,proet}$ between the $\mathcal{O}\mathbb{B}_{cris}$-modules $\mathsf{V}_{cris} \otimes_{\mathcal{O}_{X_{K_v}}} \mathcal{O}\mathbb{B}_{cris}$ and $\mathsf{V}_{et} \otimes_{\mathbb{Q}_p} \mathcal{O}\mathbb{B}_{cris}$.

and an increasing filtration W_i of all four objects, compatible with all the isomorphisms, such that all this data satisfies the axioms:

- $\mathrm{Fil}^i \mathsf{V}_{dR}$ and ∇ satisfy Griffiths transversality.
- The connection ∇ is induced under the isomorphism (2) by the trivial connection on $\mathcal{O}_X$.
- For each point of $X_{\mathbb{C}}$, the ith associated graded under W_i of the stalk of $(\mathsf{V}_{Sing}, \mathsf{V}_{DR} \otimes_K \mathbb{C}, \mathrm{Fil}^i, (2))$ at that point is a pure Hodge structure of weight i.
- The ith associated graded under W_i of V_{et} is pure of weight i, i.e. for each closed point x of $\mathcal{X} \times_{\mathrm{Spec}\,\mathbb{Z}} \mathrm{Spec}\,\mathbb{Z}[1/p]$ with residue field κ_x, the eigenvalues of Frob_{κ_x} on the ith associated graded of $\mathsf{V}_{et,x}$ are $|\kappa_x|$-Weil numbers of weight i.
- The ith associated graded of V_{cris} under W_i is pure of weight i, i.e. for each closed point x of $\mathcal{X}$ lying over p with residue field κ_x, the eigenvalues of Frobenius on the ith associated graded of $\mathsf{V}_{cris,x}$ are $|\kappa_x|$-Weil numbers of weight i.
- The connection ∇ has regular singularities in a smooth simple normal crossings compactification of X_K.
- The isomorphism (3) is compatible with the connection.
- The isomorphism (4) is compatible with connection, filtration, and Frobenius.

We note that the isomorphism (2) and the first three axioms make up the definition of a variation of Hodge structures. The isomorphism (4) and the last axiom make up the definition of a crystalline local system [59, §1]. (Faltings calls these objects "dual-crystalline sheaves" [19, Theorem 2.6], at least in the situation where the Hodge–Tate weights are bounded between 0 and $p-2$.)

We say a Hodge-Deligne system is *pure* of weight w if W_{w-1} vanishes and W_w is the whole system.

The *rank* of a Hodge–Deligne system V is the rank of the local system V_{sing} of $\mathbb{Q}$-vector spaces. By the various isomorphisms, this is equal to the ranks of V_{et}, V_{dR}, and V_{cris}.

We will also need to work with polarized and integral variations of Hodge structure, and for that we need the following slight modifications of the notion of Hodge–Deligne system.

Definition 6.3. Let K, X, S, $\mathcal{X}$, v be as above. An *integral Hodge–Deligne system* on X consists of a Hodge-Deligne system on X together with an integral structure on V_{sing} (i.e. a singular local system V_{int} of free $\mathbb{Z}$-modules on $X_{\mathbb{C}}$ together with an isomorphism $\mathsf{V}_{int} \otimes_{\mathbb{Z}} \mathbb{Q} \cong \mathsf{V}_{sing}$.)

[3] We do not use the étale-singular comparison; we could have left it out.

Definition 6.4. Let K, X, S, $\mathcal{X}$, v be as above. A *polarized Hodge-Deligne system* on X consists of a Hodge-Deligne system on X, pure of some weight, together with a polarization of the variation of Hodge structures $(\mathsf{V}_{sing}, \mathsf{V}_{dR}, (2))$ (i.e. a morphism of local systems $\mathsf{V}_{sing} \otimes \mathsf{V}_{sing} \to \mathbb{Q}$ which restricted to the stalk at any point of $X(\mathbb{C})$ defines a polarization of the pure Hodge structure at that point.)

Definition 6.5. We say that a Hodge–Deligne system V *has integral Frobenius eigenvalues* if the Weil numbers appearing as eigenvalues of Frobenius on $\mathsf{V}_{et,x}$ and $\mathsf{V}_{cris,x}$ are integral, for all closed points $x \in \mathcal{X} \times_{\operatorname{Spec} \mathbb{Z}} \operatorname{Spec} \mathbb{Z}[1/p]$ and all closed points x of $\mathcal{X}$ lying over p, respectively.

Definition 6.6. The *differential Galois group* of a Hodge–Deligne system V is the differential Galois group of the underlying vector bundle with connection V_{dR}. (For the definition of differential Galois group, see [49, §§1.2 – 1.4]. A vector bundle with connection gives rise to a linear differential equation; by the differential Galois group of the vector bundle with connection, we mean the differential Galois group of a Picard–Vessiot ring of the corresponding differential equation.)

The differential Galois group is the Zariski closure of the monodromy group of the variation of Hodge structure V_H; this follows from the Riemann–Hilbert correspondence, and the fact that the period map has regular singularities along a smooth normal crossings compactification.

Remark 6.7. Let k_v be the residue field of K at v. The "Frobenius automorphism" of K_v is the element of $\mathrm{Gal}_{K_v/\mathbb{Q}_p}$ that acts as the p-th power map on k_v.

A filtered F-isocrystal V_{cris} gives, for every $\overline{x} \in X(k_v)$, a pair

$$(V_{\overline{x}}, \phi_{\overline{x}}),$$

where $V_{\overline{x}}$ is a K_v-vector space, and $\phi_{\overline{x}}$ is an endomorphism of $V_{\overline{x}}$, semilinear over Frobenius. Furthermore, for every $x \in X(\mathcal{O}_{K_v})$ belonging to the residue class of $\overline{x}$, the object V_{cris} defines a filtration on $V_{\overline{x}}$, with an isomorphism to the filtered VS $\mathsf{V}_{dR} \otimes K_v$. We'll call the resulting data

$$\mathsf{V}_{cris,x} = (V_{cris,x}, \phi_{cris,x}, F_{cris,x}).$$

Example 6.8. *(The trivial Hodge–Deligne system.)*

Let K and E be number fields. Take $X = \operatorname{Spec} K$, and define the trivial Hodge–Deligne system O_E on $\operatorname{Spec} K$ *by:*

- $\mathsf{O}_{E,sing} = E$.
- $\mathsf{O}_{E,et} = E \otimes_{\mathbb{Q}} \mathbb{Q}_p$, *with the trivial Galois action.*
- $\mathsf{O}_{E,dR} = E \otimes_{\mathbb{Q}} K$, *with trivial connection and filtration (i.e.* $\mathrm{Fil}^1 \mathsf{O}_{E,dR} = \mathsf{O}_{E,dR}$, *and* $\mathrm{Fil}^0 \mathsf{O}_{E,dR} = 0$*).*
- $\mathsf{O}_{E,cris}$ *is determined by* V_{dR} *and the requirement that Frobenius act on* $E \otimes_{\mathbb{Q}_p} K_v$ *through the trivial action on E and the Frobenius automorphism of K_v.*
- *The weight filtration is such that O_E is concentrated in weight zero.*

When X is an arbitrary smooth K-variety, we define the system O_E on X by pullback from $\operatorname{Spec} K$.

In general, Hodge–Deligne systems will come from families of varieties by taking cohomology.

Example 6.9. *(Pushforward of Hodge–Deligne systems.)*

Let X be a variety over a number field K, and let $\pi\colon Y \to X$ be a smooth, projective family of relative dimension n. Let V be a Hodge–Deligne system on Y, and choose some k with $0 \le k \le n$. We define a Hodge–Deligne system $\mathsf{W} = \mathsf{R}^k\pi_*(\mathsf{V})$ on X, as follows.

- *Spread $Y \to X$ out to a smooth projective family $\pi : \mathcal{Y} \to \mathcal{X}$ over $\mathcal{O}_{K,S}$, for some finite set S of places of K. Call a prime p of $\mathbb{Q}$ good if no element of S lies over p.*
- *Fix a good prime p and a place v lying over p.*
- *Take $\mathsf{W}_{sing} = R^k\pi_*\mathsf{V}_{sing}$, with the pushforward taken in the analytic topology on X and Y. (This is again a local system, by Ehresmann's theorem.)*
- *Take $\mathsf{W}_{et} = R^k\pi_*\mathsf{V}_{et}$, with the pushforward taken in the étale topology. (See* [24] *for an introduction to the étale topology.)*
- *Take W_{dR} to be the relative de Rham cohomology of V_{dR} over X, i.e. the pushforward of V_{dR} as a D-module, with its Hodge filtration Fil^i. This is a filtered vector bundle by Hodge theory.*
- *Take W_{cris} to be the v-adic crystalline cohomology of Y. This is a filtered F-isocrystal by p-adic Hodge theory (*[55, Thm. 8.8]*). (See* [9] *or* [10] *for the construction of crystalline cohomology, and* [46, §1] *for its structure as filtered F-isocrystal.)*
- *The isomorphisms (1), (2), (3), (4) follow from Artin's comparison theorem, Hodge theory, the de Rham-crystalline comparison, and relative p-adic Hodge theory (*[55, Thm. 8.8], [59, Thm. 5.5]*) respectively.*
- *The filtration W_i is induced from the filtration of V after shifting by k. In particular, if V is pure of weight w, then W is pure of weight $w + k$.*

Definition 6.10. If V and W are two Hodge–Deligne systems on X, a morphism from V to W consists of:

- A map of analytic local systems $\mathsf{V}_{sing} \to \mathsf{W}_{sing}$,
- A map of étale local systems $\mathsf{V}_{et} \to \mathsf{W}_{et}$,
- A map of vector bundles $\mathsf{V}_{dR} \to \mathsf{W}_{dR}$, flat with respect to the connections on V_{dR} and W_{dR}, and respecting the filtrations $\mathrm{Fil}^* \mathsf{V}_{dR}$ and $\mathrm{Fil}^* \mathsf{W}_{dR}$, and
- A map of filtered F-isocrystals $\mathsf{V}_{cris} \to \mathsf{W}_{cris}$,

compatible with all the comparison isomorphisms (1), (2), (3), (4).

Lemma 6.11. *The systems of compatible periods on X at v form a Tannakian category with fiber functor given by $\mathsf{V}_{sing,x}$ for some $x \in X(\mathbb{C})$.*

In this Tannakian category, the tensor product of two systems will be defined by separately tensoring the individual objects $\mathsf{V}_{sing}, \mathsf{V}_{et}, \mathsf{V}_{dR}$, and V_{cris}, and similarly for the dual of a system.

Proof. Let us first check that they form an abelian category. Kernels and cokernels will be taken separately on the four components f_{sing}, f_{et}, f_{dR}, and f_{cris}. The verification of most of the axioms is routine; we'll explain only two steps which are not immediate.

The first is existence of a cokernel for f_{dR}. Let $f\colon \mathsf{V} \to \mathsf{W}$ be a morphism of Hodge–Deligne systems. In general the cokernel of a morphism of vector bundles need not be a vector bundle. But the cokernel of $f_{dR}\colon \mathsf{V}_{dR} \to \mathsf{W}_{dR}$ must be a vector bundle because f_{dR} is a flat map of vector bundles with connection.

The second is the equality of images and coimages. A priori, the image of f_{dR} has two possibly different filtrations, the image filtration and the coimage filtration. These filtrations agree because variations of Hodge structure form an abelian category.

So the category of Hodge–Deligne systems is abelian.

Furthermore, the category of Hodge-Deligne systems is manifestly $\mathbb{Q}$-linear.

To check it is a $\mathbb{Q}$-linear symmetric monoidal category, we use the symmetric monoidal structure on the categories of singular local systems, étale local systems, filtered vector bundles with flat connections, and F-isocrystals separately. To define the tensor product of two objects we individually tensor the $\mathsf{V}_{sing}, \mathsf{V}_{et}, \mathsf{V}_{dR}, \mathsf{V}_{cris}$ and then tensor the ismorphisms (1), (2), (3), (4). This requires that the various pullback functors we use in defining these isomorphisms are compatible with tensor product, which is standard. We also need to check that the tensor product satisfies the axioms, which is clear in all cases. In addition to defining a product, the symmetric monoidal structure consists of isomorphisms and commutative diagrams. The fact that these isomorphisms are compatible with (1), (2), (3), (4) follows from their naturality and the fact that the pullback functors are symmetric monoidal.

The unit object is given by $\mathsf{O}_{\mathbb{Q}}$ from Example 6.8.

For rigidity, we use a similar argument. Each individual category has a dual object functor, unit and counit natural transformations, and commutative diagrams satisfied by these isomorphisms. Because the pullback functors are compatible with this data, we can define the dual of a Hodge-Deligne system, and the unit and counit are compatible with the isomorphisms (1), (2), (3), (4) and thus form a valid morphism of Hodge-Deligne systems. The commutative diagrams follow from the commutative diagrams on the four individual objects.

To check that $\mathsf{V}_{sing,x}$ is an exact tensor functor to $\mathbb{Q}$-vector spaces , it suffices to check that V_{sing} is an exact tensor functor to $\mathbb{Q}$-local systems. The fact that it is a tensor functor follows from how we defined tensor product; the fact that it is exact follows from how we constructed kernels and cokernels. □

If $f\colon X' \to X$ is a map, then for any system V on X, we can define the *pullback* $f^*\mathsf{V}$ on X' by pulling back the four components separately. When the map f is clear, we'll sometimes write $\mathsf{V}|_{X'}$ instead of $f^*\mathsf{V}$.

Definition 6.12. Let X be a K-scheme. A *constant* Hodge–Deligne system on X is a system of the form $f^*\mathsf{V}$, where V is a Hodge–Deligne system on $\operatorname{Spec} K$.

Constant Hodge-Deligne systems will be much more general than most notions of motives. For instance nothing prevents us from combining the étale and crystalline cohomology of one variety with the Hodge structure of a different variety, as long as they have the same Hodge numbers. Despite this, the notion of Hodge-Deligne system is strong enough for the arguments that we will make.

6.1. **H^0-algebras.** In order to make the arguments of [42] work, we need bounds on the centralizer of Frobenius.

The paper [42] works with Hodge–Deligne systems of the form $\mathsf{R}^k\pi_*(\mathsf{O}_{\mathbb{Q}})$, for $\pi\colon Y \to X$ a family with geometrically disconnected fibers. In this context, the zeroth cohomology $H^0(Y_x)$ has nontrivial Galois structure. The action of $H^0(Y_x)$ on $H^k(Y_x)$ gives rise to the bounds we need on the Frobenius centralizer, by means of the semilinearity of Frobenius.

For this work involving abelian varieties, we'll have a group G acting on the fibers of $Y \to X$; this G will give a bound on the Frobenius centralizer in an analogous way.

The notion of module over an H^0-algebra generalizes both pictures. Loosely speaking, an H^0-algebra is a weight-zero algebra object in the category of Hodge–Deligne systems. Motives over extensions of K (Example 6.20), motives with coefficients in a number field E (Examples 6.14 and 6.22), and motives with an action of a finite algebraic group (Examples 6.19 and 6.22) will all turn out to be modules over various H^0-algebras.

Definition 6.13. A commutative H^0-algebra is a Hodge–Deligne system E on a variety X over a number field K, equipped with maps

$$e \colon \mathsf{O}_{\mathbb{Q}} \to \mathsf{E}$$

and

$$m \colon \mathsf{E} \otimes \mathsf{E} \to \mathsf{E}$$

satisfying the following properties.

- E is pure of weight 0.
- The filtration on E_{dR} is trivial: $\mathrm{Fil}^0 \mathsf{E}_{dR} = \mathsf{E}_{dR}$ and $\mathrm{Fil}^1 \mathsf{E}_{dR} = 0$.
- The maps e and m make E into a commutative algebra object.

When we say "H^0-algebra", we will mean "commutative H^0-algebra".

Example 6.14. *(Trivial Hodge–Deligne system O_E.)*

Let K and E be number fields. The trivial Hodge–Deligne system O_E of Example 6.8 has an algebra structure coming functorially from the algebra structure on E.

Definition 6.15. If $X = \operatorname{Spec} K$, we say that E is *étale* if E_{sing} is an étale $\mathbb{Q}$-algebra.

Example 6.16. *(Étale H^0-algebras over a field.)*

Let $X = \operatorname{Spec} K$ be the spectrum of a number field. In this setting we can give a concrete description of étale H^0-algebras E over $\operatorname{Spec} K$.

The singular realization $E = \mathsf{E}_{sing}$ has the structure of $\mathbb{Q}$-algebra, which we assume is étale; E_{dR} is determined by

$$\mathsf{E}_{dR} = E \otimes_{\mathbb{Q}} K$$

with trivial filtration.

The étale realization E_{et} is the $\mathbb{Q}_p$-algebra

$$E \otimes_{\mathbb{Q}} \mathbb{Q}_p,$$

equipped with a continuous action of the Galois group Gal_K. By assumption, E is an étale $\mathbb{Q}$-algebra, so $\mathrm{Aut}(E \otimes_{\mathbb{Q}} \mathbb{Q}_p)$ is a finite group. The action of Gal_K descends to the maximal quotient $\mathrm{Gal}_{K,S}$ of Gal_K unramified outside some finite set S of places of K, which (by p-adic Hodge theory) must include the place v of K.

Finally we turn to E_{cris}. The structure of K_v-algebra is given by an isomorphism

$$\mathsf{E}_{cris} \cong E \otimes_{\mathbb{Q}_p} K_v.$$

The filtration is trivial, and the Frobenius (which we will notate σ) is the endomorphism $\sigma_1 \otimes \sigma_2$ of $E \otimes_{\mathbb{Q}_p} K_v$, where σ_1 gives the action of $\mathrm{Frob}_v \in G_K$ on E, and σ_2 is the endomorphism of K_v that acts as the p-th power map on residue fields. The

Frobenius $\mathrm{Frob}_v \in G_K$ *is only well-defined up to conjugacy, but that is enough to determine* E_{cris} *up to isomorphism.*

Example 6.17. *(H^0 of a family.)*

If $\pi \colon Y \to X$ *is a proper map, then the degree-zero cohomology of* Y*, equipped with the cup product, gives an* H^0*-algebra on* X*.*

The underlying Hodge-Deligne system is constructed as in Example 6.9 as $\mathsf{R}^0\pi_*(\mathsf{O}_{\mathbb{Q}})$*. The Hodge filtration is trivial because the Hodge filtration on* H^0_{dR} *of any smooth scheme is trivial. The map* $\mathsf{O}_{\mathbb{Q}} \to \mathsf{E}$ *is the unit of the adjunction between* π_* *and* π^**, and the map* $\mathsf{E} \otimes \mathsf{E} \to \mathsf{E}$ *is given by cup product.*

Example 6.18. *(Group algebra of a finite group.)*

For a finite group G*, define the* H^0*-algebra* $\mathsf{O}[G]$ *as follows.*

- $\mathsf{O}[G]_{sing} = \underline{\mathbb{Q}[G]}$.
- $\mathsf{O}[G]_{et} = \underline{\mathbb{Q}_p[G]}$.
- $\mathsf{O}[G]_{dR} = \overline{K[G]}$ *is the trivial vector bundle, with trivial connection.*
- *The filtration on* $\mathsf{O}[G]_{dR}$ *is* $\mathsf{O}[G]_{dR} = \mathrm{Fil}^0 \supseteq \mathrm{Fil}^1 = 0$.
- *The filtered* F*-isocrystal is the constant vector bundle* $\mathcal{O}[G]$ *with trivial connection. Its fiber at any point is the group algebra* $K_v[G]$*, with Frobenius action coming from the Frobenius on* K_v*.*

Note that, for a group G, it is natural to view the group algebra $\mathbb{Q}[G]$ as a space of measures on G, and thus dual to the space of functions on G. Then the multiplication in the group algebra corresponds to convolution of measures. This suggests the right way to generalize the group algebra to group schemes, as the dual to their ring of functions. (Of course, the trace map makes their ring of functions self-dual.)

Example 6.19. *(Group algebra of a finite group scheme over* K*.)*

Let G *be a finite étale group scheme over a number field* K*. The group operation* $G \times G \to G$ *defines a Hopf algebra comultiplication* $\Gamma(G, \mathcal{O}_G) \to \Gamma(G, \mathcal{O}_G) \otimes \Gamma(G, \mathcal{O}_G)$*. The dual map* $(\Gamma(G,\mathcal{O}_G))^\vee \otimes (\Gamma(G,\mathcal{O}_G))^\vee \to (\Gamma(G,\mathcal{O}_G))^\vee$ *gives* $(\Gamma(G,\mathcal{O}_G))^\vee$ *the structure of an algebra.*

Denoting by $\pi \colon G \to \operatorname{Spec} K$ *the structure map, we define* $\mathsf{E} = (\pi_* \mathsf{O}_{\mathbb{Q}}|_G)^\vee$*. A concrete description is as follows.*

- $\mathsf{E}[G]_{sing} = \mathbb{Q}[G(\mathbb{C})]$ *with the usual algebra structure.*
- $\mathsf{E}[G]_{et} = \mathbb{Q}_p[G(\overline{K})]$ *with its natural Galois action and structure of* $\mathbb{Q}_p$*-algebra.*
- $\mathsf{E}[G]_{dR} = (\Gamma(G,\mathcal{O}_G))^\vee$ *with is natural algebra structure.*
- $\mathsf{E}[G]_{cris} = \left(\Gamma(G_{K_v}, \mathcal{O}_{G_{K_v}})\right)^\vee$ *with its natural algebra structure and Frobenius coming from the Galois action on* G*.*

After passing to an extension of K *over which* G *splits, an element* $a \in G$ *gives an element of each of* $\mathsf{E}[G]_{sing}$*,* $\mathsf{E}[G]_{et}$*,* $\mathsf{E}[G]_{dR}$*, and* $\mathsf{E}[G]_{cris}$*, and each of these four realizations is generated (as a vector space over the appropriate field) by* G*.*

Example 6.20. *Let* L/K *be an extension of fields, and* $\pi \colon \operatorname{Spec} L \to \operatorname{Spec} K$ *the corresponding map of schemes. If* E *is an* H^0*-algebra on* L*, then* $\pi_*\mathsf{E}$ *is an* H^0*-algebra on* K*.*

6.2. **Modules over an H^0-algebra.**

Definition 6.21. Let E be an H^0-algebra. An E-module is a system of compatible periods V with a map $m_{\mathsf{V}} \colon \mathsf{E} \otimes \mathsf{V} \to \mathsf{V}$, such that the composition

$$\mathsf{V} \cong \mathsf{O}_{\mathbb{Q}} \otimes \mathsf{V} \xrightarrow{e \otimes 1} \mathsf{E} \otimes \mathsf{V} \xrightarrow{m_{\mathsf{V}}} \mathsf{V}$$

is the identity map, and the diagram

$$\begin{tikzcd} \mathsf{E} \otimes \mathsf{E} \otimes \mathsf{V} \arrow[r, "1 \otimes m_{\mathsf{V}}"] \arrow[d, "m \otimes 1"] & \mathsf{E} \otimes \mathsf{V} \arrow[d, "m"] \\ \mathsf{E} \otimes \mathsf{V} \arrow[r, "m_{\mathsf{V}}"] & \mathsf{V} \end{tikzcd} \tag{4}$$

commutes.

Example 6.22. *An O_E-module (Example 6.14) is analogous to a motive with coefficients in E. An $\mathsf{O}[G]$-module (Example 6.19) is analogous to a motive with an action of the group scheme G.*

If V and W are E-modules, then we define the tensor product

$$\mathsf{V} \otimes_{\mathsf{E}} \mathsf{W}$$

as the coequalizer of two maps

$$\mathsf{E} \otimes \mathsf{V} \otimes \mathsf{W} \rightrightarrows \mathsf{V} \otimes \mathsf{W},$$

the first of which is induced from $\mathsf{E} \otimes \mathsf{V} \to \mathsf{V}$, and the second from $\mathsf{E} \otimes \mathsf{W} \to \mathsf{W}$. (See [13, §2.3].)

Definition 6.23. Let E be a finite étale algebra over a field E_0. We say an E-module V is *equidimensional* if it is free of finite rank.

Equivalently, writing E as a product of fields E_i, we say that V is equidimensional if

$$\dim_{E_i}(V \otimes_E E_i)$$

is independent of i.

In this case, we call that dimension the *rank* of V.

Definition 6.24. Suppose E is a constant étale H^0-algebra, and V is an E-module. We say V is *equidimensional* if the stalks of V_{sing} are equidimensional modules over the stalks of E_{sing}. In this case, the ranks of the stalks as E_{sing}-modules are constant; we call that rank the rank of V.

Lemma 6.25. *If V is an equidimensional E-module of rank N on X, then the following statements hold.*

- *The stalks of V_{et} are equidimensional modules of rank N over the stalks of E_{et}.*
- *The stalks of V_{dR} are equidimensional modules of rank N over the stalks of E_{dR}.*
- *The stalks of V_{cris} are equidimensional modules of rank N over the stalks of E_{cris}.*

Proof. Follows from the “comparison isomorphisms” in the definition of Hodge–Deligne system. ∎

Definition 6.26. Suppose V is an equidimensional E-module of rank N. Then we say that V is an E-*module with* GL_N-*structure*.

Suppose additionally that there exists a Hodge-Deligne system L of rank 1, with a nondegenerate bilinear pairing

$$\mathsf{V} \otimes_{\mathsf{E}} \mathsf{V} \to \mathsf{L}.$$

If the pairing is alternating, we say that V has GSp_N-*structure*; if symmetric, we say that V has GO_N-*structure*.

Below (Definition 6.37) we will define a notion of "object with G-structure" in a Tannakian category. We warn the reader that Definition 6.26 is not consistent with Definition 6.37: Definition 6.26 applies to E-modules, for E an H^0-algebra, while Definition 6.37 applies to objects of an E-linear tensor category, with E a field.

In the next few sections, we'll study the structure of E-modules with GL, GSp, or GO-structure, where E is a constant, étale H^0-algebra.

6.3. **Local systems on an abelian variety: construction of a Hodge–Deligne system.** Let A be an abelian variety over a number field K, X an arbitrary smooth variety over $\mathbb{Q}$, and X_K its base change to K. Let

$$Y \subseteq X \times_{\mathbb{Q}} A = X_K \times_K A$$

be a subscheme, smooth, proper and flat over X_K. Let $f\colon Y \to X$ and $g\colon Y \to A$ be the projections. Fix a positive integer r. Let L be a field, containing K, Galois over $\mathbb{Q}$, and over which $A[r]$ splits. Let χ be an order-r character of $\pi_1^{et}(A)$, where r is prime to p; let $\mathcal{L}_\chi$ be the corresponding character sheaf on A. (It is a $\mathbb{Q}_p[\mu_r]$-local system on the étale site of A_L.) Let $k = n - 1 = \dim Y$. We want to create a Hodge–Deligne system on X whole base change to L has $R^k f_* g^* \mathcal{L}_\chi$ as a direct summand.

The tensor product $K \otimes_{\mathbb{Q}} L$ splits as a direct sum

$$K \otimes_{\mathbb{Q}} L \cong \bigoplus_{\iota} L^{(\iota)},$$

indexed by the $[K\colon \mathbb{Q}]$ embeddings of K into L. Here each $L^{(\iota)}$ is an isomorphic copy of L; the superscript (ι) is merely an index. We have the corresponding splitting

$$A \times_{\operatorname{Spec} \mathbb{Q}} \operatorname{Spec} L \cong \coprod_{\iota} A_\iota,$$

where for each ι, we define

$$A_\iota = A \otimes_{\operatorname{Spec} K, \iota} \operatorname{Spec} L.$$

Similarly, define

$$Y_\iota = Y \otimes_{\operatorname{Spec} K, \iota} \operatorname{Spec} L,$$

the base change of Y along ι. Then we have the Cartesian diagram

$$\begin{tikzcd}
\coprod_\iota Y_\iota \arrow[r] \arrow[d] & Y \arrow[d] \\
\coprod_\iota X_L \times_L A_\iota \arrow[r] \arrow[d] & X \times A \arrow[d] \\
X_L \arrow[r] & X.
\end{tikzcd}$$

Let $f_\iota \colon Y_\iota \to X_L$ and $g_\iota \colon Y_\iota \to A_\iota$ be the projections.

Let r be prime to p, and let $\Pi^{K/\mathbb{Q}}(A)[r]$ be the set of all pairs (ι, χ), where $\iota \colon K \to L$ is a $\mathbb{Q}$-linear embedding, and χ is a character of $\pi_1(A_\iota)$ of order dividing r. For fixed ι, the set of characters χ is naturally identified with the set of maps

$$A_\iota[r] \to \mathbb{Q}[\mu_r];$$

thus, $\Pi^{K/\mathbb{Q}}(A)[r]$ has a natural action of

$$\mathrm{Gal}_\mathbb{Q} \times \mathrm{Gal}_{\mathbb{Q}[\mu_r]/\mathbb{Q}},$$

where $\mathrm{Gal}_\mathbb{Q}$ acts on the pairs (ι, χ) via its action on L, and $\mathrm{Gal}_{\mathbb{Q}[\mu_r]/\mathbb{Q}}$ acts on $\mathbb{Q}[\mu_r]$.

Lemma 6.27. *Let χ be character of $\pi_1^{et}(A)$ of some finite order r. Then there exists an $\mathsf{O}_{\mathbb{Q}[\mu_r]}$-module L_χ on $A|_L$ such that $(\mathsf{L}_\chi)_{et}$ is the character sheaf associated with the character χ, and $(\mathsf{L}_\chi)_{sing}$ is the analytic $\mathbb{Q}[\mu_r]$-local system associated with χ.*

Proof. Descent. □

Lemma 6.28. *(Construction of E_I and V_I.)*

Let I be an orbit of $\mathrm{Gal}_\mathbb{Q} \times \mathrm{Gal}_{\mathbb{Q}[\mu_r]/\mathbb{Q}}$ on $\Pi^{K/\mathbb{Q}}(A)[r]$. There exist an H^0-algebra E_I on K and an E_I-module V_I on X with the following properties.

- *After base change to L and extension of coefficients to $\mathbb{Q}[\mu_r]$, we have the direct sum decomposition*

$$\mathsf{E}_I|_L \otimes_{\mathsf{O}_\mathbb{Q}} \mathsf{O}_{\mathbb{Q}[\mu_r]} \cong \bigoplus_{(\iota,\chi) \in I} \mathsf{O}_{\mathbb{Q}[\mu_r]}^{(\iota,\chi)},$$

  where each $\mathsf{O}_{\mathbb{Q}[\mu_r]}^{(\iota,\chi)}$ is a copy of $\mathsf{O}_{\mathbb{Q}[\mu_r]}$.
- *The Galois representation $\mathsf{E}_{I,et}$ is compatible with the isomorphism*

$$\mathsf{E}_{I,et} \otimes_\mathbb{Q} \mathbb{Q}[\mu_r] \cong \bigoplus_{(\iota,\chi) \in I} \mathbb{Q}_p \otimes_\mathbb{Q} \mathbb{Q}[\mu_r],$$

  where Gal_K acts on the right-hand side by permutation of the characters χ.

  Similarly, the Frobenius endomorphism of $\mathsf{E}_{I,cris}$ is compatible with the isomorphism

$$\mathsf{E}_{I,cris} \otimes_\mathbb{Q} \mathbb{Q}[\mu_r] \cong \bigoplus_{(\iota,\chi) \in I} \mathbb{Q}_p \otimes_\mathbb{Q} \mathbb{Q}[\mu_r],$$

  where Frobenius acts on the right-hand side by permuting the pairs (ι, χ), via the $\mathrm{Gal}_\mathbb{Q}$ action, with trivial action on $\mathbb{Q}[\mu_r]$.

- *After base change to L and extension of coefficients to $\mathbb{Q}[\mu_r]$, the module V_I decomposes as the direct sum*
$$\mathsf{V}_I|_L \otimes_{\mathsf{O}_{\mathbb{Q}}} \mathsf{O}_{\mathbb{Q}[\mu_r]} = \bigoplus_{(\iota,\chi)\in I} R^k f_{\iota *} g_\iota^* \mathsf{L}_\chi.$$
Furthermore, this decomposition is compatible with the decomposition of
$$\mathsf{E}_I|_L \otimes_{\mathsf{O}_{\mathbb{Q}}} \mathsf{O}_{\mathbb{Q}[\mu_r]}$$
into fields.
- V_I *can be made into a polarized, integral Hodge–Deligne system.*

Proof. We have the map
$$id_X \times [r] \colon X \times_{\mathbb{Q}} A \to X \times_{\mathbb{Q}} A$$
where $[r] \colon A \to A$ is multiplication by r; let
$$Y_r = (id_X \times [r])^{-1}(Y).$$

$$\begin{tikzcd}
Y_r \arrow[r] \arrow[d] \arrow[dd, bend right, "h"'] & Y \arrow[d] \arrow[dr, bend left, "g"] \arrow[dd, bend left, "f"] & \\
X \times A \arrow[r, "id_X \times [r]"] \arrow[d] & X \times A \arrow[r] \arrow[d] & A \\
X \arrow[r, "="] & X &
\end{tikzcd}$$

We have an isomorphism of Hodge–Deligne systems
$$R^k h_* \mathsf{O}_{\mathbb{Q}} \cong R^k f_* g^* [r]_* \mathsf{O}_{\mathbb{Q}}.$$
Furthermore, on $A \times_{\operatorname{Spec}\mathbb{Q}} \operatorname{Spec} L$, we have the direct sum decomposition of Hodge–Deligne systems
$$[r]_* \mathsf{O}_{\mathbb{Q}}|_{A\times_{\mathbb{Q}} L} \otimes_{\mathsf{O}_{\mathbb{Q}}} \mathsf{O}_{\mathbb{Q}[\mu_r]} \cong \bigoplus_{(\iota,\chi)\in \Pi^{K/\mathbb{Q}}(A)[r]} \mathsf{L}_{\iota,\chi},$$
where $\mathsf{L}_{\iota,\chi}$ is the system L_χ on A_ι, and the trivial system on all other components of $A \times_{\operatorname{Spec}\mathbb{Q}} \operatorname{Spec} L$. This gives a decomposition on X_L
$$(R^k h_* \mathsf{O}_{\mathbb{Q}})|_L \otimes \mathsf{O}_{\mathbb{Q}}[\mu_r] = \bigoplus_{(\iota,\chi)} R^k f_* g^* \mathsf{L}_{\iota,\chi} = \bigoplus_{(\iota,\chi)} R^k f_{\iota *} g_\iota^* \mathsf{L}_\chi.$$

Let
$$\mathsf{V} = R^k h_* \mathsf{O}_{\mathbb{Q}} \cong R^k f_* g^* [r]_* \mathsf{O}_{\mathbb{Q}}.$$

Let E be the pullback to X of the Hodge–Deligne system on $\mathbb{Q}$ coming from
$$\mathsf{E} = (\pi_* \mathsf{O}_{\mathbb{Q}}|_{A[r]})^\vee,$$
where $\pi \colon A \to \mathbb{Q}$ is the projection. This E has a structure of H^0-algebra, coming from the structure of group scheme on $A[r]$ (see Examples 6.19 and 6.20). Furthermore, the group action $A[r] \times Y_r \to Y_r$ makes V into a module over E.

After base change from $\mathbb{Q}$ to L and extension of coefficients, we have a decomposition of H^0-algebras on X
$$\mathsf{E}|_L \otimes_{\mathbb{Q}} \mathsf{O}_{\mathbb{Q}[\mu_r]} = \bigoplus_{(\iota,\chi)\in \Pi^{K/\mathbb{Q}}(A)[r]} \mathsf{O}_{\mathbb{Q}[\mu_r]}^{(\iota,\chi)}.$$

The set $\Pi^{K/\mathbb{Q}}(A)[r]$ has commuting actions of $\mathrm{Gal}_{\mathbb{Q}}$ and $\mathrm{Gal}_{\mathbb{Q}^{cyc}/\mathbb{Q}}$, the former coming from the base field $\mathbb{Q}$, and the latter from the coefficient field $\mathbb{Q}^{cyc}$. For each orbit I of $\mathrm{Gal}_{\mathbb{Q}} \times \mathrm{Gal}_{\mathbb{Q}^{cyc}/\mathbb{Q}}$, there exists an H^0-algebra E_I over $\mathbb{Q}$ such that

$$\mathsf{E}_I|_L \otimes E_p = \bigoplus_{(\iota,\chi)\in I} \mathsf{O}^{(\iota,\chi)}_{\mathbb{Q}[\mu_r]}.$$

Over $\mathbb{Q}$ and with $\mathbb{Q}$-coefficients, E splits as the direct sum of the algebras E_I.

The E-module V also splits as the direct sum of the objects $\mathsf{V}_I = \mathsf{V} \otimes_{\mathsf{E}} \mathsf{E}_I$, and after base change and extension of coefficients we have

$$\mathsf{V}_I|_L \otimes_{\mathsf{O}_{\mathbb{Q}}} \mathsf{O}_{\mathbb{Q}[\mu_r]} = \bigoplus_{(\iota,\chi)\in I} R^k f_{\iota *} g_\iota^* \mathsf{L}_\chi.$$

Finally we have to explain the polarization and integral structure on $\mathsf{V}_{I,sing}$. On V_{sing} we have a standard polarization and integral structure: the polarization comes from Poincaré duality, and the integral structure is simply the integral structure on the singular cohomology of Y_r. These structures induce a polarization and integral structure on $\mathsf{V}_{I,sing}$, by restricting the polarization and intersecting the integral lattice with $\mathsf{V}_{I,sing}$. □

Lemma 6.29. *Fix notation as in Lemma 6.28. Then V_I is an E_I-module with GL_N-structure, in the sense of Definition 6.26.*

Furthermore, if Y is equal to a translate of $[-1]^*Y$, then V_I has GSp_N-structure if n is even and GO_N-structure if n is odd.

Proof. To prove that V_I has GL_N-structure we only need to check that V_I is equidimensional; this is a consequence of the transitive Galois action on the index set I.

Suppose Y is equal to a translate of $[-1]^*Y$. This equality gives an involution

$$\iota\colon \mathsf{V} \to \mathsf{V}.$$

The cup product pairing and the trace map compose to give a map

$$\langle -,-\rangle\colon \mathsf{V}\otimes\mathsf{V} = R^k h_* \mathsf{O}_{\mathbb{Q}} \otimes R^k h_* \mathsf{O}_{\mathbb{Q}} \to R^{2k} h_* \mathsf{O}_{\mathbb{Q}} \to \mathsf{O}_{\mathbb{Q}}(-k).$$

This pairing does not factor through $\mathsf{V}\otimes_{\mathsf{E}}\mathsf{V}$, which would be equivalent to the identity

$$\langle ev, w\rangle = \langle v, ew\rangle \tag{5}$$

of maps $\mathsf{V}\times\mathsf{E}\times\mathsf{V} \to \mathsf{O}_{\mathbb{Q}}(-k)$ (where v, e, and w are local sections of the sheaves underlying V, E, and V.) Instead, for $a \in A[r]$, the pairing satisfies

$$\langle v, w\rangle = \langle av, aw\rangle.$$

However, when Y is equal to a translate of $[-1]^*Y$, and ι is the involution of V described above, we can form the pairing

$$(v\otimes w) \mapsto \langle v, \iota w\rangle$$

as the composition

$$\mathsf{V}\otimes\mathsf{V} \xrightarrow{1\otimes\iota} \mathsf{V}\otimes\mathsf{V} \xrightarrow{\langle -,-\rangle} R^{2k} f_* g^* \mathsf{O}_{\mathbb{Q}}|_Y.$$

I claim that this pairing does satisfy (5), so that it descends to a pairing

$$\mathsf{V}\otimes_{\mathsf{E}}\mathsf{V} \to R^{2k} f_* g^* \mathsf{O}_{\mathbb{Q}}|_Y.$$

It is enough to check (5) after extending both the base field and the coefficient field, so we can assume E is the group algebra of the finite abelian group $A[r]$, and that E splits as a direct sum over characters of that group. Now (5) follows from

$$\langle av, \iota w\rangle = \langle v, a^{-1}\iota w\rangle = \langle v, \iota(aw)\rangle,$$

where $a \in A[r]$.

We have

$$\langle w, \iota v\rangle = \langle \iota w, \iota^2 v\rangle = \langle \iota w, v\rangle = (-1)^k \langle v, \iota w\rangle$$

so this pairing is symplectic if n is even (and thus k is odd) and symmetric if n is odd (and thus k is even). $\square$

For a fixed $(\mathrm{Gal}_{\mathbb{Q}} \times \mathrm{Gal}_{\mathbb{Q}^{cyc}/\mathbb{Q}})$-orbit I on $\Pi^{K/\mathbb{Q}}(A)[r]$, define the reductive group $\mathbf{H}$ as follows. Let N be the rank of V_I. If Y is equal to a translate of $[-1]^*Y$ and n is even, let $\mathbf{H} = GSp_N$. If Y is equal to a translate of $[-1]^*Y$ and n is odd, let $\mathbf{H} = GO_N$. Otherwise, let $\mathbf{H} = GL_N$. Then by Lemma 6.29, V_I has $\mathbf{H}$-structure. In the following sections we will analyze this structure in some detail.

6.4. **An algebraic group.** Let E_0 be a field – we will mostly be interested in the case $E_0 = \mathbb{Q}_p$ – and let E be a finite étale E_0-algebra. Let $\mathbf{H}$ be a reductive group over E_0, and choose a representation V_{simp} of $\mathbf{H}$. We will assume that $\mathbf{H}$ is one of GL_N, GSp_N, or GO_N, and V_{simp} is the standard representation.

Let $\mathbf{G}^0$ be the Weil restriction $\mathbf{G}^0 = \mathrm{Res}^E_{E_0}\mathbf{H}$, and let $V = E \otimes_{E_0} V_{simp}$. By restriction of scalars, we will view V as an E_0-vector space with actions of E and $\mathbf{G}^0$. Let $\mathbf{G}$ be the normalizer of $\mathbf{G}^0$ in the group of E_0-linear automorphisms of V.

Definition 6.30. Let σ be a E_0-linear automorphism of E. We say that an E_0-linear endomorphism $\phi\colon V \to V$ is *σ-semilinear* (or *semilinear over σ*) if

$$\phi(\lambda v) = \sigma(\lambda)\phi(v)$$

for all $\lambda \in E$ and $v \in V$.

Lemma 6.31. *If $\mathbf{H} = GL_N$, then $\mathbf{G}$ is the algebraic group whose E_0-points correspond to endomorphsims ϕ of V, semilinear over some E_0-linear automorphism σ of E.*

If $\mathbf{H}$ is GSp_N or GO_N, then $\mathbf{G}^0$ preserves, up to scaling, an (alternating or symmetric) E-valued pairing $\langle -, -\rangle$ on V. In this setting $\mathbf{G}$ is the group of endomorphisms ϕ of V, semilinear over some E_0-linear automorphism σ of E, and satisfying

$$\langle \phi(v_1), \phi(v_2)\rangle = C\sigma(\langle v_1, v_2\rangle) \tag{6}$$

for some $C \in E$, for all $v_1, v_2 \in V$.

In particular, we have an exact sequence of groups

$$1 \to \mathbf{G}^0 \to \mathbf{G} \to \mathrm{Aut}_{E_0} E \to 1.$$

Proof. Any element of the normalizer of $\mathbf{G}$ normalizes the center of $\mathbf{G}^0$, which is E, acting by scalar multiplication. The E_0-linear automorphisms of V that normalize the action of E are exactly the semilinear automorphisms.

If $\mathbf{H}$ is GSp_N or GO_N, then for $\phi \in \mathbf{G}$ which is E-semilinear over σ, the bilinear form $\langle \phi(v_1), \phi(v_2)\rangle$ is E-semilinear over σ in both variables and is preserved up to scaling by $\mathbf{G}^0$. Thus $\sigma^{-1}\left(\langle \phi(v_1), \phi(v_2)\rangle\right)$ is E-bilinear and preserved up to scaling by $\mathbf{G}^0$. Hence it is a scalar multiple of $\langle v_1, v_2 \rangle$, which is the unique E-linear $\mathbf{G}^0$-equivariant form. This gives (6).

Conversely, any E-semilinear automorphism which, if $\mathbf{H}$ is GSp_N or GO_N, satisfies (6), manifestly normalizes $\mathbf{G}^0$. $\square$

We need a generalization of [42, Lemma 2.1].

Lemma 6.32. *Suppose $\sigma \in \mathrm{Aut}_{E_0} E$ is such that $E^\sigma = E_0$, and suppose $\phi \in \mathbf{G}$ is semilinear over σ. Then the centralizer $Z_{\mathbf{G}}(\phi)$ satisfies*

$$\dim Z_{\mathbf{G}}(\phi) \leq \dim \mathbf{H}.$$

Proof. The proof goes through exactly as in [42]. By passing to the algebraic closure we may assume that $E = E_0^d$, so $\mathbf{G}^0 = \mathbf{H}^d$. The hypothesis implies that σ acts transitively on the d factors $\mathbf{H}$ of $\mathbf{H}^d$. A calculation shows that the projection $Z_{\mathbf{G}^0}(\phi) \to \mathbf{H}$ onto any single factor $\mathbf{H}$ of $\mathbf{H}^d$ is injective. The result follows because $Z_{\mathbf{G}^0}(\phi)$ has finite index in $Z_{\mathbf{G}}(\phi)$. $\square$

6.5. **Disconnected reductive groups.** We need to apply p-adic Hodge theory in a Tannakian setting: we'll work with Galois representations valued in the disconnected algebraic group $\mathbf{G}$, and the corresponding filtered ϕ-modules. To this end, we need to study filtrations valued in a disconnected algebraic group.

We will use a notion of parabolic subgroup due to Richardson [51] (see also [43] and the forthcoming survey [6]). Let G be an algebraic group over a field of characteristic zero whose identity component G^0 is reductive.

Definition 6.33. Let $f\colon \mathbb{G}_m \to G$ be a map of schemes. We say that $\lim_{t\to 0} f(t)$ exists if f extends to a map $\tilde{f}\colon \mathbb{A}^1 \to G$; in this case we write

$$\lim_{t\to 0} f(t) = \tilde{f}(0).$$

Let $\mu\colon \mathbb{G}_m \to G$ be a cocharacter. Define the subgroups P_μ, L_μ, and U_μ of G as follows.

- $P_\mu = \{g \in G | \lim_{t\to 0} \mu(t) g \mu(t)^{-1} \text{ exists}\}$
- $U_\mu = \{g \in G | \lim_{t\to 0} \mu(t) g \mu(t)^{-1} = 1\}$
- $L_\mu = \{\lim_{t\to 0} \mu(t) g \mu(t)^{-1} | g \in P_\mu\}$

We say that a subgroup $P \subseteq G$ is *parabolic* if it is of the form P_μ for some μ; in this case we say L_μ is an associated Levi subgroup, for any μ such that $P_\mu = P$.

In this setting, U_μ is unipotent, and P_μ is the semidirect product of L_μ by U_μ. Furthermore, L_μ is the centralizer of μ in G [43, Prop. 5.2].

Example 6.34. *The purpose of this example is to show that a parabolic subgroup $P \subseteq G$ is not uniquely determined by the parabolic subgroup $P^0 = P \cap G^0$ of G^0.*

Let $E = E_0^2$ and $\mathbf{H} = GL_N$, and let $G = \mathbf{G}$ be the group defined in Section 6.4. Then G^0 is the product of two copies of GL_N. Let μ_1 be the cocharacter that acts as t^2 on one copy of GL_N and t^3 on the other. Let μ_2 be the cocharacter that acts as t^2 on both copies of GL_N. Then $P_{\mu_1} = G^0$ but $P_{\mu_2} = G$.

Now we turn to the notion of semisimplification of subgroups of G. See [57] for a discussion of this notion in the connected setting; it is applied in [42, §2.3]. For the general theory of complete reducibility for disconnected reductive groups, see [5], [7], [4], and [6]. We warn the reader that the theory is developed there over an arbitrary field, and many complications arise in positive characteristic that are irrelevant to us here.

Definition 6.35. We say that an algebraic subgroup $H \subseteq G$ is *G-completely reducible* if, for every parabolic P containing H, there is some associated Levi subgroup L that also contains H.

If $H \subseteq G$ is an arbitrary algebraic subgroup, we define its *semisimplification* with respect to G as follows. Let P be a parabolic subgroup of G, minimal containing H. Choose a Levi subgroup L of P, and let H^{ss} be the image of H under the projection $P \to L$.

Lemma 6.36. *Let $H \subseteq G$ be an algebraic subgroup. Then H is G-completely reducible if and only if the identity component of H is reductive. If we choose an embedding of G into GL_n, then H is G-completely reducible if and only if it is GL_n-completely reducible.*

For a general algebraic subgroup $H \subseteq G$, the G-semisimplification H^{ss} is well-defined up to G-conjugacy, and it is G-completely reducible.

Proof. (This is proven in the work in progress [6, §4] by Bate, Martin, and Röhrle. Martin shared with us the argument, which we explain below.)

We begin with the claim that $H \subseteq G$ is G-completely reducible if and only if the identity component of H is reductive. Over an algebraically closed field, this is [5, §2.2]. If G and H are algebraic over a field K that is not algebraically closed, then H is G-completely reducible over K if and only if it is so over $\overline{K}$; see [4, Theorem 5.7(i)] and [4, Theorem 9.3]. This proves the claim over arbitrary K (of characteristic zero).

If $G \subseteq GL_n$, then by the above, H is GL_n-completely reducible if and only if its identity component is reductive.

Now we turn to the final claim, that H^{ss} is well-defined and G-completely reducible. Over an algebraically closed field K, this is [7, Prop. 5.10 (i)]. The notion of G-complete reducibility is independent of extension of base field in characteristic zero [7, Remark 5.3]. The fact that H^{ss} is well-defined up to $G(K)$-conjugacy, when K is not algebraically closed, follows from [4, Theorem 1.3]. □

6.6. **Some Tannakian formalism.** We recall the notion of object with G-structure. See [44, Def. 1.3]; a general reference for the Tannakian formalism is [52].

Definition 6.37. Let G be an algebraic group over a field E of characteristic zero, and let C be an E-linear rigid abelian tensor category. An *object in C with G-structure* is an exact faithful tensor functor $\omega\colon \operatorname{Rep} G \to C$, where $\operatorname{Rep} G$ is the category of finite-dimensional E-linear representations of G.

If $G_1 \to G_2$ is a map of algebraic groups, then an object with G_1-structure gives rise to an object with G_2-structure by functoriality.

Remark 6.38. If V is a faithful representation of G, then an object ω with G-structure is determined by $\omega(V)$. In practice, we will find it useful to specify ω by describing $\omega(V)$.

6.6.1. *Filtrations.* (Filtrations on G.)

We want to define a notion of "filtered vector space with G-structure" or "filtration on G." Definition 6.37 does not apply, because the category of filtered vector spaces is not abelian. Instead, we will use the formalism of filtrations from [52, §IV.2]. Throughout this section, G will be an algebraic group over a field E of characteristic zero.

Definition 6.39. A *filtration* on G (or G-filtration) is an exact tensor filtration of the forgetful functor from $\underline{\mathrm{Rep}}_G$ to $\underline{\mathrm{Vec}}_E$, in the sense of [52, IV.2.1.1].

Loosely speaking, a G-filtration is a choice of descending filtration on every finite-dimensional representation of G, compatible with tensor products, duals, and passage to subquotients.

If $G_1 \to G_2$ is a map of algebraic groups, then a G_1-filtration gives rise to a G_2-filtration by functoriality.

Since the base field E is of characteristic zero, every G-filtration is "scindable" [52, Théorème IV.2.4] and hence also "admissible" [52, IV.2.2.1]. In particular, every filtration comes from a gradation (in general not unique) on G.

Let us explain this more carefully. A representation of $\mathbb{G}_m$ on a finite-dimensional vector space V gives a decomposition $V = \bigoplus V_j$, where V_j is the t^j-eigenspace of $\mathbb{G}_m$ on V. A cocharacter $\mu\colon \mathbb{G}_m \to G$ defines a filtration on G by defining, for all representations V of G,

$$\mathrm{Fil}^i V = \bigoplus_{j \geq i} V_j.$$

That every filtration on G is "scindable" means that every filtration comes from some cocharacter μ (in general not unique).

Given a filtration on G, we can define two distinguished subgroups P and U of G [52, §IV.2.1.3]. The subgroup $P \subseteq G$ is the group of elements that stabilize the filtration on every representation V of G, and $U \subseteq P$ is the group of elements that stabilize the filtration and furthermore act as the identity on the associated graded. When G is a group whose identity component is reductive, and the filtration is defined by a cocharacter μ, the group P coincides with the Richardson parabolic P_μ (Definition 6.33), and U is its unipotent radical U_μ.

Again supposing G is a group whose identity component is reductive, define an equivalence relation on the set of cocharacters of G as follows: we say that $\mu_1 \sim \mu_2$ if $P_{\mu_1} = P_{\mu_2}$ and μ_1 is conjugate to μ_2 in this parabolic. Then μ_1 and μ_2 give rise to the same filtration on G if and only if $\mu_1 \sim \mu_2$.

If G acts faithfully on V, then a G-filtration is determined by the corresponding filtration on V, as in Remark 6.38; we will use this without comment.

We mention in passing that, even if G is reductive, the parabolic P does not quite determine the filtration. If G is reductive, then P determines a filtration on every representation of G *only up to indexing.*

6.6.2. $\mathbf{G}$-*filtrations vs* $\mathbf{G}^0$-*filtrations.* Let $\mathbf{G}$ and $\mathbf{G}^0$ be as in Section 6.4. (The discussion in this section applies to any group G whose identity component is reductive, but we will apply it only to $\mathbf{G}$ and $\mathbf{G}^0$.)

By definition, a $\mathbf{G}^0$-filtration F^0 gives rise to a $\mathbf{G}$-filtration F, by composition

$$\mathbb{G}_m \to \mathbf{G}^0 \to \mathbf{G}.$$

Since $\mathbb{G}_m$ is connected, every $\mathbf{G}$-filtration arises in this way. Conversely, given a $\mathbf{G}$-filtration, we will see shortly (Lemma 6.40) that there is a unique $\mathbf{G}^0$-filtration giving rise to it.

Lemma 6.40. *Suppose*

$$\mu_1, \mu_2 \colon \mathbb{G}_m \to \mathbf{G}$$

are two characters defining the same filtration on $\mathbf{G}$. *Then* μ_1 *and* μ_2 *define the same filtration on* $\mathbf{G}^0$.

Proof. We know there is some $g \in P_{\mu_1}$ such that

$$g\mu_1 g^{-1} = \mu_2.$$

We need to show that we can in fact take

$$g \in P^0_{\mu_1} = P_{\mu_1} \cap \mathbf{G}^0.$$

Let g_L be the image of g under the projection $P_{\mu_1} \twoheadrightarrow L_{\mu_1}$. Then g_L centralizes μ_1 by [43, Prop. 5.2(a)], and from the limit formula

$$g_L = \lim_{t\to 0} \mu(t) g \mu(t)^{-1}$$

we see that $gg_L^{-1} \in \mathbf{G}^0$. Thus $gg_L^{-1} \in \mathbf{G}^0$ conjugates μ_1 to μ_2. □

Definition 6.41. We say that a $\mathbf{G}^0$-filtration F^0 is *associated* to a $\mathbf{G}$-filtration F if there is a single cocharacter

$$\mu\colon \mathbb{G}_m \to \mathbf{G}^0 \subseteq \mathbf{G}$$

giving rise to both F^0 and F.

By the discussion above, if F and F^0 are associated, then any μ giving rise to one necessarily gives rise to the other. Being associated defines a bijection between $\mathbf{G}^0$-filtrations and $\mathbf{G}$-filtrations.

6.6.3. *G-objects in the category of weakly admissible filtered ϕ-modules.* Let G be an algebraic group over $\mathbb{Q}_p$. (It is important that we work over $\mathbb{Q}_p$, and not an extension.) A G-object in the category of weakly admissible filtered ϕ-modules gives rise to a G-filtration and an element (the Frobenius endomorphism) in G. In the spirit of Remark 6.38, we can describe such an object as a triple (V, ϕ, F), where V is a vector space on which G acts faithfully, $\phi \in G$ is an automorphism of V, and F is a G-filtration on V, such that (V, ϕ, F) is weakly admissible. Somewhat imprecisely, we will call such objects "filtered ϕ-modules with G-structure."

6.6.4. *p-adic Hodge theory.* The next result is a Tannakian form of the crystalline comparison theorem, for representations valued in an arbitrary algebraic group G. In order to avoid problems with semilinearity in the target category, we restrict to representations of $\mathrm{Gal}_{\mathbb{Q}_p}$.

Lemma 6.42. *Let $G \subseteq GL_N$ be an algebraic group over $\mathbb{Q}_p$. We say that a local Galois representation*

$$\mathrm{Gal}_{\mathbb{Q}_p} \to G$$

is crystalline *if the representation*

$$\mathrm{Gal}_{\mathbb{Q}_p} \to G_N$$

is crystalline in the usual sense. This property is independent of the choice of embedding $G \hookrightarrow GL_N$.

The functor $\underline{D}_{\mathrm{cris}}$ of p-adic Hodge theory [22, Expose III] *associates to every crystalline representation of* $\mathrm{Gal}_{\mathbb{Q}_p}$ *a G-object in the category of admissible filtered ϕ-modules over $\mathbb{Q}_p$.*

Proof. This follows from standard Tannakian properties of $\underline{D}_{\mathrm{cris}}$. □

6.7. **Structure of E-modules.** Let E be an étale H^0-algebra, and V an E-module with $\mathbf{H}$-structure, where $\mathbf{H}$ is GL_N, GSp_N, or GO_N. We described the structure of E in Example 6.16. Now we are ready to describe V.

The local system E_{et} is pulled back from a Galois representation on $\operatorname{Spec} K$, which we'll also call E_{et}. By definition this is a finite étale $\mathbb{Q}_p$-algebra E with an action of Gal_K.

Let $E_0 = \mathbb{Q}_p$, and let $\mathbf{G}$ and $\mathbf{G}^0$ be as in Section 6.4. At any $x \in X(K)$, the fiber $V = \mathsf{V}_{et,x}$ of the étale local system is a representation of the Galois group Gal_K. This V has the structure of E-algebra, and if $\mathbf{H}$ is GSp or GO then there is a pairing

$$V \otimes_E V \to L,$$

where L is a one-dimensional Q_p-vector space with an action of Gal_K. The action of Gal_K respects the pairing, and acts on V semilinearly over its action on E. It follows (Lemma 6.31) that the representation of Gal_K on V is given by a map

$$\mathrm{Gal}_K \to \mathbf{G},$$

and the quotient

$$\mathrm{Gal}_K \to \mathbf{G}/\mathbf{G}^0 \to \operatorname{Aut}_{E_0} E$$

is exactly the representation of Gal_K on E given by the structure of E_{et}.

Now we turn to de Rham cohomology and its cousins. For any point $x \in X(K)$, the group $\mathbf{G}^0$ can be realized as the group of automorphisms of $\mathsf{V}_{dR,x}$ respecting the E_{dR}-action and (where applicable) the bilinear pairing.

The differential Galois group $\mathbf{G}_{mon}$ satisfies $\mathbf{G}_{mon} \subseteq \mathbf{G}^0$. The weaker inclusion $\mathbf{G}_{mon} \subseteq \mathbf{G}$ is immediate from the Tannakian formalism; the reason $\mathbf{G}_{mon}$ is in fact contained in the finite-index subgroup $\mathbf{G}^0$ is that the monodromy action on E_{sing} is trivial.

Lastly, we describe the filtered ϕ-module $\mathsf{V}_{cris,x}$. For this, we assume that $K_v = \mathbb{Q}_p$. Recall the structure of E_{cris} from Example 6.16: it is the $\mathbb{Q}_p$-algebra E, equipped with a Frobenius endomorphism σ. Then $\mathsf{V}_{cris,x}$ is a vector space over E, and it is naturally a filtered ϕ-module with $\mathbf{G}$-structure. Its Frobenius endomorphism ϕ is semilinear over $\sigma \in \operatorname{Aut} E$.

6.8. **Galois representations.** The next result is a form of the Faltings finiteness lemma. Compare also [42, Lemma 2.4], which applies when G is a *connected* reductive group.

Lemma 6.43. *Let $E_0 = \mathbb{Q}_p$, and let E, N, $\mathbf{H}$, $\mathbf{G}$, $\mathbf{G}^0$ be as in Section 6.4. The group $\mathbf{G}$ has a standard embedding into $GL_{N[E:\mathbb{Q}_p]}(\mathbb{Q}_p)$, coming from its action by endomorphisms on a free E-module of rank n.*

Fix a number field K, a finite set S of primes of $\mathcal{O}_K$, and an integer w. There are, up to $\mathbf{G}$-conjugacy, only finitely many Galois representations

$$\rho\colon \mathrm{Gal}_K \to \mathbf{G}$$

satisfying the following conditions.

- *The representation ρ is semisimple (in the sense of Section 6.6).*
- *The representation ρ is unramified outside S.*
- *The representation ρ is* pure of weight w *with integral Weil numbers: for every prime $\ell \notin S$, the characteristic polynomial of Frob_ℓ has all roots algebraic integers of complex absolute value $q_\ell^{w/2}$.*

Proof. If ρ is semisimple in the sense of Section 6.5, then it is semisimple in the usual sense (as a representation into $GL_{n[E:\mathbb{Q}_p]}(\mathbb{Q}_p)$); see Definition 6.35 and Lemma 6.36.

Let $\mathrm{Gal}_L \subseteq \mathrm{Gal}_K$ be the kernel of

$$\mathrm{Gal}_K \to \mathbf{G} \to \mathbf{G}/\mathbf{G}^0.$$

The representation ρ restricts to a $\mathbf{G}^0$-valued representation

$$\rho|_L \, \mathrm{Gal}_L \to \mathbf{G}^0,$$

which is also semisimple (in the usual sense). Now $\mathbf{G}^0$ is a connected reductive group, so we can use [42, Lemma 2.6] to conclude that there are only finitely many possibilities for $\rho|_L$.

For each fixed choice of $\rho|_L$, if $\rho|_L$ extends to a map ρ compatible with the map $\mathrm{Gal}_K \to \mathbf{G}/\mathbf{G}^0$, then the possible extensions form a torsor under $H^1(\mathrm{Gal}_{L/K}, Z_{\mathbf{G}^0}(\mathrm{Im}(\rho|_L)))$. This cohomology group is a subgroup of $H^1(\mathrm{Gal}_K, Z_{\mathbf{G}^0}(\mathrm{Im}(\rho|_L)))$, which is finite by Serre's H^1 lemma ([56, §III.4.3, Théorème 4]). $\square$

Lemma 6.44. *Let G be an algebraic group over a field E whose identity component is reductive.*

Let $H \subseteq G$ be an algebraic subgroup.

Then the set of Levi subgroups of G containing H and defined over E forms finitely many orbits under conjugation by the E-points of the centralizer $Z_G(H)$.

Proof. A Levi subgroup L of G is the centralizer of a cocharacter $\mu \colon \mathbb{G}_m \to G$; the subgroup L contains H if and only if μ takes values in $Z_G(H)$. Since all maximal E-split tori in $Z_G(H)$ are conjugate, we can assume that μ takes values in a fixed torus T.

So we need to know that cocharacters μ of G, taking values in a given torus T, define only finitely many different Levi subgroups L of G. It is well-known that there are only finitely many possibilities for $L^0 = L \cap G^0$. But now we have

$$L^0 \subseteq L \subseteq N_G(L^0),$$

and since $N_G(L^0)$ contains L^0 with finite index, there are only finitely many possibilities for L. $\square$

Remark 6.45. Even if G^0 is a torus, and thus has only a single Levi subgroup, there can be multiple Levi subgroups of G if the component group of G acts nontrivially on the cocharacters of G^0, because a component will be in the Levi if and only if it fixes the cocharacter μ. However, there will only be finitely many.

Lemma 6.46. *(only finitely many Levis to give the semisimplification)*

Let

$$\rho_0 \colon \mathrm{Gal}_K \to \mathbf{G}$$

be a semisimple Galois representation. Then there exists a finite collection of Levi subgroups $L \subseteq \mathbf{G}$ with the following property: for any Galois representation

$$\rho \colon \mathrm{Gal}_K \to \mathbf{G}$$

whose semisimplification is $\mathbf{G}$-conjugate to ρ_0, there exist $g \in \mathbf{G}$, an L from the finite collection, and a parabolic subgroup P containing L, such that $g\rho g^{-1}$ takes values in P, and the composite map

$$\mathbf{G} \overset{g\rho g^{-1}}{\to} P \to L$$

is ρ_0.

Proof. This is an immediate consequence of Lemma 6.44. $\square$

7. Period maps and monodromy

7.1. **Compatible period maps and Bakker–Tsimerman.** The Bakker–Tsimerman theorem [3] is a strong result on the transcendence of complex period mappings. It implies a p-adic analogue, which we recall here. (See also [42, §9].)

The p-adic Bakker-Tsimerman theorem is an unlikely intersection statement for the p-adic period map attached to a Hodge–Deligne system. (For a detailed discussion of this period map, see [42, §3.3-3.4].) Suppose V is a Hodge–Deligne system on X, and let $\Omega \subseteq X(K_v)$ be a v-adic residue disk. For all $x \in \Omega$,

$$\mathsf{V}_{cris,x}$$

is a filtered ϕ-module $(V_x, \phi_x, \mathrm{Fil}_x)$. The structure of F-isocrystal means that, for all $x \in \Omega$, the vector spaces V_x are canonically identified, in such a way that ϕ_x is constant on Ω. Using this identification, the filtration Fil_x varies p-adically in x. A priori, this defines a map

$$\Phi_p \colon \Omega \to \mathcal{H} \cong GL_N/P$$

into some flag variety, where N is the rank of V. We will show that the period map actually takes values in a smaller flag variety $\mathbf{G}_{mon}/P_{mon}$, where $\mathbf{G}_{mon}$ is the differential Galois group of V.

Lemma 7.1. *If $\mathbf{G}_{mon}$ is the differential Galois group of V, then in fact the image of Φ_p is contained in a single orbit of $\mathbf{G}_{mon}$.*

Proof. (This argument was suggested by Sergey Gorchinskiy.)

Fix a basepoint $x_0 \in X(K')$, for some field K' with fixed embeddings into both $\mathbb{Q}_p$ and $\mathbb{C}$. [4] Let $\mathcal{H}$ be the flag variety classifying filtrations with the same dimensional data as the filtration on V_{dR,x_0}; we have an isomorphism $\mathcal{H} \cong GL_N/P$, for P some parabolic subgroup of GL_N.

Consider the complex period map $\Phi_{\mathbb{C}}$ from the universal cover $\tilde{X}^{an}$ of X^{an} to $\mathcal{H}$. The theorem of the fixed part ([53, §7]) implies that $\Phi_{\mathbb{C}}$ lands in a single orbit of $\mathbf{G}_{mon}$.

To transfer the result from $\Phi_{\mathbb{C}}$ to Φ_p, we use Picard–Vessiot theory. (For an introduction to Picard–Vessiot theory, see [49].)

To the vector bundle with connection underlying V, Picard–Vessiot theory attaches a $\mathbf{G}_{mon}$-torsor $PV \to X_{K'}$, whose fiber over any L-point x classifies vector space isomorphisms

$$\mathsf{V}_{dR,x} \cong \mathsf{V}_{dR,x_0}$$

respecting the $\mathbf{G}_{mon}$-structure on both sides (but not, in general, the filtrations). Furthermore, we obtain a $\mathbf{G}_{mon}$-equivariant map

$$\Phi_{PV} \colon PV \to GL_N/P$$

where a point of PV over $x \in X$ gives an isomorphism $\mathsf{V}_{dR,x} \cong \mathsf{V}_{dR,x_0}$, and we use that isomorphism to identify the filtration on $\mathsf{V}_{dR,x}$ with a point of the flag variety $\mathcal{H}$.

In the complex setting, integrating the connection gives a complex-analytic map

$$\iota_{\mathbb{C}} \colon \tilde{X}^{an} \to PV^{an};$$

[4] In our setting we are trying to prove finiteness of integral points, residue disk by residue disk, so there is no harm in taking $K' = K$: if the residue disk contains no rational points, there is nothing to prove.

in the p-adic setting, integration gives a rigid-analytic map

$$\iota_p \colon \tilde{X}^{rig} \to PV^{rig}.$$

By definition, the complex and p-adic period maps are given by $\Phi_{PV} \circ \iota_{\mathbb{C}}$ and $\Phi_{PV} \circ \iota_p$, respectively.

Since the image of $\Phi_{\mathbb{C}}$ is contained in a single $\mathbf{G}_{mon}$-orbit on $\mathcal{H}$, the same is true of the image of Φ_{PV} by $\mathbf{G}_{mon}$-equivariance. Hence, the image of Φ_p is itself contained in a single $\mathbf{G}_{mon}$-orbit. □

Lemma 7.2. *The construction above defines a p-adic period map*

$$\Phi_p \colon \Omega \to \mathbf{G}_{mon}/P_{mon},$$

where P_{mon} is a parabolic subgroup of $\mathbf{G}_{mon}$.

Proof. By Lemma 7.1, the p-adic period map takes values in $\mathbf{G}_{mon}/(P \cap \mathbf{G}_{mon})$, where $P \subseteq GL_N$ is the parabolic subgroup determined by the Hodge cocharacter μ. What remains is to show that $P \cap \mathbf{G}_{mon}$ is parabolic in $\mathbf{G}_{mon}$.

We know that μ lies in the generic Mumford-Tate group, and hence normalizes $\mathbf{G}_{mon}$ [1, §5 Thm. 1]. Since the outer automorphism group of $\mathbf{G}_{mon}$ is finite, some power μ^r acts on $\mathbf{G}_{mon}$ by inner automorphisms. Thus the adjoint action of μ^r by conjugation on $\mathbf{G}_{mon}$ gives a map

$$\mathbb{G}_m \to \operatorname{Aut} \mathbf{G}_{mon}.$$

Further increasing r if necessary, we can lift this to a cocharacter

$$\nu \colon \mathbb{G}_m \to \mathbf{G}_{mon}.$$

Then the parabolic defined by ν in $\mathbf{G}_{mon}$ is exactly $P \cap \mathbf{G}_{mon}$. □

From now on, the parabolic subgroup $P_{mon} \subseteq \mathbf{G}_{mon}$ will simply be called P. We have defined a p-adic period map, valued in $\mathbf{G}_{mon}/P$.

Theorem 7.3. *(p-adic Bakker–Tsimerman theorem). Let V be a polarized, integral Hodge–Deligne system on a smooth variety X over K, and let $\mathbf{G}_{mon}$ be the differential Galois group of V.*

Choose a p-adic residue disk $\Omega \subseteq X(\mathcal{O}_{K_p})$, and a basepoint $x_0 \in \Omega$. By Lemma 7.2, these give rise to a period map

$$\Phi_p \colon \Omega \to \mathbf{G}_{mon}/P,$$

where P is a parabolic subgroup of $\mathbf{G}_{mon}$.

Suppose $Z \subseteq \mathbf{G}_{mon}/P$ is a closed subscheme such that

$$\operatorname{codim}_{\mathbf{G}_{mon}/P} Z \geq \dim X.$$

Then $\Phi_p^{-1}(Z)$ is not Zariski-dense in X.

Proof. This is [42, Lemma 9.3]. That lemma is stated only for G the full orthogonal or symplectic group, but the proof given there applies verbatim for all reductive G, as does the underlying theorem of Bakker and Tsimerman [3] on which it is based. □

7.2. **Complex monodromy.** Our next goal is Lemma 7.6, which shows that the differential Galois group of our V is strongly c-balanced. This is a "big monodromy" statement, analogous to [42, Lemma 4.3] and [42, Theorem 8.1].

Lemma 7.4. *Let H be one of the algebraic groups SL_N, Sp_N, or SO_N. Let G be a subgroup of H^d such that each of the d coordinate projections $\pi_i\colon G \to H$ (for $1 \le i \le d$) is surjective.*

Define a relation $\sim$ on the index set $\{1, \dots, d\}$ by declaring that $i \sim j$ if and only if the projection

$$(\pi_i, \pi_j)\colon G \to H^2$$

is not *surjective.*

1. *The relation $\sim$ is an equivalence relation.*
2. *If $i_1, \dots, i_c$ are a complete set of representatives for $\sim$, then the map*

$$(\pi_{i_1}, \dots, \pi_{i_c})\colon G \to H^c$$

is surjective with finite kernel.

Proof. This is an algebraic version of Goursat's lemma. See [42, Lemma 2.12] and [50, Lemma 5.2.1].

The group H has a finite center; call it Z. For any two indices i and j (possibly equal), there are two possibilities for the image of the projection

$$G \stackrel{(\pi_i, \pi_j)}{\to} H^2 \to (H/Z)^2.$$

Either the map is surjective, or its image is the graph of an automorphism of H/Z. In the former case, (π_i, π_j) must surject (onto H^2) as well.

A calculation shows that $\sim$ is an equivalence relation. If $i_1, \dots, i_c$ is a complete system of representatives for $\sim$, then repeated application of Goursat's lemma shows that

$$(\pi_{i_1}, \dots, \pi_{i_c})\colon G \to H^c$$

is surjective with finite kernel. $\square$

Definition 7.5. Let H be one of the algebraic groups SL_N, Sp_N, or O_N, and G a subgroup of H^d. For $1 \le i \le d$, let $\pi_i\colon G \to H$ be the coordinate projection, and suppose that each π_i is surjective. The *index classes* of G are the equivalence classes of the relation $\sim$ of Lemma 7.4. We say that G is *c-balanced* (as a subgroup of H^d) if its index classes are all of equal size, and there are at least c of them.

Suppose now we are given a permutation σ of the index set $\{1, \dots, d\}$. We say G is *strongly c-balanced* (with respect to σ) if it is c-balanced, each orbit of σ on $\{1, \dots, d\}$ contains elements of at least c of the index classes of G, and σ preserves the partition of $\{1, \dots, d\}$ into index classes.

Finally, let $E_0, E, \mathbf{H}, \mathbf{G}, \mathbf{G}^0$ be as in Section 6.4, and let $\mathbf{G}^1$ be:

- in the case $\mathbf{H} = GL_N$, the kernel of the determinant map $\mathbf{G}^0 \to \mathbb{G}_{m,E}$
- in the case $\mathbf{H} = GSp_N$, the kernel of the similitude character $\mathbf{G}^0 \to \mathbb{G}_{m,E}$
- in the case $\mathbf{H} = GO_N$, the intersection of the kernels of the determinant map $\mathbf{G}^0 \to \mathbb{G}_{m,E}$ and the similitude character $\mathbf{G}^0 \to \mathbb{G}_{m,E}$.

Then $\mathbf{G}^1$ is a form of SL_N^d, Sp_N^d, or O_N^d. We say that an algebraic subgroup $G \subseteq \mathbf{G}^0$ is *c-balanced* (resp. *strongly c-balanced* with respect to σ) if $(G \cap \mathbf{G}^1)(\overline{K})$ is *c-balanced* (resp. *strongly c-balanced* with respect to σ) as a subgroup of SL_n^d, Sp_N^d or O_N^d.

Lemma 7.6. *Suppose we have K, p, v, A, X, Y, L as above. Fix some embedding $\iota_0 \colon K \to L$. Suppose χ_0 is as in Corollary 5.9, with $G^* = SL_N, Sp_N$, or SO_N, and let $E_0, E, \mathbf{H}, \mathbf{G}, \mathbf{G}^0$ be as in Section 6.4. Let I be the full $(\mathrm{Gal}_{\mathbb{Q}} \times \mathrm{Gal}_{\mathbb{Q}^{cyc}/\mathbb{Q}})$-orbit containing (ι_0, χ_0), and let $\mathsf{V} = \mathsf{V}_I$ be the corresponding Hodge–Deligne system.*

The Frobenius at v, an element of Gal_K, acts on the set I; call this permutation σ.

Then the differential Galois group $\mathbf{G}_{mon}$ of V (base-changed to $\mathbb{Q}_p$) is a strongly c-balanced subgroup of $\mathbf{G}^0$, with respect to σ.

Proof. The differential Galois group, after base change to $\mathbb{C}$, is the Zariski closure of the monodromy of the variation of Hodge structures. The variation of Hodge structure V_H splits as the direct sum of $\mathsf{V}_{H,(\iota,\chi)} = R^k f_{\iota *} g_\iota^* \mathcal{L}_\chi$. Clearly monodromy acts trivially on the set of pairs (ι, χ), and when $\mathbf{H}$ is GSp or GO, there is a bilinear pairing on $R^k f_{\iota *} g_\iota^* \mathcal{L}_\chi$. Thus, $\mathbf{G}_{mon} \subseteq \mathbf{G}^0$.

Now Corollary 5.9 implies that the geometric monodromy group of $\mathsf{V}_{H,(\iota_0,\chi_0)}$ is all of $\mathbf{H}(\mathbb{C})$. By symmetry under the action of $(\mathrm{Gal}_{\mathbb{Q}} \times \mathrm{Gal}_{\mathbb{Q}^{cyc}/\mathbb{Q}})$, the same is true for all $(\iota, \chi) \in I$, so we can talk about the relation $\sim$ and the index classes of Definition 7.5.

The action of $(\mathrm{Gal}_{\mathbb{Q}} \times \mathrm{Gal}_{\mathbb{Q}^{cyc}/\mathbb{Q}})$ respects the relation $\sim$; in particular the index classes are all of the same size, and Frobenius respects the relation $\sim$. Corollary 5.9 shows that every σ-orbit in I contains elements of at least c index classes. $\square$

8. Hodge–Deligne systems and integral points

In this section we'll finish the proof of Theorem 8.21. Recall the setup from Section 6: we have a smooth variety X over $\mathbb{Q}$ and a Hodge–Deligne system V on X, satisfying certain conditions. Lemma 6.43 tells us that, as x ranges over the integral points of X, there are only finitely many possibilities, up to semisimplification, for the global Galois representation $\mathsf{V}_{x,et}$. We will use this to bound the integral points of X.

A significant technical obstacle (both in [42] and here) is that Lemma 6.43 only applies to semisimple global Galois representations; there may be many different Galois representations arising as fibers of V_{et}, all of which have the same semisimplification. If we assume the Grothendieck–Serre conjecture—that all Galois representations that arise are semisimple—this difficulty does not arise. Under this assumption, the argument is very simple; see Section 9.

To apply Lemma 6.43 without assuming semisimplicity, we need to recognize when two global Galois representations might have the same semisimplification. We only have access to the local representations, which are generally far from semisimple. The key idea is that any filtration on the global representation, when restricted to the local representation, must be of a special form.

Passing from local Galois representations via p-adic Hodge theory to filtered ϕ-modules, we have the following situation. We're given a ϕ-module (V, ϕ) with varying filtration $F = F^\bullet$; the variation of F is described by a "monodromy group" $\mathbf{G}_{mon}$ (the differential Galois group attached to our Hodge–Deligne system). We will consider the associated graded of (V, ϕ, F) with respect to a ϕ-stable filtration $\mathfrak{f}$. By Lemma 8.8, if $\mathfrak{f}$ comes from a filtration on the global representation, then the associated graded of F with respect to $\mathfrak{f}$ is a balanced filtration (Definition 8.6). So we want to know, for how many choices of F does there exist a ϕ-stable $\mathfrak{f}$, such that the associated graded of (V, ϕ, F) with respect to $\mathfrak{f}$ lies in a given balanced isomorphism class? We will bound the dimension of such F in the flag variety. This material is very similar to the combinatorial arguments in [42, §10-11].

To start with, we'll recall some results from [42] to limit the reducibility of global representations. Recall the following notion from [42, §2.2].

Definition 8.1. For a decreasing filtration $F^\bullet V$ on a vector space V (with $F^0V = V$) we define the *weight* of the filtration to be

$$\text{weight}_F(V) = \frac{\sum_{p \geqslant 0} p \dim \mathrm{gr}^p(V)}{\dim V}, \tag{7}$$

where $\mathrm{gr}^p(V) = F^p(V)/F^{p+1}(V)$ is the associated graded.

Lemma 8.2. *Let K be a number field having no CM subfield, and let v be a place of K, unramified over $\mathbb{Q}$, and lying over the rational prime p. Let V be a representation of Gal_K on a $\mathbb{Q}_p$-vector space which is crystalline at all primes above p, and such that at all primes ℓ outside of a finite set S, the characteristic polynomial of Frobenius has algebraic coefficients and all roots rational q_ℓ-Weil numbers of weight w.*

Let $V_{dR} = (V \otimes_{\mathbb{Q}_p} B_{\mathrm{cris}})^{\mathrm{Gal}_{K_v}}$ be the filtered K_v-vector space that is associated to $\rho|_{K_v}$ by the p-adic Hodge theory functor $\underline{D}_{\mathrm{cris}}$ of [22, Expose III].

Then the weight of the Hodge filtration on V_{dR} equals $w/2$.

Proof. This is [42, Lemma 2.9]; when K has no CM subfield, the condition that v be a "friendly place" simply means that K_v is unramified over $\mathbb{Q}_p$. □

Next we'll rephrase this in terms of filtrations on reductive groups. We'll work with filtrations and semisimplifications relative to a group G whose identity component is reductive. We will apply these results to the groups $\mathbf{G}$ and $\mathbf{G}^0$ from Section 6.4, as well as similar groups. When G is disconnected, recall the notions of "parabolic subgroup," "filtration" and so forth from Sections 6.4-6.6.

We'll use the following notation (consistent with [42]): Q is the parabolic subgroup of G corresponding to the Hodge filtration F, while P corresponds to the semisimplification filtration $\mathfrak{f}$. The group M is the Levi quotient of P, corresponding to the associated graded of $\mathfrak{f}$.

Fix G, P, and M. In the reductive case [42, Lemma 11.2] defines a map from filtrations F on G to filtrations F_M on M; we'll need to extend this to the group $\mathbf{G}$ as well. Recall (Section 6.6) that for any G-filtration F, there is a cocharacter $\mu\colon \mathbb{G}_m \to G$ defining F. The substance of [42, Lemma 11.2] is that μ can be chosen to take values in P. Projecting from P to the Levi quotient M gives a filtration on M, which is independent of the choice of μ.

Lemma 8.3. *Suppose G is a group whose identity component is reductive. Let P be a parabolic subgroup of G, and M its Levi quotient.*

Fix a filtration F on G. Then there exists a cocharacter $\mu\colon \mathbb{G}_m \to P$ (with image in P) defining F. Furthermore, if F_M is the filtration on M defined by the composite map

$$\mathbb{G}_m \to P \to M,$$

then F_M is independent of the choice of μ.

Proof. For G connected reductive this is [42, Lemma 11.2].

In the general case, by [42, Lemma 11.2] applied to the identity component G^0, we know that μ can be chosen with image in $P \cap G^0$, and the corresponding filtration on $M \cap G^0$ is independent of the choice of μ. But the filtration μ defines on $M \cap G^0$ determines the filtration μ defines on M; see Section 6.6.2. □

Definition 8.4. Given G, P, M, F as above, we call the filtration F_M on G the *associated graded* filtration, and write it as $F_M = \mathrm{Gr}_M F$.

This has a natural Tannakian meaning.

Lemma 8.5. *Suppose F_P is a P-filtration, and F is the G-filtration obtained from F_P by functoriality with respect to the inclusion $P \to G$. Then F_M is the M-filtration obtained from F_P by means of the quotient map $P \to M$.*

Proof. Follows immediately from the definitions. □

We'll need a generalization of the notion of "balanced filtration" from [42, §11.1, 11.4]. Given a group S whose identity component is reductive, we define

$$\mathfrak{a}_S = X_*(Z_{S^0}) \otimes \mathbb{Q} = (X^*(S^0) \otimes \mathbb{Q})^\vee,$$

where Z_{S^0} is the center of the identity component S^0 of S. A cocharacter μ of S defines a class in $\mathfrak{a}_S$, which we call its *weight* $w(\mu) = w_S(\mu)$.

Let G, P, M be as above. Then the inclusion $Z(G^0) \to Z(M^0)$ gives a map $\iota_{GM}\colon \mathfrak{a}_G \to \mathfrak{a}_M$. Furthermore, the parabolic P defines a preferred element of

$(\mathfrak{a}_M)^\vee$, the *modular character* γ_P, defined as the inverse of the determinant of the adjoint representation of M on the Lie algebra of P.

Definition 8.6. Suppose given G an algebraic group whose identity component is reductive, P a parabolic subgroup, and M its Levi quotient. Let F_M be a filtration on M, given by a cocharacter μ.

- We say that F_M is *balanced* with respect to P if $w_M(\mu) = \iota_{GM}(w_G(\mu))$.
- We say that F_M is *weakly balanced* if $\gamma_P(\mu) = 0$ for γ_P the modular character of P.
- We say that F_M is *at least balanced* if $\gamma_P(\mu) \leq 0$, for γ_P the modular character of P.

We say a G-filtration F is balanced (resp. weakly balanced, at least balanced) with respect to P (or M, or $\mathfrak{f}$) if the associated graded $\mathrm{Gr}_M F$ is so.

We remark that a G-filtration F is balanced with respect to P if and only if the associated G^0-filtration is balanced with respect to P^0.

Remark 8.7. Balanced implies weakly balanced because

$$\gamma_P(\iota_{GM}(w_G(\mu))) = 0;$$

this identity boils down to the fact that the center of G acts trivially through the adjoint representation on the Lie algebra of P.

Furthermore, weakly balanced implies at least balanced. ("At least balanced" should more accurately be called "at least weakly balanced", but we'll call it "at least balanced" to save space.)

Lemma 8.8. *(Filtered ϕ-modules coming from global representations are weakly balanced.)*

Let G be an algebraic group over $\mathbb{Q}_p$ whose identity component is reductive, and fix an embedding $G \hookrightarrow GL_N$. Let

$$\rho\colon \mathrm{Gal}_{\mathbb{Q}} \to G(\mathbb{Q}_p) \subseteq GL_N(\mathbb{Q}_p)$$

be a representation satisfying the hypotheses of Lemma 8.2. Suppose ρ in fact has image contained in some parabolic subgroup $P(\mathbb{Q}_p) \subseteq G(\mathbb{Q}_p)$, and let M be a Levi subgroup of P.

Let (V, ϕ, F) be the filtered ϕ-module over $\mathbb{Q}_p$ that is associated to the local representation $\rho|_{\mathbb{Q}_p}$. Then F is in fact a G-filtration on V, and the associated graded $F_M = \mathrm{Gr}_M(F)$ is a balanced filtration on M.

Proof. (Compare [42, Prop. 10.6(b), §11.4, §11.6].)

That F is a G-filtration on V is a consequence of the Tannakian formalism (see Lemma 6.42).

To see that F_M is balanced, let $\mu\colon \mathbb{G}_m \to P$ be a cocharacter defining F. It is enough to show that every character $\chi\colon P \to \mathbb{G}_m$ annihilating $Z(G)$ also kills μ.

For every such character χ, a calculation shows that $\chi \circ \rho$ is pure of weight zero: the Frobenius eigenvalues at unramified primes are rational Weil numbers of weight zero.

Indeed, let ρ' be the representation of P obtained from summing ρ with χ, with χ. Let Frob_ℓ^{ss} be the semisimplification of a Frobenius element at ℓ acting on ρ'. Then Frob_ℓ^{ss} lies in P. Fix an eigenbasis of Frob_ℓ^{ss} and let T be the subgroup of P consisting of elements which have each element of this eigenbasis as eigenvectors. This is an algebraic subgroup of the torus with coordinates $\lambda_1, \ldots, \lambda_{N+1}$, hence is defined by relations $\prod_i \lambda_i^{e_i} = 1$ for $e_i = \mathbb{Z}$. Letting λ_{N+1} be the eigenvalue on χ, because $Z(G)$ acts trivially on χ, we know that every element of T whose eigenvalues $\lambda_1, \dots, \lambda_N$ are all equal acts by scalars on ρ, hence lies in $Z(G)$, and thus acts trivially on χ. It follows by restricting all relations to the subtorus where $\lambda_1 = \cdots = \lambda_N$ that there must be some relation with $e_{N+1} \neq 0$ and $\sum_{i=1}^N e_i = 0$. Applying this relation to the eigenvalues of Frob_ℓ^{ss}, we see that

$$|\chi(\mathrm{Frob}_\ell^{ss})|^{e_{N+1}} = |\chi(\mathrm{Frob}_\ell^{ss})^{e_N}| = |\prod_{i=1}^N \lambda_i(\mathrm{Frob}_\ell^{ss})^{-e_i}| = \prod_{i=1}^N |\lambda_i(\mathrm{Frob}_\ell^{ss})|^{-e_i} = \prod_{i=1}^N (p^{w/2})^{-e_i}$$

$$= p^{-(w/2)\sum_{i=1}^N e_i} = p^0 = 1$$

and so χ is pure of weight 0.

By Lemma 8.2, the weight of the corresponding filtered ϕ-module is also zero, which is what we needed to prove. □

Definition 8.9. Let E_0, E, $\mathbf{G}^0$ and $\mathbf{G}$ be as in Section 6.4, and fix $\sigma \in \mathrm{Aut}_{E_0} E$. A $\mathbf{G}$-*bifiltered* ϕ-*module* is a quadruple $(V, \phi, F, \mathfrak{f})$, with V as in Section 6.4, F and $\mathfrak{f}$ two $\mathbf{G}$-filtrations on V, and $\phi \in \mathbf{G}$ a σ-semilinear endomorphism of V respecting $\mathfrak{f}$. In this setting, let P and Q denote the parabolic subgroups of $\mathbf{G}$ corresponding to $\mathfrak{f}$ and F, respectively; to say that ϕ respects $\mathfrak{f}$ means that $\phi \in P$.

A *graded* $\mathbf{G}$-*bifiltered* ϕ-*module* is a quadruple $(V, \phi, \mathfrak{f}, F_M)$, where V is as in Notation 6.4, $\mathfrak{f}$ is a $\mathbf{G}$-filtration on V with associated Levi quotient M, and F_M is a filtration on M.

We say that two graded $\mathbf{G}$-bifiltered ϕ-modules $(V_1, \phi_1, \mathfrak{f}_1, F_{M,1})$ and $(V_2, \phi_2, \mathfrak{f}_2, F_{M,2})$ are *equivalent* if they agree up to $\mathbf{G}$-conjugacy. More precisely, let P_1 and P_2 be the parabolic subgroups attached to $\mathfrak{f}_1$ and $\mathfrak{f}_2$, and let M_1 and M_2 be their Levi quotients. We require that there exist $g \in \mathbf{G}$ satisfying the following conditions.

- $gP_1g^{-1} = P_2$
- The filtrations $gF_{M,1}g^{-1}$ and $F_{M,2}$ on M_2 agree.
- The two elements $g\phi_1 g^{-1}$ and ϕ_2 of P_2 project to the same element of M_2.

There is an obvious functor

$$(V, \phi, F, \mathfrak{f}) \to (V, \phi, \mathfrak{f}, \mathrm{Gr}_M F)$$

from $\mathbf{G}$-bifiltered ϕ-modules to graded $\mathbf{G}$-bifiltered ϕ-modules. We say that two $\mathbf{G}$-bifiltered ϕ-modules are *semisimply equivalent* if the corresponding graded $\mathbf{G}$-bifiltered ϕ-modules are equivalent.

We say that a $\mathbf{G}$-filtered ϕ-module (V, ϕ, F) is *of the semisimplicity type* $(V_0, \phi_0, F_0, \mathfrak{f}_0)$ if there exists a $\mathbf{G}$-filtration $\mathfrak{f}$ on V such that $(V, \phi, F, \mathfrak{f})$ and $(V_0, \phi_0, F_0, \mathfrak{f}_0)$ are semisimply equivalent.

Remark 8.10. Loosely speaking, the objects introduced should be understood as follows.

A $\mathbf{G}$-bifiltered ϕ-module comes (by p-adic Hodge theory) from a filtered Galois representation (i.e. a Galois representation on a vector space V, and a Galois-stable filtration on V).

A graded $\mathbf{G}$-bifiltered ϕ-module comes from the associated graded to a filtered Galois representation.

Two graded $\mathbf{G}$-bifiltered ϕ-modules are equivalent if the corresponding Galois representations are isomorphic; two $\mathbf{G}$-filtered ϕ-modules are of the same semisimplicity type if the associated gradeds of the corresponding filtered Galois representation agree up to isomorphism.

Lemma 8.11. *Let p be a prime. A representation* [5]

$$\rho_0 \colon G_{\mathbb{Q}} \to \mathbf{G}$$

of the global Galois group $G_{\mathbb{Q}}$, crystalline at p, gives rise by p-adic Hodge theory to an admissible filtered ϕ-module (V_0, ϕ_0, F_0) with $\mathbf{G}$-structure. Suppose ρ_0 is semisimple. Suppose another crystalline global representation $\rho \colon G_{\mathbb{Q}} \to \mathbf{G}$ has semisimplification ρ_0, and call the corresponding filtered ϕ-module (V, ϕ, F). Then there exist $\mathbf{G}$-filtrations $\mathfrak{f}_0$ on V_0 and $\mathfrak{f}$ on V such that $(V, \phi, F, \mathfrak{f})$ and $(V_0, \phi_0, F_0, \mathfrak{f}_0)$ are semisimply equivalent.

Furthermore, we can take $\mathfrak{f}_0$ one of a list of finitely many candidates, depending only on ρ_0, and the filtration F_0 is balanced with respect to $\mathfrak{f}_0$.

Proof. By restriction to G_{K_v} and the results of p-adic Hodge theory (Lemma 6.42), a crystalline representation $G_K \to \mathbf{G}$ gives rise to a $\mathbf{G}$-valued filtered ϕ-module (V, ϕ, F), with ϕ semilinear over σ.

To say that ρ_0 is the semisimplification of ρ means (see Lemma 6.36 and surrounding discussion) that there exist a parabolic subgroup $P \subseteq \mathbf{G}$ with associated Levi L, and a Levi $L_0 \subseteq \mathbf{G}$, such that ρ_0 takes values in L_0, and ρ takes values in P, and the image of ρ under the quotient map $P \to L$ is $\mathbf{G}$-conjugate to ρ_0. Lemma 6.46 shows that, given ρ_0, we can take L_0 to be one of finitely many possible subgroups.

Fix ρ_0 and P, and let M be the Levi quotient attached to P. By Lemma 6.42, we find that ρ and ρ_0 must give rise to filtered ϕ-modules with P-structure, such that the corresponding filtered ϕ-modules with M-structure (obtained by functoriality via the quotient map $P \to M$) are isomorphic. It follows formally that $(V, \phi, F, \mathfrak{f})$ and $(V_0, \phi_0, F_0, \mathfrak{f}_0)$ are semisimply equivalent.

Finally, F_0 is weakly balanced with respect to $\mathfrak{f}_0$ by Lemma 8.8. $\square$

Remark 8.12. Recall (Definition 6.41 and surrounding discussion) that there is an equivalence between $\mathbf{G}$-filtrations and $\mathbf{G}^0$-filtrations; we will use this without comment, and work with $\mathbf{G}^0$-filtrations on V.

Definition 8.13. (Adjoint Hodge numbers; see beginning of §10 of [42].)

Let G be a reductive group, and suppose F is a filtration on G.

In this setting, we define the *adjoint Hodge numbers* as follows: The filtration F on G gives, by definition, a filtration on every representation of G. We apply this to the adjoint representation of G on $\operatorname{Lie} G$, and call the adjoint Hodge numbers h^p the dimensions of the associated graded of the resulting filtration:

$$h^p = \dim F^p V / F^{p-1} V.$$

For any *real* $x \in [0, \sum_p h^p]$, we define the "sum of the topmost x Hodge numbers" $T(x)$ by the continuous, piecewise-linear function $T \colon [0, \sum_p h^p] \to \mathbb{R}$ satisfying

[5]The restriction to $\mathbb{Q}$, instead of an arbitrary number field K, is for two reasons. First, in the general setting, a filtered ϕ-module would be *semilinear* over K_v, and we have not defined filtered ϕ-modules with G-structure in the semilinear setting; this restriction is inessential. Second, we will need to apply Lemma 8.2, and for that we need K to have no CM subfield.

$T(0) = 0$ and

$$T'(x) = k$$

for $\sum_{p=k+1}^{\infty} h^p < x < \sum_{p=k}^{\infty} h^p$. (The sums are finite because only finitely many h^p are nonzero.)

When we want to emphasize the dependence on the group G, we'll write $T_G(x)$ and h_G^p instead of $T(x)$ and h^p.

Definition 8.14. (Uniform Hodge numbers.)

Recall notation from Section 6.4. In particular, $\mathbf{H}$ is a reductive group over $\mathbb{Q}_p$, and $\mathbf{G}^0 = \mathrm{Res}_{\mathbb{Q}_p}^E \mathbf{H}$ is a form of $\mathbf{H}^d$; the groups $\mathbf{G}^0$ and $\mathbf{G}$ act on the vector space V. Let F be a $\mathbf{G}^0$-filtration on V.

After base change to $\overline{\mathbb{Q}_p}$, we obtain

$$E \otimes \overline{\mathbb{Q}_p} \cong \bigoplus_{\iota} (\overline{\mathbb{Q}_p})_\iota,$$

the direct sum taken over all maps $\iota \colon E \to \overline{\mathbb{Q}_p}$. (The subscript on $(\overline{\mathbb{Q}_p})_\iota$ is purely for notational convenience: each $(\overline{\mathbb{Q}_p})_\iota$ is a copy of $\overline{\mathbb{Q}_p}$.) Similarly, the group $\mathbf{G}^0$ splits as a direct sum of copies of $\mathbf{H}$, and (V, F) splits as a direct sum of filtered $\overline{\mathbb{Q}_p}$-vector spaces (V_ι, F_ι), indexed by ι.

For each ι, let h_ι^p be the adjoint Hodge numbers of the $\mathbf{H}$-filtration F_ι. We say that F is *uniform* if the numbers h_ι^p are independent of ι. In this situation, we write the Hodge numbers and associated T-function as

$$h^p = h_{\mathbf{H}}^p(V_\iota) \text{ and } T(x) = T_{\mathbf{H}}(x).$$

Lemma 8.15. *In the setting of Definition 8.14, let $d = [E \colon \mathbb{Q}_p]$. We have*

$$h_{\mathbf{G}^0}^p = d h_{\mathbf{H}}^p$$

and

$$T_{\mathbf{G}^0}(dx) = d T_{\mathbf{H}}(x).$$

Proof. The Hodge numbers can be computed after base change to $\overline{\mathbb{Q}_p}$. ☐

Lemma 8.16. *Take notation as in Section 6.4, and suppose $E = E_0^c$.*

Let P be a parabolic subgroup of $\mathbf{G}^0 = \mathbf{H}^c$, corresponding to a $\mathbf{G}^0$-filtration $\mathfrak{f}$, and let M be the corresponding Levi quotient.

Let Q_0 be another parabolic subgroup of $\mathbf{G}^0$, corresponding to some filtration F_0, so that $\mathbf{G}^0 / Q_0$ parametrizes filtrations F that are $\mathbf{G}^0$-conjugate to F_0. Suppose F_0 is uniform in the sense of Definition 8.14, and let $h^p = h_{\mathbf{H}}^p$ be the adjoint Hodge numbers on $\mathbf{H}$. Let t be the dimension of a maximal torus in $\mathbf{H}$. Suppose e is a positive integer such that:

- *(First numerical condition.)*

$$\sum_{p>0} h^p \geq \frac{e}{c}$$

- *(Second numerical condition.)*

$$\sum_{p>0} p h^p > T\left(\frac{e}{c}\right) + T\left(\frac{1}{2}(h^0 - t) + \frac{e}{c}\right).$$

Then for any at-least-balanced filtration $F_{0,M}$ on M, the set of filtrations F on $\mathbf{G}^0$ that are $\mathbf{G}^0$-conjugate to F_0 and satisfy $\mathrm{Gr}_M F = F_{0,M}$ is of codimension at least e in $\mathbf{G}^0/Q_0$.

Proof. This is essentially [42, Prop. 11.3], applied to $\mathbf{G}^0 = \mathbf{H}^c$. Note that the Hodge numbers h^p and the function T used in [42] are $h_{\mathbf{H}^c}$ and $T_{\mathbf{H}^c}$. They are related to $h_{\mathbf{H}}$ and $T_{\mathbf{H}}$ by

$$h^p_{\mathbf{H}^c} = ch^p_{\mathbf{H}}$$

and

$$T_{\mathbf{H}^c}(cx) = cT_{\mathbf{H}}(x).$$

Similarly, the dimension of a maximal torus in $\mathbf{H}^c$ is ct.

Two small modifications need to be made to the proof in [42]. First, we have replaced $\frac{1}{2}h^0$ with $\frac{1}{2}(h^0 - t)$. To get this stronger bound, replace the final inequality of [42, Equation 11.15] with

$$\dim(Q/B) + e \leq \frac{1}{2}(a_0 - t) + e.$$

(Here Q is a parabolic subgroup, a_0 the dimension of its Levi quotient, and B a Borel. There is no new idea here; this bound is stronger only because the bound in [42] was not sharp.)

Second, our hypothesis is weaker: in the above-referenced proposition, $F_{0,M}$ is assumed to be balanced, while here it is only assumed to be at least balanced. This is not a problem, since the inequalities work in our favor. In our context, [42, Equation 11.14] is replaced with the inequality

$$\sum_{\gamma \in \Sigma - \Sigma_P} \langle w\mu, \gamma \rangle \leq 0,$$

and [42, Equation 11.16] becomes

$$\sum_{\beta \in X} \langle \mu, w^{-1}\beta \rangle \leq \sum_{X'} -\langle \mu, w^{-1}\beta \rangle.$$

The rest of the proof goes through as in [42]. □

Our next goal (Lemma 8.19) is a slight generalization of [42, Equation 11.18]: the result in [42] only holds with ϕ contained in a connected reductive group (e.g. $\mathbf{G}^0$), but here ϕ is semilinear, so it is contained in $\mathbf{G}$, but not in $\mathbf{G}^0$.

Lemma 8.17. *Let U be a unipotent algebraic group over a field of characteristic zero, and ψ an automorphism of U, such that $\psi^r = \mathrm{id}_U$. Suppose $u \in U$ is such that*

$$u\psi(u)\psi^2(u) \cdots \psi^{r-1}(u) = 0.$$

Then there exists $v \in U$ such that

$$u = v^{-1}\psi(v).$$

Proof. By induction on $\dim U$; reduce to the case where $U = \mathbb{G}_a^k$ is abelian. □

Lemma 8.18. *Let $\mathbf{G}^0$ and $\mathbf{G}$ be as in Section 6.4, P a parabolic subgroup of $\mathbf{G}^0$, M the associated Levi quotient, and choose a section $M \hookrightarrow P$. Suppose $\phi \in \mathbf{G}$ is semisimple and normalizes P.*

Then ϕ is P-conjugate to an element that normalizes M.

Proof. Suppose $\phi^r \in \mathbf{G}^0$. Since ϕ^r is semisimple, it is contained in some maximal torus T of $\mathbf{G}^0$; we may conjugate by an element of P to assure that $T \subseteq M$.

Since ϕ normalizes P, and all Levi subgroups in P are P-conjugate, we can write $\phi M \phi^{-1} = u^{-1} M u$ for some $u \in P$. Since u is well-defined up to left multiplication by M, we may take u unipotent. Then $(u\phi)^r$ is an element of $U\phi^r$ that normalizes M, so

$$(u\phi)^r = \phi^r.$$

Now apply Lemma 8.17, with

$$\psi(x) = \phi x \phi^{-1},$$

to conclude that

$$u = v^{-1}\psi(v) = v^{-1}\phi v \phi^{-1}$$

for some $v \in P$. Then $v^{-1}\phi v$ normalizes M. □

Lemma 8.19. *(Compare* [42, Equation 11.18]*.)*

Assume we are in the setting of Section 6.4. Fix a $\mathbf{G}$*-bifiltered* ϕ*-module* $(V_0, \phi_0, F_0, \mathfrak{f}_0)$*, and another* ϕ*-module* (V, ϕ) *with* $\mathbf{G}$*-structure; suppose both* ϕ *and* ϕ_0 *are semilinear over some* $\sigma \in \operatorname{Aut}_{E_0} E$*.*

Consider all pairs $(\mathfrak{f}, F_M)$*, where* $\mathfrak{f}$ *is a filtration on* $\mathbf{G}$*,* M *is the Levi quotient, and* $(V, \phi, \mathfrak{f}, F_M)$ *is equivalent to* $(V_0, \phi_0, \mathfrak{f}_0, (F_0)_{M_0})$*, in the sense of Definition 8.9. The set of such pairs has dimension at most* $\dim Z_{\mathbf{G}^0}(\phi^{ss})$*.*

(Even though ϕ^{ss} is only defined up to conjugacy, the dimension $\dim Z_{\mathbf{G}^0}(\phi^{ss})$ is well-defined, so the statement of the lemma makes sense.)

Proof. This is a question about the dimension of a variety, so we can pass to a finite extension of K. In particular, we may assume $E = K^d$ and σ acts by permutation on the factors of $E = K^d$. We'll index the factors K_1 through K_d, and we'll regard σ as a permutation of $\{1, \ldots, d\}$. Write $V_i = V \otimes_E K_i$ for the i-th factor of V, and write $\mathbf{H}_i$ for the corresponding factor of $\mathbf{G}^0 = \mathbf{H}^d$ (it satisfies $\mathbf{H}_i \cong \mathbf{H}$). The semilinearity condition means that ϕ decomposes as a sum of maps

$$\phi_i \colon V_i \to V_{\sigma i}.$$

As in [42, third paragraph of section 11.6], we can assume ϕ is semisimple; if not, passing to the semisimplification will only increase the dimension of the set described. (If $\phi \in N_{\mathbf{G}} P$ is semilinear over σ, then its semisimplification can be taken to have the same properties.)

A filtration $\mathfrak{f}$ on $\mathbf{G}^0$ (which is the same as a filtration on $\mathbf{G}$) decomposes as the product of filtrations $\mathfrak{f}_i$ on $\mathbf{H}_i$. If $\mathfrak{f}$ is ϕ-stable, then $\mathfrak{f}_i$ determines $\mathfrak{f}_{\sigma i}$. So to specify $\mathfrak{f}$ it is enough to specify $\mathfrak{f}_i$, for a single i in each σ-orbit.

Since everything in sight splits as a direct product over σ-orbits, we may as well restrict attention to a single σ-orbit; call it $\{1, \ldots, r\}$. A ϕ-stable filtration $\mathfrak{f}$ on $\mathbf{G}^0$ is uniquely determined by a ϕ^r-stable filtration $\mathfrak{f}_1$ on $\mathbf{H}_1$. As in [42], the dimension of the set of such filtrations is exactly

$$\dim Z_{\mathbf{G}^0}(\phi) - \dim Z_P(\phi),$$

where $Z_{\mathbf{G}^0}(\phi) = Z_{\mathbf{G}}(\phi) \cap \mathbf{G}^0$, and $Z_P(\phi)$ is defined similarly.

To conclude, we need to show that, given $\mathfrak{f}$ (and its associated Levi quotient M), the set of filtrations F_M such that $(V, \phi, \mathfrak{f}, F_M)$ is equivalent to $(V_0, \phi_0, \mathfrak{f}_0, (F_0)_{M_0})$ has dimension at most $\dim Z_P(\phi)$. The proof of this is the same as in [42], using Lemma 8.18 in place of [42, Equation 2.1]. □

Lemma 8.20. *Assume we are in the setting of Section 6.4. Fix a $\mathbf{G}^0$-bifiltered ϕ-module $(V_0, \phi_0, F_0, \mathfrak{f}_0)$, and another ϕ-module (V, ϕ) with $\mathbf{G}$-structure; suppose both ϕ and ϕ_0 are semilinear over some $\sigma \in \operatorname{Aut}_{E_0} E$, and F_0 is balanced with respect to $\mathfrak{f}_0$.*

Let $\mathbf{G}_{mon}$ be a subgroup of $\mathbf{G}^0$, strongly c-balanced with respect to σ.

Suppose F_0 is uniform in the sense of Definition 8.14, and let $h^p = h^p_{simp}$ be the adjoint Hodge numbers on $\mathbf{H}$. Let t be the dimension of a maximal torus in $\mathbf{H}$. Suppose e is a positive integer satisfying the following numerical conditions.

- *(First numerical condition.)*
$$\sum_{p>0} h^p \geq \frac{1}{c}(e + \dim \mathbf{H})$$
- *(Second numerical condition.)*
$$\sum_{p>0} p h_p > T\left(\frac{1}{c}(e + \dim \mathbf{H})\right) + T\left(\frac{1}{2}(h^0 - t) + \frac{1}{c}(e + \dim \mathbf{H})\right).$$

Let $\mathcal{H} = \mathbf{G}_{mon}/(Q^0 \cap \mathbf{G}_{mon})$ be the flag variety parametrizing filtrations on $\mathbf{G}^0$ that are conjugate to F_0 under the conjugation of $\mathbf{G}_{mon}$. Then the filtrations F such that (V, ϕ, F) is of semisimplicity type $(V_0, \phi_0, F_0, \mathfrak{f}_0)$ are of codimension at least e in $\mathcal{H}$.

Proof. Let P be the parabolic of $\mathbf{G}^0$, and M the Levi quotient, associated to a hypothetical semisimplification filtration $\mathfrak{f}$. There are only finitely many possibilities for the parabolic group P, up to $\mathbf{G}^0$-conjugacy, so we may as well as fix a single P.

The dimension in question can be calculated after base change to an extension of K, so we can assume that $\mathbf{G}^0 = \mathbf{H}^d$. As in the proof of Lemma 8.19, we'll call the d factors $\mathbf{H}_1, \ldots, \mathbf{H}_d$. The whole setup factors over these d factors, so we can write P as the direct sum of parabolics $P_i \subseteq \mathbf{H}_i$, and so forth. Again, $\sigma \in \operatorname{Aut}_{E_0} E$ gives a permutation of the index set $\{1, \ldots, d\}$, which we also call σ. Semilinearity over σ means that the map ϕ permutes the d factors according to the permutation σ.

The strategy is as follows. We want to apply a result like Lemma 8.16, but we don't have full monodromy group $\mathbf{G}^0$. Instead, we know the group $\mathbf{G}_{mon}$ is c-balanced; we'll project onto c of the d factors, so that $\mathbf{G}_{mon}$ projects onto the full group $(\mathbf{H}/Z(\mathbf{H}))^c$. We need the projection of F_0 to the c factors to be at least balanced, and we can apply Lemmas 8.16 and 8.19 to the group $\mathbf{H}^c$ to finish.

For any subset I of the index set $\{1, \ldots, d\}$, let $\mathbf{H}^I = \prod_{i \in I} \mathbf{H}_i$; define P^I, M^I, and F_0^I similarly. Any filtrations F and $\mathfrak{f}$ on $\mathbf{G}^0$ can be written as products of factors F_i and $\mathfrak{f}_i$, so we can define F^I and $\mathfrak{f}^I$.

I claim that, for any $\mathfrak{f}$ and any F that is balanced with respect to $\mathfrak{f}$, we can find $I \subseteq \{1, \ldots, d\}$ satisfying the following properties.

- I consists of exactly c elements, from c distinct index classes for $\mathbf{G}_{mon}$.
- The elements of I belong to a single orbit of σ on $\{1, \ldots, d\}$.
- F^I is at least balanced with respect to $\mathfrak{f}^I$.

To see this, let μ be a cocharacter defining F as in Definition 8.6. We can write μ as a product over μ_i with $i \in \{1, \ldots, d\}$. Similarly, the character γ_P splits over the factors, and we have:

$$\gamma_P(\mu) = \sum_i \gamma_{P_i}(\mu_i).$$

Hence there exists an orbit J of σ on the d factors such that F^J is at least balanced. Since $\mathbf{G}_{mon}$ is strongly c-balanced, we can find a subset

$$I \subset J \subset \{1, \ldots, d\}$$

of the index set such that $\#I = c$, the elements of I belong to c distinct index classes, and F^I is at least balanced. Since the elements of I belong to distinct index classes, the projection

$$\mathbf{G}_{mon} \to (\mathbf{H}/Z(\mathbf{H}))^I$$

has image a union of connected components of the target, so it is smooth with equidimensional fibers, and the same is true of

$$\mathbf{G}_{mon}/(Q^0 \cap \mathbf{G}_{mon}) \to \mathbf{H}^I/(Q^{0,I} \cap \mathbf{H}^I).$$

We want to estimate the codimension in $\mathcal{H} = \mathbf{G}_{mon}/(Q^0 \cap \mathbf{G}_{mon})$ of the set of F such that $(V, \phi, F, \mathfrak{f})$ is semisimply equivalent to $(V_0, \phi_0, F_0, \mathfrak{f}_0)$, for some choice of $\mathfrak{f}$. Consider the projection

$$(V, \phi, F, \mathfrak{f}) \to (\mathfrak{f}, F_M).$$

By Lemma 8.19 applied to $\mathbf{H}^J$, the set of pairs $(\mathfrak{f}^J, F_M^J)$ such that $(V_0^J, \phi_0^J, (F_0)_{M_0}^J, \mathfrak{f}_0^J)$ is equivalent to $(V^J, \phi^J, F_M^J, \mathfrak{f}^J)$ has dimension bounded by $\dim Z_{G^J}((\phi^J)^{ss})$; this dimension is at most $\dim \mathbf{H}$ by Lemma 6.32.

Now fix $(\mathfrak{f}^J, F_M^J)$. By Lemma 8.16 applied to $\mathbf{H}^I$, the set of F^I such that

$$\mathrm{Gr}_M F^I = F_M^I$$

has codimension at least $e + \dim \mathbf{H}$ among all $\mathbf{H}^I$-filtrations F^I. The map $F^J \mapsto F^I$ from $\mathbf{H}^J$-filtrations in a given $\mathbf{G}_{mon}$-conjugacy class to $\mathbf{H}^I$-filtrations is smooth with equidimensional fibers, so the set of F^J such that

$$\mathrm{Gr}_M(F^J) = F_M^J$$

again has codimension at least $e + \dim \mathbf{H}$.

It follows that the set of F satisfying the desired condition has codimension at least e. ☐

Theorem 8.21. *(Basic theorem giving non-density of integral points.)*

Let X be a variety over $\mathbb{Q}$, let S be a finite set of primes of $\mathbb{Z}$, and let $\mathcal{X}$ be a smooth model of X over $\mathbb{Z}[1/S]$.

Let E be a constant H^0-algebra on X, and let $\mathbf{H}$ be one of GL_N, GSp_N, or GO_N. Let V be a polarized, integral E-module with $\mathbf{H}$-structure, in the sense of Definition 6.26, having integral Frobenius eigenvalues (Def. 6.5). Suppose the Hodge numbers of V are uniform in the sense of Definition 8.14, and let $h^p = h^p_{simp}$ be the adjoint Hodge numbers on $\mathbf{H}$. Let t be the dimension of a maximal torus in $\mathbf{H}$. Let p be as in Definition 6.2. Let $\mathbf{G}^0$ and $\mathbf{G}$ be as in Section 6.4.

Let $\Omega \subseteq X$ be a residue disk modulo p. Suppose there is a positive integer c such that V satisfies the following conditions.

- *(Big monodromy.) If $\mathbf{G}_{mon} \subseteq \mathbf{G}^0$ is the differential Galois group of V, then $\mathbf{G}_{mon} \subseteq \mathbf{G}^0$ is strongly c-balanced with respect to Frobenius. (The Frobenius is determined from the structure of E; see Section 6.7.)*

- *(First numerical condition.)*
$$\sum_{p>0} h^p \geq \frac{1}{c}(\dim X + \dim \mathbf{H})$$
- *(Second numerical condition.)*
$$\sum_{p>0} ph_p > T_G\left(\frac{1}{c}(\dim X + \dim \mathbf{H})\right) + T_G\left(\frac{1}{2}(h^0 - t) + \frac{1}{c}(\dim X + \dim \mathbf{H})\right)$$

Then the image of $\mathcal{X}(\mathbb{Z}[1/S]) \cap \Omega$ in X is not Zariski dense.

In particular, if this holds for all (finitely many) mod-p residue disks Ω, then the image of $\mathcal{X}(\mathbb{Z}[1/S])$ is not Zariski dense in X.

Proof. For every $x \in \mathcal{X}(\mathbb{Z}[1/S])$, the fiber $\rho_x = \mathsf{V}_{et,x}$ of the p-adic étale local system is a global Galois representation valued in $\mathbf{G}$, having good reduction outside S and all Frobenius eigenvalues Weil numbers, by hypothesis. By Faltings's finiteness lemma (in the form of Lemma 6.43), there are only finitely many possible isomorphism classes for the semisimplified representation ρ_x^{ss}. So it is enough to show, for any fixed ρ_0, that the set
$$\mathcal{X}(\mathbb{Z}[1/S], \rho_0) := \{x \in X(\mathcal{O}_{K,S}) | \rho_x^{ss} \cong \rho_0\}$$
is not Zariski dense in X.

Lemma 8.11 gives a finite list of semisimplicity types such that, for every $x \in \mathcal{X}(\mathbb{Z}[1/S], \rho_0)$, the filtered ϕ-module $\mathsf{V}_{cris,x}$ belongs to one of them. Let $\Omega \subseteq \mathcal{X}(\mathbb{Z}[1/S], \rho_0)$ be a mod-p residue disk, and fix a semisimplicity type $(V_0, \phi_0, F_0, \mathfrak{f}_0)$. By Lemma 8.11, F_0 is weakly balanced with respect to $\mathfrak{f}_0$. It is enough to show that the set
$$X(\Omega, (V_0, \phi_0, F_0, \mathfrak{f}_0)) = \{x \in \Omega | \mathsf{V}_{cris,x} \text{ is of semisimplicity type } (V_0, \phi_0, F_0, \mathfrak{f}_0)\}$$
is not Zariski dense in X.

We are now in the setting of Theorem 7.3. For $x \in \Omega$, the filtered ϕ-module $\mathsf{V}_{cris,x}$ is of the form (V, ϕ, F_x), where (V, ϕ) is independent of x. (This is a property of F-isocrystals in general; it reflects the fact that if (V, ϕ, F) is the crystalline cohomology of a scheme over $\mathbb{Z}_p$, then (V, ϕ) can be recovered from its reduction modulo p. See [42, Section 3.3] for further discussion.) The variation of F_x with x is classified by a p-adic period map
$$\Phi_p \colon \Omega \to \mathbf{G}^0/Q,$$
where $\mathbf{G}^0/Q$ is the flag variety classifying $\mathbf{G}^0$-filtrations on V.

In fact, the p-adic period map lands in a single $\mathbf{G}_{mon}$-orbit on $\mathbf{G}^0/Q$; we can write the orbit as $\mathbf{G}_{mon}/(Q \cap \mathbf{G}_{mon})$. By Lemma 8.20 with $e = \dim X$, there is a Zariski-closed set $Z \subseteq \mathbf{G}_{mon}/(Q \cap \mathbf{G}_{mon})$ of codimension at least $\dim X$ such that, if (V, ϕ, F_x) is of semisimplicity type $(V_0, \phi_0, F_0, \mathfrak{f}_0)$, then $x \in \Phi_p^{-1}(Z)$. We conclude by Theorem 7.3. □

9. Hodge–Deligne systems and integral points, assuming global semisimplicity

In this section we prove Theorem 9.3, a variant of Theorem 8.21 that assumes the semisimplicity of certain global Galois representations. This material is not logically needed for the main argument; we include it to illustrate the main ideas of Section 8, without the complications coming from semisimplification.

Lemma 9.1. *(Compare Lemma 8.11.)*

Let p be a prime. A semisimple *representation*

$$\rho_0 \colon G_{\mathbb{Q}} \to \mathbf{G}$$

of the global Galois group $G_{\mathbb{Q}}$, crystalline at p, gives rise by p-adic Hodge theory to an admissible filtered ϕ-module (V_0, ϕ_0, F_0) with $\mathbf{G}$-structure. Suppose another crystalline global representation $\rho \colon G_K \to \mathbf{G}$ is isomorphic to ρ_0, and call the corresponding filtered ϕ-module (V, ϕ, F). Then there is an isomorphism of filtered ϕ-modules

$$(V_0, \phi_0, F_0) \cong (V, \phi, F).$$

In particular, if $(V_0, \phi_0) = (V, \phi)$, then there exists an automorphism $f \colon V \to V$ such that

- *f commutes with ϕ, and*
- *$fF_0 = F$.*

Proof. The functors of p-adic Hodge theory take isomorphic objects to isomorphic objects. □

Lemma 9.2. *(Compare Lemma 8.20.)*

Assume we are in the setting of Section 6.4. Fix an admissible filtered ϕ-module with $\mathbf{G}$-structure (V_0, ϕ_0, F_0), and another ϕ-module (V, ϕ) with $\mathbf{G}$-structure; suppose both ϕ and ϕ_0 are semilinear over some $\sigma \in \operatorname{Aut}_{E_0} E$.

Let $\mathbf{G}_{mon}$ be a subgroup of $\mathbf{G}^0$, strongly c-balanced with respect to σ.

Suppose F_0 is uniform in the sense of Definition 8.14, and let $h^p = h^p_{simp}$ be the adjoint Hodge numbers on $\mathbf{H}$. Suppose e is a positive integer satisfying the following numerical condition.

- *(Numerical condition.)*

$$\sum_{p>0} h^p \geq \frac{1}{c}(e + \dim \mathbf{H})$$

Let $\mathcal{H} = \mathbf{G}_{mon}/(Q^0 \cap \mathbf{G}_{mon})$ be the flag variety parametrizing filtrations on $\mathbf{G}^0$ that are conjugate to F_0 under the conjugation of $\mathbf{G}_{mon}$. Then the filtrations F such that (V, ϕ, F) is isomorphic to (V_0, ϕ_0, F_0) are of codimension at least e in $\mathcal{H}$.

Proof. This is a question about the dimension of a variety over $\mathbb{Q}_p$; by passing to an extension, we may assume that $\mathbf{G}^0 = \mathbf{H}^d$ is split, and σ acts by permuting the factors. Increasing c if necessary, we obtain an isomorphism $\mathbf{G}_{mon} \cong \mathbf{H}^c$ by projecting onto some c of the d factors, and we may assume that σ acts transitively on these c factors.

By Lemma 9.1, the filtrations F satisfying the condition described form at most one orbit under the action of $Z(\phi)$ on $\mathcal{H}$. By Lemma 6.32, we have

$$\dim Z(\phi) \leq \dim \mathbf{H}.$$

On the other hand,

$$\dim \mathcal{H} = c \sum_{p>0} h^p.$$

The result follows. □

Theorem 9.3. *(Compare Theorem 8.21.)*

Let X be a variety over $\mathbb{Q}$, let S be a finite set of primes of $\mathcal{O}_K$, and let $\mathcal{X}$ be a smooth model of X over $\mathbb{Z}[1/S]$.

Let E be a constant H^0-algebra on X, and let $\mathbf{H}$ be one of GL_N, GSp_N, or GO_N. Let V be a polarized, integral, E-module with $\mathbf{H}$-structure, in the sense of Definition 6.26, having integral Frobenius eigenvalues (Def. 6.5). Suppose the Hodge numbers of V are uniform in the sense of Definition 8.14, and let $h^p = h^p_{simp}$ be the adjoint Hodge numbers on $\mathbf{H}$. Let p be as in Definition 6.2. Let $\mathbf{G}^0$ and $\mathbf{G}$ be as in Section 6.4.

Let $\Omega \subseteq X$ be a residue disk modulo p. Suppose there is a positive integer c such that V satisfies the following conditions.

- *(Big monodromy.) If $\mathbf{G}_{mon} \subseteq \mathbf{G}^0$ is the differential Galois group of V, then $\mathbf{G}_{mon} \subseteq \mathbf{G}^0$ is strongly c-balanced with respect to Frobenius. (The Frobenius is determined from the structure of E; see Section 6.7.)]*
- *(Numerical condition.)*

$$\sum_{p>0} h^p \geq \frac{1}{c}(\dim X + \dim \mathbf{H})$$

Let $\mathcal{X}(\mathbb{Z}[1/S])^{ss}$ be the subset of $\mathcal{X}(\mathbb{Z}[1/S])$ consisting of those x for which the global Galois representation $\mathsf{V}_{et,x}$ is semisimple. Then the image of $\mathcal{X}(\mathbb{Z}[1/S])^{ss} \cap \Omega$ in X is not Zariski dense.

In particular, if this holds for all (finitely many) mod-p residue disks Ω, then the image of $\mathcal{X}(\mathbb{Z}[1/S])$ is not Zariski dense in X.

Proof. This follows from Lemmas 9.1 and 9.2 in the same way that Theorem 8.21 follows from Lemmas 8.11 and 8.11. We will sketch the proof here; the reader can find more details in the proof of Theorem 8.21.

For every $x \in \mathcal{X}(\mathbb{Z}[1/S])$, consider the semisimple global Galois representation $\rho_x = \mathsf{V}_{et,x}$. By Lemma 6.43, there are only finitely many possible isomorphism classes for the semisimple representation ρ_x. So it is enough to show, for any fixed ρ_0, that the set

$$\mathcal{X}(\mathbb{Z}[1/S], \rho_0) := \{x \in X(\mathcal{O}_{K,S}) | \rho_x \cong \rho_0\}$$

is not Zariski dense in X. By Lemma 9.1, it is enough to show that the set

$$X(\Omega, (V_0, \phi_0, F_0)) = \{x \in \Omega | (V, \phi, F) \cong (V_0, \phi_0, F_0)\}$$

is not Zariski dense in X.

We are now in the setting of Theorem 7.3. For $x \in \Omega$, the filtered ϕ-module $\mathsf{V}_{cris,x}$ is of the form (V, ϕ, F_x), where (V, ϕ) is independent of x. The variation of F_x with x is classified by a p-adic period map

$$\Phi_p \colon \Omega \to \mathbf{G}^0/Q,$$

where $\mathbf{G}^0/Q$ is the flag variety classifying $\mathbf{G}^0$-filtrations on V.

In fact, the p-adic period map lands in a single $\mathbf{G}_{mon}$-orbit on $\mathbf{G}^0/Q$; we can write the orbit as $\mathbf{G}_{mon}/(Q \cap \mathbf{G}_{mon})$. By Lemma 9.2 with $e = \dim X$, there is a

Zariski-closed set $Z \subseteq \mathbf{G}_{mon}/(Q \cap \mathbf{G}_{mon})$ of codimension at least $\dim X$ such that, if $(V, \phi, F_x) \cong (V_0, \phi_0, F_0)$, then $x \in \Phi_p^{-1}(Z)$. We conclude by Theorem 7.3. $\square$

10. Verifying the numerical conditions

The goal of this section is to verify the two numerical conditions of Theorem 8.21. This is very similar to the estimates in [42, Section 10.2].

Our Hodge-Deligne systems arise from pushforwards of character sheaves of an abelian variety along families of hypersurfaces in that abelian variety. Thus, the Hodge structure on the stalk at a point arises from the cohomology of a hypersurface, twisted by a rank-one character sheaf.

For A an abelian variety of dimension n, Y a degree d hypersurface in A, χ a finite-order character of $\pi_1(A)$ such that $H^i(Y, \mathcal{L}_\chi)$ vanishes for $i \neq n-1$ vanishes for $i \neq n-1$, the Hodge structure $H^{n-1}(Y, \mathcal{L}_\chi)$ has Hodge numbers numbers $h^k = dA(n,k)$, where $A(n,k)$ is the Eulerian number, by Lemma 2.4. (The *degree* of a hypersurface in an abelian variety is the degree of the corresponding polarization.)

Recall from Section 6.3 that the various realizations of $H^{n-1}(Y, \mathcal{L}_\chi)$ form a Hodge–Deligne system with $\mathbf{H}$-structure, for $\mathbf{H}$ one of GL, GSp, and GO. This system has uniform Hodge numbers (Definition 8.14), as computed in Lemma 2.4, so it makes sense to talk about the Hodge filtration as a filtration on $\mathbf{H}$.

Lemma 10.1. *Let Y be a hypersurface in an n-dimensional abelian variety A, let χ be a finite-order character of $\pi_1(A)$ such that $H^i(Y, \mathcal{L}_\chi)$ vanishes for $i \neq n-1$, and let V be the Hodge structure on $H^{n-1}(Y, \mathcal{L}_\chi)$, regarded as a filtered vector space with $\mathbf{H}$-structure, for $\mathbf{H}$ one of* GL*,* GSp*,* GO*. Let h^p be the adjoint Hodge numbers (Definition 8.13), and let t be the dimension of a maximal torus in $\mathbf{H}$.*

If $n \geq 2$, we have $\frac{1}{2}(h^0 - t) < \sum_{p>0} h^p$.

Proof. (See also [42, Lemma 10.3].)

Since the adjoint Hodge numbers satisfy $h^p = h^{-p}$, we have

$$\sum_{p>0} h^p = \frac{1}{2}\left((\sum_{p\in\mathbb{Z}} h^p) - h^0 \right) = \frac{1}{2}\left(\dim \mathbf{H} - h^0\right).$$

Thus it is enough to show that

$$2h^0 < \dim \mathbf{H} + t. \tag{8}$$

In each case, we will calculate h^0 and $\dim \mathbf{H}$ in terms of Eulerian numbers $A(n,k)$, and then prove (8) by using the following inequality of Eulerian numbers, to be established later:

$$2\sum_p d^2 A(n,p)^2 \leq d^2 \left(\sum_p A(n,p)\right)^2 \tag{9}$$

We adopt the convention that $A(n,k) = 0$ when $k \notin \{0, \dots, n-1\}$.

- If $\mathbf{H} = \mathrm{GL}$ then

$$\begin{aligned} h^0 &= \sum_p d^2 A(n,p)^2 \\ \dim \mathbf{H} &= d^2 \left(\sum_p A(n,p)\right)^2 \\ t &= d \sum_p A(n,p). \end{aligned}$$

so

$$2h^0 = 2\sum_p d^2 A(n,p)^2 \le d^2 \left(\sum_p A(n,p)\right)^2 < d^2 \left(\sum_p A(n,p)\right)^2 + d\sum_P A(n,p) = \dim \mathbf{H} + t.$$

- If $\mathbf{H} = \mathrm{GSp}$ then

$$\begin{aligned} h^0 &= \frac{1}{2}\left[\left(\sum_p d^2 A(n,p)^2\right) + dA(n, \frac{n-1}{2})\right] + 1 \\ \dim \mathbf{H} &= \frac{1}{2} d \left(\sum_p A(n,p)\right)\left[d\left(\sum_p A(n,p)\right) + 1\right] + 1 \\ t &= \frac{1}{2} d \sum_p A(n,p) + 1. \end{aligned}$$

so

$$2h^0 = \left(\sum_p d^2 A(n,p)^2\right) + dA(n, \frac{n-1}{2}) + 2 \le \frac{d^2}{2}\left(\sum_p A(n,p)\right)^2 + dA(n, \frac{n-1}{2}) + 2$$

$$< \frac{d^2}{2}\left(\sum_p A(n,p)\right)^2 + d\left(\sum_p A(n,p)\right) + 2 = \dim \mathbf{H} + t$$

- If $\mathbf{H} = \mathrm{GO}$ then

$$\begin{aligned} h^0 &= \frac{1}{2}\left[\left(\sum_p d^2 A(n,p)^2\right) - dA(n, \frac{n-1}{2})\right] + 1 \\ \dim \mathbf{H} &= \frac{1}{2} d \left(\sum_p A(n,p)\right)\left[d\left(\sum_p A(n,p)\right) - 1\right] + 1 \\ t &= \frac{1}{2} d \sum_p A(n,p) + 1, \end{aligned}$$

(The formula for t in the GO case holds only for even-dimensional orthogonal groups, but $d\sum_p A(n,p) = d \cdot n!$ is even.) Thus

$$2h^0 = \left(\sum_p d^2 A(n,p)^2\right) - dA(n, \frac{n-1}{2}) + 2 < \left(\sum_p d^2 A(n,p)^2\right) + 2 \le \frac{d^2}{2}\left(\sum_p A(n,p)\right)^2 + 2 = \dim \mathbf{H} + t.$$

We now prove (9). It is known that, as n grows large, the Eulerian numbers approximate a normal distribution with variance $\sqrt{n}/12$. This purely qualitative result implies that (9) holds for sufficiently large n.

To get precise bounds, we'll use log concavity, together with a calculation of the second moment. The key idea is that a sequence of numbers that (i) is log-concave, and (ii) has large second moment, cannot be too concentrated at the middle term. Let

$$a_i = \frac{\sum_p A(n,p) A(n,p-i)}{(\sum_p A(n,p))^2};$$

this is normalized so that $\sum a_i = 1$.

Now we'll prove that $a_0 \le 1/2$ for all $n \ge 2$; this will be Lemma 10.6. This will be a consequence of log-concavity and a formula for the second moment.

Lemma 10.2. *The sequence (a_i) is log-concave and satisfies $a_{-i} = a_i$.*

Proof. This is proved in the first paragraph of the proof of [42, Lemma 10.3]. Symmetry is elementary; log-concavity follows from the classical fact that the Eulerian numbers are log-concave (see, for example, [47, Problems 4.6 and 4.8]) and the fact that log-concavity is preserved under convolution [32, Thms. 1.4, 3.3]. □

Lemma 10.3. *The second moment of (a_i) is*

$$\sum_i i^2 a_i = \frac{n+1}{6}.$$

Proof. This is [42, Eqn. 10.10]. □

Lemma 10.4. *Suppose $a_0 > 1/2$. Then for all $k \geq 1$ we have*

$$\sum_{i=k}^{\infty} a_i < \sum_{i=k}^{\infty} \frac{1}{2 \cdot 3^i}.$$

Proof. Because $a_0 > 1/2$, by symmetry, we have

$$\sum_{i=1}^{\infty} a_i = \frac{1-a_0}{2} < \frac{1}{4} = \sum_{i=1}^{\infty} \frac{1}{2 \cdot 3^i}.$$

so if

$$\sum_{i=k}^{\infty} a_i \geq \sum_{i=k}^{\infty} \frac{1}{2 \cdot 3^i}$$

we must have $a_j < \frac{1}{2 \cdot 3^j}$ for some $1 \leq j < k$.

But then by log-concavity and the fact that $a_0 > 0$, we have $a_{<\frac{1}{2\cdot 3^j}}$ for all $i > j$, and thus in particular for all $i \geq k$, so

$$\sum_{i=k}^{\infty} a_i < \sum_{i=k}^{\infty} \frac{1}{2 \cdot 3^i}.$$

□

Lemma 10.5. *If $a_0 > 1/2$ then $\sum_k k^2 a_k < \frac{3}{2}$.*

Proof. We have

$$\sum_k k^2 a_k = 2\sum_{k=1}^{\infty} k^2 a_k = 2\sum_{k=1}^{\infty}(2k-1)\sum_{i=k}^{\infty} a_i > 2\sum_{k=1}^{\infty}(2k-1)\sum_{i=k}^{\infty}\frac{1}{2\cdot 3^i} = 2\sum_{k=1}^{\infty} k^2 \frac{1}{2 \cdot 3^i} = \frac{3}{2}.$$

Here we use symmetry, Lemma 10.4, and the identity

$$\sum_{k>0} k^2 c^k = \frac{c(1+c)}{(1-c)^3}.$$

□

Lemma 10.6. *If $n \geq 2$ then $a_0 \leq 1/2$.*

Proof. If $n \geq 9$, this is immediate from Lemmas 10.3 and 10.5.

For $2 \leq n \leq 8$, we can prove this by computation. For even n this follows immediately from the symmetry $A(n,k) = A(n, n-1-k)$, so it suffices to check $n = 3, 5, 7$, which can be done by hand. □

This proves Lemma 10.1. □

11. Proof of main theorem

Definition 11.1. Let A be an abelian variety over a number field K, and $H \subseteq A$ a hypersurface. We say that H is *primitive* if it is not invariant under translation by any $x \in A(\overline{\mathbb{Q}})$.

Recall the sequence $a(i)$ defined in Theorem 4.1.

Theorem 11.2. *Let A be an abelian variety of dimension n over a number field K. Let S be a finite set of primes of $\mathcal{O}_K$ including all the places of bad reduction for A.*

Let ϕ be an ample class in the Neron-Severi group of A, and let $d = \phi^n/n!$.

Suppose that either $n \geq 4$ or $n = 3$ and d is not $\binom{a(i)+a(i+1)}{a(i+1)/6}$ for any $i \geq 2$.

Then there are only finitely many smooth primitive hypersurfaces $H \subseteq A$ representing ϕ, defined over K with good reduction outside S, up to translation.

Proof. Choose a smooth proper model of A over $\mathcal{O}_{K,S}$. Working over $\mathcal{O}_{K,S}$, let Hilb be the Hilbert scheme of smooth hypersurfaces of class ϕ, and let $H_{\text{univ}} \subseteq \text{Hilb} \times A$ be the universal family over Hilb. Let

$$\text{Hilb}(\mathcal{O}_{K,S})^{prim} \subseteq \text{Hilb}(\mathcal{O}_{K,S})$$

be the open subscheme of *primitive* hypersurfaces in A (Definition 11.1). (Note that, by definition, a hypersurface defined over K with good reduction outside S extends to an $\mathcal{O}_{K,S}$-point of Hilb.) The group A acts on Hilb by translation; we need to show that $\text{Hilb}(\mathcal{O}_{K,S})^{prim}$ is contained in the union of finitely many orbits of A.

Theorem 8.21 only works over $\mathbb{Q}$; we'll pass from K to $\mathbb{Q}$ by Weil restriction. (See [11, Theorem 7.6.4] for the notion of Weil restriction relative to a finite étale extension of rings.) Enlarging S if necessary, we can arrange that $\mathcal{O}_{K,S}$ is finite étale over $\mathbb{Z}[1/S']$, where S' is a finite set of primes of $\mathbb{Z}$. Over the Weil restriction

$$\text{Res}_{\mathbb{Z}[1/S']}^{\mathcal{O}_{K,S}} \text{Hilb}$$

we have the family

$$H_{\text{univ},\mathbb{Q}} = H_{\text{univ}} \times_{\text{Hilb}} (\text{Res}_{\mathbb{Z}[1/S']}^{\mathcal{O}_{K,S}} \text{Hilb} \times_{\mathbb{Z}[1/S']} \mathcal{O}_{K,S});$$

and the fibers of this family are the same as the fibers of the original family H_{univ}. More precisely, the sets $\text{Hilb}(\mathcal{O}_{K,S})$ and

$$\text{Res}_{\mathbb{Z}[1/S']}^{\mathcal{O}_{K,S}} \text{Hilb}(\mathbb{Z}[1/S'])$$

are in canonical bijection. Let

$$\text{Res}_{\mathbb{Z}[1/S']}^{\mathcal{O}_{K,S}} \text{Hilb}(\mathbb{Z}[1/S'])^{prim} \subseteq \text{Res}_{\mathbb{Z}[1/S']}^{\mathcal{O}_{K,S}} \text{Hilb}(\mathbb{Z}[1/S'])$$

be the subset corresponding to $\text{Hilb}(\mathcal{O}_{K,S})^{prim}$ under this bijection. For any $x \in \text{Hilb}(\mathcal{O}_{K,S})$, call $\text{Res}\, x$ the corresponding point of $\text{Res}_{\mathbb{Z}[1/S']}^{\mathcal{O}_{K,S}} \text{Hilb}(\mathbb{Z}[1/S'])$. Then we have a canonical isomorphism of schemes

$$(H_{\text{univ}})_x \cong (H_{\text{univ},\mathbb{Q}})_{\text{Res}\, x};$$

here $(H_{\mathrm{univ}})_x$ is an $\mathcal{O}_{K,S}$-scheme, $(H_{\mathrm{univ},\mathbb{Q}})_{\operatorname{Res} x}$ is a $\mathbb{Z}[1/S']$-scheme, and the structure maps are related by the fact that the diagram

$$\begin{tikzcd} (H_{\mathrm{univ}})_x \arrow[r, "\cong"] \arrow[d] & (H_{\mathrm{univ},\mathbb{Q}})_{\operatorname{Res} x} \arrow[d] \\ \operatorname{Spec} \mathcal{O}_{K,S} \arrow[r] & \operatorname{Spec} \mathbb{Z}[1/S'] \end{tikzcd} \tag{10}$$

commutes.

Let X_{sing} be an irreducible component of the Zariski closure of $\operatorname{Res}_{\mathbb{Q}}^K \operatorname{Hilb}(\mathbb{Z}[1/S])^{prim}$ in $\operatorname{Res}_{\mathbb{Q}}^K \operatorname{Hilb}$, and choose a resolution of singularities $X \to X_{sing}$. Let $Y = H_{\mathrm{univ},\mathbb{Q}} \times_{\operatorname{Res}_{\mathbb{Q}}^K \operatorname{Hilb}} X$ be the pullback of our family of hypersurfaces to X. Note that Y is a hypersuface in $X \times_{\mathbb{Q}} A = X_K \times_K A$.

We will apply Theorem 8.21 to show that $Y \in X_K \times_K A$ is the translate of a constant hypersurface $H_0 \in A$ by a section $s \in A(X_K)$. In this case, all the points of $\operatorname{Res}_{\mathbb{Q}}^K \operatorname{Hilb}(\mathbb{Z}[1/S])^{prim}$ lying in X would correspond to points in the orbit H_0.

The map $X \to X_{sing}$ is proper, so it spreads out to a proper map

$$\mathcal{X} \to \operatorname{Res}_{\mathbb{Z}[1/S']}^{\mathcal{O}_{K,S}} \operatorname{Hilb},$$

possibly after further enlarging the finite sets S and S'. Possibly after further enlargement of S and S', the family Y spreads out to a smooth proper family $\mathcal{Y} \to \mathcal{X}$, which is a hypersurface in $A \times_{\mathbb{Z}[1/S']} X$. Now, by properness, every S-integral point of Hilb lifts to an S-integral point of $\mathcal{X}$, so by construction the S-integral points are Zariski dense in $\mathcal{X}$.

Fix some $p \notin S'$. Assume that $Y \subseteq X_K \times_K A$ is not equal to the translate of a constant hypersurface $H_0 \in A$ by a section $s \in A(X_K)$. Let η be the generic point of X_K. The fibers of Y over points in a dense subset of X are primitive hypersurfaces, so $Y_{\overline{\eta}}$ is not translation-invariant by any nonzero element of A. Let G be the Tannakian group of the constant sheaf on $Y_{\overline{\eta}}$, and let G^* be the commutator subgroup of the identity component of A. By Theorem 4.1, because of our assumptions on n and d, we have $G^* = SL_N, Sp_N$, or SO_N, acting by its standard representation. Furthermore the Sp_N case occurs exactly when $Y_{\overline{\eta}}$ is equal to a translate of $[-1]^* Y_\eta$ and n is even and the SO_N case occurs exactly when $Y_{\overline{\eta}}$ is equal to a translate of $[-1]^* Y_\eta$ and n is even. In particular, G^* is a simple algebraic group acting by an irreducible representation.

We can now apply Corollary 5.9. This gives us the existence of an embedding $\iota\colon K \to \mathbb{C}$ and a torsion character χ of $\pi_1(A_\iota)$, satisfying the big monodromy condition that is needed for Lemma 7.6. In Section 6.3 and Lemma 6.28 we have constructed a Hodge–Deligne system V_I on $X_{\mathbb{Q}}$ attached to the orbit I containing (ι, χ) and the family $Y \subseteq X_K \times_K A$. By Lemma 7.6, the differential Galois group of V_I is a strongly c-balanced subgroup of G^0. Corollary 5.9 gives vanishing of cohomology outside degree $n-1$, so the Hodge numbers of V_I are given by Lemma 2.7.

We apply Theorem 8.21 to $\mathsf{V} = \mathsf{V}_I$. That the eigenvalues of Frobenius on V are integral Weil numbers is a consequence of the Weil numbers, since V comes from geometry; the polarization and integral structure are given in Lemma 6.28. Lemma 6.29 gives $\mathbf{H}$-structure, with $\mathbf{H}$ chosen as in Lemma 6.29. (Note that this matches the group G^* by our earlier calculation, because $Y_{\overline{\eta}}$ is equal to a translate of $[-1]^*Y_\eta$ if and only if Y is equal to a translate of $[-1]^*Y$.) Lemma 7.6 gives that G_{mon} is strongly c-balanced with respect to Frobenius. The Hodge numbers of V are uniform because we have explicitly computed them, independently of ι, in Lemma 2.7.

The numerical conditions are satisfied because every term except c is independent of c, so taking c arbitrarily large it suffices to have

$$\sum_{p>0} h^p > 0$$

$$\sum_{p>0} ph_p > T_G\left(\frac{1}{2}(h^0 - t)\right).$$

By Lemma 10.1, we have

$$\sum_{p>0} h_p > \frac{1}{2}(h^0 - t).$$

This implies the first condition because $t \leq h_0$. It implies the second condition because, by the definition of T_G, $T_G(x)$ is strictly increasing for $x \leq \sum_{p>0} h_p$ and $T_G(\sum_{p>0} h_p) = \sum_{p>0} ph_p$.

Hence the hypotheses of Theorem 8.21 are satisfied, showing that the integral points are not Zariski dense, contradicting our earlier assumption. $\square$

Corollary 11.3. *Let K, A, S, ϕ be as in Theorem 11.2.*

Suppose that either $n \geq 4$ or $n = 3$ and d is not a multiple of $\binom{a(i)+a(i+1)}{a(i+1)}/6$ for any $i \geq 2$

There are only finitely many smooth hypersurfaces $H \subseteq A$ representing ϕ, with good reduction outside S, up to translation.

Proof. Any hypersurface $H \subseteq A$ is of the form $\pi^{-1}H'$, where $\pi \colon A \to A'$ is an isogeny defined over K, and $H' \subseteq A'$ is a primitive hypersurface defined over K. In this case ϕ is the pullback of an ample class ϕ' along π, and we have $d = \phi^n/n! = (\phi'^n/n!)\cdot\deg\pi$. Hence $\deg\pi$ is bounded, so there are only finitely many possibilities for π. For each one there is a unique ϕ' with $\phi = \pi^*\phi'$. Furthermore, $d(\phi') \neq \binom{a(i)+a(i+1)}{a(i+1)}/6$ for any $i \geq 2$. We conclude by applying Theorem 11.2 to each (A', ϕ'). $\square$

Theorem 11.4. *Suppose* $\dim A \geq 4$. *Fix an ample class ϕ in the Neron-Severi group of A. There are only finitely many smooth hypersurfaces $H \subseteq A$ representing ϕ, with good reduction outside S, up to translation.*

Proof. This is one case of Corollary 11.3 . $\square$

Theorem 11.5. *Suppose* $\dim A = 3$. *Fix an ample class ϕ in the Neron-Severi group of A. Assume that the intersection number $\phi\cdot\phi\cdot\phi$ is not divisible by $d(i)$ for any $i \geq 2$. There are only finitely many smooth hypersurfaces $H \subseteq A$ representing ϕ, with good reduction outside S, up to translation.*

Proof. This is one case of Corollary 11.3, once we cancel the factor of $3! = 6$ from the denominators. $\square$

Theorem 11.6. *Suppose* $\dim A = 2$. *Fix an ample class ϕ in the Neron-Severi group of A. There are only finitely many smooth curves $C \subseteq A$ representing ϕ, with good reduction outside S, up to translation.*

Proof. This is a consequence of the Shafarevich conjecture for curves. By the Shafarevich conjecture, there are only finitely many possibilities for the isomorphism class of C. Consider a fixed C; enlarging K if necessary, we may assume that $C(K)$ is nonempty, and choose a basepoint $x_0 \in C(K)$. It is enough to show that there are only finitely many maps $C \to A$ taking x_0 to the origin of A, for which the image of C represents the class ϕ.

By the Albanese property, pointed maps $(C, x_0) \to (A, 0)$ are in bijection with maps of abelian varieties $\operatorname{Jac} C \to A$; the set of such maps forms a finitely generated free abelian group. Fix an ample class ψ on A; define the degree of any map $f\colon \operatorname{Jac} C \to A$ by

$$\deg f = f(C) \cdot \psi.$$

This intersection number is a positive definite quadratic form on $\operatorname{Hom}(\operatorname{Jac} C, A)$, so there are only finitely many maps of given degree, any any map representing ϕ has degree $\phi \cdot \psi$. $\square$

Appendix A. Combinatorics involving binomial coefficients and Eulerian numbers

This section is devoted to proving Proposition 4.11. Thus, throughout this section, we preserve the notation and assumptions of Proposition 4.11, which we review here.

We use $A(n,q)$ for the Eulerian numbers. We adopt the convention that $A(n,q) = 0$ unless $0 \leq q < n$; similarly, we take $\binom{n}{q}$ to vanish whenever n is positive but $q < 0$ or $q \geq n$.

Recall from the introduction that $a(i)$ is the sequence satisfying

$$a(1) = 1, a(2) = 5, a(i+2) = 4a(i+1) + 1 - a(i).$$

Proposition A.1 (Proposition 4.11)**.** *Let $n \geq 2$ and $d \geq 1$ be integers. Suppose that there exists a natural number k, function m_H from the integers to the natural numbers and an integer s. Write $m = \sum_i m_H(i)$, and suppose that $1 < k < m - 1$. Suppose the equation*

$$\sum_{\substack{m_S:\mathbb{Z}\to\mathbb{Z} \\ 0\leq m_S(i)\leq m_H(i) \\ \sum_i m_S(i)=k \\ \sum_i i m_S(i)=s+q}} \prod_i \binom{m_H(i)}{m_S(i)} = dA(n,q). \tag{11}$$

is satisfied for all $q \in \mathbb{Z}$. Then we have one of the cases

1. *$m = 4$ and $k = 2$*
2. *$n = 2$ and $d = \binom{2k-1}{k}$ for some $k > 2$*
3. *$n = 3$ and $d = \binom{a(i)+a(i+1)}{a(i)}/6$ for some $i \geq 2$.*

A tuple (n, d, m, k, m_H, s) satisfying the conditions of the proposition will be called a *solution*.

We first handle the cases $n = 2, 3, 4$ directly, then give a general argument that handles cases $n \geq 5$. Thus Proposition A.1 will follow immediately once we have proven Lemmas A.4 (the $n = 2$ case), A.8 (the $n = 3$ case), A.9 (the $n = 4$ case), A.18, A.21, A.25, A.27, and A.31 (which collectively handle the $n \geq 5$ case).

We can assume without loss of generality that $k \leq m/2$.

Let m_{max} and m_{min} be the functions $m_S : \mathbb{Z} \to \mathbb{Z}$ satisfying $0 \leq m_S(i) \leq m_H(i), \sum_i m_S(i) = k$, and maximizing, respectively, minimizing $\sum_i i m_S(i)$.

Lemma A.2. *There is a unique w such that $m_{min}(i) = m_H(i)$ for all $i < w$, $m_{min}(i) = 0$ for all $i > w$, and $m_{min}(i) > 0$ for $i = w$.*

Similarly, there is a unique w' such that $m_{max}(i) = m_H(i)$ for all $i > w'$, $m_{max} = 0$ for all $i < w'$ and $m_{max} > 0$ for $i = w'$.

Furthermore $w \leq w'$.

Proof. We take w to be the largest i such that $m_{min}(i) > 0$. The last two conditions are then obvious and the fact that $m_{min}(i) = m_H(i)$ for $i < w$ follows by minimality - if it were not so, we could increase $m_{min}(i)$ by 1, reduce $m_{min}(w)$ by 1, and thereby reduce $\sum_i i m_S(i)$ by $w - i$.

We take w' to be the least i such that $m_{max}(i) > 0$, and make a symmetrical argument.

Finally, for contradiction, assume $w > w'$. Then

$$\begin{aligned}\sum_i m_H(i) &= m \geq 2k = \sum_i m_{min}(i) + \sum_i m_{max}(i) \\ &> \sum_{i<w} m_{min}(i) + \sum_{i>w'} m_{max}(i) = \sum_{i<w} m_H(i) + \sum_{i>w'} m_H(i) \geq \sum_i m_H(i),\end{aligned}$$

a contradiction. ☐

Lemma A.3. *We have*

$$\sum_{i<w}(w-i)m_H(i) + \sum_{i>w'}(i-w')m_H(i) + k(w'-w) = n-1 \tag{12}$$

Note that all the terms on the left side of this equation are nonnegative.

Proof. Because the function $q \mapsto dE(n,q)$ is supported on q ranging from 0 to $n-1$, we must have

$$\begin{aligned} n-1 = (s+n-1)-s = \sum_i i m_{max}(i) - \sum_i i m_{min}(i) = \sum_i (i-w')m_{max}(i) - \sum_i (i-w)m_{min}(i) + k(w'-w) \\ = \sum_{i<w}(w-i)m_H(i) + \sum_{i>w'}(i-w')m_H(i) + k(w'-w)\end{aligned}$$

☐

A.1. **The case $n \leq 4$.**

Lemma A.4. *Suppose $n = 2$. Then we must have $d = \binom{2k-1}{k}$. Furthermore if $k = 2$ then $m = 4$.*

Proof. In (12), because the summands on the left side are nonnegative, one must be 1 and the others must vanish. The one that is 1 can only be the summand associated to $i = w-1$ or $i = w'+1$ as the other summands are integer multiples of something at least 2. By symmetry, we may assume the 1 comes from $i = w+1$. Because the last summand vanishes, we must have $w = w'$.

This gives $m_H(i) = 0$ unless $i = w$ or $w+1$ and $m_H(w+1) = 1$.

Then the only possible solutions to $0 \leq m_S(i) \leq m_H(i)$ and $\sum_i m_S(i) = k$ are $m_S(w) = k, m_S(w+1) = 0$ and $m_S(w) = k-1, m_S(w+1) = 1$. These have $\sum_i m_S(i)$ equal to kw and $kw+1$ respectively, so we must have $s = kw$. This implies

$$\begin{aligned}\binom{m_S(w)}{k} &= \binom{m_S(w)}{k}\binom{1}{0} = dE(2,0) = d \\ \binom{m_S(w)}{k-1} &= \binom{m_S(w)}{k-1}\binom{1}{1} = dE(2,1) = d\end{aligned}$$

and thus

$$\binom{m_S(w)}{k-1} = \binom{m_S(w)}{k}$$

which implies $m_S(w) = 2k-1$ and thus $d = \binom{2k-1}{k}$.

In the $k = 2$ case, we have $m = m_S(w) + m_S(w+1) = 2k = 4$, as desired.

☐

Lemma A.5. *Suppose $n = 3$. Then we must have $m_H(i) = 0$ for $i \neq w-1, w, w+1$ and $m_H(w-1) = m_H(w+1) = 1$ unless $m = 4$ and $k = 2$.*

Proof. In (12), either two terms are 1 and the rest zero or one term is 2 and the rest are zero. In the first case, since the only terms that can be 1 are $i = w-1$ and $i = w'+1$, implying in particular that $w = w'$ we have the stated conclusion. So it suffices to eliminate the case that one term is 2. The only possibilities are $i = w-2, w-1, w'+1, w'+2$, and the last term if $k = 2$. By symmetry, we are reduced to eliminating $i = w'+1$, $i = w'+2$, and the final term. In the first two cases we have $w' = w$ and in the last case we have $k = 2, w' - w = 1$.

If $m_H(i) = 0$ for all i except $w, w+1$, and $m_H(w+1) = 2$, then there are three possibilities for m_S: $(m_S(w), m_S(w+1))$ must equal $(k, 0), (k-1, 1)$, or $(k-2, 2)$. Using $m_H(w) = m-2$, this gives

$$\binom{m-2}{k} = \binom{m-2}{k}\binom{2}{0} = dE(3,0) = d$$

$$2\binom{m-2}{k-1} = \binom{m-2}{k-1}\binom{2}{1} = dE(3,1) = 4d$$

$$\binom{m-2}{k-2} = \binom{m-2}{k-2}\binom{2}{2} = dE(3,2) = d$$

and this implies

$$\frac{1}{2} = \frac{\binom{m-2}{k}}{\binom{m-2}{k-1}} = \frac{m-k-1}{k}$$

$$\frac{1}{2} = \frac{\binom{m-2}{k-2}}{\binom{m-2}{k-1}} = \frac{k-1}{m-k}$$

so $m - k = 2k - 2$ and $k = 2m - 2k - 2$ giving $m = 4, k = 2$.

If $m_H(i) = 0$ for all i except $w, w+2$, then $\sum_i i m_S(i) \equiv kw \mod 2$ whenever $0 \leq m_S(i) \leq m_H(i)$ and $\sum_i m_S(i) = k$. Hence the left side of (3) is nonzero only when $s + q \equiv kw \mod 2$. This contradicts the fact that $E(2, q)$ is nonzero for q of both parities.

If $m_H(i) = 0$ for all i except $w, w+1$ and $k = 2$, then there are three possibilities for m_S: $(m_S(w), m_S(w+1))$ must be $(2,0), (1,1)$, or $(0,2)$. This gives

$$\frac{m_H(w)(m_H(w)-1)}{2} = \binom{m_H(w)}{2}\binom{m_H(w+1)}{0} = dA(3,0) = d$$

$$m_H(w)m_H(w+1) = \binom{m_H(w)}{1}\binom{m_H(w+1)}{1} = dA(3,1) = 4d$$

$$\frac{m_H(w+1)(m_H(w+1)-1)}{2} = \binom{m_H(w)}{0}\binom{m_H(w+1)}{2} = dA(3,2) = d$$

Because $d > 0$ this implies $m_H(w) = m_H(w+1)$ and thus $2m_H(w)(m_H(w)-1)) = m_H(w)^2$ which implies $m_H(w) = m_H(w+1) = 2$ and thus $m = 4$, $k = 2$. ☐

Lemma A.6. *Suppose $n = 3$. Then unless $m = 4$ and $k = 2$, we have*

$$(m-k)^2 - 4(m-k)k + k^2 = m.$$

Proof. By Lemma A.5, we have $m_H(i) = 0$ for $i \neq w-1, w, w+1$ and $m_H(w-1) = m_H(w+1) = 1$. This means there are four possibilities for m_S: $(m_S(w-1), m_S(w), m_S(w+1))$ must equal $(0,k,0)$, $(0,k-1,1)$, $(1,k-1,0)$, or $(1,k-2,1)$. Using $m_H(w) = m-2$, and $\binom{1}{0} = \binom{1}{1} = 1$, we get

$$\binom{m-2}{k-1} = dE(3,0) = d$$

$$\binom{m-2}{k} + \binom{m-2}{k-2} = dE(3,1) = 4d$$

$$\binom{m-2}{k-1} = dE(3,2) = d$$

so in other words we have

$$\binom{m-2}{k} + \binom{m-2}{k-2} = 4\binom{m-2}{k-1}$$

which dividing by $(m-2)!$ and multiplying by $k!(m-k)!$ is

$$(m-k)(m-k-1) + k(k-1) = 4k(m-k)$$

which is exactly the stated Diophantine equation. □

Lemma A.7. *The positive integer solutions to $a^2 - 4ab + b^2 = a + b$ with $a \leq b$ have the form $(a,b) = (a(i), a(i+1))$ for some $i \geq 1$.*

Proof. Let a and b be positive integers with $a^2 - 4ab + b^2$ and $b \geq a$. We will show that there exists $i \geq 1$ such that $a = a(i)$ and $b = a(i+1)$. By induction, it suffices to prove that either $a = 1, b = 5$ or that there exists an integer b' with $0 < b' < a$, $b = 4a + 1 - b'$, and (b', a) solving the same equation, as if $b' = a(i'), a = a(i'+1)$ then $b = a(i'+2)$.

To do this, let $b' = 4a + 1 - b$. Then because we can rewrite the equation as

$$b(4a + 1 - b) = a^2 - a,$$

as b is a solution then b' is a solution as well. Furthermore, if $a > 1$ then $a^2 - a > 0$ so $b' = \frac{a^2-a}{b} > 0$, and we must have $b' < a$ because if $b' \geq a$ we have

$$a^2 - a = b(4a + 1 - b) \geq a \cdot a > a^2,$$

so if $a = 1$ we always have a solution. On the other hand, if $a = 1$ then $b(5-b) = 0$ so, because $b > 0$, we have $b = 5$, the base case. □

Lemma A.8. *Suppose $n = 3$. Then either $m = 4$ and $k = 2$ or $m = a(i) + a(i+1)$ and $k = a(i)$ for some $i \geq 2$.*

Proof. This follows from Lemma A.6 and A.7 once we observe that because $k \leq m/2$, we have $k \leq m - k$, and because $k \geq 2$, the case $k = a(1), m = a(2)$ cannot occur. □

Lemma A.9. *There are no solutions for $n = 4$.*

Proof. We first consider the contribution to (12) from $k(w' - w)$. Because $k \geq 2$ and this contribution is at most $(n-1) = 3$, we can only have $w' = w$ or $w' = w+1$, $k = 2$ or 3. Let us eliminate the $w' = w + 1$ cases first.

In the $k = 3$ case we have $m_H(i) = 0$ unless $i = w$ or $w + 1$. Thus we have four possibilities for m_S - we must have $(m_S(w), m_S(w+1)) = (3,0), (2,1), (1,2)$, or $(0,3)$. This gives

$$\binom{m_H(w)}{3}\binom{m_H(w+1)}{0} = dE(4,0) = d$$

$$\binom{m_H(w)}{2}\binom{m_H(w+1)}{1} = dE(4,1) = 11d$$

$$\binom{m_H(w)}{1}\binom{m_H(w+1)}{2} = dE(4,2) = 11d$$

$$\binom{m_H(w)}{0}\binom{m_H(w+1)}{3} = dE(4,3) = d$$

Combining the first and last equations with $d > 0$, we see that $m_H(w) = m_H(w+1)$. Dividing the second equation by the first, we get

$$\frac{3m_H(w)}{m_H(w) - 2} = 11$$

which implies $m_H(w) = \frac{11}{4}$, a contradiction.

In the $k = 2$ case, one more term must be 1, which is $i = w - 1$ or $i = w' + 1$. Without loss of generality, it is $i = w - 1$. Then we have $m_H(i) = 0$ unless $i = w - 1, w, w + 1$ and $m_H(w - 1) = 1$. Then the possible values of $(m_S(w-1), m_S(w), m_S(w+1))$ are $(1,1,0)$, $(1,0,1)$, $(0,2,0)$, $(0,1,1)$, and $(0,0,2)$. This gives (ignoring factors of the form $\binom{n}{0}$ or $\binom{1}{1}$)

$$\binom{m_H(w)}{1} = dE(4,0) = d$$

$$\binom{m_H(w+1)}{1} + \binom{m_H(w)}{2} = dE(4,1) = 11d$$

$$\binom{m_H(w)}{1}\binom{m_H(w+1)}{1} = dE(4,2) = 11d$$

$$\binom{m_H(w+1)}{2} = d$$

Dividing the third equation by the first, we get $m_H(w + 1) = 11$. Thus by the fourth equation $d = \binom{11}{2} = 55$, and by the first equation $m_H(w) = 55$. Then the second equation gives $11 + \binom{55}{2} = 11 \cdot 55$, which is false, so there are no solutions.

We now handle the case $w = w'$. In this case, because the total sum of (12) is 3, and only the two terms $i = w + 1, i = w - 1$ can contribute a 1, we must have one term contributing 3 or one 2 and one 1. There are four terms that might contribute 3: $i = w+3$, $i = w+1$, $i = w-1$, and $i = w-3$, and four that might contribute 2: $i = w+2$, $i = w+1$, $i = w-1$, and $i = w-2$. This gives a total of 10 possibilities for m_H, or 5 up to symmetry: we may assume $(m_H(w-1), m_H(w), m_H(w+1), m_H(w+2), m_H(w+3)) = (0, m-1, 0, 0, 1), (0, m-3, 3, 0, 0), (0, m-2, 1, 1, 0), (1, m-2, 0, 1, 0)$, or $(1, m-2, 2, 0, 0)$.

The case $(0, m-1, 0, 0, 1)$ is easy to eliminate as it implies that the left side of (3) is nonvanishing only in a single residue class mod 3, but we know that the right side does not have that possibility.

The case $(0, m-3, 3, 0, 0)$ gives us four possibilities for m_S, implying

$$\binom{m-3}{k}\binom{3}{0} = dE(4,0) = d$$
$$\binom{m-3}{k-1}\binom{3}{1} = dE(4,0) = 11d$$
$$\binom{m-3}{k-2}\binom{3}{2} = dE(4,0) = 11d$$
$$\binom{m-3}{k-3}\binom{3}{3} = dE(4,0) = d$$

The first and fourth equations gives $\binom{m-3}{k} = \binom{m-3}{k-3}$, which implies that $m-3-k = k-3$ or $m = 2k$. Dividing the third equation by the second, we then obtain

$$11 = \frac{\binom{2k-3}{k-2}}{\binom{2k-3}{k-3}}\frac{\binom{3}{2}}{\binom{3}{3}} = \frac{k}{k-2}\cdot 3$$

whose unique solution is $k = \frac{11}{4}$, a contradiction.

The case $(0, m-2, 1, 1, 0)$ gives us four possibilities for m_S, implying

$$\binom{m-2}{k}\binom{1}{0}\binom{1}{0} = dE(4,0) = d$$
$$\binom{m-2}{k-1}\binom{1}{1}\binom{1}{0} = dE(4,1) = 11d$$
$$\binom{m-2}{k-1}\binom{1}{0}\binom{1}{1} = dE(4,2) = 11d$$
$$\binom{m-2}{k-2}\binom{1}{1}\binom{1}{1} = dE(4,3) = d$$

By the first and fourth equations we have $\binom{m-2}{k} = \binom{m-2}{k-2}$ which implies $m-2-k = k-2$ or $m = 2k$. Dividing the second equation by the first we get $11 = \frac{k}{k-1}$ so $k = \frac{11}{10}$, a contradiction.

The case $(1, m-2, 0, 1, 0)$ gives us four possibilities for m_S, implying

$$\binom{1}{1}\binom{m-2}{k-1}\binom{1}{0} = dE(4,0) = d$$
$$\binom{1}{0}\binom{m-2}{k}\binom{1}{0} = dE(4,1) = 11d$$
$$\binom{1}{1}\binom{m-2}{k-2}\binom{1}{1} = dE(4,2) = 11d$$
$$\binom{1}{0}\binom{m-2}{k-1}\binom{1}{1} = dE(4,3) = d$$

By the second and third equations, we have $\binom{m-2}{k} = \binom{m-2}{k-2}$ which again implies $m = 2k$. But then $\binom{2k-2}{k-1} > \binom{2k-2}{k-2}$ which means $d > 11d$, a contradiction.

The case $(1, m-2, 2, 0, 0)$ has six possibilities for m_S, implying

$$\binom{1}{1}\binom{m-2}{k-1}\binom{2}{0} = dE(4,0) = d$$

$$\binom{1}{0}\binom{m-2}{k}\binom{2}{0}+\binom{1}{1}\binom{m-2}{k-2}\binom{2}{1}=dE(4,1)=11d$$
$$\binom{1}{1}\binom{m-2}{k-3}\binom{2}{2}+\binom{1}{0}\binom{m-2}{k-1}\binom{2}{1}=dE(4,2)=11d$$
$$\binom{1}{0}\binom{m-2}{k-2}\binom{2}{2}=dE(4,3)=d$$

By the first and fourth equations, we have $\binom{m-2}{k-1}=\binom{m-2}{k-2}$ which implies $m=2k-1$. But then since $\binom{2k-2}{k-3}\leq\binom{2k-2}{k-2}$ we have by the third and fourth equations

$$11d=\binom{2k-2}{k-3}+2\binom{2k-2}{k-1}\leq\binom{2k-2}{k-2}+2\binom{2k-2}{k-2}=3d$$

a contradiction. □

A.2. **The case $n\geq 5$: general setup.** We'll write $m_0=m_{min}$ for the rest of this section. The key equality for $q=0$ gives

$$d=\binom{m_H(w)}{m_0(w)}. \tag{13}$$

For any $m_S:\mathbb{Z}\to\mathbb{Z}$ as in Equation 11 we'll write

$$N(m_S)=N(m_H,m_S)=\prod_i\binom{m_H(i)}{m_S(i)}.$$

Now the key equality (11) becomes

$$\sum_{\substack{m_S:\mathbb{Z}\to\mathbb{Z}\\0\leq m_S(i)\leq m_H(i)\\\sum_i m_S(i)=k\\\sum_i im_S(i)=s+q}}\frac{N(m_S)}{N(m_0)}=A(n,q). \tag{14}$$

We're going to get a contradiction from combinatorial considerations involving the terms associated to small q in Equation (3).

By abuse of notation, we'll let $[i]$ denote the function $m_{\{i\}}$ taking the value 1 on i and zero elsewhere; so any function $\mathbb{Z}\to\mathbb{Z}$ is a linear combination of the elementary functions $[i]$.

There are at most two functions m_1 that contribute to the $q=1$ case in Equation (14). These are

$$m_1^a=m_0+[w]-[w-1]$$

and

$$m_1^b=m_0+[w+1]-[w].$$

We compute

$$\frac{N(m_1^a)}{N(m_0)}=\frac{m_H(w-1)(m_H(w)-m_0(w))}{(m_0(w)+1)}$$
$$\frac{N(m_1^b)}{N(m_0)}=\frac{m_H(w+1)m_0(w)}{(m_H(w)-m_0(w)+1)};$$

the $q=1$ case of Equation (14) gives

$$\frac{N(m_1^a)}{N(m_0)}+\frac{N(m_1^b)}{N(m_0)}=A(n,1).$$

Similarly, there are at most five nonzero terms in the $q = 2$ case Equation (14):

$$\begin{aligned} m_2^a &= m_0 + [w] - [w-2] \\ m_2^b &= m_0 + 2[w] - 2[w-1] \\ m_2^c &= m_0 + [w+1] - [w-1] \\ m_2^d &= m_0 + 2[w+1] - 2[w] \\ m_2^e &= m_0 + [w+2] - [w]. \end{aligned}$$

We have the following equalities.

$$\begin{aligned} \frac{N(m_2^a)}{N(m_0)} &= \frac{m_H(w-2)(m_H(w) - m_0(w))}{(m_0(w)+1)} \\ \frac{N(m_2^b)}{N(m_0)} &= \frac{m_H(w-1)(m_H(w-1)-1)(m_H(w)-m_0(w))(m_H(w)-m_0(w)-1)}{2(m_0(w)+1)(m_0(w)+2)} \\ \frac{N(m_2^c)}{N(m_0)} &= m_H(w-1)m_H(w+1) \\ \frac{N(m_2^d)}{N(m_0)} &= \frac{(m_H(w+1))(m_H(w+1)-1)(m_0(w))(m_0(w)-1)}{2(m_H(w)-m_0(w)+1)(m_H(w)-m_0(w)+2)} \\ \frac{N(m_2^e)}{N(m_0)} &= \frac{m_0(w)m_H(w+2)}{m_H(w)-m_0(w)+1}. \end{aligned}$$

Equation (14) gives

$$\sum_{*=a,b,c,d,e} \frac{N(m_2^*)}{N(m_0)} = A(n,2). \tag{15}$$

We conclude with a lemma that will be useful at several points in the argument.

Lemma A.10. *We have*

$$\frac{N(m_1^a)}{N(m_0)} + \frac{N(m_1^b)}{N(m_0)} = A(n,1)$$

and

$$\left(\frac{N(m_1^a)}{N(m_0)}\right)\left(\frac{N(m_1^b)}{N(m_0)}\right) < m_H(w-1)m_H(w+1).$$

In particular, if

$$m_H(w-1)m_H(w+1) < \frac{1}{4}A(n,1)^2,$$

let $\alpha < \beta$ be the real roots of

$$X^2 - A(n,1)X + m_H(w-1)m_H(w+1).$$

Then we have

$$\min\left(\frac{N(m_1^a)}{N(m_0)}, \frac{N(m_1^b)}{N(m_0)}\right) \le \alpha$$

and

$$\max\left(\frac{N(m_1^a)}{N(m_0)}, \frac{N(m_1^b)}{N(m_0)}\right) \ge \beta$$

Proof. The first equality is the $q = 1$ case of Equation (14). The second follows from the explicit formulas for the two quotients $\frac{N(m_1^*)}{N(m_0)}$.

The bounds in terms of α and β follow from the first two inequalities. $\square$

A.3. **The case $n \geq 5$, with $N(m_1^a)$ big.** In the following sections we will use without proof a number of inequalities involving Eulerian numbers. The proofs of such inequalities are routine; we discuss the technique in Section B.

Recall notation from the beginning of Section A, and the beginning of Section A.2.

Lemma A.11. *If $n \geq 5$ and*

$$\frac{N(m_1^a)}{N(m_0)} \geq \frac{A(n,1)}{2} \tag{16}$$

then $m_H(w-1) = 1$.

Proof. We'll estimate $N(m_2^b)$. The key point is that, combining (16) and (15), we have

$$\frac{N(m_2^b)N(m_0)}{N(m_1^a)^2} \leq 4\frac{A(n,2)}{A(n,1)^2}. \tag{17}$$

We will focus on proving a lower bound for the left side, assuming $m_H(w-1) > 1$, which which will give a contradiction for sufficiently large n.

By hypothesis, we have

$$\frac{A(n,1)}{2} \leq \frac{m_H(w-1)(m_H(w) - m_0(w))}{(m_0(w)+1)}.$$

By Lemma A.3 we have $m_H(w-1) \leq n-1$. Also, we know $m_0(w) \geq 1$, so

$$m_H(w) - m_0(w) \geq \frac{A(n,1)}{n-1}.$$

If $n \geq 6$, then $\frac{A(n,1)}{n-1} \geq 9$, which gives the bound

$$m_H(w) - m_0(w) - 1 \geq \frac{8}{9}(m_H(w) - m_0(w)) \tag{18}$$

Assuming $m_H(w-1) \geq 2$, we find

(19)

$$\begin{aligned}
&\frac{N(m_2^b)N(m_0)}{N(m_1^a)^2} \\
=&\frac{m_H(w-1)(m_H(w-1)-1)(m_H(w)-m_0(w))(m_H(w)-m_0(w)-1)}{2(m_0(w)+1)(m_0(w)+2)} \cdot \frac{(m_0(w)+1)^2}{m_H^2(w-1)(m_H(w)-m_0(w))^2} \\
=&\frac{1}{2} \cdot \frac{m_0(w)+1}{m_0(w)+2} \cdot \frac{m_H(w-1)-1}{m_H(w-1)} \cdot \frac{m_H(w)-m_0(w)-1}{m_H(w)-m_0(w)} \\
\geq&\frac{1}{2} \cdot \frac{2}{3} \cdot \frac{1}{2} \cdot \frac{m_H(w)-m_0(w)-1}{m_H(w)-m_0(w)} = \frac{1}{6} \cdot \frac{m_H(w)-m_0(w)-1}{m_H(w)-m_0(w)}
\end{aligned}$$

By combining this with (18), we conclude that

$$\frac{N(m_2^b)N(m_0)}{N(m_1^a)^2} \geq 4/27 \tag{20}$$

which combines with (17) to give

$$\frac{A(n,2)}{A(n,1)^2} \geq 1/27$$

which is impossible for $n \geq 11$. (See the discussion in Appendix B, and `bound-A3a` in the Python code.)

For smaller n, we do a more precise version of the above analysis. For $5 \leq n \leq 10$, we will improve on the bound (16) by using Lemma A.10. For $n = 5$, we will also need to replace the bound (18) by a slightly weaker one without the assumption $n \geq 6$. With these modifications, the argument will work for n from 5 to 10.

To apply Lemma A.10, we must give an upper bound for $m_H(w-1)m_H(w+1)$, showing that the two roots α, β are far apart. Consider

$$m_4^a = m_0 + 2[w+1] - 2[w-1].$$

We have

$$A(n,4) \geq \frac{N(m_4^a)}{N(m_0)} = \frac{[m_H(w+1)(m_H(w+1)-1)]\,[m_H(w-1)(m_H(w-1)-1)]}{4},$$

so (still assuming $m_H(w-1) \geq 2$) we conclude that

$$2A(n,4) \geq m_H(w+1)(m_H(w+1)-1).$$

For $5 \leq n \leq 10$ we have

$$m_H(w+1) \leq \sqrt{2A(n,4)} + 1 \leq \frac{1}{48}\frac{A(n,1)^2}{n-1}$$

(see `bound-A3b` in the Python code). Thus we have

$$m_H(w-1)m_H(w+1) \leq (n-1) \cdot \frac{1}{48}\frac{A(n,1)^2}{n-1} = \frac{1}{48}A(n,1)^2,$$

so by Lemma A.10,

$$\frac{N(m_1^a)}{N(m_0)} \geq \frac{45}{46}A(n,1). \tag{21}$$

When $6 \leq n \leq 10$ we have from above (20) that

$$\frac{N(m_2^b)N(m_0)}{N(m_1^a)^2} \geq 4/27,$$

so

$$\frac{A(n,2)}{A(n,1)^2} \geq \frac{4}{27} \cdot \left(\frac{45}{46}\right)^2,$$

which does not hold for any $6 \leq n \leq 10$.

Finally when $n = 5$ we use the estimate

$$m_H(w) - m_0(w) \geq \frac{A(n,1)}{n-1} = \frac{13}{2}$$

to deduce

$$m_H(w) - m_0(w) - 1 \geq \frac{11}{13}(m_H(w) - m_0(w)).$$

Now (19) implies

$$\frac{N(m_2^b)N(m_0)}{N(m_1^a)^2} \geq \frac{1}{6} \cdot \frac{11}{13};$$

combining this with (21), we get

$$\frac{A(n,2)}{A(n,1)^2} \geq \frac{1}{6} \cdot \frac{11}{13} \cdot \left(\frac{45}{46}\right)^2,$$

which is not true for $n = 5$.

Thus we arrive at a contradiction in every case. □

Lemma A.12. *If*

$$\frac{N(m_1^a)}{N(m_0)} \geq \frac{A(n,1)}{2}$$

then $m_H(w-2) \leq 2$.

Proof. Assume $m_H(w-2) \geq 3$, and consider

$$m_7^a = m_0 - 3[w-2] - [w-1] + 4[w].$$

We will show that $N(m_7^a)$ is too big.

First, note that we must have $n \geq 8$ by Lemma A.3.

By hypothesis, we have $\frac{N(m_1^a)}{N(m_0)} > \frac{A(n,1)}{2}$. From Lemma A.11, we know $m_H(w-1) = 1$, so

$$\frac{N(m_1^a)}{N(m_0)} = \frac{(m_H(w) - m_0(w))}{(m_0(w)+1)}$$

and thus

$$(m_H(w) - m_0(w)) \geq A(n,1).$$

Since $n \geq 8$, this gives

$$(m_H(w) - m_0(w)) \geq 247.$$

Now

$$\frac{N(m_7^a)}{N(m_0)} = C \prod_{i=1}^{4} \frac{(m_H(w) - m_0(w) - i + 1)}{(m_0(w)+i)},$$

where

$$C = \frac{m_H(w-2)(m_H(w-2)-1)(m_H(w-2)-2)(m_H(w-1))}{6} \geq 1.$$

Using the bounds $(m_H(w) - m_0(w) - 3) > .98(m_H(w) - m_0(w))$ and

$$\frac{(m_0(w)+1)^4}{(m_0(w)+1)(m_0(w)+2)(m_0(w)+3)(m_0(w)+4)} \geq \frac{2}{15},$$

we deduce that

$$\frac{N(m_7^a)}{N(m_0)} \geq \frac{.92A(n,1)^4}{120} \geq \frac{A(n,1)^4}{140}.$$

The inequality

$$A(n,7) < \frac{1}{140} A(n,1)^4$$

(see Appendix B, and `bound-A3c` in the Python code) gives a contradiction. □

Lemma A.13. *If* $n \geq 5$ *and*

$$\frac{N(m_1^a)}{N(m_0)} \geq \frac{A(n,1)}{2}$$

then

$$m_H(w+1)(m_H(w+1)-1) \leq \frac{48}{11} A(n,3) \frac{N(m_1^a)}{N(m_0)} \leq \frac{48}{11} A(n,3) A(n,1).$$

Furthermore, if $m_0(w) \geq 2$ *then we have the stronger bound*

$$m_H(w+1)(m_H(w+1)-1) \leq \frac{36}{11} A(n,3) \frac{N(m_1^a)}{N(m_0)} \leq \frac{36}{11} A(n,3) A(n,1).$$

Proof. Let

$$m_3^a = m_0 + 2[w+1] - [w] - [w-1].$$

The result will follow from

$$\frac{N(m_3^a)}{N(m_0)} \leq A(n,3).$$

We have

$$A(n,3) \geq \frac{N(m_3^a)}{N(m_0)} = \frac{m_H(w+1)(m_H(w+1)-1)m_H(w-1)m_0(w)}{2(m_H(w)-m_0(w)+1)}.$$

On the other hand,

$$\frac{A(n,1)}{2} \leq \frac{N(m_1^a)}{N(m_0)} = \frac{m_H(w-1)(m_H(w)-m_0(w))}{(m_0(w)+1)}.$$

We know $m_H(w-1) = 1$ by Lemma A.11, so $m_H(w) - m_0(w) \geq A(n,1) \geq 11$. Thus we can estimate

$$\frac{m_0(w)}{2(m_H(w)-m_0(w)+1)} \cdot \frac{(m_H(w)-m_0(w))}{(m_0(w)+1)} \geq \frac{1}{4} \cdot \frac{11}{12} = \frac{11}{48}.$$

The first bound follows.

To get the second bound, note that if $m_0(w) \geq 2$ then

$$\frac{m_0(w)}{m_0(w)+1} \geq \frac{2}{3},$$

so

$$\frac{m_0(w)}{2(m_H(w)-m_0(w)+1)} \cdot \frac{(m_H(w)-m_0(w))}{(m_0(w)+1)} \geq \frac{1}{3} \cdot \frac{11}{12} = \frac{11}{36}.$$

□

Lemma A.14. *If $n \geq 5$ and*

$$\frac{N(m_1^a)}{N(m_0)} \geq \frac{A(n,1)}{2}$$

then

$$m_H(w+2)(m_H(w+2)-1) \leq \frac{48}{11} A(n,5) \frac{N(m_1^a)}{N(m_0)} \leq \frac{48}{11} A(n,5)A(n,1).$$

Furthermore, if $m_0(w) \geq 2$ then we have the stronger bound

$$m_H(w+2)(m_H(w+2)-1) \leq \frac{36}{11} A(n,5) \frac{N(m_1^a)}{N(m_0)} \leq \frac{36}{11} A(n,5)A(n,1).$$

Proof. Let

$$m_5^a = m_0 + 2[w+2] - [w] - [w-1].$$

The result follows from

$$\frac{N(m_5^a)}{N(m_0)} \leq A(n,5);$$

the proof is exactly analogous to the proof of Lemma A.13. □

Lemma A.15. *If*

$$\frac{N(m_1^a)}{N(m_0)} \geq \frac{A(n,1)}{2}$$

then $n \leq 10$.

Proof. Assume $n \geq 5$. We'll estimate each $\frac{N(m_2^*)}{N(m_0)}$, and show that for large n their sum is too small.

We have the following bounds. By Lemma A.12, we have $m_H(w-2) \leq 2$, so

$$\frac{N(m_2^a)}{N(m_0)} \leq 2\frac{N(m_1^a)}{N(m_0)} \leq 2A(n,1).$$

Lemma A.11 tells us that $m_H(w-1) = 1$ (if $n \geq 5$), so

$$\frac{N(m_2^b)}{N(m_0)} = 0.$$

By Lemma A.13, we know that

$$m_H(w+1)(m_H(w+1)-1) \leq \frac{48}{11}A(n,3)\frac{N(m_1^a)}{N(m_0)} \leq \frac{48}{11}A(n,3)A(n,1),$$

so

$$\frac{N(m_2^c)}{N(m_0)} = m_H(w-1)m_H(w+1) = m_H(w+1) \leq \sqrt{\frac{48}{11}A(n,3)A(n,1)} + 1$$

and

$$\begin{aligned}\frac{N(m_2^d)}{N(m_0)} &= \frac{(m_H(w+1))(m_H(w+1)-1)(m_0(w))(m_0(w)-1)}{2(m_H(w)-m_0(w)+1)(m_H(w)-m_0(w)+2)} \\ &\leq \frac{1}{2}\left(\frac{48}{11}A(n,3)\frac{N(m_1^a)}{N(m_0)}\right)\left(\frac{N(m_1^a)}{N(m_0)}\right)^{-2} \\ &\leq \frac{48}{11}\frac{A(n,3)}{A(n,1)}.\end{aligned}$$

Finally, from Lemma A.14, we find

$$m_H(w+2) \leq \sqrt{\frac{48}{11}A(n,5)\frac{N(m_1^a)}{N(m_0)}} + 1$$

so

$$\begin{aligned}\frac{N(m_2^e)}{N(m_0)} &= \frac{m_0(w)m_H(w+2)}{m_H(w)-m_0(w)+1} \\ &\leq \left(\frac{N(m_1^a)}{N(m_0)}\right)^{-1}\left(\sqrt{\frac{48}{11}A(n,5)\frac{N(m_1^a)}{N(m_0)}} + 1\right) \\ &\leq \sqrt{\frac{96}{11}\frac{A(n,5)}{A(n,1)}} + 1.\end{aligned}$$

These five bounds combine (see Appendix B, and `bound-A3d` in the Python code) to give

$$\frac{N(m_2^a)}{N(m_0)} + \frac{N(m_2^b)}{N(m_0)} + \frac{N(m_2^c)}{N(m_0)} + \frac{N(m_2^d)}{N(m_0)} + \frac{N(m_2^e)}{N(m_0)} < A(n,2)$$

for $n \geq 11$, a contradiction. ∎

Lemma A.16. *If* $5 \leq n \leq 19$ *and*

$$\frac{N(m_1^a)}{N(m_0)} \geq \frac{A(n,1)}{2}$$

then $m_0(w-1)=1$ *and* $m_0(w-s)=0$ *for* $s \geq 2$.

Proof. We have already seen in Lemma A.11 that $m_0(w-1)=1$. Suppose for a contradiction that $m_0(w-s)>0$ for some $s \geq 2$. Let

$$m_{s+3} = m_0 + 2[w+1] - [w-1] - [w-s].$$

Then

$$n! \geq A(n, s+3) \geq \frac{N(m_{s+3})}{N(m_0)} \geq \frac{m_H(w+1)(m_H(w+1)-1)}{2},$$

so

$$m_H(w+1) \leq \sqrt{2n!}+1.$$

We will use this stronger bound to redo the estimates in Lemma A.15.

We have

$$\frac{N(m_2^c)}{N(m_0)} = m_H(w+1) \leq \sqrt{2n!}+1$$

and

$$\begin{aligned} \frac{N(m_2^d)}{N(m_0)} &= \frac{(m_H(w+1))(m_H(w+1)-1)(m_0(w))(m_0(w)-1)}{2(m_H(w)-m_0(w)+1)(m_H(w)-m_0(w)+2)} \\ &\leq \frac{1}{2} m_H(w+1)^2 \left(\frac{N(m_1^a)}{N(m_0)}\right)^{-2} \\ &\leq \frac{1}{2}\left(\sqrt{2n!}+1\right)^2 \left(\frac{A(n,1)}{2}\right)^{-2}. \end{aligned}$$

We use the same bounds on

$$\frac{N(m_2^a)}{N(m_0)}, \frac{N(m_2^b)}{N(m_0)}, \frac{N(m_2^e)}{N(m_0)}$$

as in Lemma A.15; we conclude (see `bound-A3e` in the Python code) that

$$\frac{N(m_2^a)}{N(m_0)} + \frac{N(m_2^b)}{N(m_0)} + \frac{N(m_2^c)}{N(m_0)} + \frac{N(m_2^d)}{N(m_0)} + \frac{N(m_2^e)}{N(m_0)} < A(n,2)$$

for $6 \leq n \leq 19$.

Finally, suppose $n=5$. Then since $s+3 \geq 4$, we have the much stronger bound $A(n, s+3) \leq 1$, which implies $m_H(w+1) \leq 2$. We deduce as above

$$\frac{N(m_2^c)}{N(m_0)} = m_H(w+1) \leq 2$$

$$\frac{N(m_2^d)}{N(m_0)} \leq \left(\frac{N(m_1^a)}{N(m_0)}\right)^{-2} < 1$$

and arrive at a contradiction as before. (See `bound-A3f` in the Python code.) ☐

Lemma A.17. *If* $n \geq 5$ *and*

$$\frac{N(m_1^a)}{N(m_0)} \geq \frac{A(n,1)}{2}$$

then $m_0(w)=1$.

Proof. Suppose $m_0(w) \geq 2$. Again, we'll redo the estimates in Lemma A.15.

Since $n \geq 5$ and (by Lemma A.15) $n \leq 10$, Lemma A.16 implies that

$$\frac{N(m_2^a)}{N(m_0)} = \frac{N(m_2^b)}{N(m_0)} = 0.$$

Lemmas A.13 and A.14 give stronger bounds in case $m_0(w) \geq 2$. As in Lemma A.15, we deduce in this case that:

$$\frac{N(m_2^c)}{N(m_0)} \leq \sqrt{\frac{36}{11}A(n,3)A(n,1)} + 1$$

$$\frac{N(m_2^d)}{N(m_0)} \leq \frac{36}{11}\frac{A(n,3)}{A(n,1)}$$

$$\frac{N(m_2^e)}{N(m_0)} \leq \sqrt{\frac{72}{11}\frac{A(n,5)}{A(n,1)}} + 1.$$

Again, we conclude for n in the given range that

$$\frac{N(m_2^a)}{N(m_0)} + \frac{N(m_2^b)}{N(m_0)} + \frac{N(m_2^c)}{N(m_0)} + \frac{N(m_2^d)}{N(m_0)} + \frac{N(m_2^e)}{N(m_0)} < A(n,2),$$

a contradiction. (See Appendix B, and `bound-A3g` in the Python code.) $\square$

Lemma A.18. *We cannot have $n \geq 5$ and*

$$\frac{N(m_1^a)}{N(m_0)} \geq \frac{A(n,1)}{2}.$$

Proof. Again, we'll redo the estimates in Lemma A.15, in light of everything we now know. By Lemma A.15, we may assume that $n \leq 11$.

Now Lemma A.16 implies that

$$\frac{N(m_2^a)}{N(m_0)} = \frac{N(m_2^b)}{N(m_0)} = 0,$$

while Lemma A.17 gives us

$$\frac{N(m_2^d)}{N(m_0)} = 0.$$

Yet again (see proof of Lemma A.15) we have the bounds

$$\frac{N(m_2^c)}{N(m_0)} \leq \sqrt{\frac{48}{11}A(n,3)A(n,1)} + 1$$

and

$$\frac{N(m_2^e)}{N(m_0)} \leq \sqrt{\frac{96}{11}\frac{A(n,5)}{A(n,1)}} + 1.$$

Yet again, we conclude that

$$\frac{N(m_2^a)}{N(m_0)} + \frac{N(m_2^b)}{N(m_0)} + \frac{N(m_2^c)}{N(m_0)} + \frac{N(m_2^d)}{N(m_0)} + \frac{N(m_2^e)}{N(m_0)} < A(n,2),$$

a contradiction. (See Appendix B, and `bound-A3h` in the Python code.) $\square$

A.4. **The case $n \geq 5$, with $k \geq n$.** Recall notation from the beginning of Section A, and the beginning of Section A.2.

We'll treat the case where $k \geq n$ next. By Lemma A.3, if $k \geq n$, then $w = w'$.

Lemma A.19. *If $k \geq n$ and $n \geq 5$ then one of the two ratios*

$$\frac{N(m_1^a)}{N(m_0)}, \frac{N(m_1^b)}{N(m_0)}$$

is less than 1, and the other is greater than $A(n,1)-1$.

Proof. From Lemma A.3, we have

$$m_H(w-1) + m_H(w+1) \leq n-1$$

so

$$m_H(w-1)m_H(w+1) \leq \frac{(n-1)^2}{2}.$$

Then apply Lemma A.10, and the inequality

$$\frac{(n-1)^2}{2} < A(n,1) - 1$$

(see Appendix B). □

Lemma A.20. *If $n \geq 5$ and*

$$\frac{N(m_1^a)}{N(m_0)} > A(n,1) - 1$$

then $m_H(w-1) \leq 1$.

If $n \geq 5$ and

$$\frac{N(m_1^b)}{N(m_0)} > A(n,1) - 1$$

then $m_H(w+1) \leq 1$.

Proof. The first case follows from Lemma A.11; we'll prove the second. (As an alternative to Lemma A.11, the first case could be proven by an argument analagous to the argument below.)

So suppose $\frac{N(m_1^b)}{N(m_0)} > A(n,1) - 1$ and $m_H(w-1) \geq 2$, and consider

$$\frac{N(m_2^d)}{N(m_1^b)} = \frac{(m_H(w+1)-1)(m_0(w)-1)}{2(m_H(w)-m_0(w)+2)}.$$

$$\frac{N(m_2^d)N(m_0)}{N(m_1^b)^2} = \frac{1}{2} \cdot \frac{(m_H(w+1)-1)}{m_H(w+1)} \cdot \frac{(m_H(w)-m_0(w)+1)}{(m_H(w)-m_0(w)+2)} \cdot \frac{(m_0(w)-1)}{m_0(w)}.$$

From

$$\frac{(m_0(w))}{(m_H(w)-m_0(w)+1)} = \frac{1}{m_H(w+1)} \cdot \frac{N(m_1^b)}{N(m_0)} \geq \frac{A(n,1)-1}{n-1} > 6$$

we deduce that $m_0(w) > 6$, so by integrality $m_0(w) \geq 7$, so

$$m_0(w) - 1 > \frac{6}{7} m_0(w).$$

We also know that $(m_H(w) - m_0(w) + 2) \leq 2(m_H(w) - m_0(w) + 1)$ and $(m_H(w+1) - 1) \geq \frac{1}{2} m_H(w+1)$, so

$$\frac{A(n,2)}{(A(n,1)-1)^2} \geq \frac{N(m_2^d)N(m_0)}{N(m_1^b)^2} \geq \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{6}{7} = \frac{3}{28}.$$

For $n \geq 5$, this contradicts

$$A(n,2) < \frac{3}{28}(A(n,1)-1)^2.$$

(See Appendix B and `bound-A4a` in the Python code.) ☐

Lemma A.21. *For $n \geq 5$ we cannot have $k \geq n$.*

Proof. Assume $k \geq n$.

We'll consider the two cases given in Lemma A.19. The first case is ruled out by Lemma A.18; we'll prove the second. (Alternatively, the first case could be proven by an argument analagous to the argument below.)

So, suppose

$$\frac{N(m_1^b)}{N(m_0)} > A(n,1) - 1.$$

From

$$\frac{m_0(w)}{m_H(w) - m_0(w) + 1} > A(n,1) - 1$$

and the bounds

$$m_H(w-2), m_H(w-1) \leq n - 1 < A(n,1) - 1$$

we deduce that

$$\frac{N(m_2^a)}{N(m_0)} < m_H(w-2)\left(\frac{m_0(w)}{m_H(w) - m_0(w) + 1}\right)^{-1} < 1$$

and

$$\frac{N(m_2^b)}{N(m_0)} < \frac{1}{2} m_H(w-1)^2 \left(\frac{m_0(w)}{m_H(w) - m_0(w) + 1}\right)^{-2} < 1.$$

By Lemma A.20, we have $m_H(w+1) \leq 1$, so

$$\frac{N(m_2^c)}{N(m_0)} = m_H(w-1)m_H(w+1) \leq n - 1.$$

Also, $m_H(w+1) \leq 1$ implies that

$$\frac{N(m_2^d)}{N(m_0)} = 0.$$

Finally, Lemma A.3 gives $m_H(w+2) \leq \frac{n-1}{2}$, so

$$\frac{N(m_2^e)}{N(m_0)} \leq \frac{(n-1)}{2} A(n,1).$$

Adding these bounds, we see that

$$A(n,2) = \sum_{*=a,b,c,d,e} \frac{N(m_2^*)}{N(m_0)} \leq n + 1 + \frac{(n-1)}{2} A(n,1).$$

But this contradicts the inequality

$$n + 1 + \frac{(n-1)}{2} A(n,1) < A(n,2),$$

which holds for $n \geq 5$ (see Appendix B and `bound-A4b` in the Python code). $\square$

A.5. **The case $n \geq 5$, with $N(m_1^b)$ big and $k < n$.** Recall notation from the beginning of Section A, and the beginning of Section A.2.

Lemma A.22. *Suppose $k \leq n-1$ and*
$$\frac{N(m_1^b)}{N(m_0)} \geq A(n,1)/2.$$
Then
$$m_H(w+1) \geq \frac{A(n,1)}{2(n-1)}.$$
In particular, if $n \geq 5$ then
$$m_H(w+1) \geq 4.$$

Proof. We have
$$m_H(w+1) \geq \frac{N(m_1^b)}{N(m_0)} \cdot \frac{1}{m_0(w)} \geq \frac{A(n,1)}{2m_0(w)} \geq \frac{A(n,1)}{2(n-1)}.$$

The "in particular" follows from the fact that $m_H(w+1)$ is an integer. $\square$

Lemma A.23. *Suppose $n \geq 5$, $k \leq n-1$, and*
$$\frac{N(m_1^b)}{N(m_0)} \geq A(n,1)/2.$$
Then we cannot have simultaneously $\frac{N(m_1^a)}{N(m_0)} > 0$ and $m_0(w) \geq 2$.

Proof. Suppose $\frac{N(m_1^a)}{N(m_0)} > 0$ and $m_0(w) \geq 2$; the first condition implies $m_H(w-1) \geq 1$ and $m_H(w) - m_0(w) \geq 1$. Consider
$$m_4^b = m_0 - [w-1] - 2[w] + 3[w+1].$$
We have
$$\frac{N(m_4^b)}{N(m_1^b)} = \frac{m_H(w-1)(m_H(w+1)-1)(m_H(w+1)-2)(m_0(w)-1)}{6(m_H(w)-m_0(w)+2)}.$$

By Lemma A.10 we have
$$\left(\frac{N(m_1^a)}{N(m_0)}\right)\left(\frac{N(m_1^b)}{N(m_0)}\right) < m_H(w-1)m_H(w+1).$$

Thus we obtain
$$\frac{N(m_4^b)}{N(m_1^b)} > \frac{1}{6}\left(\frac{N(m_1^a)}{N(m_0)}\right)\left(\frac{N(m_1^b)}{N(m_0)}\right)^2 \frac{(m_H(w+1)-1)(m_H(w+1)-2)}{m_H(w+1)^2} \frac{m_H(w)-m_0(w)+1}{m_H(w)-m_0(w)+2} \frac{m_0(w)-1}{m_0(w)}.$$
Using $m_H(w+1) \geq 4$, $m_H(w) - m_0(w) \geq 1$, and $m_0(w) \geq 2$, the three fractional factors on the right can be bounded below by $\frac{3}{8}$, $\frac{2}{3}$, and $\frac{1}{2}$, respectively.

On the other hand, as soon as $N(m_1^a)$ is nonzero, we have
$$\frac{N(m_1^a)}{N(m_0)} = \frac{m_H(w-1)(m_H(w)-m_0(w))}{m_0(w)+1} \geq \frac{1}{n}.$$
We conclude that
$$\frac{N(m_4^b)}{N(m_1^b)} > \frac{1}{6}\left(\frac{N(m_1^a)}{N(m_0)}\right)\left(\frac{N(m_1^b)}{N(m_0)}\right)^2 \frac{3}{8}\frac{2}{3}\frac{1}{2} = \frac{1}{48}\left(\frac{N(m_1^a)}{N(m_0)}\right)\left(A(n,1) - \frac{N(m_1^a)}{N(m_0)}\right)^2$$

so

$$\frac{N(m_4^b)}{N(m_0)} > \frac{1}{48}\left(\frac{N(m_1^a)}{N(m_0)}\right)\left(A(n,1) - \frac{N(m_1^a)}{N(m_0)}\right)^3$$

As a function of $\frac{N(m_1^a)}{N(m_0)}$, this right-hand side is minimized when $\frac{N(m_1^a)}{N(m_0)} = \frac{1}{n}$, so we have

$$A(n,4) \geq \frac{N(m_4^b)}{N(m_1^b)} > \frac{1}{48n}\left(A(n,1) - \frac{1}{n}\right)^3.$$

This contradicts the inequality

$$A(n,4) < \frac{1}{78n}(A(n,1)-1)^3,$$

which is valid for all n. (See Appendix B and `bound-A5a` in the Python code.) □

Lemma A.24. *If* $n \geq 5$, $\frac{N(m_1^a)}{N(m_0)} = 0$, $k \leq n-1$, *and*

$$\frac{N(m_1^b)}{N(m_0)} \geq A(n,1)/2,$$

then $m_0(w) = 1$ *and* $N(m_2^d) = 0$.

Proof. Since $N(m_1^a) = 0$, we have

$$\frac{N(m_1^b)}{N(m_0)} = A(n,1).$$

Assuming $m_0(w) \geq 2$, let's look at m_2^d. We have

$$\frac{N(m_2^d)}{N(m_0)} = \frac{(m_H(w+1))(m_H(w+1)-1)(m_0(w))(m_0(w)-1)}{2(m_H(w)-m_0(w)+1)(m_H(w)-m_0(w)+2)}.$$

We want to compare this with the inequality

$$\frac{N(m_2^d)}{N(m_0)} \leq A(n,2) < \frac{1}{10}A(n,1)^2,$$

which is valid for all n. (See Appendix B and `bound-A5b` in the Python code.)

We have certainly

$$\frac{N(m_2^d)}{N(m_0)} \geq \frac{1}{8}\cdot\frac{(m_H(w+1)-1)}{m_H(w+1)}\left(\frac{N(m_1^b)}{N(m_0)}\right)^2 = \frac{1}{8}\cdot\frac{(m_H(w+1)-1)}{m_H(w+1)}A(n,1)^2.$$

So we will be done if $m_H(w+1) \geq 5$.

But as in the proof of Lemma A.22, we find (using now $\frac{N(m_1^b)}{N(m_0)} = A(n,1)$ in place of a weaker bound) that

$$m_H(w+1) \geq \frac{A(n,1)}{(n-1)},$$

and so in particular $m_H(w+1) > 5$ whenever $n \geq 5$. □

Lemma A.25. *If* $n \geq 5$, $k \leq n-1$, *and*

$$\frac{N(m_1^b)}{N(m_0)} \geq A(n,1)/2,$$

then $m_0(w) = 1$.

In particular, there must be some $r > 0$ *with* $m_0(w-r) > 0$.

Proof. That $m_0(w) = 1$ follows from Lemmas A.23 and A.24; the second claim follows because $k > 1$. □

Let $w - r$ be the highest Hodge weight below w. In other words, take $r > 0$ minimal such that $m_H(w - r) > 0$. (Such r exists by Lemma A.25.) For example, if $N(m_1^a) \neq 0$ then we must have $m_H(w - 1) \neq 0$, so $r = 1$.

A.5.1. *Case: m_1^b big and $r = 1$.* Recall the definitions of m_2^* from above ($* = a, b, c, d, e$).

Lemma A.26. *If $k \leq n - 1$,*

$$\frac{N(m_1^b)}{N(m_0)} \geq A(n, 1)/2,$$

and $m_0(w) = 1$, then $m_0(w - 1) \leq 1$.

Proof. Suppose for a contradiction $m_0(w - 1) \geq 2$, and consider

$$m_5^b = m_0 + 3[w + 1] - [w] - 2[w - 1].$$

We have

$$\begin{aligned}\frac{N(m_5^b)}{N(m_1^b)} &= \left(\frac{m_H(w-1)(m_H(w-1)-1)}{2}\right)\left(\frac{(m_H(w+1)-1)(m_H(w+1)-2)}{6}\right) \\ &\geq \frac{(m_H(w+1)-1)(m_H(w+1)-2)}{6}.\end{aligned}$$

From

$$\frac{N(m_1^b)}{N(m_0)} = \frac{m_H(w+1)m_0(w)}{m_H(w) - m_0(w) + 1}$$

and $m_0(w) = 1$, we find that

$$m_H(w + 1) \geq A(n, 1)/2.$$

In particular, since $n \geq 5$, we have $m_H(w + 1) \geq 13$, so certainly

$$(m_H(w+1) - 1)(m_H(w+1) - 2) \geq \frac{1}{2} m_H(w+1)^2.$$

We conclude that

$$\frac{N(m_5^b)}{N(m_0)} = \left(\frac{N(m_5^b)}{N(m_1^b)}\right)\left(\frac{N(m_1^b)}{N(m_0)}\right) \geq \frac{1}{12} m_H(w+1)^2 \frac{A(n,1)}{2} \geq \frac{A(n,1)^3}{96}.$$

This contradicts the inequality

$$A(n, 1)^3 > 341 A(n, 5),$$

which is valid for all n. (See Appendix B and `bound-A5c` in the Python code.) □

Lemma A.27. *We cannot have $n \geq 5$, $k \leq n - 1$,*

$$\frac{N(m_1^b)}{N(m_0)} \geq A(n, 1)/2,$$

and $r = 1$.

Proof. Assume the stated conditions hold. We will bound the five ratios $\frac{N(m_2^*)}{N(m_0)}$, and show that their sum is too small. Lemmas A.25 and A.26 tell us that $m_0(w-1) = m_0(w) = 1$. Hence, $N(m_2^b) = N(m_2^d) = 0$.

We have

$$\frac{N(m_2^a)}{N(m_0)} = \frac{m_0(w-2)(m_H(w)-1)}{2}$$
$$\frac{N(m_2^c)}{N(m_0)} = m_H(w+1)$$
$$\frac{N(m_2^e)}{N(m_0)} = \frac{m_H(w+2)}{m_H(w)}.$$

Since

$$\frac{A(n,1)}{2} \geq \frac{N(m_1^a)}{N(m_0)} = \frac{m_H(w)-1}{2}$$

and $m_0(w-2) \leq n-1$, we have

$$\frac{N(m_2^a)}{N(m_0)} \leq \frac{(n-1)A(n,1)}{2}.$$

Let

$$m_3^b = m_0 + 2[w+1] - [w] - [w-1].$$

Dividing the bound

$$A(n,3) \geq \frac{N(m_3^b)}{N(m_0)} = \frac{(m_H(w+1))(m_H(w+1)-1)}{2(m_H(w)-m_0(w)+1)}.$$

by

$$\frac{N(m_1^b)}{N(m_0)} = \frac{m_H(w+1)m_0(w)}{(m_H(w)-m_0(w)+1)} \geq \frac{A(n,1)}{2},$$

we obtain

$$(m_H(w+1)-1) \leq \frac{4A(n,3)}{A(n,1)},$$

so

$$\frac{N(m_2^c)}{N(m_0)} \leq \frac{4A(n,3)}{A(n,1)} + 1.$$

Finally, let

$$m_5^c = m_0 + 2[w+2] - [w] - [w-1].$$

We have

$$A(n,5) \geq \frac{N(m_5^c)}{N(m_0)} = \frac{m_H(w+2)(m_H(w+2)-1)}{2m_H(w)} \geq \frac{1}{2}\left(\frac{N(m_2^e)}{N(m_0)} - 1\right)^2,$$

so

$$\frac{N(m_2^e)}{N(m_0)} \leq \sqrt{2A(n,5)} + 1.$$

We conclude that

$$A(n,2) = \sum_{*=a,b,c} \frac{N(m_2^*)}{N(m_0)} \leq \left(\frac{nA(n,1)}{2}\right) + \left(\frac{4A(n,3)}{A(n,1)} + 1\right) + \left(\sqrt{2A(n,5)} + 1\right),$$

which contradicts the inequality

$$A(n,2) > \left(\frac{(n-1)A(n,1)}{2}\right) + \left(\frac{4A(n,3)}{A(n,1)} + 1\right) + \left(\sqrt{2A(n,5)} + 1\right),$$

valid for all $n \geq 5$. (See Appendix B and `bound-A5d` in the Python code.) $\square$

A.5.2. *Case: $N(m_1^b)$ big and $r \geq 2$.* Suppose $n \geq 5$ and

$$\frac{N(m_1^b)}{N(m_0)} \geq A(n,1)/2.$$

From Lemma A.25, we have $m_0(w) = 1$, and because $r \geq 2$, we have $m_0(w-1) = 0$.

Lemma A.28. *Suppose $r \geq 2$, $m_0(w) = 1$, and*

$$\frac{N(m_1^b)}{N(m_0)} \geq A(n,1)/2.$$

We have

$$m_0(w) = 1,$$
$$m_H(w+i) = dA(n,i)$$

for $0 \leq i \leq r-1$, and

$$m_H(w+r) = dA(n,r) - \frac{d(d-1)}{2} m_H(w-r).$$

Proof. We have already seen in Lemma A.25 that $m_0(w) = 1$.

For $w \leq q \leq w+r-1$, there is only one nonzero term in Equation (3); for $q = w+r$, there are only two. □

Lemma A.29. *Suppose $n \geq 5$, $r \geq 2$, $m_0(w) = 1$, and*

$$\frac{N(m_1^b)}{N(m_0)} \geq A(n,1)/2.$$

Then

$$A(n, 3r-2) \geq \frac{A(n,r-1)(dA(n,r-1)-1)}{2}.$$

Proof. Let

$$m_{3r-2} = m_0 + 2[w+r-1] - [w] - [w-r].$$

We have

$$A(n,3r-2) \geq \frac{N(m_{3r-2})}{d} = \frac{(dA(n,r-1))(dA(n,r-1)-1)m_H(w-r)}{2d(m_H(w) - m_0(w)+1)} \geq \frac{A(n,r-1)(dA(n,r-1)-1)}{2}.$$

□

Lemma A.30. *Suppose $n \geq 5$, $r \geq 2$, $m_0(w) = 1$, and*

$$\frac{N(m_1^b)}{N(m_0)} \geq A(n,1)/2.$$

Then $r = 2$ and

$$d < \frac{3A(n,4)}{A(n,1)^2}.$$

Proof. By Lemma A.29, we have

$$A(n,3r-2) \geq \frac{A(n,r-1)(dA(n,r-1)-1)}{2} \geq \frac{A(n,r-1)(A(n,r-1)-1)}{2}.$$

If $r \geq 3$, this is not possible by Lemma B.4.

If $r = 2$ the result follows from Lemma A.29 and

$$dA(n,r-1) - 1 > \frac{2}{3} dA(n,r-1).$$

□

Lemma A.31. *We cannot have $n \geq 5$, $k \leq n-1$, $r \geq 2$, and $m_0(w) = 1$.*

Proof. Assume we had a solution satisfying the given conditions. By Lemma A.30 we know that $r = 2$ and

$$d < \frac{3A(n,4)}{A(n,1)^2}.$$

This implies

$$d < \frac{A(n,2)}{2n},$$

by the inequality

$$6nA(n,4) < A(n,1)^2 A(n,2),$$

which holds for all $n \geq 3$. (See Appendix B and `bound-A5e` in the Python code.) Thus, since $m_H(w-2) < n$, we have

$$m_H(w+2) = dA(n,2) - \frac{d(d-1)}{2} m_H(w-2) \geq \frac{3}{4} dA(n,2).$$

Finally, taking

$$m_6^a = m_0 + 2[w+2] - [w] - [w-2],$$

we find that

$$\begin{aligned}
A(n,6) &\geq \frac{N(m_6^a)}{N(m_0)} \\
&\geq \frac{m_H(w+2)(m_H(w+2)-1)m_0(w)m_H(w-2)}{2(m_H(w)-m_0(w)+1)} \\
&= \frac{m_H(w+2)(m_H(w+2)-1)m_H(w-2)}{2d} \\
&\geq \frac{1}{2} \cdot \frac{3}{4} A(n,2)(\frac{3}{4} dA(n,2) - 1) \\
&\geq \frac{9}{32}(A(n,2)-1)^2.
\end{aligned}$$

This contradicts the inequality

$$(A(n,2)-1)^2 > 100A(n,6),$$

which holds for all n. (See Appendix B and `bound-A5f` in the Python code.) $\square$

Appendix B. Collected inequalities involving Eulerian numbers

The argument above used several dozen inequalities involving Eulerian numbers. We will not give detailed proofs of them; aside from the inequality in Lemma B.4, each of the inequalities used can be proven using Lemma B.1 or Lemma B.3 for large n, and then verifying by hand the finite number of remaining cases. One such proof is presented as Lemma B.5 to illustrate the method. The Python code used to verify the finite number of cases will be posted as an ancillary file alongside the arXiv submission.

The reader is encouraged to verify the plausibility of such inequalities for large n, using the asymptotic approximation $A(n,q) \sim (q+1)^n$, which is valid for fixed q and large n.

Recall our convention that $A(n,q) = 0$ if $q \geq n$.

Lemma B.1. *For all $n \geq 1$ and $q \geq 0$, we have*

$$(q+1)^n - (n+1)q^n \leq A(n,q) \leq (q+1)^n.$$

Proof. Recall ([47, §1.3]) that the Eulerian number $A(n,q)$ counts the number of permutations of $\{1,2,\dots,n\}$ with exactly q ascents.

If we label the integers 1 to n with labels 1 through $q+1$, then we get a permutation with at most q ascents by giving all the numbers of label 1 in decreasing order, then all the numbers of label 2 in decreasing order, and so on. Every permutation with at most q ascents arises in this way; this proves the right-hand inequality.

If a permutation constructed this way has fewer than q ascents, then there must exist adjacent labels i and $i+1$ where all the numbers with label i are less than all the numbers with label $i+1$. If that happens, we can record a number j from 0 to n which is the number of elements in the sequence with label at most i, then subtract one from the labeling of everything with label greater than i. There are $(n+1)q^n$ possibilities of this new data, and we can recover the original labeling by adding 1 to the label of everything that comes after the first j elements in the sequence. So the number of labelings giving permutations with fewer than q ascents is at most $(n+1)q^n$. This proves the left-hand inequality. ∎

Lemma B.2. *For all $n \geq 1$ and $q \geq 0$, we have*

$$A(n,q) \leq n!.$$

Proof. Follows from

$$\sum_q A(n,q) = n!.$$

∎

The two bounds given patch well enough for our modest needs: if $q \sim n/\log n$, then the bound in Lemma B.1 is close to $n!$, at least in a power sense. Surely more precise asymptotics are known, but these weak bounds are enough for the proof of Lemma B.4.

Lemma B.3. *If $n \geq 2$ and*

$$q < \frac{n}{log(n+1)+1} - 1$$

then

$$(1 - 1/e)(q+1)^n \leq A(n,q) \leq (q+1)^n.$$

Proof. We have

$$1/(q+1) < \log\left(\frac{q+1}{q}\right) < 1/q$$
$$e^{n/(q+1)} < \left(\frac{q+1}{q}\right)^n < e^{n/q}.$$

If $n/(q+1) > \log(n+1) + 1$ then

$$(n+1)q^n < (q+1)^n/e.$$

Now use Lemma B.1. □

Lemma B.4. *For n arbitrary and $3 \le r \le n-2$ we have*

$$A(n, r-1)(A(n, r-1) - 1) > 2A(n, 3r-2).$$

Proof. We'll assume $n \ge 21$; for smaller n there are only finitely many cases, which can be checked by hand. (See `bound-B` in the Python code.)

We can also assume $r \le n/2$; otherwise, the right-hand side is zero. We'll split into two cases: either $r < \frac{n}{log(n+1)+1}$ or $r > \sqrt{n}$. Our hypothesis on n guarantees that there is some r_0 with $\sqrt{n} \le r_0 < \frac{n}{log(n+1)+1}$; a fortiori, one of the two cases always holds.

If $r < \frac{n}{log(n+1)+1}$, then Lemma B.3 gives

$$A(n, r-1) \ge (1 - 1/e)r^n,$$

so

$$A(n, r-1)(A(n, r-1) - 1) \ge (1 - 1/e)^2 r^{2n} - (1 - 1/e)r^n.$$

On the other hand, by Lemma B.1, we can bound the right-hand side:

$$2A(n, 3r-2) \le 2(3r-1)^n.$$

The reader may verify that

$$(1 - 1/e)^2 r^{2n} > 2(3r-1)^n + (1 - 1/e)r^n$$

whenever $r \ge 3$ and $n \ge 14$; this proves the inequality we want.

Recall that we chose r_0 such that $\sqrt{n} \le r_0 < \frac{n}{log(n+1)+1}$. Lemma B.3 gives

$$A(n, r_0 - 1)(A(n, r_0 - 1) - 1) > \frac{3}{10} r_0^{2n} \ge \frac{3}{10} n^n > 2 * n!.$$

Now for any r with $r_0 \le r \le n/2$, we have

$$A(n, r) \ge A(n, r_0),$$

so

$$A(n, r-1)(A(n, r-1) - 1) > 2 * n! \ge 2A(n, 3r-2)$$

by Lemma B.2. □

We'll conclude with Lemma B.5, whose proof is given merely to illustrate a routine technique. A number of inequalities were used without proof in Appendix A. They can all be proven by asymptotic estimates using Lemma B.3 for large n, followed case-by-case verification for small n. The following bound was used in the proof of Lemma A.11; we give a full proof here to illustrate the general method.

Lemma B.5. *For $n \ge 11$, we have*

$$\frac{A(n, 2)}{A(n, 1)^2} < 1/27.$$

Proof. Lemma B.3 gives us

$$A(n,2) \leq 3^n$$

and

$$A(n,1) \geq (1-1/e)2^n,$$

provided $n \geq 6$. Hence we will be done as soon as we can show that

$$\frac{1}{(1-1/e)^2} \cdot \left(\frac{3}{4}\right)^n < 1/27,$$

which happens for $n \geq 15$. The cases $11 \leq n \leq 14$ must be checked separately. $\square$

References

[1] Yves André. *Mumford-Tate groups of mixed Hodge structures and the theorem of the fixed part*. Comp. Math. 82 (1992), 1–24.

[2] Yves André. *On the Shafarevich and Tate conjectures for hyperkähler varieties*. Math. Ann. 305, 205–248 (1996).

[3] Benjamin Bakker and Jacob Tsimerman. *The Ax-Schanuel conjecture for variations of Hodge structures*. Invent. Math. 217 (2019), 77–94.

[4] Michael Bate, Sebastian Herpel, Benjamin Martin, and Gerhard Röhrle. *Cocharacter-closure and the rational Hilbert-Mumford Theorem*. Math. Zeit. 287 (2017), 39–72.

[5] Michael Bate, Benjamin Martin, and Gerhard Röhrle. *A geometric approach to complete reducibility*. Invent. Math. 161 (2005), 177–218.

[6] Michael Bate, Benjamin Martin, and Gerhard Röhrle. *Semisimplification for subgroups of reductive algebraic groups*. Arxiv preprint 2004.08105, 2020.

[7] Michael Bate, Benjamin Martin, Gerhard Röhrle, and Rudolf Tange. *Closed orbits and uniform S-instability in invariant theory.*

[8] A. A. Beilinson, J. Bernstein, P. Deligne, and O. Gabber. *Fascieaux pervers*, Astérisque, Société Mathématique de France, Paris. **100** (1982).

[9] Pierre Berthelot. *Cohomologie cristalline des schémas de caractéristique $p > 0$*. Springer, Lecture Notes in Mathematics, 1974.

[10] Pierre Berthelot and Arthur Ogus. *Notes on Crystalline Cohomology.* Princeton University Press, 1978.

[11] Siegfried Bosch, Werner Lütkebohmert, Michel Raynaud. *Néron Models*. Springer, 1990.

[12] Eduardo Cattani, Pierre Deligne, and Aroldo Kaplan, *On the locus of Hodge classes.* Journal of the AMS **8** 483–506 (1995).

[13] Pierre Deligne, *Catégories tensorielles.* Online at `https://www.math.ias.edu/files/deligne/Tensorielles.pdf`. (see ref in comments)

[14] Pierre Deligne, *La conjecture de Weil: II*, IHÉS Mathematical Publications, **52**, 137–252 (1980).

[15] Pierre Deligne, *Le groupe fondamental de la droite projective moins trois points.* 1989.

[16] P. Deligne and J. S. Milne, *Tannakian categories.* `https://www.jmilne.org/math/xnotes/tc2018.pdf`.

[17] Vladimir Drinfeld, *On a conjecture of Kashiwara*, Mathematics Research Letters, **8**, 713–728, 2001.

[18] Hélène Esnault, Phûng Hô Hai, and Xiaotao Sun, On Nori's fundamental group scheme, In: Mikhail Kapranov, Yuri Ivanovich Manin, Pieter Moree, Sergiy Kolyada, Leonid Potyagailo, Geometry and Dynamics of Groups and Spaces. Progress in Mathematics, vol 265. Birkhäuser Basel

[19] Gerd Faltings. Crystalline cohomology and p-adic Galois representations. In *Algebraic analysis, geometry, and number theory: proceedings of the JAMI inaugural conference*, edited by Jun-Ichi Igusa, Johns Hopkins University Press, 1989.

[20] Gerd Faltings, Endlichkeitssätze für abelsche Varietäten über Zahlkörpern. Invent. math. **73** (1983), 349–366.

[21] Gerd Faltings, Diophantine approximation on abelian varieties, Annals of Mathematics, **133** (1991), 549–576.

[22] Jean-Marc Fontaine. Périodes p-adiques. In *Astérisque*, volume 223. Société Mathématique de France, 1994.

[23] J. Franecki and M. Kapranov, The Gauss map and a noncompact Riemann-Roch formula for constructible sheaves on semiabelian varieties, Duke Mathematical Journal **104** (2000), 171-180.

[24] Eberhard Freitag and Reinhardt Kiehl. *Étale cohomology and the Weil conjectures.* Springer, 1988.

[25] Ofer Gabber and François Loeser, Faisceaux perver ℓ-adique sur un tore, Duke Matehmatical Journal **83** (1996), 501–606.

[26] D. Gaitsgory, *On de Jong's conjecture*, Israel Journal of Mathematics **157**, 155–191 (2007).

[27] Ariyan Javanpeykar. Néron models and the arithmetic Shafarevich conjecture for certain canonically polarized surfaces. Bull. London Math. Soc. 47(1):55–64 (2015).

[28] A. Javanpeykar and D. Loughran. Arithmetic hyperbolicity and a stacky Chevalley–Weil theorem. Preprint. ArXiv 1808.09876 (2018).

[29] A. Javanpeykar and D. Loughran. Complete intersections: moduli, Torelli, and good reduction. Math. Ann. 368 (2017), no. 3–4, 1191–1225.

[30] A. Javanpeykar and D. Loughran. Good reduction of algebraic groups and flag varieties. Arch. Math. (Basel), 104(2):133–143 (2015).

[31] Ariyan Javanpeykar and Daniel Loughran. Good reduction of Fano threefolds and sextic surfaces. Ann. Sc. Norm. Super. Pisa Cl. Sci. (5) 18 (2018), no. 2, 509–535.

[32] Oliver Johnson and Christina Goldschmidt. Preservation of log-concavity on summation. *ESAIM Probab. Stat.*, 10:206–215, 2006.

[33] Nicholas M. Katz. *Convolution and Equidistribution: Sato-Tate theorems for finite field Mellin transforms*. Annals of Mathematics Studies **180**, 2012.

[34] Nicholas M. Katz. *Exponential Sums and Differential Equations*. Annals of Mathematics Studies **124**, 1991.

[35] Nicholas M. Katz. *Moments, Monodromy and Perversity: a Diophantine perspective*. Annals of Mathematics Studies **159**, 2006.

[36] E. R. Kolchin, Algebraic groups and algebraic dependence, American Journal of Mathematics **90** (1968), 1151-1164.

[37] Thomas Krämer., Characteristic cycles and the microlocal geometry of the Gauss map, I. `https://arxiv.org/pdf/1604.02389.pdf`, 2016.

[38] Thomas Krämer., Characteristic cycles and the microlocal geometry of the Gauss map, II. `https://arxiv.org/pdf/1807.01929.pdf`, 2018.

[39] Thomas Krämer. and Rainer Weissauer, Vanishing theorems for constructible sheaves on abelian varieties, J. Algebraic Geom. **24** (2015), 531–568

[40] T. Krämer. and R. Weissauer, On the Tannaka group attached to the theta divisor of a generic principally polarized abelian variety, Mathematische Zeitschrift **281** (2015), 723-745.

[41] Robert M. Guralnick and Pham Huu Tiep, Decompositions of small tensor powers and Larsen's conjecture, Representation Theory **9** (2005), 138–208

[42] Brian Lawrence and Akshay Venkatesh. *Diophantine problems and p-adic period mappings*. ArXiv preprint 1807.02721.

[43] Benjamin M. S. Martin. Reductive subgroups of reductive groups in nonzero characteristic. Journal of Algebra 262 (2003) 265–286.

[44] James S. Milne, *Points on Shimura varieties over finite fields: the conjecture of Langlands and Rapoport*. ArXiv preprint 0707.3173, v. 4.

[45] M. S. Narasimhan and M. V. Nori. Polarizations of an abelian variety. Proceeings of the Indian Academic Society (Math Society) **90** (1981), 125–128.

[46] A. Ogus. F-crystals and Griffiths transversality. In *Proceedings of the international symposium on algebraic geometry, Kyoto, 1977 (a Taniguchi symposium)*, Kinokuniya Book-Store Co., Ltd., 1978.

[47] T. Kyle Petersen. *Eulerian numbers*. Birkhäuser, 2015.

[48] Mihnea Popa and Christian Schnell, *Generic vanishing theory via Mixed Hodge modules*. For. Math. Sigma, 1 (2013).

[49] Marius van der Put and Michael F. Singer. *Galois Theory of Linear Differential Equations*. Springer, 2003.

[50] Kenneth A. Ribet. *Galois action on division points of abelian varieties with real multiplications*. Amer. J. Math., 98(3): 751–804 (1976).

[51] R. W. Richardson. *Conjugacy classes of n-tuples in Lie algebras and algebraic groups*. Duke (1988) Vol. 57, No. 1.

[52] Neantro Saavedra Rivano, *Catégories Tannakiennes*.

[53] Wilfried Schmid. *Variation of Hodge structure: the singularities of the period mapping*. Inv. math. 22, 211–319 (1973).

[54] A. J. Scholl. *A finiteness theorem for del Pezzo surfaces over algebraic number fields* J. London Math. Soc. (1985).

[55] Peter Scholze. *p-Adic Hodge theory for rigid analytic varieties*. For. Math. Pi, 1 (2013).

[56] Jean-Pierre Serre, *Cohomologie Galoisienne*. Springer, Lecture Notes in Mathematics, 1994.

[57] Jean-Pierre Serre, *Complète Réducibilité*. Séminaire Bourbaki, no. 932, 195–217 (2004).

[58] Yiwei She, *The unpolarized Shafarevich conjecture for K3 surfaces*. ArXiv preprint 1705.09038 (2017).

[59] Fucheng Tan and Jilong Tong. *Crystalline comparison isomorphisms in p-adic Hodge theory: the absolutely unramified case.* Alg. Num. Th. 13, 1509–1581 (2019).

[60] Evgueni V. Tevelev. *Projective Duality and Homogeneous Spaces.* Encyclopedia of Mathematical Sciences **133**, Springer-Verlag (2005).

[61] Rainer Weissauer. Brill-Noether sheaves. `https://arxiv.org/pdf/math/0610923` (2006).

[62] R. Weissauer. A remark on rigidity of BN-sheaves. `https://arxiv.org/pdf/1111.6095` (2011).

[63] Rainer Weissauer. Vanishing theorems for abelian varieties over finite fields. `https://arxiv.org/pdf/1407.0844` (2014).