Outline of Alpöge's proof of the generalised Riemann hypothesis

Siniša Slijepčević

November 21, 2026

hexagon:2611.00812v1

Abstract

We give an outline of the proof by L. Alpöge, in a formal development in Lean 4, of the Riemann hypothesis and of the generalised Riemann hypothesis for Dirichlet L-functions. The main theorem is that $L(s,\chi) \neq 0$ for $\operatorname{Re} s > \frac{1}{2}$ and every Dirichlet character χ. A positivity argument reduces it to one-sided bounds, on windows of scales $[Y, Y^{3/2}]$, for positive combinations of Riesz-weighted Chebyshev functions twisted by the powers of a primitive character. These bounds come from hermitian lattices attached to theta lifts over auxiliary abelian CM fields: the inert prime sums of these fields are bounded, prime by prime, by local gains of the lattices, and slope inequalities bound the lattice side. The losses in the inequalities are controlled by bounds, polynomial in the conductor, for Tate–Shafarevich groups of Tate twists, which are proved by Eisenstein congruences; the Galois representations that these need are constructed by theta lifts. We state the argument in standard language as far as we understand it, give many proofs in outline, and say what we do not yet understand. The paper was prepared with the help of Aleph, an AI math research assistant.

1 Introduction

1.1 The theorems

The Riemann hypothesis asserts that the non-trivial zeros of the Riemann zeta function lie on the line $\operatorname{Re} s = \frac{1}{2}$, and the generalised Riemann hypothesis for Dirichlet L-functions asserts the same of $L(s,\chi)$ for every Dirichlet character χ [31], [61]. Both are the main theorems of a formal development of L. Alpöge [1], and this paper is an outline of its proof. We begin with the notation and the statements.

Notation for Dirichlet characters and their L-functions is that of [89]. For an integer $N \geq 1$, a **Dirichlet character modulo** N is a character χ of $(\mathbb{Z}/N\mathbb{Z})^\times$, extended to $\mathbb{Z}$ by $\chi(n) = 0$ for $\gcd(n,N) > 1$; the **principal character** modulo N is denoted $\chi_{0,N}$, and χ is **primitive** if its conductor equals N. The complex variable is s. The **Dirichlet L-function** of χ is the meromorphic continuation to $\mathbb{C}$ of

$$L(s,\chi) = \sum_{n\geq 1} \chi(n)\, n^{-s} = \prod_p \bigl(1 - \chi(p)\, p^{-s}\bigr)^{-1}, \qquad \operatorname{Re} s > 1,$$

where χ is taken as a character modulo N, so that the Euler factors at the primes $p \mid N$ are equal to 1 also when χ is imprimitive. It is entire unless $\chi = \chi_{0,N}$, in which case its only pole is a simple pole at $s = 1$. The Riemann zeta function is $\zeta(s) = L(s, \chi_{0,1})$; its **trivial zeros** are the points $s = -2n$, $n \geq 1$. The von Mangoldt function is $\Lambda(n)$.

The main theorem is a zero-free half-plane for all Dirichlet L-functions at once.

Theorem 1.1. *Let $N \geq 1$ be an integer and let χ be a Dirichlet character modulo N. Then*

$$L(s,\chi) \neq 0 \qquad \textit{for every } s \in \mathbb{C} \textit{ with } \operatorname{Re} s > \tfrac{1}{2},$$

except at the pole $s = 1$ when $\chi = \chi_{0,N}$ is principal.

Proof. Let ψ be a primitive Dirichlet character. By Proposition 2.5, the hypothesis of Proposition 2.3 holds for ψ with window ratio $\frac{3}{2}$; hence $L(s, \psi) \neq 0$ for $\operatorname{Re} s > \frac{1}{2}$. Now let χ be any Dirichlet character modulo N and χ^* the primitive character inducing it. Then

$$L(s, \chi) = L(s, \chi^*) \prod_{p \mid N} \bigl(1 - \chi^*(p)\, p^{-s}\bigr),$$

and no factor of the finite product vanishes for $\operatorname{Re} s > 0$, because $|\chi^*(p) p^{-s}| < 1$. So $L(s, \chi) \neq 0$ for $\operatorname{Re} s > \frac{1}{2}$, apart from the pole at $s = 1$ for principal χ. □

Through the functional equation it gives the Riemann hypothesis and its analogue for Dirichlet L-functions.

Theorem 1.2. *Every zero of the Riemann zeta function $\zeta(s)$ other than the trivial zeros $s = -2n$, $n \geq 1$, satisfies $\operatorname{Re} s = \frac{1}{2}$.*

Proof. Let $\zeta(s) = 0$ with s not a trivial zero. By Theorem 1.1 with $N = 1$, ζ has no zero with $\operatorname{Re} s > \frac{1}{2}$, so it suffices to exclude $\operatorname{Re} s < \frac{1}{2}$. Let $\xi(s) = \pi^{-s/2}\Gamma(s/2)\zeta(s)$ be the completed zeta function, which satisfies $\xi(1 - s) = \xi(s)$. The factor $\pi^{-s/2}\Gamma(s/2)$ has no zeros, and its poles are $s = 0, -2, -4, \ldots$; since $\zeta(0) \neq 0$ and s is not a trivial zero, $\xi(s) = 0$. Hence $\xi(1 - s) = 0$, and since $\operatorname{Re}(1 - s) > \frac{1}{2}$ the Gamma factor is finite and nonzero at $1 - s$, so $\zeta(1 - s) = 0$ with $\operatorname{Re}(1 - s) > \frac{1}{2}$ and $1 - s \neq 1$, a contradiction. The functional equation and the properties of the Gamma factor are classical [89]. □

Corollary 1.3. *Let $N \geq 1$ and let χ be a Dirichlet character modulo N, primitive or not. Every zero s of $L(s, \chi)$ with $0 < \operatorname{Re} s < 1$ satisfies $\operatorname{Re} s = \frac{1}{2}$.*

Proof. Let $L(s, \chi) = 0$ with $0 < \operatorname{Re} s < 1$. By Theorem 1.1, $\operatorname{Re} s \leq \frac{1}{2}$; suppose $\operatorname{Re} s < \frac{1}{2}$. Let χ^* be the primitive character modulo the conductor of χ inducing χ. Then $L(s, \chi) = L(s, \chi^*) \prod_{p \mid N} \bigl(1 - \chi^*(p) p^{-s}\bigr)$, and no factor of the finite product vanishes for $\operatorname{Re} s > 0$, so $L(s, \chi^*) = 0$. The completed L-function of the primitive character χ^* satisfies a functional equation relating s and $1 - s$ and χ^* and $\overline{\chi^*}$, with a nonzero root number, and its Gamma factor is finite and nonzero for $\operatorname{Re} s > 0$; hence $L(1 - s, \overline{\chi^*}) = 0$ with $\operatorname{Re}(1 - s) > \frac{1}{2}$ and $1 - s \neq 1$, contradicting Theorem 1.1. The functional equation is classical [89]. □

The deductions of Theorem 1.2 and Corollary 1.3 from Theorem 1.1 are classical, and so is the passage, in the proof of Theorem 1.1, from arbitrary to primitive characters. Since the Dedekind zeta function of an abelian number field is the product of the L-functions of its characters [132], Theorem 1.1 is equivalent to the Riemann hypothesis for the Dedekind zeta functions of all abelian number fields. For comparison, the published literature gives a zero-free region $\operatorname{Re} s > 1 - c/\log\bigl(N(|\operatorname{Im} s| + 2)\bigr)$ for the characters modulo N, except possibly for one real zero of one real character [31], which Siegel's theorem bounds only ineffectively [115]; no half-plane $\operatorname{Re} s > \sigma_1$ with $\sigma_1 < 1$ is known there to be free of zeros of $\zeta(s)$ [128]; and on average over moduli, the Bombieri–Vinogradov theorem gives for primes in arithmetic progressions what the generalised Riemann hypothesis gives for each modulus [15], [130]. The proof of Theorem 1.1 reduces it to Proposition 2.3 and Proposition 2.5, both in Section 2; the rest of the paper outlines their proofs.

1.2 This outline

Besides the development itself there is only a short note of remarks by its author [2], which contains no proofs. The development [1] is written in the proof assistant Lean 4, and its main theorems are Theorem 1.2, Corollary 1.3 and Theorem 1.1; the argument is formalised there and machine-checked. We have not verified it. The development is very large, and we have read only part of it. This paper covers the path of the proof of Theorem 1.1 as far as we have been able to state it in standard language, and it leaves out the parts of the development that the main theorems do not use. Every numbered statement below is a statement of the development, written in standard language and compared with it clause by clause; a definition that omits conditions says so in a closing clause, and a statement that depends on the omitted part is to that extent an outline as well. A proof that opens with the words "In outline." gives the top level of the argument: it names the results it combines and the mechanism, and below it lies a tower of further results that we do not reproduce. The other proofs are complete deductions from the statements they cite. The mathematics is due to Alpöge; errors of exposition are ours.

1.3 The strategy

The proof is a positivity argument with an arithmetic engine. At the top it is analytic: the generalised Riemann hypothesis for a primitive character ψ is reduced to one-sided bounds for a positive combination of Riesz-weighted Chebyshev functions, needed only on windows of scales $[Y, Y^{3/2}]$ with Y arbitrarily large. These bounds are obtained one scale at a time over auxiliary abelian CM fields: their inert prime sums are bounded, prime by prime, by local gains of hermitian lattices attached to theta lifts, and slope inequalities bound the lattice side. The losses in those inequalities are controlled by bounds for Tate–Shafarevich groups of Tate twists, which are proved by Eisenstein congruences, and the Galois representations that the congruences need are constructed by theta lifts. We describe the layers in turn.

Positivity on windows. For $y \geq 1$ let

$$G(y) = \sum_{n \leq y} \Lambda(n)\Big(\frac{y}{n} - 1\Big) - y\big(\log y - 1 - \gamma\big),$$

where γ is Euler's constant. Thus $G(y)/y$ is the Riesz mean of order one of $\sum_n \Lambda(n)n^{-1}$ with its main term removed [47], and we call G the Riesz-weighted Chebyshev function. Its explicit formula converges absolutely, so the Riemann hypothesis gives $G(y) = O(y^{1/2})$. Conversely, a bound in one direction, $G(y) \leq y^\theta$ for all large y with $0 < \theta < 1$, excludes zeros of $\zeta(s)$ in $\operatorname{Re} s > \theta$: the Mellin transform of $y^\theta - G(y)$ has an eventually non-negative integrand, and Landau's theorem on such integrals applies [76], [60]. For a primitive character ψ of order o the argument forms the twisted functions $g(y, \chi)$ in the same way, with $\Lambda(n)\chi(n)$ in place of $\Lambda(n)$ and the main term $-y\,L'(1,\chi)/L(1,\chi)$ removed, and the combination

$$G_{\mathfrak{L}_\psi}(y) = G(y) + \frac{1}{3o+1} \sum_{\chi} \operatorname{Re} g(y, \chi),$$

where χ runs over the non-trivial primitive characters that induce powers of ψ. Again a one-sided bound suffices. Landau's argument uses only the sign of $y^\theta - G_{\mathfrak{L}_\psi}(y)$ where the bound holds and the positivity of the weights, through which the residue of the Mellin transform at a real zero is a positive combination of multiplicities; so a bound $G_{\mathfrak{L}_\psi}(y) \leq y^\theta$ excludes zeros of $L(s, \psi)$ in $\operatorname{Re} s > \theta$. The particular weight $(3o+1)^{-1}$ is not needed for this: it is the weight that the construction of the auxiliary fields below produces. In the terminology of the development, the

data of such a combination form a character level (Definition 2.1). The first step of the argument is its own form of Landau's theorem, in which the bound is needed only for y in windows $[Y, Y^{\lambda}]$, for a fixed $\lambda > 1$ and arbitrarily large Y (Proposition 2.2). This matters because the arithmetic below produces its auxiliary objects only at scales in such windows. What has to be proved is therefore that for every small $\varepsilon > 0$ one has $G_{\mathfrak{L}_\psi}(x) \le x^{1/2+C\varepsilon}$ for all x in a set of scales that contains windows $[Y, Y^{3/2}]$ with Y arbitrarily large, with C independent of ε.

One host at one scale. The bound at a scale x comes from one auxiliary field, which the development calls a host (Definition 3.1). It is an abelian CM field $E \subset \mathbb{Q}(\zeta_{Qqd})$ containing $\sqrt{-d}$, whose Galois group is identified with $\{\pm 1\}^2 \times A$ for a finite abelian group A of order about $x^{2+8\varepsilon}$, with its Artin map written out in these coordinates, so that the powers of ψ are among the Dirichlet characters of E. It is built on a pair of primes (q, d) with $q \equiv 3 \pmod 8$, $d \equiv 7 \pmod{16}$ and $q \mid d+1$, of size $\log q \asymp Y^{\varepsilon/5}$ and $\log d \asymp Y^{\varepsilon/5} \log^2 Y$ for a scale $Y \le x$, which the development calls a window pair (Definition 3.4). The device is an identity. By orthogonality of the characters of $\{\pm 1\}^2 \times A$, sums of $\log p$ over the primes p inert in $\mathbb{Q}(\sqrt{-d})$, weighted according to their Frobenius classes in E and cut off at scales near x, are combinations of the twisted Chebyshev functions of the characters of E. After averages over translates and a choice of weights, the characters of $\mathfrak{L}_\psi$ survive with the weights of $G_{\mathfrak{L}_\psi}$, and an upper bound for the excess of these inert prime sums over their main terms becomes an upper bound for $G_{\mathfrak{L}_\psi}(x)$ (Proposition 4.8), up to the other characters of the host. These are, roughly, twists by the quadratic characters of q and d. The pair is chosen so that each of them has a square-root bound and no zero in a box near $s = 1$, and such good pairs exist at every scale because almost all pairs are good, by a mean-square argument over quadratic twists (Proposition 3.6, Proposition 3.7).

Prime sums and lattices. The excess bound is not obtained by analysis. Its setting is what the development calls a slope datum (Definition 4.2): a family of numbers, most of which are the ranks, slopes, maximal slopes and heights of hermitian lattices over the host and of their duals, in the sense of the Harder–Narasimhan theory of hermitian vector bundles [121], [45], [19]. One entry is set equal, by definition, to the inert prime sums of the host. This is the heart of the matter: an inequality, the loss condition, compares these prime sums with a quantity of the lattices, a sum over the inert primes p of $\log p$ times the net local gain of the lattices at p, and asks that the prime contributions be dominated by the local gains up to a small total loss. A slope inequality in the sense of [19], a dual-slope bound and a divisor bound then bound the lattice side by the main term (Proposition 4.5). Only an upper bound for the prime sums is needed, and only an upper bound for $G_{\mathfrak{L}_\psi}$ results; positivity does the rest. The losses must be at most $x^{C\varepsilon}$ times the natural size of the terms, with C independent of ε. A family of hosts whose lattice data satisfy these conditions in this way at every large scale of its set of scales is called tied (Definition 6.13), and the analytic part of the argument shows that tied families give Theorem 1.1.

Lineage. According to its author [2], the argument is a descendant of Stepanov's proof of the Riemann hypothesis for curves over finite fields [120], [16], in which an auxiliary function vanishing to high multiplicity bounds the number of points from above. In the author's description, multiplicity of this kind is hard to find over the integers, and a chain of theta lifts plays its role, each lift mimicking multiplicity, with a chain of length about $\sqrt{y}$ for the von Mangoldt sum of length y. What the outline shows fits this description in three places. The function to be bounded is the sum $\sum_{n \le y} \Lambda(n)(y/n - 1)$ minus its main term, as above. The hermitian spaces of a master datum include a chain of definite spaces $X_0, \dots, X_m$ of ranks $k_j = k + j(t+1)$, with kernels between consecutive steps (Definition 6.7, Definition 6.8), and at a scale x the length m of the chain is about $x^{1/2-\varepsilon/16}$ (Definition 6.3). And the upper bound comes from a slope

inequality in the sense of Bost [19], the Arakelov form of the method of auxiliary functions, of which Stepanov's proof is a classical example. How the chain of lifts realises multiplicity we have not been able to see (Section 11).

Lattices from theta lifts. The hermitian lattices come from theta lifts on definite unitary groups over the host fields. To each scale the argument attaches families of what it calls master data (Definition 6.10): a host field with its Artin map, a tower of definite hermitian spaces with prescribed signatures, discrete-series parameters at the places of a CM type, iterated kernels of theta type with determinant characters, and local data at auxiliary primes. These families carry a definite-side family structure (Proposition 6.18), from which a general construction produces tied families once the local defects of the lattices are bounded (Proposition 6.20). The structure rests on three groups of results. The first is theta agreement and local matching for pairs of theta integrals, deduced from the Siegel–Weil formula for unitary groups [59], which goes back to [133] and [73], through the doubling method [99] (Proposition 7.3). The second is an arithmetic control of the rows of the families (Theorem 7.6), in which ratios of theta constants along a grid of hermitian spaces are, up to algebraic numbers with controlled prime factors, products of critical values of Hecke L-functions of the host field, in the shape of the Rallis inner product formula [101], [80], [49] (Theorem 7.9). The third is an upper bound for covolumes of lattices of automorphic forms (Theorem 8.1), from lower bounds for arithmetic heights of automorphic line bundles on towers of integral models of unitary Shimura varieties [44], [138], [91].

Tate–Shafarevich groups. The local defects of the lattices are bounded by the lengths of Tate–Shafarevich groups of Tate twists over the host fields. For a number field E, a prime ℓ and an integer k, let $\mathrm{Sha}^1_\ell(E,k)$ be the kernel of the restriction map from $H^1(G_{E,S_\ell},\mathbb{Q}_\ell/\mathbb{Z}_\ell(-k))$ to the product of the local groups $H^1(E_v,\mathbb{Q}_\ell/\mathbb{Z}_\ell(-k))$, $v\mid\ell$, where G_{E,S_ℓ} is the Galois group of the maximal extension of E unramified outside ℓ; it is the Tate–Shafarevich group of Poitou–Tate duality [97]. The arithmetic input of the whole argument is the window Sha bound (Theorem 9.1). For every primitive ψ, every small ε and every window pair (q,d) at a large scale, it gives a range $[Y,Y^2]$ of scales such that for every x in the range the compositum E of the field cut out by ψ, of $\mathbb{Q}(\sqrt{-q},\sqrt{-d})$ and of the real subfield of a cyclotomic field $\mathbb{Q}(\zeta_{2^{e+2}})$ with 2^e about $x^{2+8\varepsilon}$ satisfies

$$\sum_{\ell}\log\ell\sum_{k=1}^{t-1}\mathrm{len}_\ell(E,k)\le\exp\bigl(x^{\varepsilon}/2\bigr),\qquad t=\lfloor x^{401}\rfloor,$$

where $\mathrm{len}_\ell(E,k)$ is the length of $\mathrm{Sha}^1_\ell(E,k)$. By Shapiro's lemma these groups decompose over the characters of $\mathrm{Gal}(E/\mathbb{Q})$. The pieces of one parity are bounded through Bernoulli numbers, and the others, in all but a few cases, by Eisenstein congruences. A non-trivial class gives a residually reducible representation of dimension three or four, and the length is bounded by valuations of a congruence number between an Eisenstein representation and cuspidal automorphic representations of $\mathrm{U}(3)$ or GSp_4; one non-zero integer, whose logarithm is polynomial in the conductor and in the twist, controls all primes ℓ at once (Proposition 9.7); since the conductor of E is of size $\exp(O(x^{\varepsilon/5}\log^2 x))$, such bounds fit within $\exp(x^{\varepsilon}/2)$. Eisenstein congruences in the tradition of [102], [82], [9] are used to construct Selmer classes; here they bound Tate–Shafarevich groups from above, effectively and uniformly in ℓ. Kato's Euler system [65] enters the main case, and the bad primes are handled by a bound of Soulé's type for second cohomology [119] and the theorem of Ferrero and Washington [39].

Galois representations. The congruences need Galois representations attached to polarized regular algebraic cuspidal representations of GL_n over CM fields with local–global compatibility, the local Langlands correspondence for GL_n, and base change from unitary groups [50], [55], [114], [27], [22], [74], [88]. These are known theorems, and the development proves them by its

own routes. For the Galois representations the route goes through theta lifts (Theorem 10.1). Representations are first obtained for automorphic forms of a special type and lacunary weight on definite unitary groups of odd rank, by a step whose mechanism we do not know, then for all forms of odd rank as ℓ-adic limits of congruent systems, and in even rank from two theta lifts to odd rank with distinct auxiliary characters, whose difference, as the statement of that step suggests, isolates the representation sought.

1.4 Known and new; the state of our understanding

The top of the argument uses classical analytic number theory: Landau's theorem, the explicit formula, zero-density estimates and zero-free regions [31], [61]. The theta correspondence with the Siegel–Weil and Rallis inner product formulas, the arithmetic intersection theory of Shimura varieties, Galois cohomology with Poitou–Tate duality, Kato's Euler system, the theorem of Ferrero and Washington and Soulé's bounds enter as known theorems, and so do the theorems on Galois representations, local Langlands and base change listed above; the appendix restates them with references. As far as we can judge, the following are new: the window form of Landau's theorem, which makes one-sided bounds on sparse windows sufficient; the hosts, and above all the bound of their inert prime sums by local gains of hermitian lattices, through the loss condition, together with the slope inequalities that bound the lattice side; the supply of window pairs, by which almost all quadratic twists of enormous conductor have square-root bounds on long ranges of scales; bounds for Tate–Shafarevich groups of Tate twists over abelian CM fields, polynomial in the conductor, where the method of Eisenstein congruences is classical and the effectivity and uniformity look new; and the theta-lift constructions, both of the lattice data and of the Galois representations, in particular the pair of theta lifts in even rank.

Our understanding of the argument is uneven, and we say here once where it stands. The analytic reduction of Section 2 and the bound from one host in Section 4 are understood at the level of their statements and of their mechanism; the quantitative window form of Landau's argument, the passage from a bounded excess to the bound for $G_{\mathfrak{L}}$, and the size and error conditions that make it work are given in outline. We do not understand why the local gains of the lattices dominate the prime contributions up to a small loss, that is, how the loss bounds are proved, nor what these gains are in classical terms, nor why the lattice data satisfy the divisor bound. The supply of hosts in Section 3 is understood except for the mean-square estimates and the construction of hosts from good pairs. The construction of the lattice data in Sections 6–8 is the least understood part: several of its definitions name groups of conditions without stating them, and the general construction that turns the family structure into tied families, as well as the way in which the Sha bounds enter the parameters of the slope inequalities, are known to us only through their inputs. Of the covolume bound we understand the architecture but not the mechanism of its volume estimates. Of the Sha bounds we understand the decomposition into cases and the role of the congruence numbers, but not the deformation-theoretic core of the congruence bound or the automorphy statements for GSp_4 and for $\mathrm{U}(3)$ over $\mathbb{Q}(i)$. Of the route to Galois representations, the step that produces the first Galois representations is the least clear to us. Section 11 lists these points in detail.

1.5 Conventions and terminology

Throughout, ψ is a primitive Dirichlet character, ε is the small parameter of the argument, with $0 < \varepsilon < 1/52$, and x, y, X, Y are large real scales. A set of scales contains λ-windows if it contains intervals $[Y, Y^{\lambda}]$ with Y arbitrarily large; the word window is used only in this sense, and window pairs and the window Sha bound are named after the windows on which they serve.

Labels such as (E3) or (S2) refer to the conditions listed in the definition or statement where they occur. Known theorems are named and cited where a section first needs them, are not numbered, and are restated in the appendix. The explicit constants, such as the bound $1/52$ for ε, the window ratio $\frac{3}{2}$, the weights $(3o+1)^{-1}$, the exponent 401 of the Sha bound and the thresholds 1800, 100 and 44 for the grids of Section 7, are those of the development; we do not know which of them are forced by the method.

Most objects of the argument have no counterpart in the literature, and we use the development's names for them, introducing each as such where it first appears. In the order of the argument the main ones are: the character level, a finite set of primitive characters with weights (Definition 2.1); the host, an abelian CM field with an explicit description of its Galois group and Artin map (Definition 3.1); the window pair of primes on which a host is built, and the Sha bound $\mathrm{V}(E;t,B)$ for its fields (Definition 3.4); the host family over a set of scales, and the property of being tied (Definition 3.3, Definition 6.13); the arithmetic datum of a host and its inert prime sums (Definition 4.1); the slope datum with its slope conditions (Definition 4.2); and the master datum of the theta-lift configuration (Definition 6.10). Where the literature has a word we use it: slopes and heights are those of Arakelov geometry [19], [20], and the Tate–Shafarevich groups are those of Poitou–Tate duality.

1.6 Plan of the paper

Section 2 carries out the analytic reduction: it deduces Theorem 1.1 from the existence of tied host families and ends with the list of the five results on which the rest of the proof depends. Section 3 introduces the hosts and supplies them on windows, and Section 4 is the bridge between prime sums and lattices; a reader who wants the mechanism of the proof should read these three sections. Section 5 fixes the standard notation of the arithmetic part. Sections 6, 7 and 8 construct the lattice data from theta lifts; they are the largest part of the argument, and most of their statements are outlines. Section 9 outlines the proof of the Sha bounds; it can be read after Sections 3 and 5, independently of Sections 4 and 6–8. Section 10 outlines the route by which the argument obtains the Galois representations and the base change that the Sha bounds need. Section 11 says what we do not understand, and the appendix restates the known theorems used, by area.

2 The analytic reduction

This section reduces Theorem 1.1 to the existence of what the development calls tied host families. The reduction has two steps. The first is the argument's own form of Landau's theorem [76], in which a one-sided bound for a positive combination of twisted Riesz-weighted Chebyshev functions is required only on windows $[Y, Y^{\lambda}]$ with Y arbitrarily large (Proposition 2.2). The second obtains such a bound, scale by scale, at the scales of a tied family (Proposition 2.4). The arithmetic of the later sections enters only through the bound at a single scale and through the existence of the families; the section ends with the list of the results, treated later, on which the reduction rests.

The functions to be bounded are defined next. The reader should think of a character level as a finite set of primitive characters, closed under complex conjugation, together with the weights with which their twisted Chebyshev functions are added to G. The cyclic levels $\mathfrak{L}_{\psi}$ are the only levels that the proof of Theorem 1.1 uses; they have no exceptional character, and K_{ψ} contains ψ itself.

Definition 2.1. Let $\mathcal{P}$ denote the set of **primitive Dirichlet characters**, each regarded as

a pair (N,ψ) with ψ primitive modulo $N \geq 1$. Let $\mathbf{1} \in \mathcal{P}$ be the trivial character modulo 1, let $\overline{\psi} \in \mathcal{P}$ be the complex conjugate of ψ, and for a Dirichlet character χ let $\chi^* \in \mathcal{P}$ be the primitive character inducing χ. A character ψ is **real** if it takes only real values.

A **character level** is a tuple $\mathfrak{L} = (K, \psi_{\mathrm{exc}}, o)$ consisting of a finite set $K \subset \mathcal{P}$, an optional element ψ_{exc} (either absent or a character of K) and an integer $o \geq 0$, such that

(L1) $\mathbf{1} \in K$ and $\overline{\psi} \in K$ for every $\psi \in K$;

(L2) if ψ_{exc} is present, it is odd and real.

The **cyclic level** of $\psi \in \mathcal{P}$ of modulus N is $\mathfrak{L}_\psi = (K_\psi, \mathrm{absent}, o_\psi)$ with $K_\psi = \{(\psi^j)^* : j \in \mathbb{Z}\}$ and o_ψ the order of ψ in the group of Dirichlet characters modulo N.

For real y and a Dirichlet character χ modulo N with χ not principal, put

$$G(y) = \sum_{1 \leq n \leq y} \Lambda(n)\Big(\frac{y}{n} - 1\Big) - y\big(\log y - 1 - \gamma\big), \tag{2.1}$$

$$g(y,\chi) = \sum_{1 \leq n \leq y} \Lambda(n)\chi(n)\Big(\frac{y}{n} - 1\Big) + y\,\frac{L'(1,\chi)}{L(1,\chi)}, \tag{2.2}$$

where γ is Euler's constant, and $g(y,\chi) = G(y)$ if χ is principal. We call G the **Riesz-weighted Chebyshev function** and $g(\cdot,\chi)$ its **twist** by χ, both with their main terms removed: by partial summation

$$G(y) = -y\int_y^\infty \Big(\sum_{n \leq t} \Lambda(n) - t\Big)\frac{dt}{t^2},$$

and $G(y)/y$ is the Riesz mean of order one of $\sum_n \Lambda(n)n^{-1}$ minus its main term $\log y - 1 - \gamma$ [47]. For a character level $\mathfrak{L} = (K, \psi_{\mathrm{exc}}, o)$ the **weighted Chebyshev function** is the positive combination

$$G_{\mathfrak{L}}(y) = G(y) + \sum_{\psi \in K \setminus \{\mathbf{1}\}} w_{\mathfrak{L}}(\psi)\,\mathrm{Re}\,g(y,\psi), \qquad w_{\mathfrak{L}}(\psi) = \begin{cases} 1, & \psi = \psi_{\mathrm{exc}}, \\ (3o+1)^{-1}, & \text{otherwise.} \end{cases} \tag{2.3}$$

A character level $\mathfrak{L}$ is **zero-free above** $\theta \in \mathbb{R}$ if $L(s,\psi) \neq 0$ for all $\psi \in K$ and all s with $\mathrm{Re}\,s > \theta$.

Landau's argument needs only that the weights are positive. The weight $(3o+1)^{-1}$ of the cyclic level is therefore not chosen for positivity: it is the weight that the averaging over the set T of a host produces (see the remark before Definition 3.2).

Proposition 2.2. *Let $\mathfrak{L}$ be a character level, let $\lambda > 1$ and let $0 < \theta < 1$. Suppose that for every $y \in \mathbb{R}$ there is $Y \geq y$ such that*

$$G_{\mathfrak{L}}(x) \leq x^\theta \qquad \text{for all } x \in [Y, Y^\lambda].$$

Then $\mathfrak{L}$ is zero-free above θ.

Proof. In outline. The weighted Chebyshev function $G_{\mathfrak{L}}$ is the Riesz-weighted Chebyshev function of a finite family of Dirichlet characters with positive weights, namely the characters of $\mathfrak{L}$ with the weights $w_{\mathfrak{L}}$. Zero-freeness of that family above θ is the same as zero-freeness of $\mathfrak{L}$ above θ. The statement is a window version, due to the argument itself, of Landau's classical theorem on Dirichlet series with non-negative coefficients [76], for such families, applied with $\delta = \lambda - 1$

and constant 1: an upper bound $G_{\mathfrak{L}}(y) \le Cy^{\theta}$ on intervals $[Y, Y^{1+\delta}]$ for arbitrarily large Y suffices. One attaches pole data to the family (the logarithmic derivatives of the L-functions, with their poles at the zeros) and runs Landau's argument on windows: a zero ρ with $\operatorname{Re}\rho > \theta$ would force $G_{\mathfrak{L}}(y)$ to exceed y^{θ} somewhere on every far window. The argument uses the positivity of the weights, not of the coefficients of the weighted Dirichlet series: the Laplace transform of $G_{\mathfrak{L}}$ has its poles at the zeros, its residue at a real zero is a positive combination of multiplicities, and high derivatives of the transform at a real point to the right of all poles, compared with the window bound, exclude a pole of real part above θ. The statement therefore holds for every character level, with or without an exceptional character. □

The hypothesis concerns a sparse set of scales: windows of fixed logarithmic ratio λ, and nothing in between. Windows are what the arithmetic of the later sections can provide, with $\lambda = \frac{3}{2}$.

The bound on windows is produced by families of auxiliary fields, which Sections 3 and 6 define, partly in outline; for the statements of this section the following description suffices. A host family with parameter ε (Definition 3.3) has an unbounded set W, its scales, and assigns to every $x \in W$ a non-empty finite set of members. Each member carries a host, that is, an abelian CM field with an explicit description of its Galois group $\{\pm 1\}^2 \times A$ (Definition 3.1), a finite subset T of A, and lattice data. The family is admissible for a level $\mathfrak{L}$ if from some scale on its members satisfy the host conditions for $\mathfrak{L}$ (Definition 3.2): the characters of $\mathfrak{L}$ occur among the characters of the host, and the other characters of the host have square-root bounds at the scale. It is tied with exponents (e_D, e_L) (Definition 6.13) if at every large scale in W the lattice data of its members form a slope datum, with the inert prime sums of the host of one of them as its prime-sum entries, which satisfies the slope conditions of Section 4 with divisor and loss terms of size at most $x^{e_D\varepsilon}$ and $x^{e_L\varepsilon}$ relative to the parameters of the datum. The point is that e_D and e_L do not depend on ε.

Proposition 2.3. *Let $\psi \in \mathcal{P}$ and $\lambda > 1$. Suppose there are $e_D, e_L \in \mathbb{R}$ and $\varepsilon_0 > 0$ such that for every $\varepsilon \in (0, \varepsilon_0)$ there is a host family with parameter ε that is admissible for $\mathfrak{L}_\psi$, is tied with exponents (e_D, e_L), and whose scales contain λ-windows. Then $L(s, \psi) \neq 0$ for all s with $\operatorname{Re} s > \frac{1}{2}$.*

Proof. Since $\psi = (\psi^1)^*$ lies in K_ψ, it suffices to show that $\mathfrak{L}_\psi$ is zero-free above $\frac{1}{2}$. Let $\operatorname{Re} s > \frac{1}{2}$ and $m = \min(\operatorname{Re} s - \frac{1}{2}, \frac{1}{2}) > 0$. Choose $\varepsilon \in (0, \varepsilon_0)$ with $(e_D + e_L + 7)\varepsilon \le m/2$ and take a family $\mathcal{F}$ as in the hypothesis, with scales W. By Proposition 2.4 there is x_1 with $G_{\mathfrak{L}_\psi}(x) \le x^{1/2+(e_D+e_L+7)\varepsilon} \le x^{\theta}$, $\theta = \frac{1}{2} + m/2$, for all $x \in W$ with $x \ge \max(x_1, 1)$. As W contains λ-windows, so does the set of x with $G_{\mathfrak{L}_\psi}(x) \le x^{\theta}$. Since $0 < \theta < 1$, Proposition 2.2 shows that $\mathfrak{L}_\psi$ is zero-free above θ, and $\theta < \operatorname{Re} s$. □

Its proof uses the following bound at the scales of a tied family. This is where the uniformity of the exponents is spent: since e_D and e_L do not depend on ε, the exponent $\frac{1}{2} + (e_D + e_L + 7)\varepsilon$ comes as close to $\frac{1}{2}$ as we wish when ε is small.

Proposition 2.4. *Let $\mathfrak{L}$ be a character level, let $e_D, e_L \in \mathbb{R}$ and $\varepsilon > 0$, and let $\mathcal{F}$ be a host family with parameter ε and scales W that is admissible for $\mathfrak{L}$ and tied with exponents (e_D, e_L). Then there is x_1 such that*

$$G_{\mathfrak{L}}(x) \le x^{1/2+(e_D+e_L+7)\varepsilon} \qquad \text{for all } x \in W \text{ with } x \ge x_1.$$

Proof. Let C, x_0 be the constants of admissibility (Definition 3.3) and C_K, c^*, e_K, C_L, x_0' those of tiedness (Definition 6.13; C_L is not used here). Apply Proposition 4.11 to $\mathfrak{L}$ with the exponents (e_D, e_L), the constants (C_K, c^*, e_K, C) and ε; this gives a threshold x_1', and we put $x_1 = \max(x_1', x_0, x_0')$. For $x \in W$ with $x \geq x_1$ the family is chain-tied at x. By Lemma 6.14 some member $i \in I(x)$ gives a core-tied pair $(\mathfrak{H}_{x,i}, T_{x,i})$ with the same exponents and constants. By admissibility the same pair satisfies the host conditions for $\mathfrak{L}$ at scale x, so Proposition 4.11 applies at x. □

Two results combined in this proof are treated later. Proposition 4.11 is the bound at a single scale from a single host whose lattice data satisfy the slope conditions, a core-tied host in the terminology of the development; it is the output of the bridge of Section 4. Lemma 6.14 says that a family that is chain-tied at x, as a tied family is at every large scale in its set of scales, has a core-tied member at x; it unwinds the definitions and is stated in Section 6, after them. Core-tied pairs are defined in Definition 3.2; chain ties, the index sets $I(x)$ and the constants of tiedness in Definition 6.13.

It remains to produce tied families. This is the content of the next proposition, the junction of the three arithmetic parts of the argument: the supply of hosts on windows (Proposition 3.5), the bound for Tate–Shafarevich groups of their fields (Theorem 9.1), and the construction that turns this bound into tied families (Proposition 6.20). In its proof the Sha bound V is that of Definition 3.4, and the rank t and the size parameter b are those of Definition 6.1 and Definition 6.3.

Proposition 2.5. *For every $\psi \in \mathcal{P}$ there are $e_D, e_L \in \mathbb{R}$ and $\varepsilon_0 > 0$ such that for every $\varepsilon \in (0, \varepsilon_0)$ there is a host family with parameter ε that is admissible for $\mathfrak{L}_\psi$, is tied with exponents (e_D, e_L), and whose scales contain $\frac{3}{2}$-windows.*

Proof. Fix ψ. By Theorem 9.1, ψ satisfies the window Sha bound with exponent 401. Let P be the requirement on a host $\mathfrak{H}$ with set T at scale x formed by (P2) and (P3) of Proposition 3.5: the host conditions for $\mathfrak{L}_\psi$ at scale x with parameters $(\varepsilon, 4)$, and the Sha bound $\mathrm{V}(E; t, \sqrt{b})$. By Proposition 3.5, hosts satisfying P exist on a set of scales containing $\frac{3}{2}$-windows. The Sha hypothesis of Proposition 6.20 holds on the class $\mathcal{C}_P$ of hosts satisfying P (in the sense of Proposition 6.20) by (P3), which every member satisfies. Hence Proposition 6.20 with $\lambda = \frac{3}{2}$ gives e_D, e_L, ε_0 and, for each $\varepsilon < \varepsilon_0$, a tied family in $\mathcal{C}_P$. This family has scales containing $\frac{3}{2}$-windows, and it is admissible for $\mathfrak{L}_\psi$ with constant 4 because its members satisfy (P2). □

With Proposition 2.5 the proof of Theorem 1.1 is reduced to five results, which the rest of the paper treats: Proposition 4.11, in Section 4; Proposition 3.5, in Section 3; Lemma 6.14 and Proposition 6.20, in Section 6; and Theorem 9.1, in Section 9. The next section introduces the hosts and outlines the proof of Proposition 3.5 from Theorem 9.1.

3 Hosts and their supply

This section introduces the auxiliary fields of the argument and shows that they exist where the reduction needs them. Its main result is Proposition 3.5: if ψ satisfies the window Sha bound, which Section 9 establishes, in outline, for every primitive character (Theorem 9.1), then for every ε there are, at all scales of a set containing $\frac{3}{2}$-windows, hosts that satisfy the host conditions for the cyclic level of ψ and whose fields satisfy a Sha bound. The analytic input is the existence of good window pairs (Proposition 3.6), which comes from a mean-square argument over quadratic twists. The known theorems used are the class field theory of cyclotomic fields [132] and the classical bound for L'/L near $s = 1$ in the absence of nearby zeros [89].

3.1 Hosts and host families

A host is an abelian CM field inside a cyclotomic field, together with an explicit description of its Galois group and of its Artin map. The abstract part of this description is a frame: a finite abelian group A that maps onto $(\mathbb{Z}/2^{e+2}\mathbb{Z})^\times/\{\pm1\}$, with a cyclic subgroup A_0 of index o. In the proof the index is the order of ψ, and the characters of $\{\pm1\}^2 \times A$ that are trivial on A_0 carry the powers of ψ.

Definition 3.1. For an integer $e \geq 2$ let $H_e = (\mathbb{Z}/2^{e+2}\mathbb{Z})^\times/\{\pm1\}$. A **frame** is a tuple (e, A, h, a_0) of an integer $e \geq 2$, a finite abelian group A, a surjective homomorphism $h\colon A \to H_e$ and an element $a_0 \in A$ such that h is injective on the cyclic subgroup $A_0 = \langle a_0 \rangle$. Its **index** is $o = [A : A_0]$, its **order** is $|A|$, and its **group** is $\mathcal{G} = \{\pm1\} \times \{\pm1\} \times A$, with the sign characters $\epsilon_1, \epsilon_2\colon \mathcal{G} \to \{\pm1\}$ given by the two first coordinates and a fixed element $c \in \{\pm1\}^2 \times \{0\}$; for a host, (H5) below gives $c = \alpha(-1)$, so $\epsilon_2(c) = (-1/d) = -1$.

A **host** for a frame (e, A, h, a_0) is a tuple $\mathfrak{H} = (q, d, Q, E, \iota, \alpha)$ where

(H1) q and d are primes with $q \equiv 3 \pmod 8$, $d \equiv 7 \pmod{16}$ and $(-d/q) = 1$;

(H2) $Q \geq 1$ is an integer coprime to qd with $2^{e+2} \mid Q$ and $(-d/p) = 1$ for every odd prime $p \mid Q$;

(H3) E is a CM field, Galois over $\mathbb{Q}$, with $[E : \mathbb{Q}] = 4|A|$, contained in $\mathbb{Q}(\zeta_{Qqd})$ and containing $\sqrt{-d}$;

(H4) $\iota\colon \mathrm{Gal}(E/\mathbb{Q}) \to \mathcal{G}$ is an isomorphism with $\sigma(\sqrt{-d}) = \epsilon_2(\iota\sigma)\sqrt{-d}$, under which complex conjugation (for every embedding $E \to \mathbb{C}$) corresponds to c;

(H5) $\alpha\colon (\mathbb{Z}/Qqd\mathbb{Z})^\times \to \mathcal{G}$ is the composite of the Artin map of $\mathbb{Q}(\zeta_{Qqd})/\mathbb{Q}$ with restriction to E and ι; it is surjective, $\alpha(-1) = c$, $\epsilon_2(\alpha(u)) = (u/d)$, the A-component of $\alpha(u)$ maps under h to the class of u in H_e, $\alpha(u) \in \{1, (1, -1, 0)\}$ if $u \equiv 1 \pmod{Qq}$, and $\alpha(u)^2 = 1$ if $u \equiv 1 \pmod{Qd}$.

For a character θ of $\mathcal{G}$ let χ_θ be the Dirichlet character $u \mapsto \theta(\alpha(u))$ modulo Qqd and χ_θ^* the primitive character inducing it; the host $\mathfrak{H}$ is fixed in each use and is not shown in the notation χ_θ, χ_θ^*. Let $X^\pm(\mathfrak{H})$ be the set of characters θ of $\mathcal{G}$ that are trivial on $\{(1,1)\} \times A_0$ and satisfy $\theta(1, -1, 0) = \pm1$. A host is moreover subject to further conditions of the same nature on its CM type, on an element on which $\mathrm{Gal}(E/\mathbb{Q})$ acts by ϵ_1, and on the Artin images of the residue classes of d, which we do not reproduce. The reader should think of a host as an abelian CM field cut out by the characters of interest, together with an explicit description of its Galois group.

Thus every character θ of $\mathcal{G}$ is a Dirichlet character χ_θ modulo Qqd. The second sign records the action of $\mathrm{Gal}(E/\mathbb{Q})$ on $\sqrt{-d}$, and for the ranked hosts of Section 6 the first records the action on $\sqrt{-q}$. The characters in $X^+(\mathfrak{H}) \cup X^-(\mathfrak{H})$ are, up to the two signs, the characters of A/A_0, a group of order o; roughly, they are the characters of the cyclic level of ψ and their twists by the quadratic characters attached to q and d. The fields $\mathbb{Q}(\zeta_{2^{e+2}} + \zeta_{2^{e+2}}^{-1}, \sqrt{-q}, \sqrt{-d})$, whose Galois group is $\{\pm1\}^2 \times (\mathbb{Z}/2^{e+2}\mathbb{Z})^\times/\{\pm1\}$ by elementary class field theory [132], appear to be the model hosts, for the frames with $A = (\mathbb{Z}/2^{e+2}\mathbb{Z})^\times/\{\pm1\}$. A host has degree $4|A|$, and at a scale x the order of A is about $x^{2+8\varepsilon}$ by (P1) of Proposition 3.5: the auxiliary fields have degree polynomial in the scale.

The host conditions say that a host sees the characters of a level and that its other characters are harmless at the scale x. Condition (E1) puts the characters of the level among those of the

host. Condition (E3) asks the finite set $T \subseteq A$ to meet A_0 in a proportion $4/(3o+1)$ and every other coset of A_0 in a proportion $3/(3o+1)$, up to $x^{-1+\varepsilon}$. Then $\sum_{t\in T}\phi(t)$ is $|T|/(3o+1)$ for every non-trivial character ϕ of A/A_0, up to an error $o\,x^{-1+\varepsilon}|T|$, and this is where the weights of the cyclic level come from. Condition (E4) asks the remaining characters, the controlled characters, for a square-root bound at the point x and a bound for the logarithmic derivative at $s=1$. The second half of the definition, the notion of a core-tied pair, is the hypothesis under which Section 4 bounds $G_{\mathfrak{L}}(x)$; it refers to slope data and to the size and error conditions defined there, and may be skipped until then.

Definition 3.2. Let $\mathfrak{L}=(K,\psi_{\mathrm{exc}},o)$ be a character level, $\mathfrak{H}$ a host for a frame (e,A,h,a_0) of index o, $T\subseteq A$ a finite set, and $x>1$, $\varepsilon>0$, $C\geq 0$ reals. The **controlled characters** of $\mathfrak{H}$ relative to $\mathfrak{L}$ are the $\theta\in X^-(\mathfrak{H})$ together with the $\theta\in X^+(\mathfrak{H})$ with $\chi_\theta^*\notin K$; these are the characters of the host whose twisted Chebyshev functions do not enter $G_{\mathfrak{L}}$, and (E4) controls their size. Put $\kappa(b)=4/(3o+1)$ for $b\in A_0$ and $\kappa(b)=3/(3o+1)$ otherwise. The pair $(\mathfrak{H},T)$ satisfies the **host conditions for $\mathfrak{L}$ at scale x with parameters** (ε,C) if

(E1) every $\psi\in K$ is χ_θ^* for some $\theta\in X^+(\mathfrak{H})$;

(E2) if ψ_{exc} is present, it is $\chi_{\epsilon_1}^*$; if it is absent, $\chi_{\epsilon_1}^*\notin K$;

(E3) for every $b\in A$,
$$\Big|\,\#\{t\in T: t-b\in A_0\}-\kappa(b)\,|T|\,\Big|\leq x^{-1+\varepsilon}|T|;$$

(E4) for every controlled character θ, $|g(x,\chi_\theta^*)|\leq x^{1/2+\varepsilon}$ and $\big|L'(1,\chi_\theta^*)/L(1,\chi_\theta^*)\big|\leq C\log x$.

Let $\mathfrak{A}_{\mathfrak{H}}$ be the arithmetic datum of $\mathfrak{H}$ on A. The pair $(\mathfrak{H},T)$ is **core-tied at scale x with exponents** (e_D,e_L,C_K,c^*,e_K) **and parameter** ε if there exist a slope datum at scale x with parameter ε and index set U_0, a map $\mathrm{tr}\colon U_0\to A$ whose image contains A_0, an integer $P\geq 1$, a finite set $T_2\subseteq A_0$, a function $\tilde a$ on A, weights $w_{u,j}$ ($u\in A$, $1\leq j\leq m$) and a function η on the characters of $\{\pm1\}^2\times A$, such that:

(K1) $\eta(\phi)=L'(1,\chi)/L(1,\chi)$ for every non-trivial character ϕ and every primitive χ with $\chi(n)=\phi_{\mathfrak{A}_{\mathfrak{H}}}(n)$ for all $n\geq 1$;

(K2) the identity (Id) of Proposition 4.5 holds between the slope datum and $(\mathfrak{A}_{\mathfrak{H}},P,T,\tilde a,w,\eta)$, so that its inert prime sums are $S_u=\sum_{t\in T}\sum_j A_{\mathrm{in}}(\tilde a(t),t,w_{\mathrm{tr}(u),j})$;

(K3) the slope datum satisfies the slope conditions for a constant vector with c_{Div}-exponent $e_{\mathrm{Div}}\leq e_D$, loss exponent $e_{\mathrm{Loss}}\leq e_L$ and $c_{\mathrm{HH}}\leq c^*$, and the error conditions (Z3) of Definition 4.4 with constants (C_K,e_K);

(K4) $B_0\leq 2|A|mbx^{\varepsilon}$, $N_c\leq x^{2\varepsilon}|T_2|kb/|A|^2$, the size and thin-scale conditions (G1), (G2), (G4) of Definition 4.3 hold for $(A,\mathfrak{A}_{\mathfrak{H}},\eta,P,m,T,T_2,\tilde a,w)$ at x, with the clauses of these conditions that are indexed by a translate u required only for $u\in A_0$, and the cost conditions (Z6) and remainder-size conditions (Z7) of Definition 4.4 hold, with $T_2\subseteq A_0$.

Host families collect hosts over a set of scales. Their members carry, besides a host, the lattice data of Section 6; admissibility, defined here, concerns the hosts alone, while tiedness, the other property used in Section 2, concerns the lattice data and is defined in Section 6 (Definition 6.13).

Definition 3.3. Let $\lambda > 1$. A set $W \subseteq \mathbb{R}$ **contains λ-windows** if for every $y \in \mathbb{R}$ there is $Y \geq y$ with $[Y, Y^{\lambda}] \subseteq W$.

Let $\varepsilon > 0$. A **host family with parameter** ε is a tuple $\mathcal{F} = (x_1, W, I, (\mathfrak{M}_{x,i}))$ of a real number $x_1 > 1$, above which the members carry size data, an unbounded set $W \subseteq \mathbb{R}$ (its **scales**), index sets $I(x)$ for $x \in \mathbb{R}$, non-empty for $x \in W$, and for each x and $i \in I(x)$ a datum $\mathfrak{M}_{x,i}$ carrying a frame, a host $\mathfrak{H}_{x,i}$ for it, a finite subset $T_{x,i}$ of A, and size data at scale x (Definition 6.11), together with parameter choices for these data, which the parametrised host families of Section 6 make explicit; the members are instances of the master datum of Definition 6.10, whose further structure we do not reproduce here.

A host family $\mathcal{F}$ is **admissible for** a character level $\mathfrak{L}$ if there are $C \geq 0$ and x_0 such that for all $x \geq x_0$ and $i \in I(x)$ the pair $(\mathfrak{H}_{x,i}, T_{x,i})$ satisfies the host conditions for $\mathfrak{L}$ at scale x with parameters (ε, C).

3.2 Window pairs and the supply of hosts

Hosts are built on pairs of primes (q, d) of a prescribed shape, and their use in the families of Section 6 depends on one arithmetic quantity of their fields, a weighted total length of Tate–Shafarevich groups. The next definition fixes both, and also the abelian fields to which the bound is applied: by the Galois correspondence, a (ψ, e, q, d)-field is the compositum of the field cut out by ψ, of $\mathbb{Q}(\sqrt{-q}, \sqrt{-d})$ and of the real subfield of $\mathbb{Q}(\zeta_{2^{e+2}})$.

Definition 3.4. Let E be a number field, ℓ a prime, S_ℓ the set of places of E above ℓ, G_{E,S_ℓ} the Galois group of the maximal extension of E unramified outside S_ℓ, and $k \in \mathbb{Z}$. Put

$$\mathrm{Sha}^1_\ell(E,k) = \ker\Big(H^1\big(G_{E,S_\ell}, \mathbb{Q}_\ell/\mathbb{Z}_\ell(-k)\big) \to \prod_{v \in S_\ell} H^1\big(E_v, \mathbb{Q}_\ell/\mathbb{Z}_\ell(-k)\big)\Big),$$

with continuous cohomology, and let $\mathrm{len}_\ell(E,k) \in \mathbb{N} \cup \{\infty\}$ be its length as a $\mathbb{Z}_\ell$-module. For an integer $t \geq 1$ and $B \in \mathbb{R}$, the field E satisfies the **Sha bound** $\mathrm{V}(E; t, B)$ if $\mathrm{len}_\ell(E,k) < \infty$ for all primes ℓ and all $1 \leq k < t$, and

$$\sum_{\ell} \log \ell \sum_{k=1}^{t-1} \mathrm{len}_\ell(E,k) \leq B.$$

Let $0 < \varepsilon < 1/52$. A **window pair at scale** Y is a pair of primes (q, d) with

$$q \equiv 3 \ (8), \quad d \equiv 7 \ (16), \quad \tfrac{1}{2} Y^{\varepsilon/5} < \log q \leq Y^{\varepsilon/5}, \quad \tfrac{1}{2} Y^{\varepsilon/5} \log^2 Y < \log d \leq Y^{\varepsilon/5} \log^2 Y,$$

$q \mid d+1$, and $p \mid d+1$ for every odd prime $p \leq Y^{\varepsilon/5}$. A **page function** is a function $e_\varepsilon \colon \mathbb{R} \to \mathbb{N}$ with $x^{2+8\varepsilon} \leq 2^{e_\varepsilon(x)} \leq 2x^{2+8\varepsilon}$ for all $x \geq 1$.

For $\psi \in \mathcal{P}$ of modulus N_ψ, integers $e \geq 0$, primes q, d, a number field E is a (ψ, e, q, d)**-field** if for some N divisible by 2^{e+2}, N_ψ, q and d, there is an embedding $E \hookrightarrow \mathbb{Q}(\zeta_N)$ whose image is the fixed field of the automorphisms $\zeta_N \mapsto \zeta_N^a$ with

$$a \equiv \pm 1 \ (2^{e+2}), \qquad \psi(a) = 1, \qquad (a/q) = (a/d) = 1.$$

For $C_t > 0$, the character ψ satisfies the **window Sha bound with exponent** C_t if for every $\varepsilon \in (0, 1/52)$ and every page function e_ε there are $J \geq 1$ and X_0 such that for all $X \geq X_0$ and every window pair (q, d) at scale $X^{1/J}$ there is Y with $X^{1/J} \leq Y$ and $Y^2 \leq X$ such that every $(\psi, e_\varepsilon(x), q, d)$-field E with $x \in [Y, Y^2]$ satisfies $\mathrm{V}\big(E; \lfloor x^{C_t} \rfloor, \exp(x^\varepsilon/2)\big)$. The exponent used throughout is $C_t = 401$.

The group $\mathrm{Sha}^1_\ell(E,k)$ is the Tate–Shafarevich group of $\mathbb{Q}_\ell/\mathbb{Z}_\ell(-k)$ in the sense of Poitou–Tate duality [97]; it is dual to a second cohomology group of $\mathbb{Z}_\ell(k+1)$, which Soulé relates to K-theory [118]. The Sha bound $\mathrm{V}(E;t,B)$ bounds, with weights $\log\ell$, the total length of these groups over all ℓ and all twists $1\le k<t$. In a window pair at scale Y one has $\log q\asymp Y^{\varepsilon/5}$ and $\log d\asymp Y^{\varepsilon/5}\log^2 Y$, so the conductors are enormous compared with the scale. Since $d\equiv -1$ modulo q and modulo every odd prime $p\le Y^{\varepsilon/5}$, all these primes split in $\mathbb{Q}(\sqrt{-d})$, which is the kind of splitting that the conditions (H1) and (H2) on hosts require. The page function, in the terminology of the development, fixes the 2-power level 2^{e+2} of the fields at the scale x, with 2^e about $x^{2+8\varepsilon}$.

The supply of hosts is the following. In it, a rank datum (Definition 6.1, Section 6) fixes the ranks of the hermitian spaces that the theta-lift construction attaches to a host; its rank t is at most x^{400}, and $b=\exp(x^\varepsilon)$ is the size parameter of the scale x (Definition 6.3). Condition (P3) is therefore the Sha bound of the window Sha bound, with $\sqrt{b}=\exp(x^\varepsilon/2)$ and a rank below x^{401}.

Proposition 3.5. *Let $\psi\in\mathcal{P}$ satisfy the window Sha bound with exponent 401. For every $\varepsilon\in(0,1/52)$ there is a set of scales containing $\frac{3}{2}$-windows such that, at every scale x in this set beyond a threshold, there are a rank datum and a host $\mathfrak{H}$ with distinguished set $T\subseteq A$ for which*

- **(P1)** *the frame of $\mathfrak{H}$ has index o_ψ and order between $x^{2+8\varepsilon}$ and $2x^{2+8\varepsilon}$, and its field E is a (ψ,e,q,d)-field for a window pair (q,d) at a scale at most x;*
- **(P2)** *$(\mathfrak{H},T)$ satisfies the host conditions for $\mathfrak{L}_\psi$ at scale x with parameters $(\varepsilon,4)$;*
- **(P3)** *E satisfies the Sha bound $\mathrm{V}(E;t,\sqrt{b})$, with $t\le x^{400}$ the rank and b the size parameter of the scale;*
- **(P4)** *$\mathfrak{H}$ satisfies further conditions, with constants depending only on ψ and ε, on pairs of primes, on its auxiliary primes and on zero-free regions of the L-functions of its real characters (as in (Rb1) of Proposition 3.9), which we do not reproduce.*

A host with (P1) and (P4) is called ***good****.*

Proof. In outline. Fix ε and a page function e_ε. There is an integer $c\ge 1$, depending only on ψ, such that for all large e and every suitable window pair there is a frame of order $c\,2^e$ and index o_ψ whose field satisfies the conditions of the construction for that pair; put $k=c^{1/(2+8\varepsilon)}$. The scales combine as follows. Given y, choose X large. By Proposition 3.6, applied with $M=N_\psi$, there is a window pair (q,d) at scale $X^{1/J}$ with the properties (P-bound) and (P-box) of Proposition 3.6 (a **good window pair**), and the window Sha bound gives for this pair a number Y' with $X^{1/J}\le Y'$ and $Y'^2\le X$ such that every $(\psi,e_\varepsilon(x'),q,d)$-field with $x'\in[Y',Y'^2]$ satisfies $\mathrm{V}(E;\lfloor x'^{401}\rfloor,\exp(x'^\varepsilon/2))$. Put $Z=kY'$, with X chosen so large that $Y'\ge k^3$; then every $x\in[Z,Z^{3/2}]$ satisfies $x/k\in[Y',Y'^2]$ and $x\le X$, so the pair sits at the scale $X^{1/J}\in[x^{1/J},x]$, the page $e=e_\varepsilon(x/k)$ gives $x^{2+8\varepsilon}\le c\,2^e\le 2x^{2+8\varepsilon}$, and the (ψ,e,q,d)-fields satisfy the Sha bound with $t=\lfloor x^{400}\rfloor$. For each frame of this order and index there is a rank datum satisfying conditions of the construction, which we do not reproduce. By Proposition 3.9, such a field with such a pair extends to a good host whose real characters with q or d in the conductor have zero-free boxes. The set T is chosen so that the share condition (E3) holds; we do not reproduce this choice. Conditions (E1) and (E2) come from the construction of the field from ψ, and the pointwise bounds (E4) from (P-bound) of Proposition 3.6 and the zero-free boxes. The Sha bound passes to (P3) by monotonicity of V in t and B. □

The proof combines the window Sha bound with the results of the next subsection, good window pairs at every large scale and good hosts built on them, and it shows how one window pair serves a whole window $[Z, Z^{3/2}]$. The goodness conditions (P4) enter the hypothesis of Proposition 6.20.

3.3 Good window pairs

A host built on a window pair (q, d) has, besides the characters of the level, characters of the form $n \mapsto \xi(n)(n/q)^i(n/d)^j$ with $(i, j) \neq (0, 0)$ and ξ of small modulus. The host conditions ask each of them for a square-root bound at the scale x and for a bound on L'/L at $s = 1$; the goodness conditions (P4) ask in addition that the real ones have no zeros in a box to the right of $1 - 1/400$. A good window pair is one for which the square-root bounds hold on a whole range of scales $[Y, Y^A]$ and the real twists have zero-free boxes.

Proposition 3.6. *Let $M \geq 1$, $0 < \varepsilon < 1/52$ and $A \geq 1$. There is Y_1 such that for every $Y \geq Y_1$ there is a window pair (q, d) at scale Y with $\gcd(M, qd) = 1$ and the following two properties.*

(P-bound) *For every character ξ modulo M and $(i, j) \in \{0, 1\}^2 \setminus \{(0, 0)\}$, the primitive character χ inducing $n \mapsto \xi(n)(n/q)^i(n/d)^j$ modulo $2Mqd$ satisfies, for all $y \in [Y, Y^A]$,*

$$|g(y, \chi)| \leq y^{1/2+\varepsilon/2}, \qquad \Big|\frac{L'(1, \chi)}{L(1, \chi)}\Big| \leq 4 \log y.$$

(P-box) *For every real primitive character ρ of modulus dividing $8M$ and $(i, j) \neq (0, 0)$, the primitive character χ inducing $n \mapsto \rho(n)(n/q)^i(n/d)^j$ satisfies $L(s, \chi) \neq 0$ for $\operatorname{Re} s > 1 - 1/400$ and $|\operatorname{Im} s| \leq 8M \exp(Y^{\varepsilon/5}) \exp(Y^{\varepsilon/5} \log^2 Y)^j$.*

Proof. Apply Proposition 3.7 with $N_0 = 8M$, $\tau = \varepsilon/2$, $K = A$ and $z' = Y^{\varepsilon/5} \log^2 Y$. Pick $q \in \mathcal{Q}^{\mathrm{good}}$ and $d \in \mathcal{D}_q^{\mathrm{good}}$, which is possible since the good sets have proportion at least $1 - \eta$ of non-empty sets.

- The congruence and size conditions of a window pair hold by the definition of $\mathcal{Q}$ and $\mathcal{D}_q$; in particular $q \mid d + 1$ and the odd primes $p \leq Y^{\varepsilon/5}$ divide $d + 1$.
- Coprimality to M holds since q, d exceed M.
- The Chebyshev bounds of (P-bound) and the zero-free boxes of (P-box) are the goodness of the twists, with moduli dividing $8M$.
- The bound on $L'/L(1, \chi)$ follows from the zero-free box by Lemma 3.8.

□

Good pairs exist because almost all pairs are good. The next proposition makes this precise in two stages, first for q and then for d given a good q; its sets $\mathcal{Q}$ and $\mathcal{D}_q$ consist of the primes in the ranges of window pairs that satisfy their congruence conditions.

Proposition 3.7. *Call a primitive Dirichlet character χ of modulus N (Q, C, τ, Y, X)-**good** if $\chi \neq \mathbf{1}$, $N \leq Q$, $L(s, \chi) \neq 0$ for $\operatorname{Re} s > 1 - 1/400$ and $|\operatorname{Im} s| \leq Q$, $|L'(1, \chi)/L(1, \chi)| \leq C \log Y$, and $|g(y, \chi)| \leq y^{1/2+\tau}$ for all $y \in [Y, X]$; here (n/q) denotes the Jacobi symbol. There is $C > 0$ such that for all $0 < \varepsilon < 1/52$, $\tau > 0$, $K \geq 1$, $0 < \eta < 1$ and $N_0 \geq 1$ there is Y_0 with the following property. Let $Y \geq Y_0$, $X \leq Y^K$ and $Y^{\varepsilon/5} \log^2 Y \leq z' \leq K^2 Y^{\varepsilon/5} \log^2 Y$, and put $z = Y^{\varepsilon/5}$. Let*

$\mathcal{Q}$ be the set of primes $q \equiv 3 \pmod 8$ with $e^{z/2} < q \le e^z$, and for $q \in \mathcal{Q}$ let $\mathcal{D}_q$ be the set of primes $d \le e^{z'}$ with $d \equiv 7 \pmod{16}$, $d > e^{z'/2}$, and $d+1$ divisible by $q\prod_{3\le p\le z} p$, the product over primes. Then both sets are non-empty, and there is $\mathcal{Q}^{\mathrm{good}} \subseteq \mathcal{Q}$ with $|\mathcal{Q}^{\mathrm{good}}| \ge (1-\eta)|\mathcal{Q}|$ such that for every $q \in \mathcal{Q}^{\mathrm{good}}$:

(Aq) *for every character ξ modulo N_0, the primitive character inducing $n \mapsto \xi(n)(n/q)$ is $(N_0 e^z, C, \tau, Y, X)$-good;*

(Ad) *there is $\mathcal{D}_q^{\mathrm{good}} \subseteq \mathcal{D}_q$ with $|\mathcal{D}_q^{\mathrm{good}}| \ge (1-\eta)|\mathcal{D}_q|$ such that for every $d \in \mathcal{D}_q^{\mathrm{good}}$, every ξ modulo N_0 and $a \in \{0,1\}$, the primitive character inducing $n \mapsto \xi(n)(n/q)^a(n/d)$ is $(N_0 e^{z+z'}, C, \tau, Y, X)$-good.*

Proof. In outline. The good primes are found in two stages, first q and then d for a fixed good q. In each stage the bad primes, those for which some twist fails one of the three conditions, are counted by mean-square estimates over the family of quadratic twists, and their share is at most η.

□

For the mean-square estimates in this proof, large-sieve inequalities for quadratic characters [52] are the natural classical comparison. The scales $y \in [Y, X]$ at which square-root bounds are required are tiny compared with the conductors, which exceed $\exp(Y^{\varepsilon/5}/2)$; presumably it is the large number of available moduli, rather than the length of the sums, that makes square-root cancellation typical. The bound for $L'/L(1,\chi)$ in (P-bound) comes from the zero-free box by a classical argument:

Lemma 3.8. *For all $0 < \varepsilon < 1/52$ and $A > 0$ there is Y_0 such that the following holds for all $Y \ge Y_0$, all Q with $Y^{\varepsilon/6} \le \log Q \le AY^{\varepsilon/5}\log^2 Y$, and every Dirichlet character $\chi \ne \chi_{0,N}$ of modulus $N \le Q$ such that $L(s,\chi) \ne 0$ for $\operatorname{Re} s > 1 - 1/400$, $|\operatorname{Im} s| \le Q$:*

$$\Big|\frac{L'(1,\chi)}{L(1,\chi)}\Big| \le 4\log Y.$$

Proof. In outline. A zero-free box of height Q to the right of $1 - 1/400$ gives the bound $|L'(1,\chi)/L(1,\chi)| \le C\log\log Q + C$ (a $\log\log$ bound), by the standard argument from the partial-fraction expansion of L'/L and Borel–Carathéodory [89]. With $\log Q \le AY^{\varepsilon/5}\log^2 Y$ one has $\log\log Q \le \frac{\varepsilon}{5}\log Y + O(\log\log Y)$, so $C\log\log Q + C \le 4\log Y$ for large Y provided $C\varepsilon/5 < 4$. That the constant of the $\log\log$ bound satisfies this for $\varepsilon < 1/52$ (that is, $C < 20/\varepsilon$, about 1040) is assumed here and was not checked. □

Finally, for a host built on a good window pair, the real characters ϱ with q or d in the modulus have zero-free boxes near $s = 1$, and the CM fields attached to the pair extend to good hosts; the second part is given only in outline.

Proposition 3.9. *Let $M \ge 1$, $0 < \varepsilon < 1/52$, $Y \le X$, and let $\mathfrak{H}$ be a host whose modulus Q has all its prime factors dividing $2M$ and its odd part dividing M. Suppose its primes (q, d) form a window pair at scale Y with properties (P-bound) on $[Y, X]$ and (P-box) for M. Then:*

(Rb1) *for every real primitive character ϱ occurring as some χ_θ^* of $\mathfrak{H}$ whose modulus is divisible by q or by d, $L(s,\varrho) \ne 0$ for $\operatorname{Re} s > 1 - 1/400$ and $|\operatorname{Im} s| \le 8M\exp(Y^{\varepsilon/5})\exp(Y^{\varepsilon/5}\log^2 Y)^j$, where $j = 1$ if d divides the modulus and $j = 0$ otherwise, and the modulus of ϱ satisfies a bound of the same shape, which we do not reproduce;*

(Rb2) *in particular, no such $L(s, \varrho)$ has a real zero $\beta > 1 - 1/400$.*

Moreover, at every sufficiently large scale, the CM fields attached to such a pair (q, d) that satisfy the conditions of the construction extend to hosts with further properties used in the supply of hosts; we give this part only in outline and do not reproduce the conditions.

Proof. In outline. A real character of the host with q or d in its conductor is, by the Artin map of the host, a twist $\rho(n)(n/q)^i(n/d)^j$ with $(i, j) \neq (0, 0)$ of a real character ρ of modulus dividing $8M$. This uses the decomposition of real characters into a part of modulus dividing $8M$ and Jacobi-symbol parts, together with the Kronecker-symbol computations at 2, q, d for $q \equiv 3$ (4) and $d \equiv 7$ (16). Then (P-box) gives (Rb1), and (Rb2) is the special case $s = \beta$ real. The extension to a host uses the structure of the Galois groups of the two-power cyclotomic host fields, which is elementary class field theory [132], the existence of weights in the weight ranges (Lemma 6.6), the rank-size bound (Lemma 6.4) and the existence of enough injective words (Lemma 6.5), which supply the rank data of the host, together with the existence of elements of small norm and of a small prime $\equiv 1$ modulo the level, which supply its auxiliary primes. □

The three elementary lemmas on the sizes of rank data used in this proof, Lemma 6.6, Lemma 6.4 and Lemma 6.5, are stated in Section 6 together with the rank data. With this, hosts are available on windows. The next section shows how one host, once it carries lattice data that satisfy the slope conditions, bounds $G_{\mathfrak{L}}$ at one scale.

4 The bridge: inert prime sums and slope data

This section outlines the proof of Proposition 4.11: at a single large scale x, a host that satisfies the host conditions for a level $\mathfrak{L}$ and is core-tied at x gives $G_{\mathfrak{L}}(x) \leq x^{1/2+(e_D+e_L+7)\varepsilon}$. The lattice data are taken as given, through the hypothesis that the host is core-tied; Sections 6–8 construct them. The argument has two halves. The first bounds the inert prime sums of the host from above by their main terms (Proposition 4.5). By definition, the prime-sum entries S_u of a slope datum are these sums. The loss condition (S3) bounds them by a quantity Θ_u of the lattices, made of local gains at the inert primes, up to a small loss; the slope inequalities then bound Θ_u by the main term. The second half is analytic: an average of inert prime sums over translates expands, by orthogonality of characters, into twisted Chebyshev functions, so that a bound for the excess becomes the bound for $G_{\mathfrak{L}}$ (Proposition 4.8). Apart from orthogonality, the analytic half uses only elementary estimates for sums of $\Lambda(n)/n$ [89]. The slope inequality plays here the part of the auxiliary function in Stepanov's method, from which the argument descends in its author's description (Section 1.3). The first half is where the argument is new; Section 11 says what we do not understand about it.

4.1 Inert prime sums

An arithmetic datum records, for a finite abelian group H, what a host records about the primes: inertia groups and Frobenius elements in $\Gamma_H = (\mathbb{Z}/2)^2 \times H$, written additively. For the datum of a host, $H = A$, the group Γ_H is the Galois group $\{\pm 1\}^2 \times A$ of the host field with the signs written additively, and Φ is the subgroup of index two that fixes $\sqrt{-d}$. A datum is of host type if it looks like the datum of a host at the ramified primes: outside a finite set S of primes, which contains 2 and q but not d, only d is ramified, with inertia of order two, and at the primes of S inertia and Frobenius fix $\sqrt{-d}$; for a host, S is the set of prime factors of Qq, which split in

$\mathbb{Q}(\sqrt{-d})$. The definition introduces the twisted Chebyshev functions $G_{\mathfrak{A}}(y,\phi)$ of the characters of Γ_H and the inert prime sums, whose excess over a main term is the quantity to be bounded.

Definition 4.1. Let H be a finite abelian group, $\Gamma_H = (\mathbb{Z}/2)^2 \times H$, and $\Phi \subset \Gamma_H$ the set of elements whose first $\mathbb{Z}/2$-coordinate is 0. Put $c_\infty = ((1,1),0)$ and, for $t \in H$ and $g \in \Gamma_H$, $\sigma_t(g) = c_\infty + g + ((0,0),t)$. Let χ_C be the character of Γ_H given by (-1) to the first $\mathbb{Z}/2$-coordinate.

An **arithmetic datum** on H is a tuple $\mathfrak{A} = (q, d, (I_p)_p, (\mathrm{Fr}_p)_p)$ of integers q, d, subgroups $I_p \subseteq \Gamma_H$ and elements $\mathrm{Fr}_p \in \Gamma_H$ indexed by the primes p. It is **of host type** if q, d are distinct odd primes and there is a finite set S of primes with $2, q \in S$ and $d \notin S$ such that $I_p = 0$ for $p \notin S \cup \{d\}$, $I_p \subseteq \Phi$ and $\mathrm{Fr}_p \in \Phi$ for $p \in S$, and $I_d = \{0, ((1,0),0)\}$. For a character ϕ of Γ_H and $n \geq 1$ put

$$\phi_{\mathfrak{A}}(n) = \prod_{p \mid n} \frac{1}{|I_p|} \sum_{g \in I_p} \phi\big(v_p(n)\mathrm{Fr}_p + g\big),$$

and, for a function η on the characters of Γ_H and real y,

$$G_{\mathfrak{A}}(y,\phi) = \sum_{1 \leq n \leq y} \Lambda(n)\,\phi_{\mathfrak{A}}(n)\Big(\frac{y}{n} - 1\Big) + y\,\eta(\phi) \quad (\phi \neq 1), \qquad G_{\mathfrak{A}}(y,1) = G(y).$$

A prime $p \notin \{2, q, d\}$ is **inert** for $\mathfrak{A}$ if $\mathrm{Fr}_p \notin \Phi$; for a host this means inert in $\mathbb{Q}(\sqrt{-d})$. For $P \geq 1$, $a \in \mathbb{R}$, $t \in H$ and $w\colon \Gamma_H \to \mathbb{R}$, the **inert prime sum** is

$$\begin{aligned} A_{\mathrm{in}}(a,t,w) = {} & \sum_{\substack{p \leq P \\ p \text{ inert}}} 2\log p \sum_{\substack{1 \leq i \leq P \\ i\mathrm{Fr}_p \notin \Phi}} \Big(\frac{a}{p^i} - w\big(\sigma_t(i\mathrm{Fr}_p)\big)\Big)^+ \\ & + \log d \sum_{i=1}^{P} \sum_{\substack{g \notin \Phi \\ g - i\mathrm{Fr}_d \in I_d}} \Big(\frac{a}{d^i} - w\big(\sigma_t(g)\big)\Big)^+, \end{aligned}$$

with $u^+ = \max(u, 0)$, and its **main term** is

$$M_\eta(a,w) = a\Big(\log a - 1 - \gamma + \mathrm{Re}\,\eta(\chi_C) - \frac{1}{|\Phi|}\sum_{\rho \in \Phi} \log w(\rho)\Big).$$

For a finite set $T \subseteq H$, numbers a_t ($t \in T$) and weights $w_{u,j}\colon \Gamma_H \to \mathbb{R}$ ($u \in H$, $1 \leq j \leq m$), the **excess at** u and the **mean excess** are

$$\Delta(u) = \sum_{t \in T}\sum_{j=1}^{m}\Big(A_{\mathrm{in}}(a_t, t, w_{u,j}) - M_\eta(a_t, w_{u,j})\Big), \qquad \overline{\Delta} = \frac{1}{|H|}\sum_{u \in H}\Delta(u).$$

The group H, the arithmetic datum $\mathfrak{A}$, the function η and the cut-off P are fixed in each use and are not shown in the notation $G_{\mathfrak{A}}(y,\phi)$, $A_{\mathrm{in}}(a,t,w)$, $M_\eta(a,w)$, $\Delta(u)$, $\overline{\Delta}$; where several choices occur, Δ and $\overline{\Delta}$ are said to be formed with $(T, (a_t), (w_{u,j}), \eta, P)$. The arithmetic datum of a host $\mathfrak{H} = (q, d, Q, E, \iota, \alpha)$ for a frame (e, A, h, a_0) is the datum on $H = A$ with q, d the primes of $\mathfrak{H}$, I_p the image under α of the units of $\mathbb{Z}/Qqd\mathbb{Z}$ that are $\equiv 1$ modulo the prime-to-p part of Qqd, and Fr_p the image under α of the unit that is $\equiv p$ modulo the prime-to-p part of Qqd and $\equiv 1$ modulo its p-part (for $p \nmid Qqd$ this is the class of p), written in coordinates $(\epsilon_2, \epsilon_1, A)$. It is of host type, with S the set of prime factors of Qq: these primes split in $\mathbb{Q}(\sqrt{-d})$ by (H1), (H2), so their inertia and Frobenius lie in Φ, and the primes outside $S \cup \{d\}$ are unramified.

The inert prime sum is a sum of Riesz type over the prime powers p^i with $i\mathrm{Fr}_p \notin \Phi$, in which the cut-off a/w depends on the class $\sigma_t(i\mathrm{Fr}_p)$ through the weight function w. By orthogonality of the characters of Γ_H, such sums are combinations of the functions $G_{\mathfrak{A}}(y,\phi)$ at the cut-offs $y = a/w(\rho)$, $\rho \in \Phi$. The main term $M_\eta(a,w)$ collects the main terms of the Riesz sums at these cut-offs, averaged over Φ, together with a term $a\,\mathrm{Re}\,\eta(\chi_C)$ from the quadratic character χ_C. For a host, $\eta(\phi)$ is $L'(1,\chi)/L(1,\chi)$ for the Dirichlet character χ of ϕ (condition (K1) of Definition 3.2), so that $G_{\mathfrak{A}}(y,\phi)$ is the twisted function $g(y,\chi)$ of Section 2.

4.2 Slope data

The excess is bounded by geometry. A slope datum is a collection of real numbers indexed by translates u and words ω, together with the inequalities between them that the argument uses, the slope conditions. The definition treats the numbers abstractly and then says which lattices they come from in the argument, with ranks, slopes and heights in the sense of Arakelov geometry [19], [20]. Two entries carry the bridge: the prime-sum entry S_u, which is not a lattice invariant, and the quantity Θ_u of the lattice data, a sum of local gains at the inert primes, whose correction terms are made of an excess part and of counts over a set of residue classes.

Definition 4.2. A **slope datum** consists of finite sets U, $\mathcal{T}$ and non-empty finite sets $\mathcal{W}(u)$ ($u \in U$), integers $m, f \geq 1$, real parameters $x, \varepsilon, a, b, b', k, L, L_a, B_0, N_c, \gamma_0, \beta, v, l_g$, numbers a_τ ($\tau \in \mathcal{T}$) and weights $w_{u,j,\rho} > 0$ ($u \in U$, $1 \leq j \leq m$, $1 \leq \rho \leq f$), and for each $u \in U$ and $\omega \in \mathcal{W}(u)$ real numbers

$$r_{u,\omega},\ T_{u,\omega},\ \mu_{u,\omega},\ \mu^{\max}_{u,\omega},\ h_{u,\omega},\ \mu'_{u,\omega},\ \mu'^{\max}_{u,\omega},\ E_{u,\omega,j}$$

and

$$\mu^\vee_u,\ \mu^\vee_{\min,u},\ F_u,\ \Theta_u,\ S_u,\ I_u,\ \mathrm{HH}_u,\ \delta_u,\ \eta_u,\ e_u.$$

In the argument most entries are invariants of hermitian lattices attached to the members of a host family, in the sense of the Harder–Narasimhan theory of hermitian vector bundles [121], [45], [19]: $r_{u,\omega}$, $\mu_{u,\omega}$, $\mu^{\max}_{u,\omega}$ and $h_{u,\omega}$ are the rank, slope, maximal slope and a height of the divisor lattice of the word ω; $\mu'_{u,\omega}$, $\mu'^{\max}_{u,\omega}$ are the slopes of a level-zero lattice; $\mu^\vee_u$, $\mu^\vee_{\min,u}$ are the dual slopes of a reference lattice; and $T_{u,\omega}$ is a prime. The entries F_u, Θ_u, I_u and δ_u are quantities of the size and row data of the members. Among them Θ_u is the lattice counterpart of the prime sum: it is twice the sum, over the inert primes p up to a cut-off, of $\log p$ times the net local gain of the lattices at p (a sum of local gains over the lifts of the member, minus a correction term), the gains being 0 at the primes dropped by the size datum. The entry S_u is not a lattice invariant: in use, an identity between the slope datum and an arithmetic datum, the identity (Id) of the excess bound below, sets it equal to a sum of inert prime sums. The loss condition (S3) below, $S_u - \Theta_u \leq \mathrm{Lo}$, is the place where primes and lattices meet. Put $n_{\mathcal{T}} = |\mathcal{T}|$, $\log\Omega = \gamma_0 + \log(\pi/v) - \beta$,

$$s_u = \frac{\sum_\omega r_{u,\omega}(\mu_{u,\omega} - h_{u,\omega})}{\sum_\omega r_{u,\omega}}, \qquad \mathrm{Main}(u) = \sum_{j=1}^{m}\sum_{\tau\in\mathcal{T}} a_\tau\Big(\log a_\tau - 1 - \gamma_0 + \beta - \frac{1}{f}\sum_{\rho=1}^{f}\log w_{u,j,\rho}\Big).$$

For a constant vector $c = (c_{\mathrm{Div}}, e_{\mathrm{Div}}, c_E, c_H, c_{\mathrm{Count}}, c_{\mathrm{Slope}}, c_{\mathrm{Sl}}, c_{\mathrm{Loss}}, e_{\mathrm{Loss}}, c_{\mathrm{HH}})$, the datum satisfies the **slope conditions** if, besides positivity conditions,

(S1) Divisor bound. For all u, ω: $\mu^\vee_u + h_{u,\omega} - F_u \leq \sum_\tau\big(m\,\varphi(a_\tau) - a_\tau\log\Omega - a_\tau\log T_{u,\omega}\big) + c_{\mathrm{Div}} n_{\mathcal{T}} a L_a + c_{\mathrm{Div}} m(B_0+N_c)x^{e_{\mathrm{Div}}\varepsilon}$, with a fixed function φ.

(S2) Transfer of slopes (a slope inequality in the sense of [19]). $\mu^{\max}_{u,\omega} \le \mu'^{\max}_{u,\omega} + \Theta_u + F_u - E_{u,\omega} + c_{\mathrm{Sl}} m B_0$ and $\mu'_{u,\omega} + \Theta_u + F_u - E_{u,\omega} - \mathrm{Lo}_{u,\omega} \le \mu_{u,\omega}$, where $E_{u,\omega} = \sum_j E_{u,\omega,j}$ and $\mathrm{Lo}_{u,\omega}$ is an explicit loss term, which depends on an auxiliary scale X''.

(S3) Losses. For every $u \in U$, $S_u - \Theta_u \le \mathrm{Lo}$ and $I_u \le \mathrm{Lo}$, where $\mathrm{Lo} = c_{\mathrm{Loss}} n_{\mathcal{T}} m b x^{e_{\mathrm{Loss}}\varepsilon}$ is the **loss bound** of the datum.

(S4) Dual-slope bound. For each u: $0 \le \delta_u < 1$, $\eta_u \ge 0$, $0 \le e_u \le \frac{1}{4}$, $0 < \mathrm{HH}_u$, $s_u - \mu^{\vee}_{\min,u} \le \mathrm{HH}_u$, and

$$s_u - \log|\mathcal{W}(u)| - (1-\delta_u)^{-1}\Big(4\phi_u \log\big(1 + \mathrm{HH}_u/\phi_u\big) + (\eta_u + 4\sqrt{e_u})\,\mathrm{HH}_u\Big) \le \mu^{\vee}_u,$$

where $\phi_u = \max_\omega(\mu^{\max}_{u,\omega} - h_{u,\omega}) - s_u$.

(S5) Comparisons. Bounds comparing $E_{u,\omega,j}$ with $l_g \sum_\tau a_\tau$, comparing heights across ω, bounding $\log T_{u,\omega}$ above and below (counting), bounding the slopes μ' by $c_{\mathrm{Slope}} n_{\mathcal{T}} a L_a$, $\mathrm{HH}_u \le x^{c_{\mathrm{HH}}} b$, and $\delta_u \le \bar\delta$, $\eta_u \le \bar\eta$, $e_u \le \bar e$ for fixed $\bar\delta < 1$, $\bar\eta \ge 0$, $\bar e \le \frac{1}{4}$.

We also use the **variance term** $V = 2mc_E B_0 + c_H\big(n_{\mathcal{T}} a L_a + m B_0\big) + c_{\mathrm{Count}} \log k \sum_\tau a_\tau$. The function φ, the loss term $\mathrm{Lo}_{u,\omega}$ of (S2), the positivity conditions and the exact form of the comparisons (S5) are explicit expressions in the parameters, which we do not reproduce.

Read in the right order, the slope conditions bound a prime-sum entry by its main term. The entry S_u is set equal to the inert prime sums of the host by the identity (Id) of Proposition 4.5 below; this is a definition. The substance is the loss condition (S3), $S_u - \Theta_u \le \mathrm{Lo}$: the prime contributions are dominated by the local gains of the lattices, up to the loss bound. The transfer of slopes (S2) then bounds the slopes $\mu_{u,\omega}$ from below by Θ_u and the slopes of a level-zero lattice; the dual-slope bound (S4) passes the weighted mean of these slopes to the dual slope $\mu^{\vee}_u$; and the divisor bound (S1) bounds $\mu^{\vee}_u$ from above by the main term. The result is $S_u \le \mathrm{Main}(u) + \mathrm{Err}(u)$ (Lemma 4.7). Besides (S3), primes and lattices meet in the main terms: under (Id), the main term $\mathrm{Main}(u)$ of the divisor bound is the sum of the main terms of the inert prime sums. Why the loss condition holds for the lattices of theta lifts, and what the local gains are in classical terms, is not clear to us (Section 11).

4.3 Size and error conditions

The remaining hypotheses are size and error conditions on the parameters, collected in the next two definitions in the form the argument uses. Growth-inequality data also describe the setting of the analytic half. Their weights are close to b at the places that are thick at a translate u, so that the cut-offs a_t/w lie within $O(1)$ of x, and at most $2b'$ at the other places of Φ, so that the cut-offs lie beyond $X'/2$, where the thin-scale condition (G4) makes every twisted Chebyshev function negligible. Condition (G5) asks for square-root bounds at x for the three quadratic characters of Γ_H that are trivial on H; for the hosts of the argument these are the quadratic characters attached to q, d and qd. Conditions (G6) and (G7) are the cost and size conditions to which the statements below refer.

Definition 4.3. Let x, ε, C be real. Let χ'_C be the character $((a_1,a_2),h) \mapsto (-1)^{a_2}$ of Γ_H. For $T_2 \subseteq H$ and $u \in H$, an element $\rho \in \Gamma_H$ is **thick at** u if $\rho = ((0,1), t+u)$ for some $t \in T_2$.

Growth-inequality data at scale x with constant C and parameter ε consist of a finite abelian group H, an arithmetic datum $\mathfrak{A}$ of host type on H, a function η on the characters of Γ_H, integers $P, m \ge 1$, finite sets $T, T_2 \subseteq H$, numbers a_t $(t \in T)$, weights $w_{u,j}\colon \Gamma_H \to \mathbb{R}$ $(u \in H,\ 1 \le j \le m)$, and reals $b, b', \nu, X', \theta, C_d, C_L, \mathrm{Err}, k, a_E$, such that

(G1) $x \ge 3$, $b > 0$, $\nu \ge 0$, $X' > 0$, $b' > 0$, $b'X' \le xb$, $\theta \ge 0$, $C_d \ge 0$, and $|a_t/b - x| \le 1$ for $t \in T$;

(G2) for all u, j: $w_{u,j}(\rho) > 0$ for $\rho \in \Phi$; $b \le w_{u,j}(\rho) \le b + \nu + 1$ if ρ is thick at u; $w_{u,j}(\rho) \le 2b'$ if $\rho \in \Phi$ is not thick at u; and $a_t/w_{u,j}(\rho) \le P$ for $t \in T$, $\rho \in \Phi$;

(G3) $\overline{\Delta} \le \mathrm{Err}$, where $\overline{\Delta}$ is formed with $(T, (a_t), (w_{u,j}), \eta, P)$;

(G4) (thin scales) $|G_{\mathfrak{A}}(y, \phi)| \le \theta y$ for every character ϕ and every $y \ge X'/3$;

(G5) (values at x) $|G_{\mathfrak{A}}(x, \phi)| \le x^{1/2+\varepsilon}$ and $|\eta(\phi)| \le C_L \log x$ for $\phi \in \{\chi_C', \chi_C, \chi_C\chi_C'\}$;

(G6) (small costs) $C_d + 3\big(\log(x+1) + C_d + C_L \log x\big) \le x^{1/2+\varepsilon}$, $2|\Phi|(x+1)\theta \le 1/x$ and $6|\Phi|(x+1)(\nu+1) \le b/x$;

(G7) (size conditions) $\varepsilon > 0$, $C \ge 0$, $0 \le a_E \le 2xb$, $|H|x^{-\varepsilon/8} \le 2|T|$, $x^{-\varepsilon}|H| \le |T_2| \le |H|$, $x^{1/2-\varepsilon} \le m \le x^{1/2+\varepsilon/2}$, $xm \le k \le |H|^2 x^{1/2-\varepsilon}$, $x^{3\varepsilon/8} \ge 26$, and

$$\begin{aligned} &\mathrm{Err} \le x^{C\varepsilon} E(x), \quad \text{where} \\ &E(x) = |T|\, a_E\Big(1 + \frac{m^2}{k}\Big) + m\Big(2|H|\, m\, b\, x^{\varepsilon} + \frac{|T_2|\, k\, b\, x^{\varepsilon}}{|H|^2 x^{-\varepsilon}}\Big). \end{aligned} \tag{4.1}$$

They are **full growth-inequality data** if moreover

(G8) $\big|wG_{\mathfrak{A}}(Y/w, \phi) - w'G_{\mathfrak{A}}(Y/w', \phi)\big| \le 3Y|w - w'|/\min(w, w')$ for all ϕ and $Y, w, w' > 0$;

(G9) $|G_{\mathfrak{A}}(y, \phi) - G_{\mathfrak{A}}(z, \phi)| \le |y - z|\big(\log\max(y, z) + C_d + |\eta(\phi)|\big)$ for $\phi \ne 1$, and $|G(y) - G(z)| \le C_d|y - z|$, for all $y, z \ge 1$.

Zeta chain core data, in the terminology of the development, put a slope datum and an arithmetic datum on one group H at one scale. Condition (Z1) asks for the slope conditions with exponents $e_{\mathrm{Div}} \le e_D$ and $e_{\mathrm{Loss}} \le e_L$; (Z2) sets the prime-sum entries equal to inert prime sums; (Z3) collects the error conditions under which all error terms are $x^{O(\varepsilon)}$ times their natural sizes; and (Z4)–(Z7) are the size conditions of the analytic half. These are the data that a tied family provides at each large scale of its set of scales (Definition 6.13). The last paragraph of the definition, on zeta chain inputs, is not used in this paper.

Definition 4.4. Let $x, \varepsilon, e_D, e_L, C_K, c^*, e_K, C_L \in \mathbb{R}$. **Zeta chain core data at scale x with exponents (e_D, e_L) and constants (C_K, c^*, e_K, C_L)** consist of a finite abelian group H, a slope datum with index set $U = H$, parameters x, ε, constant vector c and an auxiliary real X'', an arithmetic datum $\mathfrak{A}$ of host type on H, a function η on the characters of Γ_H, an integer P, finite sets $T, T_2 \subseteq H$, numbers $\tilde{a}(t)$, weights $w_{u,j}\colon \Gamma_H \to \mathbb{R}$, bijections $\mathcal{T} \cong T$ and $\{1, \dots, f\} \cong \Phi$, and reals ν, X', θ, such that:

(Z1) $e_{\mathrm{Div}} \le e_D$, $e_{\mathrm{Loss}} \le e_L$, $c_{\mathrm{HH}} \le c^*$, and the slope datum satisfies the slope conditions for c and X'' (among them the loss conditions (S3), which bound $S_u - \Theta_u$ and I_u at every $u \in H$);

(Z2) (identification) $a_\tau = \tilde{a}(t)$ for $\tau \mapsto t$, $w_{u,j,\rho} = w_{u,j}(g)$ for $\rho \mapsto g$, $\gamma_0 = \gamma$, $\beta = \mathrm{Re}\,\eta(\chi_C)$ and $S_u = \sum_{t \in T}\sum_{j=1}^{m} A_{\mathrm{in}}(\tilde{a}(t), t, w_{u,j})$ for all $u \in H$;

(Z3) (error conditions) $1 \le x$, $0 < \varepsilon \le 1$, $1 \le b$, $xb \le a$, $1 \le n_{\mathcal{T}}$, $m \le x^{1/2+\varepsilon/2}$, $xm \le k$, $0 \le e_K$, $1 \le k \le \min(x^{e_K}, a)$, $0 \le L, L_a \le 2x^{\varepsilon}$, $0 \le b'$, $kb' \le a$, $0 \le X'' \le a$, $0 \le a_\tau \le 2a$ for all τ, $B_0, N_c \ge 0$, $\log|\mathcal{W}(u)| \le (k-3)\log 2$ for all u, $\bar{\delta} \le \frac{1}{2}$, $\bar{\eta}, \bar{e} \ge 0$; the constants $c_{\mathrm{Div}}, c_E, c_H, c_{\mathrm{Count}}, c_{\mathrm{Slope}}, c_{\mathrm{Sl}}, c_{\mathrm{Loss}}$ lie in $[0, C_K]$ and $e_{\mathrm{Div}}, e_{\mathrm{Loss}}, c_{\mathrm{HH}} \ge 0$; and

$$(\bar{\eta} + 4\sqrt{\bar{e}})\, x^{c_{\mathrm{HH}}}\, b \le C_K\, x^{\varepsilon}\, n_{\mathcal{T}}\, a;$$

(Z4) $B_0 \le 2|H|mbx^{\varepsilon}$ and $N_c \le |T_2|kbx^{\varepsilon}/(|H|^2 x^{-\varepsilon})$;

(Z5) conditions (G1), (G2), (G4) and (G5) hold for $(H, \mathfrak{A}, \eta, P, m, T, T_2, \tilde{a}, w)$ with the parameters b, b' of the slope datum, ν, X', θ and C_L;

(Z6) (cost conditions) $2|\Phi|(x+1)\theta \le 1/x$ and $6|\Phi|(x+1)(\nu+1) \le b/x$;

(Z7) (remainder size conditions) $a \le 2xb$, $|H|x^{-\varepsilon/8} \le 2|T|$, $x^{-\varepsilon}|H| \le |T_2|$, $x^{1/2-\varepsilon} \le m$ and $k \le |H|^2 x^{1/2-\varepsilon}$.

Zeta chain inputs at scale x with constant C_d and exponents (e_D, e_L) are zeta chain core data at scale x with exponents (e_D, e_L) and some constants (C_K, c^*, e_K, C_L) such that

$$x \ge 3, \qquad x^{3\varepsilon/8} \ge 26, \qquad \log x \le x^{\varepsilon},$$
$$200(1+C_K)^2(1+c^*)(1+e_K) \le x^{\varepsilon},$$
$$C_d + 3\big(\log(x+1) + C_d + C_L \log x\big) \le x^{1/2+\varepsilon}.$$

The slope conditions in (Z1) are those of Definition 4.2, which comprise further size and error conditions of the same nature that we do not reproduce.

4.4 The excess bound

The excess bound is the geometric half of the bridge. It is stated for an index set U with a map tr to H, because the lattice data come from several members of a family; condition (Id) sets the prime-sum entries equal to inert prime sums and matches the parameters of the two data.

Proposition 4.5. *Let $\mathfrak{A}$ be an arithmetic datum on a finite abelian group H, let $P \ge 1$, $T, T_2 \subseteq H$ finite, $\tilde{a}\colon H \to \mathbb{R}$, weights $w_{h,j}\colon \Gamma_H \to \mathbb{R}$ ($h \in H$, $1 \le j \le m$) and η a function on the characters of Γ_H. Let a slope datum with index set U be given, with a map $\mathrm{tr}\colon U \to H$ and bijections $\mathcal{T} \cong T$, $\{1, \dots, f\} \cong \Phi$, such that*

(Id) *$a_\tau = \tilde{a}(t)$ for $\tau \mapsto t$; $w_{u,j,\rho} = w_{\mathrm{tr}(u),j}(g)$ for $\rho \mapsto g$; $\gamma_0 = \gamma$; $\beta = \mathrm{Re}\,\eta(\chi_C)$; and $S_u = \sum_{t \in T} \sum_{j=1}^{m} A_{\mathrm{in}}(\tilde{a}(t), t, w_{\mathrm{tr}(u),j})$ for every $u \in U$.*

Suppose the slope datum satisfies the slope conditions for a constant vector c, the error conditions (Z3) of Definition 4.4 with constants (C_K, e_K), and $B_0 \le 2|H|mbx^{\varepsilon}$, $N_c \le x^{2\varepsilon}|T_2|kb/|H|^2$, and that x is large in terms of $(\varepsilon, C_K, c_{\mathrm{HH}}, e_K)$. Then for every $u \in U$, with a, b, k, m the parameters of the slope datum and Δ formed with $(T, \tilde{a}, w, \eta, P)$,

$$\Delta(\mathrm{tr}(u)) \;\le\; x^{(e_{\mathrm{Div}}+e_{\mathrm{Loss}}+3)\varepsilon}\Big(|T|\,a\,\big(1+m^2/k\big) + m\big(2|H|mbx^{\varepsilon} + x^{2\varepsilon}|T_2|kb/|H|^2\big)\Big).$$

Proof. In outline. By (Id), $\Delta(\mathrm{tr}(u)) = S_u - \mathrm{Main}(u)$: the inert prime sums are S_u, and the main terms $M_\eta(\tilde{a}(t), w_{\mathrm{tr}(u),j})$ summed over t and j coincide with $\mathrm{Main}(u)$, the logarithmic averages of the weights over Φ matching under $\{1,\dots,f\} \cong \Phi$. This is the identity that ties the inert prime sums of the arithmetic datum to the slope datum. The slope conditions bound $S_u - \mathrm{Main}(u)$ by an explicit error term: the transfer of slopes (S2) bounds s_u from below by a term containing Θ_u, the dual-slope bound (S4) passes this to the dual slope, the divisor bound (S1) bounds the dual slope from above by the main term, and the losses (S3), which hold at every translate u, replace Θ_u by S_u; this is $S_u \le \mathrm{Main}(u) + \mathrm{Err}(u)$ (Lemma 4.7). The bound is thus obtained translate by translate, which is the form used in the main argument. The error conditions bound $\mathrm{Err}(u)$ by $x^{(e_{\mathrm{Div}}+e_{\mathrm{Loss}}+3)\varepsilon}\big(n_{\mathcal{T}}a(1+m^2/k) + m(B_0+N_c)\big)$ (Lemma 4.6), and the bounds on B_0 and N_c give the displayed bracket. □

Its proof combines two lemmas. The first defines the explicit error expression $\mathrm{Err}(u)$ and bounds it under the error conditions (Z3); it is bookkeeping. The second is where the slope conditions are used.

Lemma 4.6. *For a slope datum with constant vector c and an auxiliary real X'', and $u \in U$, let $\mathrm{Err}(u)$ denote the per-translate error expression*

$$\begin{aligned}\mathrm{Err}(u) = {} & \Big(\sum_\tau a_\tau\Big) c_{\mathrm{Count}}\Big(L_a + \frac{m^2 L_a}{k}\Big) + c_{\mathrm{Div}} n_{\mathcal{T}} a L_a \\ & + c_{\mathrm{Div}} m(B_0+N_c) x^{e_{\mathrm{Div}}\varepsilon} + m c_E B_0 + c_{\mathrm{Slope}} n_{\mathcal{T}} a L_a\Big(1+\frac{m}{k}\Big) \\ & + \log|\mathcal{W}(u)| + \mathrm{Lo} + M_{\mathrm{Lo}} + (1-\bar{\delta})^{-1}\Big(4\log(1+x^{c_{\mathrm{HH}}})(\bar{U}+b) + (\bar{\eta}+4\sqrt{\bar{e}})\, x^{c_{\mathrm{HH}}} b\Big),\end{aligned}$$

where $\mathrm{Lo} = c_{\mathrm{Loss}}\, n_{\mathcal{T}}\, m\, b\, x^{e_{\mathrm{Loss}}\varepsilon}$ is the loss bound, $M_{\mathrm{Lo}} = c_{\mathrm{Sl}}\mathrm{Lo} + c_{\mathrm{Sl}} n_{\mathcal{T}} m(a/k + b') c_{\mathrm{Count}} mL + c_{\mathrm{Sl}} X'' + 1$ and $\bar{U} = c_{\mathrm{Slope}} n_{\mathcal{T}} a L_a(2 + m/k) + V + c_{\mathrm{Sl}} m B_0 + M_{\mathrm{Lo}}$ with the variance term

$$V = 2mc_E B_0 + c_H\big(n_{\mathcal{T}} a L_a + m B_0\big) + \Big(\sum_\tau a_\tau\Big) c_{\mathrm{Count}} \log k.$$

If the datum satisfies the error conditions (Z3) with constants (e_K, C_K), $\log x \le x^\varepsilon$ and $200(1+C_K)^2(1+c_{\mathrm{HH}})(1+e_K) \le x^\varepsilon$, then for every $u \in U$

$$\mathrm{Err}(u) \le x^{(e_{\mathrm{Div}}+e_{\mathrm{Loss}}+3)\varepsilon}\Big(n_{\mathcal{T}}\, a\Big(1+\frac{m^2}{k}\Big) + m(B_0+N_c)\Big).$$

Proof. In outline. Each term of $\mathrm{Err}(u)$ is bounded separately, using the error conditions (Z3) and the large-x conditions. The condition $0 \le a_\tau \le 2a$ gives $\sum_\tau a_\tau \le 2n_{\mathcal{T}}a$. The bounds $L, L_a \le 2x^\varepsilon$, $kb' \le a$, $X'' \le a$, $1 \le k \le x^{e_K}$ and $\log|\mathcal{W}(u)| \le (k-3)\log 2$ bound the counting, divisor, slope and loss terms by $x^{O(\varepsilon)} n_{\mathcal{T}} a$ or $x^{O(\varepsilon)} m(B_0+N_c)$. The constants of c lie in $[0, C_K]$, so $200(1+C_K)^2(1+c_{\mathrm{HH}})(1+e_K) \le x^\varepsilon$ absorbs them. The factor $(1-\bar{\delta})^{-1} \le 2$ follows from $\bar{\delta} \le \frac{1}{2}$. The strong condition $(\bar{\eta}+4\sqrt{\bar{e}})x^{c_{\mathrm{HH}}} b \le C_K x^\varepsilon n_{\mathcal{T}} a$ controls the last term. This condition cannot be weakened to $\bar{\eta}, \bar{e} \le 1$: there is an explicit datum satisfying the weakened conditions for which the bound fails. □

Lemma 4.7. *Let a slope datum with constant vector c and an auxiliary real X'' satisfy the slope conditions (S1)–(S5) of Definition 4.2; in particular, by the loss conditions (S3), $S_u - \Theta_u$ and I_u are at most the loss bound $c_{\mathrm{Loss}} n_{\mathcal{T}} m b x^{e_{\mathrm{Loss}}\varepsilon}$ for every $u \in U$. Then, for every $u \in U$,*

$$S_u \le \mathrm{Main}(u) + \mathrm{Err}(u),$$

with $\mathrm{Err}(u)$ the per-translate error expression of Lemma 4.6.

Proof. In outline. For each u, the divisor bound (S1) and the slope comparisons (S5), averaged over the words $\omega \in \mathcal{W}(u)$, bound the weighted mean of the heights by $\mathrm{Main}(u) + \log|\mathcal{W}(u)|$ plus the divisor and counting terms of $\mathrm{Err}(u)$. The transfer of slopes (S2) and the dual-slope bound (S4) give a weighted-mean lower bound, which controls the gap between the slope sum and $\mu_u^\vee$ by the term $(1-\bar\delta)^{-1}(4\log(1+x^{c_{\mathrm{HH}}})(\bar U + b) + (\bar\eta + 4\sqrt{\bar e})x^{c_{\mathrm{HH}}} b)$. The loss conditions (S3) turn $S_u - \Theta_u$ and I_u into the loss terms Lo and M_{Lo}. Summing these inequalities gives the claim. □

4.5 From the excess to the one-sided bound

The analytic half turns a bound for the excess at the translates $u \in A_0$ into an upper bound for a weighted sum of twisted Chebyshev functions at x. The characters in K' are trivial on A_0 and on the element $((1,0),0)$, which for a host is the element $(1,-1,0)$ of $\mathcal{G}$; so for a host they lie in $X^+(\mathfrak{H})$. Their weights, 1 for those trivial on H and c for the others, come from the distribution of T among the cosets of A_0, as explained before Definition 3.2. The characters outside K' stay on the right-hand side, and for a host condition (E4) bounds them.

Proposition 4.8. *Let $C \ge 0$, $\varepsilon > 0$, $C_E, K_o, M \in \mathbb{R}$ and $c \in \mathbb{R}$ with $|c| \le 1$. There is x_1 such that the following holds for all $x \ge x_1$. Let H be a finite abelian group with a subgroup A_0, $\mathfrak{A}$ an arithmetic datum of host type on H, η a function on the characters of Γ_H, and $P, m \ge 1$, $T, T_2 \subseteq H$, $\tilde a$, w as in Definition 4.1. Let $X(A_0)$ be the set of characters of Γ_H trivial on $\{0\} \times A_0$, and let $K' \subseteq X(A_0)$ contain the trivial character, with weight $w_{K'}(\phi) = 1$ if ϕ is trivial on $\{0\} \times H$ and $w_{K'}(\phi) = c$ otherwise. Suppose that:*

- **(C1)** *$m \le x^{1/2+\varepsilon/2}$, $xm \le k$, and the size and thin-scale conditions (G1), (G2), (G4) of Definition 4.3 and the cost and remainder-size conditions (Z6), (Z7) of Definition 4.4 hold at x for parameters $b > 0$, b', k, $a_E \ge 0$ and $T_2 \subseteq A_0$;*
- **(C2)** *$|X(A_0)| \le M$; every $\phi \in K'$ is trivial on $((1,0),0)$; $\bigl|\sum_{t\in T}\phi(0,t) - c|T|\bigr| \le M x^{-1+\varepsilon}|T|$ for $\phi \in K'$ non-trivial on $\{0\} \times H$; $|\eta(\phi)| \le K_o$ for $1 \ne \phi \in K'$; and $|\eta(\phi)| \le C_E \log x$ for $\phi \in X(A_0) \setminus K'$;*
- **(C3)** *$\Delta(u) \le x^{C\varepsilon} E(x)$ for every $u \in A_0$, where $E(x)$ is the bracket of Proposition 4.5 with $(a,b,k) = (a_E, b, k)$.*

Then

$$\sum_{\phi\in K'} w_{K'}(\phi)\,\mathrm{Re}\,G_{\mathfrak{A}}(x,\phi) \;\le\; x^{1/2+(C+3)\varepsilon} + \sum_{\phi\in X(A_0)\setminus K'} \bigl|G_{\mathfrak{A}}(x,\phi)\bigr|.$$

Proof. In outline. The average over $u \in A_0$, $t \in T$ and the weight index $j \in \{1,\dots,m\}$ of the inert prime sums $A_{\mathrm{in}}(a_t, t, w_{u,j})$, expanded by orthogonality of the characters of Γ_H, is a weighted combination of the twisted Riesz-weighted Chebyshev functions $G_{\mathfrak{A}}(y,\phi)$ at the scales $y = a_t/w$. The weights concentrate on K' by (C2), and the scales cluster near x by the size conditions. The excess bound (C3) is therefore an upper bound for the weighted combination of the $\mathrm{Re}\,G_{\mathfrak{A}}(y,\phi)$, $\phi \in K'$, at scales near x, up to the characters outside K'. The derivative bounds for $G_{\mathfrak{A}}$ (Lipschitz in y, with the log-terms controlled by (C2)) move these scales to x at a cost absorbed by the cost conditions. The moving of scales uses the Lipschitz estimates for these functions (Lemma 4.10). The final step, from the averaged inequality to the power bound, is the size-condition lemma Proposition 4.9 with Υ the weighted sum. □

Two auxiliary results enter its proof: an elementary absorption of error terms under the size conditions, and Lipschitz estimates for the twisted Chebyshev functions, which move the cut-offs a_t/w to x.

Proposition 4.9. *Let $C, \varepsilon, x, b, k, a_E, \mathrm{Err}, \Upsilon \in \mathbb{R}$ and let n, n_T, n_2, m be non-negative integers. Assume the* ***size conditions***

$$x \geq 3, \quad \varepsilon > 0, \quad C \geq 0, \quad n > 0, \quad b > 0, \quad 0 \leq a_E \leq 2xb, \quad x^{3\varepsilon/8} \geq 26,$$
$$\tfrac{1}{2} x^{-\varepsilon/8} n \leq n_T, \quad x^{-\varepsilon} n \leq n_2 \leq n,$$
$$x^{1/2-\varepsilon} \leq m \leq x^{1/2+\varepsilon/2}, \quad xm \leq k \leq n^2 x^{1/2-\varepsilon},$$
$$\mathrm{Err} \leq x^{C\varepsilon}\Big(n_T\, a_E\Big(1 + \frac{m^2}{k}\Big) + m\Big(2n\, m\, b\, x^{\varepsilon} + \frac{n_2\, k\, b\, x^{\varepsilon}}{n^2 x^{-\varepsilon}}\Big)\Big).$$

If

$$n_T\, m\, b\, \frac{n_2}{2n}\Big(\Upsilon - 4x^{1/2+\varepsilon}\Big) \leq \mathrm{Err} + \frac{2 n_T m b}{x},$$

then $\Upsilon \leq x^{1/2+(C+3)\varepsilon}$.

Proof. In outline. Divide the hypothesis by $n_T m b\, n_2/(2n) \geq \tfrac{1}{2} n_T m b\, x^{-\varepsilon}$. The right-hand side becomes at most $2x^{\varepsilon}(\mathrm{Err}/(n_T m b) + 2/x)$. Each term of the bracket bounding Err, divided by $n_T m b$, is bounded by a power $x^{1/2+O(\varepsilon)}$ using the size conditions: $a_E \leq 2xb$ with $m \geq x^{1/2-\varepsilon}$ and $m^2/k \leq m/x$; $2nmbx^{\varepsilon}/(n_T b) \leq 4x^{9\varepsilon/8} m$ with $m \leq x^{1/2+\varepsilon/2}$; and $n_2 k b x^{\varepsilon}/(n^2 x^{-\varepsilon} n_T b)$ with $k \leq n^2 x^{1/2-\varepsilon}$. The sum, together with $4x^{1/2+\varepsilon}$, is at most $x^{1/2+(C+3)\varepsilon}$ once $x^{3\varepsilon/8} \geq 26$. □

Lemma 4.10. *1. For every finite abelian group H, every arithmetic datum $\mathfrak{A}$ on H, every function η, every character ϕ of Γ_H and all $Y, w, w' > 0$,*

$$\Big| w\, G_{\mathfrak{A}}(Y/w, \phi) - w'\, G_{\mathfrak{A}}(Y/w', \phi) \Big| \leq \frac{3Y\, |w - w'|}{\min(w, w')}.$$

2. There is $C_d \geq 0$ such that for every finite abelian group H, every arithmetic datum $\mathfrak{A}$ on H, every η, every non-trivial character ϕ and all $y, z \geq 1$,

$$|G_{\mathfrak{A}}(y, \phi) - G_{\mathfrak{A}}(z, \phi)| \leq |y - z|\big(\log \max(y, z) + C_d + |\eta(\phi)|\big), \qquad |G(y) - G(z)| \leq C_d |y - z|.$$

Proof. In outline. For $\phi \neq 1$ the sum $G_{\mathfrak{A}}(y, \phi)$ is $\sum_{n \leq y} \Lambda(n) \phi_{\mathfrak{A}}(n)(y/n - 1) + y\eta(\phi)$ with $|\phi_{\mathfrak{A}}(n)| \leq 1$. Its derivative in y is bounded by $\sum_{n \leq y} \Lambda(n)/n + |\eta(\phi)| \leq \log y + O(1) + |\eta(\phi)|$, which gives the first bound of item 2. The bound for G comes from a Lipschitz estimate for the error term of the zeta prime sum $\sum_{n \leq y} \Lambda(n)/n - \log y$. Item 1 follows from the same kind of estimate applied to $w \mapsto w G_{\mathfrak{A}}(Y/w, \phi)$, using $\sum_{n \leq y} \Lambda(n) \leq 3y$. □

4.6 The bound at one host

We can now assemble the bound at one host. The proof applies the excess bound to the slope datum of the core-tied pair, and the analytic half to the characters θ of the host with χ_θ^* in the level, with $c = (3o + 1)^{-1}$; through the Artin map their weighted sum is $G_{\mathfrak{L}}(x)$, and condition (E4) bounds the remaining characters.

Proposition 4.11. *Let $\mathfrak{L}$ be a character level, let $e_D, e_L, C_K, c^*, e_K, C$ be real numbers and let $\varepsilon > 0$. There is x_1 such that for every $x \geq x_1$, every frame of the index of $\mathfrak{L}$, every host $\mathfrak{H}$ for it and every finite subset T of A the following holds: if $(\mathfrak{H}, T)$ satisfies the host conditions for $\mathfrak{L}$ at scale x with parameters (ε, C) and is core-tied at scale x with exponents $(e_D, e_L, C_K, c^*, e_K)$ and parameter ε, then*

$$G_{\mathfrak{L}}(x) \leq x^{1/2+(e_D+e_L+7)\varepsilon}.$$

Proof. In outline. Let $(\mathfrak{H}, T)$ be as in the statement, with the slope datum, tr, P, T_2, $\tilde{a}$, w, η of the core-tied condition, and let $\mathfrak{A}_{\mathfrak{H}}$ be the arithmetic datum of $\mathfrak{H}$.

1. Excess bound. By (K2)–(K4) and Proposition 4.5, applied for x beyond a threshold depending on $(\varepsilon, C_K, c^*, e_K)$, every $u \in A_0$, which lies in the image of tr, satisfies $\Delta(u) \leq x^{(e_{\mathrm{Div}}+e_{\mathrm{Loss}}+3)\varepsilon} E(x) \leq x^{(e_D+e_L+3)\varepsilon} E(x)$.

2. Chebyshev bound. Take K' to be the characters θ of $\mathcal{G}$, written on $\{\pm 1\}^2 \times A$, with $\chi_\theta^* \in K$, $c = (3o+1)^{-1}$ and $M = 4o$. Condition (C2) of Proposition 4.8 follows from (E1)–(E3) and (K1), and (C3) is step 1 with $C = \max(e_D + e_L, 0) + 3$; the core-tied condition gives $e_D + e_L \geq 0$, so $C = e_D + e_L + 3$. Hence the weighted sum over K' is at most $x^{1/2+(C+3)\varepsilon}$ plus the sum of $|G_{\mathfrak{A}_{\mathfrak{H}}}(x, \phi)|$ over the characters outside K'.

3. Identification. Via the Artin map α of the host, $G_{\mathfrak{A}_{\mathfrak{H}}}(y, \phi) = g(y, \chi_\theta^*)$ for ϕ corresponding to θ, and the weighted sum over K' is $G_{\mathfrak{L}}(x)$. The characters outside K' are controlled characters, so by (E4) their sum is at most $4o\, x^{1/2+\varepsilon}$. For $x \geq x_1$ one has $x^\varepsilon \geq 1 + 4o$, which absorbs both terms into $x^{1/2+(e_D+e_L+7)\varepsilon}$.

□

This is the first of the five results on which Section 2 rests. The rest of the paper produces, at all scales of a set containing windows, hosts that carry such lattice data. Section 5 fixes the notation this requires, and Sections 6–8 construct the data.

5 Notation for the arithmetic part

From here on the argument uses the standard notation of algebraic number theory, automorphic forms, Galois cohomology, p-adic Hodge theory and Arakelov geometry, which this section fixes with references; the reader can skip it and return to it as needed. One symbol changes its meaning. In Section 1, $L(s, \chi)$ for a Dirichlet character χ modulo N is the series $\sum_n \chi(n) n^{-s}$, which for imprimitive χ lacks the Euler factors at the primes dividing N; from here on, read as a Hecke L-function, it is the L-function of the primitive character inducing χ. The two differ by finitely many Euler factors, which do not vanish in $\mathrm{Re}\, s > 0$.

Notation for number fields, local fields and adeles is that of [96] and [25]. For a number field F with ring of integers $\mathcal{O}_F$, places are written v, finite places being identified with the nonzero primes of $\mathcal{O}_F$; F_v is the completion, $|\cdot|_v$ its normalised absolute value, and for finite v, ϖ_v is a uniformiser of $\mathcal{O}_{F_v}$ and $q_v = \mathrm{N}v$ the cardinality of the residue field. For a finite extension E/F, $w \mid v$ means that the place w of E lies above v, and $\mathrm{N}_{E/F}$ is the norm (on elements, ideles and idele classes). A **non-archimedean local field** is a field complete for a discrete valuation with finite residue field. The adele ring is $\mathbb{A}_F = F_\infty \times \mathbb{A}_{F,f}$, with $F_\infty = F \otimes_{\mathbb{Q}} \mathbb{R}$ and $\mathbb{A}_{F,f}$ the finite adeles; $|x|_{\mathbb{A}} = \prod_v |x_v|_v$ on the ideles $\mathbb{A}_F^\times$; $C_F = \mathbb{A}_F^\times / F^\times$ is the idele class group, C_F^1 its subgroup

of norm one and C_F^0 the connected component of the identity. $G_F = \mathrm{Gal}(\overline{F}/F)$; $G_{F_v} \subset G_F$ is a decomposition group, I_v its inertia subgroup and Frob_v an arithmetic Frobenius. For a finite set S of places, $G_{F,S}$ is the Galois group of the maximal extension of F unramified outside S.

Class field theory is normalised as in [96] and [6]: uniformisers correspond to arithmetic Frobenius elements. For a non-archimedean local field F_v, $\mathrm{Art}_{F_v}\colon F_v^\times \to G_{F_v}^{ab}$ is the local Artin map and F_v^{ur} the maximal unramified extension; for a number field F, $\mathrm{Art}_F\colon C_F \to G_F^{ab}$ is the global Artin map and F^{ab} the maximal abelian extension. The **norm group** of a finite abelian extension E/F (respectively E_w/F_v) is $\mathrm{N}_{E/F}C_E \subseteq C_F$ (respectively $\mathrm{N}_{E_w/F_v}E_w^\times \subseteq F_v^\times$).

Notation follows Tate's thesis [124] and [96], Ch. VII. A **Hecke character** of a number field F is a continuous character $\chi = \prod_v \chi_v\colon C_F \to \mathbb{C}^\times$; it is unitary unless said otherwise, and every unitary Hecke character can be written $\chi = \chi_0|\cdot|_{\mathbb{A}}^{i\tau}$ with a unique $\tau \in \mathbb{R}$ and χ_0 trivial on a fixed subgroup of C_F mapped isomorphically onto $\mathbb{R}_{>0}$ by $|\cdot|_{\mathbb{A}}$. Its conductor is $\mathfrak{f}(\chi)$. $\mathcal{S}(\mathbb{A}_F)$ is the Schwartz–Bruhat space, $\psi_{\mathrm{add}} = \prod_v \psi_{\mathrm{add},v}$ the standard additive character of $\mathbb{A}_F/F$ (trivial on F), and $\hat{f}$ the Fourier transform with respect to the self-dual measure. The **Tate zeta integral** is $Z(f,\chi,s) = \int_{\mathbb{A}_F^\times} f(x)\chi(x)|x|_{\mathbb{A}}^s\, d^\times x$ for $\mathrm{Re}\, s > 1$ and its meromorphic continuation, satisfying $Z(f,\chi,s) = Z(\hat{f},\chi^{-1},1-s)$. The **Hecke L-function** is $L(s,\chi) = \prod_{v\nmid\infty,\ \chi_v \text{ unramified}}(1-\chi_v(\varpi_v)q_v^{-s})^{-1}$ for $\mathrm{Re}\, s > 1$; for a Dirichlet character, viewed as a Hecke character of $\mathbb{Q}$ through class field theory, this is the Dirichlet L-function of the inducing primitive character. The **completed L-function** is $\xi(s,\chi) = (|d_F|\,\mathrm{N}\mathfrak{f}(\chi))^{s/2}L_\infty(s,\chi)L(s,\chi)$ with the archimedean gamma factors L_∞, and the **root number** $W(\chi)$, $|W(\chi)| = 1$, is defined by $\xi(s,\chi) = W(\chi)\,\xi(1-s,\overline{\chi})$. The Dirichlet density of a set of primes is in the sense of [96], VII.13.

Galois cohomology is continuous cohomology of profinite groups, as in [109] and [84]. We write $H^i(F,M)$ for $H^i(G_F,M)$ and $H^i(F_v,M)$ for $H^i(G_{F_v},M)$, for a discrete (or, for $\mathbb{Z}_\ell$-coefficients, continuous ℓ-adic) module M, and $\hat{H}^i$ is Tate cohomology of a finite group. For a prime ℓ, χ_{cyc} is the ℓ-adic cyclotomic character, $\mathbb{Z}_\ell(k)$ and $\mathbb{Q}_\ell/\mathbb{Z}_\ell(k)$ are the twists by χ_{cyc}^k, μ_{ℓ^r} the ℓ^r-th roots of unity, and for a finite $\mathbb{Z}/\ell^r\mathbb{Z}$-module M, $M^\vee = \mathrm{Hom}(M,\mu_{\ell^r})$ is the Cartier dual. The unramified classes are

$$H^1_{\mathrm{ur}}(F_v,M) = \ker\bigl(H^1(F_v,M) \to H^1(I_v,M)\bigr),$$

and inv_v is the invariant map of the local Brauer group. For a finite set S of places containing those above ℓ and ∞,

$$\mathrm{Ш}^i_S(F,M) = \ker\Bigl(H^i(G_{F,S},M) \to \prod_{v\in S} H^i(F_v,M)\Bigr).$$

Local Tate duality, the local and global Euler characteristic formulas and Poitou–Tate duality are used in the form of [84], Ch. I, and the Greenberg–Wiles formula for Selmer groups in the form of [30].

Notation for p-adic Hodge theory is that of [40] and [21]. Let K be a finite extension of $\mathbb{Q}_p$, G_K its absolute Galois group and $\mathbb{C}_p$ the completion of $\overline{\mathbb{Q}}_p$; $L^\flat$ is the tilt of a perfectoid field L. Fix a compatible system $\underline{\zeta} = (\zeta_{p^n})_{n\geq 0}$ of p-power roots of unity. The period rings are $A_{\mathrm{inf}} = W(\mathcal{O}_{\mathbb{C}_p^\flat})$ with $\theta\colon A_{\mathrm{inf}} \to \mathcal{O}_{\mathbb{C}_p}$; B_{dR}^+, the $\ker\theta$-adic completion of $A_{\mathrm{inf}}[1/p]$, and $B_{\mathrm{dR}} = B_{\mathrm{dR}}^+[1/t]$ with $t = \log[\underline{\zeta}]$ and $\mathrm{Fil}^i B_{\mathrm{dR}} = t^i B_{\mathrm{dR}}^+$; $A_{\mathrm{cris}} \subset B_{\mathrm{cris}}^+$ and $B_{\mathrm{cris}} = B_{\mathrm{cris}}^+[1/t]$ with Frobenius φ; and B_{st}. A finite-dimensional $\mathbb{Q}_p$-representation V of G_K is **de Rham**, **semistable** or **crystalline** if $(B\otimes_{\mathbb{Q}_p}V)^{G_K}$ has dimension $\dim_{\mathbb{Q}_p}V$ over B^{G_K} for $B = B_{\mathrm{dR}}, B_{\mathrm{st}}, B_{\mathrm{cris}}$ respectively. Hodge–Tate weights are normalised so that the cyclotomic character χ_{cyc} has Hodge–Tate weight 1.

Notation for Arakelov geometry follows [44], [20] and [92]. An **arithmetic variety** is an integral scheme $\mathcal{X}$ flat and projective over $\mathrm{Spec}\,\mathbb{Z}$ (or over $\mathrm{Spec}\,\mathcal{O}_F$), of absolute dimension

$D+1$, with generic fibre X. A **Hermitian line bundle** $\overline{\mathcal{L}} = (\mathcal{L}, \|\cdot\|)$ is a line bundle on $\mathcal{X}$ with a continuous metric on $\mathcal{L}_{\mathbb{C}}$ invariant under complex conjugation; it is smooth, semipositive, nef or big in the usual sense, and $\overline{\mathcal{O}}(1)$ on a product of projective spaces carries the Fubini–Study metric unless said otherwise. A **Hermitian lattice** $\overline{M} = (M, \|\cdot\|)$ is a projective $\mathcal{O}_F$-module of finite rank with Hermitian norms at the infinite places, with Arakelov degree $\widehat{\deg}\,\overline{M}$, normalised slope $\widehat{\mu}(\overline{M}) = \widehat{\deg}\,\overline{M}/([F:\mathbb{Q}]\operatorname{rk} M)$ and successive minima. $\widehat{c}_1(\overline{\mathcal{L}})$ is the arithmetic first Chern class, $\widehat{\deg}\big(\widehat{c}_1(\overline{\mathcal{L}})^{D+1}\big)$ the arithmetic self-intersection, and for a cycle $\mathcal{Z}$ the **height** is $h_{\overline{\mathcal{L}}}(\mathcal{Z}) = \widehat{\deg}\big(\widehat{c}_1(\overline{\mathcal{L}})^{\dim\mathcal{Z}} \mid \mathcal{Z}\big)$; for a point the height is normalised by $[F:\mathbb{Q}]$ only where said. $\widehat{h}^0(\mathcal{X}, \overline{\mathcal{L}}) = \log\#\{s \in H^0(\mathcal{X}, \mathcal{L}) : \|s\|_{\sup} \le 1\}$ and the **arithmetic volume** is $\widehat{\operatorname{vol}}(\overline{\mathcal{L}}) = \limsup_n \widehat{h}^0(\mathcal{X}, \overline{\mathcal{L}}^{\otimes n})/(n^{D+1}/(D+1)!)$. Sections s with $\|s\|_{\sup} \le 1$ are called **small**. Plurisubharmonic functions and currents are in the sense of [35].

Notation for abelian schemes is that of [93], [90] and [37]. For a scheme S, an **abelian scheme** $A \to S$ is a smooth proper group scheme with geometrically connected fibres, and a **semiabelian scheme** $G \to S$ is a smooth separated commutative group scheme whose fibres are extensions of abelian varieties by tori. $A^\vee$ is the dual abelian scheme, $e\colon S \to A$ the unit section, $[n]$ multiplication by n, $A[n]$ its kernel and $T_\ell A$, $V_\ell A = T_\ell A \otimes \mathbb{Q}_\ell$ the Tate modules. A **polarised abelian scheme** $(A, \mathcal{L})$ carries a relatively ample line bundle $\mathcal{L}$ with polarisation $\phi_{\mathcal{L}}\colon A \to A^\vee$, and $e_n\colon A[n] \times A[n] \to \mu_n$ is the Weil pairing of $\phi_{\mathcal{L}}$. Line bundles are given by an open cover and transition functions; $\bigotimes_i f_i^*\mathcal{L}^{\otimes n_i}$ is the tensor combination of pull-backs along morphisms f_i. A **rigidified** line bundle is a pair $(\mathcal{L}, r)$ with $r\colon e^*\mathcal{L} \cong \mathcal{O}_S$, and a **cubical structure** on it is a trivialisation of

$$\Theta(\mathcal{L}) = \bigotimes_{\varnothing \ne I \subseteq \{1,2,3\}} m_I^*\mathcal{L}^{\otimes(-1)^{|I|+1}}$$

on A^3 (m_I the partial sums) satisfying the symmetry and cocycle conditions of [90], Ch. I; $\mathcal{G}(\mathcal{L})$ is the theta group. For $S = \operatorname{Spec} V$ with V a complete discrete valuation ring with fraction field K, a **semiabelian model** of an abelian variety A_K is a semiabelian scheme G/V with $G_K \cong A_K$; its formal completion is the extension $\tilde{G}$ (**Raynaud extension**) of an abelian scheme by a torus, which is **split** if the torus is split; and **degeneration data** and the **Mumford uniformisation** $A_K^{\mathrm{an}} = \tilde{G}_K/Y$ are as in [37], Ch. II–III and [94]. A **p-divisible group** over a ring R is written $\mathcal{G}$, with Tate module $T_p\mathcal{G}$, Lie algebra $\operatorname{Lie}\mathcal{G}$, invariant differentials $\omega_{\mathcal{G}}$ and, over the ring of integers of a complete algebraically closed field C, Hodge–Tate map $\alpha_{\mathcal{G}}\colon T_p\mathcal{G} \to \omega_{\mathcal{G}^\vee}$ [105]. An action of a ring O with involution on A (or G, $\mathcal{G}$) is a ring homomorphism $O \to \operatorname{End}(A)$ compatible with the Rosati involution where a polarisation is given. This notation holds wherever abelian varieties occur in this paper.

Let E/F be a CM extension of number fields with complex conjugation c, and write $x \mapsto \bar{x}$ for c. For a finite-dimensional E-vector space Y with a nondegenerate form $\langle\,,\,\rangle$ that is hermitian ($\langle y, x\rangle = \overline{\langle x, y\rangle}$) or skew-hermitian ($\langle y, x\rangle = -\overline{\langle x, y\rangle}$), $\mathrm{U}(Y)$ denotes its unitary group over F, $[\mathrm{U}(Y)] = \mathrm{U}(Y)(F)\backslash\mathrm{U}(Y)(\mathbb{A}_F)$, and its signature at a real place w of F is that of the hermitian form $\langle\,,\,\rangle$, resp. $i\langle\,,\,\rangle$, at the complex place of E above w.

For a hermitian space V and a skew-hermitian space W, $V \otimes_E W$ with the form $\operatorname{tr}_{E/F}(\langle\,,\,\rangle_V \otimes \overline{\langle\,,\,\rangle_W})$ is a symplectic F-space, and $(\mathrm{U}(V), \mathrm{U}(W))$ is a dual pair in its symplectic group. Fix a nontrivial additive character ψ_{add} of $\mathbb{A}_F/F$ and Hecke characters χ_V, χ_W of $\mathbb{A}_E^\times/E^\times$ whose restrictions to $\mathbb{A}_F^\times$ are $\eta^{\dim V}$, resp. $\eta^{\dim W}$, with η the quadratic character of E/F. Then $\omega = \omega_{\psi_{\mathrm{add}}, \chi_V, \chi_W}$ denotes the Weil representation of $\mathrm{U}(V)(\mathbb{A}_F) \times \mathrm{U}(W)(\mathbb{A}_F)$ on a model of the

Schwartz space $\mathcal{S}(V^{\dim W}(\mathbb{A}_F))$ (Schrödinger or mixed model), [72], [49]. For φ in this space,

$$\theta(g,h;\varphi)=\sum_{x\in V^{\dim W}(F)}(\omega(g,h)\varphi)(x),\qquad \Theta(\varphi)(h)=\int_{[\mathrm{U}(V)]}\theta(g,h;\varphi)\,dg,$$

when the integral converges; $\Theta(\varphi)$ is the **theta integral** (theta lift of the trivial representation) of φ, a function on $[\mathrm{U}(W)]$. A parabolic subgroup of $\mathrm{U}(W)$ stabilising an isotropic line $Ex\subset W$ has Levi factor $\mathrm{Res}_{E/F}\mathbb{G}_m\times\mathrm{U}(W_1)$ with $W_1=\langle x,y\rangle^{\perp}$ for an isotropic y with $\langle x,y\rangle=1$; the constant term of a function f on $[\mathrm{U}(W)]$ along it is $f_P(h)=\int_{N(F)\backslash N(\mathbb{A}_F)}f(nh)\,dn$ with N the unipotent radical and total measure 1. This notation holds throughout the parts of this paper on theta lifting.

This convention fixes the notation for ℓ-adic coefficients, Galois representations and automorphic representations used in this paper, following [29] and [8].

Coefficients. ℓ is a prime, $\overline{\mathbb{Q}}_\ell$ an algebraic closure of $\mathbb{Q}_\ell$ with absolute value $|\cdot|_\ell$ and valuation v_ℓ normalised by $v_\ell(\ell)=1$, and $\iota\colon\overline{\mathbb{Q}}_\ell\to\mathbb{C}$ an isomorphism of fields. A **coefficient ring** is a $\mathbb{Z}_\ell$-subalgebra $\mathcal{O}\subset\overline{\mathbb{Q}}_\ell$ that is a discrete valuation ring finite over $\mathbb{Z}_\ell$; $e_{\mathcal{O}}=\mathrm{length}_{\mathcal{O}}(\mathcal{O}/\ell\mathcal{O})$ is its ramification index. For an $\mathcal{O}$-module M of finite length, its **normalised length** is $\mathrm{length}_{\mathcal{O}}(M)/e_{\mathcal{O}}$, so that $\ell^{[\mathcal{O}:\mathbb{Z}_\ell]\cdot\mathrm{length}_{\mathcal{O}}(M)/e_{\mathcal{O}}}=\#M$. For a character θ of a group with values in $\mathcal{O}^\times$, $\mathcal{O}^\vee(\theta)$ denotes $\mathcal{O}\otimes_{\mathbb{Z}_\ell}\mathbb{Q}_\ell/\mathbb{Z}_\ell$ with the group acting through θ (discrete topology). Dirichlet characters with values in $\mathcal{O}^\times$ are identified with characters of $G_{\mathbb{Q}}$ by class field theory.

Galois representations. An ℓ-adic representation of G_E is a continuous homomorphism $r\colon G_E\to\mathrm{GL}_n(\overline{\mathbb{Q}}_\ell)$; $r^\vee$ is its contragredient, and for a CM field E with complex conjugation c, $r^c(g)=r(cgc)$. Two representations are **isomorphic** if they are conjugate in $\mathrm{GL}_n(\overline{\mathbb{Q}}_\ell)$.

Automorphic side. For a number field E, a cuspidal automorphic representation π of $\mathrm{GL}_n(\mathbb{A}_E)$ is **regular algebraic** in the sense of [28]; for E CM, π is **conjugate self-dual** if $\pi^\vee\cong\pi^c$. An ℓ-adic representation r of G_E is **attached to** π, written $r\leftrightarrow_\iota\pi$, if for all but finitely many finite places w of E at which π is unramified, r is unramified at w and ι carries the characteristic polynomial of $r(\mathrm{Frob}_w^{-1})$ (a geometric Frobenius) to $\prod_\alpha\bigl(X-\mathrm{N}w^{(n-1)/2}\alpha\bigr)$, where α runs over the Satake parameters of π_w (the eigenvalues of the unramified character whose normalised induction contains π_w). For E CM, r is **automorphic of conjugate self-dual type** if $r\leftrightarrow_\iota\pi$ for some regular algebraic conjugate self-dual cuspidal π. A normalised newform f of weight k, level $\Gamma_1(M)$ and character χ has the Galois representation $\rho_{f,\iota}\colon G_{\mathbb{Q}}\to\mathrm{GL}_2(\overline{\mathbb{Q}}_\ell)$ of [32].

6 Master data and families from theta lifts

This section constructs the lattice data that the bridge needs, and outlines the proof of Proposition 6.20: if hosts are supplied on windows and the fields of their members satisfy the Sha bound, then tied host families exist, with exponents (e_D,e_L) independent of ε. It also states Lemma 6.14, which connects tiedness with the core-tied pairs of Section 4. The lattices are attached to master data: a host enlarged to a ranked host, a tower of definite hermitian spaces over the host field, discrete-series parameters at the places of a CM type, iterated kernels of theta type, and local data. Families of master data are tied through a definite-side family structure (Proposition 6.18), assembled from the theta-lift results of Section 7 and the covolume bound of Section 8; the Sha bounds of Section 9 enter through a linking inequality that bounds the local defects of the lattices. The known theorems used include the Siegel–Weil formula for unitary groups [59], which goes back to [133] and [73], the local Langlands correspondence [50], [55], mass formulas for hermitian lattices [113] and Green's parametrisation of the cuspidal characters of GL_n over a finite field [46]. This is the largest part of the argument and the one we

understand least: most of its definitions are long records that we state in outline, and most of its proofs name their inputs and little more.

6.1 Ranked hosts and scales

The numerical skeleton comes first. A rank datum fixes, besides a frame, three integers k, m, t from which the dimensions and signatures of all hermitian spaces are derived; a Harish-Chandra parameter for it is a discrete-series parameter whose entries interlace as in the branching laws of unitary groups [48].

Definition 6.1. A **rank datum** is a tuple $\mathfrak{R} = (e, A, h, a_0; k, m, t)$ consisting of a frame (e, A, h, a_0) and integers k, m, t such that

(R1) k and m are even and t is odd;

(R2) $2 \le m$, $m + 3 \le k$, $k + m + 4 \le t$ and $3m < t$;

(R3) $m \le 2|A| \le k$;

(R4) $m^2 \le k + m(t + 1)$.

Its **derived ranks** are

$$f = 2|A|, \qquad r = k + m - 1, \qquad n_V = r + t, \qquad r_S = \tfrac{m}{2} - 1, \qquad k_j = k + j(t + 1) \quad (j \ge 0),$$

and its **signatures** are $\mathrm{sig}_Y = (k + r_S,\, r_S + 1)$ and $\mathrm{sig}_W(j) = (k + j - 1,\, m - j)$ for $1 \le j \le m$.

A **Harish-Chandra parameter** for $\mathfrak{R}$ is a tuple $(u, v, b, \tilde{b})$ with $u \in (\frac{1}{2} + \mathbb{Z})^k$, $v \in \mathbb{Z}^{k-1}$ and $b, \tilde{b} \in \mathbb{Z}^m$ such that

(Pa1) $u_1 > v_1 > u_2 > v_2 > \cdots > v_{k-1} > u_k > -b_1$;

(Pa2) $\tilde{b}_j - b_j \ge \frac{t+3}{2}$ for $1 \le j \le m$, and $b_{j+1} - \tilde{b}_j \ge \frac{t+3}{2}$ for $1 \le j < m$.

Its **shift exponents** are the integers $s_j^\sharp = \tilde{b}_j - \frac{1}{2}k_j$ and $s_j^\flat = \tilde{b}_j - \frac{1}{2}k_{j-1}$ for $1 \le j \le m$.

A ranked host is a host with the further data that the theta-lift construction needs: a complex embedding, an ideal $\mathfrak{b}$ whose product with its conjugate is the relative different over $K = \mathbb{Q}(\sqrt{-d})$, level and auxiliary primes, and the top and thick places that index the weights.

Definition 6.2. Let $\mathfrak{R} = (e, A, h, a_0; k, m, t)$ be a rank datum, $o = [A : A_0]$, $\mathcal{G}^+ = \ker \epsilon_2 \subset \mathcal{G}$, and for $a \in A$ put $\tau_1(a) = (1, 1, a)$ and $\tau_2(a) = (-1, 1, a)$ in $\mathcal{G}^+$. A **ranked host** for $\mathfrak{R}$ is a host $\mathfrak{H} = (q, d, Q, E, \iota, \alpha)$ for the frame of $\mathfrak{R}$ together with data

$$(\varphi_0, \mathfrak{b}, S, g_1, g_2, p_1, p_2, \mathfrak{c}, T_{\mathrm{top}}, T_{\mathrm{th}}, a_1, n_{\mathrm{th}}, B)$$

such that:

(B1) $\epsilon_1(\alpha(u)) = (u/q)$ for every $u \in (\mathbb{Z}/Qqd\mathbb{Z})^\times$, and E contains an element $\sqrt{-q}$ with $\sigma(\sqrt{-q}) = \epsilon_1(\iota\sigma)\sqrt{-q}$ for all $\sigma \in \mathrm{Gal}(E/\mathbb{Q})$;

(B2) for every prime $p \mid Q$ there is $u \equiv 1 \pmod{Qqd/p}$ with $\alpha(u) \ne 1$;

(B3) $Q \le |A|^2$, $o^{64} \le 2^e$, $q \ne 3$ and $4tr + 1 < \min(q, d)$;

(B4) $B \geq 0$ is an integer such that $(-d/p) = 1$ for every odd prime $p \leq B$, and $4p \leq B$ for every odd prime $p \mid Q$;

(B5) $\varphi_0 \colon E \to \mathbb{C}$ is an embedding with $\varphi_0(\sqrt{-d}) = i\sqrt{d}$;

(B6) $\mathfrak{b}$ is an ideal of $\mathcal{O}_E$ with $\mathcal{D}_{E/K} = \mathfrak{b}\,\overline{\mathfrak{b}}$ and $\mathfrak{b} + \overline{\mathfrak{b}} = \mathcal{O}_E$, and $\sigma\mathfrak{b} = \mathfrak{b}$ whenever $\iota\sigma \in \mathcal{G}^+$;

(B7) S is a finite set of primes (the **level primes**), each coprime to Q, containing two distinct primes g_1, g_2 and two distinct primes $p_1, p_2 \equiv 1 \pmod{15}$ different from g_1, g_2;

(B8) $\mathfrak{c}$ is a non-zero ideal of $\mathcal{O}_K$, contained in no ideal $m\mathcal{O}_K$ with $m > 1$, every prime ideal containing $\mathfrak{c}$ has prime norm p with $(-d/p) = 1$, and $\mathrm{N}\mathfrak{c}$ is coprime to 2, 3, 5, q, d, Q and to every level prime;

(B9) $T_{\mathrm{top}} \subseteq A$ and $T_{\mathrm{th}} \subset \mathbb{Z}_{\geq 0}$ are finite, $a_1 \in A$, and $n_{\mathrm{th}} \geq 2$ is an integer (the **spacing integer**) with $4n_{\mathrm{th}}s < |A_0|$ for all $s \in T_{\mathrm{th}}$ and

$$j(s_2 - s_1) \neq i(s_3 - s_1)$$

for all pairwise distinct $s_1, s_2, s_3 \in T_{\mathrm{th}}$ and all integers $i \neq j$ with $0 < |i|, |j| \leq n_{\mathrm{th}}$.

The **top places** are $\tau_1(T_{\mathrm{top}})$, the **thick places** are $\tau_2(sa_0 + a_1)$ for $s \in T_{\mathrm{th}}$, and the **thin places** are the remaining elements of $\mathcal{G}^+$. The **class** of a prime p is: ram_2 if $p \mid Q$; ram_q if $p = q$; ram_d if $p = d$; split if otherwise $(-d/p) = 1$; signal if otherwise $\epsilon_1(\alpha(p)) = 1$; inert in the remaining case.

Forgetting all data except $(E, d, \sqrt{-d})$, the field of a host is a **CM host field**, the weaker notion used for CM points later on: E is a CM field, $d \equiv 7 \pmod{16}$ is prime, $\sqrt{-d} \in \mathcal{O}_E$ (so $(1 + \sqrt{-d})/2 \in \mathcal{O}_E$), and E is unramified at all finite primes over its maximal totally real subfield (the last condition is among the derived properties of the host: complex conjugation lies in no inertia group). A CM host field carries no frame, Artin map or ranked data.

A ranked host is moreover assumed to satisfy the arithmetic properties that follow from conditions (H1)–(H5) of Definition 3.1 by class field theory of cyclotomic fields: E is CM of degree $4|A|$; Frobenius, decomposition and inertia groups are given by α; the relative different is $\mathfrak{b}\overline{\mathfrak{b}}$; and the root discriminant obeys explicit bounds. These we do not reproduce, nor a real constant that enters only a size bound.

A scale fixes the sizes. At a scale x the group A has order about $x^{2+8\varepsilon}$, m is about $x^{1/2}$ and k about $x^{9/2+10\varepsilon}$, and t is determined by (Sc4), which makes it a power of x of exponent somewhat below 400; the size parameter is $b = \exp(x^{\varepsilon})$.

Definition 6.3. A **scale** is a triple $\mathfrak{s} = (x, \varepsilon, J)$ of reals with $x > 1$, $0 < \varepsilon < 1/52$ and $J \geq 1$. Its **derived quantities** are

$$L(\mathfrak{s}) = x^{\varepsilon}, \quad b(\mathfrak{s}) = e^{L(\mathfrak{s})}, \quad a(\mathfrak{s}) = x\,b(\mathfrak{s}),$$
$$X'(\mathfrak{s}) = \exp\bigl(x^{\varepsilon/3}\bigr), \quad b'(\mathfrak{s}) = \frac{a(\mathfrak{s})}{X'(\mathfrak{s})}, \quad \delta_1(\mathfrak{s}) = x^{-\varepsilon/8},$$

and the fixed exponent $e_{\mathrm{w}} = 26$.

A rank datum $\mathfrak{R}$ **has the sizes of the scale** $\mathfrak{s}$ if

(Sc1) $x^{2+8\varepsilon} \leq |A| \leq 2x^{2+8\varepsilon}$;

(Sc2) $\sqrt{x\,\delta_1(\mathfrak{s})} \leq m \leq 2\sqrt{x\,\delta_1(\mathfrak{s})}$;

(Sc3) $x^{9/2+10\varepsilon} \le k \le 2x^{9/2+10\varepsilon}$;

(Sc4) $r + t = 8fr + 4 + \lceil x^{377} \rceil \, n_W(\mathfrak{R}, \mathfrak{s})$, where

$$n_W(\mathfrak{R}, \mathfrak{s}) = 8\Big\lceil \tfrac{10}{8}\, r\Big(f + \frac{f^2(k-1)^3 L(\mathfrak{s})^2}{m}\Big)\Big\rceil.$$

Let $\mathfrak{H}$ be a ranked host for $\mathfrak{R}$ with top set T_{top}, and let $c_{\mathrm{w}}(\mathfrak{s}) \ge 1$ be a logarithmic weight attached to $\mathfrak{s}$ by a fixed choice of auxiliary constants (about twenty of them, used throughout the argument), whose definition we do not reproduce; only $c_{\mathrm{w}}(\mathfrak{s}) \ge 1$ is used below. Integer functions $a_\bullet, b_\bullet \colon A \to \mathbb{Z}$ are **weights in the ranges of** $(\mathfrak{s}, \mathfrak{H})$ if for every $u \in T_{\mathrm{top}}$

$$a(\mathfrak{s}) \le a_u \le a(\mathfrak{s}) + \frac{a(\mathfrak{s})}{k\, x^{e_{\mathrm{w}}/2}},$$
$$b'(\mathfrak{s}) \le b_u \le b'(\mathfrak{s}) + \frac{2|A|\, \delta_1(\mathfrak{s})\, x^{\varepsilon}\, c_{\mathrm{w}}(\mathfrak{s})\, b'(\mathfrak{s})}{k}.$$

Three elementary lemmas, used in the construction of hosts in Section 3, show that these sizes leave room: the product $m^5 f L$ is at most a constant times $x^{9/2+9\varepsilon}$, there are enough sign words of length $k-2$ to index the objects that the construction needs, and integer weights exist in the two prescribed intervals.

Lemma 6.4. *If a rank datum $\mathfrak{R}$ has the sizes of the scale $\mathfrak{s}$, then*

$$m^5\, f\, L(\mathfrak{s}) \le 128\, x^{9/2+9\varepsilon}.$$

Proof. By (Sc2) and $\delta_1(\mathfrak{s}) = x^{-\varepsilon/8}$,

$$m \le 2\sqrt{x\,\delta_1(\mathfrak{s})} = 2x^{1/2-\varepsilon/16}, \qquad \text{so} \qquad m^5 \le 32\, x^{5/2-5\varepsilon/16}.$$

By (Sc1), $f = 2|A| \le 4x^{2+8\varepsilon} \le 4x^{2+8\varepsilon+\varepsilon/100}$. With $L(\mathfrak{s}) = x^{\varepsilon}$,

$$m^5 f L(\mathfrak{s}) \le 128\, x^{\,5/2-5\varepsilon/16+2+8\varepsilon+\varepsilon/100+\varepsilon} \le 128\, x^{9/2+9\varepsilon},$$

because $x > 1$ and $-5\varepsilon/16 + \varepsilon/100 \le 0$. □

Lemma 6.5. *Let $0 < \varepsilon < 1/52$ and $c_G, c_F \ge 0$. There is x_0 such that for every scale $\mathfrak{s} = (x, \varepsilon, J)$ with $x \ge x_0$ and every rank datum $\mathfrak{R}$ having the sizes of $\mathfrak{s}$, the integer*

$$n_{\mathrm{w}} = \Big\lceil 3\exp\big(c_G\, m^5 f L(\mathfrak{s})\big)^2 x^{c_F} \Big\rceil$$

is positive and there is an injective map $\{1, \dots, n_{\mathrm{w}}\} \to \{0,1\}^{k-2}$.

Proof. Put $K = 2(2 + 256c_G + c_F) + 2$ and $x_0 = K^{1/\varepsilon}$, so that $K \le x^{\varepsilon}$ for $x \ge x_0$, and write $X = x^{9/2+9\varepsilon}$. By (Sc3), $k \ge x^{9/2+10\varepsilon} = x^{\varepsilon} X \ge KX$; by Lemma 6.4, $m^5 f L(\mathfrak{s}) \le 128X$; and $\log x \le X$. Hence

$$3\exp\big(c_G m^5 f L(\mathfrak{s})\big)^2 x^{c_F} \le \exp\big(2 + 256 c_G X + c_F X\big) \le \exp\big(\tfrac{k-2}{2}\big) \le 2^{k-2},$$

using $3 \le e^2$, $KX \le k$ and $\log 2 \ge \frac{1}{2}$. So $n_{\mathrm{w}} \le 2^{k-2} = \#\{0,1\}^{k-2}$, and an injection exists by comparing cardinalities; $n_{\mathrm{w}} \ge 1$ since the quantity inside the ceiling is positive. □

Lemma 6.6. *Let $0 < \varepsilon < 1/52$ and fix the auxiliary constants that define the weight c_{w} of Definition 6.3. There is x_0 such that for every scale $\mathfrak{s} = (x, \varepsilon, J)$ with $x \geq x_0$, every rank datum $\mathfrak{R}$ having the sizes of $\mathfrak{s}$ and every ranked host $\mathfrak{H}$ for $\mathfrak{R}$, there exist weights $a_\bullet, b_\bullet \colon A \to \mathbb{Z}$ in the ranges of $(\mathfrak{s}, \mathfrak{H})$.*

Proof. In outline. It suffices that both weight ranges have length at least 1 for $x \geq x_0$; then integers in them exist for every $u \in T_{\mathrm{top}}$. By (Sc3), $k \leq 2x^{9/2+10\varepsilon} \leq x^5$ for large x, and $b(\mathfrak{s}) = \exp(x^\varepsilon) \geq x^{17}$ for large x, so

$$k\, x^{e_{\mathrm{w}}/2} \leq x^5\, x^{13} = x \cdot x^{17} \leq x\, b(\mathfrak{s}) = a(\mathfrak{s}),$$

which is the first length bound. For the second range the length is $2|A|\delta_1(\mathfrak{s})x^\varepsilon c_{\mathrm{w}}(\mathfrak{s})b'(\mathfrak{s})/k$; with $|A| \geq 1$, $\delta_1(\mathfrak{s})x^\varepsilon = x^{7\varepsilon/8} \geq 1$, $c_{\mathrm{w}}(\mathfrak{s}) \geq 1$ and $b'(\mathfrak{s}) = x\, e^{x^\varepsilon}/e^{x^{\varepsilon/3}} \geq k$ for large x, it is at least 1. □

6.2 The master datum

A master datum attaches to a ranked host the hermitian spaces and the kernels from which the lattices are built. Its first piece is a tower of hermitian spaces. It is not a Witt tower in the sense of theta theory [101]: the spaces X_j are definite, of dimensions $k_j = k + j(t+1)$, and the spaces W_j are locally isomorphic to Y_0 at all finite places.

Definition 6.7. Let E/F be a CM extension such that every embedding $\tau \colon E \to \mathbb{C}$ satisfies $\tau(\bar{x}) = \overline{\tau(x)}$ and for every maximal ideal $\mathfrak{p}$ of $\mathcal{O}_E$ there is $\vartheta \in \mathcal{O}_E$ with $\vartheta - \bar{\vartheta} \notin \mathfrak{p}$. A **based hermitian space** is a nondegenerate hermitian space $(V, \langle\,,\rangle)$ over E with an E-basis, a unimodular $\mathcal{O}_E$-lattice $L_0 \subset V$ (equal to its hermitian dual) and, for every embedding $\tau \colon E \to \mathbb{C}$, the signature $\mathrm{sig}_\tau(V)$ of the hermitian matrix $\tau(\langle e_i, e_j\rangle)$. It is **definite** if $\mathrm{sig}_\tau(V) \in \{(\dim V, 0), (0, \dim V)\}$ for every τ.

A **hermitian tower** for a rank datum $\mathfrak{R}$ over E/F is a tuple

$$\begin{aligned}\mathfrak{X} = \big(&T;\, (X_j)_{j\geq 0},\, X',\, \iota',\, z;\, (Y_j)_{j\geq 0},\, (\iota_j),\, \kappa,\, z_0,\, (e_i, f_i)_{i\leq r_S};\\ &(W_j)_{1\leq j\leq m};\, (V_j, \delta_j)_{1\leq j\leq m}\big)\end{aligned}$$

of a prime T and based hermitian spaces and maps such that:

(X1) $\dim X_j = k_j$ and X_j is definite for $0 \leq j \leq m$;

(X2) $\dim X' = r + mt$, $\iota' \colon X' \to X_m$ is an isometric embedding and $z \in X_m$ satisfies $\langle z, z\rangle \neq 0$ and $z \perp \iota'(X')$;

(X3) $\dim Y_j = r + jt$ for $0 \leq j \leq m$, $\iota_j \colon X_j \to Y_j$ is an isometric embedding for $j < m$, and $\kappa \colon X' \to Y_m$ is an isometry;

(X4) $z_0 \in Y_0$ has $\langle z_0, z_0\rangle = -T$, and $e_1, \dots, e_{r_S}, f_1, \dots, f_{r_S} \in Y_0$ satisfy $\langle e_i, e_{i'}\rangle = \langle f_i, f_{i'}\rangle = 0$ and $\langle e_i, f_{i'}\rangle = \delta_{ii'}$, all orthogonal to $\iota_0(X_0)$ and to z_0; $\mathrm{sig}_\tau(Y_0) = \mathrm{sig}_Y$ for every τ;

(X5) for $1 \leq j \leq m$, $\dim W_j = r$, $\mathrm{sig}_\tau(W_j) = \mathrm{sig}_W(j)$ for every τ, and $W_j \otimes_F \mathbb{A}_{F,f}$ is isometric to $Y_0 \otimes_F \mathbb{A}_{F,f}$;

(X6) for $1 \leq j \leq m$, $\dim V_j = n_V$, $\delta_j \in \{\pm 1\}$, and the base lattice of V_j has an $\mathcal{O}_E$-basis that is orthonormal for $\delta_j\langle\,,\rangle$.

Its second piece is a chain of kernels with determinant twists between spaces of functions on finite sets of classes. The composite, the column map (6.1), presumably represents the theta lifts along the tower on these finite sets; we understand the definition but not this interpretation. Gauss data are operator-valued Gauss sums at an auxiliary prime T.

Definition 6.8. Let $\mathfrak{R}$ be a rank datum, Φ a finite set and $(\Xi_\tau)_{\tau\in\Phi}$ Harish-Chandra parameters for $\mathfrak{R}$; write $s_j^\sharp(\tau), s_j^\flat(\tau)$ for the shift exponents of Ξ_τ. For a finite set X write $\mathbb{C}^X$ for the functions $X\to\mathbb{C}$.

A **column chain** for $(\mathfrak{R},(\Xi_\tau))$ is a tuple $\mathfrak{C}=(\mathrm{Cl}_i,\Gamma_{i,c},D_\Gamma,\mathcal{Y}_i,\delta_i,\rho_i,\rho_i',U_i)_{i\ge0}$ of finite sets Cl_i (the **classes**), finite groups $\Gamma_{i,c}$ for $c\in\mathrm{Cl}_i$, an integer $D_\Gamma\ge1$ divisible by $|\Gamma_{i,c}|$ for all $i\le m$ and all c, finite sets $\mathcal{Y}_i$, functions $\delta_i\colon\mathcal{Y}_i\times\Phi\to\mathbb{C}^\times$, numbers $\rho_i(c)$ ($c\in\mathrm{Cl}_i$) and $\rho_i'(c')$ ($c'\in\mathrm{Cl}_{i+1}$) of absolute value 1 for $i<m$, and linear maps $U_i(c,c')\colon\mathbb{C}^{\mathcal{Y}_i}\to\mathbb{C}^{\mathcal{Y}_{i+1}}$. Its **kernels** are

$$K_i(c,c')=\rho_i(c)\rho_i'(c')\;D_i^\flat\circ U_i(c,c')\circ D_i^\sharp,$$

where $D_i^\sharp$ is multiplication by $y\mapsto\prod_{\tau\in\Phi}\delta_i(y,\tau)^{s_{i+1}^\sharp(\tau)}$ and $D_i^\flat$ is multiplication by $y'\mapsto\prod_{\tau\in\Phi}\delta_{i+1}(y',\tau)^{-s_{i+1}^\flat(\tau)}$. Its **transfer** maps $\mathcal{K}_i\colon\mathbb{C}^{\mathrm{Cl}_i}\otimes\mathbb{C}^{\mathcal{Y}_i}\to\mathbb{C}^{\mathrm{Cl}_{i+1}}\otimes\mathbb{C}^{\mathcal{Y}_{i+1}}$ and its **column map** are

$$(\mathcal{K}_i\varphi)(c')=\sum_{c\in\mathrm{Cl}_i}\frac{1}{|\Gamma_{i,c}|}\,K_i(c,c')\,\varphi(c),\qquad \mathrm{col}_{\mathfrak{C}}=\mathcal{K}_{m-1}\circ\cdots\circ\mathcal{K}_0. \tag{6.1}$$

Gauss data for $\mathfrak{C}$ are a tuple $\mathfrak{T}=(T,(\mathbb{F}_v,\psi_v,\rho_v)_{v\in V_T},M_0,d^*,D^*,\Gamma_P)$ of a prime T, a set V_T with f elements, for each $v\in V_T$ a field $\mathbb{F}_v$ with T elements, a nontrivial additive character ψ_v of $\mathbb{F}_v$ and a representation ρ_v of $\mathbb{F}_v^\times$ on $\mathbb{C}^{\mathrm{Cl}_0\times\mathcal{Y}_0}$, a ρ_v-stable subspace M_0 (for all v), an integer $d^*\ge1$ and reals $D^*>0$, $\Gamma_P>1$. Its **ramified part**, **Gauss operators** and **scales** are

$$M_{\mathrm{ram}}=M_0\cap\bigcap_{v\in V_T}\ker\sum_{a\in\mathbb{F}_v^\times}\rho_v(a),\qquad \mathrm{G}_v=\sum_{a\in\mathbb{F}_v^\times}\psi_v(a)\rho_v(a),$$
$$X_u=\Big(\frac{D^*}{d^*}\Big)^{1/(fk)},\quad T^*=X_u\,\Gamma_P^{-2/(fk)}.$$

A **period map** for $\mathfrak{C}$ (the name refers to the column periods it carries, not to period maps in Hodge theory) is a tuple $\mathfrak{P}=(M_t,W_z,W_{\mathrm{row}},\mathrm{ev},\mathrm{low},\mathrm{per})$ of complex vector spaces, linear maps $\mathrm{ev}\colon\mathbb{C}^{\mathrm{Cl}_m\times\mathcal{Y}_m}\to W_z$, $\mathrm{low}\colon W_z\to W_{\mathrm{row}}$ and a bilinear form $\mathrm{per}\colon W_{\mathrm{row}}\times M_t\to\mathbb{C}$; its **column period** is $\widetilde{P}(\varphi,F)=\mathrm{per}(\mathrm{low}(\mathrm{ev}(\mathrm{col}_{\mathfrak{C}}\varphi)),F)$.

The local p-adic data are expressed through an axiomatised period-ring datum, which a tie identifies with Fontaine's rings [40].

Definition 6.9. Let ℓ be a prime. A **period-ring datum** at ℓ is a tuple

$$\mathfrak{F}=\big(B,B'_{\mathrm{dR}},a,a_{\mathrm{dR}},\varphi,\iota,t,\chi_{\mathfrak{F}},\mathrm{Fil},\sigma,(\zeta_m)_{m\ge0},(v_K)_K\big)$$

consisting of a commutative $\mathbb{Q}_\ell^{\mathrm{ur}}$-algebra B, a commutative $\overline{\mathbb{Q}}_\ell$-algebra B'_{dR}, actions a, a_{dR} of $G_{\mathbb{Q}_\ell}$ on B and B'_{dR} by $\mathbb{Q}_\ell$-algebra automorphisms, a $\mathbb{Q}_\ell$-algebra endomorphism φ of B, a $\mathbb{Q}_\ell$-algebra map $\iota\colon B\to B'_{\mathrm{dR}}$, an element $t\in B$, a character $\chi_{\mathfrak{F}}\colon G_{\mathbb{Q}_\ell}\to\mathbb{Q}_\ell^\times$, a $\mathbb{Z}$-indexed filtration Fil of B'_{dR}, a Frobenius σ on each K_0, a compatible system of ℓ-power roots of unity and valuations v_K on the finite extensions $K/\mathbb{Q}_\ell$, such that for every finite extension $K/\mathbb{Q}_\ell$ in $\overline{\mathbb{Q}}_\ell$, with $G=G_K$:

(Pr1) σ is a bijective $\mathbb{Q}_\ell$-algebra endomorphism of K_0 lifting Frobenius: $|\sigma(x)-x^\ell|<1$ for $|x|\le1$; G acts on B by K_0-algebra automorphisms and $B^G=K_0$; G acts on B'_{dR} by K-algebra automorphisms and $(B'_{\mathrm{dR}})^G=K$; on $\mathbb{Q}_\ell^{\mathrm{ur}}\subset B$ and $\overline{\mathbb{Q}}_\ell\subset B'_{\mathrm{dR}}$ the actions are the Galois actions;

(Pr2) φ is injective, σ-semilinear on K_0 and commutes with G; ι is an injective G-equivariant K_0-algebra map;

(Pr3) Fil^a are K-subspaces, independent of K, decreasing, exhaustive, separated, multiplicative, G-stable, with $1 \in \mathrm{Fil}^0$;

(Pr4) t is a unit of B with $g\,t = \chi_{\mathfrak{F}}(g)\,t$, $\varphi(t) = \ell t$, $\iota(t) \in \mathrm{Fil}^1 \setminus \mathrm{Fil}^2$; and $\chi_{\mathfrak{F}}$ is cyclotomic: $g(\zeta) = \zeta^a$ for some integer a with $|\chi_{\mathfrak{F}}(g) - a| \le \ell^{-m}$ whenever $\zeta^{\ell^m} = 1$;

(Pr5) (fundamental property) if $b \in B$, $\varphi(b) = b$ and $\iota(b) \in \mathrm{Fil}^0$, then $b \in \mathbb{Q}_\ell$;

(Pr6) $v_K(\ell) = 1$ and $v_K(x) \ge 0$ exactly on $\mathcal{O}_K$.

An **A_{inf}-structure** on $\mathfrak{F}$ is an injective ring map $j\colon A_{\mathrm{inf}} \to B$ that is $G_{\mathbb{Q}_\ell}$-equivariant and commutes with Frobenius, together with the element $\underline{\zeta} = (\zeta_m)_m \in \mathcal{O}_{\mathbb{C}_\ell^\flat}$, such that $\iota j(A_{\mathrm{inf}}) \subseteq \mathrm{Fil}^0$, $\iota j(x) \in \mathrm{Fil}^1$ if and only if $x \in \big(\sum_{i=0}^{\ell-1} [\underline{\zeta}^{1/\ell}]^i\big) A_{\mathrm{inf}}$, and j sends Teichmüller lifts of roots of unity of order prime to ℓ to their images in $\mathbb{Q}_\ell^{\mathrm{ur}} \subset B$.

A **tie** of $(\mathfrak{F}, j)$ is a pair of ring isomorphisms $B \cong B_{\mathrm{cris}}$ and $B'_{\mathrm{dR}} \cong B_{\mathrm{dR}}$ onto Fontaine's rings, compatible with the $\mathbb{Q}_\ell$- and $\overline{\mathbb{Q}}_\ell$-structures, the Galois actions, Frobenius, ι and the filtrations, sending $\iota(t)$ to $t = \log[\underline{\zeta}]$ and j to the canonical map $A_{\mathrm{inf}} \to B_{\mathrm{cris}}$. The datum is subject to further side conditions of the same nature (compatibility of the unramified structures and of the data for different K), which we do not reproduce.

The reader should think of a master datum as a CM field with an explicit Artin map, a tower of definite hermitian spaces with prescribed signatures, discrete-series parameters at the places of a CM type, iterated kernels of theta type with determinant characters, and local p-adic and Galois–Hecke data. Items (D1), (D3), (D5) and (D6) are given in full; the others are given at the level of their data.

Definition 6.10. A **master datum** is a tuple $\mathfrak{D}$ consisting of the following data, subject to the conditions stated with them.

(D1) Ranks, host, tower. A rank datum $\mathfrak{R}$; a ranked host $\mathfrak{H} = (q, d, Q, E, \iota, \alpha; \varphi_0, \mathfrak{b}, S, \dots)$ for $\mathfrak{R}$, with $K = \mathbb{Q}(\sqrt{-d})$ and F the maximal totally real subfield of E; a hermitian tower $\mathfrak{X}$ for $\mathfrak{R}$ over E/F, whose spaces Y_j have $\mathrm{sig}_\tau(Y_j) = \big(r + jt - \lfloor \frac{m-j}{2} \rfloor, \lfloor \frac{m-j}{2} \rfloor\big)$ for $j \le m$ and all τ; and a positive function vol on based hermitian spaces over E.

(D2) Coefficient rows. A totally complex Galois number field L with an embedding $L \subset \mathbb{C}$ under which its involution is complex conjugation; a subfield L_0 and an embedding $E \to L_0 \subseteq L$ compatible with φ_0 and with conjugation; and an identification of the set of places of L (with conjugation, residue characteristics, ramification indices and normalised valuations) with the maximal ideals of $\mathcal{O}_L$. Together with an order $\mathcal{O} \to L$ containing $\mathcal{O}_E$, these form a **rows datum**. Its further data are:

- a set of **twist types** μ with base type μ_0 and types $\mu^*, \mu^{*\prime}$, and a type $\mu(w) \in \{\mu_0, \mu_{\mathrm{tw}}(w)\}$ at each place w, stable under conjugation;
- a finite-dimensional hermitian space $\mathcal{E}$ with an orthonormal basis of size $D_{\mathrm{rows}} = (1 - \delta_p) D_{\mathrm{all}}$ $(0 \le \delta_p \le 1)$, a second space $\mathcal{E}_0$ with a basis of size D_0, and a labelling of both bases by a finite set FT of types;
- for each twist type μ an ambient hermitian space $\mathrm{Amb}(\mu)$ with an $\mathcal{O}$-lattice and a self-adjoint idempotent, and multiplicity spaces $\mathrm{Mult}(\mu, j)$ $(0 \le j \le m)$ with $\mathrm{Mult}(\mu, 0) \cong \mathcal{E}$, $\mathrm{Mult}(\mu, m) \cong \mathrm{Amb}(\mu)$ isometrically and maps $\mathrm{Mult}(\mu, j+1) \to \mathrm{Mult}(\mu, j)$;

- for each word ω a hermitian space $\mathcal{M}_\omega$, period pairings $\mathrm{per}_{\omega,\mu}\colon \mathcal{M}_\omega \times \mathrm{Amb}(\mu) \to \mathbb{C}$ and $B_\omega\colon \mathcal{M}_\omega \times \mathcal{E} \to \mathbb{C}$, and units $z_{\omega,\mu} \in L^\times$ with $z_{\omega,\mu_0} = 1$;
- spaces $\mathcal{M}_j$ and $\mathcal{M}_t(\mu, j)$ with $\mathcal{O}$-lattices, L-structures, self-adjoint idempotents and theta maps $\mathcal{M}_t(\mu, j) \to \mathcal{M}_j$, together with isotypic subspaces of $\mathcal{M}_j$ indexed by FT;
- integer weights at the CM-type places of the signature $\mathrm{sig}_W(j)$, numbers $\lambda_P(\mu, j, \pi)$, $I_j(\mu, w)$, $j(\mu, w)$, non-negative rationals $\eta_j(w)$ and Sha lengths $\mathrm{len}(\ell, e) \geq 0$, and the order of a finite group of rows.

(D3) Parameters. A finite set Φ with a bijection $\Phi \to \mathcal{G}^+$ (the CM-type places); a finite set of **words** with a base word ω_0; Harish-Chandra parameters $\Xi_{\omega,\tau}$ for $\mathfrak{R}$ (ω a word, $\tau \in \Phi$) such that $\Xi_{\omega_0,\tau} = \Xi_\tau$ and $\Xi_{\omega,\tau}$ differs from Ξ_τ only in the interior entries $u_2, \dots, u_{k-1}$; positive weights $\delta(\tau)$.

(D4) Members. For each word ω: a column chain $\mathfrak{C}_\omega$ for $(\mathfrak{R}, (\Xi_{\omega,\tau}))$ with Gauss data $\mathfrak{T}_\omega$ at a prime T_ω and a period map $\mathfrak{P}_\omega$; an $\mathcal{O}_E$-lattice in X_i for each class of $\mathfrak{C}_\omega$ at step i; isomorphisms of $\mathcal{M}_\omega$ with the ramified part M_{ram} of $\mathfrak{T}_\omega$ and of $\mathrm{Amb}(\mu_0)$ with the space M_t of $\mathfrak{P}_\omega$, under which the base period pairing $\mathrm{per}_{\omega,\mu_0}$ is the column period $\widetilde{P}$ of $(\mathfrak{C}_\omega, \mathfrak{P}_\omega)$; and, for each twist type μ, a twisted column chain (equal to $\mathfrak{C}_\omega$ for $\mu = \mu_0$) with a period map, such that $\mathrm{per}_{\omega,\mu}$ is its column period. The pairings B_ω are bilinear. Bases of the divisor spaces map onto bases of $\mathcal{E}$ and of $\mathcal{M}_\omega$, and the rows data agree with the divisor data (D_{all}, δ_p).

(D5) Gauss primes. Each T_ω is a level prime different from p_1, p_2, $T_\omega = \pi\bar{\pi}$ with $\pi \in \mathcal{O}_E \cap K$, $\pi \equiv 1 \pmod{Qqd(g_1g_2)^2}$; the prime of $\mathfrak{X}$ is T_{ω_0}; there are $c_\omega \in E^\times$ with $c_\omega \bar{c}_\omega T_\omega = T_{\omega_0}$; $T^* \leq T_\omega \leq \left(1 + \frac{1}{fk}\right)T^*$ for the scale T^* of $\mathfrak{T}_\omega$; and Γ_P and D^* do not depend on ω.

(D6) Determinant characters. An abelian group Δ with a class map det from the classes of every $\mathfrak{C}_\omega$ to Δ, characters $\tilde{\mu}_j, \chi_\Delta\colon \Delta \to \mathbb{C}^\times$, a subgroup $\Delta_0 \subseteq \ker \chi_\Delta$, a homomorphism e_1 from the norm-one group of $E^\times$ to Δ with $\chi_\Delta(e_1(z)) = \prod_{\tau\in\Phi} \varphi_\tau(z)^{-1}$ (with $\varphi_\tau = \varphi_0 \circ \iota^{-1}(\tau)$), and a relative determinant of self-dual $\mathcal{O}_E$-lattices with values in Δ/Δ_0 which is a cocycle and equals the class of $e_1(\det\sigma)$ on the pair $(M, \sigma M)$ for unitary σ; moreover, for every class c at step i and c' at step $i+1$,

$$\rho_i(c) = \tilde{\mu}_{i+1}(\det c)^{-1}\chi_\Delta(\det c)^{-k_{i+1}/2}, \qquad \rho'_i(c') = \tilde{\mu}_{i+1}(\det c')\,\chi_\Delta(\det c')^{k_i/2}.$$

(D7) Local p-adic data. For every prime ℓ: rings $B \subset B_{\mathrm{dR}}$ over $\mathbb{Q}_\ell$ with actions of the absolute Galois group of $\mathbb{Q}_\ell$, a Frobenius φ on B, an element $t \in B$, a character χ_Q, and for each finite extension $K/\mathbb{Q}_\ell$ period data (filtrations) compatible with these; A_{inf}-data (the Galois action on $\mathbb{C}_\ell$ and on the tilt of $\mathcal{O}_{\mathbb{C}_\ell}$, a compatible system of ℓ-power roots of unity); and an identification of $(B, B_{\mathrm{dR}}, \varphi, \mathrm{Fil}, t)$ with Fontaine's B_{cris}, B_{dR}, φ, Fil and $\log[\underline{\zeta}]$, compatible with all structure; that is, period-ring data in the sense of Fontaine's theory [40], axiomatised as in Definition 6.9.

(D8) Galois and Hecke data. An algebraic closure Ω of E with a choice of places, and local Galois data at every prime p such that the Sha lengths of the rows datum are the lengths of the Tate–Shafarevich groups of these data: $\mathrm{len}(p, e) = \mathrm{len}\,\mathrm{Sha}(p, e)$ for $0 < e < t$. A finite index set of **local situations** i, each with a prime ℓ_i, a step $1 \leq j_i \leq m$, a twisted place w_i of L above ℓ_i with $\mu(w_i) \neq \mu_0$, and:

- a complete local ring $\mathcal{O}_i$ with $L \to \mathrm{Frac}(\mathcal{O}_i)$ compatible with w_i;

- a local $\mathcal{O}_i$-algebra $\mathbb{T}_i$ of functions on a finite set X_i (separating points), acting faithfully on a module $M_{t,i}$ whose rational span receives $\mathcal{M}_t(\mu_{\mathrm{tw}}(w_i), j_i)$, with non-zero Hecke eigenspaces;
- continuous characters $\chi_{i,c}\colon \mathrm{Gal}(\Omega/E) \to \mathcal{O}_i^{\times}$ $(1 \le c \le t)$ determined by $\chi_{i,1}$;
- subsets $Z_i, Z_i^{\mathrm{good}} \subseteq X_i$ and an exponent s_i determined by the ℓ_i-part of Q;
- disjoint finite sets S_i, Q_i of split places of E prime to ℓ_i, with S_i the places above g_1, g_2 and others, Hecke operators at split places generating $\mathbb{T}_i$, inertia generators at $S_i \cup Q_i$, and $\chi_{i,1}$ ramified at every place of Q_i;
- at g_1, g_2, which split completely in E: labels of the types by monic polynomials of degree r over $\mathbb{F}_g$ and characters χ_T, with the order of the finite group of rows equal to $\prod_{g \in \{g_1,g_2\}} |\mathrm{GL}_r(\mathbb{F}_g)|^f$.

Further refinements of these data (twisted, automorphic, base-change and theta refinements, among others) and, at each place of E above ℓ_i, a comparison with the data of (D7) over a coefficient field containing $\mu_{\mathrm{lcm}(\ell_i, Qqd)}$, subject to compatibility conditions; among them, Z_i and Z_i^{good} are the sets defined from the automorphic refinement, and $\eta_j(w) = \sum_{i\colon j_i = j,\, w_i = w} \eta_{\mathrm{loc}}(i)$ for local quantities $\eta_{\mathrm{loc}}(i)$.

(D9) Further data.
- **Divisor data**: a hermitian L-lattice R_0 (the reference lattice) and lattices N_ω for the words, maps $N_\omega \to R_0^\vee$, heights h_ω and a bound $\mathrm{HH} \ge s(N,h) + \mu^{\max}(R_0)$, with unit determinants at all places.
- **Member forms**, for each word: local lines with linear forms and dilation actions at the auxiliary places of $\mathfrak{T}_\omega$ compatible with ρ_v; dominant weights $\lambda_{j,\tau}$ of GL_{k_j} read from $\Xi_{\omega,\tau}$ and σ_τ of GL_{k_m-1}; inner products on the column spaces; good hermitian L-lattices $M_{\omega,i}$ embedded isometrically into the column spaces (the image of $M_{\omega,0}$ is ρ_v-stable); local lattices at each place; and a point lattice for the period map. The maps $\mathcal{M}_\omega \cong M_{\mathrm{ram}}$ are isometries for these forms, and N_ω embeds into $M_{\omega,0}$ onto its ramified sublattice, compatibly with norms at all places.
- **Local split data** at the places of F, split exactly at the places above g_1, g_2, with $\dim = D_{\mathrm{all}}$, isometries $M_0(\mathfrak{T}_\omega) \cong C_\omega$, and an isometry $\mathcal{E} \to H$ into its space H.
- **Haar data**: Haar measures on $\mathrm{U}(P,Q)$ for all P, Q, probability measures on $\mathrm{U}(N)$, and a common density constant for $\mathrm{U}(P,Q)$ and $\mathrm{U}(P+Q)$.
- **Theta jets**: two finite sets of local situations at primes of signal or inert class, at steps j, with lower and upper ranks k_{j-1}, k_j, each attached to a word, classes of its column chain, a place of L above the prime and CM-type places.
- **Global constants**: positive numbers $\Omega(\tau, s)$ and $\theta_1(s)$ for embeddings $s\colon L \to \mathbb{C}$.
- **Hecke data at good index primes**: for each word ω and place w whose residue characteristic p is a good index prime for T_ω, local Hecke data at p of dimension $\dim M_0(\mathfrak{T}_\omega)$.

Items (D2), (D4) and (D7)–(D9) are stated at the level of their data; their compatibility and tie conditions, of the same nature as those shown, we do not reproduce.

A size datum fixes, for a master datum at a scale, the numbers that enter its slope datum: top numbers, local functions on a finite support of places, the budgets B_0 and N_c, and the main term.

Definition 6.11. For a scale $\mathfrak{s}$ and an integer $n \geq 1$ put $w''(\mathfrak{s}, n) = b'(\mathfrak{s})/(xn)$, $X''(\mathfrak{s}, n) = a(\mathfrak{s})/w''(\mathfrak{s}, n)$ and $\varepsilon_J = \varepsilon/J$. For $\tau \in \Phi$ and $1 \leq j \leq m$ let $w_j(\tau) = \tilde{b}_j$, the j-th entry of $\tilde{b}$ in Ξ_τ. A **size datum** for $\mathfrak{D}$ is a scale $\mathfrak{s}$ together with

- **top numbers** $a_t^{\mathrm{top}}, b_t^{\mathrm{top}} \in \mathbb{Z}$ $(t \in T_{\mathrm{top}})$;
- a finite set supp of places of L and functions $\mathrm{gain}_j(w)$, $F_j(w)$, $E_{\omega,j}(w)$, $\tau(w)$, $\mathrm{io}(w)$ of places w, all vanishing off supp;
- a set drop of primes, an integer j_0, a real β, lattices N_ω in the divisor spaces of the words, reals δ_2, L_G, S_c, η_S, C_c, C_t, C_Γ, partial weights $w_{j,t,p,i}^{\mathrm{part}} \geq 1$ and rational lengths $\ell_\omega(w)$;

such that:

(Y1) Top windows. $a(\mathfrak{s}) \leq a_t^{\mathrm{top}} \leq a(\mathfrak{s}) + a(\mathfrak{s})/(kx^{13})$ and $b'(\mathfrak{s}) \leq b_t^{\mathrm{top}}$. If $\tau = \tau_1(t)$, then $u_1 = a_t^{\mathrm{top}} + \frac{k-1}{2}$, $u_k = -b_t^{\mathrm{top}} - \frac{k-1}{2}$ and $b_1 - b_t^{\mathrm{top}} > \frac{k-1}{2} + 3$ for Ξ_τ.

(Y2) Local functions. $\mathrm{gain}_j(w) = 0$ if the residue characteristic of w is g_1 or g_2. If w is untwisted, then $F_j(w) = 0$ and the row number $A_j(w)$ of $\mathfrak{D}$ is 0. If the class of the residue characteristic of w is neither signal nor inert, then $\mathrm{gain}_j(w) = \tau(w) = \mathrm{io}(w) = 0$. $F_j(\bar{w}) = F_j(w)$, and $F_j(w) = 0$ when the residue characteristic exceeds $X''(\mathfrak{s}, |A|)$ and is not g_1, g_2, d.

(Y3) Parameters. $b_j \leq b(\mathfrak{s})^2$ for all entries of all Ξ_τ. No Gauss prime T_ω lies in drop. $\log_5(|A|k_m/m) \leq j_0$ and $10j_0 + 20 \leq n_{\mathrm{th}}$, with n_{th} the spacing integer of the ranked host. β is the logarithmic derivative at $s = 1$ attached to the quadratic character of K. $\lfloor x^{\varepsilon_J/5} \rfloor \leq B$, $o \leq x^{\varepsilon/100}$ and $Q \leq 4|A|x^{\varepsilon/100}$.

(Y4) Thick density. $0 < \delta_2 = |T_{\mathrm{th}}|/|A| \leq 1$.

(Y5) Weight clouds. $S_c = 4k\Delta_S > 0$ with Δ_S the size of the local split data of $\mathfrak{D}$, and $\eta_S > 0$. For $1 \leq j \leq m$: $b(\mathfrak{s}) \leq w_j(\tau) \leq b(\mathfrak{s}) + \eta_S + 1$ at thick places; $b_t^{\mathrm{top}} + b'(\mathfrak{s})/x \leq w_j(\tau) \leq b_t^{\mathrm{top}} + b'(\mathfrak{s})/x + \eta_S$ at $\tau = \tau_1(t)$; $w''(\mathfrak{s}, |A|) \leq w_j(\tau) \leq w''(\mathfrak{s}, |A|) + \eta_S$ at thin places. For τ not a top place, $S_c \leq u_1 - v_1 \leq C_c S_c$. The entries v_i of Ξ_τ and the interior entries $u_2, \ldots, u_{k-1}$ of every $\Xi_{\omega,\tau}$ lie in $[0, S_c]$. $|b_i - \tilde{b}_j| \leq \eta_S$ for all i, j.

(Y6) Further sizes. For each place w a subset of T_{top} is given, equal to all of T_{top} when w is a signal place or its residue characteristic is at least $X'(\mathfrak{s})$. $1 \leq t \leq x^{C_t}$. $C_\Gamma > 0$, and $\log \Gamma_P = C_\Gamma m^5 f L(\mathfrak{s})$ for the Gauss data of every word.

Its **budgets** are $L_a = \log(a(\mathfrak{s})\,\mathrm{rd}_E)$ (rd_E the root discriminant), the explicit size budgets $B_0(\mathfrak{R}, x, \varepsilon)$ and $N_c(\mathfrak{R}, x, \varepsilon, |T_{\mathrm{th}}|)$, and the main term $\mathrm{Main}(\omega)$. These enter the slope data through the chain ties. The precise form of the place-local functions, of the budgets B_0, N_c and of the main term we do not reproduce.

A Hida setting realises the hermitian space of rank r of a master datum as an automorphic unitary datum, with the archimedean and adelic data that the theta computations need.

Definition 6.12. Let $\mathfrak{D}$ be a master datum and let V_S be its hermitian space of rank r with base lattice L_0. A **Hida setting** for $\mathfrak{D}$ is a tuple consisting of an automorphic unitary datum, an identification of it with V_S, and adelic and archimedean data, subject to the following conditions.

(Hd1) Unitary datum. A CM extension E'/F' with an isomorphism $E' \cong E$ compatible with conjugation and mapping F' onto F; an E-basis of V_S in which the hermitian matrix J of the datum is the Gram matrix of V_S; the lattice of the datum corresponding to L_0; a bijection of the real places of F with Φ under which the archimedean embeddings agree. Moreover $n = r$, the archimedean signs and Harish-Chandra parameters are those prescribed by $\mathfrak{D}$ (through Ξ_τ), and at each real place the archimedean representation is a discrete series with a lowest K-type satisfying a "low unit" condition. The level is $g = g_1 g_2$, with distinguished prime g_1.

(Hd2) Archimedean projector. A continuous operator on the L^2 space of the fundamental domain, the operator integral of the archimedean projector built from the formal degrees of the discrete series.

(Hd3) Hyperbolic frame. Vectors $e_1, \dots, e_{r_S}, f_1, \dots, f_{r_S} \in V_S$ with $\langle e_a, e_b \rangle = \langle f_a, f_b \rangle = 0$ and $\langle e_a, f_b \rangle = \delta_{ab}$; a lattice $L_b \perp \{e_a, f_a\}$ with $L_0 = \bigoplus_a \mathfrak{a}_a e_a \oplus L_b \oplus \bigoplus_a \mathfrak{a}_a^{-1} f_a$ for fractional ideals $\mathfrak{a}_a$ of the datum; for each word ω a vector $z_\omega \in L_b$ with $\langle z_\omega, z_\omega \rangle = -T_\omega$ orthogonal to all e_a, f_a; and a flag through $e_1, \dots, e_{r_S}, z_\omega$ satisfying a fixedness hypothesis on L_0.

(Hd4) Finite adeles. An additive character $\psi_{E,f}$ of $E \otimes \mathbb{A}_{F,f}$, trivial on the integral adeles and inverse to the archimedean character on E. Vectors $e_a^0, f_a^0, z^0, b_i, b_i^\vee$ $(i \le k)$ of the adelic lattice approximate the frame and satisfy Gram congruences modulo g ($\langle z^0, z^0 \rangle \equiv -1$, $\langle b_i^\vee, b_j \rangle \equiv \delta_{ij} e_w$, and others). There is an integral idempotent e_w with $e_w \bar{e}_w = 0$, and the vectors span the adelic lattice modulo g.

(Hd5) Measures and test vectors. For each word ω: Haar measures on the unipotent and stabiliser subgroups of $\mathrm{U}(V_S)(F_\infty)$ attached to the frame and z_ω, their quotient (Weil) measure, a vector W_ω that factors as a tensor product of continuous local vectors $W_{\omega,v}$ and satisfies the characterising property of the archimedean test vector, and normalised measures on $\mathrm{U}(V_S)$ and on the frame.

The low-unit condition, the archimedean projector, the fixedness of flags, the congruences of (Hd4) and the characterising property of W_ω are conditions of the same nature, which we do not reproduce; nor some further data of a Hida setting (its faces), with their level-matching conditions.

6.3 Parametrised families and ties

We can now define tiedness, which Section 2 described in words. A parametrised host family is a host family whose members are master data with size data at the scale x, uniform in the sense of (F1)–(F4); condition (F4) is long and is given at the level of its fields. The family is chain-tied at x if one member i_0 and members $\mathrm{pt}(u)$, $u \in A$, carry zeta chain core data (Definition 4.4) built from their lattices: (T2) and (T3) tie the parameters and the entries of the slope datum to the size data and divisor lattices of the members, and (T4) ties the arithmetic datum to the host of i_0.

Definition 6.13. Let $\varepsilon > 0$. A **parametrised host family with parameter** ε is a host family $\mathcal{F} = (x_1, W, I, (\mathfrak{M}_{x,i}))$ whose members $\mathfrak{M}_{x,i}$ are master data, each with a size datum (Definition 6.11) whose scale is (x, ε, J) for $x > x_1$, with J common to the family, subject to the following conditions.

(F1) Uniformity. For fixed x all members have the same rank datum, the same primes q, d, the same Q, the same Artin numbers, the same root discriminant, the same auxiliary

primes g_1, g_2, p_1, p_2, the same thick set T_{th}, and top sets with the same image $T_{\mathrm{top}}(x)$ in $\{0, \dots, |A| - 1\}$. Their sets of words are identified compatibly.

(F2) Bounds. There is C_o with $o \le C_o$ and $Q \le C_o|A|$ for every member at every $x > x_1$, and the top numbers are bounded by an explicit function $b_{\mathrm{top}}(x)$.

(F3) Covering. For $x > x_1$, every member i_0 and every $0 \le u < |A|$ there is a member i whose shift a_1 has index u and whose top numbers agree with those of i_0.

(F4) Parameters. For $x > x_1$ each member $\mathfrak{M}_{x,i}$ carries the following.

- A list of numerical constants, common to all members and all x, some of them fixed, which we do not reproduce.
- A constant C_{Tu}, size bounds, local and Galois compatibility data, divisor data, growth data (its **growth instance**) and lattice data (its **lattice instance**) with common growth constants $\mathrm{GK} \ge 0$, and a constant C_{sz}.
- *Support.* Every place w of the coefficient field whose residue characteristic is at most $X''(\mathfrak{s}, |A|)$ lies in the support of its size datum, and so does every place w with $\mu(w) \ne \mu_0$.
- *Growth fills.* The growth parameters fill every place. If the growth set at the residue characteristic of w is empty and that residue characteristic is not the Gauss prime of ω, then the length $\ell_\omega(w)$ of the size datum is 0. The lattice instance is compatible with the Galois compatibility data and the growth instance.
- *Divisor heights.* The heights h_ω of the divisor data are given by an explicit formula in the size datum, the divisor data and C_{sz}, together with definitions of the divisor-side quantities ν, N' and R_0.
- *Residual density.* The residual density ρ_B of the growth instance is at most x^{-2}.
- *Dropped primes.* With $C_r \ge 1$, a prime p lies in the drop set of its size datum if and only if it is no Gauss prime T_ω and satisfies the **no-gain condition**. That condition holds if either of the following holds.
  - $p < X'(\mathfrak{s})(1 + C_r\rho_B)$, p lies in the residue class zero of the growth instance, and in addition (i) the order of p in the frame quotient is below $x^{\varepsilon/4}$, or (ii) p is at least the threshold k^2x^{26} and some power p^i ($i \ge 1$) is within $\rho_B X'(\mathfrak{s})$ of $X'(\mathfrak{s})$, or (iii) a further condition on p, which we do not reproduce.
  - $p > X''(\mathfrak{s}, |A|)$, or p is in neither the residue class zero nor the signal class, or p is split.
- *Uniformity.* For fixed x, the following agree across members: the thick density sizes, the cloud sizes S_c, the frame classes of primes, the base and offset data of the growth instances, and the constants κ_I and the exponent of the growth instance.

A **family over all parameters** assigns to every $0 < \varepsilon < 1/52$ a parametrised host family with parameter ε.

The constants, the growth, lattice and divisor data in (F4), the height formula and the third alternative of the no-gain condition are further conditions of the same nature, which we do not reproduce.

Let $\mathcal{F}$ be a parametrised host family, $x \in \mathbb{R}$ and $e_D, e_L, C_K, c^*, e_K, C_L \in \mathbb{R}$. The family is **chain-tied at x with exponents (e_D, e_L) and constants (C_K, c^*, e_K, C_L)** if there

exist a member $i_0 \in I(x)$, whose rank datum has group A, a map $\mathrm{pt}\colon A \to I(x)$, integer sequences $(a_i^{\mathrm{top}}), (b_i^{\mathrm{top}})$, and zeta chain core data at scale x with exponents (e_D, e_L) and constants (C_K, c^*, e_K, C_L) on the group $H = A$, whose slope datum has index set A with word sets $\mathcal{W}(u)$ those of $\mathfrak{M}_{x,\mathrm{pt}(u)}$, such that:

(T1) Points. For every $u \in A$ the member $\mathfrak{M}_{x,\mathrm{pt}(u)}$ has the frame, m, f, q and d of $\mathfrak{M}_{x,i_0}$, its shift a_1 has the index of u, and its top numbers are $(a_i^{\mathrm{top}}), (b_i^{\mathrm{top}})$.

(T2) Global ties. For every u the parameters $a, b, b', L, L_a, B_0, N_c$ of the slope datum are those of the size datum of $\mathfrak{M}_{x,\mathrm{pt}(u)}$, k is its rank parameter k, $\gamma_0 = \gamma$, β is the log-derivative constant of the size datum, v is half the root discriminant of its host field, $l_g = \log(g_1 g_2)$, and X'', ν, X' are the corresponding scales of the size datum.

(T3) Local ties. For every u and word ω, the rank $r_{u,\omega}$, the slopes $\mu_{u,\omega}, \mu_{u,\omega}^{\max}$ and the height $h_{u,\omega}$ are those of the divisor lattice of ω in $\mathfrak{M}_{x,\mathrm{pt}(u)}$; $T_{u,\omega}$ is its Gauss prime; $\mu'_{u,\omega}, \mu'^{\max}_{u,\omega}$ are the slopes of its level-zero hermitian lattice; the $E_{u,\omega,j}$ are sums of local sizes over the support; $\mu_u^\vee, \mu_{\min,u}^\vee$ are the dual slopes of the reference lattice; and $F_u, \Theta_u, I_u, \delta_u$ are the corresponding quantities of the size and rows data.

(T4) Arithmetic ties. The arithmetic datum is the datum of the host of $\mathfrak{M}_{x,i_0}$. The set T consists of the $h \in A$ whose index lies in $T_{\mathrm{top}}(x)$, $T_2 = \{s\, a_0 : s \in T_{\mathrm{th}}\}$, and $\tilde{a}(h) = a^{\mathrm{top}}_{\mathrm{index}(h)}$. The weight $w_{u,j}(g)$ is the j-th weight of $\mathfrak{M}_{x,\mathrm{pt}(u)}$ at the CM-type place with the same sign-and-index code as g. Finally, $\eta(\phi) = L'(1,\chi)/L(1,\chi)$ for every non-trivial character ϕ of Γ_A and every primitive Dirichlet character χ with $\chi(n) = \phi_{\mathfrak{A}}(n)$ for all $n \geq 1$.

A parametrised host family $\mathcal{F}$ with windows W is **tied with exponents** (e_D, e_L) if there are C_K, c^*, e_K, C_L, x_0 such that $\mathcal{F}$ is chain-tied at every $x \in W$ with $x \geq x_0$, with exponents (e_D, e_L) and constants (C_K, c^*, e_K, C_L).

Unwinding the definitions, a chain-tied family has a core-tied member, which is what the bridge needs.

Lemma 6.14. *Let $\varepsilon > 0$, let $\mathcal{F}$ be a parametrised host family with parameter ε, and let $x, e_D, e_L, C_K, c^*, e_K, C_L \in \mathbb{R}$. If $\mathcal{F}$ is chain-tied at x with exponents (e_D, e_L) and constants (C_K, c^*, e_K, C_L), then there is $i \in I(x)$ such that $(\mathfrak{H}_{x,i}, T_{x,i})$ is core-tied at scale x with exponents $(e_D, e_L, C_K, c^*, e_K)$ and parameter ε.*

Proof. Take $i = i_0$, the member of the chain ties. By (T4), the set T of the zeta chain core data is the image of the top set of $\mathfrak{M}_{x,i_0}$, that is $T = T_{x,i_0}$, and the arithmetic datum is that of the host $\mathfrak{H}_{x,i_0}$. The slope datum of the chain ties has index set A; take tr to be the identity of A, whose image contains A_0. Conditions (K1) and (K2) are (T4) and the identification (Z2). (K3) is (Z1) with the error conditions (Z3), and (K4) collects (Z4)–(Z7). All are restricted to the subgroup A_0 where (K4) requires it. □

The constants of a family are fixed by a closure argument: a property is closed for a class of families if it holds whenever the constants of the family are large in a prescribed order. This is the precise form of the instruction to choose the constants large enough, in the right order.

Definition 6.15. A **family class** $\mathcal{C}$ is a property of parametrised host families (for each parameter ε). A **threshold datum** is a tuple $\mathfrak{K} = (k_{\mathrm{pre}}, t, k_{\mathrm{mid}}, k_{\mathrm{last}})$ of a real number k_{pre} and three functions from finite lists of reals to $\mathbb{R}$.

Each parametrised host family $\mathcal{F}$ carries three lists of its constants: a list $c_{\mathrm{pre}}(\mathcal{F})$; a list $c_{\mathrm{mid}}(\mathcal{F}, C_H)$, depending on a further real C_H; and a list $c_{\mathrm{last}}(\mathcal{F})$; together with an output list $c_{\mathrm{out}}(\mathcal{F})$. The family is **large for** $(\mathfrak{K}, C_H)$ if every entry of $c_{\mathrm{pre}}(\mathcal{F})$ is at least k_{pre}, every entry of $c_{\mathrm{mid}}(\mathcal{F}, C_H)$ is at least $k_{\mathrm{mid}}(c_{\mathrm{out}}(\mathcal{F}))$, and every entry of $c_{\mathrm{last}}(\mathcal{F})$ is at least $k_{\mathrm{last}}(c_{\mathrm{out}}(\mathcal{F})\, c_{\mathrm{mid}}(\mathcal{F}, C_H))$ (concatenated lists).

A property $\mathcal{Q}(\mathcal{F}, C_H)$ of parametrised host families and reals is **closed for** $\mathcal{C}$ if there is $\varepsilon_1 > 0$ with the following property: for every $0 < \varepsilon \le \varepsilon_1$ there is a threshold datum $\mathfrak{K}$ such that $\mathcal{Q}(\mathcal{F}, C_H)$ holds for every family $\mathcal{F}$ of parameter ε in $\mathcal{C}$ and every C_H for which $\mathcal{F}$ is large for $(\mathfrak{K}, C_H)$. A property $P(x, i)$ of the members of a family holds **eventually** on $\mathcal{F}$ if there is x_0 with $P(x, i)$ for all $x \ge x_0$ and $i \in I(x)$. Which constants of a family form the three lists is fixed once and for all; we do not reproduce the lists.

6.4 The definite-side family structure

The definite-side family structure is a selection of auxiliary data for the members of a class of host families, subject to about forty conditions, together with the linking property. We first state the standard family class and its four row conditions. Two of them are proved later: the covolume condition (COV) in Section 8, at the compatibly ruled points of the definition, and the toric functoriality condition (FUN), a bound for the maximal slope of a lattice built from the theta expansions of the members, in Section 7. The surjectivity condition (ON) and the lower bound (LOW) are stated; where they are established the outline does not show. The fourteen conditions of the standard class are described, not stated.

Definition 6.16. Let $\mathcal{F}$ be a host family whose members $\mathfrak{M}_{x,i}$ are master data. A **point** of $\mathcal{F}$ at (x, i), $x > x_1$, is a choice of the auxiliary data at the member $\mathfrak{M}_{x,i}$ (an extension datum, a coefficient level and the attached size constants), to which are attached two **row data** $\mathfrak{S}^-$, $\mathfrak{S}^+$ and, for real C and C_H, a **number datum** $\mathfrak{N}(C, C_H)$; a point is **ruled** if its row conditions hold together with the **tie conditions** of the point (a realisation condition in a first form and core, pull-back and type ties between the point and its member). The reader should think of a row datum as a finite list of Hermitian lattices attached to the point, and of a number datum as explicit size bounds, depending on the scale x, that these lattices are required to satisfy. Precisely, a row datum $\mathfrak{S}$ contains, over a CM number field $K_{\mathfrak{S}}$, an integer $D \ge 1$, Hermitian metrics X_E on $K_{\mathfrak{S}}^D$ with an $\mathcal{O}_{K_{\mathfrak{S}}}$-lattice R glued from local lattices at the finite places, and for each index j an $\mathcal{O}_{K_{\mathfrak{S}}}$-lattice M_j in a $K_{\mathfrak{S}}$-vector space of dimension $d_W(j)$ with Hermitian metrics $X_W(j)$, together with archimedean weights $\kappa(v) > 0$ and positive archimedean values $\varrho_j(\mu, p, v)$ indexed by types μ (with a base type μ_0), indices p and archimedean places v. The number datum $\mathfrak{N}(C, C_H)$ is formed from size functions $l_3^{\mathrm{m}}(j), l_3^{\mathrm{u}}, l_6^{\mathrm{m}}(j), l_6^{\mathrm{u}}$ of the scale and a parameter b by $l_3(j) = l_3^{\mathrm{m}}(j) + C\, l_3^{\mathrm{u}}$, $l_6(j) = l_6^{\mathrm{m}}(j) - C\, l_6^{\mathrm{u}}$ and $h_h = x^{C_H} b$. The construction of the lattices M_j, R, of the values ϱ_j and of the size functions from the member is part of the data that we do not reproduce. The **row conditions** are

(COV) the covolume condition: for every j,

$$\frac{1}{D} \cdot \frac{2}{[K_{\mathfrak{S}} : \mathbb{Q}]} \log \mathrm{covol}\big(M_j, X_W(j)\big) \;\le\; l_3(j),$$

where covol is the Hermitian covolume (height) of the lattice in its metrics;

(ON) the surjectivity condition: bounds on the local indices of the lattices M_j at the twisted places in terms of the ramification and error data of $\mathfrak{N}$, which we do not reproduce;

(LOW) the lower bound: $l_6(j) \le \log \varrho_j(\mu_0, p, v)$ for all j, p and archimedean v (for $\mathfrak{S}^+$);

(FUN) the toric functoriality condition: $\widehat{\mu}_{\max}(R, X_E) \le h_h - \frac{1}{n_\infty} \sum_v \log \kappa(v)$, with $\widehat{\mu}_{\max}$ the maximal slope and n_∞ the number of archimedean places.

A point satisfies the **compatible point conditions** if its row conditions hold together with the following **compatibility conditions** in place of the tie conditions: the realisation condition in a second, recursive form; the same core, pull-back and type ties; ties for the Eisenstein series, the level and the data of the point; integrality of the twisted characters; and conditions on ray class fields and roots of unity in the coefficient field. A property Q of points holds **eventually on** $\mathcal{F}$ at a given class of points if there is x_0 such that Q holds at every point of that class at (x, i) with $x \ge x_0$. The tie and compatibility conditions are conditions of the same nature as the row conditions, which we do not reproduce; neither kind of point is known to be a special case of the other.

The **standard family class** is the class of host families $\mathcal{F}$ that satisfy fourteen further conditions, which we do not reproduce. They fix the numerical constants of the family; they place the rank parameters of the members in ranges given by powers of the scale (one of them lies between $x^{2+8\varepsilon}$ and $2x^{2+8\varepsilon}$); they constrain the moduli, the discriminant-type invariants and the twists of the members, and require in particular that these invariants are eventually divisible by a Minkowski-type bound $\mathrm{mink}(2fN_m)$; they bound two size parameters; and they require a choice of auxiliary constants and sets of primes, with constant 3 and some real C_H, which we call an **exceptional choice** for $\mathcal{F}$ with constant C_H. A property of pairs $(\mathcal{F}, C_H)$ is **closed on the standard class** if it is closed for this class in the sense of Definition 6.15.

A point is **compatibly ruled** if it satisfies the compatible point conditions and its member satisfies two further conditions, a size bound for its field and a compatibility condition for Eisenstein series, which we do not reproduce. The **covolume property** is the statement that the following property of $(\mathcal{F}, C_H)$ is closed on the standard class: there are $C_0 \ge 0$ and x_0 such that (COV) holds for $(\mathfrak{S}^-, \mathfrak{N}(C, C_H))$ for all $C \ge C_0$ at every compatibly ruled point at (x, i) with $x \ge x_0$.

The definition describes the conditions of the structure in one sentence and gives in full the linking property, which is what Proposition 6.20 uses.

Definition 6.17. Let $\mathcal{N}$ be a property of classes of host families. A **definite-side family structure for** $\mathcal{N}$ is a tuple $\mathfrak{Y} = (\mathcal{A}, \sigma, c_T, \dots)$ consisting of a class $\mathcal{A}$ of host families, a selection σ attaching to every family $\mathcal{F} \in \mathcal{A}$, every $x > x_1$ beyond a threshold and every member $\mathfrak{M}_{x,i}$ a **selected datum** (a base level, a realisation of the member by indefinite hermitian forms, a choice of kernel, and row data as in Definition 6.16), a size function c_T, and about forty further conditions on these data, which we do not reproduce. They require that $\mathcal{A}$ contain families with size function c_T whenever the discriminant-type invariants of the members are divisible by the Minkowski-type bound of Definition 6.16; they bound the row data of the selected data and constrain their twists and periods and their compatibility with the points of the families; and they include the **definite-side condition**: for each selected datum, covolume conditions and further conditions of the same kind at its indefinite realisation imply the row conditions of its column data. The structure is **closed under** $\mathcal{N}$ when $\mathcal{N}(\mathcal{C})$ holds for the class $\mathcal{C}$ of its families. The property used below is the **linking property**: there are C and x_0 such that for every family in the class, every $x \ge x_0$, every member $\mathfrak{M}$ at x with field E and rank t for which $\mathrm{len}_\ell(E, k) < \infty$ for all ℓ and $1 \le k < t$, and every lift index j,

$$\eta_j(\mathfrak{M}) \le r_j(\mathfrak{M}) \Big(C\big(fmbL^2 + (\log x)^2 N_c\big) + \sum_\ell \log \ell \sum_{k=1}^{t-1} \mathrm{len}_\ell(E, k) \Big),$$

where $\eta_j(\mathfrak{M})$ is the sum, over the places w of the member's support at which the local slope differs from the generic one, of the ramification index of w times the j-th local defect, and r_j, f, m, b, L, N_c are size parameters of $\mathfrak{M}$.

Its existence is the main theorem of the construction. The proof assembles seven steps, and it is where the theta-lift results of Section 7, the covolume bound of Section 8 and the known theorems named at the beginning of this section meet; the input on primes in short intervals at very large scales comes from classical zero-density estimates [31], [61].

Proposition 6.18. *There is a definite-side family structure that is closed under the linking property.*

Proof. In outline. The known theorems used are the local Langlands correspondence rec for GL_n [50], [55]; the Galois representations attached to polarized regular algebraic cuspidal representations over CM fields [50], [114], [27], [22]; the distribution of primes in short intervals, from the classical zero-density estimates, zero-free regions and truncated explicit formula for Dirichlet L-functions [31], [61]; the mass formula for definite hermitian lattices [113]; and the further known theorems on theta correspondences, Siegel–Weil formulas and heights listed in the appendix. Two further ingredients are a choice of auxiliary lattices and measures by a counting argument based on the Siegel–Weil volume identity, and the integrality of theta-kernel coefficients at inert primes (Proposition 6.19). The structure is assembled in seven steps.

1. Separation. A family of functionals attached to the levels of the members is separating, by Howe duality at the unramified places and the convergent Siegel–Weil formula.

2. Rationality of rows. A choice of test vectors shows that the rows of the selected data are rational. It uses Siegel sets, real Howe duality, Green's parametrisation of the cuspidal characters of GL_n over a finite field [46], functional equations, and the theta agreement property with local matching (Proposition 7.3).

3. Toric functoriality. The row condition (FUN) holds eventually on the families (Theorem 7.6), whose separation hypothesis is supplied by step 1; the statements on CM values that this needs are supplied by Theorem 7.5, and theta agreement by Proposition 7.3.

4. Evaluation grid. The twisted Weil representation converges and the twisted local factors match; from this an evaluation grid for the row functionals exists at large scale. It is built from Weil–Deligne parameters attached to the local data of the members, the cuspidality of theta lifts along the grid, and a normalisation of the characters over a number field.

5. Openness. Certain conditions on the local parameters attached to the members by rec, depending on a constant c_{dep}, define open sets and are preserved when the auxiliary extension is enlarged; we do not reproduce them.

6. Covolume. The covolume condition (COV) holds at compatibly ruled points (Theorem 8.1); this gives the linking of the selected data at level three.

7. Large unramified exponents. For every c_{dep}, members with large unramified exponents, satisfying the conditions of the construction, exist. They come from compatible systems of Galois representations attached to the members (global Artin reciprocity, good reduction of elliptic curves with complex multiplication, ray class fields of imaginary quadratic fields); the convergent Siegel–Weil formula, compared in two adjacent ranks, gives a ratio of masses of standard hermitian lattices. A linking statement at levels two and three, proved with

Chebotarev's density theorem, local Tate duality [123] and the local Euler characteristic formula, then yields the structure together with its linking property.

□

One step of this proof rests on an integrality statement: at the primes $p \geq 5$ inert in K, the Taylor coefficients of certain theta kernels, written in the logarithmic coordinates of an elliptic curve with complex multiplication by $\mathcal{O}_K$, are p-integral. The admissibility conditions and the theta-kernel combinations to which it applies are not reproduced.

Proposition 6.19. *Let $\mathfrak{D}$ be a master datum, with host field $E \supset K = \mathbb{Q}(\sqrt{-d})$ and CM-type places Φ as in Definition 6.2 and Definition 6.10, $\mathfrak{c}$ an ideal of $\mathcal{O}_K$ that is the period lattice ideal of an elliptic curve with complex multiplication by $\mathcal{O}_K$, $\delta \in \mathbb{C}$ a twelfth root of the discriminant of the corresponding ideal lattice, and χ an admissible character of the ideles of E (admissibility, like the admissibility of lattices and the notion of theta-kernel combination below, consists of conditions that we do not reproduce). Let $p \geq 5$ be a prime inert in K (that is, $(-d/p) = -1$), $\lambda\colon \overline{\mathbb{Q}} \to \mathbb{C}_p$ an embedding, and $L \subseteq E^N$, $L' \subseteq E^{N'}$ admitted $\mathcal{O}_E$-lattices. Then for every power series F in the variables indexed by $\Phi \times \{1, \dots, N'\} \times \{1, \dots, N\}$ over $\mathbb{C}_p$ that is a theta-kernel combination for these data, every change of variables to a model of the curve over $\mathbb{C}_p$ with its formal logarithm series, and every orthonormal local basis of the lattice pair at p, the expansion of F in logarithmic coordinates, multiplied by the exponential factor of the model, has p-integral coefficients.*

Proof. In outline. Weak approximation at the places of E, together with the existence of local orthonormal bases, reduces the claim to an approximating basis, for which integrality is proved directly.

□

6.5 Ties from the Sha bounds

The next proposition joins the family construction to the Sha bounds. Given a supply of hosts on windows, as in Proposition 3.5, and the Sha bound for the fields of the members, it produces tied families. Its proof combines the definite-side family structure of Proposition 6.18 with a general construction, which the development calls a shape theorem, and with a closure statement, which says that a condition linking the lattices of the members to their growth data holds eventually on the families that arise; we describe the latter only in words. The Sha bound enters through the linking property at the end of Definition 6.17: the local defects of the lattices of a member are bounded by r_j times the sum of $C(fmbL^2 + (\log x)^2 N_c)$ and the weighted total length of its Tate–Shafarevich groups, and the Sha bound $\mathrm{V}(E; t, \sqrt{b})$ makes the latter at most $\sqrt{b}$.

Proposition 6.20. *Let $\lambda > 1$ and let P be a requirement on hosts at a scale such that, for every $\varepsilon \in (0, 1/52)$, good hosts satisfying P exist at all scales of a set containing λ-windows, as in (P1) and (P4) of Proposition 3.5. Let $\mathcal{C}_P$ be the class of host families whose windows contain λ-windows and whose members are such hosts. Suppose that for every family in $\mathcal{C}_P$ there is x_0 such that every member at $x \geq x_0$, with field E, rank t and size parameter b, satisfies $\mathrm{V}(E; t, \sqrt{b})$. Then there are e_D, e_L and $\varepsilon_0 > 0$ such that for every $\varepsilon \in (0, \varepsilon_0)$ the class $\mathcal{C}_P$ contains a family with parameter ε that is tied with exponents (e_D, e_L).*

Proof. In outline. Three kinds of statements are combined. The first is the definite-side family structure of Proposition 6.18, which is closed under the linking property of Definition 6.17. The second is a general construction, which uses Green's parametrisation of the cuspidal characters of

GL_n over a finite field [46]: from such a structure it produces tied families for every requirement P for which good hosts satisfying P exist on λ-windows. The third is a closure statement: a condition linking the lattices of the members to their growth data holds eventually on the families that arise, as a consequence of a group-compatibility condition on the Galois data of the members. The mechanism by which the hypothesis enters is the linking inequality: for a member with field E satisfying $\mathrm{V}(E; t, \sqrt{b})$, the term $\sum_\ell \log \ell \sum_{k<t} \mathrm{len}_\ell(E, k) \le \sqrt{b}$ bounds the local defects η_j of its lattices. This is the point at which the Tate–Shafarevich bounds enter the slope data. □

The family structure rests on the four row conditions. Of these, (COV) and (FUN) are the two whose proofs the outline shows: the next section outlines the proof of (FUN), together with the theta agreement on which the structure also rests, and Section 8 that of (COV) at compatibly ruled points. Where (ON) and (LOW) are established the outline does not show, and a second form of (LOW) is a hypothesis of Theorem 7.6 (Section 11).

7 Theta agreement, toric functoriality and critical L-values

This section supplies three inputs of the definite-side family structure. The first is theta agreement (Proposition 7.3): for two dual pairs sharing a skew-hermitian space, theta integrals with matched local data are proportional with a nonzero constant, and along each step of a tower of such configurations the ratio of the constants is a partial Hecke L-value times an epsilon factor, up to a factor fixed by the step, which does not depend on the characters or on the finite local data. The second is a group of statements on special values at CM points, the CM value statements of Definition 7.4, which hold by Theorem 7.5; they give algebraicity, Galois transport and denominator bounds for partial zeta values and for Eisenstein-type lattice sums at CM lattices. The third is toric functoriality (Theorem 7.6), the row condition (FUN) of Definition 6.16, which the development calls the constant-function fit. Its engine is an arithmetic reading of theta constants along a grid of hermitian spaces: the ratio of the square constants of two members, multiplied by a product of ratios of finite Hecke L-values of the host field, is an algebraic number with controlled prime factors, in the shape of the Rallis inner product formula for unitary groups [80], [49], which goes back to [101], and the CM value statements control these products prime by prime (Theorem 7.9). The section ends with the evaluation grids that the family structure also needs. The known theorems used are the Siegel–Weil formula for unitary groups [59], [49], in the form of the method of [133] and [73], the continuation of Siegel Eisenstein series [78], [122], Howe duality over p-adic fields [86], [42] and over the real numbers [58], [64], the main theorem of complex multiplication [112], the rationality and congruences of partial zeta values [116], [69], [34], and the functional equation of Hecke L-functions [124], [53]. What looks new is the quantitative control: every bound is explicit in the scale x. At the archimedean places the construction also needs discrete series of $\mathrm{U}(p, q)$ whose Harish-Chandra parameters are far from the walls; the argument obtains them, with explicit thresholds, from a spherical function of the lowest K-type whose decay is explicit in the distance to the walls, while their existence is classical [48], [103].

7.1 Theta agreement

A theta configuration consists of two dual pairs $(\mathrm{U}(V), \mathrm{U}(W))$ and $(\mathrm{U}(V'), \mathrm{U}(W))$ sharing the skew-hermitian space W, with lattices and models of their Weil representations; admissible triples add the characters and local data for which the theta integrals of Section 5 are compared.

Definition 7.1. A **theta configuration** of size (a, a') over a CM extension E/F is a tuple $\mathfrak{G} = (p, q, V, V', W, \mathcal{M}, \mathcal{M}')$ of integers $p, q \geq 0$, hermitian spaces V, V' of dimensions a, a', a skew-hermitian space W of dimension $a + a'$, each with an $\mathcal{O}_E$-lattice, archimedean Cartan involutions and split bases at a finite set P of finite places, and models $\mathcal{M}$, $\mathcal{M}'$ of the Weil representations of the dual pairs $(\mathrm{U}(V), \mathrm{U}(W))$ and $(\mathrm{U}(V'), \mathrm{U}(W))$, each with a vacuum vector at every finite place, a Haar measure and a fundamental domain for the rational points.

Character data are $\mathfrak{E} = (\alpha, \beta, S_0, Q)$ with α, β Hecke characters of E and S_0, Q finite sets of finite places of F; put $\nu = \alpha/\beta$ and $T = S_0 \cup Q$.

A triple $(\mathfrak{G}, \mathfrak{E}, (\omega, \omega'))$ with a base datum (additive character ψ_{add}, archimedean embeddings, integers $m_\alpha(w), m_\beta(w)$ for the real places w, the set P and a distinguished place v_1) is **admissible** if:

(At1) E/F is CM, $E/\mathbb{Q}$ is Galois, E/F is unramified at all finite places, ψ_{add} is standard, and every finite place of F inert in E does not lie over 2, has norm > 3 and is unramified for ψ_{add}.

(At2) $a' < a$, $a + a'$ is even, $a + a' < m_\alpha(w) - m_\beta(w)$ for every real place w, and $2(p+q)+1 < \min\bigl(a', (a-a')/2\bigr)$.

(At3) At every real place, V has signature $(a-p, p)$, V' has signature $(q, a'-q)$ and W has signature $(a+a'-p-q,\ p+q)$; the lattices are self-dual and the models have rational Darboux coordinates, vacuum vectors equal to the characteristic function of the standard lattice at almost all places, and the given Haar measures and fundamental domains.

(At4) α and β are Hecke characters with $\alpha|_{\mathbb{A}_F^\times} = \eta^{a'}$ and $\beta|_{\mathbb{A}_F^\times} = \eta^{a}$, with archimedean exponents m_α and m_β; $S_0 \cap Q = \varnothing$, $T \subseteq P$; ν is ramified at every place of E above T; α and β are unramified outside T; and $v_1 \notin T$ splits completely in E.

(At5) ω is the Weil representation of $\mathrm{U}(V) \times \mathrm{U}(W)$ on $\mathcal{M}$ with characters $(\beta, \beta/\alpha)$, and ω' that of $\mathrm{U}(V') \times \mathrm{U}(W)$ on $\mathcal{M}'$ with characters $(\alpha, \alpha/\beta)$.

For such a triple, the **polynomial data** $\mathcal{S}_0(\mathcal{M})$ is the span of the factorisable functions $f \otimes \bigotimes_v \varphi_v$ with f in the space of polynomial multiples of the archimedean Gaussian (the smallest space containing the Gaussian vectors and stable under the archimedean Heisenberg one-parameter groups), φ_v Schwartz–Bruhat and φ_v equal to the vacuum vector for almost all v; similarly $\mathcal{S}_0(\mathcal{M}')$. The subspace $\mathcal{S}^\circ(\mathcal{M}')$ consists of the spans of such functions whose component at v_1, in a linear model at v_1, is supported on matrices of full rank a'. The lattices, Darboux coordinates and fundamental domains in (At3) are subject to further compatibility conditions of the same nature, which we do not reproduce.

Matched data identify the local coinvariants of the two Weil representations, and tower steps of types A and B lower the dimension of one of the spaces by two. The theta agreement property (A1)–(A3) says that matched theta integrals are proportional, and that along a step the two constants differ by a partial Hecke L-value $L^{S_0}(z, \nu)$ times an epsilon factor ϵ_T, up to one constant $C \neq 0$ that depends only on the step and on the archimedean matching maps; it does not depend on the characters, on the Weil representations or on the finite matching.

Definition 7.2. Matched data. An admissible triple is **matched** by $(R, (R_v)_v)$ if R is a pair of linear maps p_∞, p'_∞ from the archimedean polynomial-Gaussian spaces of $\mathcal{M}$ and $\mathcal{M}'$ onto a nonzero space Π_∞, and R_v for each finite place v is a pair p_v, p'_v from the local Schwartz–Bruhat spaces onto a nonzero space Π_v, such that:

(M1) p_∞, p'_∞ kill the derivatives along $\mathrm{U}(V)_\infty$, resp. $\mathrm{U}(V')_\infty$, and are invariant under their maximal compact subgroups; p_v, p'_v kill the $\mathrm{U}(V)(F_v)$-, resp. $\mathrm{U}(V')(F_v)$-coinvariant relations;

(M2) the relation $p(\varphi) = p'(\varphi')$ is preserved by the action of $\mathrm{U}(W)$ at every place, and Π_∞ and each Π_v are irreducible under this action;

(M3) for $v \notin T$ the vacuum vectors satisfy $p_v(\varphi_v^0) = p'_v(\varphi_v'^0) \neq 0$, and for $v \in T$ the relation is given in linear models by an explicit identity of local zeta-type integrals with the characters ν, β, α at v, which we do not reproduce.

Two global data $\varphi \in \mathcal{S}_0(\mathcal{M})$, $\varphi' \in \mathcal{S}_0(\mathcal{M}')$ **match**, written $\varphi \leftrightarrow \varphi'$, if for some finite $T' \supseteq T$ they are finite sums of factorisable vectors, vacuum outside T', whose images in $\Pi_\infty \otimes \bigotimes_{v \in T'} \Pi_v$ coincide.

Tower steps. An **adapted plane** for a hermitian or skew-hermitian space Y of dimension n_1+2 and a space Y_1 of dimension n_1 is a hyperbolic pair $x, y \in Y$ ($\langle x,x\rangle = \langle y,y\rangle = 0$, $\langle x,y\rangle = 1$) with an isometric embedding $Y_1 \cong \langle x,y\rangle^\perp$ compatible with the lattices, Cartan involutions and split bases. For admissible configurations $\mathfrak{G}$ of size $(a, b+2)$ and $\mathfrak{G}_1$ of size (a,b), a **step of type B** consists of adapted planes $H \subset V'$ and $H' \subset W$ with $V_1 = V$, $p_1 = p$ and $q_1 + 1 = q$; for configurations of sizes $(b+2, a')$ and (b, a'), a **step of type A** consists of adapted planes $H \subset V$, $H' \subset W$ with $V_1' = V'$, $q = q_1 = 0$ and $p_1 + 1 = p$.

Evaluation maps. Let $\mathcal{M}$ be a model for $(\mathrm{U}(Y), \mathrm{U}(Z))$ and $\mathcal{M}_1$ one for $(\mathrm{U}(Y_1), \mathrm{U}(Z_1))$, with isometric embeddings $\iota : Y_1 \to Y$ and $\iota' : Z_1 \to Z$ (either may be the identity), and let $\mathcal{I}$ be a set of Heisenberg translations. A linear map e from the Schwartz space of $\mathcal{M}$ to that of $\mathcal{M}_1$ is an **evaluation map along** (ι, ι') **invariant under** $\mathcal{I}$ if it preserves Schwartz spaces, intertwines the Heisenberg action of z on $\mathcal{M}_1$ with that of $\iota z \iota'^{\top}$ on $\mathcal{M}$, is invariant under the translations in $\mathcal{I}$, and is a restricted tensor product of a nonzero continuous archimedean map with local maps carrying Schwartz–Bruhat spaces onto Schwartz–Bruhat spaces and vacuum vectors to vacuum vectors. For an adapted plane $H' = (x,y)$ in Z the relevant sets are $\mathcal{I}_0 = \{u \otimes x\}$ (type zero) and, for adapted planes $H = (x_0, y_0)$ in Y and H' in Z, $\mathcal{I}_A = \mathcal{I}_0 \cup \{x_0 \otimes \iota' z\}$ and $\mathcal{I}_B = \{x_0 \otimes z\} \cup \{\iota u \otimes x\}$.

Cone operators. Let $K_Y \subset \mathrm{U}(Y)(\mathbb{A}_F)$ be the **standard compact subgroup**: the elements commuting with the archimedean Cartan involutions and preserving the finite-adelic completion of the lattice of Y; fix Haar measures dk on K_Y and $d^\times l$ on $\mathbb{A}_E^\times$. For an evaluation map e, a Hecke character ν, an integer z and a map $l \mapsto u(l)$ into Heisenberg translations, the **cone operator** is

$$\mathrm{cone}_{e,u,\nu,z}(\varphi) = \int_{K_Y} \int_{\mathbb{A}_E^\times} e\big(\rho(u(l))\,\omega(k,1)\varphi\big)\,\nu(l)\,|l|^z\,d^\times l\,dk,$$

with ρ the Heisenberg action. For adapted planes $H = (x_0, y_0)$ in Y and $H' = (x,y)$ in Z the shifts are $u_A(l) = l\,x_0 \otimes y$ (type A) and $u_B(l) = \bar{l}\,y_0 \otimes x$ (type B).

Constant-term splitting. For an adapted plane $H' = (x,y)$ in W with complement W_1, functions f_α, f_β on $\mathrm{U}(W)(\mathbb{A}_F)$ **split** f if $f_P = f_\alpha + f_\beta$ for the parabolic P stabilising Ex, and for every Levi element m acting by $t \in \mathbb{A}_E^\times$ on x, by $\bar{t}^{-1}$ on y and trivially on W_1,

$$f_\alpha(mh) = \alpha(t)|t|^{a'/2} f_\alpha(h), \qquad f_\beta(mh) = \beta(t)|t|^{a/2} f_\beta(h).$$

Theta agreement. Local matching is the statement that every admissible triple is matched by some $(R, (R_v))$. The **theta agreement property** is the conjunction of:

(A1) for every admissible triple the theta integrals converge with moderate growth, and for every matching there is a unique $c = c(\mathfrak{G}, \mathfrak{E}, R) \in \mathbb{C}$ with $\Theta(\varphi) = c\,\Theta'(\varphi')$ for all matching $\varphi \leftrightarrow \varphi'$ (with $\varphi' \in \mathcal{S}^{\circ}(\mathcal{M}')$, limits of matching pairs allowed), this c is nonzero, some $\Theta'(\varphi')$ is nonzero, and the local matching maps identify the coinvariants;

(A2) there are units $e_1, e_2, s \in \{\pm 1\}$, fixed once and for all, such that for every step of type B and every choice of archimedean matching maps R for $\mathfrak{G}$ and R_1 for $\mathfrak{G}_1$ there is $C \neq 0$, depending only on the step and on R, R_1, with the following property: for all character data, Weil representations and finite matching maps making both triples matched, the constants c of $\mathfrak{G}$ and c_1 of $\mathfrak{G}_1$ satisfy

$$c_1 = C\,\epsilon_T(1 - z_B)\,L^{S_0}(z_B, \nu)\,c;$$

(A3) likewise, with units e_1, e_2, s fixed once and for all, for every step of type A and archimedean matching maps R, R_1 there is $C \neq 0$, depending only on the step and on R, R_1, such that for all character data, Weil representations and finite matching maps

$$c = C\,\epsilon_T(1 - z_A)\,L^{S_0}(z_A, \nu)\,c_1.$$

Thus the constant C does not depend on the characters α, β or on the sets S_0, Q: as the character data vary, the ratio of the constants of neighbouring members of the tower is a partial Hecke L-value times the factor ϵ_T, up to a factor fixed by the geometry of the step (including the base datum). Here $\epsilon_T(z) = \nu(\mathfrak{b})^{e_1}\nu(\mathfrak{b}')^{e_2}\prod_{v\in T}\nu_v(s)\,\epsilon(\nu_v, \psi_{\mathrm{add},v}, z)$ for the ideals $\mathfrak{b}, \mathfrak{b}'$ of the planes H, H', with local epsilon factors ϵ, and $L^{S_0}(s, \nu)$ is the partial Hecke L-function of ν with the places above S_0 removed. Here $z_B = \lfloor (a - b - 2)/2 \rfloor + 1$ for a step of type B of sizes $(a, b+2)$, (a, b), and $z_A = \lfloor (b + 2 - a')/2 \rfloor$ for a step of type A of sizes $(b+2, a')$, (b, a'). The convergence and coinvariant clauses of (A1) are conditions of the same nature as those stated, which we do not reproduce in full.

Both properties are proved together, from the Siegel–Weil formula and the continuation of Siegel Eisenstein series on the doubled unitary group, with local input on the Weil representation.

Proposition 7.3. *Local matching holds, and the theta agreement property (A1)–(A3) holds.*

Proof. In outline. Both statements are proved together. The known theorems used are the meromorphic continuation of Siegel Eisenstein series on doubled unitary groups [78], [122] and the Siegel–Weil formula for a definite space [133], [59], [49], in the unitary form of the method of [73], together with the following local and global facts on the Weil representation:

- a description of the modules of spherical vectors for the lattices of the configuration;
- finiteness and multiplicity one for coinvariants under maximal compact and orthogonal subgroups, and multiplicity one at ramified places;
- finite generation at ramified and split places, and an identity for spherical vectors at inert places coming from unramified Howe duality [86], [42];
- local lifts with their Lie-algebra derivative conditions at the archimedean places;
- an identity for the constant term of the theta integral and the existence of Weil representation data.

□

7.2 Special values at CM points

The second ingredient concerns special values at CM points: five statements which together give algebraicity, Galois transport and denominator bounds for the values that occur in the theta constants.

Definition 7.4. The following five statements concern partial zeta values of totally real fields and Eisenstein-type lattice sums at CM lattices; we call them the **CM value statements**. For a number field F of degree n, an ideal class $\mathfrak{c}$ of F and an integer $k \geq 1$ write $\zeta_F(\mathfrak{c}, s) = \sum_{I \in \mathfrak{c}} \mathrm{N} I^{-s}$ for the partial zeta function (continued meromorphically), and for a prime ℓ and an integer $N \geq 1$ put $m_\ell(N) = \min\{v_\ell(a^N - 1) : a \geq 2,\ \ell \nmid a\}$.

(ZD) Zeta denominators. Let E be a CM field with maximal totally real subfield F, $n = [F : \mathbb{Q}]$, and $s \geq 1$; put $k = 2s$. Then $z(\mathfrak{c}) = \zeta_F(\mathfrak{c}, 1 - k) \in \mathbb{Q}$ for every ideal class $\mathfrak{c}$ of F, and with $c(\ell) = \max_{\mathfrak{c}} \max\bigl(0, -v_\ell(2^{-n} z(\mathfrak{c}))\bigr)$:

1. for every prime ℓ, $c(\ell) \leq n\, v_\ell(2) + m_\ell(nk)$, where $m_\ell(nk) = 0$ if ℓ is odd and $(\ell - 1) \nmid nk$, $m_\ell(nk) = 1 + v_\ell(nk)$ if ℓ is odd and $(\ell - 1) \mid nk$, and $m_2(nk) = 2 + v_2(nk)$;
2. if $nk \geq 16$, then $\sum_\ell c(\ell) \log \ell \leq nk \log(nk + 1)$.

(TD) Toric denominators. For every abelian CM field E and every CM-type datum, the values at integers $s \geq 1$ of normalised Eisenstein-type lattice sums attached to algebraic Hecke characters of E have denominators bounded in terms of the class zeta values of (ZD).

(LT) Lattice transport. For every imaginary quadratic field K of discriminant $-d$ with $d > 4$ and every lattice M_K for an embedding φ_K of K, there are idèles t_σ, ideals $\mathfrak{a}_\sigma$ and lattices M_σ for $\sigma \in \mathrm{Aut}(\mathbb{C})$ fixing φ_K, with $j(M_\sigma) = \sigma(j(M_K))$, such that for every Galois CM field $E \supseteq K$ unramified at the finite places and every CM type of E of size at least 2 over φ_K, the normalised lattice sums of algebraic Hecke characters of E at M_K are algebraic, and σ carries them, after an explicit correction by $\mathrm{N}(\mathfrak{a}_\sigma)$, by the character at t_σ and by a power of the discriminant of M_σ, to the corresponding sums for the conjugate characters.

(TT) Twisted transport. The same transport law for the twisted lattice sums (lattice sums twisted by a finite-order character).

(RF) Reflection. For grid data over a master datum, all indices t, all nodes (i, j) and all integers s with $1 - e_0(i, j) \leq s \leq 0$, the finite Hecke L-functions (the entire continuation of the Euler product over the unramified finite places, without archimedean factors) of the level characters $\nu = \nu_{t,i,j}$ and $\nu_0 = \nu_{t_0,i,j}$ satisfy

$$L(s, \nu)\, L(1 - s, \nu_0) = \epsilon\, L(1 - s, \nu)\, L(s, \nu_0)$$

with ϵ in the coefficient field and a unit at the places prime to an explicit modulus.

The normalisation of the lattice sums, of the correction factors in (LT) and (TT), and of the denominator bounds in (TD) are part of the data that we do not reproduce.

Theorem 7.5. *The CM value statements (ZD), (TD), (LT), (TT) and (RF) of Definition 7.4 hold.*

Proof. In outline. The known theorems used are the main theorem of complex multiplication for elliptic curves and the generation of the Hilbert class field of an imaginary quadratic field by the

values of j [112], the rationality and congruences of partial zeta values of totally real fields [116], [69], [34], the functional equation of partial zeta functions, and the functional equation of Hecke L-functions [124], [53]. From them one obtains three statements on the values of Eisenstein-type lattice sums at CM lattices: integrality of their divided-power jets, their algebraicity, and their transformation under $\mathrm{Aut}(\mathbb{C})$ by Shimura reciprocity. A combination step, which uses these known theorems once more, turns these into the toric denominators (TD), the two transport statements (LT), (TT), the zeta denominators (ZD) and the reflection identity (RF).

□

7.3 Toric functoriality

The main result of the section is toric functoriality, the row condition (FUN): a bound for the maximal slope of the lattice R built from the theta expansions of the members of a host family. Its hypotheses are the CM value statements (Definition 7.4), which hold by Theorem 7.5, theta agreement and local matching (Proposition 7.3), and two conditions that belong to the family structure and that we do not reproduce: a separation property for theta functionals attached to the levels of the members, and a second form of the lower bound (LOW).

Theorem 7.6. *Assume the CM value statements (ZD), (TD), (LT), (TT) and (RF) of Definition 7.4, the theta agreement property and local matching of Definition 7.2, and two further conditions, which we do not reproduce: a separation property for theta functionals attached to the levels of the members, and a form of the lower bound (LOW) of Definition 6.16 that holds eventually on the families of the standard class. Then there is $C_{F,0}$ such that for every $C_F \geq C_{F,0}$ the following property of pairs $(\mathcal{F}, C_H)$ is closed on the standard class: if the divisor data of $\mathcal{F}$ have coefficients in the coefficient ring of its members, then there is x_0 such that, at every point of $\mathcal{F}$ at (x, i) with $x \geq x_0$ that satisfies the compatible point conditions of Definition 6.16 and admits a coefficient grid of size C_F (a condition on the point which we do not reproduce), the toric functoriality condition (FUN) holds for $(\mathfrak{S}^-, \mathfrak{N}(C, C_H))$ for every C.*

Proof. In outline. The theta agreement property, the CM value statements, the functional equation of Hecke L-functions [124], [53] and Howe duality at the archimedean places [58], [64] give the twelve square-top estimates of Theorem 7.9: denominator and height bounds for critical Hecke L-values on the grid (Proposition 7.11), nonvanishing of these values, existence of toric, denominator and bottom values, and the tower identity, which says that the ratio of the square constants of two members, multiplied by a product of ratios of finite Hecke L-values $L(s, \nu)$, is algebraic with controlled prime factors. Bounds for the archimedean height and the denominators of the coefficient maps of the members hold, and local sum estimates follow from theta agreement and local matching. The separation property makes the row functionals injective; the bound $|L'/L(1, \cdot)| \leq x^{\varepsilon}$ along the family, the twisted transport statement (TT) and the tower identity then place the square ratios in the coefficient field, which produces a function fitting the rows on the grid. A final combination of about twenty further estimates on the rows, among them mass, height, archimedean and integrality bounds, the bounds from (LOW) and rank gaps, yields (FUN).

□

The proof is a long assembly of estimates. Its engine is the square-top estimates of the next subsection; after them come some twenty estimates on the rows, which we know by name only.

7.4 Critical L-values on the grid

The theta-lift construction places the members of a family at the nodes of a grid of hermitian spaces and compares the theta constants at the nodes along its edges. Grid data record the characters, levels and scale of this comparison, and the thresholds b_{TH}, $a_{\mathfrak{J}}$, $b_{\mathfrak{J}}$ against which heights and sizes are measured. The size conditions that follow name the explicit constants $c_{\mathrm{grid}} = 1800$, $c_{\mathrm{type}} = 100$ and $m_0 = 44$, which fix the ranks from which the lifts are cuspidal and the bounds hold.

Definition 7.7. Let $\mathfrak{D}$ be a master datum.

(Gr1) Scale data. A **scale** for $\mathfrak{D}$ is a tuple $\mathfrak{x} = (x, b, \varepsilon, \beta_1, \dots)$ of reals $x, b > 1$, a parameter $\varepsilon > 0$, a real number β_1 (the logarithmic derivative at $s = 1$ of the L-function attached to the scale) and size conditions on the rows of $\mathfrak{D}$ (row sizes). It is **regular** if the row-size conditions hold, the host hypotheses of $\mathfrak{D}$ hold, F is the maximal totally real subfield of E, and $|\beta_1| \le x^{\varepsilon}$.

(Gr2) Level characters. A **level-character family** indexed by a set T consists of a character frame of $\mathfrak{D}$, Hecke characters ν_t of E ($t \in T$), unitary characters $\mu_{0,j}$ ($j \ge 0$) and η of $\mathbb{A}_E^{\times}/E^{\times}$, and a finite set of places of E, subject to conditions not transcribed; it has a distinguished index $t_0 \in T$.

(Gr3) Grid constants and grid data. Grid constants are triples $(c_{\mathrm{gap}}, c_{\mathrm{dep}}, b)$ of reals, with default value $(1, c_{\mathrm{grid}}, e)$, where

$$c_{\mathrm{grid}} = 1800.$$

Grid data for a level-character family and grid constants are a tuple of places (finite sets S_0, W_0 of places with depths $\mathrm{dep}(v)$), characters, and vertex data (Hermitian spaces at the nodes of the grid $\{\mathrm{top}(j), \mathrm{grid}(i,j)\}$, $0 \le i, j \le m$).

(Gr4) Thresholds. For grid data $\mathfrak{J}$ put $\sigma(\mathfrak{J}) = \sum_{v \in S_0} \mathrm{dep}(v) \log \mathrm{N}v$ and, for reals C, x, b,

$$\begin{aligned} b_{\mathrm{TH}}(C, x, b) &= C\, f\, b\, (\log b + \log x), \\ a_{\mathfrak{J}}(C, b) &= CfRb \log b + CmR^2 \sigma(\mathfrak{J}), \\ b_{\mathfrak{J}}(C, b) &= CfmR^2 \big(\log b + \sigma(\mathfrak{J})\big) b, \end{aligned}$$

where $R = R(\mathfrak{D}) = k + m(t + r + 7)$, with k, m, t the ranks and $r = k + m - 1$ the derived rank of $\mathfrak{R}$.

(Gr5) Large-scale regime. A family of properties $\mathcal{Q}(\mathfrak{J}, t_0; C, x, b)$ of grid data holds **uniformly at large scale** if there is $C_{\mathrm{dep},0}$ such that for all ε and all $C_{\mathrm{dep}} \ge C_{\mathrm{dep},0}$ there are C_0 and x_0 such that $\mathcal{Q}(\mathfrak{J}, t_0; C, x, b)$ holds for all $C \ge C_0$, all master data $\mathfrak{D}$ with a regular scale $(x, b, \varepsilon, \dots)$ with $x \ge x_0$, all level-character families with distinguished index t_0, and all grid data $\mathfrak{J}$ with constants $(1, C_{\mathrm{dep}}, b)$. The variant for **Hida choices** quantifies in the order $\exists C_{\mathrm{dep},0}\, \forall C_{\mathrm{dep}} \ge C_{\mathrm{dep},0}\, \exists C_0\, \forall \varepsilon, C \ge C_0\, \exists x_0$, and ranges over families of members over $\mathfrak{D}$ in a Hida setting with a face and a well-chosen choice of grid data at scale (C_{dep}, b) whose field conditions hold.

The row-size conditions and host hypotheses of (Gr1), the conditions on level-character families in (Gr2), the places, characters and vertices of grid data in (Gr3), and the Hida settings, faces and choices of (Gr5) consist of further conditions of the same nature, which we do not reproduce. The constant c_{grid} is the threshold for the ranks of the grid; it recurs in Definition 7.8.

Definition 7.8. Let $\mathfrak{D}$ be a master datum and write $b_j(\tau)$ for the entries of the parameter b of Ξ_τ. Put

$$c_{\mathrm{type}} = 100, \qquad m_0 = 44.$$

(Sz1) Cusp-row sizes. With the **cusp-row gap** $g_{\mathrm{cusp}}(\mathfrak{D}) = 2R(\mathfrak{D})^3$, $\mathfrak{D}$ satisfies the **cusp-row size condition** if $r > c_{\mathrm{grid}}$ and $b_{j+1}(\tau) - b_j(\tau) > g_{\mathrm{cusp}}(\mathfrak{D})$ for all $\tau \in \Phi$ and $0 \le j < m - 1$.

(Sz2) Row sizes with gap h_0. For $h_0 \in \mathbb{R}$, $\mathfrak{D}$ satisfies the **row-size condition with gap** h_0 if $b_{j+1}(\tau) - b_j(\tau) \ge h_0$ for all τ and $0 \le j < m - 1$, h_0 is at least the **row-size gap** $g_{\mathrm{row}}(\mathfrak{D}) = 2(r + mt)^3 + t + 3$, and $r + mt > c_{\mathrm{grid}}$.

(Sz3) Type hypothesis. $\mathfrak{D}$ satisfies the **type hypothesis** if $r \ge c_{\mathrm{type}}$ and $2r \le g_i$ for $i = 1, 2$.

(Sz4) Scale sizes. A scale $(x, b, \varepsilon, \dots)$ with logarithmic size parameter L satisfies the **scale-size condition** for $\mathfrak{D}$ if $m^4 \le f$, $m \ge m_0$, $x^2 m^3 L \log \mathrm{rd}_E \le k$, $mfk_m^2 d \le b$, $fm \le (r+1)^3$, $\log g_i \le 9L + 1$ for $i = 1, 2$, and $\log \mathrm{rd}_E \le \log \mathfrak{f}_E \le L/100$, where rd_E is the root discriminant of E and $\mathfrak{f}_E$ the conductor of the host, together with further size conditions of the same nature, which we do not reproduce.

The square-top estimates are the arithmetic core of toric functoriality. Most of them are the five results that follow; the valuation bound for the square ratios, item 2, is proved last, from the reflection statement (RF) and the other items.

Theorem 7.9. *Assume the CM value statements (TD), (LT), (ZD) and (RF) of Definition 7.4, the theta agreement property and local matching of Definition 7.2. Then each of the following holds uniformly at large scale for grid data $\mathfrak{J}$:*

1. *the denominator bounds and toric height bounds (Hb1) and (Hb4) of Proposition 7.11;*
2. *a valuation bound for the square ratios: at every finite place w of the coefficient field, the valuation of each square ratio is bounded by an explicit linear combination of the valuations of the auxiliary, torsion and denominator values, a loss function and the conductor;*
3. *smallness (Bp3), nonvanishing (Bp4), existence (Bp1) of bottom pairs and existence (Bp5) of bottom triangles, Proposition 7.13;*
4. *the field data: the host field of $\mathfrak{D}$ is a CM field abelian over $\mathbb{Q}$ with the field and base-conductor conditions of the grid, and admits an integral lattice;*
5. *existence of torsion values and of denominator functions with finite support at every node (i, j) and every $1 \le s \le h(i, j)$;*
6. *the nonvanishing of Hecke L-values Lemma 7.10;*
7. *the tower identity Proposition 7.14.*

The L-values in items 1, 2, 6 and 7 are values of finite Hecke L-functions (Euler products over the unramified finite places, continued to entire functions; no archimedean factors), as in Proposition 7.12.

Proof. In outline. Besides the hypotheses, the proof uses the functional equation of Hecke L-functions and the local theory of the theta correspondence, in particular Howe duality over p-adic fields and over $\mathbb{R}$ [42], [58], [64]. The items are proved in a fixed order, each from the earlier ones. The field data of item 4 come first. The functional equation of Hecke L-functions [124], [53] gives the nonvanishing Lemma 7.10 and the bound (Hb2) of Proposition 7.11. The lattice transport (LT) gives the torsion values of item 5, and with them (Hb3). The zeta denominators (ZD) give the denominator functions and (Hb1). The toric denominators (TD) together with these give (Hb4). Next come the smallness (Bp3) of bottom pairs, their existence (Bp1) and (Bp2) (Proposition 7.13), and the norm identity of Proposition 7.12. Real Howe duality and the theta agreement property then give the bottom triangles (Bp5), the nonvanishing (Bp4) and the tower identity Proposition 7.14. Finally the reflection statement (RF) and all of the above give the valuation bound of item 2.

□

The crux is the pair formed by the norm identity for bottom pairs (Proposition 7.12) and the tower identity (Proposition 7.14). Together they turn global theta constants into products of critical Hecke L-values, as the Rallis inner product formula for unitary groups does for a single lift [80], [49]. We state the five results in the order of the proof: nonvanishing at critical integers, height bounds, the norm identity, bottom pairs and the tower identity.

Lemma 7.10. *Uniformly at large scale, for grid data $\mathfrak{J}$ and all $0 \le i, j < m$, with the Hecke characters $\nu_{t,i,j}$ and the integers $e_0(i,j)$ of Proposition 7.12 and the half ranks $h(i,j)$ of Proposition 7.11:*

1. *for every $t \in T$ and every integer s with $1 - e_0(i,j) \le s \le h(i,j)$, $L(s, \nu_{t,i,j}) \neq 0$;*
2. *for the distinguished index t_0, the character $\nu_{t_0,i,j}$ and each of its twists $\nu_{t_0,i,j}\delta_c$ by the two auxiliary characters δ_c of the level-character family satisfy $L(s,\cdot) \neq 0$ for every integer $s \ge 1$.*

Here, as in Proposition 7.12, $L(s,\nu)$ is the finite Hecke L-function: the entire function of s that equals the Euler product over the finite places at which ν is unramified for $\operatorname{Re} s > 1$, without archimedean factors; part of the assertion is that this function is entire. The auxiliary characters δ_c are data of the level-character family, which we do not reproduce.

Proof. In outline. Both parts follow from the functional equation of Hecke L-functions [124], [53], the boundedness of completed Hecke L-functions in vertical strips, and the field data of the grid: the host field is a CM field, abelian over $\mathbb{Q}$, with an integral lattice.

□

Proposition 7.11. *For $0 \le i, j < m$ let $h(i,j) = \lfloor (\dim V_{i+1,j+1} + \dim V_{i,j})/2 \rfloor$ be the half rank of the grid node (i,j), where $V_{i,j}$ are the hermitian spaces at the grid nodes. Uniformly at large scale, for grid data $\mathfrak{J}$ with parameters (C,x,b), all $0 \le i, j < m$ and all integers $1 \le s \le h(i,j)$:*

(Hb1) *every denominator function n of $\mathfrak{J}$ at (i,j,s) takes values $n(p) \ge 0$ at primes p and satisfies $\sum_{p \in S} n(p) \log p \le b_{\mathrm{TH}}(C,x,b)$ for every finite set S of primes;*

(Hb2) *for every permutation π of Φ and every algebraic Hecke character χ of E whose infinity type at s is the critical type of (i,j,s) permuted by π and whose conductor divides the conductor of $\mathfrak{J}$, $\log |L(s,\chi)| \le b_{\mathrm{TH}}(C,x,b)$;*

(Hb3) *for every torsion value y of $\mathfrak{J}$ at (i,j,s) and every integral lattice M of the host, the algebraic number $\beta=\beta(s,y,M)$ attached to them satisfies $\log|a_0(\beta)|\le 12\,\deg(\beta)\,b_{\mathrm{TH}}(C,x,b)$, where $a_0(\beta)$ is the constant coefficient of its minimal polynomial over $\mathbb{Q}$;*

(Hb4) *every toric value (z,g) of $\mathfrak{J}$ at (i,j,s) has height at most $b_{\mathrm{TH}}(C,x,b)$ in L, and so does 1 relative to g.*

Denominator functions, torsion and toric values, the numbers $\beta(s,y,M)$, the heights relative to g and the critical infinity types are data of the grid of the same nature as those of Definition 7.7, which we do not reproduce.

Proof. In outline. Each item is proved separately at large scale: (Hb1) from a real inequality for linear combinations of logarithms, (Hb2) from the functional equation of $L(s,\chi)$ [124], [53] applied to the permuted weights, together with the boundedness of completed Hecke L-functions in vertical strips, (Hb3) by transporting the archimedean bound to the conjugates of β, and (Hb4) from the denominator bounds and mean bounds over the places.

□

Proposition 7.12. *For grid data $\mathfrak{J}$ over $\mathfrak{D}$, indices $t\in T$ and $0\le i,j<m$, let $\nu_{t,i,j}$ be the Hecke character of E attached to (t,i,j) by the grid, let $n_{i,j}=\dim V_{i,j}$ be the dimensions of the hermitian spaces at the grid nodes, let $\beta^{+}_{ij}=n_{i+1,j+1}-2\lfloor(m-j-1)/2\rfloor$ and $\beta^{-}_{ij}=n_{i,j}-2\lfloor(m-j)/2\rfloor$ be the big and small bottom indices, and put $e_0(i,j)=\lfloor(\beta^{+}_{ij}-\beta^{-}_{ij})/2\rfloor$ and $h_0(i,j)=\lfloor(\beta^{+}_{ij}+\beta^{-}_{ij})/2\rfloor$ (with $e_0(i,j)=0$ if $\beta^{+}_{ij}<\beta^{-}_{ij}$), and let $\mathfrak{m}_t$ be the modulus of $\mathfrak{J}$ at t. Then, uniformly at large scale, the following holds: for all $\delta,t_1\in\mathbb{C}$ and $0\le i,j<m$ there is a nonzero constant $A\in\mathbb{C}$ such that for every $t\in T$, every level Q and every bottom pair (P,P') of $\mathfrak{J}$ at (t,i,j) with modulus $\mathfrak{m}_t$ and parameters (δ,t_1,Q),*

$$\mathrm{N}(P)=A\prod_{s=1-e_0(i,j)}^{h_0(i,j)}L(s,\nu_{t,i,j}),$$

where N is the norm of a pair, s runs over integers, and $L(s,\nu)$ denotes the finite Hecke L-function of ν: the entire function of s that equals the Euler product over the finite places at which ν is unramified for $\mathrm{Re}\,s>1$ (no archimedean factors). Bottom pairs, their norms and the bottom indices are data of the grid of the same nature as those of Definition 7.7, which we do not reproduce.

Proof. In outline. The identity follows from the local theory of the theta correspondence (Howe duality over p-adic fields and over $\mathbb{R}$ [42], [58]), the norm-bound properties of the CM theta data of the master datum and the nonvanishing of the Hecke L-values Lemma 7.10. The norm of a bottom pair is a Petersson inner product of theta lifts on the definite bottom space of the grid, and an inner-product formula of Rallis type [80], [49] expresses it through the L-values.

□

Proposition 7.13. *Uniformly at large scale, for grid data $\mathfrak{J}$ with base level Q_0 and conductor modulus $\mathfrak{c}$, with the moduli $\mathfrak{m}_t$ of Proposition 7.12 and the levels Q_t of the indices $t\in T$:*

(Bp1) Existence at t_0. *There are $\delta,t_1\in\mathbb{C}$ such that for all $0\le i,j<m$ there is a bottom pair of $\mathfrak{J}$ at (t_0,i,j) with modulus $\mathfrak{c}$ and parameters (δ,t_1,Q_0).*

(Bp2) Existence at every index. *Whenever a pair with parameters (δ,t_1,Q) exists, then for all $t\in T$ and $0\le i,j<m$, there is a bottom pair at (t,i,j) with modulus $\mathfrak{m}_t$ and parameters (δ,t_1,Q_0Q_t).*

(Bp3) Smallness. *The level Q_0 is good, t_0 has no twisting places, and every bottom pair of $\mathfrak{J}$ is small with respect to $R(\mathfrak{D})$ and the maximal weight of the grid constants.*

(Bp4) Nonvanishing. *Both members of every bottom pair at (t_0, i, j) with modulus $\mathfrak{c}$ have a nonzero coordinate.*

(Bp5) Bottom triangles. *For every Weil datum of $\mathfrak{J}$ with the standard additive character and all $0 \le i, j < m$, there are hermitian spaces V, V', W and a bottom triangle on them at (i, j) whose theta values are constant along the grid indices, with constants $\gamma\colon T \to \mathbb{C}$.*

Here bottom pairs and their coordinates, smallness, good levels, twisting places and bottom triangles are data and conditions of the grid of the same nature as those of Definition 7.7, which we do not reproduce.

Proof. In outline. (Bp1) and (Bp2) follow from the norm-bound properties of the CM theta data of the master datum; (Bp3) is a direct verification. (Bp5) follows from real Howe duality [58], the theta agreement property and local matching of Definition 7.2, and the nonvanishing Lemma 7.10. (Bp4) follows from these together with (Bp3), (Bp5) and the norm identity of Proposition 7.12.

□

Proposition 7.14. *For $0 \le i, j < m$ put $p_j = \lfloor (m - j - 1)/2 \rfloor$, $q_j = \lfloor (m - j)/2 \rfloor$ and $e_{ij}^{\mathrm{top}} = \lfloor \frac{1}{2}(\dim V_{i+1,j+1} - \dim V_{i,j}) \rfloor$ (with value 0 if the difference is negative), where $V_{i,j}$ are the hermitian spaces at the grid nodes. Uniformly at large scale the following holds for grid data $\mathfrak{J}$ with index set T and distinguished index t_0, with the Hecke characters $\nu_{t,i,j}$ of Proposition 7.12. Let a Weil datum of $\mathfrak{J}$ with the standard additive character, a grid choice whose numbers lie in L at t_0, and a real matching be given. Let $0 \le i, j < m$ and $t \in T$, let $c, c_0 \in \mathbb{C}$ be the (nonzero) square constants at (t, i, j) and (t_0, i, j), and let $\gamma\colon T \to \mathbb{C}$ be the constants of a bottom triangle at (i, j). Then there are $u_1, \dots, u_{p_j+q_j} \in \mathbb{C}$, each a step ratio relative to the level Q_t of t and the places of $\mathfrak{J}$, and an element $z_t \in L$ that is a unit at every place prime to Q_t, such that*

$$\frac{c}{c_0} \prod_{k=0}^{q_j - 1} \frac{L(e_{ij}^{\mathrm{top}} + k + 1, \nu_{t,i,j})}{L(e_{ij}^{\mathrm{top}} + k + 1, \nu_{t_0,i,j})} = z_t \prod_{k=1}^{p_j + q_j} u_k \cdot \frac{\gamma(t)}{\gamma(t_0)} \prod_{k=0}^{p_j - 1} \frac{L(e_{ij}^{\mathrm{top}} + q_j - k, \nu_{t,i,j})}{L(e_{ij}^{\mathrm{top}} + q_j - k, \nu_{t_0,i,j})}.$$

Here the square constants, bottom triangles, step ratios, grid choices and real matchings are data of the grid of the same nature as those of Definition 7.7, which we do not reproduce.

Proof. In outline. The identity is proved by induction along the tower of the grid, from the local theory of the theta correspondence (Howe duality over p-adic fields and over $\mathbb{R}$ [42], [58]), the theta agreement property and local matching of Definition 7.2, and the nonvanishing of the Hecke L-values Lemma 7.10. The shape is that of the Rallis inner product formula along a tower [80], [49]: each step changes the square constant by a ratio of L-values and a step ratio.

□

The valuation bound of Theorem 7.9, which bounds the w-adic valuation of each square ratio by an explicit linear combination of the valuations of the auxiliary, torsion and denominator values, a loss function and the conductor, is where the arithmetic of the tower identity becomes

quantitative. Toric functoriality also uses bounds for the coefficient maps of the theta expansions: the coefficient maps of the members have archimedean height at most $m\,a_{\mathfrak{J}}(C,b)$, and denominators of total weight at most $\frac{1}{2}m\,b_{\mathfrak{J}}(C,b)$ at the finite places away from the members. These bounds rest on a dictionary: the jets of the grid theta data at the CM points are, up to period monomials and a constant, Taylor coefficients of bounded integral combinations of theta functions of vectors of self-dual hermitian lattices.

The family structure needs one more ingredient from this circle of ideas, an evaluation grid on which the row functionals are evaluated. Under theta agreement and local matching, and their twisted versions, such a grid exists at large scale; it is built from Weil–Deligne parameters attached to the local data of the members, the cuspidality of theta lifts along the grid, which holds under the cusp-row size condition (Sz1) of Definition 7.8, and a normalisation of the characters over a number field. This completes the theta-lift ingredients of the family structure; the covolume condition (COV) is the subject of the next section.

8 The covolume bound by towers and heights

This section outlines the proof of the covolume condition (COV) at the compatibly ruled points of host families (Theorem 8.1), the second of the two row conditions of the family structure whose proofs the outline shows. The covolumes are those of hermitian lattices of automorphic forms attached to a member, and the bound comes from lower bounds for arithmetic heights of automorphic line bundles on integral models of unitary Shimura varieties over the host field. The argument organises these models into towers of flag models, indexed by a bound B for the primes treated as bad. On the towers it proves a proportionality of Hirzebruch–Mumford type [95] and an arithmetic proportionality with error $O(\log B/B)$, and from these lower bounds for heights and mean slopes (Proposition 8.6). A transport to the models attached to a master datum gives a mean-slope lower bound (Proposition 8.5), and a chain of lattice estimates turns it into the covolume bound. The known theorems used are arithmetic intersection theory and the arithmetic Hilbert–Samuel formula [44], [91], [136], Zhang's lower bound for heights by successive minima [138], toroidal compactifications [37], [75], the Hodge–Tate sequence [38], [105], Zarhin's trick [137], and resolution of singularities [57]. What looks new is the tower construction with its volume properties; the inputs are standard.

Theorem 8.1. *Call a point of a host family* ***compatibly ruled*** *if it satisfies all conditions of a ruled point in Definition 6.16 except the last one, the tie conditions, and satisfies instead the following* ***compatibility conditions****: the point arises from a realisation in a second, recursive form, with the same core, pull-back and type ties as in the tie conditions, together with further ties for the Eisenstein series, the level and the data, integrality of the twisted characters, and conditions on ray class fields and roots of unity contained in the coefficient field. These are the point conditions used for (FUN) in Definition 6.16; we do not reproduce them further. We also require that its member satisfies the two further conditions of the covolume property in Definition 6.16: a size bound for its field and a compatibility condition for Eisenstein series. Then the following property of pairs $(\mathcal{F}, C_H)$ is closed on the standard class: there are $C_0 \geq 0$ and x_0 such that at every compatibly ruled point of $\mathcal{F}$ at (x,i) with $x \geq x_0$ the covolume condition (COV) holds for $(\mathfrak{S}^-, \mathfrak{N}(C, C_H))$ for all $C \geq C_0$.*

Proof. In outline. (We do not know whether compatibly ruled points are ruled, or conversely: the two realisation conditions are of different form.) The bound is obtained from the mean slope lower bound Proposition 8.5 through a chain of estimates on the Hermitian lattices of automorphic forms attached to a point. The known theorems used are:

- resolution and principalisation in characteristic 0 [57], [12];
- arithmetic bigness, the arithmetic volume being at least the self-intersection [91], [136], with the arithmetic Hilbert–Samuel formula and the inverse Poincaré–Lelong formula;
- the Hodge–Tate sequence of p-divisible groups [126], [38] and the Zarhin trick [137];
- toroidal compactifications with their charts and the log Kodaira–Spencer isomorphism [37], [75] and the Siegel toroidal integral;
- seminegativity of the Lie algebra of semi-abelian schemes [37], [90] and Grothendieck–Messing deformation theory [83];
- nef Riemann–Roch, point separation, the arithmetic degree and Green's regular characters [46].

The first three groups give Proposition 8.5. With all of them, one bounds in turn the Gram determinants of the lattices attached to the point and the measures that enter their volumes. One then shows that suitable integral models exist and have good reduction at the point, and bounds their degrees. A lower bound for heights comes next, and it is here that Proposition 8.5 enters. Summing the heights, truncating, and discarding a set of bad primes then gives a lower bound for the slope of each lattice M_j, which is an upper bound for its covolume, that is, (COV). The truncation uses an upper bound for the maximal slope of the ambient lattice.

□

The bound is reached from the mean-slope lower bound through some twenty lattice estimates, which we know only through their inputs. We now describe the integral models and the towers.

Definition 8.2. Let E be a CM field with maximal totally real subfield E^+ and complex conjugation $a \mapsto \bar{a}$, such that $\mathcal{O}_E/\mathcal{O}_{E^+}$ is unramified.

(U1) CM host. A **CM host** on E is a prime $d \equiv 7 \pmod{16}$ with an element $\sqrt{-d} \in \mathcal{O}_E$ and $w_0 = \frac{1+\sqrt{-d}}{2} \in \mathcal{O}_E$, together with a choice of $\sqrt{-d} \in \mathbb{C}$. The **CM type** Φ is the set of embeddings $\tau\colon E \to \mathbb{C}$ sending $\sqrt{-d}$ to the chosen root.

(U2) Unitary datum. A **unitary datum** of signature (p, q) is a finite projective $\mathcal{O}_E$-module M of rank $p+q$ with a perfect alternating $\mathbb{Z}$-bilinear form $\langle\,,\rangle$ satisfying $\langle ax, y\rangle = \langle x, \bar{a}y\rangle$, admitting a complex structure J on $M \otimes \mathbb{R}$ compatible with $\langle\,,\rangle$, positive, and satisfying the Kottwitz signature condition: for $a \in \mathcal{O}_E$ the trace of a on $(M\otimes\mathbb{R}, J)$ is $\sum_{\tau\in\Phi}\big(p\,\tau(a) + q\,\overline{\tau(a)}\big)$.

(U3) Level. A **level** is an odd integer $N \geq 3$ with an ideal $\mathfrak{m}$ generated by elements of $\mathbb{Z}[w_0]$ such that $\mathfrak{m}\bar{\mathfrak{m}} = N\mathcal{O}_E$ and $\mathfrak{m} + \bar{\mathfrak{m}} = \mathcal{O}_E$, and an integer $A \geq 1$.

(U4) Moduli. A **moduli scheme** for a unitary datum $\mathfrak{U} = (M, p, q)$ and a level $\mathfrak{m}$ is a smooth quasi-projective scheme $\mathrm{Sh}_{\mathfrak{m}}(\mathfrak{U})$ over a fixed number field $K_{\mathfrak{m}}$ with a universal object which represents, on reduced schemes, principally polarised abelian schemes with $\mathcal{O}_E$-action of signature (p, q) and level-$\mathfrak{m}$ structure modelled on M, and is unramified in the sense that first-order deformations of points are determined by the universal object.

(U5) Global models. For a coefficient field $F \subset \mathbb{C}$, finite over $\mathbb{Q}$, containing $K_{\mathfrak{m}}$, all $\tau(E)$ and the N-th roots of unity, a **model** of $\mathrm{Sh}_{\mathfrak{m}}(\mathfrak{U})$ over $\mathcal{O}_F$ is a normal arithmetic variety $\mathcal{X}$, flat and projective over $\mathcal{O}_F$, with a dense open immersion of $\mathrm{Sh}_{\mathfrak{m}}(\mathfrak{U})_F$ into its generic fibre, and a semi-abelian scheme G over $\mathcal{X}$ with $\mathcal{O}_E$-action and a symmetric cubical line bundle (a polarisation) extending the universal abelian scheme, whose abelian locus is exactly the image of $\mathrm{Sh}_{\mathfrak{m}}(\mathfrak{U})_F$. A model is **global** if the level is good at every prime, $A = 1$, the model descends, and at every rational prime p_0 and every place of $\mathbb{C}$ above p_0 it dominates a family of local p_0-adic systems with split fibres satisfying three regularity conditions. Its **components split** if F is algebraically closed in the function field of every irreducible component of $\mathcal{X}$ (every element of such a function field that is algebraic over $\mathbb{Q}$ lies in F).

The representability condition in (U4) and the local systems and regularity conditions in (U5) are subject to further conditions of the same nature, which we do not reproduce.

A root tower replaces the Shimura variety by flag models: arithmetic varieties of the dimension of the flag variety Fl_n^{Φ}, with one hermitian line bundle for each root. Its volume properties compare the arithmetic intersection numbers of the tower with those of the flag variety: (V1) is a proportionality of Hirzebruch–Mumford type, (V4) its arithmetic counterpart with error $O(\log B/B)$, and (V5) and (V6) are the resulting lower bounds for heights and mean slopes.

Definition 8.3. Fix a finite set Φ (in applications a CM type) and $n \geq 2$; let Fl_n be the variety of complete flags in an n-dimensional space, $d_n = \dim \mathrm{Fl}_n^{\Phi} = |\Phi|\binom{n}{2}$. A **weight** is $u \in \mathbb{R}^{\Phi \times \{1,\dots,n\}}$; it is dominant if each u_τ is non-increasing; $\deg_{\mathrm{Fl}}(u)$ is the top self-intersection of the line bundle $\mathcal{L}_u$ on Fl_n^{Φ}, $\|u\|_1 = \sum |u_{\tau,i}|$ and $\mathrm{spr}(u) = \sum_\tau (\max_i u_{\tau,i} - \min_i u_{\tau,i})$.

A **flag model** is an arithmetic variety $\mathcal{X}$ of absolute dimension $d_n + 1$, given by coordinates in a product of projective spaces, with Hermitian line bundles $\overline{\mathcal{L}}_{\tau,i}(t)$ indexed by the roots (τ, i), metrised with a regularisation parameter $t > 0$; for a weight u write $\overline{\mathcal{L}}_u(t) = \bigotimes \overline{\mathcal{L}}_{\tau,i}(t)^{u_{\tau,i}}$ (real coefficients, by linearity of intersection numbers). Its **geometric degree** $\deg(u^{d_n})$ is the degree of $\mathcal{L}_u$ on the generic fibre, and its **regularised arithmetic intersection numbers** are $\widehat{\deg}(u^{d_n+1}) = \lim_{t\to 0^+} \widehat{\deg}\big(\widehat{c}_1(\overline{\mathcal{L}}_u(t))^{d_n+1}\big)$, extended linearly to polynomials in the roots.

A **root tower** $\mathcal{T}$ for (Φ, n) consists of an index $1 \leq j < n$, a non-empty finite set Comp (components), a finite set Bad of primes containing 2, a rational weight ν, and flag models $\mathcal{X}_{B,c}$ for $B \in \mathbb{N}$ and $c \in \mathrm{Comp}$. Put $\mathrm{vol}(c) = \deg_{\mathcal{X}_{2,c}}(\rho^{d_n})/\deg_{\mathrm{Fl}}(\rho)$ with ρ the half-sum of positive roots, $\mathrm{vol} = \sum_c \mathrm{vol}(c)$, and the **slope** of u at (B, c)

$$\mu_{B,c}(u) = \frac{\widehat{\deg}_{\mathcal{X}_{B,c}}(u^{d_n+1})}{(d_n+1)\,\mathrm{vol}(c)\,\deg_{\mathrm{Fl}}(u)}.$$

For real $r > 0$ and $L \in \mathbb{R}$ put $\mathrm{Lg}(B, r, L) = \log(Br) + |L|$ and $b(B, r, L; u) = \mathrm{Lg}(B, r, L)\|u\|_1 + \min(j, n-j)\,\mathrm{spr}(u)$. The tower has the following **volume properties**, with a constant C and parameters (r, L):

(V1) Proportionality. $\mathrm{vol}(c) > 0$, and $\deg_{\mathcal{X}_{B,c}}(F) = \mathrm{vol}(c)\deg_{\mathrm{Fl}}(F)$ for every polynomial F of degree d_n in the roots, independently of B.

(V2) Equal volumes. $\mathrm{vol}(c) = \mathrm{vol}(c')$ for all components.

(V3) Convergence. For weights u satisfying a gap hypothesis with gap $g > 0$ and $B + 1 > 4|\Phi|n^2(1 + \|u\|/g)$, the regularised self-intersections and degrees of $\overline{\mathcal{L}}_{u-s\nu}$ converge as $s \to 0^+$ to $\widehat{\deg}(u^{d_n+1})$ and $\mathrm{vol}(c)\deg_{\mathrm{Fl}}(u) > 0$, with semipositivity and nefness on fibres for small s.

(V4) Arithmetic proportionality. There are C, B_1 and, for $B \geq B_1$, constants α_τ, β_τ such that for every monomial m of degree d_n+1, $\big|\sum_c \widehat{\deg}_{\mathcal{X}_{B,c}}(m) - \mathrm{vol}\cdot\big(h(m) - \sum_\tau(\alpha_\tau - \beta_\tau)\mathrm{Sh}_\tau(m) - \sum_\tau \beta_\tau \mathrm{Tot}_\tau(m)\big)\big| \leq C\log B/B$, for a fixed linear form h on the flag variety and two families of linear forms Sh_τ, Tot_τ.

(V5) Height lower bound. For u as in (V3), $B \geq 128$, $|\mathrm{Bad}| \leq B$: $\sum_c \widehat{\deg}_{\mathcal{X}_{B,c}}(u^{d_n+1}) \geq -C(d_n+1)\,\mathrm{vol}\,\deg_{\mathrm{Fl}}(u)\,b(B,r,L;u)$.

(V6) Mean slope. Under the same conditions, $|\mathrm{Comp}|^{-1}\sum_c \mu_{B,c}(u) \geq -C\,b(B,r,L;u)$.

The construction of the flag models and their regularised metrics, the gap hypothesis, and the linear forms h, Sh_τ, Tot_τ of (V4) are subject to further conditions of the same nature, which we do not reproduce.

A master datum provides, for each step j of its tower, a Shimura variety of the unitary group of W_j with deep integral models; the next definition records them and their volume slopes.

Definition 8.4. Let $\mathfrak{D}$ be a master datum with host field E.

(Co) Core. A **core** of $\mathfrak{D}$ is the following data. For each step j: a unitary datum $\mathfrak{U}_j$ modelled on W_j, a level and a moduli scheme Sh_j; a common coefficient field F_{c}; for each embedding $\sigma\colon F_{\mathrm{c}} \to \mathbb{C}$, a complex uniformisation of $\mathrm{Sh}_j \otimes_\sigma \mathbb{C}$ by finitely many arithmetic quotients of hermitian symmetric domains with frames, and a dictionary from the adelic unitary group of W_j to these frames; constants $c_E(j,\sigma)$ per component, unitary characters $\chi_W(j,\sigma)$ of $\mathbb{A}_E^\times$, base frames $x_0(j,\sigma)$ and lattice embeddings. These are subject to the **laws**: the constants are units, the characters take the prescribed values and are conjugation-compatible, the moduli schemes are integral, the uniformising quotients are models of W_j, and the base frames are orthogonal.

(Dm) Deep models. A **deep model** at step j of a core is an enlargement $F_{\mathrm{c}}' \supseteq F_{\mathrm{c}}$ of the coefficient field, an integral model $\mathcal{X}$ of Sh_j at an auxiliary level (an ideal $\mathfrak{l}$ and an index p_{d}), complex uniformisations U_σ of its fibres, a dictionary from the adelic group of W_j to their frames, and a compact open subgroup K_{f}. It is **good** for (n_1, o), with $n_1 \geq 0$ an integer and o real, if: $p_{\mathrm{d}} = 15^{n_1}$; the level is good with $A = 1$; at every good embedding σ the U_σ are genuine uniformisations satisfying the test and character-triviality conditions, the dictionary is a uniformisation for K_{f}, K_{f} has level $g_1 g_2 p_{\mathrm{d}}$, the base points have the prescribed absolute value, the minors of the frames have absolute value 1, and the quotients are models of W_j; the function field is geometric and F_{c}' is normal over F_{c} with size at most o; and the lattice of $\mathfrak{U}_j$ embeds isometrically into W_j, compatibly with the uniformisations.

(Vs) Component volume slopes. For a connected component Z of a deep model and a weight u, let $\deg_Z(u) > 0$ be the top degree of the automorphic line bundle of weight u on Z. Let $\mathrm{vol}_Z(u,t)$ be the asymptotic volume (normalised count, as for the arithmetic volume) of the cusp-form sections of weight $8nu$ ($n \to \infty$) on Z whose finite valuations are bounded below by n times the depth of the reference lattice and whose archimedean norms, at every frame point, are at most e^{2nt} times the n-th power of the Bergman reference. The **volume slope** is

$$\widehat{\mu}_Z(u) = \sup\big\{s \in \mathbb{R} :\ \mathrm{vol}_Z(u,t) \geq \deg_Z(u)\,(s+t) \text{ for all } t\big\}.$$

Uniformisations, dictionaries, frames, the test conditions and the valuation family in (Vs) are subject to further conditions of the same nature, which we do not reproduce.

The mean-slope lower bound for these models is the following. Its error term involves the root discriminant of the host field and $L'/L(1,\chi_{-d})$ for the quadratic character χ_{-d} of $K=\mathbb{Q}(\sqrt{-d})$.

Proposition 8.5. *For a master datum $\mathfrak{D}$ put $r_E=|\mathrm{disc}(E)|^{1/[E:\mathbb{Q}]}$, let χ_{-d} be the quadratic character of K (the character modulo d attached to $\mathbb{Q}(\sqrt{-d})$), and put*

$$\mathrm{ar}(\mathfrak{D})=\log r_E+\Big|\mathrm{Re}\,\frac{L'(1,\chi_{-d})}{L(1,\chi_{-d})}\Big|.$$

There is a constant $C\geq 0$ with the following property. Let $\mathfrak{D}$ be a master datum with $m^2\leq k$ and let $1\leq j<m$. Fix a core of $\mathfrak{D}$ satisfying its laws, and a deep model $\mathcal{X}$ of it at step j that is good for some parameters and whose level has at most 15 prime factors. Let (p,q) be the signature of the step-j space. Let $u=(u_{\tau,a})\in\mathbb{Z}^{\Phi\times\{1,\dots,p+q\}}$ be an integral weight, τ running over the set Φ of CM-type places of $\mathfrak{D}$, and let $g>0$ be real, such that $a\mapsto u_{\tau,a}$ is strictly decreasing for each τ and

$$|u_{\tau,a}-u_{\tau,b}|\geq g\qquad\text{for all }\tau\text{ and all }a\neq b\text{ with }a>p.$$

Then the mean over the connected components Z of $\mathcal{X}$ of the volume slopes $\widehat{\mu}_Z(u)$ satisfies

$$\frac{1}{\#\{Z\}}\sum_Z\widehat{\mu}_Z(u)\;\geq\;-C\Big(\big(\log B(u,g)+\mathrm{ar}(\mathfrak{D})\big)\|u\|_1+q\,\mathrm{spr}(u)\Big),$$

where

$$B(u,g)=4|\Phi|(p+q)^2\Big(1+\frac{\max_{\tau,i}|u_{\tau,i}|}{g}\Big).$$

Proof. In outline. A good deep model is realised over a root tower with a good origin. The tower theorem Proposition 8.6 gives the mean slope property (V6) of the tower, with $r=r_E$ and L the logarithmic derivative of $L(s,\chi_{-d})$ at $s=1$. Nef Riemann–Roch and a birational pull-back and push-forward formula for arithmetic intersection numbers transport this bound to the volume slopes of the deep model. Resolution of singularities and principalisation [57], [12] enter through the proofs of the tower theorem and of the pull-back formula.

□

It rests on the tower theorem, the crux of the section: good root towers exist over global integral models, and they have the volume properties (V1)–(V6), with the root discriminant and $\mathrm{Re}\,L'/L(1,\chi_{-d})$ as parameters.

Proposition 8.6. *1. Let E be a CM field with a CM host, $\mathfrak{U}$ a unitary datum of signature (p,q) with $p,q\geq 1$, $\mathfrak{m}$ a level and $\mathcal{X}$ a global model of $\mathrm{Sh}_{\mathfrak{m}}(\mathfrak{U})$ whose components split. Then there is a root tower $\mathcal{T}$ for $(\Phi,p+q)$ with a good origin over $\mathcal{X}$: a realisation of its flag models over the components of $\mathcal{X}$ by smooth flat reference models of degree one carrying an ample line bundle, such that near every point the automorphic line bundles give holomorphic local charts, together with a normalisation condition at finitely many primes, which we do not reproduce.*

2. There is $C\geq 0$ with the following property. Let the datum be as in (1) with moreover $E/\mathbb{Q}$ abelian and $3q+1\leq p+q$, and let $\mathcal{T}$ be a root tower with a good origin as in (1) over $\mathcal{X}$, with $|\mathrm{Bad}|\leq 4|\Phi|(p+q)^2$. Then, for every full system of arithmetic intersection

numbers on the flag models of $\mathcal{T}$ (a completeness condition on such a system, which we do not reproduce), $\mathcal{T}$ has the volume properties (V1)–(V6) with constant C and parameters

$$r = |\mathrm{disc}(E)|^{1/[E:\mathbb{Q}]}, \qquad L = \mathrm{Re}\,\frac{L'(1,\chi_{-d})}{L(1,\chi_{-d})},$$

where χ_{-d} is the quadratic character modulo the prime d of the CM host.

Proof. In outline. Both parts are proved together from the following ingredients.

1. Arakelov geometry: arithmetic intersection theory, the arithmetic Hilbert–Samuel formula, the inequality between arithmetic volume and self-intersection [91], [136], Zhang's lower bound for heights by successive minima [138], and a line bundle of large degree. From them follow the statements that the construction needs about Hodge–Tate data, GAGA for sections over $\mathbb{C}$, pull-backs and Hilbert–Samuel.

2. Abelian schemes: extension of homomorphisms of semi-abelian schemes, the Hodge–Tate sequence [126], [38], quotients of abelian schemes by finite flat subgroups and the Zarhin trick [137]. Resolution of singularities and principalisation [57], [12] give the integrability of the points of the models.

3. Nefness: at rational primes $p_0 > 2$ outside a finite set and below a nef threshold, the automorphic line bundles of dominant weight are nef on the special fibres of the reference models. This follows from the degree inequality Lemma 8.7, applied to curves in the special fibres.

From these one constructs the tower and proves the proportionality (V1)–(V2), the convergence (V3), the arithmetic proportionality (V4) through a prime-by-prime comparison of arithmetic intersection numbers, and the lower bounds (V5)–(V6).

□

The positivity in characteristic p that the tower theorem needs comes from a weighted degree inequality for eigenflags of the Lie algebra of a semi-abelian scheme over a curve, which gives nefness of automorphic line bundles on special fibres.

Lemma 8.7. *Let E be a CM field with a CM host, let X_0 be a scheme and $G \to X_0$ a semi-abelian scheme with $\mathcal{O}_E$-action and symmetric cubical polarisation whose powers induce bijections on global functions. Let $p > 2$ be a prime, $k \subset K$ fields with $\mathrm{char}\, K = p$, C a curve with function field K whose places are trivial on k, and $z\colon \mathrm{Spec}\, K \to X_0$ a point extending to C such that the fibres of G at the centres of the places of C are extensions of abelian varieties by tori to which the polarisation descends (a condition we state only in this form). Let V be the Lie algebra of G at z, with its $\mathcal{O}_E$-action r and its integral structure along C, and for a K-subspace $W \subseteq V$ let $\deg_C(W)$ be the degree on C of its saturation. Let T be a finite set and $\varphi\colon T \sqcup T \to \mathrm{Hom}(\mathcal{O}_E, k)$ a split family with $\varphi(\bar\tau)(a) = \varphi(\tau)(\bar a)$. Let $j, q \geq 1$, let $w_\tau(i)$ be real weights, and let $0 = W_{\tau,0} \subseteq \cdots \subseteq W_{\tau,j}$ and $0 = L_{\tau,0} \subseteq \cdots \subseteq L_{\tau,q}$ be flags of subspaces such that $W_{\tau,j}$ and $L_{\tau,q}$ are the $\varphi(\tau)$- and $\varphi(\bar\tau)$-eigenspaces of r. Suppose that each w_τ is non-increasing on $\{1, \ldots, j+q\}$ and that there are M, g with $|w_\tau(i)| \leq M$, $w_\tau(j) - w_\tau(j+1) \geq g$ for all τ, and $4|T|M \leq (p-2)g$. Then*

$$\sum_{\tau\in T}\Big(\sum_{i=0}^{j-1} w_\tau(i+1)\big(\deg_C W_{\tau,i+1} - \deg_C W_{\tau,i}\big) + \sum_{m=0}^{q-1} w_\tau(j+m+1)\big(\deg_C L_{\tau,q-m-1} - \deg_C L_{\tau,q-m}\big)\Big) \leq 0.$$

Proof. In outline. The inequality is obtained from degree formulas for lattices on curves, the unramifiedness of the split characters and a statement on the existence of curves with given function field; it is used for the nefness of automorphic line bundles of dominant weight on special fibres in the construction of root towers.

□

With (COV) and (FUN), the two row conditions whose proofs the outline shows, the definite-side family structure of Section 6 is in place, up to the conditions (ON) and (LOW), whose source the outline does not show (Section 11). What remains is the arithmetic ingredient that ties the families: the Sha bounds.

9 The Sha bounds by Eisenstein congruences

This section outlines the proof of the window Sha bound (Theorem 9.1): every primitive character satisfies it with exponent 401. It is the arithmetic ingredient of the whole argument, used through Proposition 3.5 and Proposition 6.20. The groups to be bounded are the Tate–Shafarevich groups $Ш^1_\ell(E,k)$ of Tate twists over the abelian fields E attached to ψ and a window pair (Definition 3.4); the author describes the resulting bound as a significant weakening of Vandiver's conjecture [2], [132]. By Shapiro's lemma they decompose over the characters of $\mathrm{Gal}(E/\mathbb{Q})$ into strict Tate–Shafarevich groups of twisted Dirichlet characters $\chi\chi_{\mathrm{cyc}}^e$. A Bernoulli bound for minus parts covers the characters with $\chi(-1) = (-1)^{k+1}$; the others are sorted into six cases by the prime ℓ, the twist e and the order of χ. In the main case a nonzero class gives a residually reducible three-dimensional representation, and the length of the group is bounded by ℓ-adic valuations of congruences between the Eisenstein representation $\Psi \oplus \mathbf{1} \oplus \Psi^{-1}$ and the Galois representations of cuspidal automorphic representations of U(3); one nonzero integer, whose logarithm is polynomial in the conductor and the twist, controls all ℓ at once. Other cases are counted with congruences on GSp_4, with a unitary lift over $\mathbb{Q}(i)$ and with a vanishing statement; case (IV) and part of case (V) have no count of their own and are handled in the final step of the proof, which we do not reproduce; and the finitely many bad primes are handled by a choice of scale. The conductors of the fields at a scale x are of size $\exp(O(x^{\varepsilon/5}\log^2 x))$, so bounds polynomial in the conductor are at most $\exp(x^\varepsilon/2)$, which is what the window Sha bound asks. The known theorems used are Kato's Euler system [65], an explicit bound for second cohomology of the kind proved by Soulé [119], the theorem of Ferrero and Washington [39], [117], local Tate and Poitou–Tate duality [123], [84], Clozel's theorems on regular algebraic cuspidal representations [28], Galois representations for Hilbert modular forms [23], [127], [13], Hilbert modular forms congruent to 1 [134], Kisin's local deformation rings [68], and the Galois representations and base change of Section 10. Bounding Selmer groups of characters by congruences on U(3) goes back to Ribet's method [102] and to [9]; what looks new is the effectivity and the uniformity in ℓ.

Theorem 9.1. *Every $\psi \in \mathcal{P}$ satisfies the window Sha bound with exponent* 401.

Proof. In outline. The proof first collects the automorphic statements listed in the appendix: Galois representations for regular algebraic cuspidal representations of GL_2 over totally real fields [23], [127], [13], Clozel's results on the cohomology and the conjugates of regular algebraic cuspidal representations [28], cyclic base change [5], and weak descent (Proposition 9.18). These statements rest on the Galois representations attached to polarized regular algebraic cuspidal representations over CM fields [50], [114], [27], [22], on base change from unitary groups to GL_n

[74], [88], [63], and on further known theorems on automorphic representations of GL_n and GL_2, listed in the appendix. The bound is then obtained in two steps.

1. Reduction to the counts. The bad primes contribute within the window bound with exponent 401 (Proposition 9.17), by Poitou–Tate duality, local finiteness and Tate duality, an explicit bound for $H^2(G_{L,S}, \mathbb{Z}_\ell(e+1))$ of the kind proved by Soulé [119] and a Ferrero–Washington input [39]. Together with the minus-part Sha bound for the CM fields attached to window pairs (Proposition 9.15), this reduces the window Sha bound to a joint bound for the strict Tate–Shafarevich groups $\mathrm{Sha}^1_n(\chi, e)$ (Definition 9.2) of the characters χ attached to the Sha-hosts of ψ, sorted into the cases of Definition 9.3.

2. The counts. The minus-part bound covers the characters χ with $\chi(-1) = (-1)^{k+1}$, and the counts cover the characters of $X_{\mathfrak{h}}(\mathcal{O}, e)$. The final step combines the following counts, in each of which the sum of $\log \ell$ times the normalised lengths is at most a polynomial in the conductor and the size parameter:

   - case (I) by Eisenstein congruences on U(3) (Proposition 9.4, through Proposition 9.6);
   - cases (II) and (III) by congruences on GSp_4 (Proposition 9.12);
   - case (V) by a unitary lift (Proposition 9.13);
   - case (Z), and the vanishing of strict Sha of even Tate twists (Proposition 9.14).

   The cases without a count of their own, case (IV) and part of case (V), are handled inside the final step, together with the bound for second cohomology and the bad-prime bound of step 1, in a way we do not reproduce. For the Sha-hosts attached to window pairs at scales $Y \le x$, $\log d \le Y^{\varepsilon/5} \log^2 Y$ and $\log q \le Y^{\varepsilon/5}$, so $\log N_{\mathfrak{h}} \le 2x^{\varepsilon/5} \log^2 x + O(\log x)$. With $t \le x^{401}$, a bound of the form $(N_{\mathfrak{h}} t)^C$ is therefore $\exp\big(O(x^{\varepsilon/5} \log^2 x)\big)$, which is at most $\exp(x^{\varepsilon}/2)$ for large x.

□

9.1 The groups and the cases

The pieces are strict Tate–Shafarevich groups of a twisted Dirichlet character: the classes that are unramified everywhere and trivial at the primes dividing n. Their normalised lengths $\nu_{\mathcal{O}}(\chi, e)$ are what the counts below bound.

Definition 9.2. Let ℓ be a prime, $\mathcal{O}$ a coefficient ring, $N \ge 1$, χ a Dirichlet character modulo N with values in $\mathcal{O}^\times$ (viewed as a character of $G_{\mathbb{Q}}$), $e \ge 0$ and $n \ge 1$ integers. Put $\Psi_{\chi,e} = \chi\, \chi_{\mathrm{cyc}}^{e}$. The **strict Tate–Shafarevich group** of (χ, e) away from n is

$$\mathrm{Sha}^1_n(\chi, e) = \ker\Big(H^1\big(\mathbb{Q}, \mathcal{O}^\vee(\Psi_{\chi,e}^{-1})\big) \longrightarrow \prod_{p \mid n} H^1\big(\mathbb{Q}_p, \mathcal{O}^\vee(\Psi_{\chi,e}^{-1})\big) \times \prod_{p} H^1\big(I_p, \mathcal{O}^\vee(\Psi_{\chi,e}^{-1})\big)\Big),$$

the classes that are unramified at every prime and locally trivial at the primes dividing n. Its **length** is $\mathrm{length}_{\mathcal{O}} \mathrm{Sha}^1_n(\chi, e) \in \mathbb{N} \cup \{\infty\}$, and the **normalised length**

$$\nu_{\mathcal{O}}(\chi, e) = \mathrm{length}_{\mathcal{O}} \mathrm{Sha}^1_{\ell N}(\chi, e) \,/\, e_{\mathcal{O}}$$

(taken with ℓN, and meaningful when the length is finite). For a subgroup $H \le (\mathbb{Z}/N\mathbb{Z})^\times$, $\mathcal{X}_{\mathcal{O}}(N, H)$ is the finite set of Dirichlet characters modulo N with values in $\mathcal{O}^\times$ whose kernel is exactly H.

The bookkeeping of the cases is the development's own. A Sha-host of ψ is a triple (a,q,d); its conductor $N_{\mathfrak{h}}$ and subgroup $H_{\mathfrak{h}}$ determine the (ψ,a,q,d)-field of Definition 3.4, and the characters of that field are the $\chi \in X_{\mathfrak{h}}(\mathcal{O},e)$. Each pair (ℓ,χ) falls into one of the cases (I)–(V), (Z).

Definition 9.3. Let $\psi \in \mathcal{P}$ have modulus N_ψ and order o_ψ. Its **auxiliary primes** are: $p_K(\psi)$, the least prime $p \geq 7$ with $p \equiv 3 \pmod 4$ and $\gcd(p, N_\psi) = 1$; $h^*(\psi)$, the class number of $\mathbb{Q}(\sqrt{-p_K(\psi)})$; and $p_1(\psi)$, the least prime p that is congruent to 1 modulo the odd part of o_ψ and prime to $2N_\psi\, p_K(\psi)$. The **bad set** B_ψ consists of 3 and the odd prime factors of $o_\psi N_\psi\, p_K(\psi)\, h^*(\psi)\, p_1(\psi)$.

A **Sha-host** of ψ is a triple $\mathfrak{h} = (a,q,d)$ with $a \geq 2$ and primes $q \equiv 3 \pmod 8$, $d \equiv 7 \pmod{16}$, $\gcd(qd, N_\psi) = 1$. Its conductor is $N_{\mathfrak{h}} = \operatorname{lcm}(2^{a+2}, N_\psi)\, qd$, and $H_{\mathfrak{h}} \leq (\mathbb{Z}/N_{\mathfrak{h}}\mathbb{Z})^\times$ is the subgroup of the u with $u \equiv \pm 1 \pmod{2^{a+2}}$, $\psi(u) = 1$, and u a square modulo q and modulo d. The host is **big at** an integer $t \geq 1$ if $2t+1 < \min(q,d)$ and each of $N_\psi, o_\psi, p_K(\psi), h^*(\psi), p_1(\psi)$ is smaller than $\min(q,d)$. For a coefficient ring $\mathcal{O}$ and $e \geq 0$, $X_{\mathfrak{h}}(\mathcal{O},e)$ is the finite set of Dirichlet characters χ modulo $N_{\mathfrak{h}}$ with values in $\mathcal{O}^\times$ such that $H_{\mathfrak{h}} \subseteq \ker\chi$ and $\chi(-1) = (-1)^e$.

For a prime ℓ, write $\chi^{(\ell)}$ for the product of the part of χ of order prime to ℓ with the inverse of its ℓ-component. For $e \geq 1$ and $\chi \in X_{\mathfrak{h}}(\mathcal{O},e)$ the **cases** are:

(I) $\ell \notin B_\psi$ and $(\ell-1) \nmid 2e$;

(II) $e \geq 2$, the order of $\chi^{(\ell)}$ is even, and $\ell \in B_\psi$ or $(\ell-1) \mid 2e$;

(III) $\ell \notin B_\psi$, $(\ell-1) \mid 2e$, $\chi \neq 1$ and the order of χ is odd;

(IV) $\ell \in B_\psi$, $e \geq 2$ and the order of $\chi^{(\ell)}$ is odd;

(V) $\ell \in B_\psi$ and $e = 1$;

(Z) $\ell \notin B_\psi$, $(\ell-1) \mid 2e$ and $\chi = 1$.

Case (V) is split by the quantity $R_1(\ell,\chi) = 16\, o_\psi(\ell N_\psi)^2\, 4^{n_0(\ell,\chi)}$ against $2^{c_2(\chi)}$, where $c_2(\chi)$ is the exponent of 2 in the conductor of the 2-component of χ, and $n_0(\ell,\chi)$ is the 2-adic valuation of $\ell^f - 1$ for f the order of ℓ modulo $\operatorname{lcm}(4, \operatorname{ord}\vartheta)$, with ϑ the part of χ of order prime to ℓ multiplied by the inverse of its 2-component.

Case (I) is the main case: ℓ outside the bad set B_ψ and $(\ell-1) \nmid 2e$. The bad set is finite and depends only on ψ, so cases (IV) and (V) concern finitely many primes ℓ.

9.2 Congruences on U(3)

Count I is the bound in the main case.

Proposition 9.4. *For every $\psi \in \mathcal{P}$ there are an integer $C \geq 0$ and primes r_1, r_2 such that the following holds. Let $\mathfrak{h} = (a,q,d)$ be a Sha-host of ψ with $q, d > \max(r_1, r_2)$, let $t \geq 1$ be such that $\mathfrak{h}$ is big at t, let $0 < e < t$, let L be a finite set of odd primes ℓ in case (I) for e, and let $\mathcal{O}_\ell$ be a coefficient ring over $\mathbb{Z}_\ell$ for each $\ell \in L$. Then $\mathrm{Sha}^1_{\ell N_{\mathfrak{h}}}(\chi, e)$ has finite length for all $\ell \in L$ and $\chi \in X_{\mathfrak{h}}(\mathcal{O}_\ell, e)$, and*

$$\sum_{\ell \in L} \log \ell \sum_{\chi \in X_{\mathfrak{h}}(\mathcal{O}_\ell, e)} \nu_{\mathcal{O}_\ell}(\chi, e) \leq \big(N_{\mathfrak{h}}\, t\big)^C.$$

Proof. In outline. The automorphic hypotheses of the bound for one character behind Proposition 9.6 are known theorems on automorphic representations, listed in the appendix. Apply Proposition 9.6 with $K = \mathbb{Q}(\sqrt{-p_K(\psi)})$, $N = N_{\mathfrak{h}}$, and the subgroups $H \supseteq H_{\mathfrak{h}}$ that occur as kernels of characters in $X_{\mathfrak{h}}(\mathcal{O}_\ell, e)$, with two fixed primes $r_1 < r_2$ larger than $30N_\psi p_K(\psi)$. For ℓ in case (I) and χ with $\ell \nmid \operatorname{ord}\chi$, the quadruple (K, ℓ, e, χ) is Eisenstein-admissible (Definition 9.5); bigness of $\mathfrak{h}$ guarantees that r_1, r_2 do not divide $30N_{\mathfrak{h}}d_K$. Summing over the possible kernels H, the right-hand side $2c_5\varphi(h)(N(e+1))^A(\log r_1 + \log r_2)$ of Proposition 9.6 is at most $(N_{\mathfrak{h}}t)^C$ for a suitable C depending on ψ.

□

It is deduced from a bound over all characters with a given kernel, Proposition 9.6, applied over the auxiliary imaginary quadratic field $\mathbb{Q}(\sqrt{-p_K(\psi)})$. The congruences are formed over an imaginary quadratic field K under the following conditions.

Definition 9.5. Let K be a number field, ℓ a prime, $e \geq 0$, $\mathcal{O}$ a coefficient ring and χ a Dirichlet character modulo N with values in $\mathcal{O}^\times$. Write $\chi = \prod_{p|N} \chi_p$ with χ_p a character of $(\mathbb{Z}/p^{v_p(N)}\mathbb{Z})^\times$, and $\chi' = \chi\chi_\ell^{-1}$. The quadruple (K, ℓ, e, χ) is **Eisenstein-admissible** if

(Ea1) $\ell \neq 2$, $e \geq 1$, $(\ell - 1) \nmid 2e$, N is even and $\ell \nmid \operatorname{ord}(\chi)$;

(Ea2) $\chi_\ell^2 = 1$ and $\chi(-1) = (-1)^e$;

(Ea3) K is an imaginary quadratic field with $\ell \nmid d_K h_K$ and $\gcd(d_K, N) = 1$, where d_K is the discriminant and h_K the class number;

(Ea4) $\ker(\chi_{\mathrm{cyc}} \bmod \ell) \cap \ker\chi' \not\subset G_K$ in $G_{\mathbb{Q}}$; that is, K is not contained in the compositum of $\mathbb{Q}(\mu_\ell)$ and the abelian field cut out by χ'.

Proposition 9.6. *For every number field K there are integers $c_5, A \geq 1$ with the following property. Let $e, N \geq 1$, $H \leq (\mathbb{Z}/N\mathbb{Z})^\times$ of index h, and $p \neq p'$ primes not dividing $30Nd_K$. Let L be a finite set of primes and, for each $\ell \in L$, let $\mathcal{O}_\ell$ be a coefficient ring over $\mathbb{Z}_\ell$ such that (K, ℓ, e, χ) is Eisenstein-admissible for some $\chi \in \mathcal{X}_{\mathcal{O}_\ell}(N, H)$. Then $\mathrm{Sha}^1_{\ell N}(\chi, e)$ has finite length for all $\ell \in L$ and $\chi \in \mathcal{X}_{\mathcal{O}_\ell}(N, H)$, and*

$$\sum_{\ell \in L} \log \ell \sum_{\chi \in \mathcal{X}_{\mathcal{O}_\ell}(N,H)} \nu_{\mathcal{O}_\ell}(\chi, e) \leq 2c_5\,\varphi(h)\,\big(N(e+1)\big)^A\big(\log p + \log p'\big).$$

Proof. Let M and M' be the integers given by Proposition 9.7 for p and for p'; the automorphic statements needed there are known theorems on automorphic representations, listed in the appendix. For $\ell \in L$ with $\ell \neq p$ use M, and for $\ell = p$ (so $\ell \neq p'$) use M'; in both cases $\sum_\chi \nu_{\mathcal{O}_\ell}(\chi, e) \leq 2v_\ell(M) + 2v_\ell(M')$. Multiplying by $\log \ell$ and summing over $\ell \in L$ gives at most $2(\log|M| + \log|M'|)$, since $\sum_\ell v_\ell(M)\log\ell \leq \log|M|$. The bounds on $\log|M|$ and $\log|M'|$ give the claim. If L is empty there is nothing to prove. □

That bound comes from congruence numbers: for each auxiliary prime p there is one integer M whose ℓ-adic valuations control all characters with the given kernel at once.

Proposition 9.7. *For every number field K there are integers $c_5, A \geq 1$ with the following property. Let $e, N \geq 1$, $H \leq (\mathbb{Z}/N\mathbb{Z})^\times$ a subgroup of index h, and p a prime with $p \nmid 30Nd_K$. Then there is a nonzero integer M with*

$$\log|M| \leq c_5\,\varphi(h)\,\big(N(e+1)\big)^A \log p$$

such that for every prime $\ell \neq p$ and every coefficient ring $\mathcal{O}$ over $\mathbb{Z}_\ell$: if (K, ℓ, e, χ) is Eisenstein-admissible for some $\chi \in \mathcal{X}_{\mathcal{O}}(N, H)$, then $\mathrm{Sha}^1_{\ell N}(\chi', e)$ has finite length for every $\chi' \in \mathcal{X}_{\mathcal{O}}(N, H)$ and

$$\sum_{\chi' \in \mathcal{X}_{\mathcal{O}}(N,H)} \nu_{\mathcal{O}}(\chi', e) \leq 2\, v_\ell(M).$$

Proof. In outline. If K is not imaginary quadratic, no quadruple (K, ℓ, e, χ) is Eisenstein-admissible and $M = 1$ works. Let K be imaginary quadratic; since $p \geq 7$, $p \nmid N$ and $p \nmid d_K$, choose a place w of K above p. Apply Proposition 9.10 to the family indexed by the complex characters χ modulo N with kernel H, with $z_\chi = \chi(\mathrm{N}w)^{-1}$; this family is permuted by $\mathrm{Aut}(\mathbb{C})$, and its size is at most $\varphi(h)$. This gives M with the stated bound. For $\ell \neq p$ and $\chi' \in \mathcal{X}_{\mathcal{O}}(N, H)$, apply Proposition 9.9 with σ the inverse of a Frobenius at w: then t is the value $\iota^{-1}(t_\chi)$ of Proposition 9.10 for the corresponding complex character, condition (Cn1) gives the hypothesis of Proposition 9.9, and summing its bound over χ' and using (Cn2) gives $\sum \nu_{\mathcal{O}}(\chi', e) \leq 2v_\ell(M)$. □

The bound for one character compares the Tate–Shafarevich group with congruences between the Eisenstein representation and the three-dimensional Galois representations of automorphic representations of U(3) over K of level N and weight e. The admissible ones are defined next, in outline.

Definition 9.8. Let K be an imaginary quadratic field, ℓ a prime, $\iota\colon \overline{\mathbb{Q}}_\ell \to \mathbb{C}$, and $e, N \geq 1$. An ℓ-adic representation $r\colon G_F \to \mathrm{GL}_n(\overline{\mathbb{Q}}_\ell)$ of a number field F has **level** M **and weights** $m = (m_1, \ldots, m_n)$ if

(Lv1) at every finite place $w \nmid \ell$ of F, the restriction of r to an inertia group I_w is diagonalisable, with diagonal characters that are trivial on $G_{F(\mu_M)} \cap I_w$;

(Lv2) at every place $w \mid \ell$ of F, the restriction of r to a decomposition group G_{F_w} is conjugate to an upper triangular representation with diagonal characters $\kappa_1, \ldots, \kappa_n$ such that $\kappa_i|_{I_w} = \chi_{\mathrm{cyc}}^{m_i}\eta_i|_{I_w}$ for characters η_i of G_F trivial on $G_{F(\mu_M)}$, and either $\kappa_i \neq \kappa_j\chi_{\mathrm{cyc}}$ for all i, j, or $r|_{G_{F_w}}$ is potentially crystalline.

The representation r is **unlinked of level** M **and weights** m if at every $w \mid \ell$ there is such a flag with $\kappa_i \neq \kappa_j\chi_{\mathrm{cyc}}$ for all i, j. A representation $\rho\colon G_K \to \mathrm{GL}_3(\overline{\mathbb{Q}}_\ell)$ is **admissible of stable type** for (ι, e, N) if ρ is irreducible, $\rho \otimes \chi_{\mathrm{cyc}}^{-1}$ is automorphic of conjugate self-dual type, ρ has level N and weights $(e, 0, -e)$, and ρ is unlinked of level N and weights $(e, 0, -e)$. It is **admissible of endoscopic type** for (ι, e, M) if there is an irreducible $W\colon G_{\mathbb{Q}} \to \mathrm{GL}_2(\overline{\mathbb{Q}}_\ell)$ with

(En1) $W \otimes \chi_{\mathrm{cyc}}^{-e} \cong \rho_{f,\iota}$ for a normalised newform f of weight $2e + 1$;

(En2) W has level M and weights $(e, -e)$, and $\det W$ is trivial on $G_{\mathbb{Q}(\mu_{M'})}$ for some $M' \geq 1$ (a Dirichlet character, cyclotomic weight 0);

(En3) ρ is conjugate to $s \mapsto \begin{pmatrix} a(s) & 0 & b(s) \\ 0 & 1 & 0 \\ c(s) & 0 & d(s) \end{pmatrix}$, where $W|_{G_K} = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$;

(En4) for every imaginary quadratic field over which W stays irreducible, the restriction of W satisfies a further automorphy condition, which we do not reproduce.

For the bounds below, $\mathcal{A}_3(K,\iota,e,N)$ denotes the set of representations of G_K that are admissible of stable type for (ι,e,N) or of endoscopic type for $(\iota,e,N\,|d_K|)$.

Proposition 9.9. *Let (K,ℓ,e,χ) be Eisenstein-admissible, with χ a Dirichlet character modulo N with values in a coefficient ring $\mathcal{O}$, and let $\iota\colon \overline{\mathbb{Q}}_\ell \to \mathbb{C}$. Let $\sigma \in G_K$ and*

$$t = \Psi_{\chi,e}(\sigma) + 1 + \Psi_{\chi,e}(\sigma)^{-1} \in \mathcal{O},$$

the trace at σ of the reducible representation $\Psi_{\chi,e} \oplus \mathbf{1} \oplus \Psi_{\chi,e}^{-1}$ restricted to G_K. Suppose that $\operatorname{tr}\rho(\sigma) \neq t$ for every $\rho \in \mathcal{A}_3(K,\iota,e,N)$. Then $\mathrm{Sha}^1_{\ell N}(\chi,e)$ has finite length, and there is a finite set $X \subset \mathcal{A}_3(K,\iota,e,N)$ of pairwise non-isomorphic representations such that

$$\nu_{\mathcal{O}}(\chi,e) \leq 2\sum_{\rho\in X} v_\ell\big(\operatorname{tr}\rho(\sigma) - t\big).$$

Proof. In outline. One first passes to a coefficient ring with $e_{\mathcal{O}} = 1$ that contains the values of χ; this is possible because $\ell \nmid \operatorname{ord}(\chi)$, and both sides of the inequality are unchanged. The bound is then obtained by an Eisenstein congruence argument for $\mathrm{U}(3)$ over K. A residually reducible deformation problem of type $\Psi_{\chi,e} \oplus \mathbf{1} \oplus \Psi_{\chi,e}^{-1}$ is set up. The length of $\mathrm{Sha}^1_{\ell N}(\chi,e)$ is bounded by twice the length of the congruence module of its reducible locus, and that is bounded by the valuations $v_\ell(\operatorname{tr}\rho(\sigma) - t)$ over its automorphic points. The ingredients are:

1. the vanishing of the Bloch–Kato Selmer groups of the two twists of newforms of weight $2e+1$, by Kato's Euler system [65];
2. an explicit bound for the second cohomology of Tate twists of the kind proved by Soulé [119], together with local Tate duality [123], Poitou–Tate duality and the global Euler characteristic formula [84], which control the cohomology of the residual pieces;
3. two rigidity statements for the deformation rings at reducible points, a point-counting lemma, and the identification of the automorphic points of type I with admissible representations of stable type; these use known theorems on automorphic representations, listed in the appendix;
4. automorphy of the points of the deformation ring over unramified coefficients, a lifting statement for automorphy obtained by patching with Kisin's local deformation rings.

□

This is the crux of the Sha bounds. A nonzero class in $\mathrm{Sha}^1_{\ell N}(\chi,e)$ gives a residually reducible representation, and the length is bounded by twice the valuations $v_\ell(\operatorname{tr}\rho(\sigma) - t)$ of the congruences. The direction is the reverse of the classical use of Eisenstein congruences, which construct Selmer classes [102], [82]: here the congruences bound the group from above. The congruence numbers that make the bound effective are the following.

Proposition 9.10. *Let K be an imaginary quadratic field. There are integers $c, A \geq 1$ with the following property. Let $e, N \geq 1$, let $p \geq 7$ be a prime with $p \nmid N d_K$, let w be a place of K above p with norm $q_w = \mathrm{N}w$, and let $(z_o)_{o\in O}$ be a finite family of roots of unity in $\mathbb{C}$ that is permuted by every automorphism of $\mathbb{C}$. Put*

$$t_o = q_w^{-e} z_o + 1 + q_w^{e} z_o^{-1} \qquad (o \in O).$$

Then there is a nonzero integer M with

$$\log|M| \le c\,\#O\left(N(e+1)\right)^A \log p$$

such that for every prime $\ell \notin \{2,p\}$, every $\iota\colon \overline{\mathbb{Q}}_\ell \to \mathbb{C}$, every Frobenius element σ at a place above w, and every family $(X_o)_{o\in O}$ of finite sets of pairwise non-isomorphic representations in $\mathcal{A}_3(K,\iota,e,N)$:

(Cn1) *$\operatorname{tr}\rho(\sigma^{-1}) \ne \iota^{-1}(t_o)$ for all o and all $\rho \in X_o$;*

(Cn2) *$|M|_\ell \le \prod_{o\in O}\prod_{\rho\in X_o}\left|\operatorname{tr}\rho(\sigma^{-1}) - \iota^{-1}(t_o)\right|_\ell$.*

Proof. In outline. Two counting statements are combined, both derived from known theorems on automorphic representations, listed in the appendix. The first bounds, polynomially in $N(e+1)$, the set of characteristic polynomials of $\rho(\sigma^{-1})$ for ρ admissible of stable type; it comes from a count of the systems of Hecke eigenvalues of the corresponding conjugate self-dual automorphic representations of $\mathrm{GL}_3(\mathbb{A}_K)$ of level N and weight $(e,0,-e)$. The second does the same for the endoscopic type, from a count of newforms of weight $2e+1$ and level dividing $N|d_K|$. The integer M is built from the finitely many possible Frobenius traces and the values t_o. The constants are $c = 5(c_1 + c_2|d_K|^{A_2})$ and $A = A_1 + A_2 + 1$, with c_i, A_i the constants of the two counts.

□

Count I also uses ℓ-adic measures built from products of Eisenstein series, for ordinary newforms of odd weight, in the manner of [56]; they belong to the part of the argument that uses Kato's Euler system [65].

Proposition 9.11. *For an integer $k \ge 2$, a prime ℓ, functions $u, w\colon \mathbb{Z} \to \mathbb{C}$ and $F\colon \mathbb{N}\times\mathbb{N} \to \mathbb{C}$ let $Z_{k,\ell}(u,w;F)$ be the q-series with coefficients*

$$Z_{k,\ell}(u,w;F)_n = \sum_{\substack{n_1+n_2=n\\ \ell\nmid n_1 n_2}} \sum_{x_1y_1=n_1} \sum_{x_2y_2=n_2} x_1^{k-2}\,u(x_1)\,w(x_2)\,F(x_1,y_2),$$

and for a linear functional φ on q-series, $\alpha\in\mathbb{C}$, $m\ge 1$ and $f\colon \mathbb{Z}/\ell^m \to \mathbb{C}$ let $\mu_\varphi(f)$ be the value at level m, normalised by α, of φ on $Z_{k,\ell}(u,w;(A,d)\mapsto f(Ad^{-1}))$. Let $\ell\ne 2$, $e\ge 1$, and let g be a newform of weight $2e+1$, level M prime to ℓ and nebentypus χ_g, ordinary at ℓ: $|\iota^{-1}(a_\ell(g))| = 1$. Let ψ_ℓ be a Dirichlet character modulo ℓ and $N_0 \ge 1$ prime to ℓ, and put $\epsilon = (-1)^e\psi_\ell(-1)$. Then there are an integer $N \ge 5$ with $M \mid N$, $N_0 \mid N$, $\ell\nmid N$, Dirichlet characters χ,θ modulo N, a root α of $X^2 - a_\ell(g)X + \chi(\ell)\ell^{2e}$ with $|\iota^{-1}(\alpha)| = 1$, a Hecke-eigen functional φ of level $N\ell$ with eigenvalues $a_q(g)$ and nebentypus χ that kills forms of other nebentypus, a nonzero $v\in\mathbb{C}$ and integers $c_1, d_1 > 1$ such that:

(Em1) *χ agrees with χ_g on integers prime to N, $\chi(-1) = -1$ and $\theta(-1) = \epsilon$;*

(Em2) *with u the function attached to θ and $w(B) = \mathbf{1}[B\equiv 1] - \epsilon\,\mathbf{1}[B\equiv -1] \pmod N$, the values μ_φ on ϵ-parity functions form a distribution on $\varprojlim \mathbb{Z}/\ell^m$ which is bounded ℓ-adically, takes values in a fixed finite extension of $\mathbb{Q}_\ell$ on functions with values there, and interpolates v: there is C with $|\iota^{-1}(\mu_\varphi(f) - v)| \le C\delta$ whenever f has parity ϵ and $|\iota^{-1}(f(ad^{-1}) - \psi_\ell(a)\overline{\psi_\ell}(d)(d/a)^e)| \le \delta$ for all a,d prime to ℓ;*

(Em3) *c_1, d_1 are prime to $6N\ell$, c_1^e is not of absolute value 1, and $d_1^2 \ne d_1^{e+1}(\theta\chi^{\pm1})(d_1)^{n_1}\psi_\ell(d_1)^{n_2}$ for $n_1 = \pm1$ and all $n_2\in\mathbb{Z}$.*

The normalisation of μ_φ at level m and the function attached to θ (a normalised Fourier transform) are fixed by conditions of the same nature, which we do not reproduce.

Proof. In outline. The Hecke-eigen functional is built from Atkin–Lehner operators, which intertwine the Hecke operators of the different nebentypus characters, and the measure is obtained from it on products of Eisenstein series, in the manner of [56].

□

9.3 The other cases

The other cases follow the same pattern with other groups. Counts II and III use congruences on GSp_4 with representations of dense image, and count V uses a unitary lift over $\mathbb{Q}(i)$ of a reduced residual representation; it covers the characters of case (V) with $R_1(\ell,\chi) \le 2^{c_2(\chi)}$.

Proposition 9.12. *For every $\psi \in \mathcal{P}$ there is an integer $C \ge 0$ such that for every Sha-host $\mathfrak{h}$ of ψ, every $t \ge 1$ at which $\mathfrak{h}$ is big, every odd prime ℓ, every coefficient ring $\mathcal{O}$ over $\mathbb{Z}_\ell$, every $0 < e < t$ and every $\chi \in X_{\mathfrak{h}}(\mathcal{O}, e)$ in case (II) or in case (III), the group $\mathrm{Sha}^1_{\ell N_{\mathfrak{h}}}(\chi, e)$ has finite length and*

$$\nu_{\mathcal{O}}(\chi, e)\, \log \ell \le \big(\ell\, N_{\mathfrak{h}}\, e\big)^{C}.$$

Proof. In outline. The bound comes from a congruence argument for GSp_4, parallel to the U(3) argument for case (I). The ingredients are:

1. automorphy of the relevant symplectic four-dimensional representations with dense image, which rests on known theorems on automorphic representations listed in the appendix, the Ferrero–Washington bound for the ℓ-rank of class groups in cyclotomic towers [39] and the existence of Hilbert modular forms congruent to 1 [108], [134];
2. adjoint lifting, a rigidity statement for Sp_4, and a count of congruence numbers for Sp_4;
3. an explicit bound for the second cohomology of Tate twists of the kind proved by Soulé [119], for the cohomology of the residual pieces, together with Poitou–Tate and local duality [84];
4. a nonvanishing statement for residual Bernoulli numbers.

□

Proposition 9.13. *For every $\psi \in \mathcal{P}$ there is an integer $C \ge 0$ such that for every Sha-host $\mathfrak{h}$ of ψ, every $t \ge 1$ at which $\mathfrak{h}$ is big, every odd prime ℓ, every coefficient ring $\mathcal{O}$ over $\mathbb{Z}_\ell$, every $0 < e < t$ and every $\chi \in X_{\mathfrak{h}}(\mathcal{O}, e)$ in case (V) with $R_1(\ell,\chi) \le 2^{c_2(\chi)}$, the group $\mathrm{Sha}^1_{\ell N_{\mathfrak{h}}}(\chi, e)$ has finite length and*

$$\nu_{\mathcal{O}}(\chi, e)\, \log \ell \le \big(\ell\, N_{\mathfrak{h}}\, e\big)^{C}.$$

Proof. In outline. The count comes from a unitary lift, over the field $\mathbb{Q}(i)$, of a reduced residual representation. A core statement asserts two things for the host frames attached to (χ, e): the universal three-dimensional deformation ring has finite weight-0 part, and its points of type I are automorphic for the unitary group over $\mathbb{Q}(i)$. It is combined with

1. local finiteness, local Tate duality and the local Euler characteristic, the global finiteness and Euler characteristic, and Poitou–Tate duality [123], [84];

2. the vanishing of the relevant host cohomology, which rests on an explicit bound for the second cohomology of Tate twists of the kind proved by Soulé [119];

3. a nonvanishing statement for residual Bernoulli numbers;

4. host rigidity and the congruence numbers of Proposition 9.10;

5. known theorems on automorphic representations, listed in the appendix.

□

In case (Z) the group vanishes.

Proposition 9.14. *Let $\ell \geq 5$ be a prime, $\mathcal{O}$ a coefficient ring over $\mathbb{Z}_\ell$, and $e \geq 2$ even with $(\ell - 1) \mid 2e$. Then $\mathrm{Sha}^1_\ell(\mathbf{1}, e) = 0$, that is, every class in $H^1\big(\mathbb{Q}, \mathcal{O}^\vee(\chi_{\mathrm{cyc}}^{-e})\big)$ that is unramified everywhere and trivial at ℓ vanishes. Consequently, for every Sha-host $\mathfrak{h}$ of every $\psi \in \mathcal{P}$, every t at which $\mathfrak{h}$ is big, every odd ℓ, every $0 < e < t$ and the character $\chi = \mathbf{1} \in X_\mathfrak{h}(\mathcal{O}, e)$ in case (Z), $\mathrm{Sha}^1_{\ell N_\mathfrak{h}}(\mathbf{1}, e) = 0$.*

Proof. In outline. The vanishing splits into two cases. The first case is proved more generally: for every odd prime ℓ and every $e \geq 1$ with $(\ell - 1) \mid e$, the strict Tate–Shafarevich group $\mathrm{Sha}^1_\ell(\mathbf{1}, e)$ vanishes. If $(\ell - 1) \mid 2e$ but $(\ell - 1) \nmid e$, an auxiliary quadratic field is used. The consequence for hosts is formal. In case (Z) the character $\chi = \mathbf{1}$ lies in $X_\mathfrak{h}(\mathcal{O}, e)$ only if e is even, and $\ell \notin B_\psi$ gives $\ell \neq 3$, so $\ell \geq 5$. Moreover $\mathrm{Sha}^1_{\ell N_\mathfrak{h}}(\mathbf{1}, e) \subseteq \mathrm{Sha}^1_\ell(\mathbf{1}, e)$, because more local conditions are imposed.

□

The minus parts of the Tate–Shafarevich groups of CM fields, on which complex conjugation acts by -1, are the sums of the isotypic parts of the characters with $\chi(-1) = (-1)^{k+1}$. A Bernoulli bound for isotypic parts controls them, in the spirit of the theorems of Herbrand–Ribet and Mazur–Wiles [102], [82]; the arithmetic input that produces the Bernoulli number is not identified in the outline.

Proposition 9.15. *For a CM field E write $\mathrm{len}^-_\ell(E, k)$ for the length of the part of $\mathrm{Sha}^1_\ell(E, k)$ on which complex conjugation acts by -1, for its natural action on cohomology with coefficients $\mathbb{Q}_\ell/\mathbb{Z}_\ell(-k)$ (conjugation of cocycles combined with the action on the coefficients). There is $C > 0$ such that the following holds. Let M be an abelian number field of degree o, let $e, t \geq 2$, let (q, d) be primes with $q \equiv 3 \pmod 8$, $d \equiv 7 \pmod{16}$ and qd prime to the conductor of M, and let E be a CM field attached to (M, e, q, d) (an abelian field of the shape of the (ψ, e, q, d)-fields of Definition 3.4, with conditions we do not reproduce), of degree n and conductor N_E. Then $\mathrm{len}^-_\ell(E, k) < \infty$ for all odd primes ℓ and $0 < k < t$, and for every finite set S of odd primes*

$$\sum_{\ell \in S} \log \ell \sum_{k=1}^{t-1} \mathrm{len}^-_\ell(E, k) \leq 3nt^2 \log(tN_E) + C \sum_{\substack{\ell \mid o \\ \ell \text{ odd}}} \ell\, n^2 t^2 \log(t\ell N_E)^2.$$

Proof. In outline. Three ingredients are combined: a degree computation for the fields attached to (M, e, q, d); a Stickelberger-type bound with an explicit constant; and a bound at each prime, obtained from Proposition 9.16 by decomposing the minus part into isotypic parts and estimating the Bernoulli numbers. With the action on the coefficients included, complex conjugation acts on the ω-isotypic part by $\chi_\omega(-1)(-1)^k$, so the minus part is the sum of the ω-parts with $\chi_\omega(-1) = (-1)^{k+1}$, which is the parity required in Proposition 9.16.

□

Proposition 9.16. *Let E be an abelian CM field, realised in $\mathbb{Q}(\zeta_N)$ so that $(\mathbb{Z}/N\mathbb{Z})^\times$ acts on E through the cyclotomic action; let ℓ be an odd prime with $\ell \nmid [E : \mathbb{Q}]$, $\mathcal{O}$ a coefficient ring over $\mathbb{Z}_\ell$, $e \geq 1$, and ω a character of $\mathrm{Gal}(E/\mathbb{Q})$ with values in $\mathcal{O}^\times$ whose associated Dirichlet character χ_ω modulo N satisfies $\chi_\omega(-1) = (-1)^{e+1}$. Let $\mathrm{Sha}^1_\ell(E, e)_\omega$ be the ω-isotypic part of the group of classes in $H^1\big(G_E, \mathcal{O}^\vee(\chi_{\mathrm{cyc}}^{-e})\big)$ that are unramified everywhere and trivial at the places above ℓ, with $\mathrm{Gal}(E/\mathbb{Q})$ acting by conjugation. Then*

$$\mathrm{length}_{\mathcal{O}} \mathrm{Sha}^1_\ell(E, e)_\omega \leq v_{\mathcal{O}}\big(\delta_N(e+1)\, B_{e+1,\chi_\omega}\big),$$

where B_{e+1,χ_ω} is the generalised Bernoulli number of the primitive character inducing χ_ω, $\delta_N(e+1)$ is a normalising integer depending only on N and $e+1$, and $v_{\mathcal{O}}$ is the valuation of $\mathcal{O}$ with $v_{\mathcal{O}}(\varpi) = 1$ for a uniformiser ϖ.

Proof. In outline. The bound is derived from a reciprocity statement for the ω-part, whose proof uses Poitou–Tate duality [84] and the orthogonality of unramified classes under local duality. Duality alone cannot produce the Bernoulli number; the arithmetic input that does (of the kind of the main conjecture, or of an Euler system of Gauss sums or cyclotomic units) we have not identified.

□

9.4 Bad primes and automorphy

For the primes of the bad set the bound is obtained by a choice of scale. An explicit bound for the second cohomology of Tate twists of the kind proved by Soulé [119], which is doubly exponential, and the bounded ℓ-rank of class groups in cyclotomic towers, a form of the theorem of Ferrero and Washington [39], [117], give ranges of scales on which these primes contribute within the bound. This is where the ranges $[Y, Y^2]$ of the window Sha bound come from.

Proposition 9.17. *There is a family of constants (a base rank $R_0(M)$ for each abelian number field M, ranks $R_\ell(L)$ for primes ℓ and abelian fields L, and a constant $c_{\mathrm{B}} > 0$) with the following property, for every $C_t \geq 1$. Let M be an abelian number field, B' a finite set of odd primes and, for $\ell \in B'$, L_ℓ an abelian field containing μ_ℓ. Put $R = R_0(M) + \sum_{\ell \in B'} R_\ell(L_\ell)$. Then for all $\varepsilon \in (0, 1/52)$ and A_1, putting $J = 2^{R+2}(2c_{\mathrm{B}}C_t/\varepsilon)^R$, there is X_0 such that for all admissible page and rank functions e, t (for A_1, C_t), all $X \geq X_0$ and all window pairs (q, d) for M with $\log(qd) \leq X^{\varepsilon/(4J)}$ there is Y with $X^{1/J} \leq Y$ and $Y^2 \leq X$ such that for every $x \in [Y, Y^2]$:*

(Wb1) *every field E attached to $(M, e(x), q, d)$ satisfies a two-place bounded-height condition with rank $t(x)$ and size $\exp(x^{4\varepsilon/5})$;*

(Wb2) *for every $\ell \in B'$ and every field L_1 in the layer of index $\lfloor \log_\ell R_\ell(L_\ell) \rfloor + 3$ of the cyclotomic $\mathbb{Z}_\ell$-tower over an abelian field of type L_ℓ, the ℓ-part of the Sha bound holds: $\sum_{k<t(x)} \mathrm{len}_\ell(L_1, k) \log \ell \leq \exp(x^{4\varepsilon/5})$ with finiteness.*

Here admissibility of (e,t), the fields attached to $(M,e(x),q,d)$, the two-place bounded-height condition and the layers are subject to further conditions of the same nature, which we do not reproduce.

Proof. In outline. The windows come from combining an explicit bound for the second cohomology of Tate twists of the kind proved by Soulé [119], the bounded-rank form of the Ferrero–Washington theorem [39], and Poitou–Tate and local duality [84]. Bounded ℓ-rank in the cyclotomic towers gives a bound that does not depend on the layer, through a fixed-level and a two-step statement. The bound of Soulé type controls the second cohomology, and hence the Sha lengths, by an explicit function of degree and discriminant. A choice of the window $[Y,Y^2]$ inside $[X^{1/J},X^{1/2}]$ then gives (Wb1) and (Wb2) simultaneously.

□

All counts rest on known theorems on automorphic representations, listed in the appendix: Clozel's purity and conjugation theorems for regular algebraic cuspidal representations [28], Galois representations for GL_2 over totally real fields [23], [127], [13], cyclic base change [5], and the Galois representations and base change of Section 10. One further automorphic statement is proved in the argument, a weak descent from GL_n to definite unitary groups in rank at most four.

Proposition 9.18. *Let E be a CM field with E/E^+ unramified at all finite places, p a prime, and S a finite set of finite places of E^+ that contains all places above p and splits completely in E. Let $2\le n\le 4$ and let π be a regular algebraic conjugate self-dual cuspidal automorphic representation of $\mathrm{GL}_n(\mathbb{A}_E)$ that is unramified at all places not above S. Then there is an automorphic representation σ of the definite unitary group $\mathrm{U}(n)$ of the form $\sum_i x_i\bar{y}_i$ on E^n, unramified at all non-split places, such that for all but finitely many split places w of E the local component of σ at w has the same Satake parameters as π_w.*

Proof. In outline. The descent is obtained from a split-rigidity statement for the trace formula of the unitary group, a comparison of the spectral sides through split places, which holds in all ranks $n\ge 1$.

□

This completes the outline of Theorem 9.1, the last of the five results of Section 2. The next section describes how the argument obtains the Galois representations and the base change on which the counts rest.

10 Galois representations by theta lifts

The Sha bounds and the family structure use Galois representations attached to polarized regular algebraic cuspidal representations of GL_a over CM fields, with local–global compatibility. This is a known theorem [50], [114], [27], [22], and the main argument could use it as such. The argument reaches it by a different route, which this section outlines (Theorem 10.1). It works first with definite unitary groups of odd rank. There Galois representations are obtained for automorphic representations of a special type and lacunary weight (Proposition 10.7, Proposition 10.5), by a step whose mechanism we do not know and which, judging from the names of its inputs, may be geometric, and then for all automorphic representations as ℓ-adic limits of congruent systems (Proposition 10.6, Proposition 10.4). Even rank is reached by two theta lifts to odd rank with distinct auxiliary characters, whose difference, as the statement suggests, isolates the

representation sought (Proposition 10.3); it rests on stable-range theta lifting (Proposition 10.2). The known theorems used are stable-range theta lifting and its nonvanishing [79], the local theta correspondence at split places [85], the holomorphic Lefschetz formula [7], comparison theorems of p-adic Hodge theory [36], the local Langlands correspondence [50], [55], and the finite covolume of unitary groups [17]. The section ends with base change from definite unitary groups to GL_n [74], [88], the other known automorphic theorem that the Sha bounds use, which the argument derives from candidate lifts (Proposition 10.9).

Theorem 10.1. *Let E be a CM field, ℓ a prime, $\iota\colon \overline{\mathbb{Q}}_\ell \to \mathbb{C}$, $a \geq 1$, π a regular algebraic cuspidal automorphic representation of $\mathrm{GL}_a(\mathbb{A}_E)$ and χ a Hecke character of E^+ such that (π, χ) is a polarized pair in the sense of [8] (that is, $\pi^c \cong \pi^\vee \otimes (\chi \circ \mathrm{N}_{E/E^+} \circ \det)$, together with a sign condition at the real places). Then there are a continuous semisimple $r\colon G_E \to \mathrm{GL}_a(\overline{\mathbb{Q}}_\ell)$, an integer w_0 and multisets HT_τ of integers, for the embeddings $\tau\colon E \to \overline{\mathbb{Q}}_\ell$, satisfying the conditions (Gp1)–(Gp4) of the Galois representations attached to polarized pairs [50], [114], [27], [22]: (Gp1) at every finite place $w \nmid \ell$ the Frobenius-semisimple Weil–Deligne representation attached to r is compatible with the local Langlands correspondence applied to π_w; (Gp2) $\mathrm{HT}_\tau = \{\frac{a-1}{2} - z\}$, with z running over the infinitesimal character of π at $\iota \circ \tau$; (Gp3) $\mathrm{HT}_{\tau \circ c} = \{w_0 - h : h \in \mathrm{HT}_\tau\}$ for complex conjugation c; (Gp4) at the places above ℓ, r satisfies a compatibility condition with the local data of π in terms of ℓ-adic Hodge theory (a crystallinity statement together with a condition on the associated Weil–Deligne data), whose precise form we do not reproduce.*

Proof. In outline. The proof has three stages.

1. A core statement for unitary groups: every automorphic representation of the definite unitary group in a variables over E/E^+, of any rank, has an attached a-dimensional Galois representation with the local properties at good split places and at places above ℓ. For odd rank this is Proposition 10.4. For even rank it is the subtraction argument of Proposition 10.3: two theta lifts (Proposition 10.2) to a unitary group of odd rank with distinct auxiliary characters, followed by subtraction of the auxiliary part. A closedness statement for Frobenius-characteristic data under congruences and a descent statement away from a finite set of places complete this stage. Throughout, the space of automorphic forms of given level and type on these unitary groups is finite-dimensional, by the finite covolume of arithmetic subgroups given by reduction theory [17].

2. The core statement is upgraded to all places: Galois compatibility on the unlinked support of spherical representations, and local-global compatibility at Iwahori-fixed places for unlinked parameters.

3. For a polarized pair (π, χ) on $\mathrm{GL}_a(\mathbb{A}_E)$, the unitary statement gives r with (Gp1)–(Gp4).

□

The core statement for definite unitary groups is proved separately in odd and even rank; even rank is reduced to odd rank by theta lifting.

Proposition 10.2. *Let E be a CM field with maximal totally real subfield F, such that E/F is unramified at all finite places and the inert finite places v of F satisfy $v \nmid 2$ and $\mathrm{N}v > 3$; fix a standard additive character ψ_{add} of $\mathbb{A}_F/F$ that is unramified at the inert places. Let $n \geq 1$ and let $X = E^n$ with the hermitian form $\sum_i x_i \bar{y}_i$, which is positive definite at every real place. Then there is d_0 such that for every $d \geq d_0$ the following holds. Let $W = E^d$ with the skew-hermitian form $\delta \sum_i x_i \bar{y}_i$ for some $\delta \in E^\times$ with $\bar{\delta} = -\delta$, of signature $(d, 0)$ at every real*

place; let χ_X, χ_W be Hecke characters of E restricting to η^n and η^d on $\mathbb{A}_F^\times$, with infinity types $m_X, m_W\colon \{\text{real places of } F\} \to \mathbb{Z}$. Let π be an automorphic representation of $\mathrm{U}(X)(\mathbb{A}_F)$ whose archimedean component at each real place u has highest weight $\mu_u = (\mu_{u,1} \geq \cdots \geq \mu_{u,n})$ with

$$d + m_W(u) \leq 2\mu_{u,k} \qquad (1 \leq k \leq n).$$

Then there is an automorphic representation π' of $\mathrm{U}(W)(\mathbb{A}_F)$ such that

(St1) *at each real place u, π'_u has highest weight μ'_u with $2\mu'_{u,k} = 2\mu_{u,k} - d - m_W(u) + n + m_X(u)$ for $k \leq n$ and $2\mu'_{u,k} = n + m_X(u)$ for $n < k \leq d$;*

(St2) *at each place v of F split in E, with $E_v \cong F_v \times F_v$ and $\mathrm{U}(X)(F_v) \cong \mathrm{GL}_n(F_v)$, $\mathrm{U}(W)(F_v) \cong \mathrm{GL}_d(F_v)$, the local components π_v and π'_v are paired by the local theta correspondence for $(\mathrm{GL}_n, \mathrm{GL}_d)$: there is a nonzero linear map β from $\mathcal{S}(M_{n\times d}(F_v))$ to the space of conjugate-linear maps from the space of π_v to that of π'_v with $\beta(\omega_v(g,h)\varphi)(\pi_v(g)x) = \pi'_v(h)\big(\beta(\varphi)(x)\big)$, where ω_v is the Weil representation of $\mathrm{GL}_n(F_v) \times \mathrm{GL}_d(F_v)$ on $\mathcal{S}(M_{n\times d}(F_v))$ twisted by (χ_W^{-1}, χ_X) at v.*

Proof. In outline. The representation π' is the global theta lift of π in the stable range. The construction uses the Weil representation of the dual pair $(\mathrm{U}(X), \mathrm{U}(W))$, half-index lattices and Heisenberg models, Siegel–Weil kernels, and a Schur-type nonvanishing argument for the lifted forms. The weight relation (St1) comes from the archimedean correspondence for compact unitary groups, and (St2) from the local correspondence at split places. Nonvanishing of global theta lifts in the stable range goes back to [79].

□

Proposition 10.3. *Let E be a CM field with E/E^+ unramified at all finite places, $[E^+ : \mathbb{Q}]$ even, $E \ni \sqrt{-D_0}$ for some positive integer D_0, and with the inert places of the kind required in Proposition 10.2. Let $n \geq 1$ be even, π an automorphic representation of the definite unitary group $\mathrm{U}(n)$ of the form $\sum_i x_i\bar{y}_i$ on E^n (similarly $\mathrm{U}(n')$ for every n'), with highest weights μ at the archimedean places, and w_0 a split place of E. Then there are an integer $t \geq 1$ with $n+t$ odd, an algebraic Hecke character κ of E of positive infinity type m, unramified at w_0, and for $i = 1, 2$ automorphic representations π_i of $\mathrm{U}(n+t)$ with weights μ_i, together with families $(\nu_{i,j})_{j\leq t}$ of Hecke characters of E unramified at w_0 with infinity types $p_{i,j}$, such that each π_i is an up-lift of π by $(\kappa, \nu_{i,\cdot})$ (a relation between Satake parameters and weights whose conditions we do not reproduce), and $\nu_{1,j} \neq \nu_{2,j'}$ for all j, j'. Consequently, if every automorphic representation of $\mathrm{U}(n')$ for every odd n' has an attached Galois representation with the local properties of Proposition 10.4, then so does π.*

Proof. In outline. The lifts come from the stable-range theta lift (Proposition 10.2) with $t = 2(d_0+n)+1$, where d_0 is the threshold of that result. They use pairs of splitting characters with distinct auxiliary characters, a choice of base data for E, and the identification of each theta lift as an up-lift. The consequence is obtained by a subtraction step. The statement suggests the mechanism: the Galois representations attached to π_1 and π_2 both contain the representation sought for π, twisted by κ, and they differ in their auxiliary characters $\nu_{1,\cdot} \neq \nu_{2,\cdot}$, which should isolate the representation for π. We have not seen this step carried out, nor whether the places at which Frobenius data are available have density one.

□

The subtraction suggested by the statement of Proposition 10.3 looks like the cleanest new idea of this part: the two lifts would share the representation sought and differ in their auxiliary characters, so that subtracting their Galois representations isolates it. We have not seen this step carried out. Odd rank is handled by ℓ-adic limits.

Proposition 10.4. *Let E be a CM field, ℓ a prime, $\iota\colon \overline{\mathbb{Q}}_\ell \to \mathbb{C}$, $a \geq 1$ odd, and d a positive integer satisfying the auxiliary condition relative to E and ℓ of Proposition 10.7. Let T be a finite set of finite places of E not split over E^+, and $w_0 \neq v'$ places of E above ℓ, split over E^+ and lying over different places of E^+. Let π be an automorphic representation of the definite unitary group $\mathrm{U}(a)$ of the hermitian form $\sum_i x_i\bar{y}_i$ on E^a that is unramified at the non-split places outside T. Then there is a continuous semisimple representation $r\colon G_E \to \mathrm{GL}_a(\overline{\mathbb{Q}}_\ell)$ such that*

(Gu1) *for every split place $w \nmid \ell$ of the type determined by d at which π has Satake parameters B_w, r is unramified at w and ι carries the characteristic polynomial of $r(\mathrm{Frob}_w^{-1})$ to $\prod_{b\in B_w}\bigl(X - \mathrm{N}w^{(a-1)/2}b\bigr)$;*

(Gu2) *at every place $w \mid \ell$ lying over a place of E^+ different from those below w_0 and v', at which π is unramified, r is crystalline and de Rham with Hodge–Tate data determined by π, and the Hodge–Tate jumps are realised by crystalline representations as in the local data of π.*

The auxiliary condition on d, the type of the split places in (Gu1) (a condition on their residue characteristics relative to d), and the precise local condition at the places above ℓ in (Gu2) are conditions which we do not reproduce.

Proof. In outline. The representation r is obtained as an ℓ-adic limit, in three steps.

1. A seed step (Proposition 10.5) provides the required representations for automorphic representations of a special type at w_0 and of lacunary weight at v'.
2. A congruence step (Proposition 10.6) shows that every π as in the statement is congruent, modulo arbitrarily high powers of ℓ and in its Hecke eigenvalues at the relevant split places, to representations of that special type and weight.
3. The representation r is the limit of the corresponding pseudo-characters, and (Gu1) and (Gu2) pass to the limit.

□

Proposition 10.5. *Let E, ℓ, ι, a odd, T, $w_0 \neq v'$ be as in Proposition 10.4, let π be an automorphic representation of the definite unitary group $\mathrm{U}(a)$ of Proposition 10.4, and let S, Y be finite sets of finite places of E. Then there is a finite set Z_0 of split places not above ℓ such that for every automorphic representation π' of $\mathrm{U}(a)$ that is unramified at the non-split places outside T, of special type at w_0, of lacunary weight relative to π at v' (conditions of the kind of (Ge2) and (Ge3) of Proposition 10.7, whose precise form we do not reproduce), spherical at the split places away from w_0 and S, unramified at v' and generic at the places of Z_0, there is a continuous semisimple $r\colon G_E \to \mathrm{GL}_a(\overline{\mathbb{Q}}_\ell)$ such that:*

(Sd1) *for all but finitely many split places $w \nmid \ell$ (of the type determined by the auxiliary integer), r is unramified at w and, for every Frobenius element at w that lies in a fixed dense open subset of G_E, ι carries the characteristic polynomial of $r(\mathrm{Frob}_w^{-1})$ to the Satake polynomial of π'_w;*

(Sd2) *at the split places $w \in Y$ not above ℓ where π'_w is generic, r is attached to π' at w;*

(Sd3) *at the places $w \in Y$ above ℓ lying over places of E^+ different from those below w_0 and v', r is crystalline and de Rham with the Hodge–Tate data of π', and at the generic ones a further condition on Frobenius holds.*

The type of the split places in (Sd1) and the Frobenius condition in (Sd3) are conditions which we do not reproduce.

Proof. In outline. The seed representations come from the geometric realization Proposition 10.7. The representation obtained there has larger dimension, and it only satisfies that Frobenius is annihilated by the Satake polynomial. Judging from the steps involved, it is cut down to dimension a by dividing out a determinant character and by choosing the generic places Z_0 so that the Satake parameters separate.

□

Proposition 10.6. *Let E be a CM field, ℓ a prime, $\iota\colon \overline{\mathbb{Q}}_\ell \to \mathbb{C}$, $a \geq 1$ odd, T a finite set of non-split finite places, π an automorphic representation of the definite unitary group $\mathrm{U}(a)$ of the form $\sum_i x_i \bar{y}_i$ on E^a, unramified at the non-split places outside T, and $w_0 \neq v'$ places of E above ℓ, split over E^+ and lying over different places of E^+. Then there are open compact levels $K_w \subset \mathrm{GL}_a(E_w)$ at the split places w apart from w_0, maximal ($\mathrm{GL}_a(\mathcal{O}_{E_w})$) where π is unramified away from w_0, v', such that π has a nonzero vector of level (K_w). Moreover, for all finite sets Z of split places where π is unramified and W of split places above ℓ where π is unramified (all apart from w_0, v'), there are nonzero constants $c(w,k)$ such that for every finite family of local Hecke operators at level (K_w) with eigenvalues θ on π:*

(Ap1) *the vector formed by θ and the normalised power sums $c(w,k)\,\mathrm{tr}_k(\pi_w)$ of Satake parameters at $w \in W$ is an ℓ-adic limit, under ι^{-1}, of polynomial expressions in the corresponding vectors of automorphic representations π' of $\mathrm{U}(a)$ of the same level and the same weight away from w_0, v', of special type at w_0, of lacunary weight relative to π at v' (conditions of the kind of (Ge2) and (Ge3) of Proposition 10.7, whose precise form we do not reproduce), and generic at the places of $Z \cup W$;*

(Ap2) *if the Hecke operators have rational values, the eigenvalues $\iota^{-1}(\theta)$ lie in a finite extension of $\mathbb{Q}_\ell$.*

Proof. In outline. The approximation rests on an existence statement for lacunary weights, that is, weights with gaps at the place v' that are congruent to the given weight in the required sense. Presumably the mechanism is the familiar one of ℓ-adic families: the weight at v' and the type at w_0 are changed within a congruence class modulo high powers of ℓ.

□

The seeds are the following. We know this step only through its statement and the names of the results its proof combines; how the Galois representation is produced is the least clear part of this section.

Proposition 10.7. *Let E be a CM field containing $\sqrt{-d}$ for a positive integer d satisfying an auxiliary condition relative to E and ℓ, let $\iota\colon \overline{\mathbb{Q}}_\ell \to \mathbb{C}$, and let a be odd. Let T be a finite set of finite places of E not split over E^+, and $w_0 \neq v'$ places of E above ℓ, split over E^+, lying over different places of E^+ and inducing the same place of $\mathbb{Q}(\sqrt{-d})$. Let π be an automorphic representation of the definite unitary group $\mathrm{U}(a)$ of the hermitian form $\sum_i x_i \bar{y}_i$ on E^a such that*

(Ge1) *π is unramified at the non-split places outside T and spherical at the split places away from w_0 and a finite set S;*

(Ge2) *π_{w_0} contains a character of a torsion-free compact open subgroup of $\mathrm{GL}_a(E_{w_0})$, factoring through a nontrivial character of $\mathbb{Z}/\ell^m$, every representation containing which is supercuspidal;*

(Ge3) *π is unramified at v', and at an archimedean place corresponding to v' under ι its highest weight has lacunary jumps.*

Then there are an integer D and a continuous semisimple representation $\rho\colon G_E \to \mathrm{GL}_D(\overline{\mathbb{Q}}_\ell)$ such that, for all but finitely many split places $w \nmid \ell$ (of a type determined by d) at which π is unramified with Satake parameters B_w, ρ is unramified at w and $\rho(\mathrm{Frob}_w^{-1})$ is annihilated by the polynomial $\prod_{b\in B_w}\bigl(X - \iota^{-1}(\mathrm{N}w^{(a-1)/2}b)\bigr)$; and such that for each place $w \mid \ell$ lying over a place of E^+ different from those below w_0 and v', and inducing the same place of $\mathbb{Q}(\sqrt{-d})$ as v', there is such a representation that is moreover crystalline at w with Hodge–Tate jumps determined by the weights of π. The auxiliary condition on d, the type of the split places (a condition on their residue characteristics relative to d), the precise form of the jump condition, and a further condition on Frobenius at the places $w \mid \ell$ are conditions of the same nature which we do not reproduce.

Proof. In outline. The proof combines three inputs: a set of geometric statements, a Kodaira–Spencer statement in weight one, and an existence statement for auxiliary data. One of the inputs used is the holomorphic Lefschetz formula of Atiyah and Bott for finite étale correspondences [7]. We have not determined the mechanism. Judging only from the names of the objects involved (families of Kuga–Sato type over compact unitary Shimura varieties of PEL type, Hodge cohomology, comparison theorems), one may guess that π is realised geometrically and ρ found in étale cohomology, but we have not verified this.

□

10.1 Base change

Base change from definite unitary groups to GL_n [74], [88] is reduced to the existence of finitely many candidate lifts which together match the unitary representation at its inert places.

Definition 10.8. Let E be a CM field with maximal totally real subfield E^+, let $n \geq 1$, and let U_n be the definite unitary group over E^+ of the standard hermitian form $\sum_i x_i\bar{y}_i$ on E^n. Let π be an irreducible automorphic representation of $\mathrm{U}_n(\mathbb{A}_{E^+})$. A finite place w of E is **split** if another place of E lies over the same place of E^+, and **inert** otherwise; π is **unramified** at w if it has a nonzero vector fixed by the hyperspecial maximal compact subgroup at w, and then has Satake parameters $\alpha_w(\pi)$, with base-change parameter $\mathrm{bc}(\alpha_w(\pi))$, a multiset of n elements of $\mathbb{C}^\times$. At an archimedean place τ, π_τ is the irreducible algebraic representation of highest weight μ_τ, with infinitesimal character $\{\mu_{\tau,k} + \frac{n-1}{2} - k : 0 \leq k < n\}$.

A **candidate lift** of size n over E is a tuple $\mathcal{C} = (c, (N_i), (\Pi_i), (\sigma_{i,w}), (m_{i,\tau}))$ with $c \geq 0$, positive integers $N_1 + \cdots + N_c = n$, irreducible square-integrable automorphic subrepresentations Π_i of $\mathrm{GL}_{N_i}(\mathbb{A}_E)$, irreducible smooth representations $\sigma_{i,w}$ of $\mathrm{GL}_{N_i}(E_w)$ occurring as local components of Π_i at every finite place w, and multisets $m_{i,\tau}$ of N_i complex numbers, the infinitesimal characters of Π_i at the embeddings $\tau\colon E \to \mathbb{C}$. It is a candidate lift **of** π if

(Cl1) for every τ, $\sum_i m_{i,\tau}$ is the infinitesimal character of π_τ;

(Cl2) for all but finitely many split places w, some irreducible subquotient of $\sigma_{1,w} \times \cdots \times \sigma_{c,w}$ is the component of π along the identification $\mathrm{U}_n(E_v^+) \cong \mathrm{GL}_n(E_w)$; such w are called **matched**.

A candidate lift of π **matches** π at an inert place w where π is unramified if every $\sigma_{i,w}$ is unramified, with Satake parameters $\beta_{i,w}$, and $\sum_i \beta_{i,w} = \mathrm{bc}(\alpha_w(\pi))$. The normalisations of the Satake parameters and of bc at inert places are fixed by further conventions, which we do not reproduce.

Proposition 10.9. *Suppose that the following holds for every CM field E such that E/E^+ is unramified at every finite place, every $n \geq 1$ and every irreducible automorphic representation π of $\mathrm{U}_n(\mathbb{A}_{E^+})$: for every set S of at most one finite place of E there are finitely many candidate lifts $\mathcal{C}_1, \ldots, \mathcal{C}_k$ of π, each of which matches π at the split places in S, such that for every finite set T of finite places of E some $\mathcal{C}_j$ matches π at every inert place of T at which π is unramified. Then every such π has a base change to $\mathrm{GL}_n(\mathbb{A}_E)$, that is, an isobaric sum of conjugate self-dual discrete automorphic representations of general linear groups, compatible with π at the split places, at the inert places where π is unramified, and at the archimedean places.*

Proof. In outline. Fix E, n and π. A single candidate lift that matches π at every inert place where π is unramified exists. Indeed, if each of $\mathcal{C}_1, \ldots, \mathcal{C}_k$ failed at some such place, the finitely many failures would form one finite set T for which no $\mathcal{C}_j$ matches, a contradiction. Call such a lift full. One then shows that any two full candidate lifts have the same Satake parameters at almost all places, and that a full lift is stable under the conjugate-dual involution. Moreover, the weights of π are realised by the infinitesimal characters in regular form, and the constituents Π_i decompose into strings of conjugate self-dual cuspidal representations. Local rigidity and a permutation argument on Satake exponents then upgrade a full candidate lift to the base change.

□

This ends the outline of the proof. The last section lists what in it we do not yet understand.

11 What is not yet understood

We list what we do not yet understand, section by section. The items are of three kinds: conceptual gaps, that is, steps that we can state but not explain; parts given only in outline, that is, definitions that omit conditions and arguments that we know only at their top level; and known-or-new questions, that is, statements that we could not place in the literature.

The analytic reduction and the hosts (Sections 2 and 3).

- The window form of Landau's theorem (Proposition 2.2) is given in outline: we know its statement and that it rests on the positivity of the weights, but not the quantitative argument by which a zero ρ with $\mathrm{Re}\,\rho > \theta$ forces $G_{\mathfrak{L}}$ above y^θ on every far window.

- The mean-square estimates behind Proposition 3.7 are known to us only at their top level; we do not know whether they follow from published large-sieve inequalities for quadratic characters [52] or need more.

- The construction of hosts from good window pairs in Proposition 3.9 (elements of small norm, a small prime congruent to 1 modulo the level), the choice of the set T behind the share condition (E3) and the goodness conditions (P4) of Proposition 3.5 are given

in outline, and the definition of hosts omits conditions on the CM type and on the Artin images of the residue classes of d.

The bridge (Section 4).

- This is the main conceptual gap. By definition, the prime-sum entries S_u of a slope datum are the inert prime sums of the host. The content of the bridge lies in the loss condition (S3), $S_u - \Theta_u \leq \mathrm{Lo}$, and in the matching of the main terms with the divisor bound (S1). Here Θ_u is a quantity of the lattice data: twice the sum, over the inert primes p up to a cut-off, of $\log p$ times the net local gain of the lattices at p, that is, the local gains over the lifts of the member minus a correction term. So (S3) compares, prime by prime over the inert primes, the contribution of p to the inert prime sums with $\log p$ times the net gain of the lattices at p, and the total shortfall is the loss. We do not understand why the local gains of the lattices attached to theta lifts dominate the prime contributions up to a small loss, that is, how the loss bounds are proved, nor what the local gains and the correction terms are in classical terms; their definitions are part of the size and row data of a master datum, which we have not opened.

- The slope conditions are stated in outline: the function φ of (S1), the loss term of (S2) and the comparisons (S5) are explicit expressions that we do not reproduce, and the roles of the individual size and error conditions (Z3)–(Z7) are not clear to us.

- The passage from a bounded excess to the one-sided bound (Proposition 4.8) is understood in its mechanism, orthogonality and then the moving of the cut-offs to x, but the derivative estimate that moves the cut-offs is known to us only by its statement.

The lattice data (Sections 6, 7 and 8).

- According to its author [2], the chain of theta lifts along the hermitian tower plays the part of the multiplicity in Stepanov's method; how it does so is not visible in what we have read.

- The definite-side family structure (Definition 6.17) is given in one sentence: its forty or so conditions are described only by what they concern. Two of the seven steps of Proposition 6.18 are known to us only by name: the separation of functionals, which supplies the separation property assumed by Theorem 7.6, and the openness of parameters. The linking statements at levels two and three, the members with large unramified exponents and the choice of auxiliary lattices and measures are described in words, not stated.

- The general construction that turns the family structure into tied families in Proposition 6.20, the closure statement for the condition linking the lattices of the members to their growth data, and the way in which the Sha bounds enter the parameters of the slope conditions (presumably through the defects and the node sizes) are known to us only through their inputs.

- Of the four row conditions of Definition 6.16, the outline shows the proofs of (COV) and (FUN) only. We have not found where the surjectivity condition (ON) and the lower bound (LOW) are established, and a second form of (LOW), holding eventually on the families of the standard class, is a hypothesis of Theorem 7.6 whose source we do not know.

- The covolume condition is proved at compatibly ruled points, and toric functoriality at points with the compatible point conditions, whereas ruled points are defined through the tie conditions. We do not know whether either kind of point is a special case of the other, and we have not traced which of them the construction of tied families needs.

- The master datum (Definition 6.10) omits the compatibility and tie conditions of its items (D2), (D4) and (D7)–(D9). Parametrised families (Definition 6.13) state (F4) at the level of its fields, and the standard family class (Definition 6.16) describes fourteen conditions and the surjectivity condition (ON) without stating them; Hida settings, cores of master data and grid data are outlines of the same kind.

- Which statements of the development enter as axioms of an interface rather than with proofs we have not determined. Among the objects printed here, the period-ring data of Definition 6.9, which item (D7) of a master datum requires, are of this kind, and we have not found a construction of such data.

- We understand the definition of column chains (Definition 6.8) but not their interpretation as theta lifts. The role of the parameter J of a scale and of the weight c_{w}, and the origin of the exponent 377 in (Sc4) of Definition 6.3, are not clear to us.

- Theta agreement (Proposition 7.3): we do not see how the properties (A2) and (A3), which relate the constants along tower steps through partial L-values up to a constant fixed by the step, follow from the constant-term identities, and we could not decide whether the statement as a whole is a known consequence of the Rallis inner product formula for unitary groups [80], [49].

- Critical values on the grid: we do not know how the explicit coefficients and the conductor term of the valuation bound for square ratios (Theorem 7.9, item 2) arise, nor whether the thresholds 1800, 100 and 44 are forced. The twenty estimates that close the proof of Theorem 7.6 are known to us by name only.

- The covolume bound: we do not see the mechanism behind the arithmetic proportionality (V4) and the lower bounds (V5) and (V6), with their dependence on the root discriminant and on $L'/L(1,\chi_{-d})$. We do not know where Zhang's inequality for successive minima [138] enters the tower argument. The transport from towers to the models of a master datum in Proposition 8.5, and the lattice estimates of Theorem 8.1, are known to us only through their inputs.

- Known or new: the integrality of theta-kernel coefficients (Proposition 6.19); the weighted degree inequality (Lemma 8.7), which may be a known semipositivity statement for Hodge bundles in characteristic p; the explicit denominator bounds of Definition 7.4, with the threshold $[F:\mathbb{Q}]k \geq 16$; the cuspidality of theta lifts along the grid, compared with the classical tower property [101]; and the explicit bounds for discrete series at large gap, whose conclusion is classical [48], [103].

The Sha bounds (Section 9).

- The deformation-theoretic core of the congruence bound (Proposition 9.9), that is, how the Tate–Shafarevich group is compared with the congruence module, is known to us only at its top level.

- The automorphy statements behind counts II and III (GSp_4 with dense image) and count V (the unitary lift over $\mathbb{Q}(i)$) are known to us only through their names, and one condition of Definition 9.8, a witness condition of endoscopic type, is omitted.

- Case (IV), and the part of case (V) that count V does not cover, have no count of their own; they are handled inside the final step of the proof of Theorem 9.1, which we have not read.

- The Bernoulli bound (Proposition 9.16): duality alone cannot produce the Bernoulli number, and we have not identified the arithmetic input that does, which should be of the kind of the main conjecture or of an Euler system of Gauss sums or cyclotomic units.
- The fields of the minus-part bound (Proposition 9.15) and the precise ranges of the bad-prime bound (Proposition 9.17) are given in outline.
- Known or new: whether the vanishing statement (Proposition 9.14) and the Bernoulli bound (Proposition 9.16) are results of Herbrand–Ribet and Mazur–Wiles type [102], [82] proved anew; whether weak descent (Proposition 9.18) and the ℓ-adic measures of Proposition 9.11 are in the literature in this form; and whether the polynomial bounds for Tate–Shafarevich groups are new in substance or only in their effectivity.

Galois representations (Section 10).

- The seed step (Proposition 10.7): we have not determined its mechanism. The names of its inputs suggest a geometric realisation of π in the cohomology of families of Kuga–Sato type over compact unitary Shimura varieties, but we have not verified this; we know the step only through its statement and the names of its ingredients.
- The relation between the two theta lifts in Proposition 10.3, the notion of approximation in Proposition 10.6 and the definition of the special type are given in outline. We have not seen the subtraction step carried out, nor whether the split places at which Frobenius data are available have density one, as the subtraction seems to need. We have not traced how local–global compatibility is upgraded in the second stage of Theorem 10.1, nor whether its third stage uses base change.
- Known or new: the stable-range theta lift with explicit archimedean weights (Proposition 10.2) may be in the literature in this form; the use of ℓ-adic limits starting from a special type and lacunary weight, and the pair of theta lifts in even rank, look new to us.

The author expects the argument to be simplified considerably, and says that he has not yet mastered all of the mathematics around it [2]. Finally, how much of the argument we have read. We have read the statements of the results printed here and compared them with the development, and of their proofs we have read the top level; for the analytic part, Sections 2–4, somewhat more. Below the printed results lies a much larger body of intermediate statements, by a rough estimate between fifteen and twenty-five thousand statements, which we have not read; the proofs given in outline mark where it begins.

A The known theorems used

The argument proves many of the following known theorems anew, often by routes of its own; we cite the literature instead. For each we give the statement in the form in which the argument uses it, references, and the section where it is used; where the form used differs from the published one, we say so.

A.1 Analytic number theory

- *Landau's theorem.* A Dirichlet integral $\int_1^\infty a(y)y^{-s-1}\,dy$ whose integrand is eventually non-negative has a singularity at the real point of its line of convergence [76]. Used in Section 2, in the window form of Proposition 2.2.

- *Zeros of Dirichlet L-functions.* For primitive characters modulo N: the truncated explicit formula for $\sum_{n\le y}\Lambda(n)\chi(n)$; the zero-free region $\operatorname{Re} s > 1 - c/\log\bigl(N(|\operatorname{Im} s|+2)\bigr)$ with at most one exceptional real zero; log-free zero-density estimates near $s=1$; and local counts of zeros [31], [81], [61]. Used in Section 6, for primes in short intervals at very large scales.

- *Primes in progressions.* The prime number theorem in arithmetic progressions with relative error and a possible exceptional zero [41], and Linnik's bound $p \ll N^L$, with an absolute constant L, for the least prime $p \equiv a \pmod N$ [81]. Used in the theta-lift construction of Sections 6 and 7.

- *Logarithmic derivatives.* If χ is a primitive character of modulus at most Q and $L(s,\chi)$ has no zeros in the box $\operatorname{Re} s > 1-c$, $|\operatorname{Im} s| \le Q$, then $|L'(1,\chi)/L(1,\chi)| \ll_c \log\log Q$, with an implied constant depending on c; this follows by the standard argument with the Borel–Carathéodory lemma [89]. Used in Section 3 (Lemma 3.8).

A.2 Algebraic number theory

- *Class field theory.* Local and global reciprocity and the existence theorems [106], [125], [6]. Used for host fields in Section 3 and throughout.

- *Two-power cyclotomic fields.* For primes $q \equiv 3 \pmod 8$ and $d \equiv 7 \pmod{16}$ with $(-d/q) = 1$, the field $\mathbb{Q}(\zeta_{2^{e+2}} + \zeta_{2^{e+2}}^{-1}, \sqrt{-q}, \sqrt{-d})$ is a CM field, Galois over $\mathbb{Q}$ with group $\{\pm1\}^2 \times (\mathbb{Z}/2^{e+2}\mathbb{Z})^\times/\{\pm1\}$, with explicit Frobenius elements and inertia groups [132]. Used in Section 3 (Proposition 3.9).

- *Hecke L-functions.* Meromorphic continuation and functional equation, nonvanishing on $\operatorname{Re} s = 1$, and the Chebotarev density theorem [124], [53], [26]. Used in Sections 6 and 7, in particular for the critical values on the grid.

- *Hasse's norm theorem* for quadratic extensions [51]. Used in Section 6, through the classification of hermitian forms.

- *Partial zeta values.* Rationality of the values of partial zeta functions of totally real fields at negative integers (Siegel–Klingen) and their congruences (Deligne–Ribet) [116], [69], [34]. Used in Section 7 (Theorem 7.5).

- *Complex multiplication.* The main theorem of complex multiplication for elliptic curves and the generation of the Hilbert class field of an imaginary quadratic field by values of j [112]. Used in Section 7 (Theorem 7.5).

- *Ferrero–Washington.* Used only in the form that the ℓ-rank of the class groups in the cyclotomic towers of an abelian field is bounded, a consequence of the vanishing of the μ-invariant [39], [117]. Used in Section 9 (Proposition 9.17, Proposition 9.12).

A.3 Galois cohomology

- *Local Tate duality* and the local Euler characteristic formula [123], [84]. Used in Sections 6 and 9.

- *Poitou–Tate duality* and the global Euler characteristic formula [84]. Used in Section 9.

- *Second cohomology of Tate twists.* An explicit bound for the order of $H^2(G_{L,S}, \mathbb{Z}_\ell(e+1))$ of the kind proved by Soulé [119]; the form used is weaker than the published one, doubly exponential in the parameters. Used in Section 9 (Proposition 9.17 and the counts).

- *Kato's Euler system.* Vanishing of the Bloch–Kato Selmer groups [14] of suitable twists of newforms of odd weight [65]. Used in Section 9 (Proposition 9.9).

A.4 p-adic Hodge theory

- *Fontaine's period rings.* Injectivity of $K \otimes_{K_0} B_{\mathrm{cris}} \to B_{\mathrm{dR}}$, the fundamental exact sequence and the regularity of B_{cris} [40]. Used for the period-ring data of Section 6.

- *p-adic monodromy.* De Rham representations are potentially semistable [10]. Used in Sections 6 and 9.

- *Hecke characters.* The ℓ-adic characters attached to algebraic Hecke characters are potentially crystalline at the places above ℓ, with the expected Hodge–Tate weights [107], [40]. Used in Sections 6 and 9.

- *p-divisible groups over $\mathcal{O}_C$.* The Hodge–Tate sequence [38], [105], and prismatic Dieudonné modules with the Hodge filtration [4]. Used in Sections 6 and 8.

- *Comparison theorems* of p-adic Hodge theory [36]. Used in Section 10.

- *Local deformation rings.* Kisin's local deformation rings of ℓ-adic representations of local Galois groups [68]. Used in Section 9.

A.5 Abelian varieties and their moduli

- *Torsion.* For an abelian variety of dimension g over a field and n prime to its characteristic, $A[n]$ has n^{2g} geometric points [93]. Used in Section 6.

- *Zarhin's trick* over valuation rings [137]. Used in Section 8.

- *Duality and periods.* Dual abelian schemes [93], the Néron extension property [18] and the crystalline–de Rham comparison [11]; periods of complex abelian varieties and the Riemann form [93], [33]; Gauss–Manin horizontality of periods in holomorphic families [66]. Used in Section 9.

- *Grothendieck–Messing* deformation theory of abelian schemes [83]. Used in Sections 8 and 9.

- *Semistable reduction.* A polarised abelian variety over a finite extension of $\mathbb{Q}_p$ acquires semiabelian reduction over a finite extension [110]; the form used is this special case. We have not traced where in the argument it enters.

- *Mumford uniformisation.* A split degeneration of abelian varieties over a complete discrete valuation ring is the quotient $\tilde{G}/Y$ of its Raynaud extension by a lattice, with the resulting description of its torsion points [94], [37]. Used in Section 8.

- *Semi-abelian schemes.* Seminegativity of the Lie algebra of a semi-abelian scheme over a curve [37], [90]. Used in Section 8.

- *Toroidal compactifications* of Siegel moduli spaces with level structure [37], and of unitary PEL Shimura varieties with the logarithmic Kodaira–Spencer isomorphism [3], [75]. Used in Section 8.

- *Mumford's good metrics* compute Chern numbers on toroidal compactifications, with the push-forward along flag bundles [95]. Used in Section 8.

A.6 Complex and arithmetic geometry

- *Vanishing and extension.* Kawamata–Viehweg vanishing [67], [129], and the L^2 extension theorem from a finite nondegenerate zero set with log-canonical weight [98], [35]. Used in Section 8.

- *Resolution of singularities.* Functorial resolution and principalisation in characteristic 0 [57], [12], [135], [71]. Used in Section 8.

- *Arithmetic Hilbert–Samuel.* The arithmetic Hilbert–Samuel formula for nef hermitian line bundles [43], [138], [92], and the lower bound of the arithmetic volume by the arithmetic self-intersection [91], [136]. Used in Section 8.

- *Successive minima.* Zhang's inequality between the height of an arithmetic variety and its successive minima [138]. Used in Section 8 (Proposition 8.6).

- *Holomorphic Lefschetz formula* of Atiyah and Bott [7], which is stated for maps; the argument uses a form for finite étale correspondences. Used in Section 10 (Proposition 10.7).

A.7 Hermitian forms and arithmetic groups

- *Landherr's theorem.* Hermitian forms over a CM field are classified by their dimension, their local invariants and their signatures [77]. Used in Section 6.

- *Arithmetic groups.* Unitary groups of hermitian spaces have finite covolume, and the lattices in a hermitian space fall into finitely many classes [17], [111]. Used in Section 10 (Theorem 10.1).

- *Mass formula.* The adelic class mass of a definite hermitian lattice [113]. Used in Section 6 (Proposition 6.18).

- *Strong approximation* for special unitary groups [70], [100]. Used in Sections 6 and 7.

A.8 Representations of real and p-adic groups

- *Local Langlands correspondence* for GL_n [50], [55], [104]; some of its compatibilities are used only for a restricted class of representations. Used in Sections 6 and 10.

- *Local converse theorem* for GL_n with twists by GL_{n-1} [54]. Used in Sections 6 and 9.

- *Green's parametrisation* of the cuspidal characters of GL_n over a finite field [46]. Used in Section 6 (Proposition 6.20) and in Section 9.

- *Casselman–Wallach* globalisation for indefinite unitary groups [24], [131]. Used in Section 7.

- *Howe duality* over p-adic fields [86], [42], in particular at unramified places, and over the real numbers [58], [64], and Kudla's splitting of the metaplectic cover over unitary groups [72]. Used in Section 7.

- *Theta lifts in the stable range.* Nonvanishing and unitarity of theta lifts in the stable range [79], and the explicit local theta correspondence for dual pairs of type II [85]. Used in Section 10 (Proposition 10.2).

- *Discrete series* of $\mathrm{U}(p,q)$: existence for Harish-Chandra parameters off the walls, with the lowest K-type of multiplicity one [48], [103]; the argument proves a quantitative form at parameters far from the walls. Used in Section 7.

A.9 Automorphic forms, theta lifts and Galois representations

- *Cusp forms.* Rapid decay of cusp forms on unitary groups [87], and strong multiplicity one for isobaric sums of cuspidal representations of GL_n [62]. Used in Sections 6 and 9.

- *Clozel's theorems.* The purity lemma for cuspidal representations over totally complex fields, and the conjugation of regular algebraic cuspidal representations by automorphisms of $\mathbb{C}$ [28]. Used in Section 9.

- *Cyclic base change* for GL_n [5]. Used in Section 9.

- *Eisenstein series and the doubling method.* Meromorphic continuation of Siegel Eisenstein series on doubled unitary groups [78], [122], and the basic identity of the doubling method [99]. Used in Section 7.

- *Siegel–Weil and Rallis.* The Siegel–Weil formula for a definite hermitian space and a doubled unitary group [59], [49], and the Rallis inner product formula for unitary groups in the convergent range [80], [49]; the method goes back to [133], [101] and [73]. Used in Sections 6 and 7 (Proposition 7.3).

- *Base change* from unitary groups to GL_n [74], [88], and the classification of automorphic representations of inner forms of unitary groups [63]. Used in Section 9; the argument's own route is Proposition 10.9.

- *Galois representations* attached to polarized regular algebraic cuspidal representations of GL_n over CM fields, with local–global compatibility [50], [114], [27], [22]. Used in Sections 6 and 9; the argument's own route is Theorem 10.1.

- *Totally real fields.* Galois representations for regular algebraic cuspidal representations over totally real fields, used only for GL_2 [23], [127], [13], and in general [8], and Hilbert modular forms congruent to 1 [108], [134]. Used in Section 9.

Acknowledgements

This paper was prepared with the assistance of Aleph, an AI research assistant for mathematics, using AI models from Anthropic (Claude), OpenAI (ChatGPT) and Google (Gemini) together with bespoke models. A team of Claude Code agents read the Lean development, traced the path of its main theorems, and transcribed the statements on that path into standard mathematical language with unified notation; the statements printed here were compared with the Lean source. References to the literature were added on a best-effort basis and have not all been checked yet

against the sources. The text was drafted and reviewed by Aleph under the direction of the author, with the aim to optimise the argument for clarity, logical structure and presentation. This is work in progress; later versions will aim at a clearer and more complete account.

References

[1] L. Alpöge, *Lean 4 formalization of the Riemann hypothesis and of the generalized Riemann hypothesis for Dirichlet L-functions*, alpo.ge (2026).

[2] L. Alpöge, *Remarks accompanying the formal development*, unpublished note (2026).

[3] A. Ash, D. Mumford, M. Rapoport et al., *Smooth compactification of locally symmetric varieties*, Math. Sci. Press (1975).

[4] J. Anschütz and A. Le Bras, *Prismatic Dieudonné theory*, Forum Math. Pi **11** (2023), Paper No. e2, 92 p.

[5] J. Arthur and L. Clozel, *Simple algebras, base change, and the advanced theory of the trace formula*, Ann. of Math. Stud. **120**, Princeton University Press (1989).

[6] E. Artin and J. Tate, *Class field theory*, W. A. Benjamin (1968).

[7] M. Atiyah and R. Bott, *A Lefschetz fixed point formula for elliptic complexes. II: Applications*, Ann. Math. (2) **88** (1968), 451-491.

[8] T. Barnet-Lamb, T. Gee, D. Geraghty et al., *Potential automorphy and change of weight*, Ann. Math. (2) **179** (2014), 501-609.

[9] J. Bellaïche and G. Chenevier, *Formes non tempérées pour* U(3) *et conjectures de Bloch–Kato*, Ann. Sci. Éc. Norm. Supér. (4) **37** (2004), 611–662.

[10] L. Berger, *Représentations p-adiques et équations différentielles*, Invent. Math. **148** (2002), 219–284.

[11] P. Berthelot and A. Ogus, *Notes on crystalline cohomology*, Princeton University Press (1978).

[12] E. Bierstone and P. Milman, *Canonical desingularization in characteristic zero by blowing up the maximum strata of a local invariant*, Invent. Math. **128** (1997), 207-302.

[13] D. Blasius and J. Rogawski, *Motives for Hilbert modular forms*, Invent. Math. **114** (1993), 55–87.

[14] S. Bloch and K. Kato, *L-functions and Tamagawa numbers of motives*, in: The Grothendieck Festschrift I, Progr. Math. **86**, Birkhäuser (1990), 333–400.

[15] E. Bombieri, *On the large sieve*, Mathematika **12** (1965), 201–225.

[16] E. Bombieri, *Counting points on curves over finite fields (d'après S. A. Stepanov)*, Séminaire Bourbaki 1972/73, Exp. 430, Lecture Notes in Math. **383** (1974), 234–241.

[17] A. Borel and Harish-Chandra, *Arithmetic subgroups of algebraic groups*, Ann. Math. (2) **75** (1962), 485-535.

[18] S. Bosch, W. Lütkebohmert, and M. Raynaud, *Néron models*, Berlin etc.: Springer-Verlag (1990).

[19] J.-B. Bost, *Périodes et isogénies des variétés abéliennes sur les corps de nombres (d'après D. Masser et G. Wüstholz)*, Séminaire Bourbaki, Exp. 795, Astérisque **237** (1996), 115–161.

[20] J. Bost, H. Gillet, and C. Soulé, *Heights of projective varieties and positive Green forms*, J. Am. Math. Soc. **7** (1994), 903-1027.

[21] O. Brinon and B. Conrad, *CMI summer school notes on p-adic Hodge theory*, lecture notes, Clay Mathematics Institute (2009).

[22] A. Caraiani, *Local-global compatibility and the action of monodromy on nearby cycles*, Duke Math. J. **161** (2012), 2311-2413.

[23] H. Carayol, *Sur les représentations ℓ-adiques associées aux formes modulaires de Hilbert*, Ann. Sci. École Norm. Sup. (4) **19** (1986), 409–468.

[24] W. Casselman, *Canonical extensions of Harish-Chandra modules to representations of G*, Can. J. Math. **41** (1989), 385-438.

[25] J. W. S. Cassels and A. Fröhlich (eds.), *Algebraic number theory*, Academic Press (1967).

[26] N. Tschebotareff, *Die Bestimmung der Dichtigkeit einer Menge von Primzahlen, welche zu einer gegebenen Substitutionsklasse gehören*, Math. Ann. **95** (1926), 191–228.

[27] G. Chenevier and M. Harris, *Construction of automorphic Galois representations. II*, Camb. J. Math. **1** (2013), 53-73.

[28] L. Clozel, *Motifs et formes automorphes: applications du principe de fonctorialité*, in: Automorphic Forms, Shimura Varieties, and L-functions I (Ann Arbor, 1988), Perspect. Math. **10**, Academic Press (1990), 77–159.

[29] L. Clozel, M. Harris, and R. Taylor, *Automorphy for some l-adic lifts of automorphic mod l Galois representations. With Appendix A, summarizing unpublished work of Russ Mann, and Appendix B by Marie-France Vignéras.*, Publ. Math., Inst. Hautes Étud. Sci. **108** (2008), 1-181.

[30] H. Darmon, F. Diamond and R. Taylor, *Fermat's last theorem*, in: Elliptic Curves, Modular Forms and Fermat's Last Theorem (Hong Kong, 1993), 2nd ed., International Press (1997), 2–140.

[31] H. Davenport, *Multiplicative number theory. Revised and with a preface by Hugh L. Montgomery.*, New York, NY: Springer (2000).

[32] P. Deligne, *Formes modulaires et représentations ℓ-adiques*, Séminaire Bourbaki 1968/69, Exp. 355, Lecture Notes in Math. **179** (1971), 139–172.

[33] P. Deligne, *Hodge cycles on abelian varieties* (notes by J. S. Milne), Lecture Notes in Math. **900** (1982), 9–100.

[34] P. Deligne and K. Ribet, *Values of abelian L-functions at negative integers over totally real fields*, Invent. Math. **59** (1980), 227-286.

[35] J. Demailly, *Analytic methods in algebraic geometry*, Somerville, MA: International Press; Beijing: Higher Education Press (2012).

[36] G. Faltings, *Crystalline cohomology and p-adic Galois-representations*, Baltimore: Johns Hopkins University Press (1989), 25-80.

[37] G. Faltings and C. Chai, *Degeneration of abelian varieties*, Berlin etc.: Springer-Verlag (1990).

[38] L. Fargues, *L'isomorphisme entre les tours de Lubin–Tate et de Drinfeld et applications cohomologiques*, in: L'isomorphisme entre les tours de Lubin–Tate et de Drinfeld, Progr. Math. **262**, Birkhäuser (2008), 1–325.

[39] B. Ferrero and L. Washington, *The Iwasawa invariant μ_p vanishes for abelian number fields*, Ann. Math. (2) **109** (1979), 377-395.

[40] J.-M. Fontaine, *Le corps des périodes p-adiques*, Astérisque **223** (1994), 59–111.

[41] P. Gallagher, *A large sieve density estimate near $\sigma = 1$*, Invent. Math. **11** (1970), 329-339.

[42] W. Gan and S. Takeda, *A proof of the Howe duality conjecture*, J. Am. Math. Soc. **29** (2016), 473-493.

[43] H. Gillet and C. Soulé, *Amplitude arithmétique*, C. R. Acad. Sci. Paris Sér. I **307** (1988), 887–890.

[44] H. Gillet and C. Soulé, *Arithmetic intersection theory*, Publ. Math., Inst. Hautes Étud. Sci. **72** (1990), 93-174.

[45] D. Grayson, *Reduction theory using semistability*, Comment. Math. Helv. **59** (1984), 600-634.

[46] J. Green, *The characters of the finite general linear groups*, Trans. Am. Math. Soc. **80** (1955), 402-447.

[47] G. H. Hardy and M. Riesz, *The general theory of Dirichlet's series*, Cambridge Tracts in Math. and Math. Phys. **18**, Cambridge University Press (1915).

[48] Harish-Chandra, *Discrete series for semisimple Lie groups. II: Explicit determination of the characters*, Acta Math. **116** (1966), 1-111.

[49] M. Harris, S. Kudla, and W. Sweet, *Theta dichotomy for unitary groups*, J. Am. Math. Soc. **9** (1996), 941-1004.

[50] M. Harris and R. Taylor, *The geometry and cohomology of some simple Shimura varieties. With an appendix by Vladimir G. Berkovich*, Princeton, NJ: Princeton University Press (2001).

[51] H. Hasse, *Beweis eines Satzes und Widerlegung einer Vermutung über das allgemeine Normenrestsymbol.*, Nachr. Ges. Wiss. Göttingen, Math.-Phys. Kl. **1931** (1931), 64-69.

[52] D. R. Heath-Brown, *A mean value estimate for real character sums*, Acta Arith. **72** (1995), 235–275.

[53] E. Hecke, *Eine neue Art von Zetafunktionen und ihre Beziehungen zur Verteilung der Primzahlen. II*, Math. Z. **6** (1920), 11–51.

[54] G. Henniart, *Caractérisation de la correspondance de Langlands locale par les facteurs ε de paires*, Invent. Math. **113** (1993), 339–350.

[55] G. Henniart, *Une preuve simple des conjectures de Langlands pour* $\mathrm{GL}(n)$ *sur un corps p-adique*, Invent. Math. **139** (2000), 439–455.

[56] H. Hida, *A p-adic measure attached to the zeta functions associated with two elliptic modular forms. I*, Invent. Math. **79** (1985), 159-195.

[57] H. Hironaka, *Resolution of singularities of an algebraic variety over a field of characteristic zero. I, II*, Ann. Math. (2) **79** (1964), 109-203, 205-326.

[58] R. Howe, *Transcending classical invariant theory*, J. Am. Math. Soc. **2** (1989), 535-552.

[59] A. Ichino, *A regularized Siegel–Weil formula for unitary groups*, Math. Z. **247** (2004), 241–277.

[60] A. E. Ingham, *The distribution of prime numbers*, Cambridge Tracts in Math. **30**, Cambridge University Press (1932).

[61] H. Iwaniec and E. Kowalski, *Analytic number theory*, Providence, RI: American Mathematical Society (AMS) (2004).

[62] H. Jacquet and J. A. Shalika, *On Euler products and the classification of automorphic representations. I*, Amer. J. Math. **103** (1981), 499–558; *II*, ibid., 777–815.

[63] T. Kaletha, A. Minguez, S. Shin et al., *Endoscopic Classification of Representations: Inner Forms of Unitary Groups*, arXiv preprint arXiv:1409.3731 (2014).

[64] M. Kashiwara and M. Vergne, *On the Segal–Shale–Weil representations and harmonic polynomials*, Invent. Math. **44** (1978), 1–47.

[65] K. Kato, *p-adic Hodge theory and values of zeta functions of modular forms*, Astérisque **295** (2004), 117–290.

[66] N. Katz and T. Oda, *On the differentiation of De Rham cohomology classes with respect to parameters*, J. Math. Kyoto Univ. **8** (1968), 199-213.

[67] Y. Kawamata, *A generalization of Kodaira-Ramanujam's vanishing theorem*, Math. Ann. **261** (1982), 43-46.

[68] M. Kisin, *Moduli of finite flat group schemes, and modularity*, Ann. Math. (2) **170** (2009), 1085-1180.

[69] H. Klingen, *Über die Werte der Dedekindschen Zetafunktion*, Math. Ann. **145** (1962), 265-272.

[70] M. Kneser, *Strong approximation*, in: Algebraic Groups and Discontinuous Subgroups, Proc. Sympos. Pure Math. **9** (1966), 187–196.

[71] J. Kollár, *Lectures on resolution of singularities*, Princeton, NJ: Princeton University Press (2007).

[72] S. Kudla, *Splitting metaplectic covers of dual reductive pairs*, Isr. J. Math. **87** (1994), 361-401.

[73] S. Kudla and S. Rallis, *A regularized Siegel-Weil formula: The first term identity*, Ann. Math. (2) **140** (1994), 1-80.

[74] J.-P. Labesse, *Changement de base CM et séries discrètes*, in: On the Stabilization of the Trace Formula, Stab. Trace Formula Shimura Var. Arith. Appl. **1**, International Press (2011), 429–470.

[75] K. Lan, *Arithmetic compactifications of PEL-type Shimura varieties*, Princeton, NJ: Princeton University Press (2013).

[76] E. Landau, *Über einen Satz von Tschebyschef.*, Math. Ann. **61** (1905), 527-550.

[77] W. Landherr, *Äquivalenz Hermitescher Formen über einem beliebigen algebraischen Zahlkörper*, Abh. Math. Sem. Univ. Hamburg **11** (1936), 245–248.

[78] R. P. Langlands, *On the functional equations satisfied by Eisenstein series*, Lecture Notes in Math. **544** (1976).

[79] J. Li, *Theta lifting for unitary representations with nonzero cohomology*, Duke Math. J. **61** (1990), 913-937.

[80] J. Li, *Non-vanishing theorems for the cohomology of certain arithmetic quotients*, J. Reine Angew. Math. **428** (1992), 177-217.

[81] U. V. Linnik, *On the least prime in an arithmetic progression. I. The basic theorem*, Mat. Sb. **15(57)** (1944), 139–178; *II. The Deuring–Heilbronn phenomenon*, ibid., 347–368.

[82] B. Mazur and A. Wiles, *Class fields of abelian extensions of* $\mathbb{Q}$, Invent. Math. **76** (1984), 179-330.

[83] W. Messing, *The crystals associated to Barsotti–Tate groups: with applications to abelian schemes*, Lecture Notes in Math. **264** (1972).

[84] J. Milne, *Arithmetic duality theorems*, Charleston, SC: BookSurge, LLC (2006).

[85] A. Mínguez, *Correspondance de Howe explicite: paires duales de type II*, Ann. Sci. Éc. Norm. Supér. (4) **41** (2008), 717–741.

[86] C. Mœglin, M.-F. Vignéras and J.-L. Waldspurger, *Correspondances de Howe sur un corps p-adique*, Lecture Notes in Math. **1291** (1987).

[87] C. Moeglin and J. Waldspurger, *Spectral decomposition and Eisenstein series. Transl. from the French by Leila Schneps*, Cambridge: Cambridge Univ. Press (1995).

[88] C. Mok, *Endoscopic classification of representations of quasi-split unitary groups*, Providence, RI: American Mathematical Society (AMS) (2015).

[89] H. Montgomery and R. Vaughan, *Multiplicative number theory. I. Classical theory*, Cambridge: Cambridge University Press (2007).

[90] L. Moret-Bailly, *Pinceaux de variétés abéliennes*, Astérisque **129** (1985).

[91] A. Moriwaki, *Continuity of volumes on arithmetic varieties*, J. Algebr. Geom. **18** (2009), 407-457.

[92] A. Moriwaki, *Arakelov geometry*, Providence, RI: American Mathematical Society (AMS) (2014).

[93] D. Mumford, *Abelian varieties*, London: Oxford University Press (1970).

[94] D. Mumford, *An analytic construction of degenerating Abelian varieties over complete rings*, Compos. Math. **24** (1972), 239-272.

[95] D. Mumford, *Hirzebruch's proportionality theorem in the non-compact case*, Invent. Math. **42** (1977), 239-272.

[96] J. Neukirch, *Algebraic number theory. Transl. from the German by Norbert Schappacher*, Berlin: Springer (1999).

[97] J. Neukirch, A. Schmidt, and K. Wingberg, *Cohomology of number fields*, Berlin: Springer (2008).

[98] T. Ohsawa and K. Takegoshi, *On the extension of L^2 holomorphic functions*, Math. Z. **195** (1987), 197–204.

[99] I. Piatetski-Shapiro and S. Rallis, *L-functions for the classical groups*, in: Explicit Constructions of Automorphic L-functions, Lecture Notes in Math. **1254** (1987), 1–52.

[100] V. Platonov and A. Rapinchuk, *Algebraic groups and number theory. Transl. from the Russian by Rachel Rowen*, Boston, MA: Academic Press (1994).

[101] S. Rallis, *On the Howe duality conjecture*, Compos. Math. **51** (1984), 333-399.

[102] K. A. Ribet, *A modular construction of unramified p-extensions of $\mathbb{Q}(\mu_p)$*, Invent. Math. **34** (1976), 151–162.

[103] W. Schmid, *Some properties of square-integrable representations of semisimple Lie groups*, Ann. of Math. (2) **102** (1975), 535–564.

[104] P. Scholze, *The local Langlands correspondence for GL_n over p-adic fields*, Invent. Math. **192** (2013), 663–715.

[105] P. Scholze and J. Weinstein, *Moduli of p-divisible groups*, Camb. J. Math. **1** (2013), 145-237.

[106] J.-P. Serre, *Local class field theory*, in: Algebraic Number Theory (J. W. S. Cassels and A. Fröhlich, eds.), Academic Press (1967), 128–161.

[107] J.-P. Serre, *Abelian ℓ-adic representations and elliptic curves*, W. A. Benjamin (1968).

[108] J.-P. Serre, *Formes modulaires et fonctions zêta p-adiques*, Lecture Notes in Math. **350** (1973), 191–268.

[109] J. Serre, *Galois cohomology. Transl. from the French by Patrick Ion*, Berlin: Springer (1997).

[110] A. Grothendieck (with M. Raynaud and D. S. Rim), *Groupes de monodromie en géométrie algébrique (SGA 7 I)*, Lecture Notes in Math. **288** (1972).

[111] G. Shimura, *Arithmetic of unitary groups*, Ann. Math. (2) **79** (1964), 369-409.

[112] G. Shimura, *Introduction to the arithmetic theory of automorphic functions*, Publ. Math. Soc. Japan **11**, Iwanami Shoten and Princeton University Press (1971).

[113] G. Shimura, *Euler products and Eisenstein series*, Providence, RI: American Mathematical Society (1997).

[114] S. Shin, *Galois representations arising from some compact Shimura varieties*, Ann. Math. (2) **173** (2011), 1645-1741.

[115] C. L. Siegel, *Über die Classenzahl quadratischer Zahlkörper*, Acta Arith. **1** (1935), 83–86.

[116] C. L. Siegel, *Über die Fourierschen Koeffizienten von Modulformen*, Nachr. Akad. Wiss. Göttingen Math.-Phys. Kl. II (1970), 15–56.

[117] W. Sinnott, *On the μ-invariant of the Γ-transform of a rational function*, Invent. Math. **75** (1984), 273-282.

[118] C. Soulé, *K-théorie des anneaux d'entiers de corps de nombres et cohomologie étale*, Invent. Math. **55** (1979), 251–295.

[119] C. Soulé, *A bound for the torsion in the K-theory of algebraic integers*, Doc. Math. **Extra Vol.** (2003), 761-788.

[120] S. Stepanov, *On the number of points of a hyperelliptic curve over a finite prime field*, Izv. Akad. Nauk SSSR, Ser. Mat. **33** (1969), 1171-1181.

[121] U. Stuhler, *Eine Bemerkung zur Reduktionstheorie quadratischer Formen*, Arch. Math. **27** (1976), 604-610.

[122] V. Tan, *Poles of Siegel Eisenstein series on $U(n,n)$*, Can. J. Math. **51** (1999), 164-175.

[123] J. Tate, *Duality theorems in Galois cohomology over number fields*, in: Proc. Internat. Congr. Mathematicians (Stockholm, 1962), Inst. Mittag-Leffler (1963), 288–295.

[124] J. T. Tate, *Fourier analysis in number fields, and Hecke's zeta-functions*, in: Algebraic Number Theory (J. W. S. Cassels and A. Fröhlich, eds.), Academic Press (1967), 305–347.

[125] J. T. Tate, *Global class field theory*, in: Algebraic Number Theory (J. W. S. Cassels and A. Fröhlich, eds.), Academic Press (1967), 162–203.

[126] J. T. Tate, *p-divisible groups*, in: Proc. Conf. Local Fields (Driebergen, 1966), Springer (1967), 158–183.

[127] R. Taylor, *On Galois representations associated to Hilbert modular forms*, Invent. Math. **98** (1989), 265–280.

[128] E. C. Titchmarsh, *The theory of the Riemann zeta-function*, 2nd ed., revised by D. R. Heath-Brown, Oxford University Press (1986).

[129] E. Viehweg, *Vanishing theorems*, J. Reine Angew. Math. **335** (1982), 1-8.

[130] A. I. Vinogradov, *The density hypothesis for Dirichlet L-series*, Izv. Akad. Nauk SSSR Ser. Mat. **29** (1965), 903–934.

[131] N. Wallach, *Real reductive groups II*, Boston, MA etc.: Academic Press, Inc. (1992).

[132] L. Washington, *Introduction to cyclotomic fields.*, New York, NY: Springer (1997).

[133] A. Weil, *Sur la formule de Siegel dans la théorie des groupes classiques*, Acta Math. **113** (1965), 1-87.

[134] A. Wiles, *The Iwasawa conjecture for totally real fields*, Ann. Math. (2) **131** (1990), 493-540.

[135] J. Włodarczyk, *Simple Hironaka resolution in characteristic zero*, J. Am. Math. Soc. **18** (2005), 779-822.

[136] X. Yuan, *Big line bundles over arithmetic varieties*, Invent. Math. **173** (2008), 603-649.

[137] Y. Zarkhin, *Endomorphisms of Abelian varieties over fields of finite characteristic*, Math. USSR, Izv. **9** (1976), 255-260.

[138] S. Zhang, *Positive line bundles on arithmetic varieties*, J. Am. Math. Soc. **8** (1995), 187-221.

(Siniša Slijepčević) Department of Mathematics, University of Zagreb, Croatia
Cantab Pi, Cambridge, UK
E-mail address: sinisa@math.hr, sinisa@cantabpi.com