The Alpöge–Claude proof of the Riemann hypothesis explained for the Riemann zeta function

Claude Fable 5.2, Codex Astra 6.2, and Gemini Carbon
at the behest of Michael O. Rubinstein

Working draft, 21 November 2026

Abstract. We explain the argument of Alpöge and Claude for the Riemann hypothesis in the case of the Riemann zeta function. The analytic step shows that a one-sided bound for a smoothed sum over prime powers, valid throughout intervals $[Y, Y^{3/2}]$ with Y arbitrarily large, excludes zeros to the right of the critical line. We derive this bound from one inequality between weighted sums over primes, the main inequality (M), and two estimates for Dirichlet characters. We prove the two estimates from theorems in the literature, and explain how the divisibility and the size of theta lifts enter (M). The proof of (M) is taken from the Lean development. This is a work in progress: Part III writes up parts of that proof and states exactly the inputs it takes from the Lean development; (M) itself is not proved here.

Note. This is a working guide to Alpöge's Lean-verified proof of the Riemann hypothesis, specialised to the Riemann zeta function. Its goal is a human-readable exposition of the proof in the style of classical analytic number theory. It is a work in progress, written as an aid to studying the proof.

Written up so far: the deduction of the Riemann hypothesis from one inequality, called (M) throughout, and two estimates for Dirichlet characters (Parts I and II); proofs of those two estimates and the construction of two auxiliary sets (§III.1–III.3); the analytic part of the bound for the size of the theta kernels (§III.4; four passages there are conditional on inputs from the Lean development); a local divisibility estimate for power series at inert primes, under one hypothesis from the Lean development, proved for orbits of length two and with a further duality input for longer orbits (§III.5); the passage from that estimate to lattices: the lemmas on lattices are proved, the prime sum is split and several of its terms are bounded (the bounds for the remaining loss sums are quoted), and the theorem on the glued lattice is conditional on two statements of the Lean development (§III.6); the deduction of the size inequality from eleven named inputs, and the value of an index-corrected mean of periods under stated hypotheses (§III.7); the deduction of the dimension bound that is used from three statements of the Lean development (on traces, ranks and pairings) and stated conditions on sizes and subspaces, with a finite model of the trace estimates (§III.8); and the step from the bound for Galois cohomology to a bound for denominators, with an account of how the Lean development proves that bound (§III.9).

Not proved here, all of it inside (M), and taken from the Lean development: (1) the integrality of the theta kernels in the coordinates in which the divisibility is proved, and the statement on valuations built on it; (2) the arithmetic inputs of the size bound: the inner-product formula for the theta lifts and its local factors, the covolumes of the lattices, the local index bounds and the passage between levels; (3) the trace-formula estimates themselves; (4) the bound for Galois cohomology itself; (5) the construction of the maps and their weights, and the assembly of (M), which in the Lean development is a mean over a family of maps (§III.6 treats a model case only).

The table at the start of Part III indicates which statements are proved in these pages and which are still taken without proof from the Lean development.

Contents (entries are links)

Part I. How the proof works

I.0 The strategy in one page 3
I.1 From zeros to a sum over primes 4
I.2 Chebyshev's method, and where it stalls 5
I.3 A divisible number with smooth exponents 6
I.4 Repairing the three defects 7
I.5 The inequality that comes out 13
I.6 From the table to $G(x)$ 14
I.7 Where (M) comes from 16
I.8 Why the exponent is one half 18
I.9 Dictionary with Alpöge's announcement 19
I.10 The same proof for Dirichlet L-functions 21
I.11 Notes on the literature 22
I.12 Summary: what is proved where 25

Part II. Details and proofs

1. What a zero forces 26
2. How divisibility counts primes 28
3. The reduction to one prime sum 29
4. The main inequality (M) 30
5. From the table to a single prime sum . . . 33
6. Where the main inequality (M) comes from 37
Sources 53

Part III. Proofs of the quoted statements

III.1 The auxiliary primes: proof of Estimate B 55
III.2 The large scale: proof of Estimate C . 61
III.3 The sets of columns T_2 and T 64
III.4 The size of the theta kernels (partly written) 66
III.5 Divisibility at inert primes 80
III.6 Divisibility: from derivatives to lattices 91
III.7 The size bound: the arithmetic inputs . 111
III.8 The images in general position: the trace-formula estimates 129
III.9 The bound for Galois cohomology . . . 144
Computations 161
References 161

What this is. In November 2026 Levent Alpöge announced a proof, found with the AI system Claude, of the Riemann hypothesis. It exists as a formal proof in the Lean proof assistant, about 17 million lines long. This paper explains how the argument works for the Riemann zeta function, for a reader who knows a first course in analytic number theory: the prime number theorem, Dirichlet characters, and Mellin transforms.

How it is organised. Part I is an exposition: the strategy first, then each step with its motivation and with examples computed on real primes. Part II gives the details and proofs to which Part I refers. A reader can stop after Part I and know how the proof goes; Part II is there to check each step. Part III, which is being written, is to prove the statements that Parts I and II quote. It proves Estimates B and C and constructs the two auxiliary sets of residue classes (the "sets of columns" of §I.5). For the main inequality (M) it proves some steps, proves others conditionally on statements of the Lean development which it names, and takes the rest from that development; the table at the start of Part III says which.

What is proved here and what is not. The argument rests on three statements. One of them, the main inequality (M), is the arithmetic heart of the matter. Two further estimates (B and C) concern auxiliary Dirichlet characters. We state B and C exactly, and (M) in the form in which it is used, and explain why one should expect them. That the three together imply the Riemann hypothesis is proved in Part II. Estimates B and C are proved in Part III, from theorems on Dirichlet L-functions that are cited there. But (M) is not yet proved here: we are working through its proof in the Lean development and adding the details to Part III section by section. Of what it needs, Part III so far constructs the two sets of columns (§III.3) and proves: the analytic part of the bound for the size of the theta kernels, with four conditional passages (§III.4); a local divisibility estimate for power series at inert primes, under the integrality of the theta kernels in the coordinates of the height-two logarithm (hypothesis (B)), for orbits of length two and with a further

duality input for longer orbits (§III.5); the lemmas on lattices that carry this estimate to lattices and part of the sum over primes, the main theorem there being conditional on two statements of the Lean development (§III.6); that the size inequality follows from eleven named inputs (§III.7); that the dimension bound follows from three statements of the Lean development on traces, ranks and pairings (§III.8); and the step from the bound for Galois cohomology to a bound for denominators (§III.9). Part II proves the elementary lemmas. The rest of (M) is for now *quoted*: hypothesis (B) and the statement on valuations built on it, the arithmetic inputs of the size estimate, the trace-formula estimates themselves, the bound for Galois cohomology itself, the construction of the maps and their weights, and the assembly of (M) from these. Statements marked [quoted] are taken from the Lean development without proof. The full proof of (M) is where almost all of the formal development goes, and writing it up is the work in progress. Every numerical example, except where stated, can be reproduced with the scripts listed under Computations, after Part III.

Part I. How the proof works

I.0 The strategy in one page

Chebyshev's idea about counting primes. Chebyshev [Che52] bounded the number of primes using the prime factorization of factorials. In the simplest form of the argument, due to Erdős [Erd32], one considers the middle binomial coefficient $\binom{2n}{n}$. It is divisible by every prime between n and $2n$, and it is at most 4^n. So the product of those primes is at most 4^n. That is an upper bound for a count of primes, obtained from two facts about one integer: what divides it, and how big it is.

One piece of notation. For a real number v we write v_+ for its *positive part*: $v_+ = v$ if $v > 0$, and $v_+ = 0$ otherwise. For example $5_+ = 5$ and $(-3)_+ = 0$. A sum of terms of the form $(\cdots)_+$ is automatically cut off where the terms would become negative.

The plan. The proof of the Riemann hypothesis described here is Chebyshev's argument, run with a far better "integer". There are five steps.

1. **Zeros to primes.** The Riemann hypothesis is equivalent to a bound $|G(x)| \le Cx^{1/2+\varepsilon}$ for a certain smoothed count of primes $G(x)$, for every $\varepsilon > 0$, with $C = C(\varepsilon)$. In fact an *upper* bound $G(x) \le x^{1/2+\varepsilon}$ is enough, and it is enough to have it on a sparse sequence of long intervals. (§I.1.)

2. **Why Chebyshev's method stalls.** For $n!$ the divisibility is known exactly: the exponent of p is $\sum_h \lfloor n/p^h \rfloor$. The floor functions are the obstacle; they are why the classical method gives constants but not asymptotics. (§I.2.)

3. **A divisible object with no floor functions.** Derivatives of functions on an elliptic curve with complex multiplication are divisible by primes to exponents of the smooth form $a/p + a/p^3 + \cdots$, but only at the primes that stay prime in the associated imaginary quadratic field. Subtracting a large weight w turns each term a/p^h into $(a/p^h - w)_+ = w\,(x/p^h - 1)_+$ with $x = a/w$: exactly the smooth kernel that G uses, magnified by w. (§I.3 and §I.4.)

4. **Size, to the precision needed.** "Divisibility $\le$ size" now gives an upper bound for a sum over primes. For it to be useful the size must be known not just to leading order $a \log a$ but through the term linear in a. This forces one to replace a single number by a linear map between lattices, built from *theta series*, whose size is governed by classical formulas

(Chowla–Selberg, the mass formula). The resulting comparison of divisibility with size is the main inequality (M). (§I.5 and §I.7.)

5. **Clean up and balance.** The sum in (M) runs over primes in certain residue classes. Averaging over the classes leaves $G(x)$ plus three sums twisted by quadratic characters, which are small for well-chosen auxiliary primes. Repeating the construction m times leaves two competing errors, x/m and m. Taking $m \approx \sqrt{x}$ gives $G(x) \le x^{1/2+\varepsilon}$. (§I.6 and §I.8.)

Logical skeleton.

$$\underbrace{\text{(M)}}_{\text{[quoted]}}\ ,\ \underbrace{\text{B, C}}_{\text{Part III}} \xrightarrow{\text{§I.6}} \text{Estimate A} \xrightarrow{\text{§I.6}} G(x) \le x^{1/2+\varepsilon} \text{ on long intervals} \xrightarrow{\text{§I.1}} \text{RH}. \tag{I.1}$$

Three statements enter: the main inequality (M), which is quoted, and two estimates for auxiliary Dirichlet characters, one at the scale x (Estimate B) and one at a much larger scale (Estimate C), which are proved in Part III from cited theorems. The order of choices is this. For each small ε there are arbitrarily large Y for which a single pair of auxiliary primes q, d serves for every x in $[Y, Y^{3/2}]$; the rest of the construction is then made for each x separately.

I.1 From zeros to a sum over primes

The statistic. Let $\Lambda(n) = \log p$ if $n = p^h$ is a prime power and 0 otherwise. Define

$$S(x) = \sum_{n \le x} \Lambda(n)\Big(\frac{x}{n} - 1\Big), \qquad G(x) = S(x) - x\log x + (1+\gamma)x, \tag{I.2}$$

where γ is Euler's constant. Equivalently $S(x) = \sum_n \Lambda(n)(x/n-1)_+$, where $v_+ = \max(v, 0)$: with the positive part, the condition $n \le x$ under the sum can be omitted. Compared with Chebyshev's $\psi(x) = \sum_{n \le x} \Lambda(n)$, the sum S weights n by the *ramp* $(x/n-1)_+$: as a function of n it is large for small n, decreases, reaches 0 at $n = x$, and is 0 for $n > x$. Also $S(x) = x\int_1^x \psi(t)\,dt/t^2$. The two subtracted terms are the main term of S, so G is a remainder.

Why this statistic. The Mellin transform of the ramp is simple: $\int_1^\infty (x/n-1)_+ x^{-s-1}dx = n^{-s}/(s(s-1))$ for $\operatorname{Re} s > 1$. Summing over n with weights $\Lambda(n)$,

$$F(s) = \int_1^\infty G(x)x^{-s-1}dx = -\frac{\zeta'(s)}{\zeta(s)} \cdot \frac{1}{s(s-1)} - \frac{1}{(s-1)^2} + \frac{1+\gamma}{s-1}. \tag{I.3}$$

The last two terms cancel the pole at $s = 1$ (both terms of its principal part). What remains are simple poles of F at the zeros ρ of ζ, with residue $-1/(\rho(\rho-1))$ at a simple zero, and a pole at $s = 0$. (The trivial zeros $-2, -4, \ldots$ also give poles; they produce the term $O(x^{-2})$ below.) Shifting the contour in the inverse transform gives the explicit formula (compare [Dav00, Ch. 17])

$$G(x) = \log 2\pi - \sum_\rho \frac{x^\rho}{\rho(\rho-1)} + O(x^{-2}), \tag{I.4}$$

the sum being over the nontrivial zeros. It converges absolutely, because of the factor $\rho(\rho-1)$; that is the benefit of smoothing. One has $\sum_\rho \frac{1}{\rho(1-\rho)} = 2 + \gamma - \log 4\pi = 0.0461\ldots$. If every zero has real part $\frac{1}{2}$ then $|\rho(\rho-1)| = \rho(1-\rho)$, and therefore, for large x, $|G(x) - \log 2\pi| \le 0.0462\sqrt{x}$. A zero with real part $\beta > \frac{1}{2}$ would contribute a term of size x^β.

A sanity check on real primes. Summing over the prime powers up to $2\cdot 10^6$ gives $G(10^4) = 1.749$, $G(10^5) = 3.671$, $G(10^6) = 11.040$, $G(2\cdot 10^6) = -12.268$, to be compared with $\log 2\pi = 1.838$ and $0.0462\sqrt{10^6} = 46.2$.

One side is enough. The terms of a zero off the line and of its complex conjugate together form a real function of x that oscillates and takes both signs. So one expects a zero with real part β to make $G(x)$ exceed $+cx^\beta$ for some arbitrarily large x. Thus an *upper* bound for G should already exclude it. That an upper bound suffices follows from a theorem of Landau on Mellin transforms of functions of one sign [Lan05]; see [MV07, §15.1].

Lemma (one-sided criterion)**.** *Let $0 < \theta < 1$. If $G(x) \le x^\theta$ for all large x, then $\zeta(s) \neq 0$ for $\operatorname{Re} s > \theta$.*

Proof. The integral $H(s) = \int_1^\infty (x^\theta - G(x))x^{-s-1}dx$ equals $1/(s-\theta) - F(s)$, and its integrand is nonnegative for large x. Landau's theorem says that a Mellin transform of an eventually nonnegative function has a singularity at the real point of its abscissa of convergence. But by (I.3) the function H is regular at every real $s > \theta$: the only real singularities of F in $s > 0$ would come from real zeros of ζ in $(0,1)$, and there are none. Hence the abscissa is at most θ, the integral converges for $\operatorname{Re} s > \theta$, and F is holomorphic there. A zero of ζ in that half-plane would be a pole of F. □

The arithmetic argument below produces the upper bound only for x in certain long intervals. The lemma survives this weakening.

Lemma (the same, on intervals)**.** *Let $0 < \theta < 1$. If there are arbitrarily large Y such that $G(x) \le x^\theta$ for all $x \in [Y, Y^{3/2}]$, then $\zeta(s) \neq 0$ for $\operatorname{Re} s > \theta$.*

The proof replaces Landau's theorem by an estimate of high derivatives of F, which see only one interval at a time; it is one page and is given in full in Part II, §1.

Conclusion of step 1. To prove the Riemann hypothesis it suffices to show: for every $\varepsilon > 0$ there are arbitrarily large Y with $G(x) \le x^{1/2+\varepsilon}$ throughout $[Y, Y^{3/2}]$. (Zeros come in pairs ρ, $1-\rho$, so no zeros to the right of the line means none to the left.) *From now on the goal is an upper bound for a sum over primes.*

I.2 Chebyshev's method, and where it stalls

The method. Let $N = \binom{2n}{n}$. Every prime p with $n < p \le 2n$ divides $(2n)!$ and not $(n!)^2$, so it divides N. Also $N \le (1+1)^{2n} = 4^n$. A positive integer divisible by distinct primes is at least their product, so

$$\sum_{n<p\le 2n} \log p \;\le\; \log N \;\le\; 2n\log 2. \tag{I.5}$$

In general: *if Δ is a nonzero integer and p^{e_p} divides Δ for each prime p in some set, then*

$$\sum_p e_p \log p \;\le\; \log|\Delta|. \tag{I.6}$$

Divisibility is a statement about each prime separately; size is a statement about the real numbers. The inequality (I.6) is the bridge, and it always points the same way: it bounds primes from *above*. By step 1 that is the direction we need. This is the arithmetic analogue of Stepanov's count of zeros with multiplicity; see Part II, §2.

The best classical integer. For $\Delta = n!$ both sides are known exactly. Legendre's formula [Leg08] says the exponent of p in $n!$ is $\sum_{h\ge 1}\lfloor n/p^h\rfloor$. Hence

$$\log n! = \sum_{r\le n} \Lambda(r)\Big\lfloor \frac{n}{r}\Big\rfloor. \tag{I.7}$$

The left side is $n\log n - n + O(\log n)$ by Stirling. If the floor were not there, the right side would be $n\sum_{r\le n}\Lambda(r)/r$, which is $S(n)+\psi(n)$: essentially our statistic.

Where it stalls. The floor *is* there. Writing $\lfloor n/r\rfloor = n/r - \{n/r\}$, the fractional parts contribute $\sum_{r\le n}\Lambda(r)\{n/r\}$, which is of size n: as large as the term $-n$ on the left. All information about $\psi(n)-n$ is hidden in the fractional parts, and (I.7) cannot separate it. This is why Chebyshev's method yields $c_1 n \le \psi(n) \le c_2 n$ and no more.

What would be needed. A nonzero "integer" Δ such that:

(W1) the exponent of p^h in Δ is a *smooth* function of p^h, like the ramp $(x/p^h - 1)_+$, with no floor function, or with rounding that is negligible;

(W2) $\log|\Delta|$ is known with an error much smaller than its linear term, in fact smaller by a factor of about $x^{-1/2}$, where x is the scale at which the primes are tested (§I.1);

(W3) the primes that occur are all primes, or a set of primes we can relate to all primes.

The rest of the proof constructs such an object. No single integer does all three; the final object is a linear map between lattices, and (I.6) becomes an inequality between the index of a lattice and the norm of a map.

I.3 A divisible number with smooth exponents

An example to compute with. The lemniscate sine $\mathrm{sl}(z)$ is the inverse function of $t \mapsto \int_0^t du/\sqrt{1-u^4}$. It satisfies $\mathrm{sl}'' = -2\,\mathrm{sl}^3$, and from this its Taylor coefficients can be computed exactly. Let $A_n = \mathrm{sl}^{(n)}(0)$, the nth derivative at the origin. These are integers, studied by Hurwitz [Hur99]; $A_n = 0$ unless $n \equiv 1 \bmod 4$:

$$A_1 = 1, \quad A_5 = -12, \quad A_9 = 3024, \quad A_{13} = -4390848, \quad A_{17} = 21224560896 = 2^8\cdot 3^6\cdot 7^2\cdot 11\cdot 211. \tag{I.8}$$

The curve $y^2 = 1 - x^4$ has an extra symmetry $x \mapsto ix$. One says it has *complex multiplication* by $\mathbb{Z}[i]$. A prime $p \equiv 3 \bmod 4$ remains prime in $\mathbb{Z}[i]$ (it is *inert*); a prime $p \equiv 1 \bmod 4$ factors (it *splits*).

	exponent of p in A_n, and $n\,p/(p^2-1)$, for inert p						split p	
n	$p=3$	$3n/8$	$p=7$	$7n/48$	$p=11$	$11n/120$	$p=5$	$p=13$
17	6	6.4	2	2.5	1	1.6	0	0
49	19	18.4	7	7.1	4	4.5	0	0
101	37	37.9	14	14.7	9	9.3	0	0
201	75	75.4	29	29.3	18	18.4	0	0

For every inert prime tested $(3, 7, 11, 19, 23, 31)$ and every $n \le 201$ with $n \equiv 1 \bmod 4$ the exponent of p in A_n is at least

$$n\cdot\frac{p}{p^2-1} = n\Big(\frac{1}{p} + \frac{1}{p^3} + \frac{1}{p^5} + \cdots\Big), \tag{I.9}$$

minus 1. In this range it exceeds $n\,p/(p^2-1)$ by at most 1.65; this ceiling does not persist for larger n (for $p=3$ the excess is 4.625 at $n=1945$). The split primes $5, 13, 17, 29$ divide none of these A_n. (Split primes are not excluded altogether: 157 divides A_{101}. The point is that nothing forces them to divide.)

Why. Near the origin the addition law of an elliptic curve is a *formal group*: a power series $X \oplus Y$ in two variables. Its logarithm $l(X)$ is the power series with $l(X \oplus Y) = l(X) + l(Y)$; here $l(t) = \int_0^t du/\sqrt{1-u^4}$ and $\mathrm{sl} = l^{-1}$. For a curve with complex multiplication and an inert prime p, multiplication by p on the formal group looks like $X \mapsto X^{p^2}$ modulo p, and the logarithm behaves like the model series

$$l(X) = X + \frac{X^{p^2}}{p} + \frac{X^{p^4}}{p^2} + \cdots . \tag{I.10}$$

(From here on we write a for the order of the derivative, as in the rest of the paper.) For this model one proves by induction that the coefficient of z^a in the inverse series has p-adic valuation at least $-(a-1)/(p^2-1)$ (Part II, §2). An ath derivative multiplies that coefficient by $a!$, whose valuation is $a/(p-1) - O(\log a)$. The difference is

$$a\Big(\frac{1}{p-1} - \frac{1}{p^2-1}\Big) = a \cdot \frac{p}{p^2-1} = a\Big(\frac{1}{p} + \frac{1}{p^3} + \cdots\Big). \tag{I.11}$$

Only odd powers of p appear. That is the signature of an inert prime. At a split prime the formal group is of a different type (height one): its logarithm has the denominator p^j already at degree p^j, the inverse series loses about $a/(p-1)$, and this cancels the whole valuation of $a!$. There is no gain. That an elliptic curve with complex multiplication has the first type at inert primes and the second at split primes is a theorem of Deuring [Deu41]; for the shape of the logarithm see Honda [Hon70] and Lubin–Tate [LT65].

What this achieves, and what it does not. We have integers whose exponents are smooth in p: wish (W1), up to a small amount of rounding. Three defects remain, and each drives a part of the real proof.

(D1) *Rounding.* An error of size about 1 in each exponent costs $\sum_{p\le n} \log p \approx n$ in (I.6): again as large as the linear term. This is the floor-function problem of §I.2 in another guise.

(D2) *Only half the primes.* Only inert primes appear, weighted by $1-\chi(p)$, that is, 2 for inert p and 0 for split p, where χ is the quadratic character of the field ($\chi(p) = -1$ for inert p).

(D3) *Size.* $|A_n|$ is about $n!/R^n$ with $R = 1.854\ldots$, so $\log|A_n| \approx n\log n$. But the inert primes account only for $\sum_{p\equiv 3(4),\, p\le n} \frac{n}{p}\log p = \frac{1}{2} n \log n + O(n)$. The inequality (I.6) is true and useless: it is off by a factor 2 in the *leading* term. (At $n = 101$: inert primes up to n contribute 167.7 of $\log|A_n| = 306.8$.)

For the objects of the real proof, the divisibility at inert primes is written up in §III.5, conditionally on the integrality of the theta kernels in the coordinates of the height-two logarithm (hypothesis (B)), which is taken from the Lean development.

I.4 Repairing the three defects

The lemniscate numbers A_n were derivatives of a function attached to an elliptic curve. The real proof keeps the idea *differentiate a function of arithmetic origin and look at what divides the result*, but changes the function. This section says which functions are differentiated, what "differentiating"

means for them, and what each choice achieves. It has four parts: the functions (§I.4.1), then the repair of (D3), (D1) and (D2) in that order (§I.4.2–I.4.4), and finally a list of what the reader can and cannot check directly (§I.4.5).

I.4.1 What is differentiated: theta series with a variable

The simplest case. Let $\Lambda = \mathbb{Z}[i] \subset \mathbb{C}$, the lattice of Gaussian integers, and put

$$\Theta(z) = \sum_{\lambda \in \Lambda} \exp\big(-\pi|\lambda|^2 + \lambda z \big) \qquad (z \in \mathbb{C}). \tag{I.12}$$

The Gaussian factor makes the sum converge for every z, so Θ is an entire function. Its nth derivative at the origin is

$$T_n = \Theta^{(n)}(0) = \sum_{\lambda \in \Lambda} \lambda^n e^{-\pi|\lambda|^2}. \tag{I.13}$$

So *differentiating n times at the origin means inserting the weight λ^n into the sum over the lattice.* More generally, for a polynomial P one has $P(\frac{d}{dz})\Theta(0) = \sum_\lambda P(\lambda) e^{-\pi|\lambda|^2}$: a weighted count of lattice points. Because Λ is preserved by multiplication by i, the sum T_n vanishes unless $4 \mid n$.

Three things visible in this example. They can all be checked by direct computation (script `e6_hecke_theta.py`).

1. *A period.* $T_0 = 1.18034\ldots = \sqrt{2}\,\varpi/\pi$ (a classical evaluation), where $\varpi = 2\int_0^1 dt/\sqrt{1-t^4} = 2.62205\ldots$ is the lemniscate constant: the *period* of the curve $y^2 = 1 - x^4$. And $T_4/T_0 = 0.12131\ldots = \frac{1}{4}(\varpi/\pi)^4$.

2. *Rationality.* After division by T_0 and by $(\varpi/\pi)^n$, the values are rational numbers:

$$r_k = \frac{T_{4k}}{T_0\,(\varpi/\pi)^{4k}}: \qquad r_1 = \tfrac{1}{4}, \quad r_2 = \tfrac{17}{16}, \quad r_3 = -\tfrac{111}{64}, \quad r_4 = \tfrac{12513}{256}, \quad r_5 = -\tfrac{146079}{1024}, \ \ldots \tag{I.14}$$

   The r_k were identified numerically: for $k \le 30$ the value agrees with a rational number of small denominator to 250 digits. The reason is the identity $\Theta(z)/T_0 = \sigma(u)^2 \wp(u)$ with $u = \varpi z/2\pi$, where σ and $\wp$ are the Weierstrass functions of the lattice with $g_2 = 4$, $g_3 = 0$, whose Taylor coefficients are rational; this identity was observed numerically, to more than 400 digits, and we have not written out a proof of it. It gives r_k with denominator 4^k. That lattice sums of this kind are algebraic multiples of powers of a period is a classical theorem; compare Damerell [Dam70]. So derivatives of Θ are *arithmetic* objects, once the period is removed. Each derivative costs one factor ϖ/π; this is why a term linear in the order of differentiation, $-a\log\Omega$, appears in the size estimate of §I.7.

3. *Gaussian size.* By the Cauchy–Schwarz inequality, $|T_n|^2 \le \big(\sum_\lambda |\lambda|^{2n} e^{-\pi|\lambda|^2}\big)\, T_0$, and the first factor is comparable to $\int_{\mathbb{C}} |z|^{2n} e^{-\pi|z|^2} dz = n!/\pi^n$. So $|T_n|$ is at most about $\sqrt{n!/\pi^n}$: the square root of a factorial, not a factorial. Numerically, $\log|T_{120}| = 119.0$, while $\frac{1}{2}\log(120!/\pi^{120}) = 160.2$ and $\log(120!) = 457.8$.

The general case. In the proof, Λ is replaced by a lattice of the following kind. Let E be a number field with complex multiplication (it is specified in §I.4.4 and in Part II, §6.1), and let L, L' be lattices over its ring of integers, of ranks r and r', each with a positive definite Hermitian form. The role of Λ is played by the lattice of linear maps from L to L'. The variable z becomes a complex $r \times r'$ matrix, one for each embedding of E in $\mathbb{C}$ from a fixed set of $2n$ embeddings (a *CM type*; $2n$ is half the degree of E, and n is the number of columns of the table of §I.5). The

theta series $\Theta_{L,L'}(z)$ is a holomorphic function of all these matrix entries, and one applies to it a differential operator $P(\partial)$, where P is a polynomial in the entries:

$$P(\partial)\,\Theta_{L,L'}(0) = \sum_{\nu} P_\nu\,\nu!\,c_\nu, \qquad \Theta_{L,L'}(z) = \sum_{\nu} c_\nu z^\nu, \quad P(z) = \sum_{\nu} P_\nu z^\nu. \tag{I.15}$$

Here ν runs over exponent vectors and $\nu!$ is the product of the factorials of its entries. As in item 2 above, these numbers become algebraic only after the period has been removed; in the proof this is done by writing the theta series in algebraic coordinates from the start. *The numbers* (I.15)*, so normalised, are the divisible objects of the proof* [quoted].

Why differentiate at all? There are three separate reasons.

- *It is forced by the construction.* The numbers (I.15), for suitable P, are exactly the entries of the matrix of a theta lift (§I.7) between spaces of forms of prescribed *weights*. The degree of P at a given embedding is one of the parameters (weights) of these spaces. We call it the *order* a, and reserve the word weight for the parameters w of §I.4.3. It can be chosen freely, and it will be chosen huge.
- *Differentiation creates divisibility.* This is the mechanism of §I.3: in a suitable algebraic coordinate the function has p-integral Taylor coefficients, an ath derivative multiplies a coefficient by $a!$, and at an inert prime the net gain is $a\,(1/p + 1/p^3 + \cdots)$.
- *Differentiation costs little in size.* By the Gaussian computation in item 3 above, an ath derivative has squared size about $a!$.

A warning, with numbers. The simplest theta function above does *not* have the full divisibility. At $n = 120$ the exponent of 3 in r_{30}, as identified above, is 14, whereas $n \cdot \frac{3}{8} = 45$. (For the lemniscate sine of §I.3 it was the full amount; for that function the size was too large.) Having both properties at once, full divisibility at inert primes *and* Gaussian size, is a theorem about the particular theta series used in the proof, written in particular coordinates. It is the deepest local statement in the argument, and it is quoted, not reproved, here [quoted]. (See §III.5 for what is proved: the divisibility at inert primes is written up in §III.5, conditionally on the integrality of the theta kernels in the coordinates of the height-two logarithm (hypothesis (B)), which is taken from the Lean development.)

I.4.2 Repair of (D3): squared Gaussian norms

Recall the defect: for the lemniscate numbers, divisibility at inert primes accounted for only half of $\log|A_n|$. The repair is point (i): the size becomes about $\sqrt{a!}$ in place of $a!$. Point (ii) fixes the units in which the two sides are then compared.

(i) Size is measured by a squared norm, and it is Gaussian. Give the polynomials in a complex variable z the inner product with weight $\pi^{-1}e^{-|z|^2}$. Then the monomials are orthogonal and

$$\|z^a\|^2 = \int_{\mathbb{C}} |z|^{2a}\,\pi^{-1}e^{-|z|^2}\,dz = \int_0^\infty t^a e^{-t}\,dt = a!. \tag{I.16}$$

For several variables the norms multiply: $\|z^\nu\|^2 = \nu!$. Put $\|P\|^2 = \sum_\nu |P_\nu|^2\nu!$. Applying the Cauchy–Schwarz inequality to (I.15), written as $\sum_\nu (P_\nu\sqrt{\nu!})(c_\nu\sqrt{\nu!})$, gives

$$\bigl|P(\partial)\Theta(0)\bigr|^2 \;\le\; \|P\|^2 \cdot \sum_\nu \nu!\,|c_\nu|^2, \tag{I.17}$$

the second sum being over the monomials that occur in P. For $P = z^a$ the first factor is $a!$. The second factor is a Gaussian-weighted norm of the theta function; bounding it is the analytic part of the size estimate. So the *squared* size of an entry is about $a!$ times lower-order factors: its logarithm is $a \log a + O(a)$, the term of order a being computed exactly in §I.7.

(ii) Both sides are measured for the squared norm. Here and below d is an auxiliary prime, $d \equiv 3 \bmod 4$, and the inert primes are those inert in $\mathbb{Q}(\sqrt{-d})$; see §I.4.4. After the period normalisation the entries (I.15) are algebraic numbers, in a field containing $\mathbb{Q}(\sqrt{-d})$. If such a number β is a nonzero algebraic integer divisible by p^e, for any prime p, then the *average* of $\log|\sigma(\beta)|^2$ over the complex embeddings σ of the field is at least $2e \log p$. This is the product formula. (At a single embedding it can fail, because of units; only the average is controlled.) In an imaginary quadratic field there is just one pair of embeddings and the statement is elementary: if $\beta = p^e\beta_0$ with β_0 a nonzero integer of that field, then $|\beta|^2 = p^{2e}|\beta_0|^2 \geq p^{2e}$. So divisibility by p^e contributes $2e \log p$ to the logarithm of the squared size.

Result. With $e \approx a/p$ at each inert prime $p \leq a$, the divisibility side is

$$\sum_{p \text{ inert}} 2 \cdot \frac{a}{p} \log p \;=\; a \log a + O(a), \tag{I.18}$$

since the inert primes are half of all primes and $\sum_{p \leq a} \frac{\log p}{p} = \log a + O(1)$ (Mertens' theorem for the residue classes modulo d; the $O(a)$ depends on d). The size side is $a \log a + O(a)$, the term of order a being computed exactly in §I.7. *The leading terms now agree*, and the comparison has become a statement about the terms of order a. Compare (D3): there the size was $\log|A_n|^2 \approx 2n \log n$ against the same $n \log n$ from divisibility.

I.4.3 Repair of (D1): a huge order, and a weight

What the exponent really is. For the entries (I.15) the exponent of an inert prime is not $a\,(1/p + 1/p^3 + \cdots)$ but a modification of it. Fix one of the $2n$ embeddings; call it the *place*. Let a be the order (the degree of P) at that place. Then, schematically,

$$\text{exponent of an inert prime } p \;\;\geq\;\; \sum_{h \text{ odd}} \Big(\frac{a}{p^h} - w_h\Big)_+ \;-\; \text{(small losses)}, \tag{I.19}$$

where $w_h > 0$ is a *weight* attached to another place: the one reached from the given place by the Frobenius automorphism of p^h followed by complex conjugation. The display keeps only the main term. The statement proved in the source is an inequality for sums of such exponents over the m lifts and over the places, with several loss terms: from the digits of the integers involved, from repeated visits of a Frobenius orbit to the same place, and from indices of lattices (Part II, §6.4) [quoted] (see §III.5 for what is proved).

The Frobenius and the places. The field E lies inside a field generated by roots of unity. The Frobenius of p is the automorphism of E induced by raising those roots of unity to the pth power. It permutes the $4n$ embeddings of E, and it depends only on residue classes of p. The $2n$ places are chosen embeddings, one from each pair of complex conjugates. At a prime that is inert in $\mathbb{Q}(\sqrt{-d})$ an odd power of the Frobenius carries each chosen embedding to one that is not chosen; composing with complex conjugation brings it back. This is the reason for the conjugation above. Since the Frobenius depends only on a residue class of p^h, residue classes of prime powers will appear in §I.5.

Where the subtraction of w comes from. The polynomial P is a product, over the $2n$ places, of minors (determinants of sub-matrices) of the matrix variable at each place. Its shape at

each place is described by a diagram of boxes. At the chosen place the diagram has a first row of length about a. At another place τ' the polynomial contains about $w_{\tau'}$ minors, where $w_{\tau'}$ is the weight parameter at τ'. At the scale p^h the Frobenius links the chosen place to the place τ_h, and each minor there can absorb only one derivative in a linked direction: in $\det(X + sb,\ Y + tb)$ there is no term in st, s^2 or t^2. Each derivative absorbed at τ_h costs one unit of the gain at the scale p^h. There are about w_h minors at τ_h, each absorbing at most one, and at most a/p^h units are available; so the loss at that scale is at most $\min(a/p^h, w_h)$, and what survives is $(a/p^h - w_h)_+$ (Part II, (6.1)). In the simplest case (a single index; Part II, (6.1a)) the loss comes from the digits of the integers involved, by Legendre's formula, and is of size $O((\log a)^2)$ for each prime (Part II, §6.4). The losses in the statement the source proves are larger and of several kinds, as said above.

Example. Take $a = 81$, $p = 3$, and $w_h = 2$ for all h.

h	1	3	5	7
p^h	3	27	243	2187
a/p^h	27	3	0.33	0.04
$a/p^h - w$	25	1	−1.67	−1.96
$(a/p^h - w)_+$	25	1	0	0

The exponent is at least $25 + 1 = 26$, up to the losses. Every term with $p^h > a/w = 40.5$ is zero.

The choice of a and w, and what it achieves. Let x be the scale at which we want information about primes. Put

$$b = \exp(x^{\varepsilon}), \qquad w \approx b, \qquad a \approx xb. \tag{I.20}$$

Then $a/w \approx x$, and the elementary identity

$$\Big(\frac{a}{p^h} - w\Big)_+ = w\,\Big(\frac{x}{p^h} - 1\Big)_+ \qquad (x = a/w) \tag{I.21}$$

shows what has been gained. (In the example, $x = 40.5$ and for $p^h = 3$ the right side is $2 \cdot (40.5/3 - 1) = 25$.)

- *Only prime powers up to x are involved,* at the places that carry the weight b. The right side of (I.21) vanishes for $p^h \geq x$. Although we differentiate about xb times, the primes that matter at those places are those below x. (At the places with small weights the cut-off is much larger; see below.)
- *The kernel is the ramp of §I.1.* The factor $(x/p^h - 1)_+$ is exactly the weight with which prime powers enter $G(x)$. This is where the smoothing in the definition of G comes from.
- *Rounding becomes invisible.* Each prime power below x now contributes about b times its ramp value. The bound (I.19) is used only where its main term is positive; elsewhere one uses only that the exponent is not negative, which is the integrality of the entries. The loss terms are of lower order, and after dividing the whole inequality by b they are absorbed in the error of (M). The precise accounting is part of the quoted statement. This removes defect (D1), and with it the floor-function problem of §I.2.

Why b must be exponentially large. To defeat rounding, a power of x would do. One other step needs more. There are $2n$ places and each needs a weight, but only some of them can be given the weight b (the columns that carry it form the set T_2 constructed in §III.3). The others are given very small weights w', so that they test primes at a scale $a/w' \geq \frac{1}{3}\exp(x^{\varepsilon/3})$; only at

such a scale is the distribution of primes in the relevant progressions known unconditionally with a good error term (§I.6(b)). The weights in the construction are positive integers, so $w' \geq 1$; this forces $a \geq \frac{1}{3}\exp(x^{\varepsilon/3})$, and hence b exponential in a power of x. The size of b also enters the control of denominators (§I.7). What is bounded there is a weighted sum of logarithms of orders, $\sum_\ell \log \ell \sum_a h_{\ell,a}$ in the notation of Part II, §6.8, and b is chosen so that the available bound is at most $\sqrt{b}$. After division by b this is negligible.

I.4.4 Repair of (D2): characters, and a table

Only inert primes: use a character. Let d be a prime with $d \equiv 3 \bmod 4$ and let $\chi_d(r) = (r/d)$ be the Legendre symbol. The primes inert in $\mathbb{Q}(\sqrt{-d})$ are those with $\chi_d(p) = -1$. A sum over odd powers of inert primes is a sum over all prime powers with the coefficient $1 - \chi_d(r)$:

$$1 - \chi_d(p^h) = \begin{cases} 2, & p \text{ inert, } h \text{ odd,} \\ 0, & \text{all other prime powers with } p \neq d. \end{cases} \tag{I.22}$$

(With this coefficient 2 the sum matches the factor 2 in $2e \log p$ of §I.4.2(ii).) So the divisibility side is $\sum_r \Lambda(r)\,(1 - \chi_d(r))\,(\cdots)$: the sum for ζ *minus* the sum for $L(s, \chi_d)$. To isolate ζ one must know that the χ_d-sum is small. This is arranged by choosing d well (§I.6(e)), and it is not circular.

Many places: a table. By (I.19) the weight met by a prime power $r = p^h$ depends on the place to which the Frobenius of r leads. The field used in the proof is

$$E = \mathbb{Q}(\zeta_N + \zeta_N^{-1},\ \sqrt{-q},\ \sqrt{-d}), \qquad \zeta_N = e^{2\pi i/N}, \quad N = 4n, \tag{I.23}$$

with n a large power of 2 and $q \equiv 3 \bmod 4$ a second auxiliary prime. Starting from the fixed place, the place reached by a prime power r is determined by two things: the residue class of r modulo N up to sign (the Frobenius sends $\zeta_N + \zeta_N^{-1}$ to $\zeta_N^r + \zeta_N^{-r}$), and the quadratic symbol $\chi_q(r)$ (the Frobenius multiplies $\sqrt{-q}$ by $\chi_q(r)$; the complex conjugation, which enters for the odd powers of inert primes that contribute here, then changes its sign, which only interchanges the names of the two rows). For general r prime to $2q$ this is a labelling of residue classes; the prime d itself is treated separately, with coefficient $\log d$ (Part II, §4.1). So the places can be arranged in a *table*, with two rows (the value of χ_q) and n columns (the residue class), and each prime power is assigned to a cell of the table. The inequality that comes out of the construction, stated in §I.5, is a bound for a sum over prime powers in which each prime power carries the weight of its cell. §I.6 then removes the dependence on the cell by averaging.

I.4.5 What the reader can check, and what is quoted

Statement in this section	Status
T_n, the period ϖ, the rational numbers r_k, the size of T_n	computed; script `e6_hecke_theta.py`
$\|z^a\|^2 = a!$ and inequality (I.17)	proved above
Identity (I.21) and the example $a = 81$	elementary
Leading terms agree: $\sum_{p \text{ inert}} \frac{2a}{p} \log p = a \log a + O(a)$	Mertens' theorem in progressions
A determinant absorbs one derivative per direction; the factorial count	proved in Part II, §6.4
The entries of the lift are the numbers (I.15)	quoted
The exponent bound (I.19) for the theta series of the proof	written up in §III.5, conditionally on the integrality of the theta kernels in the coordinates of the height-two logarithm (hypothesis (B)), which is taken from the Lean development

I.5 The inequality that comes out

Here is the statement. It has the form: there exist data (weights, two sets of columns, integers m, k) with the properties listed, such that the inequality (M) holds. The properties are given here in outline and exactly in Part II, §4.2–4.3. The statement looks elaborate; the example after it shows that it is elementary to state.

Data. Fix a small $\varepsilon > 0$ and a scale x. Let $q \equiv 3 \bmod 8$ and $d \equiv 7 \bmod 16$ be primes with $q \mid d+1$ (how they are chosen: §I.6). They are very large: if x lies in the interval $[Y, Y^{3/2}]$, then $\log q$ is about $Y^{\varepsilon/5}$ and $\log d$ about $Y^{\varepsilon/5}(\log Y)^2$. Let n be a power of 2 with $x^{2+8\varepsilon} \le n \le 2x^{2+8\varepsilon}$, and $N = 4n$. Let H be the group of odd residue classes modulo N with r and $-r$ identified; it has n elements, written $[r]$. Let $b = \exp(x^{\varepsilon})$.

The table. Picture a table with two rows $v \in \{0,1\}$ and n columns indexed by H. An odd prime power $r = p^h$, $p \neq q$, is assigned to the cell

$$\text{row } v(r) = \tfrac{1}{2}(1 + \chi_q(r)), \qquad \text{column } [r]. \tag{I.24}$$

The $2n$ cells are the $2n$ places of §I.4.3. Each cell i carries a positive weight w_i. A cell of row 1 is *distinguished* when its column lies in a set of columns of density δ; its weight is then (very nearly) b. All other weights are tiny by comparison: so small that $a/w_i \ge \frac{1}{3}\exp(x^{\varepsilon/3})$.

The prime sum and its main term. For a derivative order a with $|a/b - x| \le 1$ (it depends on the label t introduced below, and is written a_t in Part II), put

$$P = \sum_{p \neq 2,q} \sum_{h \ge 1} (1 - \chi_d(p^h)) \log p \Big(\frac{a}{p^h} - w_{\text{cell}(p^h)} \Big)_+, \tag{I.25}$$

$$M = a\Big(\log a - 1 - \gamma + \frac{L'}{L}(1, \chi_d) - \langle \log w \rangle \Big), \tag{I.26}$$

where $\langle \log w \rangle$ is the average of $\log w_i$ over the $2n$ cells. P is the divisibility side: by (I.21) it is a sum of ramps, one for each cell, at the scale a/w_i. M is its expected value if primes were perfectly distributed among the cells (§I.6(a); proved in Part II, §5.1). So $P - M$ measures the irregularity of primes in the cells of the table, at the scales a/w_i.

The primes 2 and q can be omitted without changing anything: $d \equiv 7 \bmod 8$ gives $\chi_d(2) = 1$, and $q \mid d+1$ gives $\chi_d(q) = 1$ by quadratic reciprocity, so the coefficient $1 - \chi_d$ vanishes on their powers. This is what the congruence conditions on q and d are for (Part II, §4.1). (The powers of d do occur, with coefficient $\log d$, but only at the large scale, since $d > x$.)

The inequality. The construction depends on three further choices: a translate $u \in H$ (which moves the distinguished cells to other columns), a label t in a set $T \subset H$ (which shifts the column of every prime power), and an index $j = 1, \dots, m$ numbering the m repetitions. Write $P_{u,t,j}$, $M_{u,t,j}$ accordingly.

$$\boxed{\frac{1}{n} \sum_{u \in H} \sum_{t \in T} \sum_{j=1}^{m} \big(P_{u,t,j} - M_{u,t,j}\big) \;\le\; \mathcal{E}, \qquad \frac{\mathcal{E}}{B} \le x^{C\varepsilon}\Big(\frac{x}{m} + \frac{xm}{k} + m + \frac{k}{n^2}\Big)} \tag{M}$$

Here $B = |T|\, m\, b\, \delta/2$ is the natural normalisation (δ is the proportion of distinguished columns), k is a further parameter of the construction (a rank), and C is an absolute constant. The inequality is asserted for $x \ge x_0(\varepsilon)$ [quoted]. In words: *on average over translates, the primes are not more numerous than expected in the table, up to the four errors shown.* It is an upper bound, as every instance of (I.6) is. The set T and the set of distinguished columns (called T_2 in Parts II and III) are constructed in Part III, §III.3; how the construction uses them is part of the quoted statement.

More precisely: let Y be large, let q, d be the primes of Estimate B for Y (§I.6(e)), and let $Y \le x \le Y^{3/2}$. Then the construction provides $m = x^{1/2+O(\varepsilon)}$, $k = x^{9/2+O(\varepsilon)}$, sets T, T_2 with $|T_2|/n = \delta \ge x^{-\varepsilon}$ and $|T|/n \ge \frac{1}{2}x^{-\varepsilon/8}$, orders a_t and weights as above, for which (M) holds. The constants C and $x_0(\varepsilon)$ do not depend on Y, q, d.

A table small enough to write out. Take $q = 3$, $d = 23$, $N = 16$, so $n = 4$ and $H = \{[1], [3], [5], [7]\}$. The prime 7 is a square modulo 3 and $7 \equiv 7 \bmod 16$: it goes in row 1, column [7], with coefficient $2 \log 7$ because 7 is a non-square modulo 23. The primes up to 61 with non-zero coefficient fall as follows.

v\column	[1]	[3]	[5]	[7]
0	17	–	5, 11, 53	23
1	–	19, 61	37, 43	7

(I.27)

With $x = 1000$, $b = 10$, $a = 10^4$, weight 10 in cell $(1, [1])$ and 0.05 elsewhere, summing over the prime powers gives $P = 95461.616$ and $M = 95444.464$: the excess $P - M$ is 17.152, against a main term of nearly 10^5. Moving the distinguished cell to the other three columns gives excesses 368.21, −129.25, −204.65. The mean of the four excesses is 12.86. (The example ignores one feature of the construction, that the weights are integers. The identities it illustrates hold for any positive weights.)

I.6 From the table to $G(x)$

This step is classical analytic number theory, and it is proved in full, from (M), B and C, in Part II, §4–5. Here are the ideas.

(a) Each cell is an arithmetic progression. The prime powers in a given cell are those in one pair of residue classes $\pm c$ modulo N with a prescribed value of χ_q; the coefficient $1 - \chi_d$ then selects the odd powers of primes inert in $\mathbb{Q}(\sqrt{-d})$. For fixed moduli and scale tending to infinity, the prime number theorem in progressions gives each cell the fraction $1/(2n)$ of the primes, half of them inert, and M is defined as the resulting main term, with its constants. The identity

$P - M = \sum_i w_i R_i(a/w_i)$, where $R_i(y)$ is the remainder of cell i at scale y (Part II, (5.2)), is then exact (Part II, (5.3)). No uniformity is claimed at this point. At the scale x there are far more cells than primes, and a single $R_i(x)$ need not be small; the argument uses only the average over translates in (c). At the large scale the uniformity is Estimate C.

(b) Two scales. A cell with weight w_i tests primes up to a/w_i. For a distinguished cell this is $a/b \approx x$: the scale we care about. For every other cell it is at least $\frac{1}{3}\exp(x^{\varepsilon/3})$, an enormous scale. *Why arrange this?* The construction needs weights on all $2n$ cells, but we want information at scale x only. At the enormous scale the distribution of primes in these progressions can be controlled with a saving better than any power of x. This is Estimate C below. It rests on the classical zero-free region for Dirichlet L-functions; it is available because $\log q$ and $\log d$ are much smaller than $x^{\varepsilon/3}$, and because q and d are chosen so that no exceptional real zero interferes. So those cells contribute a negligible amount. In the example they account for 4.18 of the mean excess 12.86.

(c) Averaging over translates. As u runs over H, the distinguished cells visit every column equally often. So the average over u no longer sees the residue class modulo N. It sees only the two quadratic conditions: row 1 means $\chi_q(r) = 1$, and the coefficient is $1 - \chi_d(r)$. The average of the distinguished part is therefore

$$\frac{\delta b}{2} \sum_{r \le x} \Lambda(r)\big(1 + \chi_q(r)\big)\big(1 - \chi_d(r)\big)\Big(\frac{x}{r} - 1\Big) \quad \text{minus its main term.} \tag{I.28}$$

In the example this accounts for the other 8.68 of the 12.86.

(d) Expanding the product. Since $(1+\chi_q)(1-\chi_d) = 1 + \chi_q - \chi_d - \chi_q\chi_d$, the last expression is $\frac{\delta b}{2} R_{q,d}(x)$ with

$$R_{q,d}(x) = G(x) + G_{\chi_q}(x) - G_{\chi_d}(x) - G_{\chi_q\chi_d}(x), \tag{I.29}$$

where, for a primitive character $\chi \neq 1$, $G_\chi(x) = \sum_{r \le x} \Lambda(r)\chi(r)(x/r - 1) + x\,\frac{L'}{L}(1,\chi)$ is the analogue of G for $L(s,\chi)$. *The function G appears with coefficient $+1$.* (The construction gives this at the scale a/b, with $|a/b - x| \le 1$; the passage to x costs $O(1 + \log x)$, see Part II, §5.4.) Dividing (M) by B, and using Estimates B and C below to bound the error terms (Part II, §5.4), gives:

Estimate A. *Let Y be large, q, d the primes of Estimate B, and $Y \le x \le Y^{3/2}$. For the parameters of the construction, with $m \approx x^{1/2}$, $k \approx x^{9/2}$, $n \approx x^{2+8\varepsilon}$,*

$$R_{q,d}(x) \;\le\; x^{C\varepsilon}\Big(\frac{x}{m} + \frac{xm}{k} + m + \frac{k}{n^2}\Big) \;\le\; x^{1/2+C'\varepsilon} \qquad (x \ge x_0(\varepsilon)). \tag{I.30}$$

(e) The three twisted sums. To pass from $R_{q,d}$ to G we need:

Estimate B. *Let $0 < \varepsilon < \frac{1}{52}$. For all large Y there are primes $q \equiv 3 \bmod 8$, $d \equiv 7 \bmod 16$ with $q \mid d+1$, $\log q \asymp Y^{\varepsilon/5}$, $\log d \asymp Y^{\varepsilon/5}(\log Y)^2$, such that $|G_\chi(x)| \le x^{1/2+\varepsilon}$ for $\chi = \chi_q,\ \chi_d,\ \chi_q\chi_d$ and all x in the interval $[Y, Y^{3/2}]$. (Proved in Part III as Theorem III.1.1(d), in a stronger form: exponent $\frac{1}{2} + \frac{\varepsilon}{2}$, and any interval $[Y, Y^A]$.)*

Estimate C. *For the same primes q, d and every $x \in [Y, Y^{3/2}]$, Y large, every one of the $4n$ primitive Dirichlet characters χ that occur when the cells of the table are expanded in characters (their conductors divide $4nqd$) satisfies $|G_\chi(y)| \le y\exp(-\kappa x^{\varepsilon/8})$ for all $y \ge \frac{1}{3}\exp(x^{\varepsilon/3})$, with a constant $\kappa > 0$; here $G_\chi = G$ for the trivial character. (Proved in Part III as Theorem III.2.1, where κ is called c' and may be any positive number once x is large. The proof uses a further property of q and d, part (e) of Theorem III.1.1.)*

Is this circular? It looks like assuming the Riemann hypothesis for three L-functions. It is not, for a reason of scale. The moduli q and d are astronomically larger than x: their logarithms are powers of x. The quantity $G_\chi(x)$ has two parts. One is the constant $\frac{L'}{L}(1,\chi)$, an infinite sum over prime powers. The other is the finite sum over $r \le x$. The tail of the constant beyond $Z = Y^{600}$ is shown to be small using a zero-density estimate for zeros close to the line $\operatorname{Re} s = 1$; this discards a small proportion of the moduli, and uses nothing about zeros near $\operatorname{Re} s = \frac{1}{2}$. What remains is a sum over $r \le Z$, which is still very short compared with q and d. Mean-square estimates over the modulus show that for *most* pairs (q, d) such a sum is as small as a random sum, about $\sqrt{x}$. One only needs one good pair. For C see (b) above. Estimates B and C are proved in Part III, from standard theorems on Dirichlet L-functions that are cited there.

Conclusion. From A and B (A itself uses (M), B and C): $G(x) \le R_{q,d}(x) + 3x^{1/2+\varepsilon} \le x^{1/2+C''\varepsilon}$ on the interval. By §I.1 this proves the Riemann hypothesis.

I.7 Where the main inequality (M) comes from

This is the part quoted from the Lean development; Part III has begun its proof: §III.4 (the analytic part of item 2 below) and §III.5 (item 1 below, conditionally on one hypothesis from the Lean development). Sections III.6–III.9 prove further steps, each from named statements of the Lean development, which are quoted: the passage from derivatives to lattices, in part (§III.6); the deduction of the size inequality from eleven inputs (§III.7); the deduction of the dimension bound of item 3 from three statements (§III.8); and the step from the bound for Galois cohomology to the bound for denominators (§III.9). The table at the start of Part III says what each section proves, proves conditionally, or quotes. The construction of the maps and weights and the assembly of (M) are not begun, and (M) is still quoted. We describe the objects and the three things that have to be proved about them.

Theta series. A theta series counts representations by a quadratic form. For the form $a^2 + b^2$,

$$\Theta(z) = \sum_{a,b\in\mathbb{Z}} s^{a^2+b^2} = 1 + 4s + 4s^2 + 4s^4 + 8s^5 + \cdots, \qquad s = e^{2\pi i z}. \tag{I.31}$$

Theta series have two properties at once: they are sums over a lattice, so their coefficients are integers with arithmetic meaning; and they are Gaussian in nature (the terms have absolute value $e^{-2\pi y(a^2+b^2)}$, where y is the imaginary part of z), so their sizes are governed by Gaussian integrals. These are the two properties needed in §I.4: divisibility of derivatives (§I.4.1) and Gaussian size (the repair of (D3), §I.4.2).

A theta lift is a map defined by a theta series in two sets of variables, used as the kernel of an integral: $(\mathcal{T}f)(h) = \int \Theta(g,h)\overline{f(g)}\,dg$. (It is conjugate-linear in f.) In the case used here both groups are *definite unitary groups*, for which everything is finite: a "function on the group" is a function on a finite set of lattices $L_1, \ldots, L_c$ (the classes in a genus), the integral is a finite sum weighted by $1/|\operatorname{Aut} L_i|$, and $\mathcal{T}$ is a matrix. Its entries are derivatives, at the origin, of the theta series attached to pairs of lattices. So *a theta lift is a matrix of derivatives of theta functions*, between two spaces that carry natural integer lattices (functions with integer values) and natural norms.

The role of (I.6)**.** Let $\mathcal{T} : L \to M$ be an injective linear map between lattices in Euclidean spaces. If $\mathcal{T}(L) \subset p^e M$ for some prime power, then, comparing volumes,

$$e \log p \;\le\; \log \|\mathcal{T}\| + \mu_{\max}(M) - \mu(L), \tag{I.32}$$

where $\|\mathcal{T}\|$ is the operator norm and $\mu = -\frac{1}{\text{rank}}\log(\text{covolume})$ is the *slope* of a lattice ($\mu_{\max}$: the largest slope of a sublattice). For $L = M = \mathbb{Z}$ this is (I.6). In the construction the lattices are modules over the integers of a field with complex multiplication, and one works with squared norms: (I.32) is doubled, so that $2e\log p$ is compared with $\log\|\mathcal{T}\|^2$, as in §I.4. Three things must be established.

1. Divisibility. *The matrix of the lift is divisible by p^e with e of the form $\sum_{h\ odd}(a/p^h - w)_+$ at inert primes.* The mechanism is that of §I.3, applied to theta functions of lattices over a field with complex multiplication: derivatives of order a gain $a(1/p + 1/p^3 + \cdots)$. The subtraction of w has an elementary source, explained in §I.4.3: each minor can absorb only one derivative in a given direction. Composing m lifts multiplies the matrices, so the exponents add: this is the analogue of a function vanishing to order m in Stepanov's proof of the Riemann hypothesis for curves over finite fields. The divisibility at inert primes is written up in §III.5, conditionally on the integrality of the theta kernels in the coordinates of the height-two logarithm (hypothesis (B)), which is taken from the Lean development. The passage from this estimate to lattices is the subject of §III.6. Proved there: the lemmas on slopes of lattices, the form of the prime sum and its splitting by classes of primes, and bounds for several of its terms. The theorem on the glued lattice is conditional on two statements of the Lean development. The bounds for the eleven loss sums and for their total are quoted, and so is the statement on valuations itself.

2. Size. *The squared norm of the lift has logarithm at most $a(\log a - 1 - \log\Omega_E) + \cdots$, with the linear term known exactly.* The leading term is the Gaussian integral $\|z^a\|^2 = a!$. The constants Ω and Ω_E are *periods*: Ω belongs to an elliptic curve, and Ω_E, defined below, to the field E. A derivative with respect to the algebraic invariant differential and one with respect to the analytic variable differ by a constant factor, the period; an ath derivative acquires its ath power, and Ω is the square of that factor. Its value is

$$\log\Omega = \gamma + \log\frac{2\pi}{\sqrt{d}} - \frac{L'}{L}(1,\chi_d). \tag{I.33}$$

In this form the identity is Kronecker's limit formula; Ω is then a geometric mean, over the ideal classes of $\mathbb{Q}(\sqrt{-d})$, of squared periods of the corresponding elliptic curves. Expressing it through values of the Gamma function is the formula of Lerch and of Chowla and Selberg [CS67], whose name the source uses. For the field $\mathbb{Q}(i)$, with d replaced by 4 and χ_d by the character modulo 4, exponentiating this formula gives $\Omega = 2\pi(\varpi/\pi)^2 = \Gamma(\frac{1}{4})^4/(4\pi^2)$ exactly: that is 2π times the square of the factor ϖ/π met in §I.4.1. The constant Ω above is the period for the elliptic curve. The constant that enters the size bound for the field E is Ω_E, with

$$\log\Omega_E = \log\Omega - \log\frac{\mathrm{rd}_E}{\sqrt{d}} = \gamma + \log\frac{2\pi}{\mathrm{rd}_E} - \frac{L'}{L}(1,\chi_d), \tag{I.34}$$

where rd_E is the root discriminant of E. Part II, §6.6 gives both normalisations; the cancellation of the linear terms is the same. This is how γ and $L'/L(1,\chi_d)$ enter the main term M. They are the same constants that appear in the prime sum, by §I.6(a); that the two sides produce the same linear term is the reason the method can work at all. The remaining linear terms come from the slopes in (I.32): the number of lattice classes counted with weights $1/|\mathrm{Aut}|$ (the Siegel mass formula) and the dimensions of the spaces (the Weyl dimension formula). After an exact cancellation of factorials, the total is M.

3. Denominators and rank. Two things can spoil (I.32).

- *Denominators.* The lift need not map integer-valued functions to integer-valued functions; one must find the right lattices. They are assembled prime by prime, using the freedom to twist

the lift by characters, and what cannot be removed is measured by certain Galois cohomology groups of the field (generalisations of class groups). A bound, by $\sqrt{b}$, for the logarithm of the product of the orders of these groups (over all primes ℓ and all twists up to x^{400}) is proved separately in the Lean development (it is quoted here); it uses the theory of automorphic forms and Galois representations, and is a large part of the formal development. Part III, §III.9 proves the step from this bound to the bound for denominators, and the vanishing of the groups in some cases over $\mathbb{Q}$ from cited theorems; the bound itself is quoted, with an account of its proof in the Lean development.

- *Rank.* A single map of small rank controls only the *maximal* slope of its target, which can be far from the average and would lose the main term. One therefore uses many maps. Their images cannot all be perpendicular to one another: the dimensions of the images add up to more than the dimension of the target. What is shown is a bound for their total overlap (a second moment; a trace formula gives it). It says that the images are spread evenly: the weighted sum, over the maps, of the adjoint of the map followed by the map is close in mean square to a multiple of the identity of the target. A lemma of linear algebra then says the images fill almost the whole target, so that the *mean* slope is controlled. Part III, §III.8 proves that the dimension bound of this lemma follows from three statements of the Lean development (on traces, ranks and pairings) and stated size conditions; the three statements, and the theorems behind them, are quoted.

The elementary parts. Several steps in this section are elementary and are proved, with worked examples, in Part II, §6: the persistence of integrality through a chain of maps; the minor congruence; the factorial count (from Legendre's formula) giving the gain $a\,p/(p^2-1)$; the Gaussian norm; the cancellation that leaves M; the elementary slope inequality; and the linear-algebra lemma on images that are spread evenly in this sense. The analytic part of the size estimate is proved in part in Part III, §III.4: the Gaussian bound for one block, for one map and for a chain, with four passages conditional on inputs from the Lean development. Part III, §III.7 proves that the size inequality (III.75) follows from eleven named inputs, and that an index-corrected mean of periods equals $\log \Omega_E$ under hypotheses stated there. The eleven inputs are quoted, among them the inner-product formula, the covolumes and the local index bounds. So the size bound itself (Part II, (6.5), first line) remains quoted.

I.8 Why the exponent is one half

After division by B, the errors in (M) are, up to factors $x^{C\varepsilon}$,

$$\frac{x}{m} \qquad \frac{xm}{k} \qquad m \qquad \frac{k}{n^2}. \tag{I.35}$$

- x/m: errors made *once*, in the sizes of the lattices at the two ends of the chain of maps. The gain from divisibility is collected m times, so these are divided by m. This is why one composes many lifts.
- m: errors accumulated over the m steps. The budget of each step (ranks, weights, denominators) itself grows in proportion to m, so the total is of order m^2 before normalisation; the factor m contained in B leaves m.
- xm/k and k/n^2: costs of the auxiliary parameters k (the rank) and n (the size of the table). With $k \approx x^{9/2}$ and $n \approx x^2$ the first is negligible (x^{-3}) and the second is again of order $\sqrt{x}$.

The terms x/m and m are balanced when $m \approx \sqrt{x}$, and both equal $\sqrt{x}$. *That is where the exponent* $\frac{1}{2}$ *comes from.* If the second cost were m^2 instead of m after normalisation, one would have $x/m + m^2$, balanced at $m = x^{1/3}$, and the same analytic argument would give only that $\zeta(s) \neq 0$ for $\operatorname{Re} s > \frac{2}{3}$, provided the corresponding inequality could be proved with the other errors below $x^{2/3+\varepsilon}$. (Whether such an inequality can be proved without the estimate for Galois cohomology is a separate question, which we do not address.) So the precision of the comparison of lattices is decisive.

I.9 Dictionary with Alpöge's announcement

The announcement[1] lists twenty steps, in the order in which they were found between 6 October and 16 November 2026. It is a record of discovery, so it contains earlier formulations and motivating ideas as well as steps of the final argument. This section says where each step appears in this paper. The correspondence is our reading. For steps 1, 2, 5 and 7–20, and for the Poincaré bundle in step 3, it rests on statements of the Lean development that were read in the course of this work. Steps 4 and 6, and the rest of step 3, were not found in the Lean development and are matched from the description in the announcement.

Steps that are in this paper.

[1] https://alpo.ge/remark.pdf, pages 3–4.

	Step in the announcement	In this paper
1	Stepanov determinant	§I.2: a nonzero integer divisible by $\prod p^{e_p}$ is at least that large, (I.6); the model is Stepanov's count of points on curves
5	Switch to $\sum_{n\le X}\Lambda(n)(X/n-1)-X\log X+(1+\gamma)X$	§I.1: this is $G(X)$
8	Iterated theta lifts between totally definite unitary groups; a chain of length about $\sqrt{X}$	§I.7 (theta lift; composition); §I.8 ($m\approx\sqrt{x}$)
9	Norms of the lifts; cancellation using Chowla–Selberg, Weyl, Siegel mass	§I.7, "Size"; Part II, §6.6–6.7; the norms, in part: Part III, §III.4; the size inequality from eleven quoted inputs: §III.7
10	Lower bound for the divisibility of the m-fold lift	§I.4.3, (I.19); §I.7, "Divisibility"; written up in Part III, §III.5, conditionally on the integrality of the theta kernels in the coordinates of the height-two logarithm (hypothesis (B)), which is taken from the Lean development; the passage to lattices, in part: §III.6
12	Seesaw: the period of the iterated lift is a Bessel period independent of the Hecke characters	§I.7, "Denominators" (the freedom to twist by characters); Part II, §6.8
13	Bost slope inequality	§I.7, the inequality for a map of lattices; the elementary case is proved in Part II, §6.8; the form with minimal slopes used in the chain is quoted (§6.10)
14	Cost of each lift; "one would stop at 2/3"; need for "division"	§I.8: $x/m+m^2$ gives $\frac{2}{3}$, $x/m+m$ gives $\frac{1}{2}$
15	Glue an integral structure across Hecke characters (a "gcd")	§I.7, "Denominators": lattices assembled prime by prime; Part II, §6.8
16	Relative trace formula; approximate orthogonality	§I.7, "Rank"; Part II, §6.9–6.10
17	Classes in H^1 not killed by twisting; a weakening of Vandiver's conjecture	§I.7, "Denominators": the bound by $\sqrt{b}$ for Galois cohomology; Part II, §6.8
20	GRH	here only ζ, the case of the trivial character

Steps that appear here in another form, or only in part.

- *Step 2* (the invariant derivation D on a curve with complex multiplication satisfies $D^p\equiv 0$ at an inert prime; divisibility of $D^a\,\Omega\cdot\theta(0)$). This is the idea of §I.3 and §I.4.1: derivatives on such a curve, the period (ϖ in §I.4.1, Ω in §I.7), and exponents involving odd powers of p only. We could not confirm the stated exponent on the simplest example: for the theta function of $\mathbb{Z}[i]$ the exponent we computed is smaller (§I.4.1). The object or its normalisation in the announcement must differ from that model. For what is proved in this paper about the divisibility at an inert prime, see Part III, §III.5.
- *Steps 4 and 6* (positivity $W(f)\ge 0$ in Weil's explicit formula, obtained from a determinant; a zero-free region from such a positivity). These are an earlier form of our first step. In the final argument the statement "an upper bound excludes zeros" is the Landau-type lemma of §I.1.
- *Step 7* (determinants containing Gan–Gross–Prasad periods; interlacing weights). In the Lean development the period pairing at the end of the chain of lifts is identified with a Bessel period (§I.7; Part II, §6.2 and §6.8; source file [19]), and a condition of interlacing on the weights governs the passage between consecutive unitary groups (source file [23]). We do not use the Gan–Gross–Prasad terminology in the text.

- *Steps 11 and 19* (an Iwasawa tower; a construction of Behrend). The tower is the field E with n a large power of 2, which gives the columns of the table (§I.4.4, §I.5). Behrend's construction is used to choose the set T_2 of distinguished columns (Part II, §4.2). The set is constructed in Part III, §III.3, following source file [27], which uses the sphere construction of Behrend as formalised in Mathlib. (The definitions in [17] are an earlier formulation; the final theorem does not use them.) The announcement itself expects Behrend's construction to be unnecessary.
- *Step 18* (covolumes; the index of theta series; congruences between lifts and forms). These are the slopes of the lattices in §I.7. The elementary inequality for slopes is proved in Part II, §6.8; the form with minimal slopes, the covolumes and the indices are quoted.

A step largely absent from the final proof. *Step 3* (an identity $qN = M^k + R$; a quadric; automorphy through a theta function on a sixfold cover of GL_5; the Poincaré bundle). We found no trace of the identity or of a sixfold cover in the Lean development (quadrics occur in some file names, in what appear to be unrelated lemmas on orbit measures); step 16 of the announcement refers to them as "a different half-plane argument". A Poincaré bundle does occur in the archive, on an auxiliary abelian scheme, in a construction of the line bundle whose section gives the theta kernel in algebraic coordinates (§I.4.1; Part II, §6.5). The declaration we tested from that file is not among those the final theorem depends on, so we cannot say that the final proof uses it.

Steps of the proof that the announcement does not list. The choice of the auxiliary primes q and d, with the three character sums and the zero-density and mean-square estimates that control them (§I.6(e)); the two scales, and the size of b (§I.4.3, §I.6(b)); and the form of Landau's lemma on intervals (§I.1; Part II, §1). These are in the Lean development; file names are given in Part II. The first is proved in Part III, §III.1, the estimate at the large scale in §III.2, and the lemma in Part II, §1.

I.10 The same proof for Dirichlet L-functions

The Lean proof treats all Dirichlet L-functions at once, and ζ is the case of the trivial character. This section says what changes for another character. It is described from the statements of the Lean development and is not worked out in this paper.

Let ψ be a primitive Dirichlet character of conductor M. Its powers $1, \psi, \psi^2, \dots$ form a finite cyclic group; let o be the number of distinct powers, that is, the order of ψ. For $1 \le j < o$ let ψ_j be the primitive character that induces ψ^j (the power ψ^j itself need not be primitive), and let G_{ψ_j} be the function of §I.6 attached to it:

$$G_{\psi_j}(y) = \sum_{n \le y} \Lambda(n)\psi_j(n)\Big(\frac{y}{n} - 1\Big) + y\,\frac{L'}{L}(1, \psi_j). \tag{I.36}$$

The statistic G is replaced by the combination

$$G^{\psi}(y) = G(y) + \frac{1}{3o+1} \sum_{j=1}^{o-1} \mathrm{Re}\; G_{\psi_j}(y). \tag{I.37}$$

For n prime to M, the term $\Lambda(n)(y/n - 1)$ occurs in it with the coefficient $1 + \frac{1}{3o+1}\sum_{j=1}^{o-1} \mathrm{Re}\; \psi^j(n)$, which equals $\frac{4o}{3o+1}$ if $\psi(n) = 1$ and $\frac{3o}{3o+1}$ otherwise. For n not prime to M the coefficient is $1 + \frac{1}{3o+1}\sum_{j=1}^{o-1} \mathrm{Re}\; \psi_j(n)$, which can differ from this formula, since $\psi_j(n)$ may be nonzero where

$\psi^j(n) = 0$; it is still at least $\frac{3o}{3o+1}$. So every coefficient is positive. Multiplied by $3o+1$, the combination is the statistic attached to the product of $\zeta(s)^{3o+1}$ and the functions $L(s,\psi_j)$, $1 \le j < o$; at a zero of this product the multiplicities add, so every zero of ζ or of an $L(s,\psi_j)$ gives a pole of the transform. The lemma of §I.1 then applies with one addition. The proof for ζ uses that there is no pole on the real segment $0 < s < 1$; for L-functions a real zero β cannot be excluded in advance. At such a zero the residue of the transform is a positive multiple of $1/(\beta(1-\beta))$. If it is the only pole nearest to the real point σ, the derivatives at σ have size about $(\sigma-\beta)^{-j-1}$, and the fixed sign of the residue makes them positive; this contradicts the bound of Part II, §1 when $\beta > \theta$; the Lean proof of the lemma has this case as a separate branch. The conclusion is: if for every $\varepsilon > 0$ the bound $G^{\psi}(y) \le y^{1/2+\varepsilon}$ holds on intervals $[Y, Y^{3/2}]$ with Y arbitrarily large, then ζ and every $L(s,\psi_j)$ have no zero with real part greater than $\frac{1}{2}$. (For $o = 1$ this is G and ζ.)

The rest changes as follows. The field E is enlarged so that ψ becomes one of its characters. The columns of the table are then indexed by a larger group: it consists of o cosets of the subgroup on which the label coming from ψ is trivial, and the common size of these cosets depends on how much ψ shares with the residue classes modulo $4n$ (for the character $(\frac{2}{\cdot})$ modulo 8, for instance, the number of columns does not grow). The set T is chosen with slightly different densities on different groups of columns; this is where the weight $\frac{1}{3o+1}$ comes from. Estimate B is needed for the twists $\xi\chi_q^i\chi_d^j$ of all characters ξ modulo M, and Estimate C for the characters of the larger field; the theorems of §III.1 and §III.2 are stated in that generality. Everything else is the same: the main inequality (M), with its theta lifts, divisibility, size, trace formula and Galois cohomology, is the same construction for the larger field, and the passage from the table to one prime sum is the same computation with more characters.

I.11 Notes on the literature

This paper explains an argument due to Alpöge and Claude [Alp26]. Apart from that argument it uses, or quotes, results of many mathematicians. This section says which, and where. Two kinds of use should be distinguished.

(a) Published results used in the proofs given in this paper.

- *Chebyshev's method* (§I.0, §I.2): Chebyshev [Che52]; the form with the middle binomial coefficient is due to Erdős [Erd32]. The exponent of a prime in $n!$ is due to Legendre [Leg08]. Textbook accounts: [Apo76, Ch. 4], [Dav00, Ch. 7].
- *Landau's theorem* on Mellin transforms (or Dirichlet series) of functions of constant sign (§I.1): Landau [Lan05]; see [MV07, §15.1]. The version on intervals (Part II, §1) is proved in full in this paper, following the Lean development.
- *The explicit formula* and the value $\sum_\rho \frac{1}{\rho(1-\rho)} = 2 + \gamma - \log 4\pi$ (§I.1): von Mangoldt; see [Dav00, Ch. 12 and 17]. The use of positivity in an explicit formula goes back to Weil [Wei52].
- *The prime number theorem in arithmetic progressions*, for a fixed modulus, and *Mertens' theorem* in progressions (§I.4.2, §I.6; Part II, §5): de la Vallée Poussin; see [Dav00, Ch. 20 and 22], [MV07, Ch. 11].
- *Theorems on Dirichlet L-functions used in Part III.* For Estimate B (§III.1): the zero-free region with at most one exceptional zero, Landau and Page [Pag35], [Dav00, Ch. 14]; primes in progressions to large moduli, Thorner and Zaman [TZ21], after Linnik and Gallagher [Gal70], [Bom87]; a zero-density estimate, Montgomery [Mon71]; Landau's bound for L'/L in

a zero-free region, [Tit86], [MV07]. For Estimate C (§III.2): the zero-free region, the count of zeros in a window and the truncated explicit formula, [Dav00, Ch. 14, 16, 19], [MV07, Ch. 11, 12]. These are cited, not proved. The mean value over the auxiliary primes is obtained from primes in progressions and not from a large sieve; compare [Bom65], [DH66], [HB95].

- *Sets without three-term progressions* (§I.9, steps 11 and 19; Part III, §III.3): Behrend [Beh46]. The construction is carried out, with proof, in §III.3.
- *Stepanov's method* (§I.2, §I.7; Part II, §2), as the model for counting with multiplicity: Stepanov [Ste69]; see Bombieri [Bom74].
- *The lemniscate functions* and their arithmetic (§I.3): Hurwitz [Hur99]. *Formal groups with complex multiplication*, which explain the exponent $a\,p/(p^2-1)$ at inert primes: Lubin and Tate [LT65], Honda [Hon70]. That the reduction is of this type exactly at the inert primes: Deuring [Deu41]. Algebraicity of the lattice sums of §I.4.1: compare Damerell [Dam70].
- *The Gaussian inner product on polynomials*, $\|z^a\|^2 = a!$ (§I.4.2; Part II, §6.5): the Fock, or Bargmann, space [Bar61].
- Hadamard's inequality, the Chinese remainder theorem and the Cauchy–Schwarz inequality are used in Part II, §6 without further reference.

(b) Results behind the quoted statements. The statements marked [quoted] are proved in the Lean development, which contains its own proofs of the theorems it needs. The underlying mathematics has the following sources. We cite them for the classical statements; we have not compared the formalised statements with the published ones in detail.

- *Theta series as kernels* and the representation that governs them (§I.4.1, §I.7): Weil [Wei64]; the correspondence between forms on a pair of groups: Howe [How79]. *Seesaw identities*: Kudla [Kud84]. *The inner product formula* for theta lifts: Rallis [Ral87]. *Bessel periods* and the restriction problems in which they occur: Gan, Gross and Prasad [GGP12].
- *The period formula* (§I.7; Part II, §6.6). In the form used, it is Kronecker's limit formula (see [Wei76]); the evaluation through Gamma values is due to Lerch, and to Chowla and Selberg [CS67].
- *The mass formula* for definite lattices (§I.7; Part II, §6.6): Siegel [Sie35]; for Hermitian lattices, Shimura, see Gan, Hanke and Yu [GHY01]. *The dimension formula*: Weyl [Wey25].
- *Slopes of Euclidean lattices* and the maximal slope (§I.7; Part II, §6.8, §6.10): Stuhler [Stu76], Grayson [Gra84]. *The slope inequality* for maps of lattices: Bost [Bos96].
- *The relative trace formula* (§I.7, "Rank"): Jacquet [Jac86].
- *The bound for Galois cohomology* (§I.7, "Denominators"; Part II, §6.8). The announcement describes it as a weakening of Vandiver's conjecture. According to our reading of the Lean development its proof uses the vanishing of the Iwasawa μ-invariant (Ferrero and Washington [FW79]): a statement under that name is proved in the development and the final theorem depends on it. The development also contains statements named after Kato's Euler systems [Kat04]; some of these are among the declarations the final theorem depends on and one that we tested is not. The proof also uses modularity lifting theorems. We did not find the main conjecture of Iwasawa theory (Mazur and Wiles [MW84]) stated under that name. An upper bound for a cohomology length in terms of the valuation of a Bernoulli number with explicit integral factors is present; it corresponds to a consequence of the main conjecture, and we have not compared the two. A bound for the whole group H^2 at every prime is also

present. Finiteness of these groups is due to Soulé ([Sou79], from memory). Soulé [Sou03, Thms. 2 and 3] gives a quantitative bound away from small primes: for the primes $p \geq 4n' - 1$, $p \geq [L : \mathbb{Q}] + 2$, where L is the field and n' is the twist (Soulé writes n; it is not the n of this paper). This was compared with the article by a reviewer, not by us. We have not identified a published estimate in the full small-prime range. Part III, §III.9 describes the statement and the structure of its proof in the development, and says which published theorems the cases correspond to; we have not checked how closely the formal proofs follow the published ones.

What is new. The combination described in this paper, namely iterated theta lifts as a source of multiplicity, the comparison of divisibility with size through slopes of lattices carried to the term linear in the order, and the resulting one-sided bound for $G(x)$, is the contribution of [Alp26]. The exposition, the worked examples and the computations in this paper are ours.

I.12 Summary: what is proved where

Statement	Status
One-sided bound for G on intervals $\Rightarrow$ RH (§I.1)	proved (Part II, §1)
Model divisibility $a\,p/(p^2-1)$; lemniscate numbers (§I.3)	proved for the model series (Part II, §2); the numbers computed exactly
(M), B and C $\Rightarrow$ Estimate A $\Rightarrow$ the bound for G (§I.6)	proved (Part II, §3–5); the identities used are checked numerically on an example
Estimate B: the auxiliary primes (§I.6)	proved (Part III, §III.1), from four cited theorems
Estimate C: the characters at the large scale (§I.6)	proved (Part III, §III.2), from three cited theorems and Theorem III.1.1(e)
The sets of columns T_2 and T (§I.5)	constructed (Part III, §III.3); their use in (M) is quoted
(M), size (§I.7, item 2)	the analytic estimates for one map and for a chain are proved in an abstract setting (Part III, §III.4); that the maps of the construction are instances of it is quoted; four passages of §III.4 are conditional; the arithmetic inputs are quoted
(M), divisibility at inert primes (§I.7, item 1)	written up in §III.5, conditionally on the integrality of the theta kernels in the coordinates of the height-two logarithm (hypothesis (B)), which is taken from the Lean development; elementary lemmas proved (Part II, §6)
(M), from derivatives to lattices (§I.7, item 1)	proved (Part III, §III.6): lemmas on slopes of lattices, the splitting of the prime sum, bounds for several of its terms. Conditional on two statements of the Lean development: the theorem on the glued lattice. Quoted: the bounds for the eleven loss sums and their total; the statement on valuations
(M), arithmetic inputs of the size bound (§I.7, item 2)	proved (Part III, §III.7): the size inequality follows from eleven named inputs; a mean of periods equals $\log \Omega_E$ under stated hypotheses. The eleven inputs are quoted
(M), rank (§I.7, item 3)	proved (Part III, §III.8): the dimension bound follows from three statements of the Lean development and stated size conditions. The three statements are quoted; elementary lemmas proved (Part II, §6)
(M), denominators (§I.7, item 3)	proved (Part III, §III.9): the step from the bound for Galois cohomology to the bound for denominators. The bound itself is quoted; elementary lemmas proved (Part II, §6)
(M), the construction of the maps and weights, and the assembly	quoted; not begun in Part III
Error budget and exponent $\frac{1}{2}$ (§I.8)	proved from the stated form of (M)

The formal proof in Lean establishes the Riemann hypothesis for all Dirichlet L-functions at once, and with it the statements quoted here, in its own formulation. The case of ζ described here is the case of the trivial character; it uses the same construction.

Sources (the Lean files) are at the end of Part II; the list of scripts follows Part III.

Part II. Details and proofs

Part II can be read after §I.1 and §I.5: it uses $G(x)$ from §I.1 and otherwise repeats its definitions. It is numbered separately. Its sections are 1–6. Its displays are numbered (II.n); those referred to often carry section tags instead: (1.1), (4.1), (5.1), ..., (6.7), and (A), (M). The main inequality is called (M) in all three parts, and Estimate A of Part I is formula (A) of Part II. Section 3 of Part II states the reduction to one prime sum again, with all definitions, so that sections 4–6 can be read on their own. References in square brackets with numbers are to the list of source files at the end of Part II. Some letters are used with more than one meaning in Part II (for instance N, L, T, h and a); each is defined where it is used. Formula (5.6) of Part II is Estimate C of Part I. Where sections 3–5 attribute a statement to "the source", a proof is now in Part III for the following: the choice of the primes q, d and the bound for the three twists (section 3; Part III, §III.1); formula (5.6) (Part III, §III.2); the existence of the sets T, T_2 with the stated densities (section 4; Part III, §III.3). The divisibility at inert primes of sections 6.3–6.4 is written up in Part III, §III.5, conditionally on the integrality of the theta kernels in the coordinates of the height-two logarithm (hypothesis (B)), which is taken from the Lean development. Part III, §III.6–III.9 prove further steps towards (M), each from named statements of the Lean development: the passage to lattices in part (§III.6), the size inequality from eleven quoted inputs (§III.7), the dimension bound of section 6.9 from three quoted statements (§III.8), and the step from the bound for Galois cohomology to the bound for denominators (§III.9); the table at the start of Part III gives the status of each. The main inequality (M) is still quoted.

1. What a zero forces

For $\operatorname{Re} s > 1$, termwise integration gives

$$F(s) = \int_1^\infty G(x) x^{-s-1}\, dx = -\frac{\zeta'(s)}{s(s-1)\zeta(s)} - \frac{1}{(s-1)^2} + \frac{1+\gamma}{s-1}. \tag{II.1}$$

Indeed, the transform of $(x/n-1)_+$ is $n^{-s}/[s(s-1)]$, where $v_+ = \max(v,0)$. The Laurent expansion of ζ at 1 cancels both terms of the principal part there. A zero ρ of multiplicity ν in the critical strip gives F a simple pole of residue $-\nu/[\rho(\rho-1)]$. The poles at $s=0$ and at the trivial zeros $-2,-4,\ldots$ lie outside the half-plane $\operatorname{Re} s > \theta$ used below. The poles in the critical strip are nonreal, since the alternating zeta series gives $\zeta(\sigma) < 0$ for $0 < \sigma < 1$.

A bound valid for every large x would allow the usual oscillation argument of Landau. We know the bound only on certain intervals. To use one such interval, we differentiate the transform many times. Its new weight is concentrated on that interval, and a nearby pole forces large derivatives.

Lemma (the one-sided criterion on intervals; §I.1). *Let $0 < \theta < 1$. If $G(x) \le x^\theta$ throughout $[Y, Y^{3/2}]$ for arbitrarily large Y, then ζ has no zero in $\operatorname{Re} s > \theta$.*

Write $u = \log x$, $g(u) = G(e^u)$ and, for each integer $j \ge 0$, let $A_j(s) = (-1)^j F^{(j)}(s)/j!$ be the signed jth Taylor coefficient. Let $\eta = \max(0.8e^{0.2}, 1.2e^{-0.2}) = 0.9825\ldots$ and $c = 0.008$, which is less than $\log((1+\eta)/(2\eta)) = 0.00887\ldots$; these are absolute constants. Fix $\sigma > 2$ so large that $(\sigma-\theta)/(\sigma-2) \le e^c$. With the constants above this holds for $\sigma \ge 252$ and every θ in $(0,1)$. Differentiating the Mellin integral gives

$$A_j(\sigma+it) = \int_0^\infty g(u) e^{-itu} K_j(u)\, du, \qquad K_j(u) = u^j e^{-\sigma u}/j!. \tag{II.2}$$

On each interval satisfying the lemma's bound, put $U = \log Y$ and choose j nearest to $\frac{5}{4}(\sigma-2)U$. Then

$$|A_j(\sigma+it)| + A_j(\sigma) \le 2\int_0^\infty \max(g(u),0)K_j(u)\,du \ll (\sigma-\theta)^{-j}, \tag{1.1}$$

uniformly for real t. The first inequality follows from the triangle inequality and $|g|+g = 2\max(g,0)$. On $[U, 3U/2]$ integrate $e^{\theta u}K_j(u)$.

Outside the window, use $|g(u)| \ll e^{2u}$, since partial summation of Chebyshev's bound $\sum_{n\le x}\Lambda(n) \ll x$ gives $\sum_{n\le x}\Lambda(n)/n \ll \log x$ and hence $|G(x)| \ll x\log x$ for $x \ge 2$. The density proportional to $u^j e^{-(\sigma-2)u}$ is concentrated about $j/(\sigma-2) \sim 5U/4$.

Claim. *For U large and j nearest to $\frac{5}{4}(\sigma-2)U$, the part of $\int_0^\infty u^j e^{-(\sigma-2)u}\,du$ outside $[U, 3U/2]$ is at most $C_0 e^{-cj}$ times the whole, with C_0 independent of U.* Indeed, on putting $v = (\sigma-2)u/j$ the density is proportional to $(ve^{1-v})^j$, and the window becomes an interval in v whose ends tend to $4/5$ and $6/5$ as $U \to \infty$. The base ve^{1-v} has its unique maximum 1 at $v = 1$. Off the window it is bounded by a number $\eta' < 1$ that tends to η as $U \to \infty$, so the tail integral is at most $\eta'^{\,j-1}\int_0^\infty ve^{1-v}\,dv$. A small fixed interval about 1 has base at least $(1+\eta')/2$. So the fraction outside is $O\big((2\eta'/(1+\eta'))^j\big)$, and this is $O(e^{-cj})$ for U large, because $c < \log((1+\eta)/(2\eta))$. The tail is therefore $O(e^{-cj}(\sigma-2)^{-j-1}) = O((\sigma-\theta)^{-j})$, by the choice of σ. This proves (1.1) along a sequence $j \to \infty$.

Suppose now there is a pole to the right of θ. Let R be the smaller of $\sigma-\theta$ and the distance from σ to the nearest pole in $\operatorname{Re} s > 0$. Some pole has horizontal distance from the line $\operatorname{Re} s = \sigma$ strictly less than R. If $R = \sigma-\theta$, this follows from the supposition. Otherwise take a nearest pole to σ and use its nonzero imaginary part. Take $t_0 = \operatorname{Im}\rho$ for that pole ρ; then the pole nearest to $\sigma + it_0$ is at distance less than R. Only finitely many poles lie within $R+1$ of it, and each pair is equidistant from at most one point of the line $\operatorname{Re} s = \sigma$ (both lie to its left). So t near t_0 can be chosen so that $z = \sigma + it$ has a unique nearest pole, at distance $d < R$.

Choose $d < r < R$ so that the disk of radius r about z contains just that pole. Subtracting its principal part and applying Cauchy's estimate gives $|A_j(z)| \ge c_1 d^{-j} - O(r^{-j})$ with $c_1 > 0$ (the principal part $c/(s-\rho)$ contributes exactly $c\,(z-\rho)^{-j-1}$ to $A_j(z)$). At the real centre σ the same estimate gives $|A_j(\sigma)| \ll r^{-j}$. This contradicts (1.1) for large j, since $d < r < \sigma-\theta$. This proof uses only the poles in two bounded disks. The Lean proof [1] reaches the same statement by a slightly different route.

2. How divisibility counts primes

Start with a polynomial $P(X)$ of degree D. If P has distinct roots $\alpha_1, \dots, \alpha_s$, each of multiplicity at least $m \geq 1$, then $\prod_{j=1}^{s}(X-\alpha_j)^m$ divides P. Comparing degrees gives $ms \leq D$.

Stepanov uses the same count for a **nonzero rational function on a smooth projective curve over a finite field**. Count points over the algebraic closure of that field (the curve is assumed geometrically irreducible), and include the points at infinity. For a nonconstant function the zeros and poles are the inverse images of 0 and of infinity under its map to the projective line. Each inverse image, counted with multiplicity, has the degree of that map. Thus the total multiplicity of the zeros equals the total order D of the poles. Hence at most $\lfloor D/m \rfloor$ points can be zeros of order at least m. On the projective line the function is a quotient of polynomials. A polynomial of positive degree D has just one pole, of order D at infinity.

Over the integers, write $v_p(c)$ for the exponent of a prime p in a nonzero rational number c (numerator exponent minus denominator exponent), and put $v_p(0) = +\infty$. The corresponding count is

$$\sum_p e_p \log p \leq \log|\Delta|, \qquad \Delta \in \mathbb{Z} \setminus \{0\}, \quad 0 \leq e_p \leq v_p(\Delta). \tag{II.3}$$

To bound the left side one needs an independent upper bound for $|\Delta|$. For a square integer matrix of nonzero determinant Δ, for example, divisibility gives a lower bound, while the lengths of its columns give an upper bound by Hadamard's inequality. Section 6 explains the corresponding comparison for lattices. (The file [2] states such an inequality for an earlier form of the argument. That form is stated in the archive but, according to our dependency check, is not used by the final theorem.)

Why odd powers of a prime occur. Here is the local model in rational arithmetic. For a prime p form the series

$$l(X) = X + \sum_{j \geq 1} X^{p^{2j}}/p^j. \tag{II.4}$$

It has a unique inverse under substitution, $U(z) = z + \cdots$, given recursively by $U = z - \sum_{j\geq 1} U^{p^{2j}}/p^j$. Put $L = p^2 - 1$. We claim that the coefficient of z^a in U has valuation at least $-(a-1)/L$ for each integer $a \geq 1$. This holds for $a = 1$. In a product of $r = p^{2j}$ copies of U, the degrees $i_1, \dots, i_r$ contributing to z^a are positive and sum to a. Each is less than a. By induction their coefficients have product of valuation at least $-(a-r)/L$. Division by p^j costs $j \leq (r-1)/L$, proving the claim.

For a series f with integer coefficients, the same product calculation gives the bound $-(a-1)/L$ for the coefficient of z^a in $f(U(z))$. Its ath derivative at zero is $a!$ times this coefficient. Counting multiples of $p, p^2, \dots$ in $a!$ gives $v_p(a!) = \sum_{h\geq 1}\lfloor a/p^h \rfloor = a/(p-1) + O(1 + \log_p a)$, where $\lfloor y \rfloor$ is the greatest integer at most y and $\log_p a = \log a / \log p$. The derivative therefore has valuation at least

$$a\left(\frac{1}{p-1} - \frac{1}{p^2-1}\right) - O(1+\log_p a) = a\left(\frac{1}{p} + \frac{1}{p^3} + \frac{1}{p^5} + \cdots\right) - O(1+\log_p a). \tag{II.5}$$

The series models differentiation on an elliptic curve with complex multiplication (extra endomorphisms) at primes remaining prime in its imaginary quadratic field. The source proves a weighted version for its arithmetic maps. At odd powers p^h it yields the term $(a/p^h - w)_+$ for a positive weight w (quoted; see section 6.4). When $w = b$ and $a = bx$, this is precisely $b(x/p^h - 1)_+$. [3]

3. The reduction to one prime sum

For an odd prime D let $\chi_D(r)$ equal 0 if $D \mid r$, 1 if r is a nonzero square modulo D, and -1 otherwise. For a nonprincipal Dirichlet character χ put

$$c_\chi = \frac{L'}{L}(1,\chi), \qquad G_\chi(x) = \sum_{r\le x} \Lambda(r)\chi(r)(x/r-1) + c_\chi x. \tag{II.6}$$

A Dirichlet character is a homomorphism from the invertible classes modulo an integer to the nonzero complex numbers, extended by zero to other classes. It is nonprincipal if it takes a value different from 1 on the invertible classes. The function $L(s,\chi)$ is the continuation of $\sum_{r\ge1}\chi(r)r^{-s}$ from $\operatorname{Re} s > 1$, and the prime denotes differentiation in s.

For distinct odd primes q, d write $\eta = \chi_q\chi_d$ for the pointwise product. Consider the single prime sum

$$S_{q,d}(x) = \sum_{r\le x} \Lambda(r)(1+\chi_q(r))(1-\chi_d(r))(x/r-1). \tag{II.7}$$

Away from q, d, the coefficient of $\Lambda(r)$ before smoothing is 4 if r is a square modulo q and a nonsquare modulo d, and zero otherwise. The exact identity obtained by expanding the two factors is

$$\begin{aligned} R_{q,d}(x) &= S_{q,d}(x) - x\log x + (1+\gamma + c_{\chi_q} - c_{\chi_d} - c_\eta)x \\ &= G(x) + G_{\chi_q}(x) - G_{\chi_d}(x) - G_\eta(x). \end{aligned} \tag{II.8}$$

Thus a one-sided bound for $R_{q,d}$ and absolute bounds for three quadratic character sums suffice.

The reduction. Fix $0 < \varepsilon < 1/52$. By Theorems III.1.1 and III.2.1 of Part III, for every large Y there are primes $q \equiv 3 \bmod 8$, $d \equiv 7 \bmod 16$ with $q \mid d+1$ for which statement 2 below (Estimate B) and formula (5.6) (Estimate C) hold for $Y \le x \le Y^{3/2}$. Statement 1 (Estimate A) is then deduced in sections 4 and 5 from the main inequality (M), which is quoted, together with Estimates B and C.

1. The residue sum. There are positive integers m, k, n depending on x such that

$$R_{q,d}(x) \le C_\varepsilon x^{C\varepsilon}\left(\frac{x}{m} + \frac{xm}{k} + m + \frac{k}{n^2}\right). \tag{A}$$

Here C is absolute and C_ε depends only on ε. The orders are $m = x^{1/2+O(\varepsilon)}$, $k = x^{9/2+O(\varepsilon)}$, $n = x^{2+O(\varepsilon)}$. Sections 4 and 5 derive this from (M).

2. The three twists. For $\chi = \chi_q, \chi_d, \eta$ the selection gives $|G_\chi(x)| \le x^{1/2+\varepsilon}$. To see what is estimated, put $\Psi_\chi(t) = \sum_{r\le t}\Lambda(r)\chi(r)$. Partial summation and $\sum_{r\ge1}\Lambda(r)\chi(r)/r = -c_\chi$ give

$$G_\chi(x) = -x\int_x^\infty \Psi_\chi(t)\,\frac{dt}{t^2}. \tag{II.9}$$

The convergence here follows from the prime number theorem for a fixed nonprincipal character. The proof bounds the proportion of bad choices by mean-square estimates, first in q, then in d for each good q. This gives a common pair on the whole interval. It is given in Part III, §III.1, with the exponent $\frac12 + \frac{\varepsilon}{2}$; the Lean proof is in [9, 10, 29].

Conclusion. The terms x/m and m balance at $\sqrt{x}$. With the stated orders, $k/n^2 = x^{1/2+O(\varepsilon)}$ and xm/k is smaller. By the identity above, $G(x) \le C'_\varepsilon x^{1/2+C'\varepsilon}$ on these intervals, with an absolute C'. Given $1/2 < \theta < 1$, choose ε so that $1/2 + C'\varepsilon < \theta$. On all sufficiently large selected intervals $G(x) \le x^\theta$. Section 1 excludes every zero to the right of θ. Letting θ decrease to $1/2$ and using the symmetry $\rho \mapsto 1-\rho$ from the functional equation proves RH from (M) and Estimates B and C. [1, 6, 12]

4. The main inequality (M)

4.1 Which prime powers are counted?

Fix an interval and primes q, d as in section 3. Fix x in the interval. For an odd prime power $r = p^h$ with $p \neq q$, attach two labels. The first is the row $v(r) = (1 + \chi_q(r))/2$, equal to 1 if r is a nonzero square modulo q, and 0 otherwise. For the second, choose $n = 2^e$ at the scale $x^{2+8\varepsilon} \leq n \leq 2x^{2+8\varepsilon}$ and put $N = 4n$. Write $[r] = \{r, -r\}$ mod N. The n pairs of odd residues form a group H under multiplication. Its elements will label columns. Choosing a fixed $t \in H$ puts r in column $t[r]$.

The coefficient at $r = p^h$, before inserting a weight, is

$$c(r) = (1 - \chi_d(r)) \log p. \tag{II.10}$$

If $p \neq d$ it is $2 \log p$ for an odd power of a nonsquare modulo d and zero otherwise. At $p = d$ it is $\log d$ for every power. The two omitted primes $2, q$ have $\chi_d(2) = \chi_d(q) = 1$. Indeed $d \equiv 7$ mod 8 gives the first identity. For the second, $q, d \equiv 3$ mod 4 and $d \equiv -1$ mod q, so quadratic reciprocity gives $(q/d) = -(d/q) = -(-1/q) = 1$. So $c(r)$ vanishes at every power of 2 and of q; omitting these primes changes nothing.

A table small enough to write out. Take $q = 3$, $d = 23$, $N = 16$ and $t = [1]$. The four columns are $[1], [3], [5], [7]$. For example $[3]^2 = [9] = [7]$ and $[3]^3 = [5]$. The nonzero squares modulo 23 are $1, 2, 3, 4, 6, 8, 9, 12, 13, 16, 18$. In each cell below are the primes at most 61 with nonzero coefficient. A dash denotes an empty cell in this range.

$v \backslash s$	$[1]$	$[3]$	$[5]$	$[7]$
0	17	–	5, 11, 53	23
1	–	19, 61	37, 43	7

(II.11)

Among primes the density in each cell is 1/16, as shown in section 5. To place 11, observe that $11 \equiv 2$ mod 3 and $[11] = [5]$. To place a higher power one repeats both tests on that power. The prime 7 goes in $(1, [7])$ with coefficient $2 \log 7$. Its square disappears because $\chi_{23}(7^2) = 1$. Its cube has the same row and column as 7. At the exceptional prime 23, its first power lands in $(0, [7])$, its square in $(1, [1])$, and each has coefficient $\log 23$. The example only illustrates the identities of section 5, which hold for any positive weights; its parameters do not satisfy the conditions of (M).

4.2 How a weight determines a cutoff

Let $b = \exp(x^{\varepsilon})$. The construction supplies nonempty subsets T, T_2 of H, an integer $a_t > 0$ for each $t \in T$ with $|a_t/b - x| \le 1$, and m maps of initial rank k (the chain of theta lifts of section 6.2; m and k are the parameters in (A)). Write $|S|$ for the cardinality of a finite set S and $\delta = |T_2|/n$. For each translate $u \in H$ and map $1 \le j \le m$ there are $2n$ positive weights $w_{u,j}(v,s)$. In the construction they are integers. The distinguished cells are $(1,s)$ with $s \in uT_2$. Their weights satisfy $b \le w_{u,j}(1,s) \le b + O(b/(nx^2))$. At every other cell the ratio $a_t/w_{u,j}(v,s)$ is at least $\exp(x^{\varepsilon/3})/3$. Thus the same table tests primes at two markedly different scales.

Fix u, t, j, write $a = a_t$ and $w = w_{u,j}$. Each power in cell $i = (v,s)$ contributes

$$c(r)(a/r - w_i)_+ = c(r)w_i(a/(rw_i) - 1)_+, \qquad z_+ = \max(z,0). \tag{II.12}$$

The contribution ends at $r = a/w_i$. Adding all cells gives the finite sum

$$P_{u,t,j} = \sum_{\substack{p \text{ odd prime} \\ p \ne q}} \sum_{h \ge 1} (1 - \chi_d(p^h)) \log p \, (a_t/p^h - w_{u,j}(v(p^h), t[p^h]))_+. \tag{4.1}$$

In the running example take $T = T_2 = \{[1]\}$, one table, $x = 1000$, $b = 10$, $a = 10000$ and $u = [1]$. Here $\delta = 1/4$. The weights and cutoffs, in the same column order as above, are

w_i	[1]	[3]	[5]	[7]
0	0.05	0.05	0.05	0.05
1	10	0.05	0.05	0.05

a/w_i	[1]	[3]	[5]	[7]
0	200000	200000	200000	200000
1	1000	200000	200000	200000

(II.13)

For example 7 contributes $2\log 7\,(10000/7 - 0.05)$, whereas $23^2 = 529$ contributes $\log 23\,(10000/529 - 10)$. Translating by $u = [7]$ moves the weight 10 to column [7] in row 1. Now the subtracted weight at 7 is 10, and at 529 it is 0.05.

For a reproducible finite calculation run through the primes $p \le 200000$ and the successive powers $p, p^2, \ldots \le 200000$. Drop $2, 3$, compute the coefficient and the cell as above, and add its positive part. This gives $P_{[1],t,1} = 95461.6156$ to four decimals. The same number will be obtained by grouping the terms in a different order in section 5.

4.3 What the arithmetic construction must prove

The quantity to subtract from (4.1) is

$$\langle \log w \rangle = \frac{1}{2n} \sum_{i} \log w_i, \qquad M_{u,t,j} = a_t \big(\log a_t - 1 - \gamma + c_{\chi_d} - \langle \log w \rangle \big). \tag{4.2}$$

Section 5 derives each term by ordinary prime summation. Section 6 explains why the arithmetic estimate on the maps yields the same expression. The match through the term linear in a_t is essential.

Every translate has seven weights equal to 0.05 and one equal to 10. Substitution in (4.2) gives

$$\begin{aligned} M = 10000 \big(\log 10000 - 1 - \gamma - 0.422120948038\ldots \\ - (\log 10 + 7 \log 0.05)/8 \big) = 95444.4636\ldots. \end{aligned} \tag{II.14}$$

For $u = [1]$ the excess is 17.1520 to four decimals. The other three excesses are 368.2081, -129.2546, and -204.6531, giving a mean of 12.8631. (M) bounds their average over u, summed over the labels t and the maps j.

The main inequality (M). On the selected intervals the source constructs the above data with $\delta \ge x^{-\varepsilon}$, $|T|/n \ge x^{-\varepsilon/8}/2$, and an upper error $\mathcal{E} \ge 0$. (Sets T_2, T with these densities are constructed in Part III, §III.3. The rest of the data is quoted.) With $B = |T| m b \delta/2$ it proves

$$\frac{1}{n} \sum_{u \in H} \sum_{t \in T} \sum_{j=1}^{m} (P_{u,t,j} - M_{u,t,j}) \le \mathcal{E}, \qquad \frac{\mathcal{E}}{B} \le x^{C\varepsilon} \left(\frac{x}{m} + \frac{xm}{k} + m + \frac{k}{n^2} \right). \tag{M}$$

Here C is absolute, and the inequality is asserted for $x \ge x_0(\varepsilon)$. The first inequality compares integral divisibility with complex size. To check the powers in the second, put $h = |T|$, $h_2 = |T_2|$ and let a_E be the aggregate scale in the source with $0 \le a_E \le 2xb$. An unnormalized error estimate stated in [6] (a parallel formulation; see the start of Part III) is

$$\mathcal{E} \le x^{C\varepsilon} \left[h a_E (1 + m^2/k) + m \left(2nmbx^{\varepsilon} + \frac{h_2 k b x^{\varepsilon}}{n(nx^{-\varepsilon})} \right) \right] \tag{II.15}$$

for an absolute C. The bound for $\mathcal{E}/B$ in (M) is thus taken from a statement of the archive on which the final theorem does not depend in that form (see the start of Part III). Division by $B = hmb\delta/2$ bounds the three bracketed contributions respectively by

$$\frac{4}{\delta} \left(\frac{x}{m} + \frac{xm}{k} \right), \quad \frac{4n}{h\delta} m x^{\varepsilon}, \quad 2 \frac{n}{h} \frac{k}{n^2} x^{2\varepsilon}. \tag{II.16}$$

Here $h_2 = n\delta$. The bounds on n/h and $1/\delta$ give (M). This computation explains why multiplicity matters. A cost incurred once yields x/m after division, while the lattice comparison incurs a cost of order m after division. Their balance occurs at $\sqrt{x}$. [3, 5, 6]

5. From the table to a single prime sum

5.1 Subtracting the main term cell by cell

Keep t fixed; the quantities S_i, Ψ_i, C_i, R_i below depend on t, through the column $t[r]$. Let $\sum_{r\in i}$ denote a sum over prime powers assigned to cell i by section 4. The smoothed sum in that cell and its unsmoothed counting function are

$$S_i(y) = \sum_{r\in i} c(r)(y/r-1)_+, \qquad \Psi_i(z) = \sum_{\substack{r\in i\\ r\le z}} c(r). \tag{II.17}$$

For $y \ge 1$, integrating each step of Ψ_i proves the first identity below. The second follows term by term from (4.1).

$$S_i(y)/y = \int_1^y \Psi_i(z)\frac{dz}{z^2}, \qquad P_{u,t,j} = \sum_i w_i S_i(a/w_i). \tag{5.1}$$

For fixed moduli, among primes not dividing $2qd$ the fraction contributing to a given cell is $1/(4n)$. One half are nonsquares modulo d, one half have the specified sign modulo q, and $1/n$ have the specified pair modulo N. The Chinese remainder theorem makes these conditions independent among the invertible residue classes. The prime number theorem in progressions gives their density among primes. The coefficient $2\log p$ doubles this density in Ψ_i. Thus $\Psi_i(z)$ has main term μz, where $\mu = 1/(2n)$.

The fixed-modulus prime number theorem gives an error $O_A(z/(\log z)^A)$ as $z\to\infty$ for each fixed $A>0$. Higher prime powers and the one exceptional prime fit within this bound. The constants may depend on the moduli.

Set $C_i = \int_1^\infty (\Psi_i(z)-\mu z)\,dz/z^2$, which converges by that estimate. Subtract the integral from y to infinity to obtain the exact formula

$$S_i(y) = \mu y\log y + C_i y + R_i(y), \qquad R_i(y) = -y\int_y^\infty (\Psi_i(z)-\mu z)\frac{dz}{z^2}. \tag{5.2}$$

In particular $R_i(y) = o(y)$ for fixed moduli, meaning $R_i(y)/y \to 0$. The uniform estimate needed when the moduli vary with x comes later.

What is $\sum_i C_i$? Write $S_1(y)$ for the ordinary smoothed von Mangoldt sum and $S_\chi(y)$ for the same sum with coefficient $\chi(r)$. Then

$$\sum_i S_i(y) = S_1(y) - S_{\chi_d}(y). \tag{II.18}$$

This identity includes powers of d, and both sides vanish termwise at $p = 2, q$. For a nonprincipal character, partial summation and the convergent series for $-L'(1,\chi)/L(1,\chi)$ give $S_\chi(y)/y \to -c_\chi$. For the ordinary sum, the constant in $\sum_{r\le y}\Lambda(r)/r - \log y$ is $-\gamma$, by the Laurent expansion of $-\zeta'/\zeta$ at 1 and partial summation. Subtracting $y^{-1}\sum_{r\le y}\Lambda(r)\to 1$ gives the linear coefficient $-1-\gamma$. Comparison with (5.2) yields

$$\sum_i C_i = -1-\gamma+c_{\chi_d}, \qquad P_{u,t,j} - M_{u,t,j} = \sum_i w_i R_i(a/w_i). \tag{5.3}$$

For the last equality expand $\mu a\sum_i \log(a/w_i) = a\log a - a\langle\log w\rangle$ and insert the sum of the constants.

5.2 Arithmetic check of the centering

For $q = 3, d = 23, n = 4, t = [1]$, every cell has $\mu = 1/8$. The constants and sums below allow (5.1)–(5.3) to be checked on the individual cells. Here $y_0 = 1000$ and $y_1 = 200000$.

v	s	C_i	$S_i(y_0)$	$S_i(y_1)$	$R_i(y_0)$	$R_i(y_1)$
0	[1]	−.372219	523.602	230456.699	32.352	−251.387
0	[3]	−.631093	224.352	178879.838	−8.025	−53.464
0	[5]	.565967	1441.249	418604.451	11.813	259.312
0	[7]	−.406444	417.613	224057.714	−39.412	194.742
1	[1]	−.575353	291.947	190566.123	3.831	484.900
1	[3]	−.202447	694.308	264017.194	33.286	−645.209
1	[5]	−.272849	578.281	250761.014	−12.339	179.021
1	[7]	−.104899	737.266	284065.998	−21.305	−106.090

(II.19)

Summing the eight constants before rounding gives $-1.999336613\ldots = -1 - \gamma + c_{\chi_{23}}$. For $u = [1]$ multiply the $(1, [1])$ entry at y_0 by 10, and each of the other seven entries at y_1 by 0.05. Doing this in the S columns, with the entries before rounding, gives 95461.6156. Doing it in the R columns gives 17.1520. The same operations at the four translates give

u	$P_{u,t,1}$	$M_{u,t,1}$	$P_{u,t,1} - M_{u,t,1}$
[1]	95461.6156	95444.4636	17.1520
[3]	95812.6718	95444.4636	368.2081
[5]	95315.2090	95444.4636	−129.2546
[7]	95239.8106	95444.4636	−204.6531

(II.20)

How to compute C_i without fitting data. For this table let ξ be the character of the pairs modulo 16 with $\xi([3]) = \sqrt{-1}$. For $i = (v, s)$ put $\beta = 2v - 1$. As a function of a prime power, the coefficient of cell i divided by Λ has the following finite character expansion:

$$\frac{1}{8} \sum_{j,l=0}^{1} \sum_{k=0}^{3} \beta^j (-1)^l \overline{\xi(s)^k} \, (\chi_3^j \chi_{23}^l \xi^k)_{\mathrm{prim}}. \tag{II.21}$$

The subscript uses the character of smallest modulus inducing the product on units, so a factor with exponent zero is the constant one even at its own prime. For units, the two sign sums are each 2 when the signs agree and zero otherwise. The four powers of ξ sum to 4 precisely in column s. This gives coefficient 2. At 23, terms with $l = 1$ vanish, giving coefficient 1 in the prescribed cell. At $2, 3$ the χ_{23} factor sums to zero. Substituting the linear coefficient $-c_\chi$ for each nontrivial character and $-1 - \gamma$ for the trivial one gives the displayed C_i. For general t, n use the column $t^{-1}s$, the n characters of H, and coefficient $1/(2n)$. [3]

5.3 What translation preserves

Fix t, j. First replace each distinguished weight by b. For a given column s, the equation $s = uz$, $z \in T_2$, gives exactly $|T_2|$ choices $u = sz^{-1}$. Thus for each admissible prime power r

$$\frac{1}{n}\sum_{u\in H} \mathbf{1}_{\{v(r)=1,\ t[r]\in uT_2\}} = \delta\, \mathbf{1}_{\{\chi_q(r)=1\}} = \frac{\delta}{2}(1+\chi_q(r)). \tag{II.22}$$

Here $\mathbf{1}$ equals one when its condition holds and zero otherwise. Multiplication by $1-\chi_d(r)$ gives the right contribution at the three special primes as well, with zero understood at the two omitted primes. Hence, with $a = a_t$,

$$\frac{1}{n}\sum_{u}\sum_{i \text{ distinguished at } u} bS_i(a/b) = \frac{\delta b}{2} S_{q,d}(a/b). \tag{5.4}$$

This is an identity for every common argument $y = a/b$. The map sending a sum with expansion $\alpha y \log y + \beta y + o(y)$ to its remainder after subtracting those two terms is linear. The coefficients are unique by taking limits. Applying it to (5.4) gives

$$\frac{1}{n}\sum_{u}\sum_{i \text{ distinguished at } u} bR_i(a/b) = \frac{\delta b}{2} R_{q,d}(a/b). \tag{5.5}$$

In particular, with $\eta = \chi_q\chi_d$, the right side involves $G + G_{\chi_q} - G_{\chi_d} - G_\eta$. The plus sign on G comes from expanding $(1+\chi_q)(1-\chi_d)$.

In the example each of the four row-1 cells is distinguished once. From the $S_i(1000)$ column we get

$$\tfrac{10}{4}(291.947014 + 694.307989 + 578.280814 + 737.265890) \simeq 5754.504268. \tag{II.23}$$

Direct summation with the two quadratic signs gives the same value $(10/8)S_{3,23}(1000)$. For (5.5), the four remainders give $8.679631\ldots$ to six decimals, or $(10/8)R_{3,23}(1000)$. There is a separate check using the three L-values:

y	G	G_{χ_3}	$G_{\chi_{23}}$	$G_{\chi_3\chi_{23}}$	$R_{3,23}$
1000	1.531	1.040	1.332	−5.704	6.944
1000000	11.040	−16.153	−97.818	−657.145	749.850

(II.24)

To six decimals the three constants are $c_{\chi_3} = 0.368282$, $c_{\chi_{23}} = -0.422121$, and $c_\eta = 0.056523$. They can be computed by grouping the Dirichlet series into residue classes and differentiating its Laurent expansion at 1. The mean excess in section 5.2 is 12.863118 to six decimals. Its distinguished part accounts for 8.679631 and the other cells for 4.183488, rounded separately. The latter need an estimate in the asymptotic proof. [3, 6]

5.4 Discarding the other scale with a uniform estimate

Divisibility uses all the cells, but only the distinguished cells should measure primes near x. Small weights at the other cells push their cutoffs far enough out for the prime number theorem to control them. The fixed-modulus limit $R_i(y)/y \to 0$ gives no rate uniform in q, d, N. For the moduli of section 3 one has, for each primitive character in the cell expansion and for $y \geq \exp(x^{\varepsilon/3})/3$,

$$|G_\chi(y)| \leq y\,\vartheta, \qquad \vartheta = \exp(-\kappa x^{\varepsilon/8}), \qquad \kappa > 0, \tag{5.6}$$

with G for the trivial character. The constant κ does not depend on x, q or d, and the bound is for $x \geq x_0(\varepsilon, \kappa)$. This is Estimate C. It is proved in Part III, §III.2 (Theorem III.2.1), where κ is called c'. At any nondistinguished cell i use (5.6) at $y = a/w_i$. Each remainder is a combination $R_i = \sum_\chi f_\chi G_\chi$ of the functions G_χ for the characters of the table, with $\sum_\chi |f_\chi| = 2$: the sum of the absolute values of the $4n$ Fourier coefficients is $4n/(2n) = 2$. Therefore

$$|w_i R_i(a/w_i)| \leq 2a\vartheta, \qquad \frac{1}{Bn} \sum_{u,t,j} \sum_{i \text{ nondistinguished}} |w_i R_i(a_t/w_i)| \ll \frac{nx}{\delta}\,\vartheta = o(1). \tag{II.25}$$

Indeed there are at most $2n$ cells per table, $a_t/b \leq x + 1$, and nx/δ grows polynomially in x. Exponential decay in (5.6) absorbs this factor.

For a distinguished cell let $w_i = b + \Delta_i$ with $0 \leq \Delta_i \ll b/(nx^2)$. The function $w \mapsto (a/r - w)_+$ changes by at most Δ_i and only $r \leq a/b$ can contribute. Since $0 \leq c(r) \leq 2\Lambda(r)$, Chebyshev's estimate bounds the total change in the prime sum by $O(x\Delta_i)$. In its main term aC_i is unchanged, and $\mu a \log(w_i/b) \leq \mu a \Delta_i / b$. Consequently

$$|w_i R_i(a/w_i) - b R_i(a/b)| \ll x\Delta_i, \qquad \frac{1}{\delta b/2} \sum_{i \text{ distinguished}} x\Delta_i \ll 1/x. \tag{II.26}$$

For the second bound, there are $\delta n = |T_2|$ distinguished cells in each table. Summing over u, t, j and dividing by Bn gives $O(1/x)$.

Finally $|a_t/b - x| \leq 1$. When y moves by at most 1 from x, $S_{q,d}(y)$ and $y \log y$ change by $O(1 + \log x)$, since $\sum_{r \leq x+1} \Lambda(r)/r \ll \log x$. For $\chi = \chi_q, \chi_d, \eta$, Theorem III.1.1(d) of Part III gives $|c_\chi| \leq 4 \log x$ (it also follows from Estimate B and the definition of G_χ). The same estimate for the variation thus holds for the centering terms. Combining (5.3), (5.5), and the two error bounds gives

$$\frac{1}{n} \sum_{u,t,j} (P_{u,t,j} - M_{u,t,j}) = B\,R_{q,d}(x) + O_\varepsilon\big(B(1 + \log x)\big). \tag{5.7}$$

With (M), this gives $R_{q,d}(x) \leq \mathcal{E}/B + O_\varepsilon(1 + \log x)$, proving (A). Only an upper bound follows. In the finite example the changes of distinguished weight and argument are both zero, so before estimation (5.7) is the numerical split of the mean excess into 8.679631 and 4.183488. [6, 11]

6. Where the main inequality (M) comes from

Sections 6.1–6.2 construct the maps from theta series. Sections 6.3–6.4 explain why their derivatives are divisible by powers of inert primes. Sections 6.5–6.7 bound the size of the same derivatives and identify the main term. Sections 6.8–6.10 control denominators and show that enough independent images survive. We prove the elementary calculations in each step and say which arithmetic statements are quoted.

6.1 A theta lift in classical terms

A theta series records representations of integers by a quadratic form. For instance put $L = \mathbb{Z}[i]$, $Q(a+bi) = a^2+b^2$ and z in the complex upper half-plane. With $s = \exp(2\pi i z)$, its theta series is

$$\Theta_L(z) = \sum_{\lambda \in L} s^{Q(\lambda)} = 1 + 4s + 4s^2 + 4s^4 + 8s^5 + \cdots . \tag{II.27}$$

The coefficient of s^5 is 8 because the solutions are $(\pm 1, \pm 2)$ and $(\pm 2, \pm 1)$. Absolute convergence follows from $\sum_{a,b} \exp(-2\pi\, \Im z\, (a^2+b^2)) < \infty$. Fourier orthogonality recovers a coefficient by pairing: for each $y > 0$,

$$e^{10\pi y} \int_0^1 \Theta_L(u+iy) e^{-2\pi i 5u}\, du = 8. \tag{II.28}$$

Indeed $\int_0^1 e^{2\pi i(k-5)u}\, du$ is one for $k = 5$ and zero for every other integer k.

To make this a lift, allow the lattice to vary within a finite collection of isometry classes L_i. Attach to each class a value $f(i)$. Weight it by $h_i = 1/|\operatorname{Aut} L_i|$, the reciprocal of the number of its symmetries in the chosen group. The theta series are now a kernel depending on both i and z. The lifted function is

$$\Theta(f)(z) = \sum_i h_i \overline{f(i)}\, \Theta_{L_i}(z). \tag{II.29}$$

In the rank-one Hermitian example, unitary symmetries of $\mathbb{Z}[i]$ are multiplication by $1, -1, i, -i$. On its single class with $f = 1$ this normalization gives $\Theta_L/4$, so the same Fourier pairing gives $8/4 = 2$. This example shows what integration against a theta kernel does. The high-rank construction has different groups and test functions.

For the source let $E = \mathbb{Q}(\zeta_N + \zeta_N^{-1}, \sqrt{-q}, \sqrt{-d})$, with $\zeta_N = \exp(2\pi i/N)$, and let F denote its maximal real subfield. Write $^-$ for complex conjugation on E. A Hermitian matrix J satisfies $\overline{J}^t = J$. Its unitary group consists of matrices g with $\overline{g}^t J g = J$. It is definite if its form at every real embedding of F is definite. At a fixed integral level and scalar weight, forms on a definite group can be viewed as functions on finitely many lattice classes. At higher weight the values are vectors and must transform under the finite stabilizers.

For a pair of unitary groups, the general lift has the same two operations:

$$\Theta_\varphi(g,h) = \sum_{\lambda \in V(F)} (\omega(g,h)\varphi)(\lambda), \qquad \mathcal{T}_\varphi f(h) = \int_{[U(J)]} \Theta_\varphi(g,h) \overline{f(g)}\, dg. \tag{II.30}$$

Here V is a finite-dimensional coordinate space for the theta sum; $V(F)$ denotes its rational points over F. The notation $[U(J)]$ denotes integration over the group at all places modulo its rational points, with its invariant measure. The test function φ decays rapidly at infinity and is locally constant of compact support at finite primes. The Weil representation ω is the action of the two groups on these test functions. Lattice indicators at the finite primes restrict the sum to lattices.

In the displayed convention the lift is linear in φ and conjugate-linear in f. [13] (The linearity statement of [13] is a statement in the archive, not used by the final theorem.)

6.2 Composing maps and clearing their denominators

For a finite form of the calculation let I_j be the finite set of classes at stage j, $h_j(i)$ its weights, and $K_j(i,i')$ the matrix of the jth map. In the linear convention write

$$(T_j f)(i') = \sum_{i \in I_j} h_j(i) K_j(i,i') f(i). \tag{II.31}$$

Complex conjugates from a pairing are absorbed by taking the conjugate source space. Expanding two such sums gives the composite kernel $\sum_{i'} K_j(i,i') h_{j+1}(i') K_{j+1}(i',i'')$. Thus successive maps really do use the same intermediate index.

In the source the definite ranks are $N_j = k + j(\tau_0 + 1)$ for an odd integer τ_0. The rank conditions [26] include m even and $m \geq 2$, k even and $k \geq m+3$, $\tau_0 \geq k+m+4$, and $\tau_0 > 3m$. For instance $m = 2, k = 6, \tau_0 = 13$ gives ranks $6, 20, 34$ and satisfies these numerical constraints. The full chain also needs field, level, and lattice data. Since m is even and $m \geq 2$, the chain has at least two steps.

At the end of the chain the complex pairing of the entire chain can be expressed as a single period integral over a subgroup. The source proves this identity with prescribed normalization factors and coordinate changes. Its use is that the final period can be compared at different choices of the intermediate character twists. [7, 14]

The integrality assertion for a finite chain. Let S be a subring of a commutative ring R. Suppose all $K_j(i,i')$ and starting values belong to S. If scalars $D_j \in R$ satisfy $D_j h_j(i) \in S$ for every i, then every final value multiplied by $\prod_{1 \leq j \leq m} D_j$ belongs to S. For one map this follows by moving D_1 into its finite sum. Applying the same argument to the already cleared values proves the statement by induction. For vector values, replace S by a lattice at each stage and require each $K_j(i,i')$ to carry the preceding lattice into the next. This is the assertion stated as ChainInt, with its vector version, in [14] (a statement in the archive, not used by the final theorem).

For example, on one class let the successive weights be $1/2, 1/3$, the kernel entries $5, 7$, and the starting value 1. The output is $35/6$, and clearing the two weights gives 35. At the prime 5, two entries divisible respectively by 5^2 and 5^3 would give a cleared output divisible by 5^5. In a sum the exponent can increase through cancellation. It is always at least the common lower bound on the summands. This elementary induction explains how divisibility persists. It requires separate local estimates to put significant prime powers into each kernel entry.

Where a period enters. A kernel evaluated in complex coordinates need not be integral. An algebraic local coordinate and a complex analytic coordinate on the same elliptic curve differ at the tangent space by a nonzero scale. An ath derivative acquires its ath power. Taking logarithms gives a term linear in a. In the global comparison the local gains, the period, and the lattice indices must all pertain to the same normalized maps. We next give a local example in which one can track every derivative, and then the estimates that take its place in the chain.

6.3 A genuine local example and an elementary minor calculation

Let $s(z)$ be the inverse near 0 of $t \mapsto \int_0^t (1-u^4)^{-1/2}\,du$, and let $A_j = s^{(j)}(0)$. The lemniscate curve $y^2 = 1 - x^4$ has the symmetry $(x, y) \mapsto (ix, y)$ fixing the origin $(0, 1)$. This yields complex multiplication by $\mathbb{Z}[i]$. Differentiating the inverse function gives

$$s(0) = 0, \quad s'(0) = 1, \quad (s')^2 = 1 - s^4, \quad s'' = -2s^3. \tag{II.32}$$

The product rule gives an integer recursion, with $A_0 = 0, A_1 = 1$,

$$A_{j+2} = -2 \sum_{a+b+c=j} \frac{j!}{a!\,b!\,c!}\, A_a A_b A_c. \tag{II.33}$$

For $j = 3$, only $(1, 1, 1)$ contributes, giving $A_5 = -12$. For $j = 7$ there are three permutations of $(5, 1, 1)$, giving $A_9 = -2 \cdot 3 \cdot 42 \cdot (-12) = 3024$. Uniqueness of the series solution gives $s(iz) = is(z)$, so only indices $j \equiv 1 \bmod 4$ can contribute. The exact computation gives

$$A_{17} = 21224560896 = 2^8\, 3^6\, 7^2\, 11\, 211. \tag{II.34}$$

The primes $3, 7, 11$ remain prime in $\mathbb{Z}[i]$. Their exponents are $6, 2, 1$; the model in section 2 predicts the respective leading terms $17 \cdot 3/8$, $17 \cdot 7/48$, $17 \cdot 11/120$. At the split primes $5, 13$ the exponents here are zero. At $j = 201$ the computed valuations at $3, 7$ are $75, 29$. These valuations are computed, not proved in general; section 2 proves the bound only for its model series.

There is a quantitative gap between this example and (4.1). At $j = 101$ the contribution to $\log|A_j|$ from inert primes at most j is 167.67 out of 306.77. At $j = 121$ it is 203.58 out of 388.67. The density calculation explains the scale: $\sum_{p \le j,\ p \equiv 3 \bmod 4} (\log p)/p \sim \frac{1}{2} \log j$, so the model exponents j/p at these primes give about $\frac{1}{2} j \log j$. The main inequality (M) requires agreement even to the term proportional to j. So the numbers A_j cannot serve in (M): they are too large by a factor 2 in the leading term (defect (D3) of §I.3).

A local calculation from the source. Let A be a commutative ring with an element π_0 satisfying $\pi_0^2 = p$, where p is prime. For an integer $e \ge 0$ put $P = p^e$. Let M, N' be square matrices with $N'_{bc} \equiv M_{bc}^P \bmod p$. Suppose all but the zeroth coordinates of row vectors u, u' are divisible by π_0. Write $X = uM$, $Y = u'N'$. The minor used in [15] satisfies

$$\pi_0 \mid Y_c X_0^P - Y_0 X_c^P. \tag{II.35}$$

Indeed, modulo π_0, $X_c = u_0 M_{0c}$ and $Y_c = u'_0 M_{0c}^P$. Substitution makes the two products equal. For $p = 3$, $\pi_0 = \sqrt{3}$, take $P = 3$, $M = N' = \left(\begin{smallmatrix} 1 & 1 \\ 0 & 1 \end{smallmatrix}\right)$, $u = (1, \pi_0)$, $u' = (1, 0)$. The minor for $c = 1$ is $1 - (1 + \pi_0)^3 = -3\pi_0 - 3\pi_0^2 - \pi_0^3$, visibly divisible by π_0. The estimate on the full lift requires many such congruences and their behavior under differentiation.

6.4 How the weight cuts down the local exponent

Here is a calculation which accounts for the positive part in (4.1). A directional Hasse derivative of a polynomial Q is the coefficient of Z^{β} in $Q(X+\sum_j b_j Z_j)$, where b_j are vectors, Z_j indeterminates, and β a list of nonnegative exponents. It has no factorial denominator. For a product $\prod_c Q_c$ the coefficient is the sum of products over decompositions $\beta=\sum_c \beta_c$. If for a given set S of directions the coefficient for Q_c vanishes whenever $\sum_{j\in S}(\beta_c)_j > m_c$, then the coefficient for the product vanishes whenever $\sum_{j\in S}\beta_j > \sum_c m_c$: each term exceeds the allowance of some factor. This is the product-vanishing lemma in [21].

For a determinant the allowance has a simple origin. In $\det(X+sb, Y+tb)$, where X, Y, b are columns of length two, the coefficient of st is $\det(b,b)=0$, and there is no term in s^2 or t^2. Each minor can take at most one such hit. Products of C minors can take at most C. With several directions confined to specified columns, the number of available columns provides another upper bound. These are constraints on nonvanishing terms before any prime estimate is made.

For an inert prime p, write $\gamma_p = p/(p^2-1) = \sum_{h\ge 1,\ h\ \mathrm{odd}} p^{-h}$. Suppose a local expansion of derivative order a supplies a valuation lower bound $\gamma_p a - \sum_h i_h - D_0$, where D_0 is a loss, and the nonvanishing terms have

$$0 \le i_h \le \min(a/p^h, w_h) \qquad (h \text{ odd}). \tag{II.36}$$

Then the gain after charging these terms is at least

$$\sum_{h\ \mathrm{odd}} \Big(a/p^h - \min(a/p^h, w_h) \Big) - D_0 = \sum_{h\ \mathrm{odd}} (a/p^h - w_h)_+ - D_0. \tag{6.1}$$

The first bound $p^h i_h \le a$ is a degree bound. The second is supplied by column and minor capacities in the translated weight data. For example $p=3$, $a=81$, $w_h=2$ gives 25 at $h=1$, 1 at $h=3$, and zero thereafter. Obtaining all these bounds together is the arithmetic content. In the source the factorial valuations in the polynomial substitution give the initial lower bound, other terms have to be paid for, and a local lattice index must also be subtracted. [19, 21]

The factorial count, with proof. The initial lower bound in the last paragraph comes from valuations of factorials. Here is the source's statement in full, with an elementary proof. Let p be prime and put $\rho_p = 1/(p^2-1)$, so that $\gamma_p + \rho_p = 1/(p-1)$. For an integer $n \ge 0$ let $\lambda(n)$ be the number of digits of n in base p, with $\lambda(0)=0$. Let S be a finite set and $S_0 \subset S$. Suppose given nonnegative integers as follows:

(i) α_s and ℓ_s for $s \in S$, with $\alpha_s = 0$ outside S_0;

(ii) lists $e_{s,1}, \dots, e_{s,r_1}$ with sum α_s;

(iii) lists $a_{s,1}, \dots, a_{s,r_0 r_1}$ with sum at most ℓ_s;

(iv) $i_{s,h}$ for h in a finite set of nonnegative integers, zero outside S_0, with $p^{2h+1} i_{s,h} \le \alpha_s$;

(v) with $A = \sum_s \alpha_s$ and $\Lambda = \sum_s \ell_s$, one has $\sum_s \sum_j a_{s,j} + \sum_{s,h} i_{s,h} = \Lambda$.

Let V be the sum of $v_p(n!)$ over all the $e_{s,j}$, $a_{s,j}$, $i_{s,h}$. Then for every integer i_2 with $p i_2 \le \Lambda$,

$$V \ \ge\ \gamma_p A + \rho_p (A+\Lambda) + i_2 - \Delta, \qquad \Delta = \sum_{s\in S} \big(r_1 \lambda(\alpha_s) + r_0 r_1 \lambda(\ell_s) \big) + \sum_{s\in S_0} \lambda(\alpha_s)^2. \tag{6.1a}$$

Proof. Legendre's formula is $v_p(n!) = (n - \sigma_p(n))/(p-1)$, where $\sigma_p(n)$ is the sum of the digits of n in base p. Each digit is at most $p-1$, so $v_p(n!) \ge n/(p-1) - \lambda(n)$. The integers in the three families add up to $A+\Lambda$, by the two hypotheses on sums. Hence $V \ge (A+\Lambda)/(p-1) - \sum \lambda(n)$.

Now count digits. Each $e_{s,j}$ is at most α_s, and there are r_1 of them: at most $r_1\lambda(\alpha_s)$. Each $a_{s,j}$ is at most ℓ_s, and there are $r_0 r_1$ of them: at most $r_0 r_1 \lambda(\ell_s)$. If $i_{s,h} \geq 1$ then $p^{2h+1} \leq \alpha_s$, so $h < \lambda(\alpha_s)$ and at most $\lambda(\alpha_s)$ values of h occur; each such $i_{s,h}$ is at most α_s. These contribute at most $\lambda(\alpha_s)^2$, and only for $s \in S_0$. Thus $\sum \lambda(n) \leq \Delta$. Finally $(A+\Lambda)/(p-1) = \gamma_p A + \gamma_p \Lambda + \rho_p(A+\Lambda)$, and $\gamma_p \Lambda \geq \Lambda/p \geq i_2$. □

In (6.1a) the term $\gamma_p A$ is the model gain $A(1/p + 1/p^3 + \cdots)$ of section 2. The term $\rho_p(A+\Lambda)$ pays exactly for the denominators $1/(p^2-1)$ per unit of degree met there in the inverse series, and i_2 pays for one further division. The loss Δ is small: per index s it is at most $(r_1 + r_0 r_1 + \lambda)\lambda$, where λ is the larger of the digit counts $\lambda(\alpha_s)$, $\lambda(\ell_s)$. For example, with $p = 3$, one index s, $r_0 = r_1 = 1$, $\alpha = 81$, $\ell = 0$ and no $i_{s,h}$: $V = v_3(81!) = 40$, while $\gamma_3 A + \rho_3 A = 81/2 = 40.5$ and $\Delta = \lambda(81) + \lambda(81)^2 = 30$. The inequality $40 \geq 40.5 - 30$ holds with room to spare; the main terms agree to within one. This lemma supplies valuations of individual factorials. It does not identify which factorials occur in the theta coefficients, which is part of the quoted local theorem. [21]

The coefficient $2\log p$**.** For the coefficient field L_0 and a prime ideal $\mathfrak{p}$ over p, write $e_\mathfrak{p}, f_\mathfrak{p}$ for the ramification index and residue degree. The normalized local weight is $\lambda_\mathfrak{p} = \log \mathrm{Norm}(\mathfrak{p})/[L_0 : \mathbb{Q}]$. The source measures a local gain y in units $2\lambda_\mathfrak{p} e_\mathfrak{p} y$. Since $\mathrm{Norm}(\mathfrak{p}) = p^{f_\mathfrak{p}}$ and $\sum_{\mathfrak{p}|p} e_\mathfrak{p} f_\mathfrak{p} = [L_0 : \mathbb{Q}]$, one unit at every place over p sums to $2\log p$. For example in $\mathbb{Q}(i)$ above 3, $e = 1$, $f = 2$ and $\lambda = (\log 9)/2 = \log 3$. The factor 2 is the squared-norm normalisation of §I.4.2(ii). At the ramified prime d the splitting condition has half weight and gives $\log d$. These are the coefficients $c(r)$ of section 4.1; section 5.2 checked that the character expansion reproduces them. [3, 22]

The concrete minor, the Hasse product rule and (6.1) are three elementary ingredients of the local argument. Applying them to all the normalized theta kernels, choosing the usable places, and bounding the losses after summation requires the source's local divisibility theorem. In its statement the left side is the sum of the odd-prime-power expressions **minus** the orbit and rounding losses. The right side is the sum of actual local gains over the m maps **minus** the local lattice correction. [19] The divisibility at inert primes is written up in Part III, §III.5, conditionally on the integrality of the theta kernels in the coordinates of the height-two logarithm (hypothesis (B)), which is taken from the Lean development; under that hypothesis §III.5 proves the lower bound for the valuation of a product of minors applied to one such function (for Frobenius orbits of length two; for longer orbits it also uses a duality statement cited from the Lean), with the positive parts of (6.1), and it proves without hypothesis the estimate of section 2 for the inverse of the logarithm in several variables; it does not prove the passage to lattices, and the congruence of section 6.3 is not used there. The passage to lattices is treated in Part III, §III.6: the lemmas on lattices and the splitting of the prime sum are proved there, the theorem on the glued lattice is conditional on two statements of the source, and the bounds for the loss sums are quoted.

6.5 Which theta function is differentiated?

At step j let $N_j = k + j(\tau_0 + 1)$ be the rank of the definite unitary group, where τ_0 is the odd rank parameter. The coordinates of the holomorphic theta function, for a pair of lattice classes L, L' at consecutive stages and a complex embedding of the coefficient field, are the entries of a rectangular matrix at each weight place:

$$z = (z_{\tau,\alpha,\beta}), \qquad 1 \le \alpha \le N_j, \quad 1 \le \beta \le N_{j+1}. \tag{II.37}$$

Write its Taylor series at zero as $\Theta_{L,L'}(z) = \sum_\nu c_\nu z^\nu$, using a multi-index $\nu = (\nu_g)$ with $g = (\tau, \alpha, \beta)$ and $\nu! = \prod_g \nu_g!$. For vectors in the consecutive spaces of forms there is a polynomial $P(z) = \sum_\nu P_\nu z^\nu$ determined by their weights and pairing. In the formulation of [23] (a statement in the archive; see the start of Part III) the local expression in a matrix entry is

$$P(\partial)\Theta_{L,L'}(0) = \sum_\nu P_\nu \, \nu! \, c_\nu. \tag{6.2}$$

This explains why the polynomial minors and the factorials occur together. At each place the monomials in P have prescribed total degree. The global matrix coefficient sums the displayed expressions over preceding classes with reciprocal stabilizer weights. The source also describes the theta function with algebraic coordinates at finite places so that the same normalized pairing has the local divisibility. That identification of rational structures is a theorem of the construction. That the entries used by the final theorem are of the form (6.2) is quoted. [23]

On the complex side the polynomial has the squared Gaussian norm

$$\|P\|_{\mathrm{F}}^2 = \sum_\nu |P_\nu|^2 \nu!. \tag{II.38}$$

Here is a proof of its simplest case. In the complex plane give area the density $\pi^{-1} e^{-|z|^2}$. Integration in angle makes different powers orthogonal, and the squared norm of z^a is

$$2\int_0^\infty r^{2a+1} e^{-r^2}\, dr = \int_0^\infty t^a e^{-t}\, dt = a!. \tag{II.39}$$

The last integral follows by a integrations by parts. For several independent variables the integrals multiply. For a positive width Ω, the density $(\pi\Omega)^{-1}\exp(-|z|^2/\Omega)$ gives the squared monomial norm $a!\Omega^a$. In the dual coefficient estimate the scale appears with the inverse power. Weighted Cauchy–Schwarz gives the inequality stated in the archive [23]

$$|P(\partial)\Theta(0)|^2 \le \|P\|_{\mathrm{F}}^2 \sum_{\nu \in \operatorname{supp} P} \nu! |c_\nu|^2, \tag{II.40}$$

where $\operatorname{supp} P$ consists of its monomials with nonzero coefficients. Apply Cauchy–Schwarz to the two lists $P_\nu\sqrt{\nu!}$ and $c_\nu\sqrt{\nu!}$.

The analytic estimates bound the coefficient factor using the Gaussian-weighted norm of the theta function on a fundamental domain. The polynomial factor is bounded by a product of hook factors of the weights times the squared norms of the two input vectors. In one row of length a the hook factor is exactly $a!$. The full operator bound also contains the period, level, multiplicities and theta mass. [23, 24] The estimates for the two factors are proved in Part III, §III.4, apart from the four conditional passages listed at the start of that section.

Comparison with section 6.3. In the lattice comparison the archimedean map height is the average over infinite places v of $\log \|T_v\|_{\mathrm{op}}^2$, where the operator norm uses the two chosen Hermitian

metrics. Divisibility by p^e thus contributes $2e \log p$ when measured at all places above p, and a degree-a monomial contributes $\log(a!) = a \log a - a + O(\log(a+1))$ in the squared Gaussian norm. The difference from section 6.3 is visible in $A_a = a![z^a]s(z)$, where $[z^a]$ means the coefficient of z^a. The nearest pole of s is at finite positive distance R; the radius formula gives $\limsup_a |[z^a]s|^{1/a} = 1/R$. Along a subsequence, therefore, $\log |A_a|^2 = 2\log(a!) + O(a)$, while the squared Gaussian norm contributes just one factorial. [24, 25]

A natural question. Why not differentiate a classical entire theta function, which already has Gaussian size? For the lemniscatic lattice the Weierstrass function σ of $y^2 = 4x^3 - 4x$ has integer derivatives $\sigma^{(n)}(0)$; they vanish unless $n \equiv 1 \bmod 4$, and the nonzero ones satisfy $\log|\sigma^{(n)}(0)| \le \frac{1}{2} n \log n + O(n)$, by the Gaussian growth of σ and Cauchy's formula. Exact computation for $n \le 241$ shows that their divisibility at inert primes falls well short of the model: the exponent of 3 is 13 at $n = 101$ and 28 at $n = 241$, against $3n/8 = 37.9$ and 90.4. So Gaussian size does not by itself bring the divisibility with it. The source's local theorem establishes both properties together for its theta kernels, in coordinates chosen for that purpose, and it is quoted here. The computation is in the file `e4_sigma.py` listed under Computations. [21, 23]

6.6 Size of a derivative, its period, and the count of forms

The model for the size of a derivative is this. Let f be an entire function with $|f(z)| \le S\, e^{|z|^2/2}$ for all z, where $S > 0$. Cauchy's formula on the circle $|z| = \sqrt{a}$ gives $|f^{(a)}(0)| \le S\, a!\, (e/a)^{a/2}$ for $a \ge 1$. By Stirling's formula, if $f^{(a)}(0) \ne 0$, then $\log|f^{(a)}(0)|^2 \le a \log a - a + 2 \log S + O(\log(a+1))$: one factorial for the square. Rescaling the variable by the period adds $-a \log \Omega_E$, with Ω_E as defined below. This is the model at the start of Part III, §III.4. It identifies the factorial and scale terms of (6.5). The bound for the norm of an actual composed map needs additional estimates.

Here is the exact period normalization appearing in the comparison. It is the normalization for the field E, which is the one of the source. Let $\mathrm{rd}_E = |\mathrm{disc}\, E|^{1/[E:\mathbb{Q}]}$ be the root discriminant of E, and put $v = \mathrm{rd}_E/2$. The normalized period Ω_E is the positive number with

$$\log \Omega_E = \gamma + \log \pi - \log v - c_{\chi_d} = \gamma + \log(2\pi/\mathrm{rd}_E) - c_{\chi_d}. \tag{6.3}$$

This is the Chowla–Selberg identity in the form used. The period Ω_E measures squared size. In [16] the period is not constructed from a specific curve: the statement is that a curve and a normalisation exist for which (6.3) holds. It is important here which period normalization is used. Rescaling the differential changes the linear term. [16]

Remark. For the elliptic curve itself, with complex multiplication by $\mathbb{Q}(\sqrt{-d})$, the same identity holds with $v = \sqrt{d}/2$, the imaginary part of the quadratic lattice parameter $(1+i\sqrt{d})/2$, and with the period Ω of the curve: $\log \Omega = \gamma + \log(2\pi/\sqrt{d}) - c_{\chi_d} = \gamma + \log \pi - \log v - c_{\chi_d}$. In this form the identity is Kronecker's limit formula, and Ω is a geometric mean over the ideal classes of $\mathbb{Q}(\sqrt{-d})$. For $\mathbb{Q}(i)$, with d replaced by 4, exponentiating the same expression gives $\Omega = 2\pi(\varpi/\pi)^2$, with ϖ the lemniscate constant. For $d = 23$ the formula gives $\log \Omega = \gamma + \log(2\pi/\sqrt{23}) + 0.422120948038\ldots$. The two normalizations differ by $\log \Omega - \log \Omega_E = \log(\mathrm{rd}_E/\sqrt{d})$, and the sum of the logarithm of the period and $\log v$ is the same in both: $\log \Omega_E + \log(\mathrm{rd}_E/2) = \log \Omega + \log(\sqrt{d}/2)$. Only this sum enters (6.6). So (6.6) is the same in either normalization, and the numerical check in section 6.7 may use either. The two lines of (6.5), taken separately, hold only with Ω_E and $v = \mathrm{rd}_E/2$.

The other source of linear terms is a count of forms. The mass of a definite genus is the sum of reciprocal stabilizer orders over its lattice classes, as in section 6.1. The dimension of a representation with strictly decreasing shifted integral weights $z_1, \ldots, z_r$ is

$$D_r(z) = \frac{V_r(z)}{\prod_{j=0}^{r-1} j!}, \qquad V_r(z) = \prod_{1 \le i < j \le r} (z_i - z_j). \tag{II.41}$$

The scale factor in the normalized mass changes from rank $r-1$ to r according to the mass formula. Write $f = 2n$ for the number of weight places. In the algebraic identity stated in [17] (the statement

of [17] is not used by the final theorem; a statement of the same name elsewhere in the archive is), M_r denotes the mass and $\kappa_r > 0$ collects the extra local factors:

$$\frac{M_r}{M_{r-1}} = \big((r-1)!(v/\pi)^r\big)^f (2v)^{-f/2}\kappa_r. \tag{II.42}$$

Here $v = \mathrm{rd}_E/2$, as in (6.3), and M_r is unrelated to the subtracted main term $M_{u,t,j}$. For two ranks $r_0 < r_1$ (for instance two consecutive ranks N_j, N_{j+1} of the chain of section 6.2), let u_ρ, w_ρ be their decreasing weight lists at each of the f places ρ. Multiplying the mass ratios and the f dimension ratios gives precisely

$$\frac{M_{r_1}\prod_\rho D_{r_1}(w_\rho)}{M_{r_0}\prod_\rho D_{r_0}(u_\rho)} = \prod_\rho \frac{V_{r_1}(w_\rho)}{V_{r_0}(u_\rho)} (v/\pi)^{f\sum_{r=r_0+1}^{r_1} r} (2v)^{-(r_1-r_0)f/2} \prod_{r=r_0+1}^{r_1} \kappa_r. \tag{6.4}$$

Indeed the extra $(r-1)!$ in the mass cancels the extra $(r-1)!$ in each dimension denominator. Taking logarithms is the count identity on ordered weights. For a numerical algebra check take one place, $r_0 = 2$, $r_1 = 3$, and $u = (3,1), w = (5,3,0)$. The two dimensions are 2 and 15. The ratio of dimensions is $15/2$. The factorial from the mass is 2, leaving $15 = V_3(w)/V_2(u) = 30/2$ before the scale and local factors. Equation (6.4) follows directly from the displayed mass recurrence. The automorphic mass formula supplying that recurrence is quoted. The passage from the count identity (6.4) to the second line of (6.5) below (the definition of the level $\mathcal{T}$, the factors κ_r, the error E_{cnt}) is also quoted.

6.7 Why the linear term has precisely the required sign

Write $\mathcal{T}$ for the positive auxiliary level attached to one choice of lattice data. It is different from the set T of derivative labels. Let $\ell_g = \log(g_0 g_1)$ for the two auxiliary level primes g_0, g_1. The size and count bounds [7, 19] have the following form at a fixed translate, with Ω_E and $v = \mathrm{rd}_E/2$ as in (6.3). Here h_B stands for the combination $\mu_{\mathrm{CC}} + h - F_0$ of three quantities defined after the display, and $E_{\mathrm{size}}, E_{\mathrm{cnt}}$ are the respective error allowances:

$$\begin{aligned} h_B &\le \sum_{t\in T}\sum_{j=1}^{m} a_t(\log a_t - 1 - \log\Omega_E) - \sum_{t\in T} a_t \log\mathcal{T} + E_{\mathrm{size}}, \\ \log\mathcal{T} &\ge \sum_{j=1}^{m} \big(\langle \log w_{u,j}\rangle + \log v - \log\pi\big) + m\ell_g - E_{\mathrm{cnt}}. \end{aligned} \tag{6.5}$$

The three quantities in $h_B = \mu_{\mathrm{CC}} + h - F_0$ are these. First, h is an upper bound for the height of the pairing map, the average over the archimedean places of the logarithm of the squared operator norm (section 6.5). Second, μ_{CC} is the mean slope of the common lattice in which all the pairing maps take their values. Third, F_0 is the part of the local divisibility that comes from the places other than those of the top weights.

Neither offset survives. The term F_0 also appears, with the opposite sign, in the slope of the source lattice after gluing (section 6.8), and the two cancel. The term μ_{CC} is the slope of the target in the inequality of section 6.10, where it appears on both sides. So one may read (6.5) as a bound for the height h alone, as in Part I. An earlier formulation [18] writes the same pair of inequalities for a single quantity.

The local divisibility comparison bounds the prime sum by $h_B + m\ell_g \sum_t a_t$ plus losses; the second term is a level correction. Insert the *lower* bound for $\log\mathcal{T}$ in the first line of (6.5). Since each a_t is positive, this gives an upper bound. The terms $m\ell_g \sum_t a_t$ cancel, and the coefficient of a_t at step j is

$$\begin{aligned} &\log a_t - 1 - \log\Omega_E - \langle \log w_{u,j}\rangle - \log v + \log\pi \\ &\qquad = \log a_t - 1 - \gamma + c_{\chi_d} - \langle \log w_{u,j}\rangle. \end{aligned} \tag{6.6}$$

The equality is (6.3). Thus the sum of the surviving terms is exactly $\sum_{t,j} M_{u,t,j}$. The two errors in this substitution are $E_{\mathrm{size}} + E_{\mathrm{cnt}} \sum_t a_t$. Other local and lattice losses must still be added to them. The main term M needs all three: Stirling's formula ($\log a - 1$), the period (6.3) ($-\gamma + c_{\chi_d}$), and the count ($-\langle \log w\rangle$).

In the finite table one can check the cancellation even though no level or lattice has been constructed for that table. For $a = 10000$ take the single-step expression on the first line of (6.6), multiply by a, and use the normalization of the elliptic curve (the remark in section 6.6): put $v = \sqrt{23}/2$. The period combination $-\log\Omega - \log v + \log\pi$, which is the same as $-\log\Omega_E - \log(\mathrm{rd}_E/2) + \log\pi$, equals $-\gamma - 0.422120948038\ldots$. The remaining two entries are $\log 10000 - 1$ and $-(\log 10 + 7\log 0.05)/8$. This is the 95444.4636 computed independently by prime summation. An uncancelled constant per derivative in (6.6) would survive multiplication by $a_t \simeq bx$; division by B would leave an error at least of order x.

What is quoted in this comparison. The period formula (6.3), the mass formula, the construction and count of the auxiliary levels, and the local level correction are the arithmetic inputs; they are quoted. The analytic norm bound is proved in part in Part III, §III.4. The factorial cancellation (6.4) and all signs in (6.6) are algebraic consequences proved above. The next comparison explains why the divisibility can be applied to enough independent vectors at once.

6.8 From one determinant to lattices

A lattice L of rank r in a Euclidean space is the integer span of r independent vectors. Its covolume, measured in its span, is the volume of the parallelepiped on a basis. Put $\mu(L) = -\log(\operatorname{covol} L)/r$. Let T be an injective linear map with $T(L) \subset M$, and let $\|T\|$ be its Euclidean operator norm. (This section uses the plain normalisation, μ and $\log\|T\|$. In (6.5)–(6.6) the height h is the logarithm of the *squared* norm, and μ_{CC} is normalised to match, so it is twice a slope in the present sense, up to an additive constant depending only on the field. The two normalisations must not be mixed.) The volume of $T(L)$ is at most $\|T\|^r$ times that of L, as one sees using an orthonormal basis of the span and Hadamard's inequality. Taking logarithms,

$$\mu(L) \le \mu(T(L)) + \log\|T\|. \tag{II.43}$$

Suppose first that M has the same rank r. Then $T(L)$ has finite index in M and $\operatorname{covol} T(L) = [M : T(L)] \operatorname{covol} M$, so

$$\mu(L) + \frac{1}{r}\log[M : T(L)] \le \mu(M) + \log\|T\|. \tag{II.44}$$

This is the exact form of the principle that divisibility is bounded by size: the index is on the left and the norm on the right. For $T = \operatorname{diag}(6, 10)$ and $L = M = \mathbb{Z}^2$ the index is 60, both slopes are zero and $\|T\| = 10$. The inequality reads $\frac{1}{2}\log 60 \le \log 10$, that is, $60 \le 100$. The factors $2, 3, 5$ of the index and the two column lengths refer to the same map.

If L has smaller rank than M, the image $T(L)$ is only a sublattice of smaller rank, and the first inequality gives

$$\mu(L) \le \mu_{\max}(M) + \log\|T\|, \tag{II.45}$$

where $\mu_{\max}$ is the supremum of slopes of nonzero sublattices. A single map of small rank therefore controls only the maximal slope of its target. Sections 6.9 and 6.10 use many maps together in order to reach the mean slope $\mu(M)$.

A rational matrix presents a further problem: local divisibility may apply at different primes in different bases. The source uses equality of a period pairing under changes of intermediate twists to place the local lattices in a common rational vector space. One can then choose a favorable lattice at each prime and put these choices together.

Over $\mathbb{Q}$ the last operation has a short proof. Let L_p be a full lattice over $\mathbb{Z}_p$ inside $\mathbb{Q}_p^r$, equal to $\mathbb{Z}_p^r$ outside a finite set. Here $\mathbb{Q}_p$ is the completion of $\mathbb{Q}$ for the p-adic absolute value, and $\mathbb{Z}_p$ its ring of integers. Choose a positive integer D with $D\mathbb{Z}_p^r \subseteq L_p \subseteq D^{-1}\mathbb{Z}_p^r$ at all primes. Let L be the set of x in $D^{-1}\mathbb{Z}^r$ whose image in $D^{-1}\mathbb{Z}_p^r/D\mathbb{Z}_p^r$ lies in $L_p/D\mathbb{Z}_p^r$ for every $p \mid D$. The Chinese remainder theorem applied after multiplying by D identifies the ambient finite quotient with $(\mathbb{Z}/D^2\mathbb{Z})^r$, the product of its prime parts. The resulting lattice has the prescribed localization at every prime. For example allow a denominator 2 in the first coordinate only at 2, and a denominator 3 in the second only at 3. The result is $\frac{1}{2}\mathbb{Z} \oplus \frac{1}{3}\mathbb{Z}$. Its covolume is $1/6$. This proves the gluing step over $\mathbb{Q}$. The source uses the corresponding construction over a number field, with its locally chosen character twists.

The denominators that remain. For a prime ℓ and positive twist index a, the relevant finite group consists of first Galois cohomology classes over E with values in $\mathbb{Q}_\ell/\mathbb{Z}_\ell(-a)$, unramified away from ℓ and zero locally above ℓ. The twist means that Galois acts by the inverse ath power of its action on roots of unity of ℓ-power order. A finite group of order ℓ^h costs $h\log\ell$ in a logarithmic index. Write $h_{\ell,a}$ for the length of this group as a $\mathbb{Z}_\ell$-module. On the selected window the supplied bound, after passage to the working scale, is $h_{\ell,a} < \infty$ for $1 \le a < \lfloor x^{400} \rfloor$ (the number of twists

used is at most x^{400}; see Part III, §III.9) and

$$\sum_{\ell \in S} (\log \ell) \sum_{1 \le a < \lfloor x^{400} \rfloor} h_{\ell,a} \le \sqrt{b} \tag{II.46}$$

for every finite set S of primes. This is the uniform assertion needed here. (In [8] the bound is stated for $a < x^{401}$ at a slightly different scale; the form above is what is used.) Even for ζ the bound is needed over the field E, not over $\mathbb{Q}$. The statement on these extension groups, the local estimates relating their lengths to actual lattice indices, and the independence of the normalized global period are quoted. [8, 19] The slope inequalities proved in this section are the elementary case; the form with minimal slopes used in the chain is quoted (section 6.10).

6.9 Why many independent images improve the estimate

A large norm for one map could be concentrated on one direction. A second moment distinguishes this possibility. For a positive Hermitian matrix Q of size D, Cauchy–Schwarz on its nonzero eigenvalues gives

$$\operatorname{rank} Q \geq (\operatorname{tr} Q)^2 / \operatorname{tr}(Q^2) \quad (Q \neq 0). \tag{II.47}$$

Two maps with images the coordinate axes in $\mathbb{C}^2$ give $Q = I$ when one adds their orthogonal projections. The bound gives rank two. Two maps onto the same axis give $Q = \operatorname{diag}(2, 0)$, with the same trace 2 but second moment 4, and the bound gives rank one.

Here is the version needed when some source directions have already been discarded. Let $B_i : V_i \to H$ be maps of finite-dimensional complex inner product spaces, with $D = \dim H > 0$, $S = \sum_i \dim V_i > 0$ and positive weights c_i. A superscript $\dagger$ denotes the adjoint. Put $K_i = c_i B_i^\dagger B_i$ and $Q = \sum_i c_i B_i B_i^\dagger$. Let U_i be subspaces of V_i with $\sum_i (\dim V_i - \dim U_i) \leq \alpha S$, and suppose

$$\sum_i \|K_i - I\|_{\mathrm{HS}}^2 \leq eS, \quad 0 \leq e \leq 1/4, \qquad \operatorname{tr}(Q^2) \leq (1+\eta)(\operatorname{tr} Q)^2 / D. \tag{II.48}$$

Here $\|A\|_{\mathrm{HS}}^2 = \operatorname{tr}(A^\dagger A)$. **Lemma.** *Under these hypotheses the codimension of $\sum_i B_i U_i$ in H satisfies*

$$\operatorname{codim}\Big(\sum_i B_i U_i\Big) \leq \big(2\eta + 4\alpha + 4\sqrt{e}\big) D. \tag{II.49}$$

(The proof below gives the sharper bound with η in place of 2η. The form with η is the one that is used: in section 6.10 for the chain of the source, and in Part III, §III.8, Lemma T1(d).) The display, with 2η, is the finite-dimensional assertion FrameLA stated in [20] (a statement in the archive, not used by the final theorem; the chain uses a variant, FrameLA1, stated for subspaces and with η in place of 2η; see Part III, §III.8, §2).

Proof. We may take $0 \leq \alpha \leq 1$, since losing the entire space makes a larger bound automatic. Let P project onto the missing target space, of dimension zD. Then PB_i vanishes on U_i. Hence $\operatorname{tr}(PQ)$ is at most the sum of the traces of K_i over the omitted source directions, since $c_i B_i^\dagger P B_i \leq K_i$ and it vanishes on U_i. Write $K_i = I + (K_i - I)$ there and apply Cauchy–Schwarz. Also apply it to the total trace:

$$\operatorname{tr}(PQ) \leq (\alpha + \sqrt{\alpha e})S, \qquad \operatorname{tr} Q \geq (1 - \sqrt{e})S \geq S/2. \tag{II.50}$$

Writing $\bar{q} = \operatorname{tr} Q / D$, the second-moment hypothesis gives $\|Q - \bar{q} I\|_{\mathrm{HS}}^2 \leq \eta (\operatorname{tr} Q)^2 / D$. Pairing with P yields

$$z \leq \operatorname{tr}(PQ) / \operatorname{tr} Q + \sqrt{\eta z} \leq 2\alpha + 2\sqrt{e} + (\eta + z)/2. \tag{II.51}$$

This proves even $z \leq \eta + 4\alpha + 4\sqrt{e}$. The specialization to the actual maps requires a lower first moment and an upper second moment of their period pairings; the latter comes from a relative trace formula. A trace formula equates a sum over a basis of functions with a sum of geometric integrals. Here it controls correlations between the images of different maps. These automorphic estimates are quoted. [7, 19, 20] In the source the index i is a word: a string of $k-2$ binary letters which chooses, at every weight place, one of the weights that interlace a fixed weight of the group in $k-1$ variables; each word also has its own level prime. The space V_i is attached to the word, and H is common to all words. The character twists of section 6.8 are a different index (one twist for each place) and do not enter this family. There is one such family for each translate u of section 4; the weights $w_{u,j}$ depend on the translate and not on the word. See Part III, §III.8.

6.10 From the rank bound to (M)

It remains to see why a count of directions can control a mean slope. In this section H is a bound for heights and u is a spread of slopes; they are not the target space H of section 6.9 and the translate u of section 4. Let $d_1, \dots, d_D$ be real slopes, a a reference value, and assume $d_j \geq a - H$ with $H \geq 0$. Put $N(t) = \#\{j : d_j < a - t\}$. Integrating the indicator of each deficit gives

$$\frac{1}{D}\sum_j d_j \geq a - \frac{1}{D}\sum_j (a - d_j)_+ = a - \int_0^H \frac{N(t)}{D}\, dt. \tag{6.7}$$

For example, slopes $2, 3, 5$ and $a = 4$ give $N(t) = 2$ on $0 < t < 1$, 1 on $1 < t < 2$, and 0 thereafter. The mean deficit is $(2 + 1)/3 = 1$. The lower bound 3 is below the true mean $10/3$ because the largest slope exceeded 4. This is the layer-cake formula; the form used by the final theorem is in [19].

Here is how an integrable count is obtained in [20]. If R slopes y_j have mean at least s and all are at most $s + u$, $u > 0$, the fraction below $s - t$ is at most $u/(u + t)$. Indeed, with fraction z below, their mean is at most $z(s - t) + (1 - z)(s + u)$. Apply the rank bound of section 6.9 after discarding those directions; so $\alpha = u/(u + t)$ in the lemma of section 6.9, which gives the term $4u/(u + t)$ below.

(Passing from the bound on the missing directions to a count of the slopes of the target below a given level uses the inequality for the minimal slope of the image of a map, and properties of the filtration of a lattice by slopes. These are in [25]; they are quoted here, and are not the inequality proved in section 6.8.)

In the notation of [20], for numbers s, h, L, u, H, η, e with $u > 0$ and $L, H, \eta, e \geq 0$, the hypotheses take the form

$$d_j \geq s - h - H, \quad \frac{\#\{j : d_j < s - h - L - t\}}{D} \leq 2\eta + \frac{4u}{u + t} + 4\sqrt{e} \quad (0 \leq t \leq H). \tag{II.52}$$

Putting $a = s - h - L$ in (6.7) and integrating gives an error at most $(2\eta + 4\sqrt{e})H + 4u \log(1 + H/u)$. In particular it proves the looser bound stated in [20] (the chain of the source uses a variant of it, which is a hypothesis in [7] and is applied for each translate in [28]; see below)

$$\frac{1}{D}\sum_j d_j \geq s - h - L - u(1 + 8\log(1 + H/u)) - (4\eta + 8\sqrt{e})H. \tag{II.53}$$

The trace estimates make $\sqrt{e}$ exponentially small on the chosen intervals: $e = 2e_0$ with $\sqrt{e_0} \leq \exp(-x^{\varepsilon/8})$, where e_0 is the constant in the bound for the diagonal Gram blocks (Part III, §III.8). For η the source has $\eta = 5x^{-C_H} + 3\sqrt{\alpha_{\mathrm{ram}}}$. Here α_{ram} is a bound for the proportion of source directions lost at the level primes (it is not the $\alpha = u/(u+t)$ above), the source takes $\sqrt{\alpha_{\mathrm{ram}}} = \exp(-x^{\varepsilon/8})$, and C_H is the exponent in the bound $H \leq x^{C_H} b$. For all sufficiently large x, $(3+4\sqrt{2})\exp(-x^{\varepsilon/8})\, x^{C_H} \leq 1$. Since $\eta = 5x^{-C_H} + 3\sqrt{\alpha_{\mathrm{ram}}}$ and $e = 2e_0$, this gives $(\eta + 4\sqrt{e})x^{C_H} \leq 6$, and so $(\eta + 4\sqrt{e})H \leq 6b$. So η is bounded only by a negative power of x, chosen to match H. The power of x comes from the number of maps, $\lceil 3\mathsf{G}^2 x^{C_H} \rceil$, where $\mathsf{G} = \exp(C_{\mathsf{G}}\, m^5 f\, x^{\varepsilon})$ is the constant in the quoted bounds for the Gram blocks, with a constant $C_{\mathsf{G}} > 0$ (Part III, §III.8). In the variant the chain uses, the error bracket is $4u \log(1 + H/u) + (\eta + 4\sqrt{e})H$, with η and not 2η. It is multiplied by $1/(1 - \delta_{\mathrm{p}})$, where δ_{p} is the proportion of the target not spanned by its lattice (exponentially small, quoted). The lower bound also has $-\log|\mathsf{W}|$ as a separate term, which this factor does not multiply. Here W is the set of words and $\log|\mathsf{W}| \leq (k - 3)\log 2$, which is negligible; the sizes of H and of E_{cnt} are part

of the quoted statements and are not derived here. The bound for the spread u and the arithmetic control of L are also needed. This argument uses a mean slope of the target, where the elementary inequality of section 6.8 by itself would only control a maximal slope.

Assembly. The local derivative bounds place the prime sum in the integral lattices. The period identity makes it possible to compare the different local choices in one space. The height and counting comparisons (6.5)–(6.6) give the prime sum's main term. Finally the mean-slope bound accounts for losses from indices and directions which must be discarded. The exact output of the source's comparison is this: for each translate, the required divisibility sum is at most its main term plus an aggregated error [28]. The hypotheses of that statement are defined in [7]. The archive also has the form averaged over the translates, in [7]; the final theorem does not use it. The size estimates bound the error by the unnormalized expression of section 4.3. These are the inputs [5, 7, 19, 28] used at (M).

Once the quoted estimates are granted, sections 4 and 5 give (A), and section 3 concludes by balancing x/m and m. A cost of m^2 would instead give $m = x^{1/3}$ and exponent $2/3$, if the corresponding inequality were available with the other errors under control.

Proved in Part II, section 6. Finite-chain integrality, the minor congruence, the Hasse product rule, the reduction (6.1), the factorial count (6.1a), the Gaussian norm and the Cauchy–Schwarz bound of section 6.5, the count identity (6.4) and the cancellation (6.6), the elementary slope inequality, rational lattice gluing, the rank bound and its integration. (Sections 1–5 are proved in full, given (M) and Estimates B, C.)

Quoted from the Lean source. The normalized theta maps and period identity, their simultaneous local divisibilities and losses, the period and mass formulas, the level bounds, and the norm bounds (apart from the analytic estimates of Part III, §III.4); the denominator estimates; and the two moments for this family. Part III, §III.6–III.9 deduce parts of these from named statements of the source, which remain quoted; see the table at the start of Part III.

Sources

Alpöge published the Lean source as one archive, `lean.tar.gz`, split into fourteen download parts at alpo.ge (listed in footnote 1 of his announcement). Unpacked, it contains 175,187 Lean files. The SHA-256 digest of the joined archive is `9631b4f020199e580ba86178cb56c6e9b09f690c2369a80dc4334aa93f49c7fc`. The numbered references below are files of that archive; all except [4] are cited in this paper. Each is given by its path from the root of the archive.

[1] `lean/src/Development/AutomorphicArithmetic/DedekindLFunctions/LandauTheorems/landauWindow.lean`

[2] `lean/src/Formalism/ProductFormula/StatementOCCM.lean`

[3] `lean/src/Formalism/SizeWindowBrackets/FirstPrimeNeSecondPrime.lean`

[4] `lean/src/Formalism/StatementCM.lean` (definitions for the earlier form of the argument; see the remark on [2] in section 2)

[5] `lean/src/Development/AutomorphicArithmetic/DedekindLFunctions/LandauTheorems/ChTop.lean`

[6] `lean/src/Formalism/SizeWindowBrackets/GrowthBoundGivesRiemannHypothesis.lean`

[7] `lean/src/Formalism/RankGapBounds/MeanBoundSecondForm.lean`

[8] `lean/src/Development/AutomorphicArithmetic/CharacterBoxes/ExistsPageN.lean`

[9] `lean/src/Development/AutomorphicArithmetic/MeanSquareBounds/pairEx.lean`

[10] `lean/src/Development/AutomorphicArithmetic/CharacterBoxes/WinPairs.lean`

[11] `lean/src/Development/AutomorphicArithmetic/PrimitiveCosets/CoreGm.lean`

[12] `lean/src/Supplements/Kernels/MainTheorems.lean`

[13] `lean/src/Development/AutomorphicArithmetic/AdelicQuotients/ThetaLiftLinear.lean`

[14] `lean/src/Formalism/ThetaLiftIndependence/Statement.lean`

[15] `lean/src/Formalism/SharpRestriction/GrowthBounds/CoarseMultiplicity/KeyMinorDvd.lean`

[16] `lean/src/Development/AutomorphicSpectrum/Proofs/GeometryProofs/ChowlaSelbergPeriodStatement.lean`

[17] `lean/src/Formalism/BehrendSets/StatementE.lean`

[18] `lean/src/Formalism/ThickAverageLoss/PartnersToFactorsAt.lean`

[19] `lean/src/Formalism/SharpRestriction/GlobalTheoremInputs.lean`

[20] `lean/src/Formalism/CorankShare/StatementFR.lean`

[21] `lean/src/Formalism/ThetaLogCoordinates/UniformThetaSubstitution/StatementAB.lean`

[22] `lean/src/Formalism/LocalRingCardinality/DataParams₂.lean` (the file name contains the subscript character U+2082)

[23] `lean/src/Formalism/SharpRestriction/GlobalSchurBounds/FlipTwoThree.lean`

[24] `lean/src/Formalism/SharpRestriction/GlobalSchurBounds/IntegrableOfContinuousCircleMeasure.lean`

[25] `lean/src/Formalism/HarderNarasimhan/StatementPl.lean`

[26] `lean/src/Formalism/LocalRingCardinality/Params.lean`

[27] `lean/src/Formalism/FrobeniusOrbits/FrobeniusOrbitsProofs/ThickExists.lean`

[28] `lean/src/Formalism/SizeWindowBrackets/ChainSlopeWindows/LevelOfCongruence.lean`

[29] `lean/src/Development/AutomorphicArithmetic/CharacterBoxes/PairGood.lean`

Part III. Proofs of the quoted statements

Parts I and II deduce the Riemann hypothesis from statements marked [quoted]. Part III is meant to prove those statements, in the same classical language, so that the three parts together form a complete proof. It is being written in stages, and the table below records what has been done.

Statement of Parts I and II	Status in Part III
Estimate B: the auxiliary primes q, d (§III.1)	proved, from four cited theorems (Q1)–(Q4)
Estimate C: the characters of E at the large scale (§III.2)	proved, from three cited theorems (Z1)–(Z3) and Theorem III.1.1(e)
The sets T_2 and T of columns (§III.3)	constructed, with proof (elementary); their use in (M) is not yet written
The inequality (M), size of the theta kernels (§I.7, item 2; §III.4)	the analytic estimates for one block, one map and a chain are proved in an abstract setting; that the maps of the construction are instances of it is quoted (item (vi) of the list in §III.4); four passages are conditional (the reference mean square, the eigenvalues of a lift, the row-by-row comparison, the pairing at the end of the chain); the arithmetic inputs listed at the end of the section are not proved
The inequality (M), divisibility at inert primes (§I.7, item 1; §III.5)	written up in §III.5, conditionally on the integrality of the theta kernels in the coordinates of the height-two logarithm (hypothesis (B)), which is taken from the Lean development. Under (B) the local estimate for power series is proved for orbits of length two; for longer orbits it also uses a duality statement cited from the Lean. That the theta kernels satisfy (B), the frames and labels, and the remaining primes are not proved
The inequality (M), from derivatives to lattices (§III.6)	proved: the lemmas on slopes of lattices, the form of the prime sum and its splitting by classes of primes, the digit losses, the inequality at one class of primes, and bounds for several terms of the sum over primes. Quoted: the bounds for the eleven loss sums and for their total. Conditional on two statements of the source (on valuations and on lattices): the theorem on the glued lattice. A model case only: the comparison of the two sides. Quoted: the statement on valuations itself
The inequality (M), arithmetic inputs of the size bound (§III.7)	proved: the size inequality (III.75) follows from eleven named inputs; an index-corrected mean of periods equals $\log \Omega_E$, under the hypotheses on the periods and the field stated there (its identification with the periods of the input (S4) rests on two further statements of the source and one that is proved there). The eleven inputs are quoted, among them the inner-product formula, the covolumes and the local index bounds
The inequality (M), the trace-formula estimates (§III.8)	proved: the dimension bound of Part II, §6.9 follows from three statements of the source (on traces, ranks and pairings) and the size and subspace conditions of Corollary T2; a finite model of the trace estimates. The three statements, and the theorems behind them, are quoted

Statement of Parts I and II	Status in Part III
The inequality (M), the bound for Galois cohomology (§III.9)	proved: the step from the bound to the bound for denominators, and vanishing in some cases over $\mathbb{Q}$ from cited theorems. The bound itself is quoted, with an account of its proof in the source
The inequality (M), the rest: the construction of the maps and weights; the identification of the abstract objects of §III.4–III.9 with those of the construction; the assembly of (M)	not yet begun; quoted

So at present Parts I–III prove the Riemann hypothesis from the inequality (M) and from the cited theorems, and (M) is quoted. Of what (M) needs, §III.4–III.9 prove the parts stated in the table, conditionally as stated there.

The statements below are translated from the Lean development of the announcement. In Part III "the source" always means this Lean development. For §III.1–III.3 each is a statement that the final theorem actually depends on (for §III.4–III.9 see the remarks in those sections): we computed the set of all declarations on which the Lean theorem `riemannHypothesis` depends, and checked that these belong to it.

For §III.1–III.3 the check was also made for the lemmas of the Lean proofs that the sections follow (for instance `QH.thinH`, `twistK`, `meanSqK`, `bad_share`, `tailK`, `count_bad`, `hostThickExists`, `topsChoiceH`): all are among the declarations the final theorem depends on.

A correction to the source references of Part II. The same computation shows that several statements cited in Part II are present in the archive but are *not* among the 1,478,198 declarations on which `riemannHypothesis` depends: the Fock norm and jet definitions of [23]; the statements `ChainInt` of [14], `FrameLA` and `FrameLast` of [20], and `CountIdentity` of [17] (a statement of the same name elsewhere is used); the definitions concerning Behrend sets in [17]; and the inequality of [6] in the form quoted. These are parallel or earlier formulations. The same holds for the linearity statement of [13]. The mathematics of Part II does not depend on them, since what Part II proves it proves directly; but for these items the phrase "the source's statement" should be read as "a statement in the archive". The statements that are used include: the Landau lemma on intervals [1], the auxiliary primes [9, 29], the weighted sums `ASK` and `MAIN` [3], the height of a map [25], the Chowla–Selberg period [16], the minors congruence [15], the digit inequality `StarArith` [21], the layer-cake formula, the Bessel identity and the local comparisons of [19], the bound for each translate `each_unit_bound_statement` of [28] with the hypotheses defined in [7] (the averaged statement `MeanBoundSecondForm` of [7] is not used), and the bound for Galois cohomology [8].

III.1 The auxiliary primes: proof of Estimate B

For a primitive Dirichlet character $\chi \neq 1$ put, as in Part I,

$$G_\chi(y) = \sum_{n \le y} \Lambda(n)\chi(n)\Big(\frac{y}{n} - 1\Big) + y\,\frac{L'}{L}(1,\chi). \tag{III.1}$$

For an odd prime p write χ_p for the Legendre symbol modulo p.

Theorem III.1.1 (auxiliary primes; `QH.pairEx` in [9], statement in [29]). *Let $M \geq 1$ be an integer, $0 < \varepsilon < \frac{1}{52}$ and $A \geq 1$. There is Y_1 such that for every $Y \geq Y_1$ there are primes q and d with the following properties.*

(a) $q \equiv 3 \bmod 8$ *and* $\frac{1}{2}Y^{\varepsilon/5} < \log q \leq Y^{\varepsilon/5}$.

(b) $d \equiv 7 \bmod 16$ *and* $\frac{1}{2}Y^{\varepsilon/5}(\log Y)^2 < \log d \leq Y^{\varepsilon/5}(\log Y)^2$.

(c) *q divides $d+1$, and so does every odd prime $p \leq Y^{\varepsilon/5}$. Neither q nor d divides M.*

(d) *Let ξ be a Dirichlet character modulo M and let $(i,j) \in \{0,1\}^2$, $(i,j) \neq (0,0)$. Let χ be the primitive character that induces $\xi\chi_q^i\chi_d^j$. Then for all y with $Y \leq y \leq Y^A$*

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

(e) *Let ρ be a real primitive character whose conductor divides $8M$, let $(i,j) \neq (0,0)$, and let χ be the primitive character that induces $\rho\chi_q^i\chi_d^j$. Then $L(s,\chi) \neq 0$ in the region*

$$\operatorname{Re} s > 1 - \tfrac{1}{400}, \qquad |\operatorname{Im} s| \leq 8M \exp\bigl(Y^{\varepsilon/5}\bigr) \exp\bigl(j\, Y^{\varepsilon/5}(\log Y)^2\bigr). \tag{III.3}$$

Remarks. (1) For ζ one takes $M = 1$ and $A \geq \frac{3}{2}$. Then (d) gives Estimate B of Part I, with $x = y$, the interval $[Y, Y^{3/2}]$ contained in $[Y, Y^A]$, and the exponent $\frac{1}{2} + \frac{\varepsilon}{2}$ in place of $\frac{1}{2} + \varepsilon$.

(2) The conductors are far larger than y: $\log q$ is a power of Y. So (d) cannot be a statement about a fixed character as $y \to \infty$. It is a statement about well-chosen q and d.

(3) Condition (c) has a consequence that is not visible in Part I. If p is an odd prime with $p \leq Y^{\varepsilon/5}$, then $d \equiv -1 \bmod p$, and since $d \equiv 3 \bmod 4$ quadratic reciprocity gives $\chi_d(p) = \left(\frac{-1}{p}\right)(-1)^{(p-1)/2} = 1$. Also $\chi_d(2) = 1$. So *every prime up to $Y^{\varepsilon/5}$ splits in $\mathbb{Q}(\sqrt{-d})$*; the inert primes, which carry the argument of Part I, all exceed $Y^{\varepsilon/5}$.

(4) Part (e) is used for Estimate C: it excludes exceptional real zeros for the real characters whose conductor is divisible by q or d.

Proof of Theorem III.1.1

The proof chooses q at random from a set of primes, then d at random from a second set, and shows that most choices work. Two things must hold for the chosen primes: a sum over $n \leq Z$ must be small (this comes from a mean square over the primes), and a tail $n > Z$ must be small (this comes from a zero-free box, which holds for all but few primes by a zero-density theorem). Throughout, $\tau = \varepsilon/2$,

$$Z = Y^{400A}, \qquad z = Y^{\varepsilon/5}, \qquad z' = Y^{\varepsilon/5}(\log Y)^2, \tag{III.4}$$

and Y is large in terms of M, ε, A. Constants C, c are absolute, or depend only on A, unless indicated.

Cited theorems.

(Q1) *Zero-free region, with uniqueness* (Landau, Page; [Dav00, Ch. 14], [Pag35]; in this explicit form [TZ21, §1]). Let $c_1 = \frac{1}{13}$. For every modulus $m \geq 2$, the functions $L(s,\chi)$, χ modulo m, have no zero with $\beta \geq 1 - c_1/\log(m(|\gamma|+3))$, except possibly one simple real zero β_1 of one real nonprincipal character χ_1. Moreover $\beta_1 \leq 1 - c/m$; this follows from the bound $\beta_1 \leq 1 - c'm^{-1/2}$ recorded in [TZ21, §1]. This is the exceptional zero meant in (Q2).

(Q2) *Primes in progressions to large moduli* (Thorner and Zaman [TZ21, Corollary 1.4]; results of this type go back to Linnik and to Gallagher [Gal70], see also [Bom87]). There is an absolute $c_3 > 0$ such that for coprime $m \geq 2$ and b, and $t \geq m^{12}$,

$$\sum_{p \leq t,\ p \equiv b\,(m)} \log p = \Lambda_b(t)(1 + O(\kappa(t))), \qquad \Lambda_b(t) = \frac{1}{\varphi(m)}\Big(t - \chi_1(b)\frac{t^{\beta_1}}{\beta_1}\Big), \tag{III.5}$$

with $\kappa(t) = \exp(-c_3 \log t/\log m) + \exp(-c_3(\log t)^{3/5}(\log\log t)^{-1/5})$. The term with χ_1 is present only if the exceptional zero of (Q1) exists for the modulus m.

(Q3) *Zero density* (Montgomery [Mon71, Theorem 12.2]). The number of primitive characters χ of conductor at most Q for which $L(s,\chi)$ has a zero with $\beta > 1 - \frac{1}{400}$, $|\gamma| \leq Q$, is at most $CQ^{1/4}$.

(Q4) *The logarithmic derivative in a zero-free box* (a standard consequence of Landau's lemma [Tit86, §3.9], [MV07, Lemma 6.3], and the convexity bound for $L(s,\chi)$). Let $\chi \neq 1$ be primitive of conductor at most Q, and suppose $L(s,\chi) \neq 0$ for $\operatorname{Re} s > 1 - \frac{1}{400}$, $|\operatorname{Im} s| \leq Q$. Then for each $\delta > 0$ there is K_δ with $|\frac{L'}{L}(s,\chi)| \leq K_\delta \log Q$ for $\operatorname{Re} s \geq 1 - \frac{1}{400} + \delta$, $|\operatorname{Im} s| \leq Q - 1$.

The form of (Q2) that is used. Let $U \geq m^{24}$, and let $\pi^*(U; m, b)$ be the number of primes $p \equiv b \bmod m$ with $U^{1/2} < p \leq U$. Then

$$|\pi^*(U; m, b) - \mathcal{M}_b| \leq k\,\mathcal{M}_b, \qquad \mathcal{M}_b = \frac{1}{\varphi(m)} \int_{U^{1/2}}^{U} (1 - \chi_1(b) t^{\beta_1 - 1}) \frac{dt}{\log t}, \tag{III.6}$$

with $k = A_0(e^{-c_4 \log U/\log m} + e^{-\sqrt{c_4 \log U}})$. *Proof.* On $[U^{1/2}, U]$ we have $t \geq m^{12}$, and $(\log t)^{3/5}(\log\log t)^{-1/5} \geq \frac{1}{2}\sqrt{\log U}$, so that the relative error in (Q2), with its implied constant, is at most $k/20$ for suitable A_0, c_4. Write $\theta(t)$ for the sum in (Q2). By partial summation $\pi^* = \theta(U)/\log U - \theta(U^{1/2})/\log U^{1/2} + \int_{U^{1/2}}^{U} \theta(t)\,dt/(t \log^2 t)$. Replacing θ by Λ_b gives exactly $\mathcal{M}_b$, because $\Lambda_b'(t) = (1 - \chi_1(b) t^{\beta_1 - 1})/\varphi(m)$. The function Λ_b is positive and increasing, so the error is at most $\frac{k}{20}\Lambda_b(U)(\frac{1}{\log U} + \frac{2}{\log U} + \frac{2}{\log U})$. Finally $\Lambda_b(U) \leq 4\int_{U/2}^{U} \Lambda_b' \leq 4 \log U \cdot \mathcal{M}_b$: if $\chi_1(b) = 1$ the derivative Λ_b' is increasing, and otherwise it lies between $1/\varphi(m)$ and $2/\varphi(m)$. □

The Lean development proves all four. For (Q2) and (Q3) it has its own proofs: (Q2) from a zero-density estimate and the Deuring–Heilbronn phenomenon, (Q3) by a mollifier and the Halász–Montgomery inequality. In (Q3) the published theorem is much stronger than what is used.

The sets of primes. Let m_0 be a positive integer divisible by 8, a_0 an integer prime to m_0, and $U > 1$. Put

$$P = P(m_0, a_0, U) = \{q \text{ prime} :\ U^{1/2} < q \leq U,\ q \equiv a_0 \bmod m_0\}. \tag{III.7}$$

Call (m_0, a_0, U) *admissible* if

$$\tfrac{1}{4}(\log Y)^2 \log m_0 \leq \log U, \qquad Y^{\varepsilon/6} \leq \log U \leq Y^{\varepsilon/4}, \tag{III.8}$$

and every prime factor of m_0 is either at most $Y^{1/4}$ or greater than Z^2.

Step 1: characters over the primes of P. *Let (m_0, a_0, U) be admissible, let $B > 0$, and let Y be large in terms of B as well. Then $|P| \geq U^{9/10}$. Moreover, among the odd squarefree s with $1 < s \leq Z^2$ and $(s, m_0) = 1$, at most one fails to satisfy the following bound; if there is one, call it n^*:*

$$\Big|\sum_{q \in P} \Big(\frac{q}{s}\Big)\Big| \leq Y^{-B}|P|. \tag{III.9}$$

Proof. Apply (III.6) with modulus $m = m_0 s$. By (III.8), $\log m \le \log m_0 + 800A\log Y$ and $\log U/\log m \ge c(\log Y)^2$ (if $\log m_0 \le \log Y$ use $\log U \ge Y^{\varepsilon/6}$). So $U \ge m^{24}$ and $k \le Y^{-B-1}$. Every $q \in P$ exceeds $Z^2 \ge s$, since $\log q > \frac{1}{2}Y^{\varepsilon/6}$, so $q \nmid s$. Split P into residue classes b modulo $m_0 s$ with $b \equiv a_0 \bmod m_0$, $(b,s)=1$; the symbol $(\frac{b}{s})$ is constant on each, and the class contains $\pi^*(U;m,b)$ elements of P. So

$$\sum_{q\in P}\Big(\frac{q}{s}\Big) = \sum_b \Big(\frac{b}{s}\Big)\mathcal{M}_b + O\Big(k\sum_b \mathcal{M}_b\Big). \tag{III.10}$$

The error. Since $k \le \frac{1}{2}$, (III.6) gives $\mathcal{M}_b \le 2\pi^*(U;m,b)$, so $\sum_b \mathcal{M}_b \le 2|P|$. (This is why a relative error is needed: P may be small if the modulus m_0 itself has an exceptional zero.)

The main terms. The part of $\mathcal{M}_b$ without χ_1 is the same for all b, and $b \mapsto (\frac{b}{s})$ is a nontrivial character on the classes $b \equiv a_0 \bmod m_0$ (s is squarefree, odd, prime to m_0 and greater than 1). So these parts cancel. The parts with χ_1 cancel in the same way unless $\chi_1(b)(\frac{b}{s})$ is constant on those classes. In that case call s *bad* if in addition $\beta_1 > 1 - \frac{c_1}{3\log(m_0 Z^2)}$.

At most one s is bad. If s_1, s_2 are both bad, apply (Q1) to the modulus $m_0 s_1 s_2$; since $\log(3m_0 s_1 s_2) \le 3\log(m_0 Z^2)$, both exceptional zeros lie in its region, so the two exceptional characters are induced by the same character. Hence $(\frac{u}{s_1}) = (\frac{u}{s_2})$ for $u \equiv 1 \bmod m_0$ prime to $s_1 s_2$, and $s_1 = s_2$. So at most one $s = n^*$ is bad.

The other s. If s is not bad and the parts with χ_1 do not cancel, then $t^{\beta_1 - 1} \le U^{-(1-\beta_1)/2} \le U^{-c_1/(6\log(m_0 Z^2))} \le Y^{-B-1}$ on the range of integration, so the part with χ_1 is at most Y^{-B-1} times the part without it, hence at most $2Y^{-B-1}\mathcal{M}_b$. Altogether $|\sum_{q\in P}(\frac{q}{s})| \le (2k + 4Y^{-B-1})|P| \le Y^{-B}|P|$.

The size of P. Apply (III.6) with $m = m_0$, $b = a_0$: $|P| \ge \frac{1}{2}\mathcal{M}_{a_0}$. By (Q1), $1 - \beta_1 \ge c/m_0$, so $1 - t^{\beta_1 - 1} \ge c'/m_0$ for $t \ge U/2$, and $\mathcal{M}_{a_0} \ge c'U/(2m_0^2 \log U)$. Since $\log m_0 \le 4\log U/(\log Y)^2$, this exceeds $2U^{9/10}$. □

Step 2: the mean square. For $y \ge 1$, an odd prime q and a function κ with $|\kappa(n)| \le 1$ put

$$S(y,q,\kappa) = \sum_{n\le Z} \Lambda(n)\min\Big(1,\frac{y}{n}\Big)\kappa(n)\Big(\frac{n}{q}\Big). \tag{III.11}$$

Let (m_0, a_0, U) be admissible. Then for $Y \le y \le Y^A$ and every such κ,

$$\sum_{q\in P} |S(y,q,\kappa)|^2 \le C\, y\log y\, |P|. \tag{III.12}$$

Proof. Write $c_n = \Lambda(n)\min(1, y/n)$. Expanding the square,

$$\sum_{q\in P}|S|^2 \le \sum_{n,n'\le Z} c_n c_{n'}\Big|\sum_{q\in P}\Big(\frac{nn'}{q}\Big)\Big|. \tag{III.13}$$

Every $q \in P$ exceeds $Z^2 \ge nn'$, since $\log q > \frac{1}{2}Y^{\varepsilon/6}$. Let s be the product of the primes that do not divide m_0 and occur in nn' to an odd power. All $q \in P$ lie in one residue class modulo m_0, and $8 \mid m_0$; so by quadratic reciprocity and its supplements $(\frac{nn'}{q}) = \pm(\frac{q}{s})$ with a sign that does not depend on q. By Step 1 the inner sum is at most $Y^{-A-1}|P|$ unless $s = 1$ or $s = n^*$, when we bound it by $|P|$.

The terms n, n' are prime powers. Put $\sigma(p) = \sum_{p^j \le Z} c_{p^j}$. One checks $\sigma(p) \le 3\log y$, and $\sigma(p) \le 2y\log p/p$ for $p > y$, whence $\sum_p \sigma(p)^2 \le Cy\log y$ by Chebyshev's bound. Call n *neutral* if it is an even power of a prime or a power of a prime dividing m_0; let R be the sum of c_n over such

$n \le Z$. Even powers contribute at most $C\sqrt{y}$. The primes dividing m_0 that are at most Z are at most $Y^{1/4}$, so there are at most $Y^{1/4}$ of them, each contributing at most $3\log y$. So $R \le C\sqrt{y}$.

Pairs with $s = 1$. Either n, n' are powers of the same prime, or both are neutral. Their contribution is at most $(\sum_p \sigma(p)^2 + R^2)|P| \le Cy\log y\,|P|$.

Pairs with $s = n^*$. Since n, n' are prime powers, such pairs exist only if n^* is a prime or a product of two primes; then one of n, n' is a power of a fixed prime while the other is a power of a second fixed prime or is neutral. The contribution is at most $(2\cdot 9\log^2 y + 6\log y \cdot R)|P| \le Cy|P|$.

All other pairs. Since $\sum_{n\le Z} c_n \le y\sum_{n\le Z}\Lambda(n)/n \le Cy\log Z$, they contribute at most $Y^{-A-1}(Cy\cdot 400A\log Y)^2|P| \le C'y|P|$, because $y \le Y^A$. □

Step 3: all y at once. *Let (m_0, a_0, U) be admissible and κ as before. The number of $q \in P$ for which $|S(y,q,\kappa)| > \frac12 y^{1/2+\tau}$ for some $y \in [Y, Y^A]$ is at most $C_\tau Y^{-\tau}|P|$.*

Proof. As a function of y, S is continuous and piecewise linear, with derivative $D(y) = \sum_{y<n\le Z}\Lambda(n)\kappa(n)(\frac{n}{q})/n$. The point is that $y\,D(y) = S(y,q,\kappa_y)$, where κ_y is κ restricted to $n > y$. So (III.12), which is uniform in κ, gives also $\sum_q |D(y)|^2 \le C\log y\,|P|/y$. On a block $I = [W, \min(2W, Y^A)]$ with $Y \le W$, by the Cauchy–Schwarz inequality, $|S(y)|^2 \le 2|S(W)|^2 + 2W\int_I |D|^2$, so $\sum_q \max_{y\in I}|S(y)|^2 \le CW\log W\,|P|$. By Chebyshev's inequality the number of q with $\max|S| > \frac12 W^{1/2+\tau}$ on the block is at most $CW^{-2\tau}\log W\,|P| \le (C/\tau)W^{-\tau}|P|$. Sum over the blocks $W = 2^i Y$. □

Step 4: the tail. For a primitive $\chi \ne 1$ and $w \ge 1$ put

$$t_\chi(w) = -\frac{L'}{L}(1,\chi) - \sum_{n\le w}\frac{\Lambda(n)\chi(n)}{n}, \tag{III.14}$$

the tail of the series for $-\frac{L'}{L}(1,\chi)$. *Let χ have conductor at most Q and suppose $L(s,\chi) \ne 0$ for $\operatorname{Re} s > 1 - \frac{1}{400}$, $|\operatorname{Im} s| \le Q$. Then for $Q \ge 2$, $w \ge 2$, $0 < \eta \le 1$ and $0 < \delta \le \frac{1}{800}$,*

$$|t_\chi(w)| \le K'_\delta\Big((\log Q)^2 w^{-1/400+\delta} + \frac{w\log Q}{\eta\, Q}\Big) + \frac{2(w\eta+1)\log(2w)}{w}. \tag{III.15}$$

Proof. Let G_χ be as above. For $h = w\eta$ one has directly from the definitions

$$G_\chi(w+h) - G_\chi(w) = -h\,t_\chi(w) + \sum_{w<n\le w+h}\Lambda(n)\chi(n)\Big(\frac{w+h}{n} - 1\Big), \tag{III.16}$$

and the last sum is at most $h(h+1)\log(2w)/w$ in absolute value. On the other hand, by Perron's formula with the kernel $y^s/(s(s-1))$,

$$G_\chi(y) - y\frac{L'}{L}(1,\chi) = \frac{1}{2\pi i}\int_{\operatorname{Re} s=2} -\frac{L'}{L}(s,\chi)\frac{y^s}{s(s-1)}ds. \tag{III.17}$$

Move the part $|\operatorname{Im} s| \le Q/2$ of the contour to the line $\operatorname{Re} s = \sigma_1 = 1 - \frac{1}{400} + \delta$. No zero is crossed. The pole at $s = 1$ has residue $-y\frac{L'}{L}(1,\chi)$, which cancels the term on the left. So $G_\chi(y)$ is the integral over the new contour. Take the difference at $y = w+h$ and $y = w$. On the line $\operatorname{Re} s = \sigma_1$ we have $|(w+h)^s - w^s| \le 4\eta|s|w^{\sigma_1}$, which cancels one factor $|s|$, and (Q4) bounds the logarithmic derivative; this part is at most $K_\delta\log Q\cdot\eta\, w^{\sigma_1}\log Q$. The horizontal segments and the remaining parts of the line $\operatorname{Re} s = 2$ contribute at most $K_\delta\log Q\cdot w^2/Q$. Divide by $h = w\eta$. □

Two consequences. Let χ be as in Step 4, with $Y^{\varepsilon/6} \le \log Q \le Y^{\varepsilon/4}$.

(a) $Y^A|t_\chi(Z)| \le Y^\varepsilon$. Take $w = Z = Y^{400A}$, $\eta = Y^{-2A}$, $\delta = \varepsilon/(1600A)$ in (III.15). The first term gives $Y^A \cdot Y^{\varepsilon/2} \cdot Y^{-A+\varepsilon/4}$; the second is at most $Y^{404A}e^{-Y^{\varepsilon/6}}$; the third is at most $801A\,Y^{-A}\log Y$. This is where the numbers 400 and $\frac{1}{400}$ are matched: $Z^{-1/400} = Y^{-A}$ pays for the factor $y \le Y^A$.

(b) *If* $\log Q \le 3z'$ *and* $\varepsilon < \frac{1}{52}$, *then* $|\frac{L'}{L}(1,\chi)| \le 4\log Y$. Take $w = (\log Q)^{1000}$, $\eta = (\log Q)^{-5}$ and $\delta = \frac{1}{4000}$, so that $(\log Q)^2 w^{-1/400+\delta} = (\log Q)^{-1/4}$; then (III.15) gives $|t_\chi(w)| \le C$, and $|\sum_{n\le w}\Lambda(n)\chi(n)/n| \le \log w + C = 1000\log\log Q + C$. Now $\log\log Q \le \frac{\varepsilon}{5}\log Y + 2\log\log Y + \log 3$. So $|\frac{L'}{L}(1,\chi)| \le 200\varepsilon\log Y + 2000\log\log Y + C$, and $200\varepsilon < \frac{50}{13}$. This step needs $\varepsilon < \frac{1}{50}$; it is the one that comes closest to using the hypothesis $\varepsilon < \frac{1}{52}$.

Step 5: from S and the tail to G_χ. For $y \le Z$, and $\chi(n) = \kappa(n)(\frac{n}{q})$ for $n \le Z$,

$$G_\chi(y) = -S(y,q,\kappa) - y\,t_\chi(Z). \tag{III.18}$$

Indeed

$$-S = -\sum_{n\le y}\Lambda(n)\chi(n) - y\sum_{y<n\le Z}\frac{\Lambda(n)\chi(n)}{n}, \qquad -y\,t_\chi(Z) = y\frac{L'}{L}(1,\chi) + y\sum_{n\le Z}\frac{\Lambda(n)\chi(n)}{n}. \tag{III.19}$$

Step 6: most primes of P are good. *Let (m_0,a_0,U) be admissible, let $1 \le F \le U^{1/2}$ with $\log(FU) \le Y^{\varepsilon/4}$, and let Ξ be a set of at most R primitive characters of conductor at most F. Then P is not empty, and for all $q \in P$ outside a set of at most $\frac{1}{2}|P|$ elements (for Y large in terms of R), every $\xi \in \Xi$ has the following properties. The character $\chi = \xi\chi_q$ is primitive and nontrivial of conductor at most FU; $L(s,\chi) \ne 0$ for $\operatorname{Re} s > 1 - \frac{1}{400}$, $|\operatorname{Im} s| \le FU$; and $|G_\chi(y)| \le y^{1/2+\tau}$ for $Y \le y \le Y^A$.*

Proof. Since $q > U^{1/2} \ge F$, the prime q does not divide the conductor of ξ, so $\xi\chi_q$ is primitive, and q is determined by it. Hence by (Q3) the number of $q \in P$ for which some primitive character of the form $\xi'\chi_q$, with ξ' of conductor at most F, has a zero in the box of height FU is at most $C(FU)^{1/4} \le CU^{3/8}$, which is at most $\frac{1}{4}U^{9/10} \le \frac{1}{4}|P|$. By Step 3 with $\kappa = \xi$, used for each $\xi \in \Xi$, the number of q with $|S(y,q,\xi)| > \frac{1}{2}y^{1/2+\tau}$ for some y and some ξ is at most $RC_\tau Y^{-\tau}|P| \le \frac{1}{4}|P|$. For the remaining q, (III.18) and consequence (a) of Step 4 with $Q = FU$ give $|G_\chi(y)| \le \frac{1}{2}y^{1/2+\tau} + Y^\varepsilon \le y^{1/2+\tau}$. □

Step 7: the choice of q and d.

The prime q. Take $m_0 = 8$, $a_0 = 3$, $U = e^z$. This is admissible. Take $F = 8M$ and for Ξ all primitive characters of conductor dividing $8M$, the trivial one included. Step 6 gives a prime $q \equiv 3 \bmod 8$ with $\frac{z}{2} < \log q \le z$ such that every $\xi\chi_q$ has the three properties, with box height $8Me^z$.

The prime d. Now take $m_0 = 16\,q\prod_{3\le p\le z}p$ and let a_0 be the class with $a_0 \equiv 7 \bmod 16$ and $a_0 \equiv -1$ modulo q and modulo every odd prime $p \le z$. Take $U = e^{z'}$. Then $\log m_0 \le 4z = 4z'/(\log Y)^2$, and the prime factors of m_0 are at most $z \le Y^{1/4}$, or equal to q, which exceeds Z^2. So (m_0,a_0,U) is admissible. Take $F = 8Me^z \le U^{1/2}$ and $\Xi = \{\xi\chi_q^a\}$ with ξ as before and $a \in \{0,1\}$. Step 6 gives a prime $d \equiv 7 \bmod 16$ with $\frac{z'}{2} < \log d \le z'$, with $d \equiv -1$ modulo q and modulo every odd prime up to z, such that every $\xi\chi_q^a\chi_d$ has the three properties, with box height $8Me^{z+z'}$.

Verification. Parts (a)–(c) hold by construction; $q, d > 8M$. For (d): a character ξ modulo M is induced by a primitive character of conductor dividing M, so the primitive character inducing $\xi\chi_q^i\chi_d^j$ is one of those treated, and $|G_\chi(y)| \le y^{1/2+\varepsilon/2}$ on $[Y, Y^A]$. The bound $|\frac{L'}{L}(1,\chi)| \le 4\log Y \le 4\log y$ is consequence (b) of Step 4, with $\log Q \le \log(8M) + z + z' \le 3z'$. Part (e) is the zero-free box. □

Remarks. (1) *There is no circularity.* Nothing of the strength of the Riemann hypothesis is assumed for the characters χ. The bound $y^{1/2+\tau}$ is the mean-square size of a sum of fixed length $Z = Y^{400A}$ over a family of primes q with $\log q \geq \frac{1}{2}Y^{\varepsilon/6}$, far beyond that length. As q varies, the symbols $(\frac{n}{q})$, $n \leq Z$, behave like independent signs; reciprocity turns the statement into one about primes q in progressions to moduli $m_0 s$ much smaller than q, where (Q2) applies. The zero-free box is a weak property of the same kind, and it is proved for all but few q, by (Q3).

(2) Part (e) is proved for every primitive ρ of conductor dividing $8M$, real or not.

(3) The mean value (III.12) is obtained without a large sieve inequality. A reader who knows Heath-Brown's large sieve for real characters [HB95] will recognise the statement; the route above uses primes in progressions.

III.2 The large scale: proof of Estimate C

This section proves Estimate C from three standard theorems on Dirichlet L-functions, which we cite, and from part (e) of Theorem III.1.1. No zero-density estimate and no mean value over characters is used here. In the Lean development this is the theorem `QH.thinH` with the lemmas it calls.

Notation. For a primitive character χ modulo N (the case $N = 1$, $\chi = 1$ included) put

$$\psi_0(u, \chi) = \sum_{n \leq u} \Lambda(n)\chi(n) - \begin{cases} u & \chi = 1 \\ 0 & \chi \neq 1. \end{cases} \tag{III.20}$$

Let G_χ be as in §III.1 for $\chi \neq 1$, and $G_1 = G$. Zeros $\rho = \beta + i\gamma$ are counted with multiplicity.

Theorem III.2.1 (Estimate C; `QH.thinH`)**.** *Let $B \geq 1$, $0 < \varepsilon < \frac{1}{52}$ and $c' > 0$. There is $x_0 = x_0(B, \varepsilon, c')$ such that the following holds for $x \geq x_0$. Let $q \neq d$ be odd primes and Q a positive integer prime to qd, divisible by 8, with odd part at most B, and with $\log(Qqd) \leq 2x^{\varepsilon/5}(\log x)^2$. Suppose that for every real primitive character $\chi \neq 1$ whose conductor divides Qqd and is divisible by q or by d, $L(\beta, \chi) \neq 0$ for real $\beta > 1 - \frac{1}{400}$. Then for every primitive character χ whose conductor divides Qqd,*

$$|G_\chi(y)| \leq y \exp(-c' x^{\varepsilon/8}) \qquad \text{for all } y \geq \tfrac{1}{3}\exp(x^{\varepsilon/3}). \tag{III.21}$$

In the application Q is the power of 2 equal to $4n$ (for ζ), the characters are those of the field E, and q, d are the primes of Theorem III.1.1 with $Y \leq x \leq Y^{3/2}$; then $\log(Qqd) \leq 2x^{\varepsilon/5}(\log x)^2$ by parts (a), (b) of that theorem, and the hypothesis on real zeros is its part (e) with $M = 1$ (see Step 6; for a general Q one takes M divisible by the odd part of Q). In Part I and in formula (5.6) of Part II the constant c' is called κ.

The three theorems cited. Proofs are in Davenport [Dav00, Ch. 14, 16, 19] and Montgomery and Vaughan [MV07]: Theorem 11.3 for (Z1), Theorem 10.17 for (Z2), and Theorems 12.5 and 12.10 with the discussion after Corollary 12.11 for (Z3), which is stated there in a slightly different form. The Lean development proves all three.

(Z1) *Zero-free region* (de la Vallée Poussin, Gronwall, Landau, Titchmarsh). There is an absolute constant $c_1 > 0$ such that, for every $N \geq 1$ and every character χ modulo N, if $L(\rho, \chi) = 0$ and $\beta \geq 1 - c_1/\log(N(|\gamma| + 2))$, then χ is real and nonprincipal and ρ is real. (For $N = 1$ this says that ζ has no zero in the region.)

(Z2) *Zeros in a window.* There is an absolute C_1 such that for every primitive χ modulo N and every real t, the number of zeros with $0 < \beta < 1$ and $|\gamma - t| \leq 1$ is at most $C_1 \log(N(|t| + 2))$.

(Z3) *Truncated explicit formula.* There is an absolute C_2 such that for every primitive χ modulo N and $2 \le T \le u$,

$$\Big|\psi_0(u,\chi) + \sum_{|\gamma|<T} \frac{u^\rho}{\rho}\Big| \le C_2\Big(\frac{u\log^2(NuT)}{T} + \log(Nu) + \sum_{|\gamma|<1}\frac{1}{|\rho|}\Big), \tag{III.22}$$

the sums being over zeros with $0<\beta<1$.

We also use the functional equation, in the form: if χ is primitive and $0<\beta<1$, then $1-\rho$ is a zero of $L(s,\overline{\chi})$; and Dirichlet's theorem $L(1,\chi)\neq 0$.

Step 1: an integral for G_χ. *For primitive χ and $y \ge 1$,*

$$G_\chi(y) = -y\int_y^\infty \psi_0(u,\chi)\,\frac{du}{u^2}. \tag{III.23}$$

Proof. Let $\delta = 1$ if $\chi = 1$ and $\delta = 0$ otherwise, so that $\psi_0(u,\chi) = \psi(u,\chi) - \delta u$ with $\psi(u,\chi) = \sum_{n\le u}\Lambda(n)\chi(n)$. For real $s>1$, integrating the Dirichlet series term by term,

$$H(s) = \int_1^\infty \psi_0(u,\chi)\,u^{-s-1}du = -\frac{1}{s}\frac{L'}{L}(s,\chi) - \frac{\delta}{s-1}. \tag{III.24}$$

By Step 4 below (which does not use the present step), $|\psi_0(u,\chi)|u^{-2}$ is integrable; for this one needs, for a real χ, only some interval $(1-c_\chi,1)$ free of zeros, which exists because $L(1,\chi)\neq 0$. So we may let $s\to 1^+$ under the integral. This gives $H(1) = -\frac{L'}{L}(1,\chi)$ for $\chi\neq 1$, and $H(1) = -1-\gamma$ for $\chi = 1$, from the expansion $-\frac{\zeta'}{\zeta}(s) = \frac{1}{s-1} - \gamma + O(s-1)$. On the other hand, by partial summation,

$$y\int_1^y \psi(u,\chi)\frac{du}{u^2} = \sum_{n\le y}\Lambda(n)\chi(n)\Big(\frac{y}{n}-1\Big). \tag{III.25}$$

Hence $y\int_1^y \psi_0 u^{-2}du$ equals this sum minus $\delta y\log y$. Subtracting $yH(1)$ gives $-y\int_y^\infty \psi_0 u^{-2}du = \sum_{n\le y}\Lambda(n)\chi(n)(y/n-1) - \delta y\log y - yH(1)$, which is $G_\chi(y)$ in both cases. This is also where the constant $1+\gamma$ in the definition of G comes from. □

Step 2: the zero-free region, as used. Let $c_0>0$ and $c = \min(c_1,c_0)$. *Let χ be a character modulo N and $\lambda \ge \max(1,\log N)$. If χ is real and nonprincipal, suppose $L(\beta,\chi)\neq 0$ for $1-c_0<\beta<1$. Then every zero with $\beta<1$ satisfies*

$$\beta \le 1 - \frac{c}{\lambda + \log(|\gamma|+3)}. \tag{III.26}$$

Proof. If not, $\beta > 1 - c_1/\log(N(|\gamma|+2))$, so by (Z1) χ is real and nonprincipal and $\rho=\beta$ is real. But then $\beta > 1 - c/(\lambda+\log 3) \ge 1-c_0$, since $c\le c_0$ and $\lambda\ge 1$; this is against the assumption. □

Zeros are not close to 0. If moreover χ is primitive, a zero with $0<\beta<1$ and $|\gamma|\le 2$ has $\beta \ge c/(3\lambda)$. Indeed $1-\rho$ is a zero of $L(s,\overline{\chi})$, and (III.26) applies to $\overline{\chi}$ (if $\overline{\chi}$ is real it equals χ); so $1-\beta \le 1 - c/(\lambda+\log 5)$, and $\log 5 \le 2\lambda$.

Step 3: the sum of $1/|\rho|$. *Under the assumptions of Step 2 with χ primitive, for $T\ge 1$,*

$$\sum_{|\gamma|<T}\frac{1}{|\rho|} \le C_3(\lambda+\log T)^2, \qquad C_3 = 6C_1\Big(\frac{9}{c}+4\Big). \tag{III.27}$$

Proof. Group the zeros by $i = \lfloor|\gamma|\rfloor$, $0\le i\le T$. By (Z2), applied to the two windows centred at $\pm i$, each group has at most $2C_1\log(N(i+2)) \le 6C_1(\lambda+\log T)$ zeros. For $i\le 1$ (so $|\gamma|<2$) use

$1/|\rho| \le 1/\beta \le 3\lambda/c$; for $i \ge 2$ use $1/|\rho| \le 1/i$. Since $\sum_{i\le T} 1/i \le 1 + \log T$, the sum is at most $6C_1(\lambda + \log T)(6\lambda/c + 1 + \log T)$. □

Step 4: the bound for ψ_0. *Under the same assumptions there are C_4 and $c_2 > 0$, depending only on c_0, such that for $u \ge e$, with $v = \log u$,*

$$|\psi_0(u,\chi)| \le C_4\, u\,(v+\lambda)^2 \exp\Big(-\frac{c_2\, v}{\lambda + \sqrt{v}}\Big). \tag{III.28}$$

Proof. Take $T = \exp(\sqrt{v})$ in (Z3); then $2 \le T \le u$. Put $c_2 = \min(c/3, 1)$ and $\mathcal{E} = \exp(-c_2 v/(\lambda + \sqrt{v}))$. For $|\gamma| < T$ we have $\log(|\gamma| + 3) \le \sqrt{v} + 2$, so by (III.26) $\beta \le 1 - c/(\lambda + \sqrt{v} + 2) \le 1 - c_2/(\lambda + \sqrt{v})$, and $|u^\rho| \le u\mathcal{E}$. By Step 3, $\sum_{|\gamma|<T} |u^\rho/\rho| \le C_3(\lambda + \sqrt{v})^2 u\mathcal{E}$. The error terms of (Z3): $1/T = e^{-\sqrt{v}} \le \mathcal{E}$ because $c_2 v/(\lambda + \sqrt{v}) \le \sqrt{v}$, and $\log(NuT) \le 2(v + \lambda)$, so the first is at most $4(v+\lambda)^2 u\mathcal{E}$. Since $u\mathcal{E} \ge 1$, the second is at most $(v+\lambda)^2 u\mathcal{E}$, and the third at most $C_3\lambda^2 u\mathcal{E}$. □

Step 5: integration. *Let $A > 0$ and $c' > 0$. For large x, if $y \ge \frac{1}{3}\exp(x^{\varepsilon/3})$, $1 \le \lambda \le Ax^{\varepsilon/5}(\log x)^2$, and* (III.28) *holds for $u \ge y$, then $y\int_y^\infty |\psi_0(u,\chi)| u^{-2} du \le y\exp(-c' x^{\varepsilon/8})$.*

Proof. Substitute $u = e^v$. The integral is at most $C_4 \int_V^\infty (v+\lambda)^2 \exp(-c_2 v/(\lambda + \sqrt{v}))\, dv$ with $V = \log y \ge \frac{1}{2}x^{\varepsilon/3}$. For $v \ge \frac{1}{2}x^{\varepsilon/3}$ and large x we have $\lambda \le v^{61/100}$, because $v^{61/100} \ge 2^{-1}x^{\varepsilon/5}x^{\varepsilon/300}$ and $2A(\log x)^2 \le x^{\varepsilon/300}$. Also $\sqrt{v} \le v^{61/100}$ and $\lambda \le v$. Hence the exponent is at least $\frac{c_2}{2}v^{39/100}$, and $(v+\lambda)^2 \le 4v^2$. Now $v^{39/100} \ge \frac{1}{2}x^{13\varepsilon/100}$, so

$$\exp\big(-\tfrac{c_2}{2}v^{39/100}\big) \le \exp\big(-\tfrac{c_2}{8}x^{13\varepsilon/100}\big)\exp\big(-\tfrac{c_2}{4}v^{39/100}\big), \tag{III.29}$$

and $\int_0^\infty v^2 \exp(-\frac{c_2}{4}v^{39/100})dv$ is a finite constant. So the integral is at most a constant times $\exp(-\frac{c_2}{8}x^{13\varepsilon/100})$. Since $\frac{13}{100} > \frac{1}{8}$, this is at most $\exp(-c'x^{\varepsilon/8})$ for large x. □

Where the exponents come from. The classical saving is $\exp(-c\log y/(\log N + \sqrt{\log y}))$. Here $\log N$ is about $x^{\varepsilon/5}$ and $\log y$ about $x^{\varepsilon/3}$, so the saving is $\exp(-x^{\varepsilon/3-\varepsilon/5-o(1)})$, and $\frac{1}{3} - \frac{1}{5} = \frac{2}{15} > \frac{1}{8}$.

Step 6: exceptional zeros. It remains to produce $c_0 > 0$, depending only on B, such that every real primitive $\chi \ne 1$ with conductor $N \mid Qqd$ has no zero in $(1 - c_0, 1)$.

(i) N divisible by q or d. This is the hypothesis of the theorem, with $\frac{1}{400}$.

(ii) N prime to qd. Then $N \mid Q$. The conductor of a real primitive character is not divisible by 16 or by the square of an odd prime. (If $p^2 \mid N$ with p odd, each $u \equiv 1 \bmod N/p$ is a square modulo N: $1 + \frac{N}{p}t \equiv (1 + \frac{N}{p}t\frac{p+1}{2})^2$. So $\chi(u) = 1$ and χ is defined modulo N/p. Similarly $1 + \frac{N}{2}t \equiv (1 + \frac{N}{4}t)^2$ when $16 \mid N$.) So N divides 8 times the odd part of Q, and $N \le 8B$. There are finitely many characters of conductor at most $8B$. For each, $L(1,\chi) \ne 0$ and $L(s,\chi)$ is continuous, so $L(\beta,\chi) \ne 0$ on some interval $(1 - c_\chi, 1]$. Let c_0 be the smallest of the c_χ and of $\frac{1}{400}$. (For ζ only the real characters of conductor 4 and 8 occur. The constant can be made explicit in B by (Q1) of §III.1: a real zero β of a character of conductor $N \le 8B$ has $1 - \beta \ge \min(c_1/\log(3N),\, c/N)$.)

How part (e) of Theorem III.1.1 gives (i). A real primitive character χ of conductor $N \mid Qqd$ divisible by q or d factors as $\rho\,\chi_q^i\chi_d^j$ with $(i,j) \ne (0,0)$ and ρ real primitive of conductor dividing 8 times the odd part of Q. Part (e), on the real axis, says $L(\beta,\chi) \ne 0$ for $\beta > 1 - \frac{1}{400}$.

Conclusion. Let χ be primitive of conductor $N \mid Qqd$, and $\lambda = \max(1, \log N) \le 3x^{\varepsilon/5}(\log x)^2$. By Step 6 the assumption of Step 2 holds with c_0. So (III.28) holds for all $u \ge e$, and Steps 1 and 5 give the theorem. □

Remark. The Lean proof obtains the limits in Step 1, and the convergence of the integral, from the Siegel–Walfisz theorem. The argument above uses only (III.28) for the one character in question.

III.3 The sets of columns T_2 and T

The table of §I.5 has n columns, indexed by the group H of §I.5. It is cyclic of order n; fix an isomorphism with $A = \mathbb{Z}/n\mathbb{Z}$ and write it additively, so that the translates uT_2 of Part II become $u_0 + T_2$. The construction singles out two sets of columns. Both are needed only through the combinatorial properties proved here; how (M) uses them belongs to the part not yet written. Throughout, x is large, $0 < \varepsilon < \frac{1}{52}$, and $x^{2+8\varepsilon} \le n \le 2x^{2+8\varepsilon}$.

The set T_2: a construction of Behrend

Theorem III.3.1 (`hostThickExists` in [27]; `hostThickOrbit`, `thickFrame2`)**.** *Let $L \ge 2$ be an integer with $L \le C_L \log x$. For x large in terms of ε and C_L there is a set T_2 of positive integers with the following properties.*

(a) *$4Ls < n$ for every $s \in T_2$, and $|T_2| \ge x^{-\varepsilon} n$.*

(b) No relation. *If $s_1, s_2, s_3 \in T_2$ are distinct and i, j are integers with $0 < |i|, |j| \le L$, $i \ne j$, then $j(s_2 - s_1) \ne i(s_3 - s_1)$.*

(c) Arithmetic progressions meet translates of T_2 rarely. *Let $u_0, t, h \in \mathbb{Z}/n\mathbb{Z}$. If the additive order of h exceeds $4L$, then among any L consecutive terms $t + ih$ at most two lie in $u_0 + T_2$. If the order of h is at most $4L$, at most one element of $\mathbb{Z}/n\mathbb{Z}$ on the progression lies in $u_0 + T_2$.*

For $i = 1$, $j = 2$ property (b) says that T_2 contains no three-term arithmetic progression. The proof is Behrend's [Beh46]: points on a sphere, written in a large base.

Proof. A sphere. Let $S, D \ge 2$ be integers. The D^S vectors $v \in \{0, \dots, D-1\}^S$ have $|v|^2 < SD^2$, so some sphere $|v|^2 = k$ contains at least D^{S-2}/S of them. Call this set Σ. Three points of a sphere are never collinear. More precisely: *if $v_1, v_2, v_3 \in \Sigma$ and $j(v_2 - v_1) = i(v_3 - v_1)$ with $i, j \ne 0$, $i \ne j$, then $v_2 = v_1$.* Indeed put $b = v_2 - v_1$, $c = v_3 - v_1$. Since $|v_1 + b|^2 = |v_1|^2$ we have $2v_1 \cdot b + |b|^2 = 0$, and likewise for c. Substitute $b = \frac{i}{j} c$ in the first relation and multiply by j^2/i: this gives $2j\, v_1 \cdot c + i|c|^2 = 0$. Subtract j times the second relation: $(i - j)|c|^2 = 0$. So $c = 0$ and then $b = 0$.

Base $B = 4LD$. Map $v \mapsto s(v) = \sum_m v_m B^m$, and let T be the image of Σ. Suppose $j(s(v_2) - s(v_1)) = i(s(v_3) - s(v_1))$ with i, j as in (b). The digits $c_m = j(v_{2,m} - v_{1,m}) - i(v_{3,m} - v_{1,m})$ satisfy $|c_m| \le 2L(D-1) < B$ and $\sum_m c_m B^m = 0$; hence every $c_m = 0$ (look at the lowest nonzero digit modulo B). By the statement about the sphere, $v_2 = v_1$. This is (b) for T.

Parameters. Put $u = \sqrt{\log n \, \log L}$, $S = \lfloor u / \log L \rfloor$, $a = (\log n - \log L - 3)/S$ and $D = \lfloor e^a/(4L) \rfloor$. For large x one has $S \ge 32$, $a \ge 7 + \log L$, hence $D \ge e^a/(8L) \ge 2$, and $a \le 2u$. Then $B^S \le e^{Sa} < n/(4L)$, so $4Ls < n$ on T, and

$$|T| \ge \frac{D^{S-2}}{S} \ge \frac{n}{L} \exp\big(-32\sqrt{\log n \, \log L}\big). \tag{III.30}$$

Since $\log L \le \log\log x + O(1)$, the right side exceeds $4x^{-\varepsilon} n$ for large x. Remove 0 from T to get T_2. This proves (a) and (b).

Property (c). Let the order of h exceed $4L$, and suppose $t + ah$, $t + bh$, $t + ch$ lie in $u_0 + T_2$ with $a < b < c < a + L$: they equal $u_0 + s_a$, $u_0 + s_b$, $u_0 + s_c$. The s's are distinct because $(b - a)h$, $(c - a)h$, $(c - b)h \ne 0$. Eliminating h,

$$X = (c - a)(s_b - s_a) - (b - a)(s_c - s_a) \equiv 0 \bmod n. \tag{III.31}$$

But $|X| < n$, because $4Ls < n$ for each s. So $X = 0$, against (b). If the order m of h is at most $4L$ and two points of the progression are $u_0 + s_1$, $u_0 + s_2$, then $n \mid m(s_1 - s_2)$ and $|m(s_1 - s_2)| < n$, so $s_1 = s_2$. □

The set T: a random choice

Let $\delta_1 = x^{-\varepsilon/8}$. The set $T \subset A$ is chosen so that $|T|$ is about $\delta_1 n$ and so that, for most primes p and most $t \in T$, the points of T on the orbit of t under the Frobenius of p are evenly spread. The precise meaning is this.

Good orbits. Let $n' \geq 1$ and $W \subset \{0, \dots, n'-1\}$, read cyclically. Fix integers $L_2, K, r \geq 1$, and put $u = \lceil 1/\delta_1 \rceil$, $\ell = (6r+1)u$, $M_0 = \lceil (5+8\varepsilon)\, 10^{6r+1} \log x \rceil$, $G' = (M_0+5)\ell$ and $N_1 = \lceil \max(6eG'\delta_1, (5+8\varepsilon)\log_2 x) \rceil + 1$. Cut $\{0, \dots, n'-1\}$ into $\lfloor n'/\ell \rfloor$ consecutive *tiles* of length ℓ (a remainder shorter than ℓ is ignored, and tiles are numbered cyclically), and each tile into $6r+1$ *cells* of length u, numbered $1, \dots, 6r+1$. A tile is *marked* if its odd-numbered cells contain no point of W and each even-numbered cell contains exactly one point of W, at an even position. Call W *good* if

(G1) every cyclic window of length L_2 contains at most K points of W;

(G2) among any $M_0 + 1$ consecutive tiles (cyclically) at least one is marked;

(G3) every cyclic window of length G' contains at most N_1 points of W.

Theorem III.3.2 (`topsGroup`, `topsChoiceH`)**.** *Let $L_b > 0$, $K = \lceil 8(4+8\varepsilon)/\varepsilon \rceil$ and $r = K+1$. There are C_b and x_0 such that for $x \geq x_0$ the following holds. Let $L_2 \leq L_b$. Let $\mathcal{P}$ be a finite set of primes, and for $p \in \mathcal{P}$ let $h_p \in A$ have order $n_p \geq x^{\varepsilon/4}$. For each p and each coset of $\langle h_p \rangle$ fix a base point b on it; for $t \in A$ let $W_p(t,T) = \{0 \leq i < n_p : \ b + ih_p \in T\}$, b being the base point of the coset of t. Then there is $T \subset A$ with*

$$\tfrac{1}{2}\delta_1 n \leq |T| \leq 2\delta_1 n \qquad \text{and} \qquad \sum_{p \in \mathcal{P}} \frac{\log p}{p} \#\{t \in T : \ W_p(t,T) \text{ is not good}\} \leq 4C_b x^{-2}\, \delta_1 n \sum_{p \in \mathcal{P}} \frac{\log p}{p}. \tag{III.32}$$

In the application $\mathcal{P}$ consists of the inert primes below $\exp(x^{\varepsilon/3})$ whose Frobenius has order at least $x^{\varepsilon/4}$ in A, and L_2 is bounded by a constant depending only on ε (the source gives $2(620\, C_o/\varepsilon + 1)$ for primes $p > x^{\varepsilon/(5C_o)}$, with a constant $C_o \geq 2$ that is defined in the source). That L_2 is bounded uses Remark (3) of §III.1: every inert prime exceeds $Y^{\varepsilon/5}$.

Proof. Choose T at random: each element of A belongs to T independently with probability δ_1. (For Dirichlet characters other than the trivial one the source uses two different probabilities, δ_1 and $\frac{3}{4}\delta_1$, on different cosets. For ζ there is one coset.)

Size. $|T|$ has mean $\delta_1 n$ and variance at most $\delta_1 n$.

One orbit. Fix p and t, write $n' = n_p$, and condition on $t \in T$. The other points of the orbit remain independent. We bound the probability that $W = W_p(t,T)$ is not good.

(G1) fails only if some window of length L_2 contains K points other than t. For one window this has probability at most $\binom{L_2}{K}\delta_1^K \leq L_b^K x^{-K\varepsilon/8} \leq L_b^K x^{-(4+8\varepsilon)}$. There are n' windows.

(G2). A given tile not containing t is marked with probability at least

$$\big((1-\delta_1)^u\big)^{3r+1}\Big(\lfloor \tfrac{u}{2} \rfloor \delta_1 (1-\delta_1)^{u-1}\Big)^{3r} \geq 10^{-(6r+1)}, \tag{III.33}$$

since $u\delta_1$ is between 1 and 1.01. Among M_0+1 consecutive tiles at most one contains t; the other M_0 are disjoint, hence independent. The probability that none is marked is at most $(1-10^{-(6r+1)})^{M_0} \le x^{-(5+8\varepsilon)}$. There are at most n' starting tiles. (Here and for (G1), (G3) one needs $L_2 \le n'$ and $G' = (M_0+5)\ell \le n'$, so that windows and runs of tiles do not wrap around; this holds for large x because $n' \ge x^{\varepsilon/4}$, $L_2 \le L_b$ and $\ell \le 2(6r+1)x^{\varepsilon/8}$.)

(G3) fails only if some window of length G' contains N_1 points other than t; then it contains $j = N_1 - 1$ such points. Since $j \ge 6eG'\delta_1$, the probability of the latter is at most $\binom{G'}{j}\delta_1^j \le (eG'\delta_1/j)^j \le 6^{-j}$, which is at most $x^{-(5+8\varepsilon)}$. There are n' windows.

Since $n' \le n \le 2x^{2+8\varepsilon}$, the three bounds add up to at most $C_b x^{-2}$. So the expected number of $t \in T$ with $W_p(t,T)$ not good is at most $C_b x^{-2}\delta_1 n$, for every p.

Choice. We may assume $\mathcal{P}$ is not empty. Let $V = (|T| - \delta_1 n)^2/(2\delta_1 n)$ and let $\mathcal{B}$ be the left side of the inequality of the theorem divided by four times its expected upper bound $C_b x^{-2}\delta_1 n \sum_p \frac{\log p}{p}$. Then $\mathbb{E}(V) \le \frac12$ and $\mathbb{E}(\mathcal{B}) \le \frac14$. So some T has $V + \mathcal{B} \le 1$. For it $\mathcal{B} \le 1$, which is the second assertion, and $\big||T| - \delta_1 n\big| \le \sqrt{2\delta_1 n} \le \frac12 \delta_1 n$, which gives the first. □

III.4 The size of the theta kernels (partly written)

This section concerns the upper bound for the size of the maps: inequality (6.5) of Part II, first line. It is partly written. The following are proved below: the one-variable model; the coefficient estimate in several variables and in several blocks, with the period; the bound for the polynomial on the unit sphere; the bound for one map and for a chain of maps; the bound for a theta function by its mean square; and the hook product. Four passages are conditional: the reference mean square; the eigenvalues of a lift; the comparison, row by row, of the kernel degrees with the hook product; and the norm of the pairing at the end of the chain. The last two are deductions from inputs of the source that are named where they are used. The passage from these estimates to (6.5) is not written. It needs the arithmetic statements listed at the end of the section, none of which is proved in these pages. So the first line of (6.5) is still quoted. We begin with the one-variable model. It shows where the two features of the bound come from: the size $\sqrt{n!}$ and the sign of the term with the period.

The model. *Let F be an entire function with $|F(z)| \le S\, e^{|z|^2/2}$ for all z, where $S > 0$. Then for $n \ge 1$*

$$|F^{(n)}(0)| \le S\, n! \left(\frac{e}{n}\right)^{n/2}, \tag{III.34}$$

so that, if $F^{(n)}(0) \ne 0$,

$$\log |F^{(n)}(0)|^2 \le 2\log S + n\log n - n + O(\log(n+1)). \tag{III.35}$$

Proof. By Cauchy's formula on the circle $|z| = r$, $|F^{(n)}(0)| \le n!\, S\, e^{r^2/2} r^{-n}$. Take $r = \sqrt{n}$. The second form follows from Stirling's formula, $\log n! = n\log n - n + O(\log(n+1))$. □

So a Gaussian bound for the function gives, for the *square* of its nth derivative, the size $n\log n - n$: this is $\log n!$. Compare a function with a pole at finite distance, such as $1/(1-z)$: its nth derivative at 0 is $n!$, and the logarithm of the square is $2\log n!$. That was the situation of the lemniscate sine in §I.3. Theta functions satisfy such a Gaussian bound. This is point (i) of §I.4.2.

The period. If the coordinate is changed by a factor, $z = \lambda w$ with $|\lambda|^2 = 1/\Omega$, then the nth derivative with respect to w is λ^n times that with respect to z, and the bound for the logarithm of

its square becomes

$$n\log n - n - n\log\Omega + 2\log S + O(\log(n+1)). \tag{III.36}$$

This is the shape $a(\log a - 1 - \log\Omega)$ of (6.5), with its sign. The constant relevant there is Ω_E, not the elliptic Ω of §I.7; the two differ by the factor $\mathrm{rd}_E/\sqrt{d}$.

The paragraphs that follow carry the model over to the setting of the proof: several complex variables, a polynomial with values in linear maps in place of a single derivative, several blocks of variables (one for each place), and sums over classes of lattices. Those up to and including the bound for a theta function by its mean square, and the hook product, are complete as stated; the four passages named above are conditional on inputs from the source. The first of them assumes a Gaussian bound S for the function and a bound for the polynomial on the unit sphere; both assumptions are then proved in an abstract setting; for the first, the mean square of one reference theta function is bounded under the inputs (I1)–(I5) listed at the head of that paragraph. That the actual theta kernels and polynomials of the construction are instances of this setting is taken from the source: the hypotheses of the majorant lemma with $v = \mathrm{rd}_E$, the form of the constant S with its factors β and c, and the identification of the polynomial P of the construction with a substitution of the kind treated below.

Several variables: a coefficient estimate on a sphere. Let $F(z) = \sum_\alpha c_\alpha z^\alpha$ be entire on $\mathbb{C}^d$, $d \ge 1$, with $|F(z)| \le S\exp(\|z\|^2/2)$, where $\|z\|^2 = \sum_j |z_j|^2$. Let $P(\bar z) = \sum_{|\alpha|=n} A_\alpha \bar z^\alpha$ be homogeneous, with coefficients in a finite-dimensional normed space, for example a space of linear maps. Here α is a d-tuple of nonnegative integers, $|\alpha| = \sum_j \alpha_j$, $z^\alpha = \prod_j z_j^{\alpha_j}$ and $\alpha! = \prod_j \alpha_j!$. Suppose, for unit vectors u, that the following holds, factors with $\nu_i = 0$ being omitted:

$$\|P(\bar u)\| \le \prod_i s_i(u)^{\nu_i}, \qquad s_i(u) \ge 0, \quad \sum_i s_i(u)^2 = 1, \quad \nu_i \ge 0, \quad \sum_i \nu_i = n. \tag{III.37}$$

Put $h(v) = v\log v - v$ for $v > 0$ and $h(0) = 0$. Then

$$\left\| \sum_{|\alpha|=n} c_\alpha \alpha! A_\alpha \right\| \le S\, e^{1/2}(2n+2)^d \exp\left(\tfrac{1}{2} \sum_i h(\nu_i) \right). \tag{III.38}$$

Proof. Let du be probability measure on the unit sphere in $\mathbb{C}^d$. Independent rotations of the coordinates show that the mean of $u^\beta \bar u^\alpha$ vanishes unless $\alpha = \beta$. Its value for $\alpha = \beta$ is $\alpha!(d-1)!/(n+d-1)!$. Indeed, integrate $|z^\alpha|^2$ against $\pi^{-d}e^{-\|z\|^2}$ first in the separate complex coordinates and then in polar coordinates. The two radial moments needed are $\int_0^\infty t^j e^{-t}\,dt = j!$ for $j = d-1, n+d-1$. Thus for $r > 0$ the sum to be bounded equals

$$\frac{(n+d-1)!}{(d-1)!\, r^n} \int F(ru) P(\bar u)\, du. \tag{III.39}$$

Termwise integration is valid by uniform convergence on the sphere. For $n > 0$ the weighted arithmetic–geometric mean inequality, with weights ν_i/n, gives $\prod_i s_i(u)^{\nu_i} \le \prod_{\nu_i>0} (\nu_i/n)^{\nu_i/2}$. Take $r = \sqrt{n}$. After removing $S\exp(\frac{1}{2}\sum_i h(\nu_i))$, the remaining factor is

$$n!(e/n)^n \binom{n+d-1}{d-1} \le en(n+1)^{d-1} \le e^{1/2}(2n+2)^d. \tag{III.40}$$

The first inequality follows by comparing $\sum_{j<n} \log j$ with $\int_1^n \log t\,dt$ and counting the weak compositions of n into d parts by their first $d-1$ entries. When $n = 0$ the estimate follows at once from $|F(0)| \le S$ and $\|A_0\| \le 1$. □

Several blocks, and the period. For several blocks write $z = (z_\tau)_\tau$. Assume $|F(z)| \le S e^{\|z\|^2/2}$, where $\|z\|^2 = \sum_\tau \|z_\tau\|^2$. Take P homogeneous of degree n_τ in each block and assume the product of the preceding bounds for P on the product of unit spheres. Integrate over that product, using radius $\sqrt{n_\tau}$ in each block of positive degree and setting the variables of degree-zero blocks of F to zero. This gives the product of the displayed factors, with S occurring once. Multiplying all coordinates in block τ of F by a scalar λ_τ multiplies the extracted sum by $\prod_\tau \lambda_\tau^{n_\tau}$. Thus its squared norm acquires the factor $\prod_\tau \Omega_\tau^{-n_\tau}$ if $|\lambda_\tau|^2 = \Omega_\tau^{-1}$.

The second assumption of the estimate, the bound $\|P(\bar{u})\| \le \prod_i s_i(u)^{\nu_i}$ on the unit sphere, is proved in the next two paragraphs. The variables of a block form a matrix; the $s_i(u)$ are its singular values; and $P(\bar{u})$ acts on a space of polynomials by substituting a multiple of the matrix of variables. (Lean: `fock_subst_le` and `blockBound`.)

Where the bound for the polynomial comes from. The numbers s_i are the singular values of the matrix of variables. The integers ν_i are its prescribed column degrees. We prove the estimate directly from the two properties defining the polynomials. Let $\nu_1 \ge \cdots \ge \nu_r \ge 0$ be integers. Write W_ν for the polynomials $f(X)$ on complex r by r matrices which have degree ν_i in column i (every monomial has that degree) and are unchanged when a multiple of an earlier column is added to a later column. Use the squared Gaussian norm

$$\|f\|_G^2 = \pi^{-r^2} \int |f(X)|^2 e^{-\sum_{i,j} |X_{ij}|^2}\, dX. \tag{III.41}$$

Here and below integration on a space of complex matrices uses ordinary Lebesgue measure on the real and imaginary parts of its entries. The same definition, with one factor π^{-1} per entry, applies to rectangular matrices. If A is r by b, with $b \ge r$, let $s_1 \ge \cdots \ge s_r \ge 0$ be its singular values. For a b by r variable matrix Y we claim that

$$\|f(AY)\|_G \le \Big(\prod_{i=1}^{r} s_i^{\nu_i}\Big) \|f\|_G, \qquad \sum_i s_i^2 = \sum_{i,j} |A_{ij}|^2. \tag{III.42}$$

Factors of exponent zero are taken as one.

First consider a monomial occurring in f, and let d_i be its degree in row i. Then

$$\sum_{i \le k} d_i \le \sum_{i \le k} \nu_i \quad (1 \le k < r), \qquad \sum_i d_i = \sum_i \nu_i. \tag{III.43}$$

To see this, multiply the first k rows of X by an indeterminate t. On the dense set where the upper left k by k minor is invertible, subtract combinations of the first k columns from each later column to make its first k entries zero. These operations commute with the row scaling and preserve f. Now t occurs only in the first k columns, so the resulting polynomial in t has degree at most $\sum_{i \le k} \nu_i$. This holds for all X by polynomial identity. Equality for $k = r$ is total homogeneity.

Row substitutions by unitary matrices preserve both the Gaussian norm and the two defining properties of W_ν. Hence the singular value decomposition of A reduces the claim to diagonal row scaling by the s_i, with the unused rows of Y integrated out. For positive s_i, partial summation of the degree inequalities against the decreasing numbers $\log s_i$ gives $\sum_i d_i \log s_i \le \sum_i \nu_i \log s_i$. Thus the multiplier of every monomial is at most $\prod_i s_i^{\nu_i}$. Distinct monomials are orthogonal in the Gaussian norm by coordinate rotations. This proves the claim. Zero singular values follow by continuity, and the sum of their squares is the trace of AA^*.

Why the norm factors cancel. In the application the input degree vector is μ, and the polynomial is multiplied by $\det(X)^e$ before substitution, for an integer $e \ge 0$. This gives precisely

$\nu_i = \mu_i + e$. For a degree vector μ let p_μ be the product of the leading i by i minors raised to $\mu_i - \mu_{i+1}$, where $\mu_{r+1} = 0$. The norm on W_μ is $\|f\|_G / \|p_\mu\|_G$. In this norm multiplication by $\det^e$ is an isometry from W_μ to its image in W_ν.

Indeed, write a nonsingular Gaussian matrix as $X = QR$ by Gram–Schmidt on its columns, with Q unitary and R upper triangular with positive diagonal. The defining column operations give $f(QR) = f(Q) \prod_i R_{ii}^{\mu_i}$. In computing $\|f\|_G^2$, we may first average $|f(UX)|^2$ over unitary U with its invariant probability measure, since the Gaussian measure of X is unchanged. The average of $|f(UQ)|^2$ is the average of $|f(U)|^2$. Hence the squared Gaussian norm is the unitary average of $|f(U)|^2$ times a factor depending only on μ. Singular matrices have measure zero. Dividing by the same formula for p_μ cancels the factor. Since $p_\nu = \det^e p_\mu$ and $|\det U| = 1$, the normalised norm is unchanged by the shift. Padding ν with zeros to length b preserves the norm of p_ν, because it then depends only on the upper left r by r variables. Thus the normalised substitution has norm at most $\prod_i s_i^{\nu_i}$. For independent matrix blocks the maps form a tensor product; its norm is the product of the individual norms, as follows by singular value decomposition. On each unit sphere $\sum_i s_i^2 = 1$. This supplies the polynomial bound used in the coefficient estimate.

Sums over classes, and composing the maps. Fix the number field K over which the matrices of the maps are defined. Let v run over its archimedean places, with one embedding chosen for each conjugate pair, and let s be the number of these places. For $0 \le j \le m$ let V_j be a finite-dimensional vector space over K. At v its complexification has a norm $\|\cdot\|_{j,v}$. Consider K-linear maps $K_j : V_j \to V_{j+1}$ for $0 \le j < m$. We write $\|K_j\|_v$ for the operator norm in these two norms.

Here is the precise information about one map needed for the chain. The classes at step j are partitioned into a_j groups, each containing at most H_j classes. Write I_j for their union and give each class p a weight $0 < w_j(p) \le 1$. At each place a vector x is represented by a tuple $f = (f_p)_{p \in I_j}$ in a complex normed space, with $\|x\|_{j,v}^2 = \sum_p w_j(p) \|f_p\|^2$. Suppose that $K_j x$ is represented in the corresponding tuple space at step $j+1$ by

$$g_q = \sum_{p \in I_j} w_j(p) \epsilon_{pq} T_{pq} f_p \quad (q \in I_{j+1}), \qquad |\epsilon_{pq}| = 1, \qquad \|T_{pq}\|^2 \le b_j(v). \tag{III.44}$$

The phases and the block operators may depend on j and v. Here $b_j(v) \ge 0$ is a squared bound, as supplied by the preceding coefficient estimate. Put $D_j(v) = a_j a_{j+1} H_j H_{j+1} b_j(v)$. Then $\|K_j\|_v^2 \le D_j(v)$. Indeed, Cauchy–Schwarz gives

$$\|g_q\|^2 \le b_j(v) \Big(\sum_p w_j(p) \Big) \sum_p w_j(p) \|f_p\|^2. \tag{III.45}$$

Multiply by $w_{j+1}(q)$ and sum over q. The two sums of weights are at most $|I_j| \le a_j H_j$ and $|I_{j+1}| \le a_{j+1} H_{j+1}$. This proves the bound also when the tuples representing vectors form only a subspace of the space of all tuples.

Let $P = K_{m-1} \cdots K_0$. If $P \ne 0$, define its archimedean height by $h(P) = s^{-1} \sum_v \log \|P\|_v^2$. At each v a nonzero entry of the matrix of P stays nonzero under the embedding of K. Consequently every K_j has positive norm there, and every factor of $D_j(v)$ is positive. Successive application of the operator norm gives

$$\log \|P\|_v^2 \le \sum_{j=0}^{m-1} \log \|K_j\|_v^2 \le \sum_{j=0}^{m-1} \log D_j(v). \tag{III.46}$$

Averaging over v proves

$$h(P) \le \sum_{j=0}^{m-1} \left\{ \frac{1}{s} \sum_{v} \log b_j(v) + \log H_j + \log H_{j+1} + \log a_j + \log a_{j+1} \right\}. \tag{III.47}$$

Thus the contribution of each complex place can be estimated before the logarithms are averaged. Applying this result to the arithmetic maps requires the stated tuple representation with the same norm at the output of one map and the input of the next. The final period pairing, when appended to the chain, requires its own norm bound.

Combining the two. Take for $b_j(v)$ the square of the block estimate, with the period:

$$\log b_j(v) \le 2 \log S_{j,v} + \sum_{\tau} \Big(\sum_{i} h(\nu_{\tau,i}) - n_\tau \log \Omega_\tau + 2 d_\tau \log(2n_\tau + 2) + 1 \Big), \tag{III.48}$$

where d_τ is the number of variables of block τ. Then the height of the composite of m maps is at most the sum over j of the averages over v of these quantities, plus the logarithms of the numbers of classes. Since $\sum_i h(\nu_i) \le h(n) = n \log n - n$ when $\sum_i \nu_i = n$, this has the shape of the first line of (6.5), provided the constants $S_{j,v}$ and the numbers of classes are under control, and provided the periods Ω_τ average to Ω_E.

Where the Gaussian bound comes from. The first assumption above, $|F(z)| \le S\, e^{\|z\|^2/2}$, holds for theta functions because they are quasi-periodic, and the constant S is controlled by a mean square. *Let A be a lattice in $\mathbb{C}^g$ with a fundamental domain D of volume 1, and let $v \ge 1$ be such that $\|x\|^2 \ge 1/v$ for every nonzero $x \in A$. Let G be entire on $\mathbb{C}^g$, and suppose $N(z) = |G(z)|^2 e^{-\pi\|z\|^2}$ is invariant under translation by A. Then for every z*

$$|G(z)|^2 e^{-\pi\|z\|^2} \le 3\,(2gv+1)^g \int_D N(w)\, dw. \tag{III.49}$$

Proof. Let $r^2 = 1/(4v)$ and let B be the ball of radius r about z. Two translates of B by different elements of A are disjoint, since their centres are at distance at least $2r$. As N is periodic, $\int_B N \le \int_D N$. Put $\varphi(u) = G(u) \exp(-\pi \langle u - z, z \rangle)$, where $\langle u, z \rangle = \sum_j u_j \bar{z}_j$; it is entire in u and $\varphi(z) = G(z)$. Now $|\varphi(z)|^2$ is at most the mean of $|\varphi|^2$ over B: expand φ in its Taylor series about z and integrate $|\varphi|^2$ over each sphere centred at z; the mixed terms integrate to zero and the constant term contributes $|\varphi(z)|^2$. The volume of B is $\pi^g r^{2g}/g!$. From $\|u\|^2 = \|z\|^2 + 2\,\mathrm{Re}\langle u - z, z \rangle + \|u - z\|^2$ one gets $|\varphi(u)|^2 e^{-\pi\|z\|^2} = e^{\pi\|u-z\|^2} N(u) \le 3 N(u)$ on B, because $\pi r^2 \le \pi/4 < 1$. Hence

$$|G(z)|^2 e^{-\pi\|z\|^2} \le 3\, g! \Big(\frac{4v}{\pi}\Big)^g \int_B N \le 3\,(2gv+1)^g \int_D N, \tag{III.50}$$

using $g! \le g^g$ and $4/\pi < 2$. □

After the change of variable $z \mapsto z/\sqrt{\pi}$ this is the assumption of the estimate, with $S^2 = 3(2gv+1)^g \int_D N$. In the source (`sup_le_mean_square`, `sup_translate_le`) it is applied to the theta functions attached to a pair of lattices of ranks N, N', with $g = 2nNN'$ and with v the root discriminant of E; so for one such theta function $\log S^2 \le g \log(2g \,\mathrm{rd}_E + 1) + \log 3 + \log \int_D N$.

The constant in the application is larger. The kernel whose coefficients are algebraic integers is not one such theta function but a finite combination of translates of a reference theta function. In the source the constant that enters the block estimate is

$$S = \beta\, c^{2[E:\mathbb{Q}]NN'} \sqrt{3(2g\,\mathrm{rd}_E + 1)^g}\; \Theta_0^{1/2}, \tag{III.51}$$

where c is a modulus, β bounds the coefficients of the combination, and Θ_0 is the mean square of the reference theta function at the given embedding. The factor in front of the square root bounds the sum of the absolute values of the coefficients. The bound for Θ_0, with its arithmetic normalisation, is the subject of the next paragraph. (All the Lean declarations named in this section are among the declarations the final theorem depends on, by the dependency computation: `sup_le_mean_square`, `sup_translate_le`, `fock_subst_le`, `blockBound`, `blockBound_of_fockJet`, `gaussianCoeffIntegral`, `chain_coupled_bound`, `divClausesR_of`; in the passage "Row by row", `weights_rows`, `nu_eq_originRow`, `aPlus_eq_originBig`, `origin_rewrite`, `origin_interlace`, `originRow`, `originBig`, `RowsOriginOff`, `RowsOriginTop`, `Sw`; and in the passage "The pairing at the end of the chain", `brStd`, `brConst`, `brAt`, `BranchMap`, `sum_pair_le`.)

The reference mean square Θ_0. *This paragraph is conditional on the following inputs, which are not proved here.* The notation is explained below the list.

(I1) The singular modulus j is an algebraic integer (classical).

(I2) The conjugates of a curve with complex multiplication by $\mathcal{O}_K$ run equally often through the ideal classes of K.

(I3) The class mean period formula: $\langle \log A_\rho \rangle = \log \pi + \gamma + \log(2\pi/\sqrt{d}) - c_d$. (It follows from Kronecker's limit formula, [Wei76, Ch. VIII, §8], applied to the decomposition of the zeta function of $\mathbb{Q}(\sqrt{-d})$ over ideal classes.)

(I4) The tensor-power description of the normalised reference section of ranks N, N': its squared norm is $\Theta_{0,\rho} = H_\rho^{fNN'}$.

(I5) The reference section at ρ is $\theta/\theta(0)$, for the characteristic whose zero is the marked half-period, with the metric $e^{-2\pi y v^2}$ in the coordinates $z = u + v\tau$.

Conclusion. Under (I1)–(I5),

$$\langle \log \Theta_{0,\rho} \rangle \le \frac{fNN'}{2} \bigl(\log \sqrt{d} - \gamma + c_d \bigr). \tag{III.52}$$

Let $K = \mathbb{Q}(\sqrt{-d})$ have discriminant $-d$, with d prime and $d \equiv 7 \pmod 8$ (these are the hypotheses of the application; the calculation below does not use them). Write χ_d for its quadratic character and $c_d = L'(1,\chi_d)/L(1,\chi_d)$. Let M be a number field of definition for the elliptic curve with complex multiplication by $\mathcal{O}_K$ and its marked point of order two, enlarged if necessary. Average over all $[M:\mathbb{Q}]$ embeddings $\rho : M \hookrightarrow \mathbb{C}$, denoting this average by brackets. (The chain of maps is defined over a field containing M, and $H_{\bar\rho} = H_\rho$; so an average over the archimedean places of that field, as in the bound for the composite, is the same as this average.) The periods at ρ form a lattice $\Lambda = \omega(\mathbb{Z} + \tau\mathbb{Z})$, $y = \Im\tau > 0$. The differential is scaled so that the Weierstrass invariants satisfy $g_2^3 - 27 g_3^2 = 1$. Thus $A_\rho = |\omega|^2 y$ is the area of this particular lattice. The marked point and its conjugates determine which of the three even theta functions to use.

In coordinates $z = u + v\tau$ on the torus, put

$$\theta(z) = \sum_{m \in \mathbb{Z}+\alpha} e^{\pi i m^2 \tau + 2\pi i m (z+\beta)}, \qquad (\alpha,\beta) \in \{(0,0),(0,\tfrac12),(\tfrac12,0)\}. \tag{III.53}$$

Choose the characteristic whose zero is the marked half-period in these coordinates. Write $T = |\theta(0)| > 0$. The normalisation of the section at the origin divides it by $\theta(0)$; this is input (I5). In Jacobi coordinates its squared norm, averaged over the torus with total mass one, is

$$H_\rho = \int_0^1 \int_0^1 \frac{|\theta(u+v\tau)|^2}{T^2} e^{-2\pi y v^2}\, du\, dv = \frac{1}{\sqrt{2y}\, T^2}. \tag{III.54}$$

Indeed, integration in u kills the mixed Fourier terms. The remaining integrals of $e^{-2\pi y(m+v)^2}$ tile the real line. Absolute convergence on the closed parallelogram justifies termwise integration.

The arithmetic normalisation. On the curve $Y^2 = 4x^3 - g_2x - g_3$, let e be the x-coordinate of the marked point, and e', e'' the other two roots of the cubic. Put $a = e - e'$, $b = e - e''$, $P = ab = 3e^2 - g_2/4$, and $X = P^{-3} \in M$. All three roots are distinct. The discriminant identity gives $16P^2(a-b)^2 = 1$, while the sum of the roots gives $g_2 = \frac{4}{3}((a-b)^2 + P)$. Substituting these identities into $j = 1728g_2^3$ yields

$$(X + 16)^3 - jX = 0. \tag{III.55}$$

The singular modulus j is an algebraic integer by complex multiplication. The displayed monic equation therefore makes X a nonzero algebraic integer. This is the only integrality calculation needed here.

For clarity, the classical Jacobi identities used to recover the absolute value of X are as follows. The three half-period differences, up to sign, are $\pi^2\omega^{-2}$ times the fourth powers of the three even theta constants. The two differences through e use the other two constants. The product of all three constants is $2\eta(\tau)^3$, where $\eta(\tau) = e^{\pi i\tau/12}\prod_{n\geq 1}(1 - e^{2\pi in\tau})$. Finally, the discriminant identity $1 = (2\pi/\omega)^{12}\eta(\tau)^{24}$ gives $|\omega| = 2\pi|\eta(\tau)|^2$. Hence at each embedding

$$|\rho X| = (T/|\eta(\tau)|)^{12}, \qquad A_\rho = 4\pi^2 y|\eta(\tau)|^4, \qquad \log H_\rho = \tfrac{1}{2}\log(2\pi^2) - \tfrac{1}{2}\log A_\rho - \tfrac{1}{6}\log|\rho X|. \tag{III.56}$$

Here the same algebraic number X works at every embedding because its formula uses just the marked coordinate e and g_2.

The conjugates of a curve with complex multiplication by $\mathcal{O}_K$ run equally often through the ideal classes of K (input (I2)). By the class mean period formula of §I.7 in the discriminant-one scale (input (I3)), $\langle \log A_\rho \rangle = \log\pi + \gamma + \log(2\pi/\sqrt{d}) - c_d$, where γ is Euler's constant. Since $|N_{M/\mathbb{Q}}X|$ is a positive integer, averaging gives

$$\langle \log H_\rho \rangle = \tfrac{1}{2}(\log\sqrt{d} - \gamma + c_d) - \frac{\log|N_{M/\mathbb{Q}}X|}{6[M:\mathbb{Q}]} \leq \tfrac{1}{2}(\log\sqrt{d} - \gamma + c_d). \tag{III.57}$$

The reference section of ranks N, N' has squared norm $\Theta_{0,\rho} = H_\rho^{fNN'}$, where $f = 2n$ is the number of weight places (n as in Part II). This last identity uses the tensor-power description of the normalised reference section; the integral of a product on independent elliptic coordinates factors. Thus its average logarithm is at most fNN' times the displayed bound. This is the conclusion stated at the head of the paragraph.

The step that uses L_{sc}. This step is separate from the conclusion above. In the application $|c_d| \leq 4\log x$, by Theorem III.1.1(d) with $M = 1$ and $\chi = \chi_d$. Put $v_{\mathrm{host}} = \mathrm{rd}_E/2$, and let L_{sc} be the logarithmic scale of the Lean development; it is defined in the source. In the parameters of the chain one has $\log\sqrt{d} - \gamma \leq \log v_{\mathrm{host}}$ (the inequality $\sqrt{d} \leq v_{\mathrm{host}}$ itself holds for $N \geq 8$ but fails for $N = 4$, $q = 3$) and $\log v_{\mathrm{host}} + |c_d| \leq 2L_{\mathrm{sc}}$; both inequalities are taken from the source. So the contribution is at most $fNN'L_{\mathrm{sc}}$.

The hook product. The hook product enters as an exact Gaussian integral. Given integers $\lambda_1 \geq \cdots \geq \lambda_s \geq 0$, form a diagram with λ_i boxes in row i, left justified. The hook length of a box is one plus the numbers of boxes to its right in the same row and below it in the same column. Let $H(\lambda)$ be the product of all hook lengths, with empty product one. For a square complex matrix X of size s put

$$p_\lambda(X) = \prod_{i=1}^{s} \det(X_{1:i,1:i})^{\lambda_i - \lambda_{i+1}}, \qquad \lambda_{s+1} = 0. \tag{III.58}$$

With the Gaussian probability norm used above, $\|p_\lambda\|_G^2 = H(\lambda)$.

Here is a proof by removing full columns. Write $X = QR$ as in the proof of the polynomial estimate. Conditional on the preceding columns, R_{ii}^2 is the squared length of a standard complex Gaussian vector in dimension $s-i+1$. Its density is $t^{s-i}e^{-t}/(s-i)!$ on $t>0$, by polar integration. This conditional law is constant, so the R_{ii}^2 are independent. Averaging first over unitary matrices on the left as above gives

$$\|p_\lambda\|_G^2 = \left(\int_{U(s)} |p_\lambda(U)|^2\, dU\right) \prod_{i=1}^{s} \frac{(\lambda_i + s - i)!}{(s-i)!}, \tag{III.59}$$

where dU has total mass one. Put $a = \lambda_s$ and $\mu_i = \lambda_i - a$. Then $p_\lambda = (\det X)^a p_\mu$, so the unitary integrals for λ and μ agree. The ratio of their squared Gaussian norms is consequently

$$\prod_{i=1}^{s} \frac{(\lambda_i + s - i)!}{(\lambda_i - a + s - i)!}. \tag{III.60}$$

These are exactly the hook lengths in the first a columns, since the hook at (i,j) there has length $\lambda_i - j + 1 + s - i$. Deleting those columns leaves all other hook lengths unchanged. The last row is now empty, and p_μ depends only on the upper left $(s-1)$ by $(s-1)$ matrix. Induction on s proves the formula, including $a = 0$.

The eigenvalues of a lift: what the inner-product formula contributes. *This paragraph is conditional on the following hypotheses, which are taken from the source and not proved here.* The notation is explained below the list.

(E1) At each weight place the polynomial in the theta lift is a product of two polynomials of the form p_λ on disjoint sets of variables.

(E2) On each non-empty simultaneous Hecke component of the target, the Rallis inner-product formula, evaluated with the chosen local test functions, gives $TT^* = \lambda I$ with the scalar λ of the form displayed below; in it $D_0 \geq 1$, and the level index times the squared norm of the finite test function equals $1/D_S$.

(E3) The parameter conditions: $1 \leq f \leq \varrho$, $1 \leq r \leq \varrho/2$, $t \geq r+5$, $\varrho \geq 8$, $L \geq 5.5$, $\log \varrho \leq L$, and $\log(d_\rho + 2) \leq 2L$.

(E4) The bound for the mass M_0, the bounds for the normalising factors D_S, D_{S^c}, and the formal-degree estimate for J_ρ, as displayed below.

(E5) The local parameter bound $|x| \leq (\mathrm{N}\mathfrak{p})^{(r-t-2)/2}$ for the parameters of the unramified Euler factors.

(E6) The lift used at the distinguished embedding is $S = PT$ with P an orthogonal projection on the target, and the Hecke components span the target.

(E7) At the other embeddings the lower bound of the conclusion holds as well.

Conclusion. Under (E1)–(E7) every positive eigenvalue of SS^* is at least

$$\exp\Big(\sum_\rho \log H(A_\rho^+) - 8f\varrho^2 L\Big). \tag{III.61}$$

The lower bound for these eigenvalues is one of the inputs listed at the end of this section. Fix one step of the chain. At each of the $f = 2n$ weight places ρ, the polynomial in the theta lift is a

product of two polynomials of the form p_λ on disjoint sets of variables, by (E1). Write A_ρ^+ and A_ρ^- for their diagrams; their row lengths are the column degrees of the two products of minors, arranged in decreasing order. Reversing rows or columns in a corner minor leaves its absolute value unchanged after the corresponding change of variables. The Gaussian integrals over disjoint coordinates factor, giving the squared norm $\prod_\rho H(A_\rho^+)H(A_\rho^-)$.

Let T be the full theta lift between its finite-dimensional complex inner-product spaces. On each non-empty simultaneous Hecke component of the target the Rallis inner-product formula, evaluated with the chosen local test functions, gives $TT^* = \lambda I$, by (E2). In the normalisation of the source its scalar has the form

$$\lambda = D_0\, M_0\, \frac{E_0}{D_S D_{S^c}} \prod_\rho H(A_\rho^+)H(A_\rho^-)J_\rho. \tag{III.62}$$

Here $D_0 \geq 1$ is a product of representation dimensions, M_0 is the mass (the sum of reciprocal stabilizer orders), E_0 is the absolute value of the unramified Euler product, and D_S, D_{S^c} are the positive normalising factors at the exceptional finite places and their complement. The level index times the squared norm of the finite test function equals $1/D_S$. The factor J_ρ is the integral of the squared absolute value of a unit vector's matrix coefficient on the real unitary group at ρ, for the measure of the formula; it is defined in the source.

Here are the bounds on these factors that suffice. Here k, m and t are parameters of the construction (in Part II, §6.2: the initial rank k, the number m of lifts, and the odd parameter written τ_0 there), and $f = 2n$ is the number of weight places. Write $r = k + m - 1$, $\varrho = r + t$ for the rank parameter, g_0, g_1 for the two level primes, b for the weight bound, and $L = \log(b\,\mathrm{rd}_E\, g_0 g_1 \varrho)$. (In Part II, §6.6 the letter h denotes $k + m - 1$, which is r here.) By (E3) the parameter conditions give $1 \leq f \leq \varrho$, $1 \leq r \leq \varrho/2$, $t \geq r + 5$, $\varrho \geq 8$, $L \geq 5.5$ and $\log \varrho \leq L$. The mass bound, the finite Euler factor estimates and the formal-degree estimate used in the source give

$$M_0 \geq \frac{1}{(8f^2)^\varrho\, \varrho!}, \qquad E_0 \geq e^{-2rf}, \qquad D_S, D_{S^c} \leq e^{4rf}, \qquad -\log J_\rho \leq 2 + r^2 \log\bigl(2\sqrt{2}\, r^2 \varrho (d_\rho + 2)\bigr), \tag{III.63}$$

where $d_\rho = |A_\rho^+| + |A_\rho^-|$ counts their boxes and $\log(d_\rho + 2) \leq 2L$. The bounds for M_0, D_S, D_{S^c} and J_ρ are (E4). The bound for E_0 follows from (E5): the factors of E_0 are $|1 - x|^{-1}$ with r parameters at each unramified prime ideal $\mathfrak{p}$ of E. Their absolute values satisfy $|x| \leq (\mathrm{N}\mathfrak{p})^{(r-t-2)/2} \leq (\mathrm{N}\mathfrak{p})^{-2}$, by the local parameter bound and $t \geq r + 5$. At most $[E : \mathbb{Q}] = 2f$ prime ideals lie above any rational prime p, and each has norm at least p, so $\sum_\mathfrak{p} (\mathrm{N}\mathfrak{p})^{-2} \leq 2f \sum_p p^{-2} \leq 2f$, the last sum being over the rational primes. Taking logarithms and using $\log|1 - x| \leq |x|$ proves the stated bound for E_0. The logarithmic losses are at most $4\varrho L$ from the mass, $10rf$ from the finite factors, and $f(2 + 6r^2 L)$ from the real integrals. Their sum is at most $8f\varrho^2 L$ by the preceding inequalities. Since every hook length is at least one, this gives

$$\log \lambda \geq \sum_\rho \log H(A_\rho^+) - 8f\varrho^2 L. \tag{III.64}$$

This deduction uses the stated arithmetic and analytic estimates for the test functions.

Finally, by (E6), the lift used at the distinguished embedding is $S = PT$ with P an orthogonal projection on the target, and the Hecke components span the target. Since TT^* is self-adjoint, the scalar bounds on those components give $TT^* \geq cI$, where c is the exponential of the last lower bound. On the range of P one has $\langle SS^* x, x\rangle = \|T^* x\|^2 \geq c\|x\|^2$. This bounds every positive eigenvalue of SS^*. At the other embeddings the same bound is taken from the Lean development; this is (E7).

Row by row: why only the first row survives. *This passage is conditional: it is a deduction from inputs taken from the source, listed below.* It compares the terms $\sum_i h(\nu_i)$ of the block estimate with the hook product of the preceding paragraph, one row at a time. Fix a step $1 \le j \le m$ of the chain (steps are numbered from 1 here, as in the source) and one weight place ρ. Let k and t be the parameters of the preceding paragraph, with $k \ge 2$. The terms used are these.

- The *kernel degrees* are the column degrees of the polynomial of the kernel of this step in the block at this place. We assume, as input (o) from the source, that these are the degrees ν_i to which the block estimate is applied (one block for each weight place); in the Lean development this is the case: the degrees fed to the block estimate are identified with these numbers (`weights_rows`, `nu_eq_originRow`). We call them also the lengths of the kernel rows. There are at least k of them.
- A^+_ρ is the positive-signature diagram of the preceding paragraph, and $H(A^+_\rho)$ its hook product.
- The weight shift w at this step, the weight parameters $u_1, \dots, u_k$ of the domain at this place, and the parameters $v_1, \dots, v_{k-1}$ of the positive-signature rows are as in the source (Lean: `origin_rewrite`, `origin_interlace`, with the formulas `originRow`, `originBig`). We do not define them here.
- A weight place is *ordinary* or *distinguished*, as in the source (Lean: the two cases `RowsOriginOff` and `RowsOriginTop`). At a distinguished place, which we denote by $\mathfrak{t}$, there are two integers $a_\mathfrak{t} \ge 0$ and $b_\mathfrak{t} \ge 0$; $a_\mathfrak{t}$ is the number a_t of the inequality at the end of this section. The index $\mathfrak{t}$ is not the parameter t.
- Δ is the width of the weight interval (in the source, `Sw`).
- B is a bound for the sum of the kernel degrees other than the first k.
- $L_a = \log(a \operatorname{rd}_E)$, with a as in the inequality at the end of this section.

Put

$$\varsigma = \frac{k + (2j-1)(t+1)}{2}, \quad W = w - \varsigma, \quad \delta = (j-1)t, \tag{III.65}$$

$$x_i = u_i - \frac{k+1-2i}{2}, \quad y_i = v_i - \frac{k-2i}{2}. \tag{III.66}$$

Assumed here, from the source:

(a) *The degree formulas.* $W \ge 0$, the first k kernel degrees are given by the formulas for ν_i below, and $k-1$ of the rows of A^+_ρ have the lengths R_i below:

$$\nu_i = W + x_i \quad (1 \le i \le k), \qquad R_i = W + y_i + \delta \quad (1 \le i < k). \tag{III.67}$$

There may be further rows on either side. (Lean: `nu_eq_originRow`, `aPlus_eq_originBig`.)

(b) *Strict interlacing.* $u_i > v_i > u_{i+1}$ for $i < k$.

(c) *Integrality and parity.* The degrees ν_i and R_i are nonnegative integers; $k \ge 2$ is even, $t \ge 1$ is odd, and the weight w is an integer. Then ς, the x_i and the y_i are integers. (The argument below uses only that $x_i - y_i = \nu_i - R_i + \delta$ and $y_i - x_{i+1} = R_i - \delta - \nu_{i+1}$ are integers, and this follows from the integrality of ν_i, R_i and δ.)

(d) *The endpoint bounds.* At an ordinary place, $x_1 \le \Delta$ and $x_k \ge 0$. At a distinguished place, $x_1 = a_\mathfrak{t}$, $x_k = -b_\mathfrak{t}$, $y_{k-1} \ge 0$ and $y_1 \le \Delta$.

(e) *The extra degrees and the sizes.* The sum of the kernel degrees other than the first k is at most B, and $B \le a\,\mathrm{rd}_E$; $w+\Delta \le a\,\mathrm{rd}_E$ at an ordinary place, and $w+a_{\mathfrak{t}} \le a\,\mathrm{rd}_E$ at a distinguished place. The size hypotheses of the source permit $B = m(t+1)\eta$, with η the bound used there for the additional rows.

None of (a)–(e) is proved in these pages.

The comparison of rows. Strict interlacing and integrality give $x_i \ge y_i \ge x_{i+1}$, since both differences are integers greater than $-1/2$. Consequently

$$R_i \ge \nu_{i+1} \quad (1 \le i < k). \tag{III.68}$$

In particular every one of the first k kernel rows except the first has its own row of at least the same length in A_ρ^+.

Recall that $h(v) = v\log v - v$ for $v > 0$ and $h(0) = 0$. For integers $d \ge 0$ one has $h(d) \le \log(d!)$ by integration of $\log x$. The hooks along a row of length R are at least $R, R-1, \dots, 1$. Hence $\log H(A_\rho^+) \ge \sum_{i<k} \log(R_i!) \ge \sum_{i=2}^{k} h(\nu_i)$. Each remaining kernel degree d satisfies $h(d) \le d\,L_a$ (by (e), since $d \le B$), so

$$\sum_{\text{all kernel rows}} h(\nu_i) - \log H(A_\rho^+) \le h(\nu_1) + B\,L_a. \tag{III.69}$$

At an ordinary place $x_1 \le \Delta$ and $w+\Delta \le a\,\mathrm{rd}_E$. Thus the right side is at most $(w+\Delta+B)L_a$. At a distinguished place $x_1 = a_{\mathfrak{t}}$ and $w + a_{\mathfrak{t}} \le a\,\mathrm{rd}_E$. Integrating $h'(v) = \log v$ gives $h(a_{\mathfrak{t}}+W) \le h(a_{\mathfrak{t}}) + WL_a$, also when $a_{\mathfrak{t}} = 0$. The right side is therefore at most $h(a_{\mathfrak{t}}) + (w+B)L_a$. This accounts for the one large factorial term left at each distinguished place.

The degree that multiplies the period. The same comparison explains the period term. At an ordinary place $0 \le x_k \le x_1 \le \Delta$. Here the lower bound $x_k \ge 0$ is the hypothesis in (d); it does not follow from $\nu_k \ge 0$, and the bound below fails without it. At a distinguished place the inner weights satisfy $0 \le y_{k-1} \le y_1 \le \Delta$, hence $0 \le x_i \le \Delta$ for $1 < i < k$; the last endpoint is $x_k = -b_{\mathfrak{t}}$ with $0 \le b_{\mathfrak{t}} \le W$, since $\nu_k \ge 0$. Write $a_* = a_{\mathfrak{t}}$ at a distinguished place and $a_* = 0$ otherwise. Summing the row lengths gives

$$\left| \sum_{\text{all kernel rows}} \nu_i - kw - a_* \right| \le k(\varsigma+\Delta) + w + B. \tag{III.70}$$

Thus the degree that multiplies $-\log\Omega_E$ is $k\,w + a_{\mathfrak{t}}$ at a distinguished place, and $k\,w$ at an ordinary place, up to the displayed error. The part $k\,w$ cancels only through the comparison of periods, item (iii) of the list at the end of this section. Nothing here identifies the period of the block with the period in the covolume.

What this proves, and what it assumes: from (a)–(e) it proves that each of the first k kernel rows after the first is matched by a row of A_ρ^+ of at least the same length, that the terms $\sum_i h(\nu_i)$ exceed $\log H(A_\rho^+)$ by at most $(w+\Delta+B)L_a$ at an ordinary place and $h(a_{\mathfrak{t}}) + (w+B)L_a$ at a distinguished place, and that the total degree is $k\,w + a_*$ up to $k(\varsigma+\Delta) + w + B$; it assumes (o) and (a)–(e), which are statements about the weights of the construction taken from the source and not proved here.

The pairing at the end of the chain. *This passage is conditional: the form of the last map, its level factor and the form of the final functional are inputs from the source, marked below.* Here the rank decreases from $\Psi = \psi + 1$ to ψ, with $\psi \ge 1$. The last map of the chain goes from polynomials of weight λ on Ψ by Ψ matrices (the space $W_{\lambda'}$, in the notation W_ν above) to polynomials of weight σ on ψ by ψ matrices (the space $W_{\sigma'}$); $\mathcal{T}$ is the level of Part II, §6.7. We

give the calculation at one place. Let $\lambda_1 \geq \cdots \geq \lambda_\Psi$ and $\sigma_1 \geq \cdots \geq \sigma_\psi$ be the integer weights, with $\lambda_i \geq \sigma_i \geq \lambda_{i+1}$ and $\lambda_\Psi = \sigma_\psi$. Subtract their common last entry to get polynomial degrees λ'_i, σ'_i. In the polynomial spaces used above, divide each Gaussian norm by the Gaussian norm of the corresponding highest polynomial, $p_{\lambda'}$ or $p_{\sigma'}$. Set

$$a_i = \sigma_i - \lambda_{i+1}, \quad b_i = \lambda_i - \sigma_i, \quad d_0 = \sum_{i=1}^{\psi} a_i, \quad \ell = \sum_{i=1}^{\Psi} \lambda_i - \sum_{i=1}^{\psi} \sigma_i = \lambda_1 - d_0. \tag{III.71}$$

The exponent ℓ uses the original integer weights.

Write the matrix of variables as its first ψ rows and a last row $(z_1, \ldots, z_\psi)$, omitting its last column, whose degree is zero. Let C extract the coefficient of $z_1^{b_1} \cdots z_\psi^{b_\psi}$ and project it orthogonally on the target weight space. Let D_i be the upper left i by i minor and let D'_i be that minor with row i replaced by the last row. Put

$$\xi = \prod_{i=1}^{\psi} D_i^{a_i} (D'_i)^{b_i}, \qquad \kappa = \frac{\|\xi\|_G^2}{\|p_{\lambda'}\|_G^2}. \tag{III.72}$$

Then $C\xi = p_{\sigma'}$. Indeed the exponent b_ψ forces every factor D'_ψ to use z_ψ. Descending induction forces each D'_i to contribute $z_i D_{i-1}$, with $D_0 = 1$. The remaining product is $p_{\sigma'}$.

The next identity is needed for the norm of C: $C^* p_{\sigma'} = \xi/\kappa$. The operator C commutes with unitary substitutions on the first ψ rows, as does its adjoint. Such an intertwiner also commutes with complex invertible row substitutions: differentiate on the unitary group and take complex linear combinations, since skew-Hermitian matrices span all matrices over $\mathbb{C}$. Thus $C^* p_{\sigma'}$ has row degrees σ'_i in the first ψ rows and is invariant under adding an earlier row to a later one among them. (Rows are meant here, where the definition of W_ν used columns: this step uses that $p_{\sigma'}$ has row degrees σ'_i and is unchanged when an earlier row is added to a later one, as well as the corresponding properties for columns.) There is at most one such polynomial up to scalar in the weight space of the domain. Indeed, where the leading minors are nonzero, these row operations and the allowed column operations reduce the upper ψ by ψ matrix to a diagonal matrix. On this slice row and column degrees force a single monomial, $\prod_i X_{ii}^{\sigma'_i} z_i^{b_i}$. Polynomial identity extends uniqueness to all matrices. The polynomial ξ has these properties, and pairing with ξ gives the asserted scalar. The unitary row translates of $p_{\sigma'}$ span the target. To see this, any nonzero invariant orthogonal complement would contain a vector of lexicographically largest row degree among its diagonal torus weights. Adding an earlier row to a later one fixes this vector, since every nonconstant term of the substitution raises that weight. Reduction to a diagonal matrix as above now identifies this vector with a multiple of $p_{\sigma'}$. It could not lie in the complement. It follows by equivariance that $CC^* = I/\kappa$ on the target.

The integral normalisation is an input from the source. There the map is $\vartheta\kappa C$ (Lean: `brStd`, with the constant `brConst`), where ϑ is the following nonzero integer. For each $i \leq \psi$ take all integers v from λ_{i+1} to $\min(\lambda_i, \lambda_{i+1} + d_0)$, omitting σ_i. Multiply $v - i - (\sigma_j - j)$ over these i, v and over $1 \leq j \leq \psi$. Interlacing gives $v - i > \sigma_j - j$ for $i < j$ and the reverse inequality for $i > j$. The omission excludes zero when $i = j$. Put $\Gamma = \lambda_1 - \lambda_\Psi + \Psi$ and $d_1 = \sum_{i=1}^{\psi} \min(\lambda_i - \lambda_{i+1}, d_0)$. Each factor has absolute value at most Γ and there are ψd_1 of them. What the source shows to be integral is the transposed map $\vartheta\kappa C^*$ (Lean: `BranchMap`); the map $\vartheta\kappa C$ itself can have denominators. A map and its adjoint have the same norm, so the statements about norms here are not affected. By $CC^* = I/\kappa$, the squared norm of either map is $\vartheta^2\kappa$.

One can also bound κ by the Gaussian calculation above. The two polynomials in its ratio, ξ and $p_{\lambda'}$, have identical column degrees, and both are unchanged when a multiple of an earlier

column is added to a later column. So both lie in $W_{\lambda'}$, the argument by $X = QR$ applies to both, and their radial factors cancel. Equal column degrees alone would not be enough. Every minor of a unitary matrix has absolute value at most one. In addition, the Gaussian hook formula and the radial moments give

$$\kappa \le \left(\int_{U(\Psi)} |p_{\lambda'}(U)|^2 \, dU \right)^{-1} = \frac{1}{H(\lambda')} \prod_{i=1}^{\Psi} \frac{(\lambda'_i + \Psi - i)!}{(\Psi - i)!} \le \prod_{i=1}^{\Psi} \binom{\lambda'_i + \Psi - i}{\Psi - i} \le \Gamma^{\Psi(\Psi-1)/2}. \quad \text{(III.73)}$$

The middle bound uses the row factorial lower bound for $H(\lambda')$.

The level factor is taken from the source, not derived here (Lean: `brAt`). At level $\mathcal{T}$ the map has the additional scalar $\varpi^{-\ell}$ with $|\varpi|^2 = \mathcal{T}$ in orthonormal coordinates. Its squared norm is consequently $\mathcal{T}^{-\ell} \vartheta^2 \kappa$. Factors at independent places multiply. Taking logarithms, the full bound is

$$-\sum_{\rho} \ell_\rho \log \mathcal{T} + \sum_{\rho} (2\psi d_{1,\rho} + \Psi(\Psi-1)/2) \log \Gamma_\rho. \quad \text{(III.74)}$$

The final functional pairs these images with the target by a weighted sum over lattice classes, with reciprocal stabilizer orders as weights. That it has this form is an input from the source. Cauchy–Schwarz for this finite sum gives precisely the same bound for the squared pairing on unit vectors. Here unit vectors are tuples indexed by the classes, of norm one in the weighted class norms $\sum_p w(p) \|f_p\|^2$ of the paragraph on sums over classes, with $w(p)$ the reciprocal stabilizer order; and the squared pairing is the squared absolute value of the sum over the classes p of $w(p)$ times the inner product of the image of f_p with the pth entry of the second tuple (Lean: `sum_pair_le`).

What this proves, and what it assumes: it proves that $CC^* = I/\kappa$ in the normalised norms, so that $\vartheta \kappa C$ and its adjoint have squared norm $\vartheta^2 \kappa$, that $|\vartheta| \le \Gamma^{\psi d_1}$ and $\kappa \le \Gamma^{\Psi(\Psi-1)/2}$, and the displayed bound for the squared pairing; it assumes interlacing of the two weights with a common last entry, that the last map of the construction is $\vartheta \kappa C$ up to the level factor, that this factor is $\varpi^{-\ell}$, and that the final functional is the weighted sum over classes, and it does not substitute the actual weights into ℓ and d_1.

Relation to the Lean proof. The estimate for one block, with the constant $e^{1/2}(2n + 2)^d \exp(\frac{1}{2} \sum_i h(\nu_i))$ and the factor $\Omega_\tau^{-n_\tau}$, is the statement of the Lean development that the final theorem depends on (`blockBound_of_fockJet`, with `gaussianCoeffIntegral`); the Lean statement is for the square of the norm, which is why the period enters as $\Omega_\tau^{-n_\tau}$; with the factor $|I|\,|J|$ from the sum over classes it becomes the bound for one lift. The source proves it through an integral over $\mathbb{C}^d$ against the Gaussian measure; the proof above uses the sphere instead. For a chain of m lifts the source bounds the norm of each map at each archimedean place in this way, multiplies the bounds (the norm of a composition is at most the product of the norms), and averages the logarithms over the places (`chain_coupled_bound`). The hook products and the lower bound for the eigenvalues of a lift enter the assembly as a term subtracted for each map: in the source the bound for the height has the form $h_A + \sum_j (h_{B,j} - h_{X,j})$, where $h_{B,j}$ comes from the block estimate and $h_{X,j}$ from the eigenvalue bound.

What remains to be written. In the Lean development the passage from this estimate to the first line of (6.5) is an assembly (`divClausesR_of`) of the following further statements, each proved elsewhere in the development. The assembly itself is carried out in §III.7, which proves that the first line of (6.5) follows from a list of such statements, and proves one statement on the periods. The statements themselves are not proved in these pages. Until they are, the first line of (6.5) remains a quoted statement.

(i) *The inputs* (I1)–(I5) *of the reference mean square above*: the integrality of the singular modulus (classical), that the conjugates of the curve run equally often through the ideal classes, the

formula for the mean of $\log A_\rho$ over the classes (the class mean period formula), the description of the reference section as a tensor power, and the identification of the reference section at ρ with $\theta/\theta(0)$ and its metric. Also: that the chain of maps is defined over a field containing M, with $H_{\bar\rho} = H_\rho$; the two inequalities for v_{host} and L_{sc} used in the last step there; and the bounds for β and c in the constant S.

(ii) *The pairing at the end of the chain, and the eigenvalues of a lift.* The norm of the pairing at the end of the chain and the row-by-row comparison are proved above, conditionally on the inputs named there. What remains: the integral realisation of the last map, with its level factor; the substitution of the actual weights of the construction (into ℓ and d_1 for the pairing, and into the inputs (o) and (a)–(e) of the comparison); and, for the lower bound for the eigenvalues of a lift, the hypotheses (E1)–(E7) of the conditional paragraph above: the form of the polynomial at each weight place (E1); the Rallis inner-product formula with the chosen test functions and the form of its scalar (E2); the parameter conditions (E3); the bound for the mass, the bounds for the normalising factors and the formal-degree estimate (E4); the local parameter bound (E5); the form $S = PT$ of the lift and the spanning by Hecke components (E6); and the bound at the other embeddings (E7).

(iii) *The periods*: that the average over the archimedean places of $\log \Omega_\tau$ is the constant $\log \Omega_E$ of §I.7 (the constant relevant here is Ω_E, not the elliptic Ω; the two differ by the factor $\mathrm{rd}_E/\sqrt{d}$), and that the same constant appears in the covolume of the lattices. This is where the Chowla–Selberg formula is used, and where the two occurrences of $\log \Omega$ cancel up to a small error.

(iv) *Local index bounds* at the finite places, including the term F_0 of Part II, §6.7, which is supplied by the divisibility statement (item 1 of §I.7).

(v) *The passage between levels*, and bounds for class numbers and for the constants of the kernels.

(vi) *That the kernels and polynomials of the construction are instances of the abstract setting*: the lattice hypothesis with $v = \mathrm{rd}_E$, the form (III.51) of S, the identification of P with a substitution, and the tuple representation of each map with matching norms.

The inequality that results is, in the notation of Part II, §6.7,

$$\mu_{\mathrm{CC}} + h - F_0 \le \sum_{t \in T} \Big(m(a_t \log a_t - a_t - a_t \log \Omega_E) - a_t \log \mathcal{T} \Big) + C\,|T|\,a\,L_a + C\,m\,(B_0 + N_c)\,x^{2\varepsilon}, \quad \text{(III.75)}$$

where $a = x\,b$ (the orders a_t are not assumed to be at most a; Part II allows a_t up to $(x+1)\,b$), $L_a = \log(a\,\mathrm{rd}_E)$, $B_0 = 2n\,m\,b\,x^{\varepsilon}$ and $N_c = |T_2|\,k\,b\,x^{2\varepsilon}/n^2$. (In the Lean record `RecNums.rhs` the last factor is x^{ε} times a constant, and the form above is weaker; the final Lean statement has the factor $x^{C_{\mathrm{abs}}\varepsilon}$ with a free constant C_{abs}, and the form above is its case $C_{\mathrm{abs}} = 2$. The value of N_c is the one fixed in the Lean file `DataArith.lean`.) Here h is an upper bound for the height of the pairing map, μ_{CC} is the mean slope of the lattice in which the pairing maps take their values (in the source it is the dual of a fixed lattice), in the normalisation for squared norms, as in Part II, §6.7, and F_0 is as in Part II, §6.7. After division by $|T|\,m\,b\,\delta/2$, where $\delta = |T_2|/n$, the term $C\,|T|\,a\,L_a$ gives the error x/m of (M), the summand with B_0 gives m, and the summand with N_c gives k/n^2 (each up to a factor $x^{O(\varepsilon)}$); the fourth error, xm/k, does not come from this inequality.

III.5 Divisibility at inert primes

This section concerns the lower bound for the exponent of an inert prime: inequality (I.19) of §I.4.3. It proves the local estimate behind that bound, for power series, under one hypothesis of integrality, called (B) below. In this section the prime is called ℓ, as in the Lean files; it is the inert prime p of Part I. Several letters used here have other meanings elsewhere in the paper, among them β, λ, ε, a, p, d, x, N, D, q, k, m, n, f, c, σ, e and ψ; in this section each has the meaning given where it is introduced.

Proved outright. The valuation of the coefficients of the inverse of the height-two logarithm, in several variables (Lemma 1); the integrality lemma (Lemma 2); the optimisation for the top row (Lemma 6), which produces the sum $\sum_{h\ \mathrm{odd}}(A/\ell^h - w_h)_+$; weak duality for the linear program of §5; and the closed form of that program for orbits of length two.

Proved conditionally on hypothesis (B). Lemmas 3 and 4: the passage from (B) to a bound for a series β in the exponent, with the gain from $1/(\ell^2-1)$ to $1/(\ell-1)$. Proposition 5: the bound for one Taylor coefficient of a function $U\cdot\exp(\pm\beta)$, with weights on the variables. Theorem J, for orbits of length two: the bound for a product of minors applied to such a function.

Conditional on (B) *and on strong duality, cited from the source.* Theorem J for orbits of length four or more. The cited statement is the equality of a linear program with its dual (Lean: `strongDuality`). This case was not tested end to end.

Taken from the source without proof. The transcription of the top-level inequality `LinkG_at` (§1 below); that the theta kernels of the construction have the form $U\cdot\exp(\pm\beta)$ and satisfy (B); the dictionary between sites and places; the frames and labels that Theorem J requires; the passage from Taylor coefficients (jets) to lattices and to the inequality that the final theorem uses; and the remaining primes. These are listed at the end.

1. The statement in the source

The top-level inequality. The final theorem depends on the Lean statement `LinkG`, of the file `GlobalTheoremInputs.lean`. It does so through the theorems `closed_linkG_of_duty` and `linkG_concrete5`. Its third clause is the inequality of Part II, §6.4, last paragraph. We state it. The scale is $x>1$, $0<\varepsilon<1/52$, $b=\exp(x^{\varepsilon})$, $a=xb$, $X'=\exp(x^{\varepsilon/3})$, $b'=a/X'$, $w''=b'/(xn)$, $X''=a/w''$, $p_0=k^2x^{26}$, $\delta_1=x^{-\varepsilon/8}$. Here m (the number of lifts), k and t_0 are integer parameters of the construction (the fields `m`, `k`, `t` of `D.R`); n is a field of the parent structure (`ord`); and $f=2n$ and $N_m=k+m(t_0+1)$ are defined quantities, not fields. $\mathcal{T}$ is a finite index set (`D.host.TT`). The following are *given data* (fields of the structure $\mathtt{SizeData}_2$): integers a_t for $t\in\mathcal{T}$ with $a\le a_t\le a+a/(kx^{13})$; real numbers $\mathrm{gain}(j,w)$ and $\tau(w)$ for $1\le j\le m$ and w a place of the coefficient field; real numbers $\mathrm{wPart}(j,t,p,i)\ge 1$; subsets $\mathcal{T}_w\subset\mathcal{T}$; and real numbers Sc, η, Λ_G. Put

$$V(j,t,p)=\sum_{\substack{i\ \mathrm{odd}\\ i<\lfloor\log_p a_t\rfloor+2}}\Big(\frac{a_t}{p^i}-\mathrm{wPart}(j,t,p,i)\Big)_+. \tag{III.76}$$

This is the Lean's `vonMangoldtClassSum`. For constants C and b_o put

$$\mathrm{fl}_G(p)=C\,f\,N_m^2\,\log_p a+C\,f\,(\log_p a)^2,\qquad \mathrm{lo}_G(p)=C\,(\mathrm{Sc}+k+\eta)\log_p(2a)+\mathrm{pr}(p)+\mathrm{or}(p), \tag{III.77}$$

where $\mathrm{pr}(p)$ and $\mathrm{or}(p)$ vanish unless p is of the class `inertClass` (defined below), and for such p, with $r_G = \lceil 8(4+8\varepsilon)/\varepsilon \rceil + 1$ and $e_p = \lceil \log_p(k^2 x^{78}) \rceil + 1$,

$$\mathrm{pr}(p) = \begin{cases} C\, r_G\, b' \log x, & p < p_0, \\ C(b'/x + \eta + \mathrm{Sc} + (w''+t_0)\,\delta_1^{-1} \log x) + b_o, & p_0 \le p < X', \\ 0, & p \ge X', \end{cases} \tag{III.78}$$

$$\mathrm{or}(p) = C \log x \Big(\frac{\Lambda_G\, a}{p\, k} + \begin{cases} a\, p^{-e_p}, & p < p_0 \\ a/p^2, & p \ge p_0 \end{cases} \Big) \quad (p < X'), \qquad \mathrm{or}(p) = 0 \quad (p \ge X').$$

In the definition of $\mathrm{pr}(p)$ the three cases are taken in this order, and the first that applies is used, as in the Lean (an `if ... else if ... else`). This matters only when $X' \le p_0$.

The third clause of `LinkG_at` says: for every place w whose residue characteristic p satisfies $p \le X''$ and is of class `signal` or `inertClass`, one has $0 \le \mathrm{gain}(j,w) \le 4\,|\mathcal{T}|\,a/p$ for $1 \le j \le m$, and

$$\sum_{j=1}^{m} \sum_{t \in \mathcal{T}_w} \Big(V(j,t,p) - \mathrm{lo}_G(p) \Big) \;-\; m\, \mathrm{fl}_G(p) \;\le\; \sum_{j=1}^{m} \mathrm{gain}(j,w) \;-\; \tau(w). \tag{III.79}$$

The statement `LinkG` asserts this, with four further clauses, for all large x, with a constant C and with $b_o = b^{\delta(x)}$ for some function $\delta(x) \to 0$. (The classes: a prime not dividing the conductor, inert in the field $\mathbb{Q}(\sqrt{-d})$ of Part I (this d is the prime of Part I; everywhere else in this section d is the length of an orbit), is of class `signal` or `inertClass` according to the value $+1$ or -1 of one further sign; `DataHost.lean`.)

What this statement is. The inequality (III.79) is an inequality between the numbers gain, τ, wPart. In the Lean it is not a statement about valuations. A separate field of the family (`fills`, of type `ParametersFill`) identifies these numbers with explicit formulas (`gainP`, `tauP`), and the proof of (III.79) (`linkG_at_core`) rewrites by `ParametersFill` and concludes by linear arithmetic from a hypothesis `AuxBlock8.Conclusion1`, which already is the inequality for the formulas, together with `Conclusion0`, `lo_le_loG` and `fl_le_flG`. The calculation with the formulas is in `stmtF`. The statements which identify gain with a valuation of theta kernels are other hypotheses and other theorems of the development. We do not claim more than this about (III.79). The Lean theorem in which a valuation of Taylor coefficients is bounded below by a sum of the shape (III.76) is `ThmAOp` (`StatementI.lean`, proved in `ThmAOp.lean`); the final theorem depends on its proof `thmAOp`: this is confirmed by the dependency computation (see §6 (iii)). The rest of this section is about that theorem.

One function, several places. Let K be a field of characteristic zero with a non-archimedean valuation v, normalised by $v(\ell) = 1$. Let $d \ge 2$ be even. The elements of $\mathbb{Z}/d$ are called *sites*; they stand for the places in one orbit of the Frobenius of ℓ, the site $s+1$ being reached from s by the Frobenius followed by complex conjugation (this identification is made in the source and is assumed here). Let $N_0, N_1 \ge 1$. The variables are $Z_{s,q,c}$ and $X_{s,i,c}$, for s a site, $q, i < N_0$, $c < N_1$. A *frame* is a family $\xi(s,i,q) \in K$ with $v \ge 0$; it gives the substitution $Z_{s,q,c} = \sum_i \xi(s,i,q) X_{s,i,c}$. The function is

$$\Theta = U(g(Z)) \cdot \exp\Big(\pm \sum_{q,c} \beta((Z_{s,q,c})_s) \Big), \tag{III.80}$$

where U is a power series in the variables Z with coefficients of valuation ≥ 0, where g is the inverse of the logarithm (III.84) below, taken for the following data (Lean: `entLog`): the index set $\mathcal{Y}$ of (III.84) is the set of the variables $Z_{s,q,c}$, and $\pi(s,q,c) = (s-2,q,c)$ (for $d=2$ this π is the identity, and the logarithm is the one-variable logarithm in each variable); and where β is a power series

in d variables, one for each site, subject to the hypothesis (B) of §3. A *shape* is, for each site s, a list of column lengths $\mathrm{len}(s,1),\dots,\mathrm{len}(s,\nu_1(s))$ in $[1,N_0]$. Put $\nu_2(s)=\#\{c:\mathrm{len}(s,c)\ge 2\}$ and $\mathrm{size}(s)=\sum_c \mathrm{len}(s,c)$. Let TT be a set of sites (the *top* sites; the Lean uses the letters `TT` both for this set and for the index set $\mathcal{T}$ above). The polynomial is $P=\prod_s\prod_c \det\left(X_{s,i,I_{s,c}(k)}\right)_{i,k<\mathrm{len}(s,c)}$ for chosen columns $I_{s,c}(k)<N_1$. The row $i=0$ at a top site is a *top row*; its degree in every monomial of P is $\nu_1(s)$. Put

$$\mathrm{cap}(s)=\begin{cases}\nu_2(s), & s\in TT\\ \nu_1(s), & s\notin TT,\end{cases}\qquad D(s)=\begin{cases}\mathrm{size}(s)-\nu_1(s), & s\in TT\\ \mathrm{size}(s), & s\notin TT.\end{cases}\tag{III.81}$$

Let $d_{\mathrm{top}}(s)$ be the least odd $e\ge 1$ with $s+e\in TT$, or ∞. Let $W(s)$ be the largest $\mathrm{cap}(s-i)$ over odd $i\ge 1$ with $\ell^i\le \mathrm{size}(s-i)$ (zero if there is none); here i runs over all odd integers ≥ 1, not only over $i<d$ (Lean: `Wcap`). Put $\gamma=\ell/(\ell^2-1)$, $\rho=1/(\ell^2-1)$, so that $\gamma=\sum_{h\ \mathrm{odd}}\ell^{-h}$ and $\gamma+\rho=1/(\ell-1)$. The exponent of the Lean theorem is

$$\begin{aligned}B&=\sum_{s\in TT}\left(\mathrm{top}(s)-\gamma W(s)\right)+F(D)-dN_0N_1(\log_\ell \bar a+1),\\ \mathrm{top}(s)&=\sum_{\substack{h\ \mathrm{odd}\\ h<d_{\mathrm{top}}(s)}}\Big(\frac{\nu_1(s)}{\ell^h}-\mathrm{cap}(s+h)\Big)_+ +\sum_{\substack{h\ \mathrm{odd}\\ h>d_{\mathrm{top}}(s)}}\frac{\nu_1(s)}{\ell^h},\end{aligned}\tag{III.82}$$

where $\bar a\ge 1$ is an integer with $\mathrm{size}(s)\le\bar a$ for all s, and $F(D)\ge 0$ is the value of a linear program defined in §5. `ThmAOp` asserts $v(P(\partial)\Theta(0))\ge B$ under hypotheses on the frame stated in Theorem J; here $P(\partial)\Theta(0)=\sum_\kappa P_\kappa\,\kappa!\,\Theta_\kappa$ as in (6.2) of Part II, with Θ expanded in the variables X.

One top site. If $TT=\{s_0\}$ then $d_{\mathrm{top}}(s_0)=\infty$, because d is even. With $A=\nu_1(s_0)$ and $w_h=\nu_1(s_0+h)$ the bound reads

$$v(P(\partial)\Theta(0))\ \ge\ \sum_{h\ \mathrm{odd}}\Big(\frac{A}{\ell^h}-w_h\Big)_+-\gamma W(s_0)+F(D)-dN_0N_1(\log_\ell\bar a+1).\tag{III.83}$$

This is (I.19): A is the length of the first row at the chosen place, and w_h is the number of minors at the place reached by the hth power of the Frobenius followed by conjugation.

2. Weights, and the inverse of the logarithm

Weights. Let $f=\sum_n f_nX^n$ be a power series in finitely many variables X_i, with coefficients in K. Let $r=(r_i)$ be real numbers ≥ 0 and w a real number. Write $n\cdot r=\sum_i n_ir_i$ and $|n|=\sum_i n_i$. We say that f has *weight w at radii r*, and write $f\in\mathcal{W}(r,w)$, if $v(f_n)\ge w-n\cdot r$ for all n. This says that the Gauss norm of f on the polydisc $v(X_i)\ge r_i$ is at most ℓ^{-w}. (Lean: `WtGe`.) The following rules are immediate from $v(x+y)\ge\min(v(x),v(y))$.

(W1) If $f\in\mathcal{W}(r,w)$ and $h\in\mathcal{W}(r,w')$ then $fh\in\mathcal{W}(r,w+w')$. If $r\le r'$ then $\mathcal{W}(r,w)\subset\mathcal{W}(r',w)$.

(W2) If $f\in\mathcal{W}(r,w)$ and the series a_i have no constant term and $a_i\in\mathcal{W}(r',r_i)$, then $f(a)\in\mathcal{W}(r',w)$. (A monomial f_na^n lies in $\mathcal{W}(r',\,w-n\cdot r+n\cdot r)$.)

(W3) If Y has no constant term and $Y\in\mathcal{W}(r,\frac{1}{\ell-1})$, then $\exp(Y)-1$ and $\log(1+Y)$ lie in $\mathcal{W}(r,\frac{1}{\ell-1})$.

(W4) The coefficient of X^κ in fh, in $\exp(Y)$ and in $\log(1+Y)$ depends only on the coefficients of X^n with $n\le\kappa$; so (W1) and (W3) hold with "for all n" replaced by "for all $n\le\kappa$".

For (W3): $Y^k \in \mathcal{W}(r, \frac{k}{\ell-1})$, and $v(k) \le v(k!) = (k - \sigma_\ell(k))/(\ell-1) \le (k-1)/(\ell-1)$, where $\sigma_\ell(k) \ge 1$ is the sum of the digits of k in base ℓ. So $Y^k/k!$ and Y^k/k lie in $\mathcal{W}(r, \frac{1}{\ell-1})$. (Lean: `MulWtLe`, `RadMono`, `SubstWt`, `LogStep`. The Lean statement `ExpWtLe` is not (W3): it gives weight 0 for the exponential of a series, truncated at κ, which is what Proposition 5 uses.)

Lemma 1 (inverse of the logarithm). *Let $\mathcal{Y}$ be a finite set and $\pi : \mathcal{Y} \to \mathcal{Y}$ a map. For $y \in \mathcal{Y}$ put*

$$L_y(X) = \sum_{e \ge 0} \ell^{-e}\, X_{\pi^e(y)}^{\ell^{2e}}. \tag{III.84}$$

There is a unique family $g = (g_y)$ of power series without constant term with $L_y(g) = X_y$ for all y. It satisfies $g_y \in \mathcal{W}(\rho, \rho)$, where the radius is ρ for every variable:

$$v(\text{coefficient of } X^n \text{ in } g_y) \;\ge\; -(|n|-1)\rho. \tag{III.85}$$

Proof. The equation is $g_y = X_y - \sum_{e\ge1} \ell^{-e} g_{\pi^e(y)}^{\ell^{2e}}$. It determines the coefficients by induction on $|n|$, since the terms with $e \ge 1$ involve only coefficients of lower degree. The bound holds for $|n| = 1$. Let $|n| \ge 2$ and $r = \ell^{2e}$, $e \ge 1$. The coefficient of X^n in an rth power is a sum of products of r coefficients, of degrees $|n_1|, \dots, |n_r| \ge 1$ with sum $|n|$. By induction such a product has valuation at least $-(|n| - r)\rho$. Division by ℓ^e costs e, and $e \le 1 + \ell^2 + \cdots + \ell^{2(e-1)} = (r-1)\rho$. □

This is Part II, §2, in several variables. (Lean: `GenLogWeight`, `generalized_log_weight`; the series (III.84) is `genLog`. Also $g(L(X)) = X$: `GenLogInverse`.) In one variable (π the identity) a computation confirms (III.85) for $\ell = 3, 5, 7$ and all degrees up to 801, 1201, 1201: the minimum of $v(g_n) + (n-1)\rho$ is 0, attained exactly at $n = \ell^{2e}$. The bound cannot be improved.

3. From integrality to a bound for β

In (III.80) the factor $U(g(Z))$ lies in $\mathcal{W}(\rho, 0)$, by Lemma 1 and (W2). After multiplication by $\kappa!$ this gives the exponent $|\kappa|\,(\frac{1}{\ell-1} - \rho) = |\kappa|\gamma$ of §I.3. For the exponential factor to do as well, by (W3) the series β should have weight $\frac{1}{\ell-1}$ at radii ρ. Integrality gives directly only the weight ρ. The gain from ρ to $\frac{1}{\ell-1}$ is the content of this paragraph. It uses that the degrees of the monomials are divisible by $\ell + 1$. The proof given below also uses that the coefficients lie in a field with value group $\mathbb{Z}$ (an unramified field); the source carries this hypothesis (`Unram`, in `BHyp`), and it is kept as (B1). The statements do not need it: see the Remarks after Theorem J.

Hypothesis (B). Let $K_0 \subset K$ be a subfield with $v(K_0^\times) = \mathbb{Z}$. Let $\zeta \in K$ be a primitive root of unity of order $\ell^d - 1$. For a site s, represented by $0 \le s < d$, put $\mathrm{lw}(s) = (-1)^s \ell^s$ (the letters are those of the Lean). The series β, in variables X_s ($s \in \mathbb{Z}/d$), satisfies: (B1) its coefficients lie in K_0; (B2) $\beta((\zeta^{\mathrm{lw}(s)} X_s)_s) = \beta(X)$; (B3) every monomial of β has degree ≥ 1 in the variables at even sites, degree ≥ 1 in the variables at odd sites, and total degree ≥ 3; (B4) for the two series

$$\begin{aligned} \Delta_1\beta &= \beta(x_{\mathrm{ev}} + y_{\mathrm{ev}}, x_{\mathrm{od}}) - \beta(x_{\mathrm{ev}}, x_{\mathrm{od}}) - \beta(y_{\mathrm{ev}}, x_{\mathrm{od}}), \\ \Delta_2\beta &= \beta(x_{\mathrm{ev}}, x_{\mathrm{od}} + y_{\mathrm{od}}) - \beta(x_{\mathrm{ev}}, x_{\mathrm{od}}) - \beta(x_{\mathrm{ev}}, y_{\mathrm{od}}), \end{aligned} \tag{III.86}$$

in which the variables at the even (resp. odd) sites are doubled, the series $\exp(\Delta_i\beta) \circ L$ has all coefficients of valuation ≥ 0. Here L is (III.84) for the doubled set of variables, with π the shift $s \mapsto s - 2$ of the site. (Lean: `BHyp`; Δ_1, Δ_2 are `de1`, `de2`; L is `letterLog`.) Hypothesis (B4) says that $\exp(\Delta_i\beta)$ becomes integral when written in the coordinates $T = g(X)$ of the formal group. It is assumed here, from the source.

Lemma 2 (quantization). *Let $c \in K_0$ and let N be an integer divisible by $\ell + 1$. If $v(c) \ge \rho - N\rho$, then $v(c) \ge \frac{1}{\ell-1} - N\rho$.*

Proof. Let $c \neq 0$, $k = v(c) \in \mathbb{Z}$ and $N = (\ell+1)m$. Multiply $k \geq \rho - (\ell+1)m\rho$ by $\ell^2 - 1$:

$$(\ell+1)\big(k(\ell-1)+m\big) \;\geq\; 1. \tag{III.87}$$

The number $z = k(\ell-1)+m$ is an integer and $(\ell+1)z \geq 1$; so $z \geq 1$. Divide $z \geq 1$ by $\ell - 1$: $k \geq \frac{1}{\ell-1} - \frac{m}{\ell-1} = \frac{1}{\ell-1} - N\rho$. □

Here k is the valuation of the coefficient and m is its degree divided by $\ell+1$. The hypothesis loses the factor $\ell+1$ because $N\rho = m/(\ell-1)$ has denominator $\ell-1$ only. (Lean: `Quant`, `weightQuantization`; the step $(\ell+1)z \geq 1 \Rightarrow z \geq 1$ is at lines 83–91 of `OneLetterRadiusIdentity.lean`. The Lean statement `Quant` has as hypothesis the congruence of the torus, that $\ell^d - 1$ divides $\sum_x n_x \,\mathrm{lw}(s_x)$ in the notation of the proof of Lemma 3; it does not have the hypothesis that $\ell + 1$ divides N.) Without the divisibility the conclusion fails: for $N = 1$ the value $k = 0$ is allowed and $\frac{1}{\ell-1} - \rho > 0$.

Lemma 3. *Assume* (B). *Then $\Delta_1\beta$ and $\Delta_2\beta$ lie in $\mathcal{W}(\rho, \frac{1}{\ell-1})$.*

Proof. Let Δ be one of them and $E = \exp(\Delta) \circ L$. By (B3) Δ has no constant term, so E has constant term 1. Since $L \circ g$ is the identity, $\exp(\Delta) - 1 = (E-1) \circ g$. By (B4), Lemma 1 and (W2), a monomial of degree $j \geq 1$ of $E - 1$ contributes a series in $\mathcal{W}(\rho, j\rho) \subset \mathcal{W}(\rho, \rho)$. So $\exp(\Delta) - 1 \in \mathcal{W}(\rho, \rho)$. Its coefficients lie in K_0, by (B1). By (B2), Δ and $\exp(\Delta)$ are unchanged when each variable at the site s is multiplied by $\zeta^{\mathrm{lw}(s)}$. So a monomial X^n with nonzero coefficient has $\sum n_x \mathrm{lw}(s_x) \equiv 0$ modulo $\ell^d - 1$, where s_x is the site of the variable x. Now $\ell + 1$ divides $\ell^d - 1$, because d is even, and $\mathrm{lw}(s) \equiv (-1)^s(-1)^s = 1$ modulo $\ell + 1$. Hence $\ell + 1$ divides $|n|$. Lemma 2 gives $\exp(\Delta) - 1 \in \mathcal{W}(\rho, \frac{1}{\ell-1})$, and (W3) gives the same for $\Delta = \log(1 + (\exp(\Delta) - 1))$. □

(Lean: `asmS_diff` in `ConstantCoeffEvenDerivativeEqZero.lean`, with `ExpDeInv` for the divisibility.) A computation tests the chain of Lemmas 1–3 in one variable: for $Y = \varphi(g(X))$ with φ an integral series in $T^{\ell+1}$, and $\Delta = \log(1 + Y)$, the minimum over $n \leq 240$ of $v(\Delta_n) - (\frac{1}{\ell-1} - n\rho)$ is 0 for $\ell = 3, 5, 7$, attained at $n = \ell + 1$. For $\varphi(T) = T$ the bound fails (the minimum is $-11/8$, $-29/24$, $-55/48$).

Lemma 4 (from the differences to β). *Assume* (B). *Call a monomial* exceptional *if it is $X_s^{\ell^\alpha} X_{s'}^{\ell^{\alpha'}}$ with s even, s' odd, $(\alpha, \alpha') \neq (0, 0)$ and $s + \alpha = s' + \alpha'$ in $\mathbb{Z}/d$. Then every coefficient c_n of β satisfies*

$$v(c_n) \;\geq\; \frac{1}{\ell-1} - |n|\rho - \mathrm{ex}(n), \qquad \mathrm{ex}(n) = 1 \text{ if } X^n \text{ is exceptional}, \quad \mathrm{ex}(n) = 0 \text{ otherwise.} \tag{III.88}$$

Proof. Write $X^n = X_{\mathrm{ev}}^u X_{\mathrm{od}}^{u'}$ with multi-indices u, u' on the even and the odd sites. By (III.86), the coefficient of $x_{\mathrm{ev}}^j \, y_{\mathrm{ev}}^{u-j} \, x_{\mathrm{od}}^{u'}$ in $\Delta_1\beta$ is $c_n \prod_s \binom{u_s}{j_s}$, for every j with $0 \leq j \leq u$, $j \neq 0$, $j \neq u$. By Lemma 3, $v(c_n) \geq \frac{1}{\ell-1} - |n|\rho - \mu(u)$, where $\mu(u)$ is the least valuation of such a product of binomial coefficients; this holds if $|u| \geq 2$. If u involves two sites, take j equal to u at one site and 0 elsewhere: $\mu(u) = 0$. Let u be supported at one site, with u_s not a power of ℓ. If u_s has at least two nonzero digits in base ℓ, let j be its leading digit times the corresponding power of ℓ; if u_s is ℓ^k times one digit between 2 and $\ell - 1$, let $j = \ell^k$. By Lucas's congruence $\binom{u_s}{j}$ is prime to ℓ, so $\mu(u) = 0$. If $u_s = \ell^\alpha$ with $\alpha \geq 1$ then $\mu(u) = 1$, attained at $j = \ell^{\alpha-1}$. The same holds for u' and $\Delta_2\beta$. By (B3), $|u| \geq 2$ or $|u'| \geq 2$. So $\mathrm{ex}(n) = 0$ unless u and u' are each a power of ℓ at one site, not both equal to 1. In that case $X^n = X_s^{\ell^\alpha} X_{s'}^{\ell^{\alpha'}}$, and if $c_n \neq 0$ then (B2) gives $\ell^{s+\alpha} \equiv \ell^{s'+\alpha'}$ modulo $\ell^d - 1$, that is $s + \alpha = s' + \alpha'$ in $\mathbb{Z}/d$. □

(Lean: the statement is `LemmaS`, proved as `lemmaS`; the step of Lemma 4 is `BinomExtract`. The proof of Lemma 4 above is ours; the Lean proof of `BinomExtract` was not read. The facts on binomial coefficients were checked for $u_s \leq 1500$.) So β has the weight $\frac{1}{\ell-1}$ that is needed, except

at the exceptional monomials, where one unit is missing. The exceptional monomials pair a site s with the site $s' = s + (\alpha - \alpha')$, at an odd distance. They are the reason for the weights of the next paragraph.

4. One Taylor coefficient, with weights on the variables

Proposition 5. *Assume* (B). *Let X_i ($i \in \iota$) be finitely many variables. Let $Z_y = \sum_i M_{y,i} X_i$ be linear forms with $v(M_{y,i}) \geq 0$, one for each variable Z_y of* (III.80)*, and let Θ be* (III.80) *written in the variables X. Let $\mathcal{B} = \sum_{q,c} \beta((Z_{s,q,c})_s) = \sum_n \mathcal{B}_n X^n$. Let κ be a multi-index and let $r_i \geq 0$ be real numbers such that*

$$\ell^\alpha r_i + \ell^{\alpha'} r_j \;\geq\; 1 \qquad \text{whenever } X_i^{\ell^\alpha} X_j^{\ell^{\alpha'}} \text{ divides } X^\kappa \text{ and } v(\mathcal{B}_{\ell^\alpha e_i + \ell^{\alpha'} e_j}) < \tfrac{1}{\ell-1} - (\ell^\alpha + \ell^{\alpha'})\rho. \quad \text{(III.89)}$$

The case $i = j$ is included; the monomial is then $X_i^{\ell^\alpha + \ell^{\alpha'}}$. Then, with $\lambda(k) = \lfloor \log_\ell k \rfloor + 1$ for $k \geq 1$ and $\lambda(0) = 0$,

$$v(\kappa!\,\Theta_\kappa) \;\geq\; \sum_i \kappa_i\,(\gamma - r_i) \;-\; \sum_i \lambda(\kappa_i). \quad \text{(III.90)}$$

Proof. First, $\mathcal{B}_n$ has valuation $\geq \frac{1}{\ell-1} - |n|\rho$ unless $n = \ell^\alpha e_i + \ell^{\alpha'} e_j$ for some i, j, α, α' ($i = j$ allowed); and $\geq \frac{1}{\ell-1} - 1 - |n|\rho$ always. Indeed, by (III.88) and (W2) only the exceptional monomials $c\, X_s^{\ell^\alpha} X_{s'}^{\ell^{\alpha'}}$ of β matter. For a linear form $Z = \sum_i M_i X_i$ one has $Z^{\ell^\alpha} = \sum_i M_i^{\ell^\alpha} X_i^{\ell^\alpha} + \ell R$, with R a polynomial with coefficients of valuation ≥ 0. So all monomials of $Z^{\ell^\alpha} Z'^{\ell^{\alpha'}}$ other than the $X_i^{\ell^\alpha} X_j^{\ell^{\alpha'}}$ carry a factor ℓ. Now put $r_i' = \rho + r_i$. For $n \leq \kappa$ we get $v(\mathcal{B}_n) \geq \frac{1}{\ell-1} - n \cdot r'$: this is clear if $v(\mathcal{B}_n) \geq \frac{1}{\ell-1} - |n|\rho$, and otherwise $n = \ell^\alpha e_i + \ell^{\alpha'} e_j$ and $n \cdot r' = |n|\rho + \ell^\alpha r_i + \ell^{\alpha'} r_j \geq |n|\rho + 1$ by (III.89). By (W3) and (W4) the coefficients of X^n in $\exp(\pm\mathcal{B})$, $n \leq \kappa$, have valuation $\geq -n \cdot r'$. The factor $U(g(Z))$ lies in $\mathcal{W}(\rho, 0) \subset \mathcal{W}(r', 0)$ by Lemma 1 and (W2), (W1). By (W1), $v(\Theta_\kappa) \geq -\kappa \cdot r'$. Finally $v(k!) = (k - \sigma_\ell(k))/(\ell - 1) \geq \frac{k}{\ell-1} - \lambda(k)$, and $\frac{1}{\ell-1} - \rho = \gamma$. □

(Lean: `LemmaR`, with `PureCross` for the first step and `FactVal` for the last; the Lean loss is $|\iota|\,(\log_\ell |\kappa| + 1)$. `LemmaR` has one more hypothesis than Proposition 5: each variable X_i belongs to one site. Proposition 5 does not need it. The statement `LemmaR` is in the archive, proved as `lemmaR_of`, but the dependency computation shows that `LemmaR` is not among the declarations the final theorem depends on. Proposition 5 is proved here independently. By occurrences of names the chain uses `LemmaRA`.) Each variable pays its radius r_i out of the gain γ. The radii are needed only where an exceptional monomial is present with a coefficient that is not divisible by one more power of ℓ.

One variable. Take one variable X and $\Theta = U(g(X)) \exp(\mathcal{B}(X))$, where $v(\mathcal{B}_n) \geq \frac{1}{\ell-1} - n\rho$ for $n \notin S$ and $v(\mathcal{B}_n) \geq \frac{1}{\ell-1} - 1 - n\rho$ for $n \in S = \{\ell^\alpha + \ell^{\alpha'} : \alpha + \alpha' \text{ odd}\}$. The second half of the proof of Proposition 5 applies to such a series: what is needed is $nr \geq 1$ for $n \in S$, $n \leq \kappa$. The least element of S is $\ell + 1$, so $r = \frac{1}{\ell+1}$ will do, and $\gamma - \frac{1}{\ell+1} = \rho$. Thus

$$v(\kappa!\,\Theta_\kappa) \;\geq\; \kappa\gamma - \lambda(\kappa) \;\text{ if } \mathcal{B} \text{ has no terms in } S, \qquad v(\kappa!\,\Theta_\kappa) \;\geq\; \kappa\rho - \lambda(\kappa) \;\text{ in general.} \quad \text{(III.91)}$$

In the first case the factorial contributes $\kappa/(\ell - 1) = \kappa(\gamma + \rho)$ and the function costs $\kappa\rho$. In the second the exceptional terms can cost a further $\kappa/(\ell + 1)$. (Lean, for the second: `OneLetterRadius`. The Lean statement `OneLetterR` is for several variables through an integral matrix, with exponent $(\mathrm{sgn}/2) \sum \beta(Z, \mathrm{ep} \cdot Z)$.) A computation tests (III.91) with exact rational arithmetic for $\ell = 3, 5, 7$ and all $\kappa \leq 240$, with U a random series with integer coefficients that are units at ℓ, and $\mathcal{B}_n$ of the least valuation allowed. The minimum over κ of the left side minus the right side is $\frac{1}{4}, \frac{1}{6}, \frac{1}{8}$ in the

first case and $\frac{3}{4}, \frac{5}{6}, \frac{7}{8}$ in the second. In the second case the first bound fails: at $\kappa = 240$ the left side is below $\kappa\gamma - \lambda(\kappa)$ by 57, 37, 28. So the cost $\kappa/(\ell+1)$ is real in this model ($240/(\ell+1) = 60, 40, 30$).

5. The choice of the weights

In Proposition 5 the bound is $\sum_i \kappa_i(\gamma - r_i)$, and the radii are free subject to (III.89). The exponent (III.82) comes from a choice of radii. Two elementary optimisations are involved.

Lemma 6 (the top radius). *Let $A \geq 0$, let $c_j \geq 0$ for odd $j \geq 1$, and let $J \geq 1$ be odd. There is $\omega \geq \ell^{-J}$ with*

$$A\omega + \sum_{\substack{j \text{ odd} \\ j<J}} c_j \,(1 - \ell^j\omega)_+ \;\leq\; \sum_{\substack{j \text{ odd} \\ j<J}} \min\Big(c_j, \frac{A}{\ell^j}\Big) + \frac{A}{\ell^J}. \tag{III.92}$$

Consequently

$$A(\gamma - \omega) - \sum_{\substack{j \text{ odd} \\ j<J}} c_j \,(1 - \ell^j\omega)_+ \;\geq\; \sum_{\substack{j \text{ odd} \\ j<J}} \Big(\frac{A}{\ell^j} - c_j\Big)_+ + \sum_{\substack{j \text{ odd} \\ j>J}} \frac{A}{\ell^j}. \tag{III.93}$$

Proof. Let j_0 be the least odd $j < J$ with $c_j \geq A/\ell^j$, and $j_0 = J$ if there is none. Take $\omega = \ell^{-j_0}$. On the left of (III.92) the terms with $j \geq j_0$ vanish, and each term with $j < j_0$ is at most $c_j = \min(c_j, A/\ell^j)$. The term $A\omega = A/\ell^{j_0}$ equals $\min(c_{j_0}, A/\ell^{j_0})$ if $j_0 < J$, and is the last term on the right if $j_0 = J$. For (III.93) use $\gamma A = \sum_{j \text{ odd}} A/\ell^j$ and $x - \min(c, x) = (x - c)_+$. □

(The archive contains this statement as `TopRadius`, proved as `top_radius_holds` with the same choice of ω. The dependency computation shows that neither is among the declarations the final theorem depends on. Lemma 6 is proved here independently. The declaration the final theorem uses for the top-row term has been identified by reading the proof terms: it is `asmA_top`, which the proof of `thmAOp` reaches through `thmAJetA` and `rowRadiusAssembly`; the bound for one Taylor coefficient on that route is `lemma_R_log_loss`. The lemma was checked on 20000 random cases.) Here ω is the radius of the variables of a top row of length A, and c_j bounds the degree in those variables at the site $s + j$ which are linked to the top row by an exceptional monomial. Such a variable needs the radius $(1 - \ell^j\omega)_+$, by (III.89). The right side of (III.93), with both of its sums, is $\mathrm{top}(s)$ of (III.82) for $A = \nu_1(s)$, $c_j = \mathrm{cap}(s + j)$ and $J = d_{\mathrm{top}}(s)$: the second sum of (III.93), over odd $j > J$, is the second sum in the definition of $\mathrm{top}(s)$. The lemma asserts that a suitable ω exists, not that (III.93) holds for every $\omega \geq \ell^{-J}$. If $d_{\mathrm{top}}(s) = \infty$ and only the j with $\ell^j \leq A$ occur on the left, the same choice (with $\omega = 0$ if there is no j_0) gives $\sum_{j \text{ odd}} (A/\ell^j - c_j)_+$ on the right.

The linear program. Let $D : \mathbb{Z}/d \to \mathbb{R}_{\geq 0}$. A *flow* is a family $\psi(s, i) \geq 0$, for sites s and odd $i < d$, with

$$\sum_i \ell^i \,\psi(s, i) + \sum_i \psi(s - i, i) \;\leq\; D(s) \qquad \text{for all } s. \tag{III.94}$$

Let $\mathrm{maxFlow}(D)$ be the supremum of $\sum_{s,i} \psi(s, i)$ over flows, and $F(D) = \gamma \sum_s D(s) - \mathrm{maxFlow}(D)$. A *dual* is a family $0 \leq \sigma(s) \leq 1$ with $\ell^i\sigma(s) + \sigma(s + i) \geq 1$ for all s and all odd $i < d$. (Lean: `IsFlow`, `Fnum`, `IsDual`; and `lpF_eq_Fnum` for the form used elsewhere.)

Weak duality, proved here. For a flow ψ and a dual σ,

$$\sum_{s,i} \psi(s, i) \;\leq\; \sum_{s,i} \psi(s, i)\big(\ell^i\sigma(s) + \sigma(s + i)\big) = \sum_s \sigma(s)\Big(\sum_i \ell^i\psi(s, i) + \sum_i \psi(s - i, i)\Big) \;\leq\; \sum_s \sigma(s)D(s). \tag{III.95}$$

So $\sum_s D(s)(\gamma - \sigma(s)) \leq F(D)$ for every dual σ.

Strong duality, cited. For $\ell \geq 3$ and d even there is a dual σ with $\sum_s D(s)(\gamma - \sigma(s)) = F(D)$, and with $\ell^i \sigma(s) + \sigma(s+i) \geq 1$ for *all* odd $i \geq 1$ (Lean: `strongDuality`, of type `LemmaFa`). For a dual with the last property, if $s + \alpha = s' + \alpha'$ and $\alpha + \alpha'$ is odd, then

$$\ell^{\alpha} \sigma(s) + \ell^{\alpha'} \sigma(s') \;\geq\; 1 : \tag{III.96}$$

if $\alpha > \alpha'$ this is $\ell^{\alpha'}$ times $\ell^{\alpha - \alpha'} \sigma(s) + \sigma(s + (\alpha - \alpha')) \geq 1$; if $\alpha' > \alpha$ exchange the roles of (s, α) and (s', α') (and $\alpha \neq \alpha'$, since $\alpha + \alpha'$ is odd). So such a σ, used as radius for all variables at the site s, satisfies (III.89).

Orbits of length two, proved here. Let $d = 2$. Then

$$\mathrm{maxFlow}(D) = \min\Big(D_0,\, D_1,\, \frac{D_0 + D_1}{\ell + 1}\Big), \quad F(D) = \begin{cases} \rho\,(D_0 + D_1), & D_1/\ell \leq D_0 \leq \ell D_1, \\ \gamma D_0 - (1-\gamma) D_1, & \ell D_1 \leq D_0, \\ \gamma D_1 - (1-\gamma) D_0, & \ell D_0 \leq D_1. \end{cases} \tag{III.97}$$

Indeed, in the first case $\psi = (\frac{\ell D_0 - D_1}{\ell^2 - 1}, \frac{\ell D_1 - D_0}{\ell^2 - 1})$ is a flow and $\sigma = (\frac{1}{\ell+1}, \frac{1}{\ell+1})$ a dual, both of value $\frac{D_0 + D_1}{\ell + 1}$; in the second $\psi = (D_1, 0)$ and $\sigma = (0, 1)$, of value D_1; the third is symmetric. By (III.95) these are optimal. These duals satisfy the inequalities for all odd i. (Checked exactly on 20000 cases, and strong duality numerically for $d = 2, 4, 6$.) Note that $F(D)$ is *not* $\sum_{h\ \mathrm{odd}} (D_0/\ell^h - D_1)_+$: for $\ell = 5$ and $D = (1000, 10)$ the two are 200.42 and 190; for $D = (1000, 250)$ they are 52.08 and 0. One has $F(D) \geq \sum_{h\ \mathrm{odd}} (D_0/\ell^h - D_1)_+$, with equality only for $D_1 = 0$. In the source the sum of the shape (III.76) is the term $\mathrm{top}(s)$. Here it comes from Lemma 6; on the route of the final theorem it is the statement `asmA_top`. The linear program governs the rows other than the top rows.

Theorem J. *Assume* (B)*, with ℓ an odd prime and $d \geq 2$ even. Let a shape, a set TT, the polynomial P and the integer $\bar{a}$ be as in §1. Call a row (s, i)* ordinary *if it is not a top row. Call a top site s_0, an odd integer j and an ordinary row (s, i)* linked *if $s = s_0 + j$, and $\ell^j \leq \nu_1(s_0)$ if $j > 0$, and $\ell^{-j} \leq \mathrm{size}(s)$ if $j < 0$. Let* spec *be an injective map from a set of ordinary rows to pairs (s_0, j), such that s_0, j, (s, i) are linked whenever $\mathrm{spec}(s, i) = (s_0, j)$; the pair is the* label *of the row. Here j is an integer, not a class modulo d: (s_0, j) and $(s_0, j + d)$ are different labels for the same site. Suppose that whenever s_0, j, (s, i) are linked and (s_0, j) is not the label of (s, i), the frame satisfies*

$$v\Big(\sum_q \xi(s_0, 0, q)^{\ell^j}\, \xi(s, i, q)\Big) \geq 1 \quad (j > 0), \qquad v\Big(\sum_q \xi(s_0, 0, q)\, \xi(s, i, q)^{\ell^{-j}}\Big) \geq 1 \quad (j < 0). \tag{III.98}$$

Then $v(P(\partial)\Theta(0)) \geq B'$, where $B' = \sum_{s \in TT} (\mathrm{top}(s) - \gamma W(s)) + F(D) - \max_\kappa \sum_y \lambda(\kappa_y)$, the maximum being over the monomials X^κ of P and the sum over the variables y. This is the loss that the proof gives. Since $\sum_y \lambda(\kappa_y) \leq dN_0N_1(\log_\ell \bar{a} + 1)$, one has $B' \geq B$ with B as in (III.82)*, which is the exponent of the Lean statement. For $d = 2$ this is proved here; for $d \geq 4$ the proof uses the cited strong duality.*

Proof. Since P has integer coefficients, it suffices to prove $v(\kappa! \Theta_\kappa) \geq B'$ for each monomial X^κ of P. A determinant is linear in each row. So the degree of X^κ in the variables of the row (s, i) is the number of c with $\mathrm{len}(s, c) > i$. It is $\nu_1(s)$ for $i = 0$, and at most $\nu_2(s)$ for $i \geq 1$. Hence the degree is $\nu_1(s)$ in a top row, at most $\mathrm{cap}(s)$ in an ordinary row, and $D(s)$ in all ordinary rows at s together; and each $\kappa_y \leq \bar{a}$.

Radii. Let σ be an optimal dual for D, as above. For $s \in TT$ let $\omega(s)$ be given by Lemma 6 with $A = \nu_1(s)$, $c_j = \mathrm{cap}(s + j)$ and $J = d_{\mathrm{top}}(s)$ (and by the remark after it if $d_{\mathrm{top}}(s) = \infty$). Give the variables of a top row the radius $\omega(s)$. Give the variables of an ordinary row (s, i) the radius $\sigma(s)$

if the row has no label; $\max(\sigma(s), 1-\ell^j\omega(s_0))$ if its label is (s_0, j) with $j>0$; and $\max(\sigma(s), \ell^{-|j|})$ if $j<0$. (Lean: `rowRad`.)

The condition (III.89). With $Z_{s,q,c}=\sum_i \xi(s,i,q)X_{s,i,c}$, a monomial $X_y^{\ell^\alpha}X_{y'}^{\ell^{\alpha'}}$, $y=(s,i,c)$, $y'=(s',i',c')$, has in $\mathcal{B}$ the coefficient $c_n\sum_q \xi(s,i,q)^{\ell^\alpha}\xi(s',i',q)^{\ell^{\alpha'}}$ if $c=c'$, and 0 otherwise, where c_n is the coefficient of $X_s^{\ell^\alpha}X_{s'}^{\ell^{\alpha'}}$ in β (it is 0 by (B3) if s and s' have the same parity). By (III.88) the condition concerns only the case $s+\alpha=s'+\alpha'$, $\alpha+\alpha'$ odd, and only if the sum over q has valuation <1. If both rows are ordinary, their radii are at least $\sigma(s)$, $\sigma(s')$, and (III.96) applies. If both are top rows, say with $\alpha>\alpha'$, then $d_{\mathrm{top}}(s)\le\alpha-\alpha'$, so $\ell^\alpha\omega(s)\ge\ell^\alpha\ell^{-(\alpha-\alpha')}\ge 1$. There remains the case of a top row and an ordinary row; by symmetry let $(s,0)$ be the top row and (s',i') the ordinary row, and $j=\alpha-\alpha'$, so $s'=s+j$. Let $j>0$. Since the monomial divides X^κ, $\ell^j\le\ell^\alpha\le\nu_1(s)$. If the label of (s',i') is (s,j), then $\ell^\alpha\omega(s)+\ell^{\alpha'}(1-\ell^j\omega(s))=\ell^{\alpha'}\ge 1$. If not, (III.98) holds, and $\sum_q\big(\xi(s,0,q)^{\ell^j}\xi(s',i',q)\big)^{\ell^{\alpha'}}\equiv\big(\sum_q\xi(s,0,q)^{\ell^j}\xi(s',i',q)\big)^{\ell^{\alpha'}}$ modulo ℓ has valuation ≥ 1: the condition is void. Let $j<0$. Then $\ell^{|j|}\le\ell^{\alpha'}\le\mathrm{size}(s')$. If the label is (s,j) then $\ell^{\alpha'}\ell^{-|j|}=\ell^\alpha\ge 1$. If not, (III.98) applies in the same way.

The sum. By Proposition 5 it remains to show $\sum_y\kappa_y(\gamma-r_y)\ge\sum_{s\in TT}(\mathrm{top}(s)-\gamma W(s))+F(D)$. Use $\max(\sigma,u)\le\sigma+u_+$. The terms with σ give $\sum_s D(s)(\gamma-\sigma(s))=F(D)$. The labels (s_0,j) with $j>0$ cost at most $\mathrm{cap}(s_0+j)(1-\ell^j\omega(s_0))_+$ each, because spec is injective and one row has degree at most cap; with the top row this gives $\mathrm{top}(s_0)$ by (III.93). The labels (s_0,j) with $j<0$ cost at most $\mathrm{cap}(s_0-|j|)\,\ell^{-|j|}\le W(s_0)\,\ell^{-|j|}$, and $\sum_{j\ \mathrm{odd}}\ell^{-j}=\gamma$. The loss in (III.90) is $\sum_y\lambda(\kappa_y)$, as in B'. It is at most $dN_0N_1(\log_\ell\bar a+1)$, because there are dN_0N_1 variables and each $\kappa_y\le\bar a$. □

(Lean: `ThmAJetA` for one κ, `ThmAAssembly` for the sum, `ThmARadii` for the condition, `ThmAOp` for the result. The Lean statement `ThmAOp` is more general: it allows an integral linear operator on the ordinary rows before the minors are formed. That generality is not proved here. The inequality for the sum was tested on 3000 random shapes.)

Remarks. The following are observations of the test of this section. They are recorded without proof, and nothing above depends on them.

(1) The field K_0. The field K_0, hypothesis (B1) and Lemma 2 are not used in Lemmas 3–4, Proposition 5 and Theorem J, except in the proof of Lemma 3 as written. The reason is this. In that proof the series $E=\exp(\Delta)\circ L$ is invariant under the torus, because L commutes with it. Hence every monomial of $E-1$ has degree j divisible by $\ell+1$, so $j\ge\ell+1$, and it contributes a series in $\mathcal{W}(\rho,j\rho)\subset\mathcal{W}(\rho,(\ell+1)\rho)=\mathcal{W}(\rho,\frac{1}{\ell-1})$. This is the conclusion of Lemma 3. The Lean statement `BinomExtract` has no K_0 either.

(2) Strong duality. The strong duality cited from the source is the duality theorem of linear programming together with two elementary reductions. First, replacing σ by $\min(\sigma,1)$ keeps the constraints and does not increase $\sum_s\sigma(s)D(s)$. Second, for odd $i\ge d$ write $i=i'+kd$ with i' odd and $1\le i'<d$; then $\ell^i\sigma(s)+\sigma(s+i)\ge\ell^{i'}\sigma(s)+\sigma(s+i')\ge 1$, so every dual satisfies the inequalities for all odd $i\ge 1$. The equality was confirmed in every instance computed ($d=2,\dots,10$).

(3) The prime 2. That ℓ is odd is not used in the proofs of this section, and Lemma 1 holds for $\ell=2$. No series with (B) was constructed for $\ell=2$, so Lemmas 3–4 and Theorem J were not tested for $\ell=2$.

An example. The test of this section constructed an explicit series β for orbits of length two ($d=2$). Let $\ell\equiv 3 \bmod 4$. Let E_c be the curve $Y^2=X^3-cX$, with c in $\mathbb{Z}[i]$ and $c^{(\ell^2-1)/4}\equiv -1$ modulo ℓ ($c=-i$ for $\ell=3$ and 11, $c=-1-i$ for $\ell=7$). It is a twist of the lemniscate curve, the curve of the lemniscate sine of §I.3. Let σ_c be the Weierstrass σ-function of E_c (with $g_2=4c$,

$g_3 = 0$), let $t = -X/Y$, and put $m(z) = \log(\sigma_c(z)/t(z))$ and $\tilde{\beta}(z,w) = m(z+w) - m(z) - m(w)$. The series $\beta(X_0, X_1)$ is the part of $\tilde{\beta}$ that is invariant under $(X_0, X_1) \mapsto (\zeta X_0, \zeta^{-\ell} X_1)$; the field is $K_0 = \mathbb{Q}_\ell(i)$. Hypothesis (B4) was checked in exact arithmetic: the series $\exp(\Delta_1\beta) \circ L$ has coefficients integral at ℓ up to total degree 84 for $\ell = 3$, 72 for $\ell = 7$ and 144 for $\ell = 11$. That (B) holds in all degrees was *not* proved; the test has a sketch of an argument, which was not written out as a proof. On this β, Lemmas 3 and 4 were checked up to degree 800, and equality holds in (III.88) at exceptional monomials: there the valuation of the coefficient is exactly $\frac{1}{\ell-1} - |n|\rho - 1$. So the tested coefficients attain the loss of one unit. This shows that the term $\mathrm{ex}(n)$ is sharp for the finite part of the series that was verified; sharpness under (B) in all degrees would follow if (B4) were proved for this example. Theorem J was checked on this β in exact arithmetic, for minors of size one, with $\ell = 3, 7, 11, 19$: the bound holds, both with B and with B', in each of 1 041 934 comparisons, of which 147 296 have $B > 0$ and 266 970 have $B' > 0$. It also holds in every case computed with minors of size two at one site. These checks use β in degrees beyond those to which (B4) was checked.

Comparison with §I.4.3. The description there is (III.83). The weight w_h is $\mathrm{cap}(s_0 + h)$: the number of minors at the place reached by the hth power of the Frobenius composed with conjugation. It enters in two steps. First, an exceptional monomial of β links the top row at s_0 to a row at $s_0 + h$, by (III.88); only such a link forces a radius, by (III.89). Second, a monomial of a product of minors has degree at most cap in one row; this is the statement of §I.4.3 that a minor absorbs one derivative in a linked direction. Lemma 6 then balances the radius of the top row against the radii of the linked rows, and the result is $(A/\ell^h - w_h)_+$ at each odd h. The labels with (III.98) express that at each h only one row at the place $s_0 + h$ is linked. The congruence of Part II, §6.3 is a statement of the same kind as (III.98). The source does not use it by name at this point: see §6 (ii).

6. Assumed here, from the source

(i) *The form* (III.80) *of the theta kernels, and hypothesis* (B). In the Lean this is the statement `Loc.AdmB`, of the file `ProductZerothFirstParts.lean`. It says that there are a unit e and a series β such that the kernel, in the coordinates $e \cdot L$, equals $U \cdot \exp(\pm \sum \beta) \circ L$ with U integral (clause `B0`; also `OrbitThetaExpansion`), with coefficients in the fraction field of the Witt ring (`B1`), torus invariance (`B2`), the condition on degrees (`B3`), and the integrality (B4) (`B4`). The hypothesis `hServ.pdv` (`SelectionServed`) of the final theorem does not contain `Loc.AdmB` literally: it gives `BeOK` (B1–B4 for each orbit representative) and `ThetaExpansionAllPlaces`. How these give `Loc.AdmB` was not read. Nothing of it is proved here. The passage from `Loc.AdmB` to `BHyp` (`tieB`) was not read.

(ii) *The frames.* Theorem J needs a labelling and (III.98). In the Lean the theorem for all frames at once is `ThmAGauss`: every coefficient of the polynomial `frameJet` in the frame variables has valuation $\geq B$, if $2\log_\ell \bar{a} + 2 \leq N_0 - 1$. Its proof (`thmAGauss_of`) names `residuallyGenericExists`: there are frames whose entries are residually independent units (`IsGeneric`). Three further inputs are supplied to it as hypotheses, in `thmAGauss`: a transfer of the hypotheses between fields (`admissibleTransfer`, of type `TransferB`), the evaluation of the frame polynomial (`FrameJetEval`, `GaussPow`), and the existence of labels (`labelsExist_statement`, of type `LabelsExist`). The statement `TransferB` contains no frame: it transports `Unram`, `BHyp`, the inverse g of the logarithm and the integrality of U along a norm-preserving homomorphism of fields. The proofs of these inputs were not read. The congruence `key_minor_dvd` of Part II, §6.3 is not used by name in this step: no file under

`Formalism/GaussJetBounds` mentions it. Its only use by name is in its own file; that file is imported by `Loc4ResOfLine` and `LiftSubstStart`, in the directory `CoarseMultiplicity`, which were not read.

(iii) *From Taylor coefficients to the statement used.* The comparison of $F(D)$ and $\mathrm{top}(s)$ with the weights of the construction (`normCor7`; `split_bound_eleven_one`, conditional on `Loc.AdmB`); the factorial count `StarArith` and the inequality `BCore`, which is of the form (6.1) of Part II; the identification of sites with places and of ν_1 with the weights. Then the loss terms lo_G, fl_G of (III.77); the lattice clauses of `LinkG_at`; `stLoss7`; `placesF_min'`. For `stLoss7` and `placesF_min'` only the statements were read. On the route from `LinkG` to the final theorem that was read, the clauses about valuations of kernels are proved for the zero chain (`linkG_without_kernels`); how `ThmAOp` enters the final theorem was not read. The dependency computation shows that `thmAOp` is among the declarations the final theorem depends on. By occurrences of names it reaches the final theorem through `thmAGauss`, `thmA_of`, `split_bound_eleven_one` and `sum_over_primes_service`, that is, through `SelectionServed`.

(iv) *The other primes.* The primes of class `inertClass` are at least 13 for large x (`thirteen_le`, which has a hypothesis that x is large and uses the assumed splitting of the small primes, `split_small`). The primes of the other classes, and the primes above X'', are treated by other clauses, not discussed here.

What this section proves, and what it assumes. *Proved outright:* Lemma 1 (the inverse of the logarithm), Lemma 2 (integrality), Lemma 6 (the optimisation for the top row), weak duality (III.95), and the closed form (III.97) for orbits of length two. *Proved conditionally on hypothesis* (B)*:* Lemmas 3 and 4, Proposition 5, and Theorem J for $d = 2$. *Conditional on* (B) *and on strong duality, which is cited from the source* (`strongDuality`): Theorem J for $d \geq 4$; this case was not tested end to end. *Taken from the source without proof:* the transcription of the top-level inequality `LinkG_at` in §1; that the theta kernels satisfy (B) (item (i)); the dictionary between sites and places; the frames and labels (item (ii)); the passage from jets to lattices (item (iii)); the remaining primes (item (iv)). Theorem J is proved without the operator on the ordinary rows that the Lean statement `ThmAOp` allows.

Relation to the declarations the final theorem depends on. By the dependency computation, the following declarations named or used in this section are among them: `strongDuality`, `weightQuantization`, `lemmaS_of`, `thmAOp`, `binomial_coefficient_extraction`, `lpF_eq_Fnum`, `residuallyGenericExists`, `theorem_A_generic` and `gauss_sum_powers`. The four statements `TopRadius`, `top_radius_holds`, `LemmaR` and `ThmABound` are in the archive but are not used by the final theorem. So Lemma 6 and the weighted bound for one Taylor coefficient (Proposition 5) are proved here independently of the route of the final theorem, and the declarations the final theorem uses instead are `asmA_top` and `lemma_R_log_loss`. Their statements are consistent with Lemma 6 and Proposition 5: for a finite cut-off, and for the set of all odd $j < J$, `asmA_top` is (III.93) with $\mathrm{top}(s)$ moved to the other side, and its proof takes the same radius (for an infinite cut-off it is the remark after Lemma 6; see §III.6, §4), and Proposition 5 is the stronger of the two bounds for a coefficient.

What remains to be written. The passage from the estimate of this section to lattices is taken up in §III.6; some of the items below are treated there, as that section states.

(i) A proof of strong duality for $d \geq 4$ (in the Lean: compactness and a separation theorem); see Remark (2) after Theorem J.

(ii) Theorem J with the operator on the ordinary rows, and the deduction of the statement for all frames (`ThmAGauss`); the role of the congruence of Part II, §6.3.

(iii) Hypothesis (B) for the theta kernels of the construction: the height-two formal group at an inert prime, the series β, and the integrality (B4). This is the arithmetic input of the whole section.

(iv) The passage from Theorem J to (III.79): items (iii) and (iv) above, with the loss terms.

(v) A proof that the series β of the example after Theorem J satisfies (B) in all degrees, and an example for $d \geq 4$ with an exceptional monomial; none is known to us. For $d \geq 4$ Lemmas 3–4 and Theorem J were tested only on a series without exceptional monomials and on the combinatorial steps.

(vi) To follow, in the source, how the statement for one Taylor coefficient on the route of the final theorem, `lemma_R_log_loss`, is used in place of Proposition 5. (The statements `TopRadius`, `top_radius_holds`, `LemmaR` and `ThmABound` of the archive are not used by it.)

III.6 Divisibility: from derivatives to lattices

This section concerns the passage from the local estimate of §III.5 to the prime sum $P_{u,t,j}$ of (4.1) of Part II, the left side of the main inequality (M). In §III.5 one function at one prime is studied. What (M) needs is a statement about lattices over a number field, at all primes at once, with the losses added up. The section states that passage, proves its elementary parts, and names the statements of the source that carry the arithmetic. For the prime sum it does this much: it splits the sum and bounds several of its terms. The estimates for the eleven loss sums and their total bound (III.138) are quoted. Theorem III.3.2 alone does not give that bound. The section does not prove the divisibility side of (M). The divisibility of the period pairing is assumed, as statement (Val) of §2. Also assumed are a statement on lattices, the inequality (III.79) at the primes of one of the two classes, the estimates for the eleven loss sums with their total bound, and the slope inequality for many maps.

Notation. The notation of §III.5, §1 is kept: the scale $x, \varepsilon, b, a, b', X', X'', p_0, \delta_1$; the parameters m, k, t_0, n, f, N_m; the set $\mathcal{T}$; the numbers a_t, gain, τ, wPart, $V(j,t,p)$, lo_G, fl_G, Sc, η, and the constant C of (III.77). So a is the number xb, as in §III.4 and §III.5, with $a \leq a_t$. And $\tau(w)$ is the number in (III.79); it is not the modulus τ of §III.4. The letter u is the translate of Part II, §4.2, as in $P_{u,t,j}$. The letters introduced here are listed at the end of §1.

The letters are local. The letters of this section have the meanings given here in this section only. Other sections of Part III use some of them differently, and some objects have another name there. The main cases are these. τ is a weight place in §III.7. $\mathcal{T}$ is the level in §III.4 and §III.7, and $\mathcal{T}(\mathsf{w})$ in §III.8 is the level prime of a word, which is Lv_α here. The members α are the words w of §III.8, where α is a constant. U is a power series in §III.5 and a space on the target side in §III.7. γ is $\ell/(\ell^2-1)$ in §III.5 and in the paragraph on `asmA_top` of §4, and Euler's constant in §6. ka_j is the number ς of §III.4, and c_{χ_d} is the c_d of §III.4 and §III.7. λ_w is written $\mathrm{wt}(y)$ in §III.7, and the λ_{w} of §III.8 is another number. $\mu_{0,\alpha}$ is not the μ_0 of §III.7 or of §III.9. Ξ is an operator in §III.7 and §III.8. L_0 is the field called K in §III.9, where L_0 is another field.

Proved outright. Three lemmas on lattices over the ring of integers of a number field: the slope of a sublattice (Lemma L1), the slope of a lattice multiplied by a fractional ideal, with the cost of rounding (Lemma L2), and the integrality of a pairing after such a multiplication (Lemma L3). The bound for the slope when the rank drops (Lemma L5). The identity between the

prime sum (4.1) and sums over single primes (Lemma L6). Three elementary comparisons of losses (Lemmas L7 to L9). An inequality of the shape of (III.79) at the primes of class `signal`, for the explicit formulas of the source and with explicit loss terms (Proposition L10); it gives (III.79) itself when the constants satisfy the conditions (III.130). The splitting of the prime sum by classes of primes (Proposition L11) and three of the estimates that bound its terms (Lemmas L12 to L14; the last uses Theorem III.3.2 and one cited estimate).

Proved from two statements of the source. Theorem L4. Given the valuation statement (Val) of §2, the glued lattice pairs integrally with the target. Given the lattice statement (Lat) of §2, its slope exceeds that of the source lattice by the total gain, minus explicit losses.

A model, not a statement about the source. Proposition L15 shows how the terms cancel when there is one injective map between spaces of equal dimension. That hypothesis does not hold in the source.

Taken from the source without proof. The statement (Val), that is, the divisibility of the period pairing of the chain of maps. In the Lean it rests on three assumptions (`adm2`, the hypotheses `Tied` and `CutUp` of `val2AllAtRF`, and the field `l1b`; see §2). It is where Theorem J of §III.5 is used, and it is conditional on hypothesis (B) there. The statement (Lat). The inequality (III.79) at the primes of class `inertClass`. The estimates for the eleven sums over primes of the losses, and their total bound (III.138). The slope inequality for many maps (Part II, §6.9–6.10). These are listed in §7.

1. Lattices over a number field: three lemmas

Places and weights. Let L_0 be a number field of degree n_0, totally complex, with ring of integers $\mathfrak{o}$. (In the source L_0 is the coefficient field, `LL`.) A *place* w is a nonzero prime ideal $\mathfrak{p}_w$ of $\mathfrak{o}$; p_w is its residue characteristic, e_w its ramification index over p_w, v_w the valuation with $v_w(\mathfrak{p}_w) = 1$, and $\mathfrak{o}_w$ the localisation of $\mathfrak{o}$ at $\mathfrak{p}_w$, a discrete valuation ring. For an $\mathfrak{o}_w$-submodule Λ_w of an L_0-vector space and an integer z (which may be negative) we write $\mathfrak{p}_w^z\Lambda_w$ for $(\mathfrak{p}_w\mathfrak{o}_w)^z\Lambda_w$. As in Part II, §6.4, put

$$\lambda_w = \frac{\log \mathrm{Norm}(\mathfrak{p}_w)}{n_0}, \qquad \sum_{w|p} \lambda_w e_w = \log p, \qquad \mathrm{sz}_w(\sigma) = 2\lambda_w e_w\, \sigma \quad (\sigma \in \mathbb{R}). \tag{III.99}$$

The middle identity holds because the sum over $w \mid p$ of e_w times the residue degree of w is n_0. So a gain of σ at every place above p, measured in units sz_w, has total $2\sigma \log p$. (Lean: `lw`, `SizeData₂.sz`, `FactEtaMean`. The Lean states `FactEtaMean` for a field that is Galois over $\mathbb{Q}$. The identity holds for every number field, by the argument just given.)

Lattices and slopes. Let U be an L_0-vector space of finite dimension $\dim U \geq 1$, and fix a Haar measure on the real vector space $U \otimes_{\mathbb{Q}} \mathbb{R}$. A *lattice* in U is a finitely generated $\mathfrak{o}$-submodule Λ that spans U over L_0. Its rank is $\mathrm{rk}\,\Lambda = \dim U$; it is a discrete subgroup of $U \otimes \mathbb{R}$ of rank $n_0 \dim U$, with a covolume. The slope divides by the rank. So wherever a slope is taken, in Lemmas L1, L2 and L5 and in Theorem L4, the space has positive dimension and the lattice has rank ≥ 1. Put

$$\mu(\Lambda) = -\frac{2}{n_0\,\mathrm{rk}\,\Lambda}\,\log \mathrm{covol}(\Lambda) + \mathrm{cst}_{L_0}, \qquad \mathrm{cst}_{L_0} = -\log 2 + \frac{\log|\mathrm{disc}\, L_0|}{n_0}. \tag{III.100}$$

This is the slope in the normalisation for squared norms: twice the slope of Part II, §6.8, up to the constant. (Lean: `muDef`, with the measure given by Hermitian forms at the complex places. In Lemmas L1–L5 only differences $\mu(\Lambda) - \mu(\Lambda')$ for lattices in the same space occur; they do not depend on the measure or on cst_{L_0}.) For a finite $\mathfrak{o}$-module write $\lg_w$ of it for the length of

its localisation at w: the number of factors $\mathfrak{o}/\mathfrak{p}_w$ in a composition series. For a lattice Λ write $\Lambda_w = \mathfrak{o}_w\Lambda \subset U$. Here Λ always denotes a lattice; the number Λ_G of (III.78) is not used in this section.

Lemma L1 (a sublattice). *Let $\Lambda' \subset \Lambda$ be lattices in U, with $\dim U \geq 1$. Then*

$$[\Lambda : \Lambda'] = \prod_w \mathrm{Norm}(\mathfrak{p}_w)^{\lg_w(\Lambda/\Lambda')}, \qquad \mu(\Lambda') = \mu(\Lambda) - \frac{2}{\mathrm{rk}\,\Lambda}\sum_w \lambda_w \lg_w(\Lambda/\Lambda'). \tag{III.101}$$

Proof. The quotient Λ/Λ' is a finite $\mathfrak{o}$-module. It is the direct sum of its localisations, and each composition factor $\mathfrak{o}/\mathfrak{p}_w$ has $\mathrm{Norm}(\mathfrak{p}_w)$ elements. This gives the index. Then $\mathrm{covol}(\Lambda') = [\Lambda : \Lambda']\,\mathrm{covol}(\Lambda)$ and $\log \mathrm{Norm}(\mathfrak{p}_w) = n_0\lambda_w$. □

(Lean: `muDef_index`, `mu_index_sum`.)

Lemma L2 (multiplication by an ideal). *Let Λ be a lattice in U, with $\dim U \geq 1$, let z_w be integers, zero for almost all w, and put $I = \prod_w \mathfrak{p}_w^{-z_w}$, a fractional ideal.*

(i) *$I\Lambda$ is a lattice with $(I\Lambda)_w = \mathfrak{p}_w^{-z_w}\Lambda_w$ for every w, and it is the only lattice with these localisations.*

(ii) *$\mu(I\Lambda) = \mu(\Lambda) + 2\sum_w \lambda_w z_w$.*

(iii) *Let σ_w be real numbers, zero for almost all w, let $z_w = \lfloor \sigma_w \rfloor$, and let* Rd *be a finite set of primes containing p_w for every w with $\sigma_w \notin \mathbb{Z}$. Then*

$$2\sum_w \lambda_w\sigma_w - 2\sum_{p\in \mathrm{Rd}} \log p \;\leq\; \mu(I\Lambda) - \mu(\Lambda) \;\leq\; 2\sum_w \lambda_w\sigma_w. \tag{III.102}$$

Proof. (i) Localisation commutes with products, and $I\mathfrak{o}_w = \mathfrak{p}_w^{-z_w}\mathfrak{o}_w$. A lattice is determined by its localisations: if $y \in U$ lies in Λ_w for all w, the ideal $\{\beta \in \mathfrak{o} : \beta y \in \Lambda\}$ is contained in no $\mathfrak{p}_w$, so it is $\mathfrak{o}$ and $y \in \Lambda$. (ii) Suppose first that all $z_w \leq 0$, so that $I \subset \mathfrak{o}$ and $I\Lambda \subset \Lambda$. The module Λ_w is finitely generated and torsion-free over the discrete valuation ring $\mathfrak{o}_w$, hence free of rank $\mathrm{rk}\,\Lambda$. So the localisation of $\Lambda/I\Lambda$ at w is a sum of $\mathrm{rk}\,\Lambda$ copies of $\mathfrak{o}_w/\mathfrak{p}_w^{-z_w}\mathfrak{o}_w$, of length $-z_w\,\mathrm{rk}\,\Lambda$, and Lemma L1 gives (ii). In general write $I = I_1I_2^{-1}$ with ideals $I_1, I_2 \subset \mathfrak{o}$. Then $I_2 \cdot (I\Lambda) = I_1\Lambda$. The case just proved, applied to $(I_2, I\Lambda)$ and to (I_1, Λ), gives $\mu(I\Lambda) - 2\sum_w \lambda_w v_w(I_2) = \mu(\Lambda) - 2\sum_w \lambda_w v_w(I_1)$, and $v_w(I_1) - v_w(I_2) = -z_w$. (iii) By (ii) the middle term is $2\sum_w \lambda_w z_w$. Now $0 \leq \sigma_w - z_w < 1$, and $\sigma_w - z_w = 0$ unless $p_w \in \mathrm{Rd}$. Finally $\sum_{w|p}\lambda_w \leq \sum_{w|p}\lambda_w e_w = \log p$. □

(Lean: `ShiftEx`, `ShiftMu` with its proof `shift_slope_mu`, `RoundingCost`; the ideal is `shiftIdeal`, the lattice $I\Lambda$ is `shiftLat`. The factor 2 in (ii) is the squared normalisation.) The lemma and its proof hold for any number field. For example over $\mathbb{Q}$, where $\lambda_p = \log p$: for $\Lambda = \mathbb{Z}^2$ and $I = \frac{1}{2}\mathbb{Z}$ the covolume of $I\Lambda$ is $\frac{1}{4}$, and the slope (III.100) rises by $2\log 2$.

Lemma L3 (integrality after the multiplication). *Let U, U' be L_0-vector spaces and $\langle\ ,\ \rangle : U \times U' \to L_0$ a bilinear map. Let $\Lambda \subset U$ and $\mathcal{R} \subset U'$ be lattices. For each place w let $\mathrm{Gen}_w \subset U'$ be a set that generates $\mathcal{R}_w$ over $\mathfrak{o}_w$, and let z_w be integers, zero for almost all w, such that*

$$\langle y, y'\rangle = 0 \quad \text{or} \quad v_w(\langle y, y'\rangle) \geq z_w \qquad \text{for all } y \in \Lambda,\ y' \in \mathrm{Gen}_w. \tag{III.103}$$

Let $I = \prod_w \mathfrak{p}_w^{-z_w}$. Then $\langle \tilde{y}, \tilde{y}'\rangle \in \mathfrak{o}$ for all $\tilde{y} \in I\Lambda$ and $\tilde{y}' \in \mathcal{R}$. If moreover $v_{w_0}(\langle y, y'\rangle) = z_{w_0}$ for some place w_0, some $y \in \Lambda$ and some $y' \in \mathrm{Gen}_{w_0}$, then $\mathfrak{p}_{w_0}^{-1}I\Lambda$ does not have this property.

Proof. An element of L_0 lies in $\mathfrak{o}$ if its valuation is ≥ 0 at every place. Fix w. The element $\tilde{y}$ is a finite sum of products βy with $\beta \in I$ and $y \in \Lambda$. The element $\tilde{y}'$ is a finite sum of products $\beta' y'$ with $\beta' \in \mathfrak{o}_w$ and $y' \in \mathrm{Gen}_w$. So $\langle \tilde{y}, \tilde{y}'\rangle$ is a sum of terms $\beta\beta'\langle y, y'\rangle$, and each term is zero

or has valuation at least $-z_w + 0 + z_w = 0$ at w. For the second assertion take $\beta \in \mathfrak{p}_{w_0}^{-1} I$ with $v_{w_0}(\beta) = -z_{w_0} - 1$, and $\beta' \in \mathfrak{o} \setminus \mathfrak{p}_{w_0}$ with $\beta' y' \in \mathcal{R}$. Then $v_{w_0}(\langle \beta y, \beta' y' \rangle) = -1$. □

No slope occurs in Lemma L3, and it holds also for spaces of dimension 0. (Lean: `ShiftPairing`, `DivIntoDualShift`. There the hypothesis is stated with real numbers and integers at most their ceilings, which is the same since valuations are integers; and the local forms differ from $\langle\ ,\ \rangle$ by units at w, which changes nothing.)

Remarks. (1) *Places are independent.* By Lemma L2 (i) the localisation of $I\Lambda$ at w depends only on z_w, and by (ii) the gains add over the places. This holds also for two places above the same prime. It is the form over a number field of the gluing over $\mathbb{Q}$ of Part II, §6.8, where the Chinese remainder theorem was used. (2) *Maps compose.* Suppose that there are m maps and $m+1$ lattices, and that for $1 \le j \le m$ the jth map sends the localisation at w of the lattice before it into $\mathfrak{p}_w^{z_{w,j}}$ times the localisation of the lattice after it. Then the composite sends the first localisation into $\mathfrak{p}_w^{z_{w,1}+\cdots+z_{w,m}}$ times the last. So the exponents of the m maps of the chain add. This is the integrality assertion of Part II, §6.2. (3) *A sublattice costs.* If $\Lambda'_w \subset \mathfrak{p}_w^z \Lambda_w$ with an integer $z \ge 0$, then $\lg_w(\Lambda/\Lambda') \ge z \operatorname{rk} \Lambda$, and by Lemma L1 the slope drops by at least $2\lambda_w z$ at w.

Checks. The three lemmas were tested in exact arithmetic over $\mathbb{Q}(\sqrt{-5})$, whose class number is 2, with the ramified prime above 2, the two primes above 3 and above 7, and the inert prime 11; lattices of rank 1 to 3, not all free. In 60 cases the two sides of Lemma L2 (ii) agree to $2 \cdot 10^{-15}$; with $\sum \lambda_w z_w$ in place of $2 \sum \lambda_w z_w$ they do not. In 25 cases the index formula of Lemma L1 holds exactly, the lengths being computed place by place. In 12 cases the pairing of Lemma L3 is integral on $I\Lambda$ and is not on $\mathfrak{p}_{w_0}^{-1} I\Lambda$.

Letters introduced in this section. The field L_0, $\mathfrak{o}$, n_0; a place w with $\mathfrak{p}_w, p_w, e_w, v_w, \mathfrak{o}_w$; λ_w, sz_w; spaces U, U'; a lattice Λ, rk, μ, cst_{L_0}, $\lg_w$; integers z_w, real numbers σ_w, an ideal I, a set of primes Rd; elements $y, \tilde{y}$ of U and $y', \tilde{y}'$ of U', and scalars β, β' in L_0; a member α, with U_α, $\Lambda_\alpha^{\mathrm{src}}$, $\Lambda_\alpha^{\mathrm{div}}$, $\Lambda_\alpha^{\mathrm{gl}}$, I_α, Lv_α, rk_α^*, Rd_α; the lattice $\mathcal{R}$, the sets Gen_w; $\mathrm{sg}_\alpha(w)$, $\phi(j,w)$, $\mathrm{err}(\alpha,j,w)$, $\mathrm{io}(w)$; G_{top}, F_0, $E_{\mathrm{err}}(\alpha)$, $\mathrm{Lo}(\alpha)$; the group H, h_p, n_p, an exponent i of a prime, cell, wt_j, V_∞, an orbit $\mathcal{Q}$, $\mathcal{T}(\mathcal{Q})$, $n(\mathcal{Q})$, $\kappa(\mathcal{Q})$, $\mathrm{Srv}(p)$, ka_j, $\mathrm{gain}^{\mathrm{plain}}$, $\mathrm{gain}^{\mathrm{srv}}$, Pr, Ex, dl, C_1, C_2, dig_p, the constants C_{pl}, C_{lv}, C_{den} of §2 and C_{r} of §5, odd_p, real numbers A, c, c_i, ψ_i, counts n_{var}, n_{st}, n_{top}; sets of primes Ξ, Ξ_{s}, Ξ_1, Ξ_{c}; $\mathrm{def}(p)$, $\mathcal{T}_p$, $\mathrm{Sig}(p)$, bw, θ, a', Y; and in §6 Per_α, ht_α and the error terms $E_{\mathrm{size}}, E_{\mathrm{lev}}, E_{\mathrm{cnt}}, E_{\mathrm{src}}$. The number of digits of an integer in base p, written $\lambda(k)$ in §III.5, is written $\mathrm{dig}_p(k)$ here, because λ_w is the weight of a place. The level of a member, written $\mathcal{T}$ in Part II, §6.7, is written Lv_α here, because $\mathcal{T}$ is the set of labels, as in §III.5.

2. The glued lattice: statement and proof

The data. Fix x and one choice of the data of the construction (a "record" of the source). The record contains one translate $u \in H$ (Lean: the field `Host.u0`). It is the translate u of Part II, §4.2, and the weights of the record are the weights for this u (§4 below). The maps of the record are indexed by a finite set. Its elements α are the *members* (Lean: `Word`). A member is not a translate. All members of a record have the same translate, and the prime sum $P_{u,t,j}$ does not depend on the member (the Lean statement `Loss_at` contains no member). A member enters through its lattices, its level and its error terms. For each member α the source gives:

- an L_0-vector space U_α of dimension ≥ 1 with a lattice $\Lambda_\alpha^{\mathrm{src}}$, the *source lattice* (Lean: `MramL`; it is the preimage of the first lattice of the chain of maps; the positive dimension is the field `a_pos` of the inputs of `divClausesR_of`, §3);
- a lattice $\Lambda_\alpha^{\mathrm{div}} \subset \Lambda_\alpha^{\mathrm{src}}$ (Lean: `NuLat`);

- a space U' and a lattice $\mathcal{R} \subset U'$, the same for all α (Lean: `RR`), and a bilinear pairing $\langle\ ,\ \rangle_\alpha : U_\alpha \times U' \to L_0$, the period pairing of the whole chain (Lean: `perCol`; the bilinear form is `col.B`);
- a prime Lv_α, the level of the member (the level $\mathcal{T}$ of Part II, §6.7), and an integer $\mathrm{rk}^*_\alpha \geq \mathrm{rk}\,\Lambda^{\mathrm{src}}_\alpha$ (Lean: `duS`).

Independently of α except where shown, there are real numbers, zero outside a finite set of places: $\mathrm{gain}(j,w) \geq 0$, $\phi(j,w)$, $\mathrm{err}(\alpha,j,w)$, $\tau(w) \geq 0$, $\mathrm{io}(w) \geq 0$, for $1 \leq j \leq m$ (Lean: the fields `gain`, `F`, `E`, `tau`, `io` of $\mathtt{SizeData}_2$). Here gain and τ are the numbers of (III.79). Put

$$\mathrm{sg}_\alpha(w) = \sum_{j=1}^{m} \big(\mathrm{gain}(j,w) + \phi(j,w) - \mathrm{err}(\alpha,j,w)\big), \tag{III.104}$$

$$G_{\mathrm{top}} = \sum_w \mathrm{sz}_w\Big(\sum_j \mathrm{gain}(j,w) - \tau(w)\Big), \quad F_0 = \sum_{j,w} \mathrm{sz}_w\big(\phi(j,w)\big), \quad E_{\mathrm{err}}(\alpha) = \sum_{j,w} \mathrm{sz}_w\big(\mathrm{err}(\alpha,j,w)\big). \tag{III.105}$$

(Lean: `sg`, `Ga`, `Fo`, `Ee`. In the Lean the gain is first set to zero at the places above a set of "dropped" primes, see §5; we assume this done.) So

$$\sum_w \mathrm{sz}_w\big(\mathrm{sg}_\alpha(w) - \tau(w)\big) = G_{\mathrm{top}} + F_0 - E_{\mathrm{err}}(\alpha). \tag{III.106}$$

The number F_0 is the F_0 of Part II, §6.7. The *glued lattice* is

$$\Lambda^{\mathrm{gl}}_\alpha = I_\alpha\, \Lambda^{\mathrm{div}}_\alpha, \qquad I_\alpha = \prod_w \mathfrak{p}_w^{-\lfloor e_w\, \mathrm{sg}_\alpha(w)\rfloor}. \tag{III.107}$$

(Lean: `Nu`, `Iu`, and the statement `NprimeDef`, which is (III.107).) The factor e_w converts an exponent of p_w into an exponent of $\mathfrak{p}_w$.

The two statements taken from the source.

(Val) *Valuation.* For every member α and every place w there is a set $\mathrm{Gen}_w \subset U'$ generating $\mathcal{R}_w$ over $\mathfrak{o}_w$ with

$$\langle y, y'\rangle_\alpha = 0 \quad \text{or} \quad v_w\big(\langle y, y'\rangle_\alpha\big) \ \geq\ \lfloor e_w\, \mathrm{sg}_\alpha(w)\rfloor \qquad (y \in \Lambda^{\mathrm{div}}_\alpha,\ y' \in \mathrm{Gen}_w). \tag{III.108}$$

(Lat) *Lattices.* For every member α and every place w with $p_w \neq \mathrm{Lv}_\alpha$: $\tau(w)$ is an integer, $(\Lambda^{\mathrm{div}}_\alpha)_w \subset \mathfrak{p}_w^{e_w\tau(w)}(\Lambda^{\mathrm{src}}_\alpha)_w$, and

$$\lg_w\,(\Lambda^{\mathrm{src}}_\alpha/\Lambda^{\mathrm{div}}_\alpha) \ \leq\ \mathrm{rk}^*_\alpha\, e_w\,\big(\tau(w) + \mathrm{io}(w)\big). \tag{III.109}$$

Note what (Lat) does not give. It has rk^*_α, not $\mathrm{rk}\,\Lambda^{\mathrm{src}}_\alpha$, on the right of (III.109). And it says nothing at the places above Lv_α.

The statement (Val) *in the Lean.* It is not one Lean statement. It is assumed here; in the Lean it is obtained from three assumptions.

(1) `LocalQuadraticParts.adm2`. Let μ_w be the twist chosen at w (`muAt`). The letters of this item are those of the Lean: μ_w, with a place as index, is a twist and not a slope; the element F below has nothing to do with the number F_0; and $\widetilde{M}$, V_0 have nothing to do with $M_{u,t,j}$ and V. The statement is: for every member α, every place w, every $y \in \Lambda^{\mathrm{div}}_\alpha$, and every F that lies both in the lattice $\widetilde{M}(\mu_w)$ (`Mtilde`) and in the subspace $V_0(\mu_w)$ (`V0`) of the ambient

space of the twist, there is an element of L_0 whose image under the fixed embedding of L_0 in $\mathbb{C}$ (`embLL`) is the period of y against F; and this element is zero or has valuation at least $\lfloor e_w \mathrm{sg}_\alpha(w) \rfloor$ at w. Both restrictions on F are part of the statement. It says nothing for an element of the lattice outside $V_0(\mu_w)$, and nothing for an element of $V_0(\mu_w)$ outside the lattice. That the period is the embedded value of an element of L_0 is also part of the statement; (Val) presupposes it.

(2) The hypotheses of the theorem `val2AllAtRF`. This theorem turns (1) into the same bound for the matrix form of the pairing, on a lattice `Nu` given in coordinates. Besides (1) it has three hypotheses. Two are carried by structures. `ColBrF.Tied` says that the matrix form is tied to the period: the twist at w is the one of the matrix form, each element of the lattice of the matrix form maps into V_0 and is an integral combination of the generators of $\widetilde{M}$, and the matrix pairing has the period as its embedded value. `CutUp`, for every twist that is used, says that an element of V_0 that is an integral combination of the generators of $\widetilde{M}$ is an integral combination of elements of $\widetilde{M} \cap V_0$. This is what the two restrictions in (1) require. The third hypothesis is the compatibility with the source of (1): every element of `Nu` is mapped by the coordinate map `eNu` into `NuLat`, that is, into $\Lambda_\alpha^{\mathrm{div}}$. It is not an assumption. Where `val2AllAtRF` is applied (in the files `RecBridge.lean` and `RecBridgeTied.lean`) it is supplied by the theorem `nu_into`, from the definition (III.107) (`NprimeDef`) and two identities of lattices: `shift`, a field of the package, and `lat`, a field of the structure `HLatTie`, which the package carries for each member as its field `hNu`.

(3) The structure field `l1b` (of type `PairingFactorsThroughPeriods`). It says that the local form at w equals a unit at w times $\langle y, y' \rangle_\alpha$, where y' is the image of the element of the local lattice under a fixed linear map.

That the sets Gen_w generate $\mathcal{R}_w$ is not an assumption. The set Gen_w is the image of the local lattice of (1). The lattice $\mathcal{R}$ is *defined* as the lattice glued from these images (`RR`, by `glueLat`), and the theorem `latticeGlue` proves that this is a lattice with these localisations (Lemma L2 (i) is the case of an ideal). The passage to one bilinear pairing is faithful: the Lean theorem `intoDual` applies `DivIntoDualShift` (Lemma L3) with one L_0-bilinear form. The quantifiers of (Val) are those of the Lean: all members, all places, all $y \in \Lambda_\alpha^{\mathrm{div}}$. The final theorem depends on `adm2` (the dependency computation confirms this). It also depends on `val2AllAtRF` and on `latticeGlue`: in the seventh run of the dependency computation both are marked as used. Whether the hypotheses `Tied` and `CutUp` of (2) and the field of (3) are discharged for the construction on the route of the final theorem was not determined. The dependency computation does not say how hypotheses are discharged, and the proof terms were not followed that far.

The statement (Lat) *in the Lean.* It is the part of the first clause of `LinkG_at` that is used here, read with `NuDef`. That clause is for the places w with $p_w \neq \mathrm{Lv}_\alpha$ and has three parts. The first is the inclusion, with the exponent $\lfloor e_w \tau(w) \rfloor$. The second is $\texttt{lenMN}(\alpha, w) \leq \mathrm{rk}_\alpha^* (\tau(w) + \mathrm{io}(w))$. The third says: if the set `VV` of the Lean at the prime p_w is empty, then $\tau(w) = 0$, $\mathrm{io}(w) = 0$ and $\texttt{lenMN}(\alpha, w) = 0$. The third part is not used in Theorem L4. The number `lenMN` is not the local length. The last clause of `NuDef` says that $\mathrm{lg}_w(\Lambda_\alpha^{\mathrm{src}} / \Lambda_\alpha^{\mathrm{div}}) = \texttt{lenMN}(\alpha, w) \cdot e_w$. With the second part this gives (III.109), and it is where the factor e_w there comes from. For the integrality of $\tau(w)$ two statements are needed. The theorem `exists_tau_eq_natCast` says that the number given by the formula `tauP` is a natural number. It does not mention the τ of the record. That $\tau(w)$ equals `tauP` at p_w is part of the hypothesis `fills` of §3.

Theorem L4 (the glued lattice). *Let α be a member. Write* $\mathrm{rk} = \mathrm{rk}\, \Lambda_\alpha^{\mathrm{src}}$ *and* $\mathrm{rk}^* = \mathrm{rk}_\alpha^*$*; so* $\mathrm{rk}^* \geq \mathrm{rk} \geq 1$.

(i) *Assume* (Val). *Then* $\langle y, y'\rangle_\alpha \in \mathfrak{o}$ *for all* $y \in \Lambda_\alpha^{\mathrm{gl}}$ *and* $y' \in \mathcal{R}$. *That is, the pairing maps* $\Lambda_\alpha^{\mathrm{gl}}$ *into the dual lattice* $\mathcal{R}^\vee = \{\xi \in \mathrm{Hom}_{L_0}(U', L_0) : \xi(\mathcal{R}) \subset \mathfrak{o}\}$.

(ii) *Without any assumption,*

$$\mu(\Lambda_\alpha^{\mathrm{gl}}) = \mu(\Lambda_\alpha^{\mathrm{src}}) + \sum_w 2\lambda_w \lfloor e_w \,\mathrm{sg}_\alpha(w)\rfloor - \frac{2}{\mathrm{rk}} \sum_w \lambda_w \,\mathrm{lg}_w\,(\Lambda_\alpha^{\mathrm{src}}/\Lambda_\alpha^{\mathrm{div}}). \tag{III.110}$$

(iii) *Assume* (Lat). *Let* Rd_α *be the set of primes* p_w *with* $e_w \,\mathrm{sg}_\alpha(w) \notin \mathbb{Z}$. *Then*

$$\mu(\Lambda_\alpha^{\mathrm{gl}}) \;\geq\; \mu(\Lambda_\alpha^{\mathrm{src}}) + G_{\mathrm{top}} + F_0 - E_{\mathrm{err}}(\alpha) - \mathrm{Lo}(\alpha), \tag{III.111}$$

where

$$\begin{aligned}\mathrm{Lo}(\alpha) = &\sum_{w:\ p_w \neq \mathrm{Lv}_\alpha} \mathrm{sz}_w\Big(\frac{\mathrm{rk}^*\,\mathrm{io}(w) + (\mathrm{rk}^* - \mathrm{rk})\,\tau(w)}{\mathrm{rk}}\Big) \\ &+ \sum_{w:\ p_w = \mathrm{Lv}_\alpha} \Big(\frac{2}{\mathrm{rk}}\,\lambda_w\,\mathrm{lg}_w\,(\Lambda_\alpha^{\mathrm{src}}/\Lambda_\alpha^{\mathrm{div}}) - \mathrm{sz}_w(\tau(w))\Big) + 2\sum_{p \in \mathrm{Rd}_\alpha} \log p.\end{aligned} \tag{III.112}$$

The three terms of $\mathrm{Lo}(\alpha)$ are: the loss of the lattice $\Lambda_\alpha^{\mathrm{div}}$ beyond τ, with the cost of the rank $\mathrm{rk}^* \geq \mathrm{rk}$; the places above the level, where (Lat) gives no bound and the length is kept as it is; and the rounding. If $\mathrm{rk}^* = \mathrm{rk}$ the first term is the sum of $\mathrm{sz}_w(\mathrm{io}(w))$ over the places w with $p_w \neq \mathrm{Lv}_\alpha$.

Proof. (i) This is Lemma L3 with $\Lambda = \Lambda_\alpha^{\mathrm{div}}$ and $z_w = \lfloor e_w \mathrm{sg}_\alpha(w)\rfloor$; the hypothesis (III.103) is (III.108).

(ii) Lemma L2 (ii) for the pair $(I_\alpha, \Lambda_\alpha^{\mathrm{div}})$ gives $\mu(\Lambda_\alpha^{\mathrm{gl}}) = \mu(\Lambda_\alpha^{\mathrm{div}}) + \sum_w 2\lambda_w \lfloor e_w \mathrm{sg}_\alpha(w)\rfloor$. Lemma L1 for $\Lambda_\alpha^{\mathrm{div}} \subset \Lambda_\alpha^{\mathrm{src}}$ gives $\mu(\Lambda_\alpha^{\mathrm{div}}) = \mu(\Lambda_\alpha^{\mathrm{src}}) - \frac{2}{\mathrm{rk}} \sum_w \lambda_w \,\mathrm{lg}_w(\Lambda_\alpha^{\mathrm{src}}/\Lambda_\alpha^{\mathrm{div}})$.

(iii) We bound the two sums in (III.110). *The first sum.* Apply Lemma L2 (iii) with $\sigma_w = e_w \,\mathrm{sg}_\alpha(w)$ and $\mathrm{Rd} = \mathrm{Rd}_\alpha$. These numbers vanish outside a finite set of places, and $2\lambda_w\sigma_w = \mathrm{sz}_w(\mathrm{sg}_\alpha(w))$. So

$$\sum_w 2\lambda_w \lfloor e_w \,\mathrm{sg}_\alpha(w)\rfloor \;\geq\; \sum_w \mathrm{sz}_w(\mathrm{sg}_\alpha(w)) - 2\sum_{p \in \mathrm{Rd}_\alpha} \log p. \tag{III.113}$$

The second sum, at a place w *with* $p_w \neq \mathrm{Lv}_\alpha$. By (III.109) and Lemma L5 below (whose proof is one line and uses nothing of this theorem),

$$\frac{2}{\mathrm{rk}}\,\lambda_w\,\mathrm{lg}_w\,(\Lambda_\alpha^{\mathrm{src}}/\Lambda_\alpha^{\mathrm{div}}) \;\leq\; \mathrm{sz}_w(\tau(w)) + \mathrm{sz}_w\Big(\frac{\mathrm{rk}^*\,\mathrm{io}(w) + (\mathrm{rk}^* - \mathrm{rk})\,\tau(w)}{\mathrm{rk}}\Big). \tag{III.114}$$

The second sum, at a place w *with* $p_w = \mathrm{Lv}_\alpha$. Here nothing is assumed. We write the term as $\mathrm{sz}_w(\tau(w))$ plus the difference $\frac{2}{\mathrm{rk}}\lambda_w \,\mathrm{lg}_w(\Lambda_\alpha^{\mathrm{src}}/\Lambda_\alpha^{\mathrm{div}}) - \mathrm{sz}_w(\tau(w))$. *Conclusion.* Insert (III.113), (III.114) and the last identity into (III.110). All sums are finite. The result is

$$\mu(\Lambda_\alpha^{\mathrm{gl}}) \;\geq\; \mu(\Lambda_\alpha^{\mathrm{src}}) + \sum_w \mathrm{sz}_w\big(\mathrm{sg}_\alpha(w) - \tau(w)\big) - \mathrm{Lo}(\alpha), \tag{III.115}$$

with $\mathrm{Lo}(\alpha)$ as in (III.112). By (III.106) this is (III.111). □

Of (Lat) the proof of (iii) uses only (III.109). The inclusion $(\Lambda_\alpha^{\mathrm{div}})_w \subset \mathfrak{p}_w^{e_w\tau(w)}(\Lambda_\alpha^{\mathrm{src}})_w$ explains the signs. The gain $\mathrm{gain}(j, w)$ enters $\Lambda_\alpha^{\mathrm{gl}}$ with a plus sign and $\tau(w)$ with a minus sign for the following reason. The divisibility (III.108) holds on $\Lambda_\alpha^{\mathrm{div}}$, not on $\Lambda_\alpha^{\mathrm{src}}$. By the inclusion and Remark (3), $\Lambda_\alpha^{\mathrm{div}}$ is smaller than $\Lambda_\alpha^{\mathrm{src}}$ by at least $\tau(w)$ at w, in the units sz_w. The net exponent at w is $\sum_j \mathrm{gain}(j, w) - \tau(w)$: the right side of (III.79).

Lemma L5 (when the rank drops). *Let $\Lambda' \subset \Lambda$ be lattices, w a place, $\mathrm{rk} = \mathrm{rk}\,\Lambda \geq 1$, and let $\mathrm{rk}^* \geq \mathrm{rk}$, $\tau(w) \geq 0$ and $\mathrm{io}(w) \geq 0$ be real numbers with $\lg_w(\Lambda/\Lambda') \leq \mathrm{rk}^* e_w(\tau(w) + \mathrm{io}(w))$. Then*

$$\frac{2}{\mathrm{rk}}\,\lambda_w\,\lg_w(\Lambda/\Lambda') - \mathrm{sz}_w(\tau(w)) \;\leq\; \mathrm{sz}_w\Big(\frac{\mathrm{rk}^*\,\mathrm{io}(w) + (\mathrm{rk}^* - \mathrm{rk})\,\tau(w)}{\mathrm{rk}}\Big). \tag{III.116}$$

Proof. The left side is at most $2\lambda_w e_w(\frac{\mathrm{rk}^*}{\mathrm{rk}}(\tau(w)+\mathrm{io}(w)) - \tau(w))$, and $\frac{\mathrm{rk}^*}{\mathrm{rk}}(\tau(w)+\mathrm{io}(w)) - \tau(w) = \frac{\mathrm{rk}^*\mathrm{io}(w) + (\mathrm{rk}^* - \mathrm{rk})\tau(w)}{\mathrm{rk}}$. □

(Lean: `ClaimLow`. There the rank deficit is a number `de` with $\mathrm{rk} = (1 - \mathtt{de})\,\mathrm{rk}^*$, and the right side of (III.116) reads $\mathrm{sz}_w((\mathrm{io}(w) + \mathtt{de}\,\tau(w))/(1 - \mathtt{de}))$, which is the same number.)

Checks. Theorem L4 was tested in exact arithmetic over $\mathbb{Q}(\sqrt{-5})$ on 200 lattices of rank 2, with rk^* from 2 to 5, with the level prime 2, 3, 7 or 11, with a sublattice that is smaller at a place above the level, and with every hypothesis of (Lat) verified on the data. The identity (III.110) holds exactly in all cases, and (III.111) holds in all cases. In 60 cases the data were chosen with equality in (III.109) and with $e_w \mathrm{sg}_\alpha(w)$ integral; then (III.111) is an equality, exactly, in all 60. On these 60 cases three weaker forms of $\mathrm{Lo}(\alpha)$ were tried. With $\sum_w \mathrm{sz}_w(\mathrm{io}(w))$ in place of the first two terms (the loss one would get from (III.109) with rk in place of rk^* at all places) the inequality fails in 48. Without the term at the level it fails in 15, which are all the cases where that term is positive. Without the factor for the rank it fails in 46, which are all the cases where $\mathrm{rk}^* > \mathrm{rk}$ and $\tau + \mathrm{io}$ is not zero. On the other 140 cases, without the rounding term it fails in 15. Lemma L5 was also tested on 20000 random data in the form of the Lean statement `ClaimLow`; without the factor for the rank the inequality fails in a third of the cases.

The Lean statement, and what it compares with. The Lean statement of the slope bound (`SlopesN_at`, proved from `Places_at` by `lower_bound_simplified`) is not (III.111) word for word. It reads

$$\mu(\Lambda_\alpha^{\mathrm{gl}}) \;\geq\; \mu_{0,\alpha} + G_{\mathrm{top}} + F_0 - E_{\mathrm{err}}(\alpha) - \Big(C_{\mathrm{pl}} \sum_w \lambda_w e_w \mathrm{io}(w) + C_{\mathrm{pl}}\,|\mathcal{T}|\,m\Big(\frac{a}{k} + b'\Big)\log \mathrm{Lv}_\alpha + C_{\mathrm{pl}}\,X'' + 1\Big), \tag{III.117}$$

where C_{pl} is a constant of the source (it is not the constant C of (III.77)), and $\mu_{0,\alpha}$ is the slope of the *first lattice of the chain*, not of $\Lambda_\alpha^{\mathrm{src}}$, which is its preimage. It comes with an upper bound of the same shape for the largest slope of a sublattice of $\Lambda_\alpha^{\mathrm{gl}}$. The passage from (III.111) to (III.117) uses four statements of the source, which are assumed here (Lean: `adm26`, `HypLow2`, `places_at_of`).

(a) *The rank.* $\mathrm{rk} = (1 - 1/(\mathrm{Lv}_\alpha - 1))^f\,\mathrm{rk}^*$ (`adm26`). Hence $(\mathrm{rk}^* - \mathrm{rk})/\mathrm{rk}^* \leq 2f/\mathrm{Lv}_\alpha$, which the source bounds by $\frac{1}{2}$. So $\mathrm{rk}^*/\mathrm{rk} \leq 2$, and the part of the first term of (III.112) with io is at most $4\sum_w \lambda_w e_w \mathrm{io}(w)$. This factor is absorbed in the constant C_{pl} of (III.117); it is not part of the "+1".

(b) *The places above the level.* There $\tau(w) = 0$; this is the hypothesis `htauT` of `places_at_of`, not a clause of `adm26`. The second term of (III.112) is then at most $C_{\mathrm{pl}}|\mathcal{T}|m(a/k + b')\log \mathrm{Lv}_\alpha$ (the last clause of `adm26`). This is the second term of (III.117).

(c) *The rounding.* $2\sum_{p \in \mathrm{Rd}_\alpha} \log p \leq C_{\mathrm{pl}} X''$. The theorem `places_at_of` does not prove this: it has the bound as a hypothesis (`NprimeMuR`, the statement that $\mu(\Lambda_\alpha^{\mathrm{gl}}) - \mu(\Lambda_\alpha^{\mathrm{div}})$ differs from $\sum_w \mathrm{sz}_w(\mathrm{sg}_\alpha(w))$ by at most $C_{\mathrm{pl}}X''$). The source proves that hypothesis elsewhere (`round_le`, `nprimeMu_at`, `nprimeMuR_at`, `linkMu`). There the rounding is counted at every place with $\mathrm{sg}_\alpha(w) \neq 0$; the primes below these places include Rd_α. The set Rd_α is not contained in $[1, X'']$. By two statements of the source (`ZeroAbove` for gain and err, the field `F_above` for ϕ; together `sg_eq_zero`), $\mathrm{sg}_\alpha(w) = 0$ if $p_w > X''$, unless p_w is Lv_α, g_0, g_1, d or a prime factor

of the denominator D_α of §3. So

$$\sum_{p\in \mathrm{Rd}_\alpha} \log p \;\le\; X''\log 4 + \log \mathrm{Lv}_\alpha + \log g_0 + \log g_1 + \log d + \log D_\alpha, \tag{III.118}$$

by Chebyshev's bound for the primes up to X'' (`sum_log_primes_le`), and because the sum of $\log p$ over the prime factors of D_α is at most $\log D_\alpha$ (`sum_log_primeFactors_le`). This display is the statement `nprimeMu_at`. The source then uses that each of $\log g_0, \log g_1, \log d$ is at most X'' (from the size conditions `SizeG`, `SizeQD`), and two further hypotheses on the family: $\log \mathrm{Lv}_\alpha \le C_{\mathrm{lv}} X''$ and $\log D_\alpha \le C_{\mathrm{den}} X''$ (`MuTu`, `MuDGam`). This gives the constant $2(\log 4 + 3 + C_{\mathrm{lv}} + C_{\mathrm{den}})$ in `NprimeMuR` (`nprimeMuR_at`, `linkMu`). It also assumes that the glued lattices are finitely generated and have positive rank (`MuFG`, `MuRank`). These hypotheses are assumed here. The names of this item were not put to the dependency computation.

(d) *The first lattice of the chain, and the "+1".* The source has $\mu(\Lambda_\alpha^{\mathrm{src}}) = \mu_{0,\alpha} - f\log(\mathrm{Lv}_\alpha - 1)/(\mathrm{Lv}_\alpha - 2)$ exactly (`adm26`), and $\sum_w \mathrm{sz}_w(\tau(w)) \le b^2$ (a hypothesis of `places_at_of`; the source proves it as `ta_le`). With (a), the part of the first term of (III.112) with τ is at most $4fb^2/\mathrm{Lv}_\alpha$. The source requires $4fb^2/\mathrm{Lv}_\alpha + f\log(\mathrm{Lv}_\alpha - 1)/(\mathrm{Lv}_\alpha - 2) \le 1$ (`HypLow2`). So the cost of the rank on τ and the change from $\mu(\Lambda_\alpha^{\mathrm{src}})$ to $\mu_{0,\alpha}$ are together at most 1. This is the single "+1" of (III.117). There is one such term, not two.

3. What `LinkG` is, and where the valuation statement lives

The statement `LinkG` of the source has five clauses (`LinkG_at`). They are: (1) the lattice clause, of which (Lat) is the part used in §2; (2) a bound for the auxiliary numbers `Sv`; (3) the inequality (III.79) of §III.5, with $0 \le \mathrm{gain}(j,w) \le 4|\mathcal{T}|a/p_w$, at the places of class `signal` or `inertClass` with $p_w \le X''$; (4) a clause that holds by definition; (5) a bound for the mean of certain numbers $\kappa(\mathcal{Q})$, used in §5. None of them mentions a theta function, a Taylor coefficient, or a pairing. As §III.5 already said, (III.79) is an inequality between the numbers $\mathrm{gain}, \tau, \mathrm{wPart}$. A hypothesis on the family (the field `fills`, of type `ParametersFill`) says that these numbers are given by explicit formulas (`gainP`, `tauP`, `fill`, `partner`, `TTP`); they are recalled in §4. With that hypothesis, clause (3) is arithmetic about the formulas. Proposition L10 below carries this out at the primes of class `signal`, with its own loss terms.

The zero chain. The proof of `LinkG` in the source (`linkG_without_kernels`) goes through an intermediate statement `admG1`, one part of which does speak of the valuation of a chain of maps on a local lattice (`vKge`). On the route to the final theorem that part is applied to the chain in which every map is zero (`zeroCh`). The valuation of 0 is infinite, so the part holds for a trivial reason (`vKge_zero`). So `LinkG` is proved without any information on the theta kernels. This is consistent: `LinkG` is the bookkeeping, and the lattice clause (1).

Where the divisibility is asserted. It is the statement (Val). The final theorem depends on `LocalQuadraticParts.adm2`. The route, read downward from the place where it is consumed, is this.

(a) `divClausesR_of` takes a structure of inputs, `DivInputsR`, and produces four clauses. One field of `DivInputsR` is `val2`, that is (III.108). The other fields are independent of it. From `val2` comes the first clause only: the integrality of Theorem L4 (i), by `intoDual` and `divIntoDualShift` (Lemma L3), which use the field `l1b`. The second clause bounds the squares of the operator norms of the pairing at the complex places by numbers whose mean logarithm is the height of the record minus a constant. It comes from the height inputs (the fields `ch`, `hn`, `nP`, `nKcb`, `eig6`). The third clause is the inequality of sizes, first line of

(6.5) of Part II, with its term $-F_0$. It comes from the size inputs (among them `L3main`, `Qb`, `deQ`, `errB`, and the inequality `constant_ge_twenty_eight`). The fourth clause is the slope identity of Lemma L2 (ii), by `shiftMu`; it uses that `Nu` is a lattice of positive rank (`Nu_lat`, `a_pos`). So the valuation statement does not imply the height bound or the inequality of sizes. The inequality of sizes is the subject of §III.4.

(b) `val2AllAtRF` deduces `val2` from `adm2`, by linearity, under the hypotheses described in §2. In the seventh run of the dependency computation `val2AllAtRF` is marked as used by the final theorem. The archive has a second name, `DivisionPackage.val2`. The query for it returned an empty list. This is inconclusive: it does not show which fully qualified name was intended, and it does not show that such a declaration is unused.

(c) `admissible_level_two` proves `adm2`, place by place. At the places above Lv_α and above the two auxiliary level primes g_0, g_1 of Part II, §6.7 it uses separate inputs. At all other places it uses the statement `Conclusion1` of the file `WeilJacquetShalikaLaw.lean`: *for every y in the local lattice at w that defines $\Lambda_\alpha^{\mathrm{div}}$, the image of y under the composite of the m maps of the chain is divisible, in the last lattice of the chain, by $\mathfrak{p}_w$ to the power $e_w\big(\sum_j(\mathrm{gain}(j,w)+\phi(j,w)-\mathrm{err}(\alpha,j,w))+\mathtt{sp}\cdot v_{p_w}(D_\alpha)\big)$.* Here D_α is an integer denominator of the member, v_{p_w} is the p_w-adic valuation, and `sp` is a real parameter of the Lean statement. On the route `sp` is 1: the statement `Serv` contains `Conclusion1` with this value. The term with D_α does not reach (III.108). The theorem `period_geTU`, which passes from the chain to the period, lowers the exponent by exactly $v_{p_w}(D_\alpha)$ (`vDG`), and the two cancel. This is the statement about kernels.

(d) `Conclusion1` is part of `ServF`, which `sum_over_primes_service` proves from the structure `SumOverPrimesHypotheses`, which has the fields `twU` and `pdv`, a bound for the denominators D_α (`DGamSize`), and the theorem `split_bound_eleven_one`. The latter is the local estimate: it rests on `thmAOp`, the Lean form of Theorem J of §III.5. The hypothesis `pdv` contains hypothesis (B) of §III.5 for the kernels of the construction.

So there are two branches. One (`LinkG`) says what the numbers gain and τ are and that they dominate the sums V; it needs no kernel. The other (`ServF`, then (Val)) says that the kernels are divisible by these numbers; it needs Theorem J and hypothesis (B), and it is assumed here. They meet in Theorem L4 and in §5.

4. From Theorem J to a gain at a place

Cells, and sums over one prime. Let H be the group of Part II, §4.1, of order n. It is the group of §III.3, and it is written additively here, as there and as in the Lean; $\mathcal{T} \subset H$. A *cell* is a pair $(0,t)$ or $(1,t)$ with $t \in H$; the first entry is the *row*. There are $f = 2n$ cells. For an odd prime p let $h_p \in H$ be its class, of order n_p. The classes of primes are those of the Lean (`primeClass`): a prime $p \nmid 2qd$ is `splitK` if $\chi_d(p) = 1$; if $\chi_d(p) = -1$ it is `signal` when $\chi_q(p) = 1$ and `inertClass` when $\chi_q(p) = -1$. For an odd prime $p \neq q$, $t \in H$ and an integer $i \geq 1$ put

$$\mathrm{cell}(p,t,i) = \Big(\frac{1+\chi_q(p)^i}{2},\ t + i\,h_p\Big). \tag{III.119}$$

This is the cell of the prime power p^i in the table of Part II, §4.1, for the column t. For odd i the row is 1 if p is `signal` and 0 if p is `inertClass`. (Lean, for odd i: `partner`.) For each j there are weights $\mathrm{wt}_j > 0$ on the cells (Lean: `fill`); these are the weights of Part II, §4.2, for the translate u

of the record and the map j. The hypothesis `fills` says $\mathrm{wPart}(j,t,p,i) = \mathrm{wt}_j(\mathrm{cell}(p,t,i))$. Put

$$V_\infty(j,t,p) = \sum_{i \text{ odd}} \Big(\frac{a_t}{p^i} - \mathrm{wt}_j\big(\mathrm{cell}(p,t,i)\big) \Big)_+ , \tag{III.120}$$

the sum being over all odd $i \geq 1$. The sum $V(j,t,p)$ of (III.76) has the same terms, but only those with $p^{i-1} \leq a_t$. *If all the weights are* ≥ 1, *then* $V_\infty = V$: the terms with $p^i > a_t$ vanish, and (III.76) contains every i with $p^i \leq a_t$. In the source the weights that occur are ≥ 1 (§III.5, §1). Without this hypothesis the two sums can differ; see the example below.

Lemma L6 (the prime sum). *Let the weights be any positive numbers. The sum* (4.1) *of Part II, for these weights, is*

$$P_{u,t,j} = \sum_{\substack{p \text{ } \texttt{signal} \\ \text{or } \texttt{inertClass}}} 2\log p \; V_\infty(j,t,p) \; + \; \log d \sum_{i \geq 1} \Big(\frac{a_t}{d^i} - \mathrm{wt}_j\big(\mathrm{cell}(d,t,i)\big) \Big)_+ . \tag{III.121}$$

Proof. In (4.1) the coefficient of p^i is $(1 - \chi_d(p^i)) \log p$. For $p \neq d$ it is $2\log p$ if $\chi_d(p) = -1$ and i is odd, and 0 otherwise. The row and the column of p^i are those of (III.119). The prime d gives the last term, with coefficient $\log d$ for every i. The primes 2 and q are absent, and split primes have coefficient 0. □

In the running example of Part II, §4.2 ($q = 3$, $d = 23$, $n = 4$, $a_t = 10000$) the weights are 0.05 and 10. Computing the right side of (III.121) gives 95461.6156 for the translate [1], the value of Part II, and the four values for the four translates agree with (4.1) to 10^{-10}. The example uses V_∞, not V: its weights 0.05 are below 1, and with V in place of V_∞ the right side of (III.121) is smaller by 0.309064 for the translate [1] (the powers 5^7 and 11^5 are lost). With the weights 10 and 1 the difference is 0. The primes of class `inertClass` contribute 51234.4133 for every translate: their odd powers lie in row 0, which contains no distinguished cell. The primes of class `signal` contribute 42833.7633 for the translate [1]. The distinguished weights carry the information on the zeros. Among the unramified primes, only `signal` primes sample a distinguished weight; this is the reason for the name. The powers of d contribute the separate last term of Lemma L6, and that term does depend on the translate. Here $23^2 \equiv 1$ modulo 3 and modulo 16, so the cell of 23^2 for the column t is $(1,t)$, in row 1, and $10000/23^2 > 10$. The last term is 1393.4390 for the translate [1] and 1424.6372 for the translates [3], [5], [7]. The difference is $(10 - 0.05)\log 23$.

The formulas for the gain. From here on the weights are ≥ 1, as in the source, so that $V_\infty = V$. Let p be `signal` or `inertClass`. The group generated by h_p (by $2h_p$ if p is `signal`) acts on H by translation. For an orbit $\mathcal{Q}$ write $\mathcal{T}(\mathcal{Q}) = \mathcal{T} \cap \mathcal{Q}$. (The Lean fixes a base point on each orbit, `reps`.) Put $\mathrm{ka}_j = \frac{1}{2}(k + (2j-1)(t_0+1))$, with the integer parameter t_0 of §III.5. The source has $0 \leq \mathrm{ka}_j \leq \mathrm{Sc}$ for $1 \leq j \leq m$ (`kaJ_nonneg` for the lower bound, `kaJ_le` for the upper bound). For an orbit $\mathcal{Q}$ put

$$\mathrm{gain}^{\mathrm{plain}}(j,\mathcal{Q}) = \sum_{t \in \mathcal{T}(\mathcal{Q})} \sum_{\substack{i \text{ odd} \\ i \notin \mathrm{Ex}(p,t)}} \Big(\frac{a_t - 8\,\mathrm{Sc}}{p^i} - \mathrm{wt}_j\big(\mathrm{cell}(p,t,i)\big) + \mathrm{ka}_j - 4\,\mathrm{Sc} \Big)_+ - \mathrm{dl}(\mathcal{Q}), \tag{III.122}$$

$$\mathrm{dl}(\mathcal{Q}) = C_1\, n(\mathcal{Q})\, N_m^2 \log_p(2a) + C_2\, |\mathcal{T}(\mathcal{Q})| \, \big(\log_p(2a)\big)^2 . \tag{III.123}$$

Here $n(\mathcal{Q})$ is the number of *sites* of the orbit. For an `inertClass` prime the sites are the points of the orbit, and $n(\mathcal{Q}) = n_p$. For a `signal` prime a site is a pair consisting of the parity of an exponent $i \geq 0$ and the column $t + i\,h_p$, with $t \in \mathcal{Q}$; there are two parities and, for each, as many columns as $\mathcal{Q}$ has points, so $n(\mathcal{Q})$ is twice the order of $2h_p$. (Lean: the sites are the values of

`partner`, and $n(\mathcal{Q})$ is `nP`. The sites are not the cells (III.119) of the powers of p: the two agree for odd i only.) The numbers $C_1, C_2 \geq 0$ are constants, and $\mathrm{Ex}(p,t)$ is the set of odd i for which $\mathrm{cell}(p,t,i) = (0,t')$ with $t' \in \mathcal{T}$, $t' \neq t$, and i is the least positive exponent with this property for t'. (Lean: `g0`, `dl`, `Excl`, `al`, `capU`, `kaJ`; the number $4\,\mathrm{Sc}$ is `cG`.) The gain is

$$\begin{aligned}\mathrm{gain}(j,p) = &\sum_{\mathcal{Q}\in \mathrm{Srv}(p)} \max\left(\mathrm{gain}^{\mathrm{srv}}(j,\mathcal{Q}) - \Pr(j,\mathcal{Q}),\, \mathrm{gain}^{\mathrm{plain}}(j,\mathcal{Q}),\, 0\right)\\ &+ \sum_{\mathcal{Q}\notin \mathrm{Srv}(p)} \max\left(\mathrm{gain}^{\mathrm{plain}}(j,\mathcal{Q}), 0\right)\end{aligned} \tag{III.124}$$

for $p \leq X''$, and $\mathrm{gain}(j,p) = 0$ for $p > X''$ and for `inertClass` primes with $h_p = 0$. Here $\mathrm{Srv}(p)$ is the set of *served* orbits, defined in §5; it is empty unless p is `inertClass` and $p < X'$. The numbers $\mathrm{gain}^{\mathrm{srv}}$ and Pr are further explicit expressions (Lean: `gU`, `Pr`): $\mathrm{gain}^{\mathrm{srv}}$ is (III.122) without the exclusion of Ex and with the full $(a_t - 8\,\mathrm{Sc})\,p/(p^2-1)$ for the points t at even positions on the orbit, by the identity (6.1) of Part II. By `fills`, $\mathrm{gain}(j,w) = \mathrm{gain}(j,p_w)$ and $\tau(w)$ depends only on p_w; we write $\tau(p)$ for it.

Comparison with Theorem J. Theorem J of §III.5 is applied to one orbit: the sites are the sites of the orbit just defined, the top sites are the sites whose column lies in $\mathcal{T}(\mathcal{Q})$ (for a `signal` prime: those of even parity among them), the number $a_t - 8\,\mathrm{Sc}$ plays the role of the length of the first row at the top site t, and $\mathrm{wt}_j - \mathrm{ka}_j + 4\,\mathrm{Sc}$ that of the capacity at a site. The term top of (III.82) then has the shape of the double sum in (III.122), and the last term of (III.82) has the shape of the first term of (III.123) (Lemma L8). This dictionary, the treatment of the excluded exponents, the other terms of (III.82), and the passage from one function to the lattices of the chain are in the source (`normCor7`, `split_bound_eleven_one`, `orbit_entry_bounds_hold`, `lift_bounds_of_essential`, then `WeilJacquetShalikaLaw.Proof.stmt_proved`); they are assumed here.

The statement `asmA_top`. The final theorem uses this statement for the top row (it is marked as used in the seventh run of the dependency computation). It was read and compared with Lemma 6 of §III.5. In this paragraph s is a site, ℓ is the prime, and ν_1, cap, top, d_{top} are as in §III.5, §1; i is an odd integer. The Lean statement is: for every site s and every finite set of odd integers $i \geq 1$ there is a real number $\omega \geq 0$, with $\omega \geq \ell^{-d_{\mathrm{top}}(s)}$ if $d_{\mathrm{top}}(s)$ is finite, such that

$$\mathrm{top}(s) + \nu_1(s)\,\omega + \sum_{i \text{ in the set}} \mathrm{cap}(s+i)\left(1 - \ell^i\omega\right)_+ \;\leq\; \nu_1(s)\,\frac{\ell}{\ell^2-1}. \tag{III.125}$$

It relates to Lemma 6 of §III.5 as follows.

(a) *If $d_{\mathrm{top}}(s)$ is finite*, take the set of all odd $i < d_{\mathrm{top}}(s)$. Then (III.125) is (III.93) of that lemma, with $A = \nu_1(s)$, $c_i = \mathrm{cap}(s+i)$ and with $d_{\mathrm{top}}(s)$ as the odd bound J of the lemma: the same statement. Here $\mathrm{top}(s)$ is the whole of the second line of (III.82): the sum over odd $h < d_{\mathrm{top}}(s)$ and the sum of $\nu_1(s)/\ell^h$ over odd $h > d_{\mathrm{top}}(s)$ (the Lean's `topTerm` has both). It is the right side of (III.93), and $\nu_1(s)\,\ell/(\ell^2-1) = \gamma A$, with $\gamma = \ell/(\ell^2-1)$ as in §III.5. So (III.125) is (III.93) with $\mathrm{top}(s)$ moved to the left. Both assert only that some ω exists; neither holds for every $\omega \geq \ell^{-d_{\mathrm{top}}(s)}$. For example, for $\nu_1(s) = 9$, $\ell = 3$, $d_{\mathrm{top}}(s) = 1$ one has $\mathrm{top}(s) = \frac{3}{8}$ and $\gamma A = \frac{27}{8}$; both hold with equality at $\omega = \frac{1}{3}$ and both fail at $\omega = \frac{3}{8}$. A larger set gives nothing more, because the terms with $i \geq d_{\mathrm{top}}(s)$ vanish when $\omega \geq \ell^{-d_{\mathrm{top}}(s)}$. The Lean proof takes the same radius as the proof of the lemma: $\omega = \ell^{-i_0}$, where i_0 is the least odd $i < d_{\mathrm{top}}(s)$ with $\mathrm{cap}(s+i) \geq \nu_1(s)/\ell^i$, and $i_0 = d_{\mathrm{top}}(s)$ if there is none.

(b) *If $d_{\mathrm{top}}(s) = \infty$*, the Lean statement is not Lemma 6, which has a finite bound. It is the remark after Lemma 6. The Lean allows any finite set of odd i; the remark has the restriction $\ell^i \leq \nu_1(s)$, and the Lean does not.

(c) The Lean has ℓ any integer ≥ 2, and ν_1, cap natural numbers.

(Tested: 25600 exact cases for (a), 30000 for the radius; and 40000 further exact cases in which the two inequalities were evaluated from the definitions at the radius of the lemma and at an arbitrary ω, and agreed.)

The statement `lemma_R_log_loss`. The final theorem also uses this statement (seventh run). It is the bound for one Taylor coefficient on the route of the final theorem. Its comparison with Proposition 5 is in §III.5, in the paragraph on the declarations the final theorem depends on: Proposition 5 is the stronger of the two bounds for a coefficient. The formal statement (`LemmaRA`, in the file `GaussJetBounds/StatementG.lean`) is for a multi-index whose entries are all at most an integer $a \geq 1$, and it assumes a test on the pure monomials below the multi-index. Its exponent has the explicit loss $|\iota|\,(\log_\ell a + 1)$, where ι is the finite index set of the variables and ℓ is the prime. (In this sentence a is the bound of the Lean statement.)

We now prove what is elementary: three comparisons of losses, and then an inequality of the shape of (III.79) at the `signal` primes.

Lemma L7 (shifted positive parts). *Let $p \geq 2$, $A \geq 0$, $c \geq 0$, and for each odd $i \geq 1$ let $\psi_i \geq 1$ and $c_i \geq 0$ be real numbers. Let $\mathrm{odd}_p(A)$ be the number of odd $i \geq 1$ with $p^i \leq A$. Then*

$$\sum_{i\ \mathrm{odd}} \Big(\frac{A-2c}{p^i} - \psi_i + c_i - c\Big)_+ \ \geq\ \sum_{i\ \mathrm{odd}} \Big(\frac{A}{p^i} - \psi_i\Big)_+ - c\Big(1+\frac{2}{p}\Big)\mathrm{odd}_p(A). \tag{III.126}$$

The left side may be $+\infty$. If $A \geq 1$, then $\mathrm{odd}_p(A) \leq \frac{1}{2}(\log_p A + 1)$.

Proof. For each i, $\frac{A-2c}{p^i} - \psi_i + c_i - c \geq \frac{A}{p^i} - \psi_i - c(1+\frac{2}{p^i})$. The positive part is increasing, and subtracting a number ≥ 0 lowers a positive part by at most that number. If $p^i > A$ then $A/p^i - \psi_i < 0$, so the term on the right is 0 and nothing is lost. The odd i with $p^i \leq A$ are the odd i with $1 \leq i \leq \log_p A$; for $A \geq 1$ their number is at most $\frac{1}{2}(\log_p A + 1)$. □

The bound for $\mathrm{odd}_p(A)$ needs $A \geq 1$: for $0 < A < 1/p$ the count is 0 and $\frac{1}{2}(\log_p A + 1)$ is negative. In the use below $A = a_t \geq 1$. The left side of (III.126) is finite if $c_i \leq c + \psi_i$ for all large i; this holds in the use below.

Lemma L8 (three digit losses, one bound). *Let p be a prime and $\bar a \geq p$ an integer. Then $\mathrm{dig}_p(\bar a) \leq \log_p \bar a + 1 \leq 2\log_p \bar a$, and:*

(i) *(the loss of §III.5) the sum of dig_p over any n_{var} integers in $[0, \bar a]$ is at most $n_{\mathrm{var}}\,\mathrm{dig}_p(\bar a)$; in the last term of* (III.82) *the number n_{var} of variables is the number n_{st} of sites times the two numbers N_0, N_1 of §III.5;*

(ii) *(the loss Δ of* (6.1a) *of Part II) if the index set there has n_{st} elements, its subset has n_{top} elements, the multiplicities there are r_0, r_1, and all the integers whose digits are counted are $\leq \bar a$, then $\Delta \leq n_{\mathrm{st}}\,(r_1 + r_0 r_1)\,\mathrm{dig}_p(\bar a) + n_{\mathrm{top}}\,\mathrm{dig}_p(\bar a)^2$;*

(iii) *(the loss of the Lean) if $N_0, N_1, r_0, r_1 \leq N_m$, both bounds are at most $2\,n_{\mathrm{st}}\,N_m^2\,\mathrm{dig}_p(\bar a) + n_{\mathrm{top}}\,\mathrm{dig}_p(\bar a)^2$. With $n_{\mathrm{st}} = n(\mathcal{Q})$, $n_{\mathrm{top}} = |\mathcal{T}(\mathcal{Q})|$ and $\bar a \leq 2a$ this is at most* (III.123) *with $C_1 = 4$, $C_2 = 4$.*

Proof. dig_p is increasing and $\mathrm{dig}_p(\bar a) = \lfloor \log_p \bar a \rfloor + 1$. (i) and (ii) are term-by-term bounds; for (ii) recall that Δ is a sum over the index set of r_1 digit counts and $r_0 r_1$ digit counts, plus a sum over the subset of squares of digit counts. (iii) $N_0 N_1 \leq N_m^2$, $r_1 + r_0 r_1 \leq 2N_m^2$ and $\mathrm{dig}_p(\bar a) \leq 2\log_p(2a)$. □

The paper has three digit losses: one digit count per variable (§III.5); one per factorial, with a square for the top sites (Part II, (6.1a); Lean: `StarArith`, and the loss in `gainA`); and the closed

form (III.123) per orbit. They are three different quantities. The lemma shows only that (III.123) is an upper bound for the first two. For example, with $p = 3$, $\bar{a} = 80$, two sites, $N_0 = N_1 = 2$ and $r_0 = r_1 = 1$, the largest value of the first loss is 32 and that of the second, when both sites are in the subset, is 48.

Lemma L9 (the sum over the orbits). *For every prime p of class* `signal` *or* `inertClass`, $\sum_{\mathcal{Q}} n(\mathcal{Q}) \le f$ *and* $\sum_{\mathcal{Q}} |\mathcal{T}(\mathcal{Q})| = |\mathcal{T}|$, *the sums being over the orbits. Hence*

$$\sum_{\mathcal{Q}} \mathrm{dl}(\mathcal{Q}) \;\le\; C_1\, f\, N_m^2 \log_p(2a) + C_2\, |\mathcal{T}| \big(\log_p(2a)\big)^2. \tag{III.127}$$

Proof. The orbits partition H. For `inertClass` the orbit has n_p points and $n(\mathcal{Q}) = n_p$, so the sum is $n = f/2$. For `signal` the orbits are the cosets of the group generated by $2h_p$, and $n(\mathcal{Q})$ is twice their length, so the sum is $2n = f$. □

The right side of (III.127) is the term $\mathrm{fl}_G(p)$ of (III.77), up to the constant and to $\log_p(2a) \le 2\log_p a$ (which holds when $a \ge 2$), since $|\mathcal{T}| \le n = f/2$. This is why fl_G is subtracted once for each of the m maps in (III.79), and not once for each t.

Proposition L10 (the shape of (III.79) at a `signal` prime). *Let $p \le X''$ be of class* `signal`, *let the weights be ≥ 1, and let* $\mathrm{gain}(j,p)$ *be given by* (III.124) *with* $0 \le \mathrm{ka}_j \le \mathrm{Sc}$. *Then* $\mathrm{Srv}(p) = \emptyset$, *and for* $1 \le j \le m$

$$\mathrm{gain}(j,p) \;\ge\; \sum_{t\in\mathcal{T}} \Big(V(j,t,p) - 4\,\mathrm{Sc}\Big(1+\frac{2}{p}\Big)\mathrm{odd}_p(a_t)\Big) - C_1\, f\, N_m^2 \log_p(2a) - C_2\, |\mathcal{T}| \big(\log_p(2a)\big)^2, \tag{III.128}$$

and $0 \le \mathrm{gain}(j,p) \le \frac{4}{3}\sum_{t\in\mathcal{T}} a_t/p \le 4|\mathcal{T}|a/p$. *Summing* (III.128) *over j gives, at the places w above p,*

$$\sum_{j=1}^{m} \sum_{t\in\mathcal{T}} \Big(V(j,t,p) - \mathrm{lo}'(p)\Big) - m\,\mathrm{fl}'(p) \;\le\; \sum_{j=1}^{m} \mathrm{gain}(j,w) - \tau(w), \tag{III.129}$$

with $\tau(w) = 0$, $\mathrm{lo}'(p) = 2\,\mathrm{Sc}\,(1+\frac{2}{p})(\log_p(2a)+1)$ *and* $\mathrm{fl}'(p) = C_1 f N_m^2 \log_p(2a) + C_2|\mathcal{T}|(\log_p(2a))^2$. *This is* (III.79) *with $\mathcal{T}_w = \mathcal{T}$ and with* lo′, fl′ *in place of* lo_G, fl_G. *If moreover*

$$C \ge \tfrac{20}{3}, \qquad C \ge 2C_1, \qquad C \ge 2C_2, \qquad a \ge 2, \qquad \eta \ge 0, \qquad p \le 2a, \tag{III.130}$$

then $\mathrm{lo}'(p) \le \mathrm{lo}_G(p)$ *and* $\mathrm{fl}'(p) \le \mathrm{fl}_G(p)$, *and* (III.79) *itself holds at the places above p.*

Proof. Served orbits exist only for `inertClass` primes, so $\mathrm{Srv}(p) = \emptyset$ and $\mathrm{gain}(j,p) = \sum_{\mathcal{Q}} \max(\mathrm{gain}^{\mathrm{plain}}(j,\mathcal{Q}), 0) \ge \sum_{\mathcal{Q}} \mathrm{gain}^{\mathrm{plain}}(j,\mathcal{Q})$. For a `signal` prime and odd i, $\mathrm{cell}(p,t,i)$ lies in row 1, so $\mathrm{Ex}(p,t)$ is empty. Apply Lemma L7 with $A = a_t$, $c = 4\,\mathrm{Sc}$, $\psi_i = \mathrm{wt}_j(\mathrm{cell}(p,t,i))$ and $c_i = \mathrm{ka}_j$, sum over $t \in \mathcal{T}(\mathcal{Q})$ and over the orbits $\mathcal{Q}$, and use Lemma L9. Since the weights are ≥ 1, the sum on the right of (III.126) is $V(j,t,p)$. This gives (III.128). For the upper bound, $\mathrm{wt}_j - \mathrm{ka}_j + 4\,\mathrm{Sc} \ge 0$, so each inner sum in (III.122) is at most $\sum_{i\ \mathrm{odd}} a_t/p^i = a_t p/(p^2-1) \le \frac{4}{3}a_t/p$, and $a_t \le 2a$, $\frac{8}{3} \le 4$. For (III.129): $\tau(w) = 0$ when $\mathrm{Srv}(p) = \emptyset$ (the definition of `tauP`), and $\mathrm{odd}_p(a_t) \le \frac{1}{2}(\log_p(2a)+1)$ because $1 \le a_t \le 2a$. For the last assertion, a `signal` prime is odd, so $1 + \frac{2}{p} \le \frac{5}{3}$; and $p \le 2a$, so $\log_p(2a) + 1 \le 2\log_p(2a)$. Hence $\mathrm{lo}'(p) \le \frac{20}{3}\mathrm{Sc}\log_p(2a) \le C(\mathrm{Sc}+k+\eta)\log_p(2a) = \mathrm{lo}_G(p)$, since $\mathrm{Sc}, k, \eta \ge 0$ and the terms pr, or of (III.78) vanish at a `signal` prime. For $a \ge 2$, $\log_p(2a) \le 2\log_p a$. So $C_1 f N_m^2 \log_p(2a) \le 2C_1 f N_m^2 \log_p a$, and $C_2|\mathcal{T}|(\log_p(2a))^2 \le 4C_2|\mathcal{T}|(\log_p a)^2 \le 2C_2 f(\log_p a)^2$ because $|\mathcal{T}| \le n = f/2$. □

The proposition does not show that the constants of the source satisfy (III.130). In the Lean the comparison of the losses is `lo_le_loG` and `fl_le_flG`. They say: for every real number C_*, the

Lean's own losses `lo` and `fl` with the constant C_* are at most `loG` and `flG` with the constant $4|C_*|+1$. This is the source's version of the conditions (III.130). (The sums in (III.122) are finite because the capacities are ≥ 0. Proposition L10 was tested on 3000 random toy data, with H cyclic of order 8, 16, 32, and again on 30170 exact cases: no violation, and the lower bound is attained; without the loss terms the inequality fails in nearly all cases.) For the primes of class `inertClass` the same computation applies to the orbits that are not served, but there $\mathrm{Ex}(p,t)$ is not empty and the labels on these orbits do not belong to $\mathcal{T}_w$. On the served orbits the source uses $\mathrm{gain}^{\mathrm{srv}} - \mathrm{Pr}$ and pays $\tau(w)$; these orbits satisfy the conditions (G1)–(G3) of §III.3. That case is assumed (`AuxBlock8.Pf.stmtF`, `linkG_at_core`). The terms $\mathrm{pr}(p)$ and $\mathrm{or}(p)$ of (III.78) occur only there.

5. The sum over the primes

In this subsection the weights are ≥ 1, the record and its translate u are fixed, and $\mathrm{gain}(j,w)$ and $\tau(w)$ depend only on p_w. The set $\mathcal{T}_w$ of (III.79) also depends only on p_w. From here on it is written $\mathcal{T}_p$, with $p = p_w$.

Proposition L11 (splitting). *Let Ξ be a finite set of primes of class `signal` or `inertClass`. Suppose that $V(j,t,p) = 0$ for all j, t and all primes $p \notin \Xi$ of these two classes, and that*

$$\mathrm{gain}(j,w) = 0 \quad (1 \leq j \leq m) \quad \text{and} \quad \tau(w) = 0 \qquad \text{at every place } w \text{ with } p_w \notin \Xi. \tag{III.131}$$

Put $\mathrm{def}(p) = \sum_j \sum_{t\in\mathcal{T}} V(j,t,p) - \big(\sum_j \mathrm{gain}(j,p) - \tau(p)\big)$. *Then*

$$\sum_{t\in\mathcal{T}} \sum_{j=1}^{m} P_{u,t,j} - G_{\mathrm{top}} = \sum_{p\in\Xi} 2\log p\ \mathrm{def}(p) + \sum_{t\in\mathcal{T}} \sum_{j=1}^{m} \log d \sum_{i\geq 1} \Big(\frac{a_t}{d^i} - \mathrm{wt}_j\big(\mathrm{cell}(d,t,i)\big)\Big)_+ . \tag{III.132}$$

Let $\Xi_{\mathrm{s}} \subset \Xi$ be a set of primes at which (III.79) *holds with a set $\mathcal{T}_p \subset \mathcal{T}$, and suppose* $\mathrm{gain}(j,p) = \tau(p) = 0$ *for $p \in \Xi \setminus \Xi_{\mathrm{s}}$. Then*

$$\begin{aligned} &\mathrm{def}(p) \leq \sum_j \sum_{t\in\mathcal{T}\setminus\mathcal{T}_p} V(j,t,p) + m\,|\mathcal{T}_p|\,\mathrm{lo}_G(p) + m\,\mathrm{fl}_G(p) \quad (p \in \Xi_{\mathrm{s}}), \\ &\mathrm{def}(p) = \sum_j \sum_{t\in\mathcal{T}} V(j,t,p) \qquad\qquad\qquad\qquad\qquad\qquad (p \in \Xi \setminus \Xi_{\mathrm{s}}). \end{aligned} \tag{III.133}$$

Proof. By (III.131) only the places above the primes of Ξ contribute to G_{top}, and by (III.99), $G_{\mathrm{top}} = \sum_{p\in\Xi} 2\log p\,(\sum_j \mathrm{gain}(j,p) - \tau(p))$. With Lemma L6 and $V_\infty = V$ this is (III.132). For $p \in \Xi_{\mathrm{s}}$ split $\sum_{t\in\mathcal{T}}$ into $\mathcal{T}_p$ and its complement and insert (III.79). □

The hypothesis (III.131) is needed: G_{top} is a sum over all places, and Ξ contains only primes of the two classes. In the source it holds at the places above the split primes, above 2, q, d, g_0, g_1 and above the levels. The Lean field `not_inert` gives the vanishing at every place whose prime is of neither of the two classes. That the level primes, among them g_0 and g_1, are of class `splitK` is another field (`level_split`). At g_0, g_1 the gain vanishes also by the field `gain_g`. Without it (III.132) fails: a gain of 1 at one place above a split prime changes the left side and not the right. (Lean: `SplitEq`, `SplitClasses`. Tested on 15000 random data with (III.131): no failure. Without it (III.132) fails in all 15000. If the labels outside $\mathcal{T}_p$ are forgotten, (III.133) fails.) So the prime sum exceeds G_{top} by four kinds of terms: the losses $\mathrm{lo}_G, \mathrm{fl}_G$ at the primes where (III.79) is used; the labels not served at those primes; the whole of V at the other primes; and the prime d.

Lemma L12 (one prime). *For $p \geq 2$, $V(j,t,p) \leq a_t\,p/(p^2-1) \leq \frac{4}{3}\,a_t/p$. Proof.* $\sum_{i\ \mathrm{odd}} p^{-i} = p/(p^2-1)$, and $p^2/(p^2-1) \leq \frac{4}{3}$. □

The classes of primes in the sum. The source sorts the primes of class `signal` or `inertClass` as follows (`clAt`, `NoGain`). Say that a prime p is *in a band* if $p \ge p_0$ and some power p^i lies within a factor $1 \pm \mathrm{bw}$ of X', for a small number $\mathrm{bw} > 0$ of the source. Below, C_{r} is a further constant of the source (Lean: `Cr`; it is not the constant C of (III.77)). A *candidate* is an `inertClass` prime $p < X'$ whose class h_p has order $n_p \ge x^{\varepsilon/4}$ and which is not in a band. For a candidate put $\mathrm{Sig}(p) = m \sum \kappa(\mathcal{Q})$, the sum over the *good* orbits $\mathcal{Q}$, those on which $\mathcal{T}$ satisfies (G1)–(G3) of §III.3 with the parameters there; $\kappa(\mathcal{Q}) \ge 0$ is a number attached to the orbit (`kapv`). The set of served orbits $\mathrm{Srv}(p)$ of a candidate is the set of good orbits if $\mathrm{Sig}(p) \le x^{-\varepsilon}$, and is empty otherwise; for all other primes $\mathrm{Srv}(p) = \emptyset$. The set of served labels is

$$\mathcal{T}_p = \mathcal{T} \ \text{ if } p \text{ is } \texttt{signal} \text{ or } p \ge X', \qquad \mathcal{T}_p = \{t \in \mathcal{T} : \text{ the orbit of } t \text{ is in } \mathrm{Srv}(p)\} \ \text{ otherwise.} \tag{III.134}$$

There are four classes. A prime belongs to the first class in the following list whose condition it satisfies; read in this order the list is a partition. (As conditions, "high" and "first" overlap.)

1. *high*: $p > X''$. The gain is 0. All weights are at least a/X'' (Lean: the field `off_thin` of the structure `Inst`, with $a/X'' = w''$; the other cells have larger weights, by `off_top` and `off_thick`). Since moreover $a_t \le a(1 + 1/(kx^{13}))$, V vanishes unless $p \le X''(1 + 1/(kx^{13}))$: a short interval.
2. *second*: candidates with $\mathrm{Sig}(p) > x^{-\varepsilon}$. The gain is set to 0.
3. *first*: `inertClass`, $p < X'(1 + C_{\mathrm{r}}\,\mathrm{bw})$, with $n_p < x^{\varepsilon/4}$ or p in a band; and `inertClass` with n_p odd and $n_p > 1$. The gain is set to 0.
4. *served*: all others. There (III.79) holds with $\mathcal{T}_p$ as in (III.134).

The class "served" contains the `inertClass` primes with $h_p = 0$ and $p \ge X'(1 + C_{\mathrm{r}}\,\mathrm{bw})$. Their gain is 0 by (III.124), and V vanishes there, because a_t/p is below the weights of the cells $(0, t)$, $t \in \mathcal{T}$. These weights exceed $b' = a/X'$ (Lean: the field `bT_ge` of $\mathtt{SizeData}_2$ for the label, and `off_top` of `Inst` for the rest of the weight, which is at least b'/x). The primes of the classes high, second and first are the "dropped" primes of §2. The other primes need no gain: split primes have coefficient 0 in (4.1), and so do 2 and q. Every odd prime up to $x^{\varepsilon/(5C_o)}$ splits, where C_o is the constant of the source that appears after Theorem III.3.2 (Lean: `split_small`, an assumed property of d, with `three_le_split`; the Lean writes the exponent as `ez`/5, and `ez` is ε divided by a constant `J` of the scale, so C_o is `J`). The paragraph after Theorem III.3.2 writes $C_o \ge 2$. The Lean structure of the scale has only $1 \le \mathtt{J}$ (`one_le_J`). Whether $\mathtt{J} \ge 2$ on the route was not determined; the statement that these primes split does not depend on it. So the sum starts above that bound. At the split primes, at g_0, g_1 and at the levels the numbers gain, τ, io are 0, and only ϕ and err occur. The cut-offs have this meaning: $X' = a/b'$ is where a/p falls below the weights of the cells $(0, t)$, $t \in \mathcal{T}$; X'' is where it falls below the smallest weights; below p_0 the source uses a different local argument.

Lemma L13 (a label that is not served lies on a bad orbit). *Let p be a served prime and $t \in \mathcal{T} \setminus \mathcal{T}_p$. Then p is of class `inertClass`, $p < X'$, $n_p \ge x^{\varepsilon/4}$, and the orbit of t under h_p is not good.*

Proof. By (III.134), p is `inertClass` and $p < X'$. Since p is neither first nor second, $n_p \ge x^{\varepsilon/4}$, p is not in a band, and $\mathrm{Sig}(p) \le x^{-\varepsilon}$. So p is a candidate and $\mathrm{Srv}(p)$ is the set of good orbits. □

(Lean: `unserved_bad`, which the final theorem uses by the seventh run of the dependency computation.) This is where the conditions (G1)–(G3) are used in the sum over primes. Since served orbits are good by definition, the conditions are also available in the proof of (III.79) on the served orbits (§4); that proof was not read.

Lemma L14 (the three small sums). *Let Ξ_1 be the set of* `inertClass` *primes $p < X'$ with $n_p \ge x^{\varepsilon/4}$.*

(i) *Let L_b, C_b, x_0 and L_2 be as in Theorem III.3.2, with $L_2 \le L_b$, and let $x \ge x_0$. Let $\mathcal{T}$ be the set T of that theorem, for the set of primes Ξ_1. If $x^{2\varepsilon/3} \ge 16C_b$ and $\sum_{p<X'} \frac{\log p}{p} \le 2x^{\varepsilon/3}$, then*

$$\sum_{p\in\Xi_1} \frac{\log p}{p}\, \#\{t\in\mathcal{T} :\ \text{the orbit of } t \text{ under } h_p \text{ is not good}\} \ \le\ |\mathcal{T}|\, x^{-2+\varepsilon}. \tag{III.135}$$

(ii) *Hence the labels not served cost at most $\sum_{p \text{ served}} 2\log p \sum_j \sum_{t\notin\mathcal{T}_p} V(j,t,p) \le \frac{8}{3}\, m\, a'\, |\mathcal{T}|\, x^{-2+\varepsilon}$, where $a' = \max_t a_t$.*

(iii) *Let $\theta > 0$ and $\mathrm{Sig}(p) \ge 0$ on a finite set Ξ_{c} of primes. Then*

$$2 \sum_{p\in\Xi_{\mathrm{c}},\ \mathrm{Sig}(p)>\theta} \frac{\log p}{p} \ \le\ \sum_{p\in\Xi_{\mathrm{c}}} \frac{\log p}{p}\Big(2\cdot[\mathrm{Sig}(p)>\theta] + 4\,\mathrm{Sig}(p)\Big) \ \le\ \Big(\frac{2}{\theta}+4\Big) \sum_{p\in\Xi_{\mathrm{c}}} \frac{\log p}{p}\,\mathrm{Sig}(p). \tag{III.136}$$

Here $[\,\cdot\,]$ is 1 if the condition holds and 0 otherwise. Hence, if the middle sum is at most x^{-9}, the primes of the second class cost at most $\frac{8}{3}\, m\, a'\, |\mathcal{T}|\, x^{-9}$.

Proof. (i) Theorem III.3.2 bounds the left side by $4C_b x^{-2}\delta_1 n \cdot 2x^{\varepsilon/3}$, and $\delta_1 n \le 2|\mathcal{T}|$. So it is at most $16C_b x^{-2+\varepsilon/3}|\mathcal{T}| \le x^{-2+\varepsilon}|\mathcal{T}|$. (ii) By Lemma L12 each V is at most $\frac{4}{3}a'/p$; by Lemma L13 the labels counted are those of (III.135). (iii) The first inequality is clear. For the second, $[\mathrm{Sig}(p) > \theta] \le \mathrm{Sig}(p)/\theta$. The cost of the second class is, by Lemma L12, at most $\frac{8}{3} m a' |\mathcal{T}| \sum_{\text{second}} \frac{\log p}{p}$. □

(Lean: `ChBad_at` is the inequality of Theorem III.3.2; `ChBadp_at` is (III.135); `OutD` is (ii); `lhsW`, `ChW_at`, `term_le`, `SecondD` are (iii). The bound $\sum_{p<Y} \frac{\log p}{p} \le 2\log Y$ for $Y \ge 1$ is the Lean's `mertens`; it is a form of Mertens' first theorem and is cited. By the seventh run of the dependency computation the final theorem uses `ChBadp_at`.) Since $a' \le 2xb$, the two costs are at most $6\, m\, |\mathcal{T}|\, b\, x^{-1+\varepsilon}$ and $6\, m\, |\mathcal{T}|\, b\, x^{-8}$. Let $B = |\mathcal{T}|\, m\, b\, \delta/2$ be the unit of (M), with $\delta = |T_2|/n$ as in Part II, §4.2; this δ is not the δ_1 of the scale. Since $\delta \ge x^{-\varepsilon}$, the two costs divided by B are at most $12\, x^{-1+2\varepsilon}$ and $12\, x^{-8+\varepsilon}$. With $\varepsilon < \frac{1}{52}$ these are small for large x only: the first bound is 1.31 at $x = 10$.

Where the hypothesis of (iii) *comes from.* The numbers a_t and the weights are still free in short intervals. Clause (5) of `LinkG_at` bounds the mean of $\kappa(\mathcal{Q})$ over these choices by $|\mathcal{T}(\mathcal{Q})|\, x^{-12}$. So the mean of $\mathrm{Sig}(p)$ is at most $m|\mathcal{T}|x^{-12}$ for each candidate p, and the mean of the right side of (III.136) with $\theta = x^{-\varepsilon}$ is at most $(2x^{\varepsilon}+4)\, m|\mathcal{T}|x^{-12}\cdot 2x^{\varepsilon/3}$. This is at most x^{-9} if

$$m \le x^{1/2+\varepsilon/100}, \qquad |\mathcal{T}| \le 2x^{2+9\varepsilon}, \qquad x \ge 10^5. \tag{III.137}$$

Indeed the product is then at most $24\, x^{-9.5+10.35\,\varepsilon} \le 24\, x^{-9.3}$. The first two bounds of (III.137) follow from three statements of the source: $m \le x^{\varepsilon/100}\sqrt{x\delta_1}$ (`SizeM`), $|\mathcal{T}| \le 2n\delta_1$ (`SizeTT`) and $n \le x^{2+8\varepsilon+\varepsilon/100}$ (`SizeN`), with $\delta_1 = x^{-\varepsilon/8}$. They are needed, and without a bound for $m|\mathcal{T}|$ the inequality is false. Some choice of the weights does at least as well as the mean. (Lean: `chosenW_of`, `size_alt_alt`. The latter has the three statements as hypotheses, and it assumes $24 \le x^{1/4}$ where (III.137) has $x \ge 10^5$; the condition of (III.137) suffices for the inequality as stated here. The bound for the mean of κ is part of `LinkG` and is assumed.)

The other terms, and the total. There remain: the losses $m|\mathcal{T}_p|\mathrm{lo}_G(p)$ and $m\,\mathrm{fl}_G(p)$ at the served primes; the first class; the high class; and the prime d. For the terms of $\mathrm{lo}_G, \mathrm{fl}_G$ in $\log_p$ one has, for $Y \ge 2$, $\sum_{p\le Y} \log p \cdot \log_p(2a) = \#\{p \le Y\}\cdot\log(2a)$ and $\sum_{p\le Y} \log p\, (\log_p a)^2 \le \#\{p \le$

$Y\} \cdot (\log a)^2/\log 2$. With $Y = X''$, and with the bounds $\mathrm{Sc}, \eta \le \exp(x^{\varepsilon/4})$ of the source, these are negligible against $b = \exp(x^{\varepsilon})$ for large x. The term b'/x of $\mathrm{pr}(p)$, summed with weight $2\log p$ over $p < X'$, gives $2\,(b'/x)\sum_{p<X'} \log p \le 2\log 4 \cdot (b'/x)X' = 2\log 4 \cdot b$ for each pair (j,t), by Chebyshev's bound $\sum_{p\le Y} \log p \le Y \log 4$ (cited): a constant times the unit $|\mathcal{T}|mb$. So this term alone is of the size of the right side of (III.138) below. The complete statement is taken from the source: *with the constant $C \ge 0$ of* (III.77) *and with $C' = 1$, for large x*

$$\sum_{t\in\mathcal{T}} \sum_{j=1}^{m} P_{u,t,j} - G_{\mathrm{top}} \;\le\; (3C+13)\,|\mathcal{T}|\,m\,b, \qquad \sum_{w} \lambda_w e_w \,\mathrm{io}(w) \;\le\; C'\,|\mathcal{T}|\,m\,b. \tag{III.138}$$

(Lean: the constants $3C + 13$ and $C' = 1$ are those of `joint_discriminant_eventually` (`ExcessAndPrimeLogSumUnitBounds`), which contains the bounds for the eleven sums; its proof was not read. There C is $\max(C_{\mathrm{G}}, 0)$ for the constant C_{G} of `LinkG`. The statement on the route is `stLoss7`. It proves `Loss_at` with exponent 0 from `LinkG`, the size of $\mathcal{T}$, (III.135) and the hypothesis of Lemma L14 (iii). It has one constant for both inequalities of (III.138), namely $3(1 + |C_{\mathrm{G}}| + C_{\mathrm{ord}}) + 13$, where $C_{\mathrm{ord}} \ge 0$ is a constant of `joint_discriminant_eventually`. This is at least $3C + 13$ and at least 1, so it is the weaker statement. In the Lean the left side is written with a sum `ASK`; that it equals $\sum P_{u,t,j}$ is Lemma L6 together with the statement `askIsF`, not read. By the seventh run of the dependency computation the final theorem uses `stLoss7`.) The inequality (III.138) is not proved here: the estimates for the eleven sums and their total bound are quoted. Proposition L11 and Lemmas L12 to L14 split the sum and bound the terms named in them, and no more. In particular Theorem III.3.2 gives (III.135), one of the inputs of `stLoss7`; alone it does not give (III.138).

6. A model computation: how the terms would cancel

This subsection does not prove the prime side of (M), for the source or for one member. It proves an inequality under a hypothesis, called (a) below, that does not hold in the source. Its purpose is to show which term of the earlier subsections meets which term of Part II, §6.7.

Proposition L15 (a model; hypothesis (a) does not hold in the source). *Fix a member α of a record with translate u. Assume the conclusions* (i) *and* (iii) *of Theorem L4, the first inequality of* (III.138)*, and:*

(a) *(model) the pairing gives an injective map $\mathrm{Per}_\alpha : U_\alpha \to \mathrm{Hom}_{L_0}(U', L_0)$, and $\dim U_\alpha = \dim U'$; the measures on $U_\alpha \otimes \mathbb{R}$ and on $\mathrm{Hom}_{L_0}(U', L_0) \otimes \mathbb{R}$ come from Euclidean norms that are orthogonal sums over the $n_0/2$ complex places of L_0; and ht_α is the mean over these places of the logarithm of the square of the operator norm of Per_α at the place.* ***This hypothesis does not hold in the source.*** *There the rank of $\Lambda_\alpha^{\mathrm{src}}$ is smaller than the rank of $\mathcal{R}$, there are many members, and no inequality for one member is available: the source has only an inequality for a mean over the members of the record (Lean: `Frame_at`), and that inequality has further error terms (the logarithm of the number of members, a term for the spread of the slopes, two error terms from second moments multiplied by a range of slopes, and a factor close to 1);*

(b) *(size) $\mu(\mathcal{R}^\vee) + \mathrm{ht}_\alpha - F_0 \le \sum_{t\in\mathcal{T}} \big(m(a_t \log a_t - a_t - a_t \log \Omega_E) - a_t \log \mathrm{Lv}_\alpha\big) + E_{\mathrm{size}}$;*

(c) *(level) $|E_{\mathrm{err}}(\alpha) - m \log(g_0 g_1) \sum_t a_t| \le E_{\mathrm{lev}}$;*

(d) *(count) $\log \mathrm{Lv}_\alpha \ge \sum_j \big(\langle \log \mathrm{wt}_j \rangle + \log(\mathrm{rd}_E/2) - \log \pi\big) + m \log(g_0 g_1) - E_{\mathrm{cnt}}$;*

(e) *(source lattice) $\mu(\Lambda_\alpha^{\mathrm{src}}) \ge -E_{\mathrm{src}}$.*

Here Ω_E, rd_E are as in (6.3) *of Part II, g_0, g_1 are the level primes of Part II, §6.7, $\pi = 3.14159\ldots$, γ is Euler's constant, and $\langle \log \mathrm{wt}_j \rangle$ is the mean of $\log \mathrm{wt}_j$ over the $2n$ cells. Then, with $M_{u,t,j} = a_t(\log a_t - 1 - \gamma + c_{\chi_d} - \langle \log \mathrm{wt}_j \rangle)$ as in* (4.2) *of Part II,*

$$\sum_{t\in\mathcal{T}} \sum_{j=1}^{m} \left(P_{u,t,j} - M_{u,t,j}\right) \;\le\; (3C+13)|\mathcal{T}|mb + \mathrm{Lo}(\alpha) + E_{\mathrm{size}} + E_{\mathrm{lev}} + E_{\mathrm{cnt}} \sum_t a_t + E_{\mathrm{src}}. \quad \text{(III.139)}$$

This is not (M)*. The inequality* (M) *is a mean over the translates u; a record has one translate, and the members α are not translates.*

Proof. By Theorem L4 (i), $\mathrm{Per}_\alpha(\Lambda_\alpha^{\mathrm{gl}}) \subset \mathcal{R}^\vee$, and by (a) both are lattices of the same rank, say rk, in $\mathrm{Hom}_{L_0}(U', L_0)$. At each complex place the space has real dimension $2\,\mathrm{rk}$, so the map multiplies volumes there by at most the operator norm to the power $2\,\mathrm{rk}$. Taking the product over the $n_0/2$ places, the covolume of $\mathrm{Per}_\alpha(\Lambda_\alpha^{\mathrm{gl}})$ is at most $\exp(\frac{1}{2} n_0 \,\mathrm{rk}\, \mathrm{ht}_\alpha)$ times that of $\Lambda_\alpha^{\mathrm{gl}}$. With (III.100) this is $\mu(\Lambda_\alpha^{\mathrm{gl}}) \le \mu(\mathrm{Per}_\alpha(\Lambda_\alpha^{\mathrm{gl}})) + \mathrm{ht}_\alpha \le \mu(\mathcal{R}^\vee) + \mathrm{ht}_\alpha$: the first inequality of Part II, §6.8, in the squared normalisation, followed by Lemma L1. Now

$$G_{\mathrm{top}} \le \mu(\Lambda_\alpha^{\mathrm{gl}}) - \mu(\Lambda_\alpha^{\mathrm{src}}) - F_0 + E_{\mathrm{err}}(\alpha) + \mathrm{Lo}(\alpha) \le \left(\mu(\mathcal{R}^\vee) + \mathrm{ht}_\alpha - F_0\right) + E_{\mathrm{src}} + E_{\mathrm{err}}(\alpha) + \mathrm{Lo}(\alpha) \quad \text{(III.140)}$$

by (III.111) and (e). The term F_0 has cancelled against the F_0 of (b). Insert (b), then (c) and (d): the terms $m \log(g_0 g_1) \sum_t a_t$ cancel, and the coefficient of a_t for the map j is $\log a_t - 1 - \log \Omega_E - \langle \log \mathrm{wt}_j \rangle - \log(\mathrm{rd}_E/2) + \log \pi$, which is $\log a_t - 1 - \gamma + c_{\chi_d} - \langle \log \mathrm{wt}_j \rangle$ by (6.3) of Part II. So $G_{\mathrm{top}} \le \sum_{t,j} M_{u,t,j} + E_{\mathrm{size}} + E_{\mathrm{lev}} + E_{\mathrm{cnt}} \sum_t a_t + E_{\mathrm{src}} + \mathrm{Lo}(\alpha)$. Add the first inequality of (III.138). □

This is the computation of Part II, §6.7, with the lattices made explicit. It differs from the source in four ways. (1) Hypothesis (a): in the source one map controls only the largest slope of a sublattice of the target (Part II, §6.8), and the inequality $\mu(\Lambda_\alpha^{\mathrm{gl}}) \le \mu(\mathcal{R}^\vee) + \mathrm{ht}_\alpha$ is replaced by the mean-slope argument of Part II, §6.9–6.10, applied to all members at once. (2) The errors and the inputs of that argument (second moments, the filtration by slopes) have no counterpart in (III.139). (3) (III.139) is for one member and one translate, and (M) is a mean over translates. (4) The inputs (b)–(e) are modelled on Lean statements but are not those statements: (b) and (c) on clauses 7 and 8 of `LinkDIV_at`, (d) on `Count_at`. For (e), the Lean statement `SlopesM_at` bounds the slope $\mu_{0,\alpha}$ of the first lattice of the chain, not $\mu(\Lambda_\alpha^{\mathrm{src}})$; the two differ by the term $f \log(\mathrm{Lv}_\alpha - 1)/(\mathrm{Lv}_\alpha - 2)$ of §2 (d). The proposition shows what each lattice statement of this section is for: Theorem L4 (i) lets the size of the map bound the slope of $\Lambda_\alpha^{\mathrm{gl}}$; (III.111) turns that into a bound for G_{top}; (III.138) turns G_{top} into the prime sum. (The arithmetic was tested on 25000 random data satisfying the hypotheses: no violation, equality when all hypotheses are equalities, and failure with a wrong sign in three places.)

7. Assumed here, from the source

(i) *The valuation statement* (Val). In the Lean it rests on three assumptions: `LocalQuadraticParts.adm2`, for the elements that lie both in the lattice $\widetilde{M}(\mu_w)$ and in $V_0(\mu_w)$ (the final theorem depends on it); the hypotheses `ColBrF.Tied` and `CutUp` of `val2AllAtRF`; and the structure field `l1b`. The final theorem depends on `val2AllAtRF` (seventh run of the dependency computation). The third hypothesis of that theorem, the inclusion into $\Lambda_\alpha^{\mathrm{div}}$, is supplied by `nu_into`. It was not determined whether `Tied`, `CutUp` and `l1b` are discharged on the route of the final theorem. That the local images generate $\mathcal{R}_w$ is a definition (`RR`) with the theorem `latticeGlue`, not an assumption. The statement `adm2` is proved in the source by `admissible_level_two` from: the statement `Conclusion1` on the composite of the chain

(§3 (c)), which is part of `ServF`; the integrality of the chain and of the period at the places where no gain is claimed (`MemberChainIntegral`, `TwistedPointIntegral`, `PeriodClassSumWeyl`, `TwistedPeriodAdmissible`); separate statements at the places above Lv_α, g_0, g_1; and the hypothesis `fills`. None of it is proved here.

(ii) *`ServF` from Theorem J.* `sum_over_primes_service` deduces `ServF` from `split_bound_eleven_one` (which rests on `thmAOp`, §III.5, and is conditional on `Loc.AdmB`) and from the fields `twU`, `pdv` of its hypothesis `SumOverPrimesHypotheses`; the field `pdv` contains hypothesis (B) for the kernels (`EvenConditionsInert`, `JetIntegralityBis`, `ThetaKernelShapeAtD`, `Shape`). The steps between (`orbit_entry_bounds_hold`, `lift_bounds_of_essential`, `four_orbit_conditions_of_proof`, `stmt_proved`) were not read. In particular the dictionary of §4 between Theorem J and (III.122) is not proved here: the frames, the labels, the identification of the sites of an orbit with places, and the length of the first row.

(iii) *The lattice statement* (Lat), with the integrality of τ (the first two parts of the first clause of `LinkG_at`; the last clause of `NuDef`, for the factor e_w; `exists_tau_eq_natCast` together with `fills`), and the definition (III.107) (`NprimeDef`). For the passage from (III.111) to the Lean statement (III.117): the rank formula, the bound at the places above the level, the comparison with the first lattice of the chain (`adm26`), the inequality of `HypLow2`, the bound $\sum_w \mathrm{sz}_w(\tau(w)) \le b^2$ (`ta_le`), and the bound for the rounding term. The last is the hypothesis `NprimeMuR` of `places_at_of`; the source proves it by `linkMu` from `ZeroAbove`, from Chebyshev's bound up to X'', and from bounds for $\log \mathrm{Lv}_\alpha$, $\log g_0$, $\log g_1$, $\log d$ and $\log D_\alpha$ (§2 (c)), with the finite generation and the positivity of the ranks of the glued lattices (`MuFG`, `MuRank`). That $\tau(w) = 0$ above the level is the hypothesis `htauT` of `places_at_of`. The archive also has a statement `RoundingTheta`; it was not put to the dependency computation.

(iv) *The inequality* (III.79) *at the `inertClass` primes*, and the bound for the mean of $\kappa(\mathcal{Q})$ (clauses 3 and 5 of `LinkG_at`; `linkG_at_core`, `AuxBlock8.Pf.stmtF`). The hypothesis `fills`, which identifies the data with the formulas of §4, and the bounds `kaJ_nonneg`, `kaJ_le`. At the `signal` primes: that the constants of the source satisfy (III.130), or the Lean's own comparison `lo_le_loG`, `fl_le_flG`.

(v) *The sums over primes*: the estimates for the eleven loss sums and their total bound (III.138) (`stLoss7`, `joint_discriminant_eventually`); the sizes (III.137) (`SizeM`, `SizeTT`, `SizeN`); the vanishing (III.131) (`not_inert`, with `level_split` and `gain_g`); the two classical estimates cited in §5.

(vi) *The slope inequality for many maps* (Part II, §6.9–6.10; `Frame_at`), which the source has in the place of hypothesis (a) of Proposition L15, and the Lean statements on which the inputs (b)–(e) of that proposition are modelled. The inequality (b) is the subject of §III.4.

(vii) *The properties of d*: every odd prime up to $x^{\varepsilon/(5C_o)}$ splits (`split_small`), with C_o the Lean's `J`, for which the Lean has $1 \le$ `J`.

(viii) *The translates.* How the records for the different translates u enter the mean over u in (M) was not examined.

What this section proves, and what it assumes. *Proved outright:* Lemmas L1, L2, L3 and L5 (lattices); Lemma L6 (the prime sum as a sum over single primes); Lemmas L7, L8, L9 (losses); Proposition L10 (an inequality of the shape of (III.79) at the `signal` primes, for the formulas of the source, with explicit loss terms; it is (III.79) under (III.130)); Proposition L11, Lemmas L12

and L13, and Lemma L14 given Theorem III.3.2 and one cited estimate. For the prime sum this amounts to the following: the section splits the sum and bounds several of its terms. The estimates for the eleven loss sums and their total bound (III.138) are quoted; Theorem III.3.2 alone does not give that bound. *Proved from statements of the source:* Theorem L4: part (i) from (Val), part (iii) from (Lat); part (ii) needs neither. *A model:* Proposition L15, whose hypothesis (a) does not hold in the source. *Assumed:* the items of §7. The section does not prove that any theta kernel is divisible by anything: that is (Val), which rests on Theorem J and on hypothesis (B). It does not prove the prime side of (M).

What remains.

(i) The passage from Theorem J to `Conclusion1`: frames and labels, the dictionary of §4, the chain of m maps, and the local lattice that defines $\Lambda_\alpha^{\mathrm{div}}$. Then the passage from `adm2` to (Val): the hypotheses `Tied` and `CutUp` of `val2AllAtRF` and the field `l1b`, that is, the comparison of the local pairings at different twists.

(ii) The inequality (III.79) on the served orbits of the `inertClass` primes, which are good orbits in the sense of (G1)–(G3), with the terms $\mathrm{pr}(p)$, $\mathrm{or}(p)$ and τ; the bound for the mean of $\kappa(\mathcal{Q})$; and the comparison of constants (III.130) at the `signal` primes.

(iii) The lattice statement (Lat), and the four statements of §2 (a)–(d) that lead from (III.111) to (III.117).

(iv) The eleven sums over primes behind (III.138), and their total bound.

(v) The mean-slope argument for all members of a record, which the source has in the place of hypothesis (a) of Proposition L15, and the mean over the translates in (M).

III.7 The size bound: the arithmetic inputs

This section concerns the passage from the analytic estimates of §III.4 to the inequality (III.75) at the end of §III.4, which is the first line of (6.5) of Part II. In the source this passage is one theorem (`divClausesR_of`) with a long list of hypotheses.

Proved here. First: a finite list of numerical inputs, named (S1)–(S8), (W1), (W2) and (G) below, implies the inequality, with explicit constants (Theorem III.7.1). The proof includes a lemma on the norm of a map that factors through a quotient, the averaging of the bounds over the archimedean places, and the cancellation of the period. Second: a formula for a mean of logarithms of periods, defined by the integral lines of the field E (the Proposition of subsection 4). It is deduced from the class mean formula for the elliptic curve (inputs (I2), (I3) of §III.4), from four statements about ideals, three of which are proved and one cited, and from one fact about the field E. Third: (W1) and (W2) follow from the conclusions of the paragraph "Row by row" of §III.4 and three facts taken from the source (the remark after the list of inputs).

Assumed here, from the source. All the inputs (S1)–(S8), (W1), (W2), (G). For each we give the statement and say what kind of theorem it is. In particular the period input (S4) is assumed. The Proposition of subsection 4 does not prove it. (S4) follows from that Proposition only together with three further statements, (P1)–(P3), listed at the end of subsection 4. (P1) and (P3) are taken from the source. (P2) is proved in subsection 4. Nothing in this section proves that the lattices and maps of the construction exist or have the stated properties.

1. Notation

Letters introduced in this section: k_d, $\mathcal{P}$, $\mathcal{J}$, U, U_0, $\bar{U}$, Y, ι, Π, Φ, $\varkappa_v$, $\lambda_{j,p}(v)$, $\mathcal{L}$, $\mathcal{R}$, $\mathcal{R}_0$, $\mu(\cdot)$, $\mu_{\max}(\cdot)$, μ_0, $\mathcal{H}$, $\nu_{j\tau i}$, $H_{j\tau}$, $\Omega_\tau(v)$, h_Π, $\Sigma_j(v)$, r_ν, $n_{j\tau}$, p_j, $w_{j\tau}$, $\mathfrak{e}_{j\tau}$, $\mathfrak{d}_{j\tau}$, S_{c}, λ_0, C_0, c_e, C_h, $\mathfrak{F}$, E_{loc}, E_3, $l_{3,j}$, $l_{6,j}$, de, la, $H_{\max}$, H_*, $Q_0, \mathcal{U}_1, \mathcal{U}_2, \mathcal{V}, \mathcal{E}_1, \dots, \mathcal{E}_8$, E_b, L_{pass}, $\mathcal{I}$, $\mathfrak{g}_j$, v_{h}, $\beta_j^*(v)$, $\Theta_1(v)$, cl_j, cl'_j, L_g, $\lambda_{1,\tau}$, b_τ, $\mathfrak{m}_\tau$, $\mathfrak{m}'_\tau$, e_Π, the constants $C_2, C_{\mathrm{th}}, C_3, C_4, c_5, c_6, C_T, C_8, C_e, C_\Pi, C_H$, and, in subsection 4, $\mathfrak{D}$, $\mathfrak{b}$, $\mathfrak{a}_0$, $\mathfrak{l}_\tau$, $\mathcal{M}$, $\mathcal{M}_{\mathfrak{c}}$, $\mathcal{M}^*$, $\mathfrak{c}^*$. Bold letters ($\mathbf{u}$, $\mathbf{x}$, $\mathbf{y}$, $\mathbf{e}_\tau$, $\mathbf{a}_i$, $\mathbf{m}$) denote vectors. The lists contain the letters that are used in more than one subsection. A letter that is used in one proof, in one remark, or in one paragraph of subsection 5 only is introduced where it is used and is not listed. Examples: Ξ, $\mathfrak{s}_p$ in Step 1; Δ, B, ς, C_1 in remark (1); $\mathfrak{Q}$, z, $\mathfrak{A}$ in subsection 4; w_E, D_0, $\mathrm{wt}(y)$, $\mathrm{ram}(y)$ in subsection 5.

Letters that keep the meaning they have in §III.4: x, ε, b, a, m, k, $f = 2n$, η, T, a_t, $\mathcal{T}$, Ω_E, rd_E, L_a, B_0, N_c, L_{sc}, $h(\cdot)$, $H(\cdot)$, h_A, $h_{B,j}$, $h_{X,j}$, g_0, g_1, $r = k + m - 1$, ϱ, L, M, A_ρ, Λ, γ, c_d, the terms μ_{CC}, F_0 and the constant C of (III.75), and, in (S2), ℓ, d_0, d_1, λ_1 (each with the index τ of the place). *Letters taken from §III.5:* N_m, t_0, w'', b'.

Letters that need a remark.

- K is the number field over which the matrices of the maps are defined, as in the paragraph "Sums over classes" of §III.4. In the paragraph "The reference mean square" §III.4 writes K also for $\mathbb{Q}(\sqrt{-d})$; here that field is k_d.
- K_j ($1 \le j \le m$) are the maps of the chain. Steps are numbered from 1, as in "Row by row" and in the source; "Sums over classes" numbers them from 0.
- T is the set of distinguished places and $\mathcal{T}$ is the level, as in (III.75). The other uses of T in §III.4 (for $|\theta(0)|$ and for the theta lift) do not occur here. In §III.5 the letter $\mathcal{T}$ is an index set.
- τ always denotes a weight place, that is, an element of $\mathcal{P}$; t denotes an element of T; and ρ denotes an embedding of the field M into $\mathbb{C}$, as in "The reference mean square". The letter ρ occurs only in subsection 4 and in quotations from §III.4. §III.4 writes τ for the weight place in "Several blocks" and ρ in "The eigenvalues of a lift", "Row by row" and "The pairing at the end of the chain"; what is $A_\rho^\pm$, J_ρ, d_ρ there is $A_\tau^\pm$, J_τ, d_τ here.
- h is the function $h(u) = u \log u - u$, and, in the last display of Theorem III.7.1 only, the term h of (III.75); §III.4 does the same. Heights of maps are written ht. The numbers $h_{B,j}$ are not the h_B of Part II, §6.7.
- N_m is the rank of the last space of the chain; in the source $N_m = k + m(t_0 + 1)$. The letters N, N' are left to §III.4, where they are the ranks of a pair. t_0 is the parameter called t in §III.4. ϱ and L are those of "The eigenvalues of a lift" (Lean: `nV`, `Lp`).
- η is the bound for the additional rows, as in input (e) of "Row by row"; it is not Dedekind's function. π is the number. $\varkappa_v$ is not the ratio κ of §III.4. $\Theta_1(v)$ is not the Θ_0 of §III.4; see (S1) and subsection 5.
- The labels (W1), (W2) are not the rules (W1)–(W4) of §III.5, and (G) is a label, not a letter. Parts (a), (b) of Theorem III.7.1 are not the inputs (a)–(e) of "Row by row".
- §III.5 uses x, a, ε, λ, p, d, β and N with meanings of its own, as it says. Here they follow §III.4.
- Some letters have several roles inside this section. The index tells them apart.
  - λ: λ_0 is the bound for la in (S7). $\lambda_{1,\tau}$ is the first weight at τ in (S2), the λ_1 of §III.4; it is not related to λ_0. $\lambda_{j,p}(v)$ are the eigenvalue factors of (S3). λ without index is the scalar of the source's evaluation, in subsection 5 only.

– p: p_j is a number. p without index is the index of the subspaces $\bar{U}_p$, in (S3) and in Step 1 only.
– l: ℓ_τ is the exponent of the level in (S2). $l_{3,j}$ and $l_{6,j}$ are the real numbers of (S5), (S6) and of (S3). $\mathfrak{l}_\tau$ is the integral line of subsection 4.
– H: $H(\cdot)$ is the hook product of a diagram, and $\log H_{j\tau}$ is a number of the record. $H_{\max}$ and H_* are bounds for slopes in (S7); they are not hook products. H_ρ occurs once, in a quotation from §III.4 in subsection 5.
– E: E is the field. E_3, E_b, E_K, E_X, E_{loc} and $\mathcal{E}_1, \dots, \mathcal{E}_8$ are error terms. E_0, in subsection 5 only, is the Euler product of §III.4.
– d: d is the prime. $d_{0,\tau}$, $d_{1,\tau}$ are the degrees in (S2). d_τ, in subsection 5 only, is the d_ρ of §III.4. $\mathfrak{d}_{j\tau}$ is the error term of (W2). de is one symbol.
– w: $w_{j\tau}$ is the weight shift, written w in remark (1) for fixed j, τ. w'' is a parameter. w_E, in subsection 5 only, is the number of roots of unity of E.
– S_{c} is a parameter. In $D_{S^{\mathrm{c}}}$ (subsection 5) the index S^{c} is the complement of a finite set S of primes.
– φ is an element of V_0; ϕ is the finite test function of subsection 5. ρ is an embedding; ϱ is a rank. $r = k + m - 1$; r_ν is the number of kernel degrees. B_0 is the scale above; B occurs in remark (1) only. $\Sigma_j(v)$ is a number, not a sum sign.

Letters and the other sections. The letters introduced here are local to this section. The sections of Part III do not share one notation, and we did not rename across sections. Each section declares its own letters. The clashes most likely to confuse are these. C_0 and C_H are other constants in §III.8. $\mathcal{R}$ is another object in §III.6 and in §III.9. $\mathcal{M}$, $\mathcal{H}$, $\mathcal{I}$, $\mathfrak{m}$, $\varkappa$ and Θ_1 have other meanings in §III.8. U is a unitary matrix in §III.4 and another space in §III.6. Y has other meanings in §III.4–III.6, III.8 and III.9. Φ is a test function in §III.8. ι is an index set in §III.5 and an isomorphism in §III.9. Π and $\mathfrak{d}$ have other meanings in §III.9. C_2 is another constant in §III.6, and μ_0 another number in §III.9. No statement of this section uses a letter in the sense of another section, except the letters listed above as kept from §III.4 and §III.5.

Scales. $x > 1$ and $\varepsilon > 0$ are real numbers; $L_{\mathrm{sc}} = x^\varepsilon$, $b = e^{L_{\mathrm{sc}}}$, $a = x\,b$, $L_a = \log(a\,\mathrm{rd}_E)$, $B_0 = f\,m\,b\,x^\varepsilon$, and $N_c \geq 0$. (In the application $f = 2n$ and $N_c = |T_2|\,k\,b\,x^{2\varepsilon}/n^2$; the proofs below use only $N_c \geq 0$.) As in §III.4, $h(u) = u \log u - u$ for $u > 0$ and $h(0) = 0$. So $a_t \log a_t - a_t$ means 0 when $a_t = 0$ (the convention $0 \cdot \log 0 = 0$).

Index sets. $\mathcal{J} = \{1, \dots, m\}$ indexes the lifts. $\mathcal{P}$ is the set of weight places; it has at most f elements. $T \subseteq \mathcal{P}$ is the set of distinguished places, with a real number $a_t \geq 0$ for each $t \in T$. When an element of T is called τ this number is written a_τ. For a statement S the symbol $[\mathsf{S}]$ is 1 if S holds and 0 otherwise.

The field and its places. K is the number field of the first remark above. In the source it is totally complex. v runs over its archimedean places, s is their number, and $\langle u \rangle = s^{-1} \sum_v u(v)$ for a real function u of v.

The chain and its end. A Hermitian norm at v on a K-vector space W means a positive definite Hermitian norm on the complex space $W \otimes_{K,v} \mathbb{C}$. $V_0, \dots, V_m$ are finite-dimensional K-vector spaces with a Hermitian norm at each v, and $K_j : V_{j-1} \to V_j$ ($j \in \mathcal{J}$) are K-linear maps. U is a finite-dimensional K-vector space with a positive definite Hermitian norm at each v, and $U_0 \subseteq U$ is a K-subspace. $\bar{U} = U/U_0$, and $\bar{\mathbf{u}} \in \bar{U}$ is the class of $\mathbf{u} \in U$. $\bar{U}$ carries at each v the quotient norm, $\|\mathbf{x}\| = \min\{\|\mathbf{u}\| : \bar{\mathbf{u}} = \mathbf{x}\}$, where $\mathbf{x}$ is in $\bar{U} \otimes_{K,v} \mathbb{C}$ and $\mathbf{u}$ runs over $U \otimes_{K,v} \mathbb{C}$ (the minimum is attained there). Y is a K-vector space with a Hermitian norm at each v, and $\iota : \bar{U} \to Y$ is a K-linear bijection. $\Pi : V_m \to U^*$ is a K-linear map into the dual space (the pairing at the end

of the chain). Dual spaces carry the dual norms. The map of the inequality is the K-linear map $\Phi : V_0 \to Y^*$ determined by

$$\Phi(\varphi)(\iota\,\bar{\mathbf{u}}) = \Pi(K_m \cdots K_1\varphi)(\mathbf{u}) \qquad (\varphi \in V_0,\ \mathbf{u} \in U). \tag{III.141}$$

It exists as soon as the right side vanishes for $\mathbf{u} \in U_0$; this is part of input (S3). If $\Phi \neq 0$, its height is $\mathrm{ht}(\Phi) = \langle \log \|\Phi\|_v^2 \rangle$; §III.4 writes $h(P)$ for the height of a map P. In the source the chain is a chain of matrix steps, and one step may contain several lifts (Lean: a map `lf` from the lifts to the steps); the bound (S1) is then a bound for the matrix of a step. We state everything for one lift in each step. The argument is the same.

Lattices and slopes. A lattice $\mathcal{L}$ is a finitely generated $\mathcal{O}_K$-module that spans a K-vector space with norms at every v. $\mu(\mathcal{L})$ denotes the normalised slope of the source (`muDef`): twice the slope of $\mathcal{L}$ for these norms, plus a constant depending only on K. It is the normalisation for squared norms. $\mathcal{L}^\vee$ is the dual lattice in the dual space, with the dual norms, and $\mu_{\max}(\mathcal{L})$ is the largest normalised slope of a sublattice. $\mathcal{R} \subset Y$ is the lattice whose dual receives the values of Φ, and $\mu_{\mathrm{CC}} = \mu(\mathcal{R}^\vee)$. There is a second lattice $\mathcal{R}_0$, in a space Y_0 of the same kind, built from the same data at a lower level; $\mu_0 = \mu(\mathcal{R}_0^\vee)$.

The record. The following real numbers are given: $\log \mathcal{T} \geq 0$; $\log \Omega_E$; for $j \in \mathcal{J}$ and $\tau \in \mathcal{P}$ nonnegative kernel degrees $\nu_{j\tau i}$ $(1 \leq i \leq r_\nu)$ with sum $n_{j\tau} = \sum_i \nu_{j\tau i}$, a number $\log H_{j\tau}$ (in the application the logarithm of the hook product $H(A_\tau^+)$ of step j), a number $w_{j\tau}$ (the weight shift), and two error terms $\mathfrak{e}_{j\tau}$, $\mathfrak{d}_{j\tau}$; for $j \in \mathcal{J}$ a number p_j; and parameters $k, N_m, S_{\mathrm{c}}, t_0, \eta, w'', b', \lambda_0, \varrho, L$ and constants C_0, c_e, C_h. (In the source the kernel degrees are padded by zeros to $r_\nu = N_m$.) Put

$$h_A = -\sum_{t \in T} a_t \log \mathcal{T} + C_h\big(|T|\, b' N_m + f(m\eta + S_{\mathrm{c}})N_m + f N_m^2\big)L_a, \tag{III.142}$$

$$h_{B,j} = \sum_{\tau \in \mathcal{P}} \Big(\sum_i h(\nu_{j\tau i}) - n_{j\tau} \log \Omega_E \Big) + C_h f N_m^2 (L_a + m^3 L_{\mathrm{sc}}), \tag{III.143}$$

$$h_{X,j} = \sum_{\tau \in \mathcal{P}} \log H_{j\tau} - C_h f\, \varrho^2 L, \tag{III.144}$$

$$\mathcal{H} = h_A + \sum_{j \in \mathcal{J}} (h_{B,j} - h_{X,j}). \tag{III.145}$$

These are the quantities h_A, $h_{B,j}$, $h_{X,j}$ of the paragraph "Relation to the Lean proof" of §III.4 (Lean: `RecNums.hA`, `hB`, `hX`, `hu`).

2. The inputs

(S1) *The chain bound.* For $j \in \mathcal{J}$ and every v, $\log \|K_j\|_v^2 \leq \Sigma_j(v)$, where

$$\Sigma_j(v) = \sum_{\tau \in \mathcal{P}} \Big(\sum_i h(\nu_{j\tau i}) - n_{j\tau} \log \Omega_\tau(v) \Big) + |\mathcal{P}|\, \mathfrak{g}_j \log \Theta_1(v) \tag{III.146}$$

$$+ |\mathcal{P}|\, \mathfrak{g}_j \log\big(8 + 32\,|\mathcal{P}|\, \mathfrak{g}_j\, v_{\mathrm{h}}\big) + 2 \log \beta_j^*(v) \tag{III.147}$$

$$+ \sum_{\tau \in \mathcal{P}} \big(1 + 2\mathfrak{g}_j \log(2n_{j\tau} + 2)\big) + \log \mathrm{cl}_j + \log \mathrm{cl}'_j. \tag{III.148}$$

Here $\Omega_\tau(v) > 0$ (the local squared period), $\Theta_1(v) > 0$ (the mean square of the reference theta function, per variable), $\beta_j^*(v) > 0$ (the constant of the kernel), $\mathrm{cl}_j, \mathrm{cl}'_j > 0$ (numbers of classes),

$\mathfrak{g}_j$ and v_{h} are real numbers, $v_{\mathrm{h}} \ge 0$, with, for constants $C_2, C_{\mathrm{th}}, C_4, c_5, c_6, C_T, C_8, C_e \ge 0$ and a real number L_g (in the application $\log(g_0 g_1)$):

$$\log \beta_j^*(v) \le C_2 f m^2 N_m^2 (L_g + \log \mathcal{T}) + C_e f N_m^2 L_g, \qquad \langle \log \Theta_1 \rangle \le C_{\mathrm{th}} L_{\mathrm{sc}}, \tag{III.149}$$

$$\log \mathrm{cl}_j,\ \log \mathrm{cl}'_j \le C_4 f N_m^2 (L_{\mathrm{sc}} + \log \mathcal{T} + L_g), \qquad 0 \le \mathfrak{g}_j \le N_m^2, \tag{III.150}$$

$$\log\left(8 + 32|\mathcal{P}|\mathfrak{g}_j v_{\mathrm{h}}\right) \le c_5 L_a, \qquad \log(2n_{j\tau} + 2) \le c_6 L_a, \tag{III.151}$$

$$L_g \le C_8 L_{\mathrm{sc}}, \qquad \log \mathcal{T} \le C_T m L_{\mathrm{sc}}, \tag{III.152}$$

and $C_{\mathrm{th}} + c_5 + 2c_6 + 1 + 2C_4 + 2C_e C_8 \le C_h$, $2(C_2 + C_4)(C_T + C_8) \le C_h$. Since $\mathfrak{g}_j \ge 0$ and $v_{\mathrm{h}} \ge 0$, the number $8 + 32|\mathcal{P}|\mathfrak{g}_j v_{\mathrm{h}}$ is at least 8; so its logarithm is defined. The condition $v_{\mathrm{h}} \ge 0$ is assumed here for this reason only. The record of the source does not contain it; in the source's construction $v_{\mathrm{h}} = \mathrm{rd}_E/2 > 0$ (see subsection 5).

(S2) *The pairing at the end.* There is a real number h_Π with $\|\Pi\|_v^2 \le e^{h_\Pi}$ for every v, and there are real numbers $\ell_\tau, \lambda_{1,\tau}, d_{0,\tau}, d_{1,\tau}, b_\tau, \mathfrak{m}_\tau, \mathfrak{m}'_\tau$ ($\tau \in \mathcal{P}$) and e_Π, $C_\Pi \ge 0$ with

$$h_\Pi \le -\sum_{\tau \in \mathcal{P}} \ell_\tau \log \mathcal{T} + \sum_{\tau \in \mathcal{P}} \left(2\,\mathfrak{m}_\tau + \mathfrak{m}'_\tau\right) + e_\Pi, \qquad \ell_\tau = \lambda_{1,\tau} - d_{0,\tau}, \tag{III.153}$$

$$\lambda_{1,\tau} = a_\tau - \tfrac{N_m - k}{2}\ (\tau \in T), \qquad \lambda_{1,\tau} \ge S_{\mathrm{c}} - \tfrac{N_m - 1}{2}\ (\tau \notin T), \tag{III.154}$$

$$d_{0,\tau} \le S_{\mathrm{c}} + m\eta + [\tau \in T]\, b_\tau, \quad b_\tau \le 2b'\ (\tau \in T), \quad d_{1,\tau} \le 3d_{0,\tau} + S_{\mathrm{c}} + \eta, \tag{III.155}$$

$$\mathfrak{m}_\tau \le (N_m - 1)\, d_{1,\tau} \log(2a), \quad \mathfrak{m}'_\tau \le \tfrac{N_m(N_m-1)}{2} \log(2a), \quad e_\Pi \le C_\Pi f N_m^2 L_a, \tag{III.156}$$

and $2(16 + C_T + C_\Pi) \le C_h$. The number e_Π is one real number; it does not depend on v. We read the first line as the bound of the paragraph "The pairing at the end of the chain" of §III.4 at the place τ, with $\psi \le N_m - 1$, $\Psi \le N_m$ and $\Gamma \le 2a$, and with $\mathfrak{m}_\tau$, $\mathfrak{m}'_\tau$ in the place of $\log|\vartheta|$, $\log\kappa$ there. This reading is ours: in the source $\mathfrak{m}_\tau$ and $\mathfrak{m}'_\tau$ are by definition the two majorants $(N_m - 1)\, d_{1,\tau} \log(2a)$ and $\frac{N_m(N_m-1)}{2} \log(2a)$ (Lean: `logc`, `logxi` in `lastSide_of_logT_le`).

(S3) *The eigenvalues of a lift.* The right side of (III.141) vanishes for $\mathbf{u} \in U_0$. There are a finite nonempty index set $\mathcal{I}$, K-subspaces $\bar{U}_p \subseteq \bar{U}$ ($p \in \mathcal{I}$) with $\sum_p \bar{U}_p = \bar{U}$, and positive numbers $\varkappa_v$ and $\lambda_{j,p}(v)$ such that at every v

$$\langle \iota\mathbf{x}, \iota\mathbf{x}' \rangle_{Y,v} = \varkappa_v \prod_{j \in \mathcal{J}} \lambda_{j,p}(v)\ \langle \mathbf{x}, \mathbf{x}' \rangle_{\bar{U},v} \qquad (\mathbf{x} \in \bar{U}_p,\ \mathbf{x}' \in \bar{U}). \tag{III.157}$$

There are real numbers $l_{6,j}$ with $l_{6,j} \le \log \lambda_{j,p}(v)$ for all j, p, v, and $\sum_j h_{X,j} \le \sum_j l_{6,j}$.

(S4) *The periods.* For every $\tau \in \mathcal{P}$, $\langle \log \Omega_\tau \rangle = \log \Omega_E$.

(S5) *The slope at the lower level.* There are real numbers $l_{3,j}$ ($j \in \mathcal{J}$), $\mathfrak{F}$ and E_{loc} with

$$\mu_0 - \langle \log \varkappa \rangle \le \sum_{j \in \mathcal{J}} l_{3,j} + 2\mathfrak{F} + E_{\mathrm{loc}}. \tag{III.158}$$

Here $l_{3,j}$ is an upper bound for the normalised logarithmic covolume, divided by $\dim Y$, of the lattice attached to the jth lift; $\mathfrak{F}$ is the sum of the local index terms; E_{loc} is the sum of the local errors. This is the form of the source: its theorem has the bound by $l_{3,j}$ among its hypotheses and the display as its conclusion (subsection 5).

(S6) *The covolumes.* There is a real number E_3 with

$$\sum_{j \in \mathcal{J}} l_{3,j} \le \sum_{j \in \mathcal{J}} \sum_{\tau \in \mathcal{P}} p_j\, w_{j\tau} \log \Omega_E + E_3. \tag{III.159}$$

(S7) *The passage between levels.* There are real numbers $\mathrm{de} \geq 0$, la, $H_{\max}$, $Q_0 \geq 1$ with $\mu_{\max}(\mathcal{R}_0) \leq H_{\max} - \langle \log \varkappa \rangle$, and such that, if $\mathrm{de} \leq \frac{1}{3}$ and H_* is a real number with $\mu_{\max}(\mathcal{R}_0) \leq H_*$, then

$$\mu_{\mathrm{CC}} \leq \mu_0 + 3\,\mathrm{de}\,(\mu_0 + H_* + 28\,\mathrm{la}) + 28\,\mathrm{la}. \tag{III.160}$$

Moreover $3\,\mathrm{de}\,Q_0 \leq 1$, $\sum_j l_{3,j} + 2\mathfrak{F} + E_{\mathrm{loc}} + H_{\max} + 28\,\mathrm{la} \leq Q_0$, $\mathrm{la} \leq \lambda_0$ and $C_0 \geq 28$. Here $\langle \log \varkappa \rangle$ is the mean of the numbers $\log \varkappa_v$ of (S3), as in (S5). (The source uses the same scalars $\varkappa_v$ for $\mathcal{R}_0$ and for $\mathcal{R}$; this explains why one number occurs at both levels. It is a comment, not a hypothesis.)

(S8) *The size of the local errors.* $E_3 + E_{\mathrm{loc}} \leq c_e\, m\, x^{\varepsilon}(B_0 + N_c)$.

(W1) *Rows against hooks.* For all j, τ: $\sum_i h(\nu_{j\tau i}) - \log H_{j\tau} \leq [\tau \in T]\, h(a_\tau) + \mathfrak{e}_{j\tau}$, with $\mathfrak{e}_{j\tau} \leq C_0(w_{j\tau} + S_{\mathrm{c}} + N_m^2 + m(t_0+1)\eta)L_a$.

(W2) *The total degree.* For all j, τ: $n_{j\tau} = [\tau \in T]\, a_\tau + k\, w_{j\tau} + \mathfrak{d}_{j\tau}$, with $|\mathfrak{d}_{j\tau}| \leq C_0(w_{j\tau} + k(N_m + S_{\mathrm{c}}) + m(t_0+1)\eta + [\tau \in T]\, k\, b')$.

(G) *Sizes of the parameters.* $\mathrm{rd}_E \geq 1$, $\log \mathrm{rd}_E \leq L_{\mathrm{sc}}$, $\log x \leq L_{\mathrm{sc}}$; $m, f, k \geq 1$, $k + m \leq N_m$; $S_{\mathrm{c}}, t_0, \eta, b', \lambda_0, \varrho, L \geq 0$, $\varrho \leq N_m$; $0 \leq w_{j\tau} \leq 2b$; $|p_j - k| \leq m$; $|\log \Omega_E| \leq L_a$; and

$$m\lambda_0 \leq b, \quad N_m^2 \leq b, \quad m^3 N_m^2 L_{\mathrm{sc}} \leq b, \quad m\,\eta\, N_m\, L_{\mathrm{sc}} \leq b, \quad kN_m \leq b, \tag{III.161}$$

$$kS_{\mathrm{c}} \leq w'', \quad m(t_0+1)\eta \leq w'', \quad N_m w'' \leq b, \quad b' N_m \leq a, \quad m\,k\,b' \leq a, \tag{III.162}$$

$$1 \leq C_0 \leq c_e, \qquad 0 \leq C_h \leq C_0, \qquad C_h L \leq C_0 L_a. \tag{III.163}$$

Remarks on the list. (1) (W1) *and* (W2) *follow from "Row by row".* In the source $\mathfrak{e}_{j\tau}$ and $\mathfrak{d}_{j\tau}$ are defined by equality in the first halves of (W1) and (W2) (Lean: `eEN`, `dN`). Fix j and τ and write $w = w_{j\tau}$. From its inputs (o), (a)–(e) the paragraph "Row by row" of §III.4 derives

$$\mathfrak{e}_{j\tau} \leq (w + \Delta + B)\, L_a, \qquad |\mathfrak{d}_{j\tau}| \leq k(\varsigma + \Delta) + w + B, \qquad \varsigma = \frac{k + (2j-1)(t_0+1)}{2}, \tag{III.164}$$

where $\Delta \geq 0$ is the width of the weight interval and B bounds the extra degrees. (At a distinguished place it gives $(w + B)L_a$ for the first.) Three facts are taken from the source (Lean: the fields `constant_weight_sum_bound`, `constant_succ_le`, `blk` of `NumFacts`, and `Nm_cast`): $2\Delta \leq C_1 S_{\mathrm{c}}$ for a constant C_1 with $C_1 + 1 \leq C_0$; $B = m(t_0+1)\eta$; and $N_m = k + m(t_0+1)$. So $\Delta \leq C_0 S_{\mathrm{c}}$, because $S_{\mathrm{c}} \geq 0$, and $\varsigma \leq N_m$, because $2j - 1 \leq 2m$. With $w \geq 0$, $B \geq 0$, $k \geq 1$ and $C_0 \geq 1$ from (G),

$$w + \Delta + B \leq w + C_0 S_{\mathrm{c}} + B \leq C_0(w + S_{\mathrm{c}} + N_m^2 + m(t_0+1)\eta), \tag{III.165}$$

$$k(\varsigma + \Delta) + w + B \leq kN_m + C_0 k S_{\mathrm{c}} + w + B \leq C_0(w + k(N_m + S_{\mathrm{c}}) + m(t_0+1)\eta). \tag{III.166}$$

Multiply the first line by $L_a \geq 0$. These are the bounds of (W1) and (W2); the term $[\tau \in T]\, k\, b'$ of (W2) is not needed. So (W1) and (W2) hold under the inputs (o), (a)–(e) of "Row by row" and the three facts. In the source this is `stepE_num`, used in `RecNums.ofNum`. The theorem below assumes (W1) and (W2) as stated.

(2) In the source $p_j = k + (2j - 1 - m)$ (Lean: `ClosedScale3.l3m`), so that $|p_j - k| \leq m - 1$.

(3) The condition $C_0 \geq 28$ stands in (S7) because the source has it there (Lean: the field `constant_ge_twenty_eight`). The proof uses it once, in Step 3, for $28\,\mathrm{la} \leq C_0\, m\, \lambda_0$. It is not an extra condition: $2(16 + C_T + C_{\Pi}) \leq C_h$ of (S2) and $C_h \leq C_0$ of (G) give $C_0 \geq 32$. Two other conditions are also used once each: $k + m \leq N_m$ of (G) only through $m \leq N_m$ in Step 2, and $C_0 \leq c_e$ only for the last display of the theorem.

3. The assembly

Theorem III.7.1. *Assume* (S1)–(S8), (W1), (W2) *and* (G). *Then:*

(a) *if* $\Phi \neq 0$, *then* $\operatorname{ht}(\Phi) \le \mathcal{H} - \langle \log \varkappa \rangle$;

(b) *with* $\mathcal{U}_1 = |T|\, a\, L_a$ *and* $\mathcal{U}_2 = m\, x^{\varepsilon}(B_0 + N_c)$,

$$\begin{aligned} \mu_{\mathrm{CC}} - \langle \log \varkappa \rangle + \mathcal{H} - 2\mathfrak{F} \;\le\; & \sum_{t \in T} \Big(m\big(h(a_t) - a_t \log \Omega_E\big) - a_t \log \mathcal{T} \Big) \\ & + 2C_0\, \mathcal{U}_1 + (47C_0 + 7 + c_e)\mathcal{U}_2. \end{aligned} \tag{III.167}$$

In particular, with $h = \mathcal{H} - \langle \log \varkappa \rangle$, $F_0 = 2\mathfrak{F}$ *and* $C = \max(50C_0 + 10,\, 60c_e)$,

$$\begin{aligned} \mu_{\mathrm{CC}} + h - F_0 \le & \sum_{t \in T} \Big(m\big(a_t \log a_t - a_t - a_t \log \Omega_E\big) - a_t \log \mathcal{T} \Big) \\ & + C\, |T|\, a\, L_a + C\, m\, B_0\, x^{2\varepsilon} + C\, m\, N_c\, x^{2\varepsilon}. \end{aligned} \tag{III.168}$$

The last display is the inequality (III.75) at the end of §III.4. In it $a_t \log a_t$ means 0 when $a_t = 0$, since $h(0) = 0$; part (b) is written with $h(a_t)$. The constant C is at least 1610, because $C_0 \ge 32$. So the inequality has content only when m is large compared with C: otherwise the error term $C\,|T|\,a\,L_a$ can exceed the main term. The source states (b) with the error $(50C_0 + 10)\,\mathcal{U}_1 + 60c_e\,\mathcal{U}_2$ (Lean: `RecNums.rhs`); this is weaker than (b), since $47C_0 + 7 + c_e \le 55c_e$. The terms h and F_0 of the final inequality are thus $\mathcal{H} - \langle \log \varkappa \rangle$ and $2\mathfrak{F}$. The term F_0 enters in two places. In (S5) it is added in the bound for the slope, and it is subtracted on the left. It also occurs in the condition of (S7) that bounds Q_0 from below. The proof takes the rest of this subsection. It has four steps: a bound for $\|\Phi\|_v$ at one place; the average over the places; the slope; the cancellation.

Step 1: a map that factors through a quotient.

Lemma (descent). *Fix a place* v *and extend all spaces to* $\mathbb{C}$ *by the embedding at* v. *Assume the first two sentences of* (S3). *Then*

$$\|\Phi\|_v^2 \;\le\; \frac{\|\Pi\|_v^2 \prod_{j \in \mathcal{J}} \|K_j\|_v^2}{\varkappa_v \min_{p \in \mathcal{I}} \prod_{j \in \mathcal{J}} \lambda_{j,p}(v)}. \tag{III.169}$$

Proof. Write $\mathfrak{s}_p = \varkappa_v \prod_j \lambda_{j,p}(v)$ and $\mathfrak{s}_{\min} = \min_p \mathfrak{s}_p$. Let Ξ be the operator on $\bar{U} \otimes \mathbb{C}$ with $\langle \iota\mathbf{x}, \iota\mathbf{x}' \rangle_Y = \langle \Xi\mathbf{x}, \mathbf{x}' \rangle_{\bar{U}}$ for all $\mathbf{x}, \mathbf{x}'$. It is Hermitian. The hypothesis says $\langle \Xi\mathbf{x}, \mathbf{x}' \rangle = \mathfrak{s}_p \langle \mathbf{x}, \mathbf{x}' \rangle$ for $\mathbf{x} \in \bar{U}_p$ and all $\mathbf{x}'$, that is, $\Xi\mathbf{x} = \mathfrak{s}_p \mathbf{x}$ on $\bar{U}_p$. The spaces $\bar{U}_p \otimes \mathbb{C}$ span $\bar{U} \otimes \mathbb{C}$, so Ξ is diagonalisable with all eigenvalues among the $\mathfrak{s}_p$. Eigenspaces of a Hermitian operator for different eigenvalues are orthogonal. Hence $\|\iota\mathbf{x}\|_Y^2 = \langle \Xi\mathbf{x}, \mathbf{x} \rangle \ge \mathfrak{s}_{\min} \|\mathbf{x}\|_{\bar{U}}^2$ for every $\mathbf{x}$. Now let $\mathbf{y} \in Y \otimes \mathbb{C}$ and put $\mathbf{x} = \iota^{-1}\mathbf{y}$. Choose $\mathbf{u}$ with $\bar{\mathbf{u}} = \mathbf{x}$ and $\|\mathbf{u}\|_U = \|\mathbf{x}\|_{\bar{U}}$; this is possible by the definition of the quotient norm. By (III.141),

$$|\Phi(\varphi)(\mathbf{y})| = |\Pi(K_m \cdots K_1 \varphi)(\mathbf{u})| \le \|\Pi\| \prod_j \|K_j\|\, \|\varphi\|\, \|\mathbf{u}\|_U \le \|\Pi\| \prod_j \|K_j\|\, \|\varphi\|\, \frac{\|\mathbf{y}\|_Y}{\sqrt{\mathfrak{s}_{\min}}}. \tag{III.170}$$

The norm of $\Phi(\varphi)$ in the dual of Y is the supremum of $|\Phi(\varphi)(\mathbf{y})|/\|\mathbf{y}\|_Y$. □

(Lean: `HeightBoundConstant`, proved as `height_constant_bound` and used in `height_chain_atR`; there for a chain of matrix steps, as said in subsection 1.)

Assume now (S1), (S2), (S3) and $\Phi \neq 0$. Then $\|\Phi\|_v > 0$ at every v, because a nonzero entry of the matrix of Φ stays nonzero under an embedding of K. Taking logarithms in the lemma and using $l_{6,j} \le \log \lambda_{j,p}(v)$,

$$\log \|\Phi\|_v^2 \le h_\Pi + \sum_{j\in\mathcal{J}} \Sigma_j(v) - \log \varkappa_v - \sum_{j\in\mathcal{J}} l_{6,j} \qquad \text{for every } v. \tag{III.171}$$

Step 2: the average over the places.

Lemma (the height record). *Assume the numerical parts of* (S1) *and* (S2)*, the identity* (S4)*, and* (G)*. Then*

$$h_\Pi \le h_A \qquad \text{and} \qquad \langle \Sigma_j \rangle \le h_{B,j} \quad (j \in \mathcal{J}). \tag{III.172}$$

Proof. From (G): $L_{\mathrm{sc}} \ge 1$ since $x > 1$; $L_a = \log x + L_{\mathrm{sc}} + \log \mathrm{rd}_E$, so $1 \le L_{\mathrm{sc}} \le L_a \le 3L_{\mathrm{sc}}$; $0 \le \log(2a) \le 2L_a$; and $m \le N_m$.

The first inequality. For $\tau \in T$ put $\bar D_\tau = S_{\mathrm{c}} + m\eta + 2b'$, and for $\tau \notin T$ put $\bar D_\tau = S_{\mathrm{c}} + m\eta$; then $d_{0,\tau} \le \bar D_\tau$ and $d_{1,\tau} \le 3\bar D_\tau + S_{\mathrm{c}} + \eta \le 4S_{\mathrm{c}} + 4m\eta + [\tau \in T]\, 6b'$. Since $(N_m - 1)\log(2a) \ge 0$ and the last bound is nonnegative,

$$2\,\mathfrak{m}_\tau \le \big(16N_m(S_{\mathrm{c}} + m\eta) + [\tau \in T]\, 24\, b' N_m\big) L_a, \qquad \mathfrak{m}'_\tau \le N_m^2 L_a. \tag{III.173}$$

Next, $-\ell_\tau \log \mathcal{T} = (d_{0,\tau} - \lambda_{1,\tau}) \log \mathcal{T}$ and $0 \le \log \mathcal{T} \le C_T m L_a$. For $\tau \in T$ the coefficient is $-a_\tau + \frac{N_m - k}{2} + d_{0,\tau} \le -a_\tau + \frac{N_m}{2} + S_{\mathrm{c}} + m\eta + 2b'$; for $\tau \notin T$ it is at most $\frac{N_m - 1}{2} - S_{\mathrm{c}} + S_{\mathrm{c}} + m\eta \le \frac{N_m}{2} + m\eta$. With $m \le N_m$ this gives

$$-\ell_\tau \log \mathcal{T} \le -[\tau \in T]\, a_\tau \log \mathcal{T} + C_T\Big(\tfrac{N_m^2}{2} + N_m(S_{\mathrm{c}} + m\eta) + [\tau \in T]\, 2b' N_m\Big) L_a. \tag{III.174}$$

Sum over the at most f places τ, of which $|T|$ are in T, and add $e_\Pi \le C_\Pi f N_m^2 L_a$:

$$h_\Pi \le -\sum_{t\in T} a_t \log \mathcal{T} + \Big((24 + 2C_T)|T|\, b' N_m + (16 + C_T) f (m\eta + S_{\mathrm{c}}) N_m + \big(1 + \tfrac{C_T}{2} + C_\Pi\big) f N_m^2\Big) L_a. \tag{III.175}$$

Each of the three coefficients is at most $2(16 + C_T + C_\Pi) \le C_h$. This is $h_\Pi \le h_A$.

The second inequality. Average $\Sigma_j(v)$ over v. By (S4) the first sum becomes $\sum_\tau (\sum_i h(\nu_{j\tau i}) - n_{j\tau} \log \Omega_E)$, exactly. For the other terms, using $|\mathcal{P}| \le f$, $0 \le \mathfrak{g}_j \le N_m^2$, $1 \le L_{\mathrm{sc}} \le L_a$ and $L_g + \log \mathcal{T} \le (C_8 + C_T) m L_{\mathrm{sc}}$:

$$|\mathcal{P}|\, \mathfrak{g}_j\, \langle \log \Theta_1 \rangle \le C_{\mathrm{th}} f N_m^2 L_a, \tag{III.176}$$

$$|\mathcal{P}|\, \mathfrak{g}_j \log(8 + 32|\mathcal{P}|\mathfrak{g}_j \upsilon_{\mathrm{h}}) \le c_5 f N_m^2 L_a, \tag{III.177}$$

$$\textstyle\sum_\tau \big(1 + 2\mathfrak{g}_j \log(2n_{j\tau} + 2)\big) \le (1 + 2c_6) f N_m^2 L_a, \tag{III.178}$$

$$2\langle \log \beta_j^* \rangle \le 2C_2(C_T + C_8) f m^3 N_m^2 L_{\mathrm{sc}} + 2C_e C_8 f N_m^2 L_a, \tag{III.179}$$

$$\log \mathrm{cl}_j + \log \mathrm{cl}'_j \le 2C_4 f N_m^2 L_a + 2C_4(C_T + C_8) f m^3 N_m^2 L_{\mathrm{sc}}. \tag{III.180}$$

The sum of the coefficients of $f N_m^2 L_a$ is $C_{\mathrm{th}} + c_5 + 2c_6 + 1 + 2C_4 + 2C_e C_8 \le C_h$, and that of $f m^3 N_m^2 L_{\mathrm{sc}}$ is $2(C_2 + C_4)(C_T + C_8) \le C_h$. □

(Lean: `height_bound_at_set`, applied through `HtNums.le_hu`. A parallel statement for the first inequality, `stepDiiiRecord`, is in the archive but is not used by the final theorem.)

Proof of Theorem III.7.1(a). Average (III.171) over v and apply the lemma and the last inequality of (S3):

$$\mathrm{ht}(\Phi) \le h_\Pi + \sum_j \langle \Sigma_j \rangle - \sum_j l_{6,j} - \langle \log \varkappa \rangle \le h_A + \sum_j h_{B,j} - \sum_j h_{X,j} - \langle \log \varkappa \rangle = \mathcal{H} - \langle \log \varkappa \rangle. \tag{III.181}$$

This is the first place where $\log\Omega_E$ *enters:* through (S4), each $h_{B,j}$ contains $-\sum_\tau n_{j\tau}\log\Omega_E$. The hook products enter through (S3), as the subtracted terms $h_{X,j}$. □

Step 3: the slope. Put $\mu_0' = \mu_0 - \langle\log\varkappa\rangle$ and $\mu_C = \mu_{\mathrm{CC}} - \langle\log\varkappa\rangle$. By (S5),

$$\mu_0' \le \sum_j l_{3,j} + 2\mathfrak{F} + E_{\mathrm{loc}}, \tag{III.182}$$

and then by (S6) and (S8),

$$\mu_0' \le \sum_{j\in\mathcal{J}}\sum_{\tau\in\mathcal{P}} p_j\, w_{j\tau}\log\Omega_E + 2\mathfrak{F} + E_b, \qquad E_b = E_3 + E_{\mathrm{loc}} \le c_e\, m\, x^{\varepsilon}(B_0 + N_c). \tag{III.183}$$

This is the second place where $\log\Omega_E$ *enters.* For the passage, $3\,\mathrm{de}\,Q_0 \le 1$ and $Q_0 \ge 1$ give $\mathrm{de} \le \frac{1}{3}$. Apply (S7) with $H_* = H_{\max} - \langle\log\varkappa\rangle$ and subtract the number $\langle\log\varkappa\rangle$ from both sides:

$$\mu_C \le \mu_0' + 3\,\mathrm{de}\,(\mu_0' + H_{\max} + 28\,\mathrm{la}) + 28\,\mathrm{la}. \tag{III.184}$$

By (III.182) and (S7), $\mu_0' + H_{\max} + 28\,\mathrm{la} \le Q_0$; as $\mathrm{de} \ge 0$, the middle term is at most $3\,\mathrm{de}\,Q_0 \le 1$. Also $28\,\mathrm{la} \le 28\lambda_0 \le C_0 m\lambda_0$. So

$$\mu_C \le \mu_0' + L_{\mathrm{pass}}, \qquad L_{\mathrm{pass}} \le 1 + C_0\, m\, \lambda_0. \tag{III.185}$$

Step 4: the cancellation.

Lemma (cancellation). *Assume* (W1), (W2), (G), *and let* $\mu_C, \mu_0', \mathfrak{F}, E_b, L_{\mathrm{pass}}$ *be real numbers satisfying* (III.183) *and* (III.185). *Then*

$$\mu_C + \mathcal{H} - 2\mathfrak{F} \le \sum_{t\in T}\Big(m\big(h(a_t) - a_t\log\Omega_E\big) - a_t\log\mathcal{T}\Big) + 2C_0\,\mathcal{U}_1 + (47C_0 + 7 + c_e)\,\mathcal{U}_2. \tag{III.186}$$

Proof. Add (III.185), (III.183) and the definition of $\mathcal{H}$:

$$\mu_C + \mathcal{H} - 2\mathfrak{F} \le L_{\mathrm{pass}} + E_b + h_A + E_K + E_X \tag{III.187}$$

$$+ \sum_{j,\tau}\Big(p_j w_{j\tau}\log\Omega_E + \sum_i h(\nu_{j\tau i}) - n_{j\tau}\log\Omega_E - \log H_{j\tau}\Big), \tag{III.188}$$

where $E_K = m\,C_h f N_m^2(L_a + m^3 L_{\mathrm{sc}})$ and $E_X = m\,C_h f\,\varrho^2 L$. By (W1) and (W2) the summand is at most

$$[\tau\in T]\big(h(a_\tau) - a_\tau\log\Omega_E\big) + \mathfrak{e}_{j\tau} + (p_j - k)\,w_{j\tau}\log\Omega_E - \mathfrak{d}_{j\tau}\log\Omega_E. \tag{III.189}$$

Here the two occurrences of the period cancel: $+p_j w_{j\tau}\log\Omega_E$ from the slope against $-k\,w_{j\tau}\log\Omega_E$ from the height. What is left of them is $(p_j - k)w_{j\tau}\log\Omega_E$. The sum over $j\in\mathcal{J}$ and $\tau\in\mathcal{P}$ of the first bracket is $m\sum_{t\in T}(h(a_t) - a_t\log\Omega_E)$, and h_A contributes $-\sum_t a_t\log\mathcal{T}$. So the left side is at most the main term plus

$$\mathcal{E}_1 + \cdots + \mathcal{E}_8, \tag{III.190}$$

where

$$\mathcal{E}_1 = L_{\mathrm{pass}}, \qquad \mathcal{E}_2 = E_b, \qquad \mathcal{E}_3 = E_K, \qquad \mathcal{E}_4 = E_X, \tag{III.191}$$

$$\mathcal{E}_5 = C_h\big(|T|b' N_m + f(m\eta + S_{\mathrm{c}})N_m + fN_m^2\big)L_a, \qquad \mathcal{E}_6 = \textstyle\sum_{j,\tau}\mathfrak{e}_{j\tau}, \tag{III.192}$$

$$\mathcal{E}_7 = \textstyle\sum_{j,\tau}|p_j - k|\,w_{j\tau}\,|\log\Omega_E|, \qquad \mathcal{E}_8 = \textstyle\sum_{j,\tau}|\mathfrak{d}_{j\tau}|\,|\log\Omega_E|. \tag{III.193}$$

Put $\mathcal{V} = f\,m\,b\,L_{\mathrm{sc}}$. From (G): $L_{\mathrm{sc}} \geq 1$, $b \geq 1$, $L_a \leq 3L_{\mathrm{sc}}$; $S_{\mathrm{c}} \leq kS_{\mathrm{c}} \leq w'' \leq N_m w'' \leq b$ and $S_{\mathrm{c}} N_m \leq b$; $m(t_0+1)\eta \leq w'' \leq b$; $b \leq \mathcal{V}$; and

$$\mathcal{U}_2 \geq m\,x^{\varepsilon} B_0 = m\,L_{\mathrm{sc}}\,\mathcal{V} \;\geq\; m\mathcal{V} \;\geq\; \mathcal{V}. \tag{III.194}$$

There are m values of j and at most f values of τ. Then:

$$\mathcal{E}_1 \leq 1 + C_0 b \leq (1+C_0)\mathcal{V}; \tag{III.195}$$

$$\mathcal{E}_3 \leq C_0\,mf\,(b \cdot 3L_{\mathrm{sc}} + b) \leq 4C_0\mathcal{V}; \tag{III.196}$$

$$\mathcal{E}_4 \leq C_0\,mfN_m^2 L_a \leq 3C_0\mathcal{V} \qquad (\text{by } C_h L \leq C_0 L_a,\ \varrho \leq N_m); \tag{III.197}$$

$$\mathcal{E}_5 \leq C_0|T|aL_a + C_0 f\big(3b + 3bL_{\mathrm{sc}} + 3bL_{\mathrm{sc}}\big) \leq C_0\,\mathcal{U}_1 + 9C_0\mathcal{V} \quad (\text{by } b'N_m \leq a,\ m\eta N_m L_{\mathrm{sc}} \leq b); \tag{III.198}$$

$$\mathcal{E}_6 \leq mf\,C_0\,(2b + b + b + b)\,3L_{\mathrm{sc}} = 15C_0\mathcal{V}; \tag{III.199}$$

$$\mathcal{E}_7 \leq mf \cdot m \cdot 2b \cdot 3L_{\mathrm{sc}} = 6m\mathcal{V}; \tag{III.200}$$

$$\mathcal{E}_8 \leq 15C_0\mathcal{V} + m|T|\,C_0\,k\,b'L_a \leq 15C_0\mathcal{V} + C_0\,\mathcal{U}_1 \qquad (\text{by } kN_m \leq b,\ kS_{\mathrm{c}} \leq b,\ mkb' \leq a). \tag{III.201}$$

Finally $\mathcal{E}_2 \leq c_e\,\mathcal{U}_2$. The sum is at most $2C_0\,\mathcal{U}_1 + (47C_0+1)\mathcal{V} + 6m\mathcal{V} + c_e\,\mathcal{U}_2 \leq 2C_0\,\mathcal{U}_1 + (47C_0 + 7 + c_e)\,\mathcal{U}_2$. □

(Lean: `divisor_inequality_explicit`, used through `divIneqRecord` and `RecNums.divIneq`; it is `divisor_inequality_assembly`, the exact rearrangement, followed by `error_absorption`, the estimates of $\mathcal{E}_1, \ldots, \mathcal{E}_8$. The source's constants are $50C_0 + 10$ and $60c_e$.)

Proof of Theorem III.7.1(b). Steps 3 and 4. For the last display of the theorem: $h(a_t) = a_t \log a_t - a_t$; $2C_0 \leq C$; $47C_0 + 7 + c_e \leq 55c_e \leq C$ because $1 \leq C_0 \leq c_e$; and $\mathcal{U}_2 = m\,x^{\varepsilon}(B_0 + N_c) \leq m\,x^{2\varepsilon}(B_0 + N_c)$ because $x > 1$. □

Where each term comes from.

Term of the inequality	Source in the proof
$m\,h(a_t)$	(W1): at a distinguished place the hook product cancels all rows of the kernel but the first, once for each of the m lifts
$-m\,a_t \log \Omega_E$	(S4) and (W2): the part $[\tau \in T]\,a_\tau$ of the degree $n_{j\tau}$, times the mean period, for each lift
$-a_t \log \mathcal{T}$	(S2): the level factor of the last map, through $\ell_\tau = \lambda_{1,\tau} - d_{0,\tau}$ and $\lambda_{1,\tau} = a_\tau - \frac{N_m - k}{2}$
cancelled: $k\,w_{j\tau} \log \Omega_E$	$-k\,w_{j\tau} \log \Omega_E$ from (S4), (W2) in the height; $+p_j w_{j\tau} \log \Omega_E$ from (S6) in the slope; remainder $\mathcal{E}_7$
$C\,\lvert T\rvert\,a\,L_a$	$\mathcal{E}_5$ (the term $\lvert T\rvert\,b'N_m$ of the end pairing) and $\mathcal{E}_8$ (the term $k\,b'$ of the degree error at distinguished places, times $\lvert\log \Omega_E\rvert$)
$C\,m\,B_0\,x^{2\varepsilon}$	$\mathcal{E}_1$ (level passage), $\mathcal{E}_3$ (constants of the kernels, class numbers, reference mean square), $\mathcal{E}_4$ (loss in the eigenvalue bound), the rest of $\mathcal{E}_5$, $\mathcal{E}_6$ (rows against hooks), $\mathcal{E}_7$ (the periods cancel only up to $\lvert p_j - k\rvert \leq m$), the rest of $\mathcal{E}_8$; and the part of $\mathcal{E}_2$ with B_0
$C\,m\,N_c\,x^{2\varepsilon}$	$\mathcal{E}_2$ only: the local errors $E_3 + E_{\mathrm{loc}}$, by (S8)

Among the terms of the second kind, $\mathcal{E}_7 \leq 6m\mathcal{V}$ is the only one whose bound carries the factor m; so it is the only one that needs the full factor m in $\mathcal{U}_2 \geq mL_{\mathrm{sc}}\mathcal{V}$. It is the cost of $p_j \neq k$. Its bound is not the largest in general: the bounds for $\mathcal{E}_6$ and $\mathcal{E}_8$ are $15C_0\mathcal{V}$, and $6m > 15C_0$ only when $m > 2.5\,C_0$.

4. The mean of the periods of the integral lines

This subsection proves a formula for a mean of periods (the Proposition below). It does not prove (S4); what is missing is listed at the end. *In this subsection ρ denotes an embedding of a field M into $\mathbb{C}$, as in the paragraph "The reference mean square" of §III.4, and τ, τ' denote weight places.*

Let $k_d = \mathbb{Q}(\sqrt{-d})$, embedded in $\mathbb{C}$. Its discriminant is $-d$, because d is squarefree and $d \equiv 3 \bmod 4$ (the d of the paper is a prime with $d \equiv 7 \bmod 8$); Lemma A uses this. Let $E \supset k_d$ be the field of the construction, of degree $[E : k_d] = f$. The weight places are the f embeddings $\tau : E \to \mathbb{C}$ that extend the given embedding of k_d. So here $\mathcal{P}$ has exactly f elements; subsection 3 used only $|\mathcal{P}| \le f$. Let $\mathfrak{D}$ be the different of E over k_d, and let the bar be complex conjugation on E.

A fact about the field E. There is an integral ideal $\mathfrak{b}$ of E with $\mathfrak{D} = \mathfrak{b}\,\bar{\mathfrak{b}}$. This holds for the field of the paper, $E = \mathbb{Q}(\zeta_N + \zeta_N^{-1}, \sqrt{-q}, \sqrt{-d})$, where $N = 4n$, n is a power of 2, and q is a prime with $q \equiv 3 \bmod 4$ that divides $d + 1$. The proof has two steps.

First step: every prime of E that divides $\mathfrak{D}$ lies over 2 or over q. (This is (P2) below.) Put $F = \mathbb{Q}(\zeta_N + \zeta_N^{-1}, \sqrt{-q})$. Then $E = k_d F$. Since $q \equiv 3 \bmod 4$, the number $\sqrt{-q}$ lies in $\mathbb{Q}(\zeta_q)$ (the quadratic Gauss sum). So $F \subseteq \mathbb{Q}(\zeta_{Nq})$, and F is unramified over $\mathbb{Q}$ at every prime number that does not divide $2q$. Let $\mathfrak{Q}$ be a prime of E that does not lie over 2 or q. An extension that is unramified at a prime stays unramified when the ground field is enlarged; so $E = k_d F$ is unramified over k_d at $\mathfrak{Q}$. A prime that is unramified does not divide the different. (For $n = 1$ the field F is $\mathbb{Q}(\sqrt{-q})$, and only primes over q can divide $\mathfrak{D}$.) The Lean has this step as `diffSupport`.

Second step. 2 and q split in k_d, because $d \equiv 7 \bmod 8$ and $-d \equiv 1 \bmod q$. Complex conjugation exchanges the two primes of k_d over 2, and the two over q. So a prime $\mathfrak{Q}$ of E that divides $\mathfrak{D}$ and its conjugate $\bar{\mathfrak{Q}}$ lie over different primes of k_d, and $\bar{\mathfrak{Q}} \ne \mathfrak{Q}$. Conjugation maps E to E and k_d to k_d; so $\bar{\mathfrak{D}} = \mathfrak{D}$, and $\mathfrak{Q}$, $\bar{\mathfrak{Q}}$ occur in $\mathfrak{D}$ with the same exponent. Take for $\mathfrak{b}$ the product of one prime power from each pair. (Lean: `principal` in `KcDiff.lean`, from `diffSupport` and `kc_below_ne`; it is used through `Arith.bb_ex` to fill the field `b_spec` of the record `Host`.)

This $\mathfrak{b}$ is coprime to $\bar{\mathfrak{b}}$, as is the ideal of the source. There are other integral choices when an exponent e is larger than 1: one may take $\mathfrak{Q}^i \bar{\mathfrak{Q}}^{e-i}$ with $0 \le i \le e$ from a pair. The Proposition below holds for every $\mathfrak{b}$ with $\mathfrak{D} = \mathfrak{b}\,\bar{\mathfrak{b}}$; its proof uses nothing else. The coprimality is not used in this section.

The splitting of 2 and q in k_d is sufficient for the existence of $\mathfrak{b}$. It is not necessary. For a field E with complex multiplication over an imaginary quadratic field, an integral $\mathfrak{b}$ with $\mathfrak{D} = \mathfrak{b}\,\bar{\mathfrak{b}}$ exists if and only if every prime of $\mathfrak{D}$ that is equal to its conjugate has an even exponent. (If $\bar{\mathfrak{Q}} = \mathfrak{Q}$, the exponent of $\mathfrak{Q}$ in $\mathfrak{b}\,\bar{\mathfrak{b}}$ is twice its exponent in $\mathfrak{b}$. Conversely, halve the exponents of these primes and treat the other primes in pairs as above.) So the statement fails for some fields: for $E = \mathbb{Q}(\sqrt{-7}, \sqrt{-3})$ over $\mathbb{Q}(\sqrt{-7})$ the different is the prime over 3, with exponent 1, and it is its own conjugate.

Fix such a $\mathfrak{b}$ and put $\mathfrak{a}_0 = \mathfrak{b}^{-1}$. Let $M \subset \mathbb{C}$ be a number field that contains $\tau(E)$ for every τ and a field of definition of the elliptic curve of §III.4, with lattice Λ normalised by $g_2^3 - 27 g_3^2 = 1$. For an embedding ρ of M, A_ρ is the area of the lattice of the conjugate curve, as in §III.4.

Let $\mathcal{M} \subset M^f$ be the $\mathcal{O}_M$-module spanned by the vectors $(\tau(\alpha))_\tau$, $\alpha \in \mathfrak{a}_0$, and let

$$\mathfrak{l}_\tau = \{ z \in M : \; z\,\mathbf{e}_\tau \in \mathcal{M} \}, \tag{III.202}$$

where $\mathbf{e}_\tau$ is the τth coordinate vector. This is the integral line at the place τ (Lean: `ellT`). For a nonzero $z \in \mathfrak{l}_\tau$ the local squared period at (τ, ρ) is $A_\rho / (\pi |\rho(z)|^2)$, and the mean of its logarithm is

defined, with a correction for the index, by

$$\overline{\log \Omega_\tau} = \frac{1}{[M:\mathbb{Q}]}\Big(\sum_\rho \log \frac{A_\rho}{\pi\,|\rho(z)|^2} + 2\log\big[\mathfrak{l}_\tau : \mathcal{O}_M z\big]\Big). \tag{III.203}$$

(Lean: `meanLogOm`, `logMet`.) If $\mathfrak{l}_\tau = \mathcal{O}_M z_\tau$ is principal, this is the plain mean over ρ of $\log\left(A_\rho/(\pi|\rho(z_\tau)|^2)\right)$.

Proposition (the mean period). *The number* (III.203) *does not depend on z, on τ, or on M, and*

$$\overline{\log \Omega_\tau} = \frac{1}{[M:\mathbb{Q}]}\sum_\rho \log \frac{A_\rho}{\pi} - \log\frac{\mathrm{rd}_E}{\sqrt{d}}. \tag{III.204}$$

Under the inputs (I2), (I3) *of §III.4, it equals $\gamma + \log(2\pi/\mathrm{rd}_E) - c_d = \log \Omega_E$.*

The proof uses four statements about ideals. N is the absolute norm of ideals of E.

Lemma A (the tower formula). $|\mathrm{disc}\, E| = d^f\, \mathrm{N}(\mathfrak{D})$. *Consequently* $\frac{1}{[E:\mathbb{Q}]}\log \mathrm{N}(\mathfrak{D}) = \log \mathrm{rd}_E - \log\sqrt{d}$.

Proof. Cited: the different is multiplicative in towers, $\mathfrak{D}_{E/\mathbb{Q}} = \mathfrak{D}\cdot \mathfrak{D}_{k_d/\mathbb{Q}}\mathcal{O}_E$, and the absolute norm of the different of a number field is the absolute value of its discriminant (J. Neukirch, *Algebraic Number Theory*, 1999, Ch. III, (2.2)(i) and (2.9)). The norm of $\mathfrak{D}_{k_d/\mathbb{Q}}\mathcal{O}_E$ is $\mathrm{N}(\mathfrak{D}_{k_d/\mathbb{Q}})^{[E:k_d]} = d^f$, since the discriminant of k_d is $-d$. For the second statement divide the logarithm by $[E:\mathbb{Q}] = 2f$. □

Lemma B (the integral line). *For a nonzero fractional ideal $\mathfrak{c}$ of E let $\mathcal{M}_\mathfrak{c}$ be the $\mathcal{O}_M$-module spanned by the vectors $(\tau'(\alpha))_{\tau'}$, $\alpha \in \mathfrak{c}$. Then the line $\{z \in M : z\,\mathbf{e}_\tau \in \mathcal{M}_\mathfrak{c}\}$ is the $\mathcal{O}_M$-module generated by $\tau(\mathfrak{c}\,\mathfrak{D})$. In particular $\mathfrak{l}_\tau = \tau(\mathfrak{a}_0\mathfrak{D})\mathcal{O}_M = \tau(\bar{\mathfrak{b}})\mathcal{O}_M$.*

Proof. Let $\mathfrak{c}^* = \{\alpha^* \in E : \mathrm{Tr}_{E/k_d}(\alpha^*\mathfrak{c}) \subseteq \mathcal{O}_{k_d}\} = \mathfrak{c}^{-1}\mathfrak{D}^{-1}$ (the definition of the different), and let $\mathcal{M}^*$ be the $\mathcal{O}_M$-module spanned by the vectors $(\tau'(\alpha^*))_{\tau'}$, $\alpha^* \in \mathfrak{c}^*$. Use the bilinear form $\langle \mathbf{m}, \mathbf{m}'\rangle = \sum_{\tau'} m_{\tau'} m'_{\tau'}$ on M^f. For $\alpha \in \mathfrak{c}$ and $\alpha^* \in \mathfrak{c}^*$ the pairing of their vectors is $\sum_{\tau'} \tau'(\alpha\alpha^*) = \mathrm{Tr}_{E/k_d}(\alpha\alpha^*) \in \mathcal{O}_{k_d} \subseteq \mathcal{O}_M$. (Every τ' restricts to the given embedding of k_d; so the map $\alpha \mapsto (\tau'(\alpha))_{\tau'}$ is k_d-linear, and the sum over τ' is the trace to k_d.) We claim that $\mathcal{M}_\mathfrak{c}$ is exactly the set of $\mathbf{m}$ with $\langle \mathbf{m}, \mathbf{m}'\rangle \in \mathcal{O}_M$ for all $\mathbf{m}' \in \mathcal{M}^*$. It is enough to check this after localising at each prime of $\mathcal{O}_M$. Above a prime $\mathfrak{p}$ of k_d the module $\mathfrak{c}_\mathfrak{p}$ is free over the local ring of k_d, with a basis $\alpha_1, \dots, \alpha_f$, and $\mathfrak{c}^*_\mathfrak{p}$ has the dual basis $\alpha_1^*, \dots, \alpha_f^*$ for the trace form. The corresponding vectors $\mathbf{a}_i$, $\mathbf{a}_i^*$ of M^f satisfy $\langle \mathbf{a}_i, \mathbf{a}_{i'}^*\rangle = \delta_{ii'}$, so they are two bases of M^f, dual to each other, and they generate the localisations of $\mathcal{M}_\mathfrak{c}$ and $\mathcal{M}^*$. This proves the claim. Now $z\,\mathbf{e}_\tau$ pairs with the vector of α^* to $z\,\tau(\alpha^*)$. So $z\,\mathbf{e}_\tau \in \mathcal{M}_\mathfrak{c}$ if and only if $z\,\tau(\alpha^*) \in \mathcal{O}_M$ for all $\alpha^* \in \mathfrak{c}^*$, that is, if and only if z lies in the inverse of the ideal $\tau(\mathfrak{c}^*)\mathcal{O}_M$, which is $\tau(\mathfrak{c}\,\mathfrak{D})\mathcal{O}_M$. □

Lemma C (the index). *For nonzero $z \in \mathfrak{l}_\tau$,* $[\mathfrak{l}_\tau : \mathcal{O}_M z] = |\mathrm{N}_{M/\mathbb{Q}}(z)| \cdot \mathrm{N}(\bar{\mathfrak{b}})^{-[M:\mathbb{Q}]/[E:\mathbb{Q}]}$.

Proof. For a nonzero ideal $\mathfrak{A}$ of $\mathcal{O}_M$ and $z \in \mathfrak{A}$ one has $[\mathfrak{A} : z\mathcal{O}_M] = |\mathrm{N}_{M/\mathbb{Q}} z|/\mathrm{N}_M(\mathfrak{A})$, by multiplicativity of the norm. By Lemma B, $\mathfrak{A} = \mathfrak{l}_\tau$ is the extension to M of the ideal $\tau(\bar{\mathfrak{b}})$ of $\tau(E)$, and the norm of an extended ideal is the norm raised to the power $[M : \tau(E)]$. □

Lemma D (the product formula). *For nonzero $z \in M$,* $\sum_\rho \log|\rho(z)|^2 = 2\log|\mathrm{N}_{M/\mathbb{Q}}(z)|$.

Proof. $\mathrm{N}_{M/\mathbb{Q}}(z) = \prod_\rho \rho(z)$, the product over all embeddings. □

Proof of the Proposition. By Lemmas C and D the terms with z cancel:

$$2\log[\mathfrak{l}_\tau : \mathcal{O}_M z] - \sum_\rho \log|\rho(z)|^2 = -\frac{2[M:\mathbb{Q}]}{[E:\mathbb{Q}]}\log \mathrm{N}(\bar{\mathfrak{b}}). \tag{III.205}$$

In particular a unit, or any other change of z, does not change (III.203). Since $\mathrm{N}(\mathfrak{b}) = \mathrm{N}(\bar{\mathfrak{b}})$ and $\mathfrak{D} = \mathfrak{b}\bar{\mathfrak{b}}$, $2\log\mathrm{N}(\bar{\mathfrak{b}}) = \log\mathrm{N}(\mathfrak{D})$. By Lemma A,

$$\overline{\log\Omega_\tau} = \frac{1}{[M:\mathbb{Q}]}\sum_\rho \log\frac{A_\rho}{\pi} - \frac{1}{[E:\mathbb{Q}]}\log\mathrm{N}(\mathfrak{D}) = \frac{1}{[M:\mathbb{Q}]}\sum_\rho \log\frac{A_\rho}{\pi} - \log\frac{\mathrm{rd}_E}{\sqrt{d}}. \tag{III.206}$$

The first term does not depend on M. Indeed, let $M_0 \subseteq M$ be a field that contains the coefficients g_2, g_3 of the curve. The number A_ρ depends only on the restriction of ρ to M_0, and every embedding of M_0 extends to exactly $[M:M_0]$ embeddings of M. So the mean over the embeddings of M is the mean over the embeddings of M_0, and two fields M are compared through their compositum. This proves the first assertion. By (I2) and (I3) of §III.4 the mean of $\log A_\rho$ is $\log\pi + \gamma + \log(2\pi/\sqrt{d}) - c_d$. Subtracting, $\overline{\log\Omega_\tau} = \gamma + \log(2\pi/\mathrm{rd}_E) - c_d$, which is $\log\Omega_E$ by its definition in §I.7. □

So the constant Ω_E differs from the elliptic Ω by the factor $\mathrm{N}(\mathfrak{D})^{1/[E:\mathbb{Q}]} = \mathrm{rd}_E/\sqrt{d}$. Two things give this factor. The first is Lemma B: the integral line at a weight place is generated by $\tau(\mathfrak{a}_0\mathfrak{D})$, not by $\tau(\mathfrak{a}_0)$. The second is the choice $\mathfrak{a}_0 = \mathfrak{b}^{-1}$, for which $\mathrm{N}(\mathfrak{a}_0\mathfrak{D})^2 = \mathrm{N}(\mathfrak{D})$. For an arbitrary nonzero fractional ideal $\mathfrak{a}_0$ the same proof gives

$$\overline{\log\Omega_\tau} = \frac{1}{[M:\mathbb{Q}]}\sum_\rho \log\frac{A_\rho}{\pi} - \frac{2}{[E:\mathbb{Q}]}\log\mathrm{N}(\mathfrak{a}_0\mathfrak{D}); \tag{III.207}$$

with $\mathfrak{a}_0 = \mathcal{O}_E$ the factor would be $(\mathrm{rd}_E/\sqrt{d})^2$. (Lean: `meanPeriodAt`, from `mean_log_period`, `br_meanLogOm_eq` and `two_log_brN`; Lemma B is `lineR_eq`, Lemma C is `index_lineR`, Lemma D is `br_sum_log_norm_sq`, Lemma A is `natAbs_discr_eq`. The source proves Lemma A from the same two facts.)

What the Proposition proves, and what it does not. It proves that the number (III.203), which is defined by the ideal $\mathfrak{a}_0$ and the areas A_ρ, equals $\log\Omega_E$, under (I2), (I3) and the fact about E. Input (S4) is a different statement. It concerns the numbers $\Omega_\tau(v)$ of (S1), averaged over the archimedean places v of K. To deduce (S4) from the Proposition one needs three further statements. (P1) and (P3) are assumed, from the source, and not proved here. (P2) was proved above:

(P1) the period that occurs in the block estimate of §III.4 for the kernel conjugated by ρ, at the weight place τ, is $A_\rho/(\pi|\rho(z_\tau)|^2)$ for a generator z_τ of $\mathfrak{l}_\tau$ (Lean: `logOmAt`, `ExtraTie.om_eq`). Generators need not exist over a given M. An example: $M = E$, $q = 3$, $d = 23$, $N = 4$, 8 or 16 (the class number of E is 3, 6, 60). There the ideal $\tau(\bar{\mathfrak{b}})\mathcal{O}_E$ of Lemma B is not principal for any of the choices of $\mathfrak{b}$ coprime to $\bar{\mathfrak{b}}$. This was computed with Sage 10.9, with class groups computed in its mode `proof=False`; so the example is correct under GRH and was not checked without it. Nothing in this section depends on the example. It depends on the choice of $\mathfrak{b}$: for $N = 8$ two of the eight integral $\mathfrak{b}$, which are not coprime to their conjugates, give a principal line. The existence of generators over a suitable larger field is the hypothesis `IsDgen` of `mean_logMet_eq`;

(P2) every prime of E that divides $\mathfrak{D}$ lies over 2 or over q (Lean: `diffSupport`). This is the first step of the fact about E above; it is proved there for the field E of the paper. With it the ideal $\mathfrak{b}$ exists;

(P3) the mean over the embeddings ρ of M can be replaced by the mean over the places v of K: the source takes $M \supseteq K$ and for $\log\Omega_\tau(v)$ an average, with weights of sum 1, of the numbers $\log\left(A_\rho/(\pi|\rho(z_\tau)|^2)\right)$, the weights having the same total for every ρ (Lean: `kerSide_of_meansS`, `exists_weights`).

5. The remaining inputs: statements and status

Each statement below is assumed in this section. We give its form in the source and say what a proof needs. The Lean names are collected in subsection 6.

(S1), the chain bound. The analytic content is proved in §III.4: the bound for one block with the period ("Several blocks, and the period"), the bound for one map ("Sums over classes"), and the bound for the Gaussian constant by a mean square. Compare with the formula for $\log b_j(v)$ and with the constant S of (III.51) there. $|\mathcal{P}|\,\mathfrak{g}_j$ is the number $g = 2nNN'$ of variables, with N, N' the ranks of the pair as in §III.4. $\beta_j^*(v)$ is the prefactor $\beta\, c^{2[E:\mathbb{Q}]NN'}$, and $\mathrm{cl}_j, \mathrm{cl}'_j$ count classes. $\Theta_1(v)$ is a mean square per variable. In §III.4 the reference mean square at an embedding ρ is $\Theta_{0,\rho} = H_\rho^{fNN'}$, with H_ρ the mean square of one elliptic theta function. It enters S^2 once, with $\langle \log \Theta_{0,\rho} \rangle \le fNN'L_{\mathrm{sc}}$. In (S1) the same term is written $|\mathcal{P}|\,\mathfrak{g}_j \log \Theta_1(v)$, with $\langle \log \Theta_1 \rangle \le C_{\mathrm{th}} L_{\mathrm{sc}}$. So Θ_1 corresponds to H_ρ, not to $\Theta_{0,\rho}$. What is assumed: that the maps of the construction are instances (item (vi) of the list in §III.4), and the eight numerical bounds listed under (S1). Of these, the bound for $\langle \log \Theta_1 \rangle$ is the conclusion of the paragraph "The reference mean square" of §III.4 (conditional on (I1)–(I5)), divided by the number of variables; the bounds for β_j^* and for the numbers of classes are statements about the finite data of the construction (coefficients, modulus, class numbers of Hermitian lattices of bounded level). A proof needs a bound for class numbers of definite Hermitian lattices in terms of the mass and the level. The source's statements were not read for this section, with one exception, the sign of v_{h}. In the record of the source v_{h} is a real number with no condition on its sign (Lean: the field `vv` of `HtNums`, and the same in `KerSide`); the logarithm of the Lean is defined for every real number, so the record needs none. Where the source builds the record, $v_{\mathrm{h}} = \mathrm{rd}_E/2$ (Lean: `NumDatum.v`), and $v_{\mathrm{h}} \ge 0$ is proved from $\mathrm{rd}_E \ge 1$ inside the proof of the bound with c_5 (`log_eight_leS`). The source's statement of the analytic bound for one map (`LiftedMapEmbeddingLogBound`) has $v_{\mathrm{h}} > 0$ among its hypotheses. So $v_{\mathrm{h}} \ge 0$ is not a hypothesis of the record, and it is a consequence of the construction.

(S2), the pairing at the end. The inequality $\|\Pi\|_v^2 \le e^{h_\Pi}$, with h_Π bounded as in the first line of (S2) with $e_\Pi = 0$, is the conclusion of the paragraph "The pairing at the end of the chain" of §III.4, with its own inputs; there the two terms are bounded by $\log|\vartheta| \le \psi\, d_1 \log\Gamma$ and $\log\kappa \le \frac{\Psi(\Psi-1)}{2}\log\Gamma$. The source does not name $\log|\vartheta|$ and $\log\kappa$: its numbers $\mathfrak{m}_\tau$, $\mathfrak{m}'_\tau$ are these two majorants with $\psi = N_m - 1$, $\Psi = N_m$ and $2a$ in the place of Γ. The identification of $d_{0,\tau}$, $d_{1,\tau}$, ℓ_τ, $\lambda_{1,\tau}$, b_τ with the d_0, d_1, ℓ, λ_1, $b_{\mathfrak{t}}$ of §III.4 was confirmed in the Lean (`lastSide_of_logT_le`). The remaining conditions of (S2) express the exponent $\ell_\tau = \lambda_{1,\tau} - d_{0,\tau}$ and the degrees $d_{0,\tau}$, $d_{1,\tau}$ through the weights of the construction: at a distinguished place the first weight is a_τ up to the shift $(N_m - k)/2$. They are statements about the weights only; this is the "substitution of the actual weights" of item (ii) of the list in §III.4. A proof needs the weights of the last two spaces of the chain as explicit functions of the data. In the source $e_\Pi = 0$ and $C_\Pi = 0$.

(S3), the eigenvalues. This is the paragraph "The eigenvalues of a lift" of §III.4, with its hypotheses (E1)–(E7). The source's evaluation of the scalar is the following statement (Lean: `scalar_bound_lowC`; among the declarations the final theorem depends on). Let

$$\lambda = D_0 \cdot \mathrm{vol}(\mathcal{F}) \cdot \prod_\tau H(A_\tau^+) H(A_\tau^-) \cdot \frac{\|\phi\|^2}{D_{S^c}} \cdot \prod_\tau J_\tau \cdot |E_0|, \tag{III.208}$$

where $D_0 \ge 1$ is a product of dimensions of representations; $\mathrm{vol}(\mathcal{F})$ is the volume of a fundamental set for the level subgroup of the definite group; $\|\phi\|^2$ is the squared L^2-norm of the finite test function for a measure in which the rational lattice has covolume one; J_τ is the integral, over the real unitary group at τ, of the squared absolute value of a matrix coefficient of a unit vector of the

lowest type for the maximal compact subgroup; E_0 is an Euler product over the prime ideals $\mathfrak{P}$ of E outside a finite set, with parameters of absolute value at most $(\mathrm{N}\mathfrak{P})^{(r-1)/2-(1+t_0)/2}$, as in (E5); and D_{S^c} is the normalising factor outside that set. Assume the level-index identity (the index of the level subgroup times $\|\phi\|^2$ equals $1/D_S$), the lower bound $\left(w_E^{\varrho}\,\varrho!\right)^{-1}$ for the mass, with w_E the number of roots of unity of E, and the bound $-\log J_\tau \le 2 + r^2 \log\left(2\sqrt{2}\, r^2 \varrho(d_\tau+2)\right)$. Then $\lambda > 0$ and

$$\log\lambda \;\ge\; \sum_{\tau} \left(\log H(A_\tau^+) + \log H(A_\tau^-)\right) - 8f\,\varrho^2\, L, \qquad L = \log\left(b\,\mathrm{rd}_E\, g_0 g_1\, \varrho\right). \tag{III.209}$$

This is the deduction written in §III.4, with one difference. The mass bound in (III.63) is $\left((8f^2)^{\varrho}\varrho!\right)^{-1}$; here and in the Lean it is $\left(w_E^{\varrho}\varrho!\right)^{-1}$. The second implies the first, because $w_E \le 8f^2$. The reason: write $\mathrm{eul}(u)$ for Euler's function of an integer $u \ge 1$. The field E contains $\mathbb{Q}(\zeta_{w_E})$, which has degree $\mathrm{eul}(w_E)$ over $\mathbb{Q}$. So $\mathrm{eul}(w_E)$ divides $[E:\mathbb{Q}] = 2f$, and $\mathrm{eul}(w_E) \le 2f$. Next, $u \le 2\,\mathrm{eul}(u)^2$ for every $u \ge 1$: the quotient $\mathrm{eul}(u)^2/u$ is the product of the numbers $p^{e-2}(p-1)^2$ over the prime powers p^e that divide u exactly, and such a number is at least $\frac{4}{3}$ for $p \ge 3$ and at least $\frac{1}{2}$ for $p = 2$. Hence $w_E \le 2(2f)^2 = 8f^2$. The Lean has this inequality as a theorem, for the field E of its record `Host`: `host_torsionOrder_le`, which states `Units.torsionOrder H.E` $\le 8f^2$. The inner-product formula itself is used in the following form (Lean: the hypothesis `InnerProductFormulaC` of `term_eval_lowC`): for theta lifts $\theta(\phi_i, \mathfrak{f}_i)$, to the definite group, of forms $\mathfrak{f}_i$ on the group `G` lifted from,

$$\left\langle \theta(\phi_1,\mathfrak{f}_1), \theta(\phi_2,\mathfrak{f}_2)\right\rangle = \mathrm{vol}(\mathcal{F}) \int_{\mathsf{G}(\mathbb{A})} \langle \omega(g)\phi_1, \phi_2\rangle\, \overline{\langle g\cdot \mathfrak{f}_1, \mathfrak{f}_2\rangle}\, dg, \tag{III.210}$$

the integral being written as a sum over finitely many cosets at the finite places and an integral over the real points, absolutely convergent. This is the inner-product formula in the convergent range, as an integral of matrix coefficients; it is not the form with a special value of an L-function, and no Siegel–Weil formula appears in the statement. For the classical statements see S. Rallis, *L-functions and the oscillator representation*, Lecture Notes in Math. 1245 (1987) [Ral87] and *Injectivity properties of liftings associated to Weil representations*, Compositio Math. 52 (1984); J.-S. Li, *Non-vanishing theorems for the cohomology of certain arithmetic quotients*, J. reine angew. Math. 428 (1992) (the formula as an integral of matrix coefficients, in the stable range); W. T. Gan, Y. Qiu, S. Takeda, *The regularized Siegel–Weil formula (the second term identity) and the Rallis inner product formula*, Invent. Math. 198 (2014); S. Yamana, *L-functions and theta correspondence for classical groups*, Invent. Math. 196 (2014). All from memory; none was consulted. A proof of (S3) needs: this formula for the actual test functions; the unramified computation of the local integrals (the factors of E_0); the mass formula for definite unitary groups (for instance [GHY01]; its formula is for maximal lattices, and local comparisons would be needed for the levels used here); the formal degree of the discrete series at the real places (Harish-Chandra); the transport to the other embeddings (E7); and the description of the target by Hecke components (E6). The source has a theorem that derives the inner-product formula (`innerProductFormula_of_hypotheses`) from four hypotheses. Three are statements named `Published.CuspRapidDecay`, `WeilSiegelFormulaUnitary` and `SiegelEisensteinContinuation`. The fourth is the existence of a choice of parameters (Lean: a `ParameterChoice'` exists for some `SizeData`$_2$). Each of the three named statements has a closed theorem in the source. Their proofs were not read. The dependency computation shows that the final theorem depends on `innerProductFormula_of_hypotheses`; the three named statements were not put to it.

(S5), the slope at the lower level. In the source this is a theorem (`IndefSide.boundsMu`, from `slope_bounds`). Its conclusion is the display of (S5), with the sum of the $l_{3,j}$. Its hypotheses

are the following. The first is the covolume bound: the normalised logarithmic covolume of the lattice of the jth lift, divided by $\dim Y$, is at most $l_{3,j}$ (Lean: the field `l3` of `IndefSide`; in the source the covolume is minus the normalised slope). The others are all at finite places. For a finite place y of K write $\mathrm{wt}(y) = \log \mathrm{N}(y)/[K:\mathbb{Q}]$, $\mathrm{ram}(y)$ for its ramification index and $\mathrm{rc}(y)$ for its residue characteristic. There are a finite set $\mathcal{S}_{\mathrm{fin}}$ of places containing the set $\mathcal{S}_{\mathrm{tw}}$ of places where a twist is made, and real numbers $\mathrm{gd}(y)$, $\mathrm{ix}(y,j)$, $\mathrm{err}(y,j) \geq 0$, $\mathrm{jx}(y,j)$, $\mathrm{et}(y,j)$, such that: the valuation at y of a Gram determinant of the quotient lattice is 0 outside $\mathcal{S}_{\mathrm{fin}}$ and at most $\dim Y \cdot \mathrm{gd}(y)$ on $\mathcal{S}_{\mathrm{fin}}$; on $\mathcal{S}_{\mathrm{tw}}$ the sum over j of the two local indices of the lattice of the jth lift (at y and at $\bar{y}$) is at most $\big(\sum_j (2\,\mathrm{ram}(y)\,\mathrm{ix}(y,j) + \mathrm{ram}(y)\,\mathrm{err}(y,j)) + 2\,\mathrm{ram}(y)/\mathrm{rc}(y)^2\big)\dim Y$; a further local index is at most $\mathrm{ram}(y)\,\mathrm{jx}(y,j)\dim Y + \mathrm{et}(y,j)$; outside $\mathcal{S}_{\mathrm{tw}}$, for each j, the sum of these three local indices (the two at y and $\bar{y}$ and the further one) vanishes, and so does $\mathrm{ix}(y,j)$ (Lean: `IndefData.Untwisted`; it asks for the vanishing of the sum, not of each index). Then (S5) holds with $\mathfrak{F}$ as follows, and with every E_{loc} that is at least the sum written below:

$$\mathfrak{F} = \sum_j \sum_{y \in \mathcal{S}_{\mathrm{tw}}} \mathrm{wt}(y)\,\mathrm{ram}(y)\,\mathrm{ix}(y,j), \tag{III.211}$$

$$E_{\mathrm{loc}} = \sum_{y \in \mathcal{S}_{\mathrm{fin}}} \mathrm{wt}(y)\,\mathrm{gd}(y) + \sum_j \sum_{y \in \mathcal{S}_{\mathrm{tw}}} \mathrm{wt}(y)\Big(\mathrm{ram}(y)\big(\mathrm{err}(y,j) + \mathrm{jx}(y,j)\big) + \frac{\mathrm{et}(y,j)}{\dim Y}\Big) \tag{III.212}$$

$$+ \sum_{y \in \mathcal{S}_{\mathrm{fin}}} \frac{2\,\mathrm{wt}(y)\,\mathrm{ram}(y)}{\mathrm{rc}(y)^2}. \tag{III.213}$$

The display gives E_{loc} in the case of equality. In the source the four sums of E_{loc} are not computed: each is bounded by a free real number (Lean: `LG`, `E1`, `E2`, `E3`), and the conclusion has the sum of these majorants. The Lean `E3` is the last sum of E_{loc}. It is not the E_3 of (S6), which the Lean calls `El3`. It is a statement of the geometry of numbers over K: the covolume of a lattice glued from local lattices, compared with the covolumes of the lattices of the lifts through the scalar relations of (S3) (which is where $\langle \log \varkappa \rangle$ enters) and through local indices. No period occurs in it. The term $\mathfrak{F}$ is the term $F_0/2$ of Part II, §6.7. The local index bounds are separate inputs; those at the places of $\mathcal{S}_{\mathrm{tw}}$ are of the kind treated in §III.5. The proof in the source was not read.

(S6), the covolumes. In the source $l_{3,j} = \sum_\tau p_j w_{j\tau} \log \Omega_E + C_3\,B_0 \log\log x$ by definition, with a constant C_3 (Lean: `ClosedScale3.l3m`, `l3u`) and with $p_j = k + 2j - 1 - m$. So (S6) holds with $E_3 = m\,C_3\,B_0 \log\log x$ (`levelThree_bound_closed`). The content is therefore in the first hypothesis of the theorem quoted under (S5), the covolume bound (`IndefSide.l3`): the normalised logarithmic covolume of the lattice of the jth lift, divided by $\dim Y$, is at most $\sum_\tau p_j\,w_{j\tau} \log \Omega_E$ up to $C_3\,B_0 \log\log x$. This is the count of forms of Part II, §6.6 in its arithmetic form: the lattice is a space of algebraic automorphic forms of weight governed by $w_{j\tau}$, its metric is the Petersson norm, and the period appears because integral forms are normalised by the algebraic differential, with the same integral lines $\mathfrak{l}_\tau$ as in subsection 4. *That the constant here is the same Ω_E as in* (S4) *is exactly what makes the cancellation in Step 4 possible; it is assumed, not proved, in this section.* A proof needs the covolume of the lattice of integral forms on a definite unitary group for the Petersson metric (mass formula and dimension formula, as in (6.4) of Part II), with the period normalisation of subsection 4. The source derives it from a statement named `CovolumeBoundThirdPlaceSeries` (`linked_bundles_l3_of_open_alt`), whose hypotheses include statements on discrete series and on Harish-Chandra modules; it was not read beyond its header. One of its hypotheses is a covolume bound named `CovolumeBoundOfX`. The source proves that one (`topCovolSlim2_use_alt_alt`) from statements of algebraic and Arakelov geometry, which it takes as hypotheses: among them resolution of singularities and principalisation in characteristic zero, a Poincaré–Lelong formula, an arithmetic Hilbert–Samuel statement, a bigness statement named after Moriwaki, toroidal

compactifications, and a list named `PublishedHodgeBundleTheorems`. Only the headers of these two theorems were read.

(S7), the passage between levels. Two statements: the bound $\mu_{\max}(\mathcal{R}_0) \le H_{\max} - \langle \log \varkappa \rangle$ with $H_{\max} = x^{C_H} b$ for a constant C_H (Lean: `IndefExtraR.fun0`), and the comparison of $\mu(\mathcal{R}^\vee)$ with $\mu(\mathcal{R}_0^\vee)$ displayed in (S7) (`IndefExtraR.passage`). In the source la is the logarithm of the order of a finite group, and $\mathrm{de} = 8fr/\min(g_0, g_1)$ with $r = k + m - 1$. The number Q_0 is $\max(1, \cdot)$ of a majorant of the left side of its defining inequality: any upper bound for $\sum_j l_{3,j} + 2\mathfrak{F} + E_{\mathrm{loc}}$, plus $H_{\max} + 28\,\mathrm{la}$. The condition $3\,\mathrm{de}\,Q_0 \le 1$ then says that the level primes g_0, g_1 are at least $24\,f\,r\,Q_0$. (The source proves an inequality of this form, $24\,f\,r\,Q \le \min(g_0, g_1)$ for its number `QOfR`, from $g_0, g_1 \ge (kb)^8$ and further size conditions: `hg_of_sizeG`, `SizeG`. Only the statement was read.) So the passage costs a bounded amount because the level primes are large. The comparison is a statement about two lattices of forms of different level and the trace map between them; the first bound is a bound for the largest slope of a lattice of forms. Proofs were not read; the source reaches them through statements on Fourier coefficients (`fittedFunction_at_zero_close_alt`).

(S8), the size of the local errors, and **(G)**. These are inequalities between the parameters. (S8) contains the only occurrence of N_c. With $E_3 = m\,C_3\,B_0 \log\log x$ as above, and with the bounds for $\mathrm{wt}(y)\,\mathrm{gd}(y)$, err, jx, et supplied by the local statements, it holds for large x. The part of E_{loc} with $\mathrm{et}(y, j)$ is where the bound for Galois cohomology of Part II, §6.8 is used (Lean: `divinN_of_hypV`); this was read in the notes of another reader and not checked here. (G) is a list of size conditions on the parameters of the chain, valid for $x \ge x_0$ in the source; among them $|\log \Omega_E| \le L_a$ follows from $|c_d| \le 4 \log x$ (Theorem III.1.1(d)) for large x.

(W1), (W2). See remark (1) after the list of inputs: they follow from the conclusions of "Row by row", which are conditional on the inputs (o), (a)–(e) there, and from three facts of the source.

6. Assumed here, from the source

The Lean statements corresponding to the inputs. "Dependency: yes" means that the name is among the declarations the final theorem depends on, by the dependency computation; "reached" means that the name is not in the lists computed, but is used in the proof or in the definition of a name that is, either directly or through further declarations that are not in the lists. The route is given each time.

- `divClausesR_of` (dependency: yes): the assembly; its third clause is Theorem III.7.1(b) with the error $(50C_0 + 10)\,\mathcal{U}_1 + 60 c_e\,\mathcal{U}_2$, its second clause is Theorem III.7.1(a) in the form of a majorant at each place.
- (S1): fields `cb`, `nKcb` of `DivInputs0R` and `i2`–`i10`, `k2`, `k3` of `HtNums`: each matrix of the chain is a sum over classes of blocks with squared norm at most a bound whose logarithm is at most $\Sigma_j(v)$. `chain_coupled_bound` (yes) turns this into $\log \|K_j\|_v^2 \le \Sigma_j(v)$.
- (S2): field `nP` of `DivInputs0R` and `t1`–`t11`, `k1` of `HtNums`.
- (S3): fields `l1b`, `l1c`, `l6` of `IndefColsR` and `eig6` of `DivInputs0R`; `scalar_bound_lowC` (yes). The source has a theorem named `lamLow_of_classicalTheorems`; among its hypotheses are the Siegel–Weil formula for unitary groups with its convergence, the doubling method, and the unramified doubling computation for GL (`weilSiegelFormulaUnitary`, `siegelWeilUnitaryConvergent`, `doublingMethod`, `unramifiedDoublingGL`). By its name it gives the lower bound for the scalars. That it is what fills the field `l6` was not verified.

- (S4): field `i1` of `HtNums`; `meanPeriod` (yes), `mean_log_period` (yes), `chowla_selberg_eta` (yes). Reached: `two_log_brN` and `br_meanLogOm_eq`, through `meanPeriodAt`, which is the body of `meanPeriod` and is not in the lists; `index_lineR`, through `br_meanLogOm_eq`; `lineR_eq`, through `index_lineR`. The ideal `𝔟`: `principal`, `Arith.bb_ex`, field `𝔟_spec` of `Host` (dependency: not queried. The computation was asked for the name `KcDiff.principal`. It compares the last component of a name only, so this name cannot match, and the empty answer says nothing).
- (S5): `IndefSide.boundsMu` (yes), `slope_bounds` (yes), with the fields `bi`, `bLG`, `biv`, `borderedMinorImage`, `twisted_third_sum_le` of `DivInputs0R`, `on` and `l3` of `IndefSide` and `untw` of `IndefExtraR`.
- (S6): field `L3main` of `DivInputs0R`.
- (S7): fields `level`, `passage`, `fun0` of `IndefExtraR`; fields `de0`, `deQ`, `Q1`, `Qb`, `la_le`, `constant_ge_twenty_eight` of `DivInputs0R`.
- (S8): field `errB` of `DivInputs0R`.
- (W1), (W2), (G): fields `weight_height_sum_le`, `weight_sum_eq`, `sE`, `sδ`, `p1`–`p38`, `c1`–`c5` of `RecNums`; for remark (1), `NumFacts`, `stepE_num`, `RecNums.ofNum`.
- Step 1: `height_chain_atR` (yes); `height_constant_bound` (reached directly: it is used in the proof of `height_chain_atR`). Step 2: `height_bound_at_set` (yes). Step 4: `divisor_inequality_explicit` (yes), `divIneqRecord` (yes), `error_absorption` (yes); `divisor_inequality_assembly` (reached through `divisor_inequality`, which is not in the lists and is applied by `divisor_inequality_explicit`).
- The identifications $h = \mathcal{H} - \langle \log \varkappa \rangle$, $F_0 = 2\mathfrak{F}$, $\mu_{\mathrm{CC}} = \mu(\mathcal{R}^\vee)$ with the quantities of the final statement of the source: `hTie_of` (yes) and the hypotheses `tMu`, `tFo`, `tR` of `wordFacts_of_clauses`. In the source the function $h(u)$ of the final statement is defined by the formula for $\mathcal{H} - \langle \log \varkappa \rangle$ (`hExpl6`); so it already contains $-\langle \log \varkappa \rangle$.

7. What this section proves, and what it assumes

Proved: Theorem III.7.1: the inequality (III.75) at the end of §III.4 follows from (S1)–(S8), (W1), (W2), (G), with the constant $C = \max(50C_0 + 10, 60c_e)$; the descent lemma; the two lemmas on real numbers, with every constant; the table of the origin of each term. The Proposition of subsection 4: the mean (III.203) of the logarithms of the periods of the integral lines is $\log \Omega_E$, given the class mean formula (I2), (I3) of §III.4, the multiplicativity of the different and the formula for its norm (cited), and the ideal $\mathfrak{b}$; Lemmas B, C, D are proved. The existence of $\mathfrak{b}$ for the field E of the paper, with (P2), on which it rests. Remark (1) after the list of inputs: (W1), (W2) from the conclusions of "Row by row" and three facts of the source.

Assumed: (S1)–(S8), (W1), (W2), (G). The input (S4) is not proved: it would follow from the Proposition and (P1)–(P3), and (P1), (P3) are assumed from the source. Of the inputs, (S1), (S2), (S3), (W1), (W2) correspond to conclusions of conditional paragraphs of §III.4, each with inputs of its own that are named there. For (S1) and (S2) the exact match of the numerical forms (the constants of (S1), the weights in (S2)) is part of what is assumed. For (W1), (W2) the match is made in remark (1), given the three facts quoted there.

8. What remains

(i) The covolume bound with the period Ω_E of subsection 4 (the first hypothesis of the theorem behind (S5), with the values of $l_{3,j}$ in (S6)): the statement that the lattice of integral forms of the jth lift has normalised covolume at most $\sum_\tau p_j w_{j\tau} \log \Omega_E$ up to the stated error. This is the largest gap: the cancellation of the period is proved here only as an identity between two inputs.

(ii) The slope bound (S5) and its local index bounds, including the definition of $\mathfrak{F}$.

(iii) The passage between levels (S7).

(iv) The inner-product formula, the mass, the formal degrees and the local factors, for (S3).

(v) The statements (P1) and (P3) on the periods of the actual kernels, and item (vi) of the list of §III.4.

(vi) The verification of (G) and (S8) for the parameters of Part II for large x, and the bounds for the constants and the class numbers in (S1).

III.8 The images in general position: the trace-formula estimates

This section concerns the two smallness statements that the linear algebra of Part II, §6.9–6.10 consumes, and the identity that connects them with the pairings of the chain of lifts. In the Lean development these are three statements: `LinkS` (bounds for the Gram blocks of a family of maps; the trace statement), `LinkIRR` (a count; the rank statement), and `BesselId` (the pairing at the end of the chain is an inner product with the image of such a map; the pairing statement). The title needs a word of explanation. The source does not prove that the images of the maps are nearly orthogonal. The images cannot all be mutually orthogonal: by the trace statement their dimensions add up to more than the dimension of the target (Remark (1) of subsection 2). What the source bounds is this: every block $B_{\mathsf{w}}^\dagger B_{\mathsf{w}'}$ with $\mathsf{w} \neq \mathsf{w}'$ has at most $(1+\eta_0)(1+\eta_1)$ times the mean size that it has for subspaces in general position, where η_0 and η_1 are the numbers of Statement (S) of subsection 1, at most $\exp(-x^{\varepsilon/8})$. So the operator $Q = \sum_{\mathsf{w}} B_{\mathsf{w}} B_{\mathsf{w}}^\dagger$ is close to its mean eigenvalue in Hilbert–Schmidt norm (Remark (1) of subsection 2). The operator Q is not a sum of projections. *Proved here:* Lemma T1 and Corollary T2. This section deduces the dimension bound from the three trace, rank and pairing statements and the size and subspace conditions in T2, with explicit constants. It also proves Lemma T3, a statement about finite sets and weights which is the mechanism of a second moment in the simplest case (a sum over twists of $B_\chi^\dagger B_\chi$ is a count of pairs of representations, and its main term is a mass). *Quoted from the source:* the three statements themselves, transcribed in subsection 1, and the theorems from which the source deduces them, listed in subsection 4 with their Lean names. The construction of the spaces and maps, the arithmetic inputs and the assembly of (M) are quoted. None of these is proved here. The error term that Part II, §6.10 takes from this section is recalled at the end of subsection 2. Below, §1, §2, ... without a part number mean the subsections of this section.

Letters introduced in this section: W, w (words); $\mathcal{H}$, $\mathcal{C}_{\mathsf{w}}$, $\mathcal{M}_{\mathsf{w}}$, $\mathcal{E}_0$, $\mathcal{E}$, $\mathcal{U}_{\mathsf{w}}$ (spaces), with $\mathcal{M}_{\mathsf{w},\mathbb{L}}$, $\mathcal{E}_{\mathbb{L}}$ (spaces over a number field $\mathbb{L}$); P_{w}, B_{w}, $K_{\mathsf{w}\mathsf{w}'}$, s, $\mathfrak{B}_{\mathsf{w}}$, β_{w}, Ψ, Ψ_{w}, Λ, $\pi_{\mathcal{E}}$ (maps and pairings); d_{w}, $d_{\mathsf{w}}^{\mathrm{ram}}$, D_*, D_{r} (dimensions); $\eta_0, \eta_1, e_0, \delta_{\mathrm{p}}, \delta_{\mathcal{E}}, \varkappa, \alpha_{\mathrm{ram}}, \alpha', \varepsilon_z, \varepsilon_C$ (small numbers); G, C_{G}, C_0, C_H, J_z, $c_0(\mathsf{w})$, λ_{w} (constants); Wg, $\mathcal{T}(\mathsf{w})$, ξ_i, ψ_i, θ_j, e'_j, Δ (in §2, §4, §5); and, in §3 only, $\mathcal{I}, \mathcal{I}', \omega, \omega', G_i, \mathrm{Rep}, A, \mathrm{inv}, \Theta_\chi, B_\chi, \mathcal{N}, \mathcal{N}^\circ, \mathfrak{N}^\circ, \mathfrak{m}, \sigma, \ell_0, \varsigma_\iota$. The letters Q, D, S, e, η, α, u, H have the meaning of Part II, §6.9–6.10: Q is the sum of the operators $B_{\mathsf{w}} B_{\mathsf{w}}^\dagger$, D the dimension of the

target, S the total dimension of the sources, e, η, α the three constants of the lemma of §6.9, and u, H the spread and the height bound of §6.10. From Corollary T2 on, the share of directions lost at the level primes is written α_{ram}; it is one part of the constant α of the lemma, which in Corollary T2 is $\alpha_{\mathrm{ram}} + \alpha'$, and it is not the $\alpha = u/(u+t)$ of §6.10. The target space, called H in §6.9, is written $\mathcal{H}$ here. The letters k, m, t_0, f, r, g_0, g_1, q, d, b have the meaning of §III.4 and §III.5: k is the initial rank, m the number of lifts, t_0 the integer parameter written t in §III.4 and t_0 in §III.5 (the field `t` of the Lean record of ranks), $f = 2n$ the number of weight places, $r = k + m - 1$, g_0, g_1 are the two level primes, q, d the auxiliary primes, and $b = \exp(x^{\varepsilon})$ with $0 < \varepsilon < 1/52$. $\|A\|_{\mathrm{HS}}^2 = \mathrm{tr}(A^\dagger A)$ is the squared Hilbert–Schmidt norm, and $A^\dagger$ the adjoint. Inner products are conjugate-linear in the first variable and linear in the second. $V^\vee$ is the space of $\mathbb{C}$-linear forms on a complex space V. The letter ν denotes a dimension where it occurs, and j, j', l, ι are indices. The exponential function is written exp in this section, because e is a constant of the lemma of §6.9.

Other letters. Used in more than one place: φ and F (elements of $\mathcal{M}_{\mathsf{w}}$ or $\mathcal{M}_{\mathsf{w},\mathbb{L}}$, and of $\mathcal{E}_0$ or $\mathcal{E}_{\mathbb{L}}$); x_0 (a threshold for x) and c, c', C, C' (positive constants), which are not the same at each occurrence; y (a real number in the proof of Lemma T3, and $x^{\varepsilon/8}$ in §4); i (an index in c_i, ξ_i, ψ_i, an element of $\mathcal{I}$ in §3, and the imaginary unit in iF in §1); v (a finite place in (R1) and (R2), a positive real number in (R3)); X (a tuple of vectors in (R2), the variable of a polynomial in (R4)); g (a group element in (R2), a prime in (R4)); M (a rank, in (R2) and (R3)). Used in one place only and explained there: s (the inner mean of §2, a real number; it is not the map `s`); $\mathcal{X}$, $\mathcal{Z}$, $\mathcal{Y}$, h (in the proofs of Lemma T1(d) and Corollary T2); Ξ, ζ, ζ', p (in the Remarks); Y, ρ, χ (in §3); τ, $K^{(\tau)}$, K'', s_0 (in (R1)); E/E^+, J, Z, Φ, ϱ_v (in (R2)). The number e_0' of §5 is the first of the e_j'; it is not the e_0 of (S).

The letters are local. No letter of this section was changed to agree with the other sections of Part III, and several have another meaning there. The clashes most likely to confuse are these. C_0 is 3 here; in §III.7 it is a constant with $C_0 \geq 28$. $\mathcal{H}$ is the target space here; in §III.7 it is a height. G is the number $\exp(C_{\mathsf{G}} m^5 f x^{\varepsilon})$ here and in Part II, §6.10; in §III.7 it is a group. $\mathcal{T}(\mathsf{w})$ is a level prime here, as $\mathcal{T}$ is in Part II, §6.7; in §III.6 $\mathcal{T}$ is a set. ς_ι are minors here (§3); in §III.4 and §III.7 ς is a number formed from the ranks. Λ is a map here; in §III.6 the Λ_w are lattices, and elsewhere Λ is the von Mangoldt function. Ξ is a random unitary map here; in §III.6 it is a set of primes and in §III.7 an operator. Ψ, Ψ_{w} are unitary maps and ψ_i weights here; in §III.4 $\Psi = \psi + 1$ and ψ are ranks, and in §III.9 Ψ is a character. Δ is the step of the weights here (§5); in §III.4 it is the width of the weight interval. Also: α is a constant here and a member of the family in §III.6 (the word of this section; see §1); $\mathfrak{B}$ and c_0 occur in §III.9, $\varkappa_v$ in §III.7 and $\mathcal{Y}$ in §III.5, with other meanings.

1. The three statements in the source

The data. At a scale x the source works with a record (Lean: `D7`) which contains, among many other things, the following. They are fields of the record. Their identification with spaces of automorphic forms is made in the source by further hypotheses, which were not read (see §6).

- A finite set W of *words*. Each word w has a sequence of letters in $\{0,1\}^{k-2}$, and distinct words have distinct sequences (`letter`, `letter_inj`). A word fixes, at each of the f weight places, a weight $\xi = (\xi_0 > \cdots > \xi_{k-1})$ of a unitary group in k variables; see §5.
- A complex inner product space $\mathcal{H}$ of finite dimension D (`HH`, `Dall`), and a real number $D_* > 0$ (`Dstar`). The positivity of D_* is not a field of the record. It follows from two fields of other records: `Dstar_eq` and `Dstar_pos`.
- For each word a complex inner product space $\mathcal{C}_{\mathsf{w}}$ of finite dimension d_{w} (`Cu`, `duS`), a subspace

$\mathcal{M}_{\mathsf{w}} \subset \mathcal{C}_{\mathsf{w}}$ of dimension $d^{\mathrm{ram}}_{\mathsf{w}}$ (`Mram`, `du`), a linear map $P_{\mathsf{w}} : \mathcal{C}_{\mathsf{w}} \to \mathcal{H}$ (`Pu`), and a number $c_0(\mathsf{w}) \geq 0$ (`c0`, a product of two volumes divided by a dimension). When $c_0(\mathsf{w}) > 0$ put

$$B_{\mathsf{w}} = c_0(\mathsf{w})^{-1/2} P_{\mathsf{w}}, \qquad K_{\mathsf{w}\mathsf{w}'} = B^{\dagger}_{\mathsf{w}} B_{\mathsf{w}'} : \ \mathcal{C}_{\mathsf{w}'} \to \mathcal{C}_{\mathsf{w}}. \tag{III.214}$$

When $c_0(\mathsf{w}) = 0$ the definitions of the source give $K_{\mathsf{w}\mathsf{w}} = 0$. In the notation of Part II, §6.9 the maps B_i are the P_{w}, the spaces V_i are the $\mathcal{C}_{\mathsf{w}}$, and the weights are $c_i = 1/c_0(\mathsf{w})$ (the letter c_i is that of §6.9; the number $c_0(\mathsf{w})$ of the record is not one of the c_i); so $K_i = K_{\mathsf{w}\mathsf{w}}$ and $Q = \sum_{\mathsf{w}} B_{\mathsf{w}} B^{\dagger}_{\mathsf{w}}$.

- A complex inner product space $\mathcal{E}_0$ of dimension D_{r} (`EE`), and a *conjugate-linear* map $\mathsf{s} : \mathcal{E}_0 \to \mathcal{H}$ which preserves norms (`sEE`). So $\mathsf{s}(cF) = \bar{c}\,\mathsf{s}(F)$ for $c \in \mathbb{C}$, and s is injective. Its image $\mathcal{E} = \mathsf{s}(\mathcal{E}_0)$ is a complex subspace of $\mathcal{H}$ of dimension D_{r}. There is a number $\delta_{\mathrm{p}} \leq 1$ with $D_{\mathrm{r}} = (1 - \delta_{\mathrm{p}})D$ (`Drows`, `delp`, `D_eq`). The spaces $\mathcal{E}_0$ and $\mathcal{E}$ have the same elements up to `s`, with conjugate actions of the scalars. They are kept apart below.

- For each word a pairing $\mathfrak{B}_{\mathsf{w}} : \mathcal{M}_{\mathsf{w}} \times \mathcal{E}_0 \to \mathbb{C}$, linear over $\mathbb{C}$ in each variable (`Bu`, `Bu_linL`, `Bu_linR`). In the source it stands for the pairing at the end of the chain of m lifts of Part II, §6.2, between the space at the start of the chain and the space at its end (it is tied to the chain by the fields `perCol` and `PtilCol`, which were not read). Both spaces carry structures over a number field $\mathbb{L} \subset \mathbb{C}$: there are $\mathbb{L}$-vector spaces $\mathcal{M}_{\mathsf{w},\mathbb{L}}$ and $\mathcal{E}_{\mathbb{L}}$, with lattices in them, and maps $\mathcal{M}_{\mathsf{w},\mathbb{L}} \to \mathcal{M}_{\mathsf{w}}$ and $\mathcal{E}_{\mathbb{L}} \to \mathcal{E}_0$, linear over $\mathbb{L}$ (`jN`, `jEE`), under which an $\mathbb{L}$-basis of $\mathcal{M}_{\mathsf{w},\mathbb{L}}$ is a $\mathbb{C}$-basis of $\mathcal{M}_{\mathsf{w}}$ and an $\mathbb{L}$-basis of $\mathcal{E}_{\mathbb{L}}$ is a $\mathbb{C}$-basis of $\mathcal{E}_0$ (`jN_basis`, `jEE_basis`: linear independence and span). The second map goes to $\mathcal{E}_0$ and not to $\mathcal{E}$: composed with `s` it would be conjugate-semilinear over $\mathbb{L}$, and so linear over $\mathbb{L}$ only if $\mathbb{L} \subset \mathbb{R}$. In particular

$$\dim_{\mathbb{L}} \mathcal{E}_{\mathbb{L}} = D_{\mathrm{r}} \tag{III.215}$$

(`RRV_finrank`). There is an $\mathbb{L}$-bilinear pairing $\mathfrak{B}^{\mathrm{alg}}_{\mathsf{w}} : \mathcal{M}_{\mathsf{w},\mathbb{L}} \times \mathcal{E}_{\mathbb{L}} \to \mathbb{L}$ (`Bk`). The slopes of Part II, §6.8–6.10 are those of these lattices.

Comparison with the letters of §III.6. The following was checked in the types of the Lean record. The field $\mathbb{L}$ is the coefficient field L_0 of §III.6 (`LL`, with the embedding `embLL`). The words are the members α of §III.6 (`Word`). The space $\mathcal{M}_{\mathsf{w},\mathbb{L}}$ is the space U_α that carries the glued lattice $\Lambda^{\mathrm{gl}}_\alpha$ (`Nu`), and $\mathcal{E}_{\mathbb{L}}$ is the space U' with the lattice $\mathcal{R}$ (`RR`): the maps `jN` and `jEE` have the sources `(div.Nu u).V` and `div.RR.V`. Not checked in the Lean: that $\mathfrak{B}^{\mathrm{alg}}_{\mathsf{w}}$ (`Bk`) is the pairing $\langle\ ,\ \rangle_\alpha$ of §III.6 (`perCol`), and that the level prime $\mathcal{T}(\mathsf{w})$ of §2 is the prime Lv_α of §III.6; both sections describe them in the same words. The letters of §III.7 were not compared. *The index of the family is the word.* It is not a Hecke character: in the source the twists by Hecke characters form another index (`Tup`), with one twist attached to each place (`muAt`). By our reading it serves to glue the lattices (Part II, §6.8). The space $\mathcal{E}_0$ and the pairings $\mathfrak{B}_{\mathsf{w}}$ do not depend on it: `Tup` does not occur in the types of `EE` and `Bu`, nor in `LinkS_at`, `Frame_at` and `frame_at_of`.

Translates. There is one record for each translate of Part II, §4.2, and each record has its own set of words (Lean: the definition `data` in `QuadraticLogDerivCongr.lean`, whose family of index sets is `fun u0 => (Fm.D x (M.pt u0)).what₂.Word`; there `u0` is the translate and `u` the word). The weights of the table of Part II, §4.2 depend on the translate and not on the word (`w u0 j`). The level prime, the height and the rank depend on the translate and on the word (`T u0 u`, `h u0 u`, `d u0 u`).

Statement (S) (Lean: `LinkS_at`, `LinkS`, with $C_0 = 3$). *Put* $\mathsf{G} = \exp(C_{\mathsf{G}}\, m^5 f\, x^{\varepsilon})$, *where* $C_{\mathsf{G}} > 0$ *is a constant of the family. There is* x_0 *such that for every* $x \geq x_0$ *and every record of the*

family at the scale x there are real numbers η_0, η_1 and $e_0 \geq 0$ with

$$\eta_0 \leq \exp(-x^{\varepsilon/8}), \qquad \eta_1 \leq \exp(-x^{\varepsilon/8}), \qquad \sqrt{e_0} \leq \exp(-x^{\varepsilon/8}), \tag{III.216}$$

such that, if

$$C_0^{-1} \leq d_{\mathsf{w}}\, \mathsf{G}^2/D_* \leq C_0 \qquad \text{for every word } \mathsf{w}, \tag{III.217}$$

then

$$\|K_{\mathsf{w}\mathsf{w}'}\|_{\mathrm{HS}}^2 \leq (1+\eta_0)\, \frac{d_{\mathsf{w}} d_{\mathsf{w}'}}{D_*} \qquad (\mathsf{w} \neq \mathsf{w}'), \tag{III.218}$$

$$\|K_{\mathsf{w}\mathsf{w}} - 1\|_{\mathrm{HS}}^2 \leq e_0\, d_{\mathsf{w}}, \tag{III.219}$$

$$D \leq (1+\eta_1)\, D_*. \tag{III.220}$$

In the proof in the source the three numbers are $\eta_0 = b^{-6}$, $e_0 = 2C_0/\mathsf{G} + 2\exp(-kf\log k)$ and $\eta_1 = \exp(-c\, r f \log r)$ with a constant $c > 0$ (`linkS_concl`). So they are of size $\exp(-c'x^{\varepsilon})$, and (III.216) is a weaker, uniform way of recording this. When (III.217) fails for a record, that proof takes $\eta_0 = \eta_1 = e_0 = 0$, and the implication is empty (`linkS_of_standing_hypotheses`). The source proves (S) for families satisfying a further condition on the level primes of the words (`TuSize`).

Statement (I) (Lean: `LinkIRR`). *There is x_0 such that for $x \geq x_0$ and every record of the family at the scale x*

$$0 \leq \delta_{\mathrm{p}} \leq \exp\big(-x^{\varepsilon_z/7}\big), \qquad \varepsilon_z = \varepsilon/J_z, \tag{III.221}$$

where $J_z \geq 1$ is a constant of the family. So $\mathcal{E}$ is almost all of $\mathcal{H}$.

Statement (P) (Lean: `BesselId`). *For every record:* (P1) *for $\varphi \in \mathcal{M}_{\mathsf{w},\mathbb{L}}$ and $F \in \mathcal{E}_{\mathbb{L}}$, $\mathfrak{B}_{\mathsf{w}}(\varphi, F) = \mathfrak{B}_{\mathsf{w}}^{\mathrm{alg}}(\varphi, F)$, the right side being read in $\mathbb{C}$ through the fixed embedding of $\mathbb{L}$, and φ, F on the left being their images in $\mathcal{M}_{\mathsf{w}}$ and $\mathcal{E}_0$;* (P2) *there are a unitary map Ψ of $\mathcal{E}_0$ and, for each word, a complex number $\lambda_{\mathsf{w}} \neq 0$ and a unitary map Ψ_{w} of $\mathcal{M}_{\mathsf{w}}$, such that*

$$\mathfrak{B}_{\mathsf{w}}(\varphi, F) = \lambda_{\mathsf{w}} \left\langle \mathsf{s}\Psi F,\ P_{\mathsf{w}}\Psi_{\mathsf{w}}\varphi \right\rangle_{\mathcal{H}} \qquad (\varphi \in \mathcal{M}_{\mathsf{w}},\ F \in \mathcal{E}_0). \tag{III.222}$$

Both sides are linear in φ and in F. For the right side and F: replace F by iF. Then ΨF is multiplied by i, since Ψ is linear; $\mathsf{s}\Psi F$ is multiplied by $-i$, since s is conjugate-linear; and the inner product is multiplied by i, since it is conjugate-linear in its first variable. The map s cannot be omitted: for a complex subspace $\mathcal{E}$ of $\mathcal{H}$, a pairing that is linear on $\mathcal{M}_{\mathsf{w}} \times \mathcal{E}$ cannot be equal to $\lambda_{\mathsf{w}}\langle \Psi F, P_{\mathsf{w}}\Psi_{\mathsf{w}}\varphi\rangle$ with Ψ unitary on $\mathcal{E}$, unless it is zero: the left side would be multiplied by i and the right side by $-i$. In the Lean, `sEE` has the type of a conjugate-linear isometry from `EE` to the target, the pairing `besS` is `inner (sEE F) (Pu ...)`, and `BesselId` asserts `Bu u phi F = a * besS u (U phi) (V F)` with linear isometric equivalences `V` of `EE` and `U` of the space of φ. In the source (P) is a clause of the statement `LinkDIV_at`. The name refers to the realisation of $\langle \mathsf{s}F, P_{\mathsf{w}}\varphi\rangle$ in the source as a Bessel period of F; in the form (III.222) no integral appears.

2. From the three statements to the hypotheses of Part II, §6.9

The lemma of Part II, §6.9 has three hypotheses, with constants e, η, α: a bound $\sum_i \|K_i - 1\|_{\mathrm{HS}}^2 \leq eS$ for the diagonal blocks; a bound $\mathrm{tr}(Q^2) \leq (1+\eta)(\mathrm{tr}\, Q)^2/D$ for the operator Q; and subspaces of total codimension at most αS. Its conclusion is that the images span all of the target but a part of dimension at most $(\eta + 4\alpha + 4\sqrt{e})D$. This is the form that the proof in §6.9 gives. The display of §6.9 has 2η in place of η, which is weaker. This section uses the form with η everywhere, and

so does the source (`FrameLA1`; see below). The next lemma obtains the first two hypotheses from (III.218)–(III.220).

Lemma T1. *Let $\mathcal{H}$ be a complex inner product space of dimension $D \geq 1$, let W be a finite nonempty set, and for $\mathsf{w} \in \mathsf{W}$ let $B_{\mathsf{w}} : \mathcal{C}_{\mathsf{w}} \to \mathcal{H}$ be a linear map from an inner product space of dimension $d_{\mathsf{w}} \geq 1$. Put $K_{\mathsf{ww'}} = B_{\mathsf{w}}^\dagger B_{\mathsf{w'}}$, $Q = \sum_{\mathsf{w}} B_{\mathsf{w}} B_{\mathsf{w}}^\dagger$ and $S = \sum_{\mathsf{w}} d_{\mathsf{w}}$.*

(a) $\operatorname{tr} Q = \sum_{\mathsf{w}} \operatorname{tr} K_{\mathsf{ww}}$ *and* $\operatorname{tr}(Q^2) = \sum_{\mathsf{w,w'}} \|K_{\mathsf{ww'}}\|_{\mathrm{HS}}^2$.

(b) *Let $D_* > 0$, $\eta_0 \geq -1$, $e_0 \geq 0$, and assume* (III.218) *and* (III.219). *Then*

$$\operatorname{tr}(Q^2) \leq (1+\eta_0)\frac{S^2}{D_*} + (1+\sqrt{e_0})^2 S, \qquad \operatorname{tr} Q \geq (1-\sqrt{e_0})S, \qquad \sum_{\mathsf{w}} \|K_{\mathsf{ww}} - 1\|_{\mathrm{HS}}^2 \leq e_0 S. \tag{III.223}$$

(c) *Assume in addition* (III.220), *and let $0 < \varkappa \leq \frac{1}{4}$ be such that*

$$D_* \leq \varkappa S, \qquad \eta_0 \leq \varkappa/2, \qquad \eta_1 \leq \varkappa/2, \qquad \sqrt{e_0} \leq \varkappa/2. \tag{III.224}$$

Then

$$\operatorname{tr}(Q^2) \leq (1+5\varkappa)\,\frac{(\operatorname{tr} Q)^2}{D}, \qquad e_0 \leq \tfrac{1}{64}. \tag{III.225}$$

(d) *Under the hypotheses of* (c), *let $\mathcal{U}_{\mathsf{w}} \subset \mathcal{C}_{\mathsf{w}}$ be subspaces with $\sum_{\mathsf{w}}(d_{\mathsf{w}} - \dim \mathcal{U}_{\mathsf{w}}) \leq \alpha S$, let $\mathcal{E} \subset \mathcal{H}$ be a complex subspace of dimension $(1-\delta_{\mathcal{E}})D$ with $0 \leq \delta_{\mathcal{E}} < 1$, and let $\pi_{\mathcal{E}}$ be the orthogonal projection onto $\mathcal{E}$. Then*

$$\dim \mathcal{E} - \dim \sum_{\mathsf{w}} \pi_{\mathcal{E}} B_{\mathsf{w}} \mathcal{U}_{\mathsf{w}} \;\leq\; \frac{5\varkappa + 4\alpha + 4\sqrt{e_0}}{1-\delta_{\mathcal{E}}}\, \dim \mathcal{E}. \tag{III.226}$$

Proof. (a) The first identity is $\operatorname{tr}(BB^\dagger) = \operatorname{tr}(B^\dagger B)$. For the second, $K_{\mathsf{ww'}}^\dagger = K_{\mathsf{w'w}}$, so

$$\operatorname{tr}\left(B_{\mathsf{w}} B_{\mathsf{w}}^\dagger B_{\mathsf{w'}} B_{\mathsf{w'}}^\dagger\right) = \operatorname{tr}\left(B_{\mathsf{w'}}^\dagger B_{\mathsf{w}}\, B_{\mathsf{w}}^\dagger B_{\mathsf{w'}}\right) = \operatorname{tr}\left(K_{\mathsf{ww'}}^\dagger K_{\mathsf{ww'}}\right) = \|K_{\mathsf{ww'}}\|_{\mathrm{HS}}^2, \tag{III.227}$$

and $\operatorname{tr}(Q^2)$ is the sum of these over all pairs.

(b) Let K be an operator on a space of dimension ν. By the Cauchy–Schwarz inequality for the Hilbert–Schmidt inner product, $|\operatorname{tr}(K-1)| \leq \sqrt{\nu}\,\|K-1\|_{\mathrm{HS}}$, so $\operatorname{Re}\operatorname{tr} K \geq \nu - \sqrt{\nu}\,\|K-1\|_{\mathrm{HS}}$; and $\|K\|_{\mathrm{HS}} \leq \sqrt{\nu} + \|K-1\|_{\mathrm{HS}}$. With (III.219) this gives $\operatorname{tr} K_{\mathsf{ww}} \geq (1-\sqrt{e_0})d_{\mathsf{w}}$ (the trace is real) and $\|K_{\mathsf{ww}}\|_{\mathrm{HS}}^2 \leq (1+\sqrt{e_0})^2 d_{\mathsf{w}}$. Sum over w and use (a). For the pairs $\mathsf{w} \neq \mathsf{w'}$ use (III.218) and $\sum_{\mathsf{w}\neq\mathsf{w'}} d_{\mathsf{w}} d_{\mathsf{w'}} \leq S^2$; here $1+\eta_0 \geq 0$ is used.

(c) By (III.220), $1+\eta_1 \geq D/D_* > 0$ and $1/D \geq 1/((1+\varkappa/2)D_*)$. By (b) and (III.224), $\operatorname{tr} Q \geq (1-\varkappa/2)S \geq 0$, and, since $S \leq \varkappa S^2/D_*$,

$$\operatorname{tr}(Q^2) \leq \Big((1+\tfrac{\varkappa}{2}) + (1+\tfrac{\varkappa}{2})^2\varkappa\Big)\frac{S^2}{D_*}, \tag{III.228}$$

$$(1+5\varkappa)\frac{(\operatorname{tr} Q)^2}{D} \geq \frac{(1+5\varkappa)(1-\frac{\varkappa}{2})^2}{1+\frac{\varkappa}{2}} \cdot \frac{S^2}{D_*}. \tag{III.229}$$

So it suffices that $(1+\frac{\varkappa}{2})^2 + (1+\frac{\varkappa}{2})^3\varkappa \leq (1+5\varkappa)(1-\frac{\varkappa}{2})^2$. Expanding both sides, the difference of the right side and the left side is

$$\varkappa(2 - \tfrac{13}{2}\varkappa + \tfrac{1}{2}\varkappa^2 - \tfrac{1}{8}\varkappa^3), \tag{III.230}$$

which is positive for $0 < \varkappa \le \frac{1}{4}$, since then the bracket is at least $2 - \frac{13}{8} - \frac{1}{512} > 0$. Finally $e_0 \le \varkappa^2/4 \le 1/64$.

(d) By (b) and (c) the hypotheses of the lemma of Part II, §6.9 hold for the maps B_{w}, with all weights c_i equal to 1, with $e = e_0 \le \frac{1}{4}$, $\eta = 5\varkappa$, and the given α. So, by the form of that lemma with η which its proof gives, the orthogonal complement $\mathcal{Z}$ of $\mathcal{X} = \sum_{\mathsf{w}} B_{\mathsf{w}}\mathcal{U}_{\mathsf{w}}$ in $\mathcal{H}$ has dimension at most $(5\varkappa + 4\alpha + 4\sqrt{e_0})D$. Now $\pi_{\mathcal{E}}(\mathcal{X}) = (\mathcal{X} + \mathcal{E}^\perp) \cap \mathcal{E}$, and the codimension of this space in $\mathcal{E}$ equals the codimension of $\mathcal{X} + \mathcal{E}^\perp$ in $\mathcal{H}$, because $\mathcal{X} + \mathcal{E}^\perp = \pi_{\mathcal{E}}(\mathcal{X}) \oplus \mathcal{E}^\perp$ and $\mathcal{H} = \mathcal{E} \oplus \mathcal{E}^\perp$. This codimension is at most that of $\mathcal{X}$, which is $\dim \mathcal{Z}$. Finally $D = \dim \mathcal{E}/(1 - \delta_{\mathcal{E}})$. □

Remarks. (1) *The blocks are not small, and the images are not nearly orthogonal.* For every family, $\operatorname{tr}(Q^2) \ge (\operatorname{tr} Q)^2/D$, by the Cauchy–Schwarz inequality for the eigenvalues of Q. If all B_{w} are isometries, then $K_{\mathsf{ww}} = 1$, and this and (a) give

$$\sum_{\mathsf{w} \ne \mathsf{w}'} \|K_{\mathsf{ww}'}\|_{\mathrm{HS}}^2 \;\ge\; \frac{S^2}{D} - S. \tag{III.231}$$

The sum of the bounds $d_{\mathsf{w}} d_{\mathsf{w}'}/D$ over the same pairs is $(S^2 - \sum_{\mathsf{w}} d_{\mathsf{w}}^2)/D$, which exceeds the right side by at most S, that is, by the relative amount D/S at most. By (III.220), $1/D_* \le (1 + \eta_1)/D$. So (III.218) and (III.220) assert that the squared norm of every off-diagonal block is at most $(1 + \eta_0)(1 + \eta_1)$ times $d_{\mathsf{w}} d_{\mathsf{w}'}/D$. Up to this factor and up to the relative error D/S, that is the least size that the blocks can have on average. Statement (S) bounds D by D_* from above only; the bound in the other direction is in (R3) of §4 and is not used in this section. The number $d_{\mathsf{w}} d_{\mathsf{w}'}/D$ is also the mean size for subspaces in general position: if Ξ is a random unitary map of $\mathcal{H}$ for Haar measure and ζ, ζ' are unit vectors, the mean of $|\langle \zeta, \Xi\zeta' \rangle|^2$ is $1/D$, so the mean of $\|B^\dagger \Xi B'\|_{\mathrm{HS}}^2$ for isometries B, B' of ranks ν, ν' is $\nu\nu'/D$. Orthogonal images would mean $K_{\mathsf{ww}'} = 0$. Mutually orthogonal images have dimensions that add up to at most D. For injective maps this is impossible when $S > D$. The maps of the source are not known to be injective, but (III.219) bounds their ranks from below. The operator $K_{\mathsf{ww}} - 1$ is -1 on the kernel of B_{w}, so $\|K_{\mathsf{ww}} - 1\|_{\mathrm{HS}}^2 \ge d_{\mathsf{w}} - \operatorname{rank} B_{\mathsf{w}}$, and (III.219) gives $\operatorname{rank} B_{\mathsf{w}} \ge (1 - e_0)\, d_{\mathsf{w}}$. Under the hypotheses of Corollary T2 below, $e_0 \le \frac{1}{64}$, $S \ge x^{C_H} D_* \ge 4D_*$ and $D \le \frac{9}{8} D_*$. So the ranks add up to at least $\frac{63}{64} S > D$, which is more than the dimension of the target, and the images cannot all be mutually orthogonal. The Gram block estimates control their total squared overlap and put Q close to its mean eigenvalue in Hilbert–Schmidt norm. Precisely, for a Hermitian operator Q on $\mathcal{H}$ one has $\operatorname{tr}(Q^2) - (\operatorname{tr} Q)^2/D = \|Q - (\operatorname{tr} Q/D) \cdot 1\|_{\mathrm{HS}}^2$, so under the hypotheses of Lemma T1(c) the bound (III.225) is

$$\Big\| Q - \frac{\operatorname{tr} Q}{D} \cdot 1 \Big\|_{\mathrm{HS}}^2 \;\le\; 5\varkappa \, \frac{(\operatorname{tr} Q)^2}{D}. \tag{III.232}$$

The operator $Q = \sum_{\mathsf{w}} B_{\mathsf{w}} B_{\mathsf{w}}^\dagger$, with $B_{\mathsf{w}} = c_0(\mathsf{w})^{-1/2} P_{\mathsf{w}}$ for the maps of the source, is not the sum of the projections onto the images: $B_{\mathsf{w}} B_{\mathsf{w}}^\dagger$ is a projection only if B_{w} is a partial isometry. (2) *The constant* 5. The term $(1 + \sqrt{e_0})^2 S$ in (III.223) comes from the diagonal blocks. It is the reason why (III.225) needs $D_* \le \varkappa S$: the total dimension S of the sources must exceed the dimension of the target by the factor $1/\varkappa$. A term of this size is present on average for maps in general position: for isometries B_{w} composed with independent random unitary maps of $\mathcal{H}$, the mean of $\operatorname{tr}(Q^2)$ is $S + (S^2 - \sum_{\mathsf{w}} d_{\mathsf{w}}^2)/D$ by Remark (1), while $\operatorname{tr} Q = S$; so the mean of $\operatorname{tr}(Q^2) D/(\operatorname{tr} Q)^2 - 1$ is about D/S. This is a statement about a mean. Lemma T1(c) is an upper bound, and (III.218)–(III.220) do not bound $\operatorname{tr}(Q^2) D/(\operatorname{tr} Q)^2 - 1$ from below. For example, if every B_{w} is a unitary map of $\mathcal{H}$, then (III.218)–(III.220) hold with $D_* = D$ and $\eta_0 = e_0 = \eta_1 = 0$, and $Q = |\mathsf{W}| \cdot 1$, so this number is 0, while $D_*/S = 1/|\mathsf{W}|$. (Another example, with $d_{\mathsf{w}} = 1$: the $p(p+1)$ lines spanned by the vectors of $p + 1$ mutually unbiased bases of $\mathbb{C}^p$, where such bases exist. Here $Q = (p + 1) \cdot 1$ and

$D_*/S = 1/(p+1)$.) (3) The archive contains statements with these names. `CorBounds` is for (b). `TraceSquareAndDiagBounds` is for (c), with $e_0 \le \frac{1}{4}$ where (c) has $\frac{1}{64}$. `CompressTraceSquareBound` is a variant of (c), not of (d): the maps are first restricted to subspaces of total codimension at most αS, and the conclusion is the second-moment bound for the restricted maps with $\eta = 5\varkappa + 3\sqrt{\alpha}$ (for $\varkappa \le \frac{1}{20}$, $0 \le \alpha \le \frac{1}{4}$). The bound for the diagonal blocks of the restricted maps, with $e = 2e_0$, is the conclusion of another statement, `CompressDiagDefectBound` (for $\alpha \le \frac{1}{2}$). The proofs above are ours. The Lean proofs of these statements were not read. By the dependency computation the final theorem depends on `CompressTraceSquareBound` and not on `CorBounds` or `TraceSquareAndDiagBounds`. The name `CompressDiagDefectBound` was not among those queried.

Corollary T2 (the hypotheses of Part II, §6.9 for the maps of the source, and for arbitrary subspaces). *Assume* (S) *with* (III.217), (I) *with* $\delta_{\mathrm{p}} < 1$, *and* (P), *for one record. Let* $C_H > 0$ *and put* $\varkappa = x^{-C_H}$. *Let* $\alpha_{\mathrm{ram}}, \alpha' \ge 0$, *and for each word let* $\mathcal{U}_{\mathsf{w}} \subset \mathcal{M}_{\mathsf{w}}$ *be a complex subspace. Assume further*

$$|\mathsf{W}| \;\ge\; C_0\,\mathsf{G}^2 x^{C_H}, \qquad d_{\mathsf{w}} - d_{\mathsf{w}}^{\mathrm{ram}} \le \alpha_{\mathrm{ram}}\, d_{\mathsf{w}} \;\; (\mathsf{w} \in \mathsf{W}), \qquad \varkappa \le \tfrac{1}{4}, \qquad \exp(-x^{\varepsilon/8}) \le \varkappa/2, \tag{III.233}$$

$$\sum_{\mathsf{w}} \left(d_{\mathsf{w}}^{\mathrm{ram}} - \dim \mathcal{U}_{\mathsf{w}}\right) \;\le\; \alpha' \sum_{\mathsf{w}} d_{\mathsf{w}}. \tag{III.234}$$

For each word let $\beta_{\mathsf{w}} : \mathcal{M}_{\mathsf{w}} \to \mathrm{Hom}_{\mathbb{L}}(\mathcal{E}_{\mathbb{L}}, \mathbb{L}) \otimes_{\mathbb{L}} \mathbb{C}$ *be the complexification of the map* $\varphi \mapsto \mathfrak{B}_{\mathsf{w}}^{\mathrm{alg}}(\varphi, \cdot)$ *on* $\mathcal{M}_{\mathsf{w},\mathbb{L}}$; *here* $\mathcal{M}_{\mathsf{w},\mathbb{L}} \otimes_{\mathbb{L}} \mathbb{C}$ *is identified with* $\mathcal{M}_{\mathsf{w}}$ *as in §1. The target is identified with the dual* $\mathcal{E}_0^\vee$ *of* $\mathcal{E}_0$ *by the map* $\mathcal{E}_{\mathbb{L}} \to \mathcal{E}_0$ *of §1. Then the sum of the spaces* $\beta_{\mathsf{w}}(\mathcal{U}_{\mathsf{w}})$ *has codimension at most*

$$\frac{5x^{-C_H} + 4(\alpha_{\mathrm{ram}} + \alpha') + 4\sqrt{e_0}}{1 - \delta_{\mathrm{p}}}\; D_{\mathrm{r}} \tag{III.235}$$

in $\mathcal{E}_0^\vee$, *which has dimension* D_{r} *by* (III.215). *In particular, for* $\mathcal{U}_{\mathsf{w}} = \mathcal{M}_{\mathsf{w}}$ *and* $\alpha' = 0$, *the sum of the images of the* β_{w} *has codimension at most* $(5x^{-C_H} + 4\alpha_{\mathrm{ram}} + 4\sqrt{e_0})D_{\mathrm{r}}/(1 - \delta_{\mathrm{p}})$.

Proof. First, $c_0(\mathsf{w}) > 0$ for every word, so that B_{w} is defined. Indeed $d_{\mathsf{w}} \ge 1$ by the lower bound in (III.217), and $e_0 < 1$ by (III.216); if $c_0(\mathsf{w}) = 0$, then $K_{\mathsf{ww}} = 0$ and (III.219) would read $d_{\mathsf{w}} \le e_0 d_{\mathsf{w}}$. (This is `unfolding_constant_pos_of_diagonal` in the source.) By the lower bound in (III.217), $d_{\mathsf{w}} \ge D_*/(C_0\mathsf{G}^2)$, so $S \ge |\mathsf{W}| D_*/(C_0\mathsf{G}^2) \ge D_*/\varkappa$. With (III.216) and the last condition of (III.233), (III.224) holds; and $D \ge 1$, because (III.219) with $e_0 < 1$ gives $B_{\mathsf{w}} \ne 0$. (If $\eta_0 < -1$, (III.218) forces $|\mathsf{W}| = 1$, which the first condition of (III.233) excludes.) By the second condition of (III.233) and by (III.234), $\sum_{\mathsf{w}}(d_{\mathsf{w}} - \dim \mathcal{U}_{\mathsf{w}}) \le (\alpha_{\mathrm{ram}} + \alpha')S$.

The estimate in $\mathcal{H}$. The space $\mathcal{E} = \mathsf{s}(\mathcal{E}_0)$ is a complex subspace of $\mathcal{H}$, because $c\,\mathsf{s}(F) = \mathsf{s}(\bar{c}F)$. Its dimension is $D_{\mathrm{r}} = (1 - \delta_{\mathrm{p}})D$, because s is injective and maps a basis of $\mathcal{E}_0$ to a basis of $\mathcal{E}$. Let $\pi_{\mathcal{E}}$ be the orthogonal projection of $\mathcal{H}$ onto $\mathcal{E}$. Apply Lemma T1(d) with the subspaces $\Psi_{\mathsf{w}}\mathcal{U}_{\mathsf{w}} \subset \mathcal{M}_{\mathsf{w}} \subset \mathcal{C}_{\mathsf{w}}$, which have the dimensions of the $\mathcal{U}_{\mathsf{w}}$, with $\alpha_{\mathrm{ram}} + \alpha'$ in place of α, and with $\delta_{\mathcal{E}} = \delta_{\mathrm{p}}$. Since P_{w} and B_{w} differ by a positive factor, the space $\mathcal{Y} = \sum_{\mathsf{w}} \pi_{\mathcal{E}} P_{\mathsf{w}} \Psi_{\mathsf{w}} \mathcal{U}_{\mathsf{w}}$ has codimension at most (III.235) in $\mathcal{E}$. This step uses only the subspace $\mathcal{E}$ of $\mathcal{H}$, and neither s nor the scalars of $\mathcal{E}_0$.

The passage to the dual. Let $\Lambda : \mathcal{H} \to \mathcal{E}_0^\vee$ send h to the form

$$\Lambda(h) : \; F \longmapsto \langle \mathsf{s}\Psi F,\, h\rangle_{\mathcal{H}} \qquad (F \in \mathcal{E}_0). \tag{III.236}$$

The form $\Lambda(h)$ is $\mathbb{C}$-linear in F, by the computation after (III.222); so it lies in $\mathcal{E}_0^\vee$. The map Λ is $\mathbb{C}$-linear in h. Its kernel is the set of h orthogonal to $\mathsf{s}\Psi(\mathcal{E}_0) = \mathsf{s}(\mathcal{E}_0) = \mathcal{E}$, that is, $\mathcal{E}^\perp$. So $\Lambda = \Lambda \circ \pi_{\mathcal{E}}$, the restriction of Λ to $\mathcal{E}$ is injective, and since $\dim \mathcal{E} = D_{\mathrm{r}} = \dim \mathcal{E}_0^\vee$ it maps $\mathcal{E}$ isomorphically onto $\mathcal{E}_0^\vee$. Hence Λ maps $\mathcal{H}$ onto $\mathcal{E}_0^\vee$, and $\Lambda(\mathcal{Y})$ has in $\mathcal{E}_0^\vee$ the codimension of $\mathcal{Y}$ in $\mathcal{E}$. By (III.222), the

map $\varphi \mapsto \mathfrak{B}_{\mathsf{w}}(\varphi,\cdot)$ from $\mathcal{M}_{\mathsf{w}}$ to $\mathcal{E}_0^\vee$ is $\lambda_{\mathsf{w}}\,\Lambda P_{\mathsf{w}}\Psi_{\mathsf{w}}$, a $\mathbb{C}$-linear map. The image of $\mathcal{U}_{\mathsf{w}}$ under it is $\Lambda\pi_{\mathcal{E}}P_{\mathsf{w}}\Psi_{\mathsf{w}}\mathcal{U}_{\mathsf{w}}$, because $\lambda_{\mathsf{w}}\neq 0$ and the image is a complex subspace. So the sum of these images is $\Lambda(\mathcal{Y})$.

The algebraic model. The map $\mathcal{E}_{\mathbb{L}}\to\mathcal{E}_0$ is linear over $\mathbb{L}$ and maps an $\mathbb{L}$-basis to a $\mathbb{C}$-basis. So restriction of forms along it is a $\mathbb{C}$-linear isomorphism of $\mathcal{E}_0^\vee$ with $\mathrm{Hom}_{\mathbb{L}}(\mathcal{E}_{\mathbb{L}},\mathbb{L})\otimes_{\mathbb{L}}\mathbb{C}$. Under it and under $\mathcal{M}_{\mathsf{w},\mathbb{L}}\otimes_{\mathbb{L}}\mathbb{C}=\mathcal{M}_{\mathsf{w}}$, the map $\varphi\mapsto\mathfrak{B}_{\mathsf{w}}(\varphi,\cdot)$ is β_{w}: both are $\mathbb{C}$-linear, and by (P1) they agree on the image of $\mathcal{M}_{\mathsf{w},\mathbb{L}}$, which spans $\mathcal{M}_{\mathsf{w}}$. □

On the two scalar actions. Every space in this proof is a complex space, and every map in it is $\mathbb{C}$-linear, except `s`. The conjugate-linear map `s` occurs only inside the inner product, in its conjugate-linear variable. The constants are those of Lemma T1(d).

The corollary is stated for arbitrary subspaces because Part II, §6.10 needs it so: there the rank bound is applied after discarding directions, for every level of the slope filtration. The subspaces that occur are complexifications of $\mathbb{L}$-subspaces of $\mathcal{M}_{\mathsf{w},\mathbb{L}}$.

The side conditions in the source. For the second condition of (III.233): each word has a level prime $\mathcal{T}(\mathsf{w})\geq b^{m/2}$, and $d_{\mathsf{w}}^{\mathrm{ram}}=d_{\mathsf{w}}(1-1/(\mathcal{T}(\mathsf{w})-1))^f$ (`RamRank_at`); by Bernoulli's inequality this gives $\alpha_{\mathrm{ram}}=2f/b$ for $b\geq 3$ (`ramShare_at_of`). The number $\mathcal{T}(\mathsf{w})$ is the level $\mathcal{T}$ of Part II, §6.7, which is said there to be attached to one choice of lattice data. In the source it depends on the translate and on the word (`T u0 u`). The first condition of (III.233) is `ChCode_at`. In the source the number of words is fixed at that size: with $C_0=3$,

$$|\mathsf{W}|=\lceil 3\,\mathsf{G}^2x^{C_H}\rceil,\qquad \text{so that, under (III.217)},\qquad x^{C_H}\;\leq\;\frac{S}{D_*}\;\leq\;9x^{C_H}+3 \tag{III.237}$$

(the equality is a clause in `DesignS.lean`, where the exponent is the one passed to `ChCode_at`; the two bounds are ours, from $D_*/(3\mathsf{G}^2)\leq d_{\mathsf{w}}\leq 3D_*/\mathsf{G}^2$). Since the words have $k-2$ letters, this requires $2^{k-2}\geq|\mathsf{W}|$. The source has $\log|\mathsf{W}|\leq(k-3)\log 2$ (a clause of `Win_at`) and the size condition $8C_{\mathsf{G}}\,m^5f\,x^{\varepsilon}\leq k$ (a clause of `SizeKRel`), with $m\approx x^{1/2-\varepsilon/16}$, $n\approx x^{2+8\varepsilon}$ and $k\approx x^{9/2+10\varepsilon}$, each up to a factor $x^{\pm\varepsilon/100}$ (`SizeM`, `SizeN`, `SizeK`). The rounded sizes $m\approx\sqrt{x}$, $n\approx x^2$, $k\approx x^{9/2}$ of §I.8 do not satisfy this condition: they give $m^5f\,x^{\varepsilon}=2x^{9/2+\varepsilon}$. By the dependency computation the final theorem depends on `RamRank_at`, `ramShare_at_of` and `ChCode_at`. The names `Win_at`, `SizeKRel`, `SizeM`, `SizeN` and `SizeK` were not among those queried.

How the source uses the corollary. The source does not state Corollary T2. It applies a theorem on lattices (`TheoremF`). By the names of the declarations, its proof uses the lemma of Part II, §6.9 in the form `FrameLA1`, which is stated for arbitrary subspaces and has η and not 2η (the chain of names from `frame_at_of` to `frame_linear_algebra_first` was followed; the proof terms were not read). The conclusion is the mean-slope inequality of Part II, §6.10, for each translate, with the error term

$$\frac{1}{1-\delta_{\mathrm{p}}}\Big(4u\log\big(1+\tfrac{H}{u}\big)+(\eta+4\sqrt{e})H\Big),\qquad \eta=5x^{-C_H}+3\sqrt{\alpha_{\mathrm{ram}}},\quad e=2e_0, \tag{III.238}$$

and with $-\log|\mathsf{W}|$ as a separate term of the lower bound, outside the bracket and not multiplied by $(1-\delta_{\mathrm{p}})^{-1}$ (`Frame_at`, `frame_at_of`). Here u and H are the spread and the height bound of Part II, §6.10, and α_{ram} is any number with $\alpha_{\mathrm{ram}}\leq\frac14$ for which the second condition of (III.233) holds (`de` in the Lean); the statement also requires $x^{-C_H}\leq\frac{1}{20}$. The values of η and e in (III.238) are the ones that the proof of `frame_at_of` supplies. The source then takes $\alpha_{\mathrm{ram}}=e^{-2x^{\varepsilon/8}}$, so that $\sqrt{\alpha_{\mathrm{ram}}}=\exp(-x^{\varepsilon/8})$, and uses $\sqrt{e_0}\leq\exp(-x^{\varepsilon/8})$ from (III.216) (`matRealF_of`). For all sufficiently large x, $(3+4\sqrt{2})\,\exp(-x^{\varepsilon/8})x^{C_H}\leq 1$. Since $\eta=5x^{-C_H}+3\sqrt{\alpha_{\mathrm{ram}}}$ and $e=2e_0$, this gives

$$(\eta+4\sqrt{e})\,x^{C_H}\leq 6,\qquad \text{so that}\qquad (\eta+4\sqrt{e})H\leq 6b\quad\text{for } H=x^{C_H}b \tag{III.239}$$

(`FrameF`, `frameF_of`). Indeed $(\eta+4\sqrt{e})x^{C_H}=5+(3\sqrt{\alpha_{\mathrm{ram}}}+4\sqrt{2e_0})x^{C_H}$, and the bracket is at most $(3+4\sqrt{2})\exp(-x^{\varepsilon/8})$. The weaker bound $\exp(-x^{\varepsilon/8})\le x^{-C_H}/2$, which is the last condition of (III.233), would not suffice: it leaves $(3+4\sqrt{2})/2>1$ for the terms beyond 5. In the Lean the hypothesis of `frameF_of` is $(3\sqrt{\mathtt{de}}+4\sqrt{\mathtt{eB}})\,x^{C_H}\le 1$ with $0\le \mathtt{de}\le\frac14$ and $0\le \mathtt{eB}\le\frac14$ (`FrameRealF`), where `de` is α_{ram} and `eB` is an upper bound for e; `matRealF_of` proves it for $x\ge 18+|a_0|+|a_1|$. Here a_0 is the threshold in its hypothesis (`MatSF`), and a_1 is a number such that $x^{C_H+1}\le\exp(x^{\varepsilon/8})$ for all $x\ge\max(a_1,1)$. For such x the proof deduces $\exp(-x^{\varepsilon/8})x^{C_H}\le\frac{1}{18}$, and it uses the cruder constant 11 in place of $3+4\sqrt{2}$. So the threshold is not 18; it depends on ε and C_H. The exponent C_H is the one in the bound $H\le x^{C_H}b$ of the source (a clause of `LinkDIV_at`).

What is known about the use of these declarations. `blockReal_at_of` is in the dependency lists as a declaration that the final theorem depends on. `frame_at_of` and `matRealF_of` were first reached by reading proof terms from `blockReal_at_of`, and a later dependency computation confirms that the final theorem depends on both. The chain of proof terms is this. `blockReal_at_of` is used only in the proof of `matKF_of`. `matKF_of` is used in four places. In each, its value is an argument of `frameRealF_of"`, and the value of `frameRealF_of"` is an argument of `frameF_of`. `frameRealF_of"` uses `matRealF_of` through `frameRealF_of'`, and `frameF_of` uses `frame_at_of`. The same computation shows that the final theorem depends on `LinkIRR`, `TheoremF`, `FrameLA1`, `frame_linear_algebra_first`, `Frame_at`, `FrameF`, `matSF_of`, `ChCode_at` and `CodeUpF`, and does not depend on `MeanBoundSecondForm`. The statement `FrameLA` and its proof `frame_linear_algebra_bound` are in the lists as not used by the final theorem, as Part II, §6.9 notes. The final theorem also depends on `matKF_of`, on `unfolding_constant_pos_of_diagonal` and on `TheoremFSharp`.

The error term and the two means. Part II, §6.9–6.10 states these points; here they are given with the names of the source. *The error.* For each translate, the mean-slope inequality of the source has the form

$$\text{(mean slope of the target)}\ \ge\ s-\log|\mathsf{W}|-\frac{1}{1-\delta_{\mathrm{p}}}\Big(4u\log\big(1+\tfrac{H}{u}\big)+(\eta+4\sqrt{e})H\Big),\qquad\text{(III.240)}$$

with $\eta=5x^{-C_H}+3\sqrt{\alpha_{\mathrm{ram}}}$ and $e=2e_0$ as in (III.238). Here s is the mean over the words, weighted by the ranks, of the slopes of the sources minus their heights (`sFrame`), and the target is the dual lattice (`muDual`); this is the statement `Frame_at`. The factor $(1-\delta_{\mathrm{p}})^{-1}$ multiplies the error bracket only. It is there because the lattice of the target spans a space of dimension D_{r}, not D, and the second moment is taken in $\mathcal{H}$; this is where (I) is used, as in Lemma T1(d). The term $-\log|\mathsf{W}|$ is separate. It is at most $(k-3)\log 2$ in absolute value (`Win_at`), which is negligible against b. The archive has a statement with the same hypotheses and no such term (`TheoremFSharp`). In η only $\sqrt{\alpha_{\mathrm{ram}}}$ is exponentially small. The power of x comes from the number of words (III.237): by Lemma T1(c) a bound $5\varkappa$ for η needs $D_*\le\varkappa S$, and $D_*/S\ge 1/(9x^{C_H}+3)$. So no bound for η better than a negative power of x follows from (S) by Lemma T1. This is enough, by (III.239). We do not assert that the number $\mathrm{tr}(Q^2)D/(\mathrm{tr}\,Q)^2-1$ of the family is of that size: it was not bounded from below (Remark (2)). *The two means.* Each translate of Part II, §4.2 has its own record, with its own set of words (§1). The mean over the words is an inner mean, weighted by the ranks, inside (III.240) for one translate: the number s is, in the file `MeanBoundSecondForm.lean` (source [7] of Part II),

```
def s (u0 :  U0) :  ℝ := wmean (D.d u0) (fun u => D.muN u0 u - D.h u0 u).
```

That file defines the record and its hypotheses, among them the slope inequality (III.240) for each translate (`Data.SlopeIneq`; the final theorem depends on it). It also states a bound for the mean

over the translates, with an error `ErrX` that contains the mean of $\log|\mathsf{W}|$ (`MeanBoundSecondForm`). The final theorem does not depend on that statement. The route of the final theorem keeps the translate fixed: `each_unit_bound_statement`, in the file `LevelOfCongruence.lean` (source [28] of Part II), has the conclusion

$$\texttt{D.ASK u0} \le \texttt{D.Main u0 + ErrX5at D C Xpp u0} \quad \text{for every } \texttt{u0},$$

and $\log|\mathsf{W}|$, for the set of words of the translate `u0`, is a term of `ErrX5at`. The final theorem depends on `each_unit_bound_statement`; its two uses in the archive, in the files `ChTop.lean` (source [5]) and `PChain.lean`, are again for one translate. So the bound of the source holds for each translate, and the mean over the translates in (M) is an outer mean, taken afterwards. The inner mean is not visible in (M). The "two moments" of Part II, §6.9–6.10 are what the lemma there needs. What the source assumes are the bounds (III.218)–(III.220) for the Gram blocks; the two moments follow from them by Lemma T1(b),(c).

3. The mechanism in a finite model

Statement (S) is proved in the source by expanding each block $K_{\mathsf{ww'}}$ as a sum of terms and computing its squared norm as a sum over pairs of terms (see §4). The simplest instance of this kind of computation is the following, for theta lifts between two definite groups as in Part II, §6.1–6.2. It is a statement about finite sets.

The setting. Let $\mathcal{I}$ and $\mathcal{I}'$ be nonempty finite sets with positive weights ω and ω' (classes of lattices L_i and $L'_{i'}$, with $\omega(i) = 1/|\mathrm{Aut}\, L_i|$). For $i \in \mathcal{I}$ let G_i be a finite group of order $1/\omega(i)$. For $i \in \mathcal{I}$, $i' \in \mathcal{I}'$ let $\mathrm{Rep}(i,i')$ be a finite set on which G_i acts freely (the isometric embeddings $\rho : L_i \to L'_{i'}$, with $\sigma \in \mathrm{Aut}\, L_i$ acting by $\rho \mapsto \rho\sigma$). Let A be a finite abelian group and $\mathrm{inv} : \mathrm{Rep}(i,i') \to A$ maps with $\mathrm{inv}(\rho\sigma) = \mathrm{inv}(\rho)$ for $\sigma \in G_i$. For a character χ of A put

$$\Theta_\chi(i,i') = \sum_{\rho \in \mathrm{Rep}(i,i')} \chi(\mathrm{inv}(\rho)), \qquad (B_\chi f)(i') = \sum_{i \in \mathcal{I}} \omega(i)\, \Theta_\chi(i,i')\, f(i). \tag{III.241}$$

So $\Theta_1(i,i')$ is the number of representations of L_i by $L'_{i'}$, and Θ_χ is a representation number twisted by a character. The map B_χ goes from $L^2(\mathcal{I},\omega)$ to $L^2(\mathcal{I}',\omega')$, where $\langle f,g\rangle = \sum_i \omega(i)\overline{f(i)}g(i)$, conjugate-linear in the first variable as everywhere in this section. An operator Y on $L^2(\mathcal{I},\omega)$ has a kernel, also written Y, with $(Yf)(i_1) = \sum_{i_2} \omega(i_2) Y(i_1,i_2) f(i_2)$; the kernel of the identity is $1/\omega(i_1)$ for $i_1 = i_2$ and 0 otherwise, and $\|Y\|_{\mathrm{HS}}^2 = \sum_{i_1,i_2} \omega(i_1)\omega(i_2)|Y(i_1,i_2)|^2$. For $i_1, i_2 \in \mathcal{I}$ put

$$\mathcal{N}(i_1,i_2) = \sum_{i' \in \mathcal{I}'} \omega'(i')\, \#\big\{(\rho_1,\rho_2) \in \mathrm{Rep}(i_1,i') \times \mathrm{Rep}(i_2,i') :\ \mathrm{inv}(\rho_1) = \mathrm{inv}(\rho_2)\big\}, \tag{III.242}$$

and let $\mathcal{N}^\circ(i_1,i_2)$ be the same sum restricted to the pairs that are *not* of the form $(\rho,\rho\sigma)$ with $i_1 = i_2$ and $\sigma \in G_{i_1}$.

Lemma T3. (a) *The operator $\sum_\chi B_\chi^\dagger B_\chi$ on $L^2(\mathcal{I},\omega)$, the sum being over all characters of A, has the kernel $|A|\,\mathcal{N}$.*

(b) *Assume that there is a number $\mathfrak{m}$ with*

$$\sum_{i' \in \mathcal{I}'} \omega'(i')\, |\mathrm{Rep}(i,i')| = \mathfrak{m} \qquad \text{for every } i \in \mathcal{I}. \tag{III.243}$$

Then $\sum_\chi B_\chi^\dagger B_\chi = |A|\,(\mathfrak{m}\cdot 1 + \mathfrak{N}^\circ)$, where $\mathfrak{N}^\circ$ is the operator with kernel $\mathcal{N}^\circ \ge 0$, and

$$\Big\| \sum_\chi B_\chi^\dagger B_\chi - |A|\,\mathfrak{m}\cdot 1 \Big\|_{\mathrm{HS}}^2 = |A|^2 \sum_{i_1,i_2 \in \mathcal{I}} \omega(i_1)\omega(i_2)\, \mathcal{N}^\circ(i_1,i_2)^2. \tag{III.244}$$

(c) *Under* (III.243) *and* $\mathfrak{m} > 0$*, the family of the adjoint maps* $B_\chi^\dagger : L^2(\mathcal{I}', \omega') \to L^2(\mathcal{I}, \omega)$ *satisfies the second-moment hypothesis of Part II, §6.9: with* $Q = \sum_\chi B_\chi^\dagger B_\chi$*,*

$$\operatorname{tr}(Q^2) \le (1+\eta)\frac{(\operatorname{tr} Q)^2}{|\mathcal{I}|}, \qquad \eta = \frac{1}{\mathfrak{m}^2|\mathcal{I}|}\sum_{i_1,i_2}\omega(i_1)\omega(i_2)\,\mathcal{N}^\circ(i_1,i_2)^2. \tag{III.245}$$

Proof. (a) The adjoint of B_χ is $(B_\chi^\dagger g)(i) = \sum_{i'}\omega'(i')\overline{\Theta_\chi(i,i')}g(i')$, as one sees by writing out $\langle B_\chi f, g\rangle = \langle f, B_\chi^\dagger g\rangle$. So $B_\chi^\dagger B_\chi$ has the kernel $\sum_{i'}\omega'(i')\overline{\Theta_\chi(i_1,i')}\Theta_\chi(i_2,i')$. Insert (III.241) and sum over χ: by the orthogonality of the characters of A, $\sum_\chi \overline{\chi(\operatorname{inv}(\rho_1))}\chi(\operatorname{inv}(\rho_2))$ is $|A|$ if $\operatorname{inv}(\rho_1) = \operatorname{inv}(\rho_2)$ and 0 otherwise. This gives $|A|\,\mathcal{N}(i_1,i_2)$.

(b) Let $i_1 = i_2 = i$. The pairs $(\rho, \rho\sigma)$ with $\rho \in \operatorname{Rep}(i,i')$ and $\sigma \in G_i$ are distinct for distinct (ρ,σ), because the action is free; they satisfy $\operatorname{inv}(\rho) = \operatorname{inv}(\rho\sigma)$; and their number is $|\operatorname{Rep}(i,i')|/\omega(i)$. By (III.243) their contribution to $\mathcal{N}(i,i)$ is $\mathfrak{m}/\omega(i)$, which is the kernel of $\mathfrak{m}\cdot 1$. The rest is $\mathcal{N}^\circ$, by definition, and $\mathcal{N}^\circ \ge 0$ because it is a weighted count. (III.244) is the formula for the Hilbert–Schmidt norm of an operator with a kernel.

(c) For a Hermitian operator Q on a space of dimension ν and any real y, $\operatorname{tr}(Q^2) - (\operatorname{tr} Q)^2/\nu = \|Q - \frac{\operatorname{tr} Q}{\nu}\|_{\mathrm{HS}}^2 \le \|Q - y\|_{\mathrm{HS}}^2$. Take $\nu = |\mathcal{I}|$ and $y = |A|\,\mathfrak{m}$ and use (III.244). Also $\operatorname{tr} Q = |A|(\mathfrak{m}|\mathcal{I}| + \operatorname{tr}\mathfrak{N}^\circ) \ge |A|\,\mathfrak{m}|\mathcal{I}|$, because $\operatorname{tr}\mathfrak{N}^\circ = \sum_i \omega(i)\mathcal{N}^\circ(i,i) \ge 0$. Divide. □

What the lemma shows. The sum over the twists is a count of pairs of representations subject to one condition, $\operatorname{inv}(\rho_1) = \operatorname{inv}(\rho_2)$: this is the form a relative trace formula takes when both groups are definite. The pairs that are forced, $(\rho, \rho\sigma)$, give a multiple of the identity, and the multiple is the mean number of representations (III.243). The other pairs are coincidences, and η is their normalised mean square. Part (b) uses three hypotheses, and each is needed. The first is that the action of G_i is free: with stabilisers of order 2 the identity of (b) fails. The second is the invariance $\operatorname{inv}(\rho\sigma) = \operatorname{inv}(\rho)$: without it the forced pairs do not all satisfy the condition, and the identity of (b) and (III.244) fail. (The number $\mathcal{N}^\circ$ is a restricted count and stays nonnegative; it is the kernel $\mathcal{N} - \mathfrak{m}/\omega$ on the diagonal that can become negative.) The third is (III.243), which is not formal. For lattices it is Siegel's theorem: if $\mathcal{I}'$ is a full genus and the L_i belong to one genus, the left side of (III.243) is the mass of the genus $\mathcal{I}'$ times a product of local densities, which depend only on the genus of L_i (C. L. Siegel, *Über die analytische Theorie der quadratischen Formen*, Ann. of Math. 36 (1935); in the adelic form, A. Weil, *Sur la formule de Siegel dans la théorie des groupes classiques*, Acta Math. 113 (1965); both cited from memory).

An example. Let L_1, L_2 be the binary lattices with Gram matrices $\left(\begin{smallmatrix}2 & 1\\ 1 & 12\end{smallmatrix}\right)$ and $\left(\begin{smallmatrix}4 & 1\\ 1 & 6\end{smallmatrix}\right)$ (twice the forms $X_1^2 + X_1X_2 + 6X_2^2$ and $2X_1^2 + X_1X_2 + 3X_2^2$ of discriminant -23, which form one genus and represent the two classes under $\mathrm{GL}_2(\mathbb{Z})$), with automorphism groups of orders 4 and 2, and let $\mathcal{I}'$ consist of the lattice $\mathbb{Z}^5$, which is alone in its genus, with $\omega' = 1/(2^5\,5!)$. A count gives $|\operatorname{Rep}(L_1,\mathbb{Z}^5)| = |\operatorname{Rep}(L_2,\mathbb{Z}^5)| = 1920$, so (III.243) holds with $\mathfrak{m} = \frac12$; for $\mathbb{Z}^6$ both numbers are 19200. (For two forms of determinant 60 in different genera, $\left(\begin{smallmatrix}2 & 0\\ 0 & 30\end{smallmatrix}\right)$ and $\left(\begin{smallmatrix}6 & 0\\ 0 & 10\end{smallmatrix}\right)$, the numbers for $\mathbb{Z}^5$ are 4480 and 13440.) Let $\varsigma_0, \dots, \varsigma_9$ be the ten 2×2 minors of the pair of vectors ρ, in lexicographic order. Take for $\operatorname{inv}(\rho)$ the 55 products $\varsigma_\iota\varsigma_{\iota'}$ modulo a prime ℓ_0; they are invariant under G_i, and $A = (\mathbb{Z}/\ell_0)^{55}$. For $\ell_0 = 3$ one finds $\mathcal{N}^\circ = \left(\begin{smallmatrix}2 & 0\\ 0 & 3\end{smallmatrix}\right)$. Then the true value $\operatorname{tr}(Q^2)|\mathcal{I}|/(\operatorname{tr} Q)^2 - 1$ is $\frac19$, and the bound (III.245) gives $\eta = 5$. For $\ell_0 = 5$ and $\ell_0 = 11$ there are no coincidences, $\mathcal{N}^\circ = 0$, and $\sum_\chi B_\chi^\dagger B_\chi = |A|\,\mathfrak{m}$ exactly. With the single invariant $\operatorname{inv}(\rho) = \varsigma_0^2 + 2\varsigma_9^2 + 3\varsigma_1\varsigma_4$ modulo 5 (so $|A| = 5$) the coincidences dominate: $\mathcal{N}^\circ \approx \left(\begin{smallmatrix}242 & 173\\ 173 & 209\end{smallmatrix}\right)$. So the family of twists must be large compared with the number of representations for the count $\mathcal{N}^\circ$ to be small. The identities (III.244), (III.245) were also checked numerically on random finite models.

4. The statements behind (S), (I) and (P) in the source

These are assumed here. For each we give the statement that was read, its Lean name, the kind of theorem it is, and the parameter that produces the saving. References to the literature are from memory.

(R1) The expansion of a block and its four parts (`ClKppOff`, `ClKppDiag`, `ClOff`, `ClDiag`; they give (S) by `linkS_concl`). Each block is an absolutely convergent sum $K_{\mathsf{w}\mathsf{w}'} = \sum_\tau K^{(\tau)}_{\mathsf{w}\mathsf{w}'}$ over a set of *terms* τ (`Term`, `aTerm`). A term has an invariant in the totally real field (`Gterm`) and may be the identity term, or *generic*. Write $K''_{\mathsf{w}\mathsf{w}'}$ for the sum of the terms with invariant 0 other than the identity term. The four statements are: $K''_{\mathsf{w}\mathsf{w}'} = 0$ for $\mathsf{w} \neq \mathsf{w}'$; $\|K''_{\mathsf{w}\mathsf{w}}\|^2_{\mathrm{HS}} \le d_{\mathsf{w}} \exp(-kf \log k)$;

$$\|K_{\mathsf{w}\mathsf{w}'} - K''_{\mathsf{w}\mathsf{w}'}\|^2_{\mathrm{HS}} \le (1+b^{-6})\frac{d_{\mathsf{w}} d_{\mathsf{w}'}}{D_*} \quad (\mathsf{w} \neq \mathsf{w}'), \qquad \|K_{\mathsf{w}\mathsf{w}} - K''_{\mathsf{w}\mathsf{w}} - 1\|^2_{\mathrm{HS}} \le \mathsf{G}\,\frac{d_{\mathsf{w}}^2}{D_*}. \tag{III.246}$$

With the upper bound in (III.217), $\mathsf{G}\, d_{\mathsf{w}}^2/D_* \le (C_0/\mathsf{G})\, d_{\mathsf{w}}$: this is (III.219), and it shows where e_0 comes from. *The diagonal bound is small because each $\mathcal{C}_{\mathsf{w}}$ is a small part of $\mathcal{H}$*, of relative dimension about G^{-2}, while the constant in the second bound of (III.246) is only G. This is an expansion of the kind of Petersson's formula for the inner products of Poincaré series, here for the Bessel periods of a unitary group in $2(\frac{m}{2}-1)+k+1$ variables along a unitary group in k variables; this number is $k+m-1$, the r of §III.4 (the periods of W. T. Gan, B. Gross and D. Prasad, Astérisque 346 (2012)); it is a relative trace formula in the sense of Jacquet. In the source the squared norm of a part of the sum is expressed as a sum, over *pairs* of terms, of inner products of vectors attached to the terms (`pairSum`; compare (III.242)), and the pairs are divided by a rank $s_0 \ge 1$ and by being "scalar" or not. The pairs with $s_0 = 1$ that are scalar and generic give the main term. The statements read are bounds for the classes of pairs: `RowNGP` (non-generic terms: at most $2^{-(k-m-3)} \exp(Cm^5 f \log b)\, d_{\mathsf{w}} d_{\mathsf{w}'}/D_*$) and `RowMainEqP` (the main pairs of a diagonal block: at most $\exp(Cm^5 f \log b)\, d_{\mathsf{w}}^2/D_*$). Both are deduced in the source from (R2). The saving in $\eta_0 = b^{-6}$ comes from k and from the level primes: the non-generic and the higher-rank pairs carry factors 2^{-k} and $\exp(-ckf \log k)$ with $k \ge C' m^5 f x^{\varepsilon}$, and the factor $\prod_v (1 + 3/|\mathbb{F}_v|)$ over the places above g_0, g_1 is $1 + O(b^{-7})$ when $g_0, g_1 \ge (kb)^8$ (`NumPrice`). How these are combined in the proof of the first bound of (III.246) was not read.

(R2) The Siegel–Weil formula in the convergent range (`SiegelWeilUnitaryConvergent`; the theorem `siegelWeilUnitaryConvergent` is among the declarations the final theorem depends on). *Let E/E^+ be a CM extension, J a nondegenerate Hermitian matrix of size M over E, and ν an integer with $2\nu < M$. The quotient of the adelic points of the unitary group of J by its rational points has finite volume, and for every nondegenerate Hermitian Z of size ν and every factorisable Schwartz function $\Phi = \Phi_\infty \otimes \bigotimes_v \Phi_v$, the mean over this quotient of $\sum_X \Phi(gX)$, the sum being over the ν-tuples X of rational vectors with Gram matrix Z, equals $\varrho_\infty(Z) \prod_v \varrho_v(Z)$, where ϱ_v is the density at Z of the image of the measure $\Phi_v\, dX$ under the Gram map.* (The description of the theta sum is our reading of the names; its definition was not read.) This is Siegel's theorem in Weil's form, for a unitary group, in a range of absolute convergence; for unitary groups beyond that range see A. Ichino, *A regularized Siegel–Weil formula for unitary groups*, Math. Z. 247 (2004), and S. Yamana, *On the Siegel–Weil formula: the case of singular forms*, Compos. Math. 147 (2011). It plays the role of (III.243): it evaluates the mean number of representations, and so the main pairs, by local densities. It contains no saving by itself.

(R3) The count of $\mathcal{H}$ (`CountAllPointsClass`; it gives (III.220)). *$|D - D_*| \le \exp(-crf \log r)\, D_*$, where D is the multiplicity of a discrete series representation, at a fixed level, in the automorphic spectrum of the unitary group in r variables, and D_* is the product of*

its formal degree, the volume of the quotient and the value at 1 *of the idempotent defining the level,* provided the Harish-Chandra parameter has all gaps at least $Cr^{19}f^5\log(b\,\mathrm{rd}_E)$, and under two conditions on the sizes of f and b. (The description of D and D_* is our reading of the definitions `Dcount` and `Dstar`.) This is a limit multiplicity statement with a rate (in the tradition of D. L. de George and N. R. Wallach, Ann. of Math. 107 (1978), and L. Clozel, Invent. Math. 83 (1986)); D_* is the identity term of the trace formula. The saving comes from the gaps of the weights. The volume in D_* is a mass. The mass formula of the source is `siegel_mass_ratio`: the quotient of the masses in ranks M and $M-1$ is $((M-1)!\,(v/\pi)^M)^f(2v)^{-f/2}$ times a further factor (here v is a positive real number, not a place, and π is the number), which is the recurrence displayed before (6.4) in Part II, §6.6 (H. Braun, Abh. Math. Sem. Hamburg 14 (1941); W. T. Gan, J. Hanke and J.-K. Yu, Duke Math. J. 107 (2001)). Here we read the source's `vInv`, `fDeg`, `lValEta` as v, f and that factor; their definitions, and the place where the formula is applied to D_*, were not read.

(R4) The count behind (I) (`closed_linkIRR_of_join_count`, `delp_of_allRows`). The space $\mathcal{H}$ has a basis each of whose D elements carries, at each of the places above g_0 and above g_1, a *label*: a monic polynomial of degree r over $\mathbb{F}_{g_\iota}$ ($\iota=0,1$) with nonzero constant term. A tuple of labels at the places above one g_ι is *good* if the polynomials are squarefree, their roots satisfy two conditions of non-coincidence, for every j with $1\le j\le r/2$ some place has a label with no factor of degree j, and some place has a label with an irreducible factor of prime degree j' with $r/2<j'\le r-3$ (`IsGood`). D_{r} is the number of basis elements whose tuples are good at both primes. The proof of `delp_of_allRows` contains the step

$$\delta_{\mathrm{p}}\le\frac{1+\varepsilon_C}{1-\varepsilon_C}\big(\mathrm{Wg}(g_0)+\mathrm{Wg}(g_1)\big),\qquad \varepsilon_C=\exp(-c\,r f\log r),\tag{III.247}$$

$$\mathrm{Wg}(g)=\frac{3f}{g}+\frac{23f\log g}{\sqrt{g}}+2fm\Big(\frac{1}{4\min(q,d)}+\frac{1+\log r}{g-1}\Big)+\frac{r}{2}\Big(\frac{7}{8}\Big)^f+\exp\Big(-\frac{f}{5\log r}\Big),\tag{III.248}$$

where q,d are the auxiliary primes of §III.1. The statement `delp_of_allRows` has as a hypothesis that the right side of (III.247) is at most a number of the source (`E7`), and as its conclusion $0\le\delta_{\mathrm{p}}\le\exp(-x^{\varepsilon_z/7})$; (III.247) itself is not a statement of the source. Two inputs of that proof: the count (R3), applied to the part of $\mathcal{H}$ with prescribed labels, whose expected dimension is D_* times a product over the places of local weights; and the fact that the local weight of a squarefree label is $g^{-r}(1+O(r/g))$, the labels that are not squarefree having total weight at most $3/g$ (`admLq`, `lawAt`), so that the labels at each place are distributed, up to these errors, like uniformly random monic polynomials of degree r with nonzero constant term. The latter rests on the parametrisation of the irreducible characters of $\mathrm{GL}_r(\mathbb{F}_g)$ by J. A. Green, *The characters of the finite general linear groups*, Trans. Amer. Math. Soc. 80 (1955), which the source takes as a hypothesis `Green` (the cuspidal characters of $\mathrm{GL}_j(\mathbb{F}_g)$ correspond to the monic irreducible polynomials of degree j with nonzero constant term and have degree $\prod_{i=1}^{j-1}(g^i-1)$; for $j=1$ these are the $g-1$ polynomials $X-a$ with $a\neq0$, and the polynomial X is excluded; Lean: `IsPrimeLab`). The terms of $\mathrm{Wg}(g)$ are then statements about the factorisation of uniformly random monic polynomials with nonzero constant term, taken at f independent places. The bound (III.221) follows from (III.247) and the sizes $g_0,g_1\ge\exp(8x^{\varepsilon_z})$, $f\ge x^2$, $f,m,r\le x^C$ and $\min(q,d)\ge\exp(x^{\varepsilon_z/6})$ (`SizesAt`). *The term that limits the bound is* $fm/(2\min(q,d))$; the others are much smaller. The good tuples are those for which the source can prove that an attached residual Galois representation is irreducible; this use was not read. By the dependency computation the final theorem depends on `delp_of_allRows` and on `Green`.

(R5) Uniqueness of Bessel models, and decay (`besselMapUniqGap`, `wuDecayOfGap`, `adm13_of_kernel_join`). *For a discrete series representation of a real unitary group in* r *variables*

whose Harish-Chandra parameter has all gaps at least $120r^3$*, a representation of the unitary group in* k *variables of a given type, and a generic character, any two Bessel maps on the smooth vectors are proportional.* (The source states this for $k+2j+1$ variables with any $j \geq 0$; here $j = \frac{m}{2} - 1$, which gives r variables.) Compare Theorem A of D. Jiang, B. Sun and C.-B. Zhu, *Uniqueness of Bessel models: the archimedean case*, Geom. Funct. Anal. 20 (2010): it gives multiplicity at most one for Bessel models at an archimedean place. The statement of the source has in addition a hypothesis on the gaps. Theorem A does not give the quantitative decay used here. The decay is a further statement of the source: from the uniqueness it deduces an integrability statement for a function attached to a word (`WuDecay`; the name refers to this function), under the condition that $r \geq 1801$ and that consecutive weights θ_j of §5 differ by at least $16r^{19}$ (`GapHyp`); by our reading this serves the convergence of the expansion in (R1). Statement (P2) is proved by `adm13_of_kernel_join`, from hypotheses that were not read.

Where the exponents come from.

number	value in the source	large parameter
η_0	b^{-6}	k, and $g_0, g_1 \geq (kb)^8$
e_0	$2C_0/\mathsf{G} + 2\exp(-kf\log k)$	$\mathsf{G} = \exp(C_{\mathsf{G}} m^5 f x^{\varepsilon})$: the ratio $D_*/d_{\mathsf{w}} \approx \mathsf{G}^2$
η_1	$\exp(-c\,r f \log r)$	the rank $r \geq k$ and the degree f
α_{ram}	$2f/b$	the level primes $\mathcal{T}(\mathsf{w}) \geq b^{m/2}$
δ_{p}	$fm/(2\min(q,d))$: leading term in the upper bound for δ_{p}	the auxiliary primes $q, d \geq \exp(x^{\varepsilon_z/6})$
bound for η	$5x^{-C_H} + 3\sqrt{\alpha_{\mathrm{ram}}}$	the number of words, $\lvert\mathsf{W}\rvert = \lceil 3\mathsf{G}^2 x^{C_H} \rceil$

The exponent $x^{\varepsilon/8}$ in (III.216) is not the true size of any of these numbers. It is chosen in the source so that one bound, $\exp(-y)$ with $y = x^{\varepsilon/8}$ and $x^{\varepsilon} = y^8$, covers η_0, η_1 and $\sqrt{e_0}$ (`eF_nonneg_and_le_exp_neg`). The exponent $\varepsilon_z/7$ in (III.221) is just below the $\varepsilon_z/6$ of the auxiliary primes. The last row is the bound for the constant η of Part II, §6.9 proved in the source (`frame_at_of`); see the paragraph on the error term at the end of §2 for what it does and does not say about η.

5. The interlacing of the weights

At each weight place the source fixes rational numbers

$$\xi_0 > \psi_0 > \xi_1 > \psi_1 > \cdots > \psi_{k-2} > \xi_{k-1} > -\theta_0, \qquad \theta_0 < e'_0 < \theta_1 < e'_1 < \cdots < \theta_{m-1} < e'_{m-1}, \tag{III.249}$$

with $\xi_i \in \mathbb{Z} + \frac{1}{2}$ and $\psi_i, \theta_j, e'_j \in \mathbb{Z}$ (`HCParam`; the condition is named `Interlace`; we write $\xi_i, \psi_i, \theta_j, e'_j$ for the numbers called `u`, `v`, `b`, `e` there; e'_0 is not the e_0 of (S)). The first chain is the condition under which the representation of the unitary group in $k-1$ variables with parameter ψ occurs in the restriction of the one in k variables with parameter ξ: the classical branching rule, here for parameters shifted by half the sum of the positive roots. In this section it is used in one place, the definition of the family: *the words are the weights* ξ *that interlace one fixed* ψ. Precisely, consecutive ψ_l differ by 4Δ for a fixed Δ, the numbers ξ_0 and ξ_{k-1} are the same for all words, and for $0 \leq l \leq k-3$

$$\xi_{l+1} = \psi_l - \tfrac{1}{2} - \Delta \cdot (\text{the } l\text{th letter of the word}), \tag{III.250}$$

and for both values of the letter this number lies between ψ_{l+1} and ψ_l, since (III.249) is part of the data for every word (`cloud_gap`, `u_letter`, `hcW_v`, `hcW_first`, `hcW_last`). So two words differ in

the weight ξ at the positions where their letters differ. That the bounds (III.246) for $\mathsf{w} \neq \mathsf{w}'$ use this (through the archimedean factors of the terms) is our reading of the definitions; the proofs were not read. The second chain of (III.249), with gaps at least $(t_0+3)/2$ (`gap_be`, `gap_eb`), concerns the weights along the chain of lifts and belongs to the divisibility and size estimates (§III.4, §III.5).

6. Assumed here, from the source

(i) *Statement* (S): `LinkS` with $C_0 = 3$, proved by `link_alt_alt_bis` from `theoremS_transport_link_zero`, for families with `TuSize`. Its inputs, as hypotheses of `link_alt_alt_bis`: the decay `WuDecayOfGap`; a finiteness statement `RowsFinDim`; the bounds `TypeP`, `RowNGP`, `RowMainEqP`, `LoopRowS`; the off-diagonal bound `TheoremSSelected`; the count `CountAllPointsClass`; and compatibility conditions. Only the statements were read, and for `RowNGP`, `RowMainEqP` the deduction from (R2) by name.

(ii) *Statement* (I): `LinkIRR`, proved by `closed_linkIRR_of_join_count` from `Green` and a count hypothesis (`CountAllRows`).

(iii) *Statement* (P): `BesselId`, the conjunction of `BuRational` and `adm13`; the latter by `adm13_of_kernel_join`.

(iv) *The side conditions of Corollary T2*: the proviso (III.217) (`ChProviso_at`) and the number of words (`ChCode_at`), which are conditions on the family; the formula for $d_{\mathsf{w}}^{\mathrm{ram}}$ (`RamRank_at`) and $\mathcal{T}(\mathsf{w}) \geq b^{m/2}$ (`TuGe`). That $c_0(\mathsf{w}) > 0$ is not assumed: it follows from (S), see the proof of Corollary T2.

(v) *The identification of the data of §1* with spaces of automorphic forms, with Poincaré series and with the pairing of the chain of lifts; and the passage from Corollary T2 to the mean-slope inequality (`TheoremF`, `frame_at_of`), which uses the slope filtration of Part II, §6.10. For what is known about the dependence of the final theorem on these two declarations see the end of §2.

(vi) (R1)–(R5) of §4.

What this section proves, and what it assumes. This section deduces the dimension bound from the three trace, rank and pairing statements (S), (I), (P) and the size and subspace conditions (III.233), (III.234) in T2. *Proved outright:* Lemma T1, including the polynomial inequality (III.230); Corollary T2, for arbitrary subspaces, as that deduction; Lemma T3; and the arithmetic of the example of §3 (by computation). *Quoted from the source without proof:* (S), (I), (P) and everything in §4 and §6. So the construction of the spaces and maps, the arithmetic inputs and the assembly of (M) are quoted. Lemma T3 is a model. It is not a step of the source's proof: there the family is the set of words, the terms are attached to a Bessel subgroup, and the larger group is, by our reading, not definite (the parameter `sigY0` of the source is the signature $(k+\frac{m}{2}-1, \frac{m}{2})$; it is the signature of the space `Y 0` of the record at every embedding, by the field `baseSpace_signature_eq`). The statements `FrameLA`, `frame_linear_algebra_bound`, `ThmSFree` and `KernelFree` of the archive are not used by the final theorem; the declarations `LinkS`, `LinkS_at`, `BesselId`, `blockReal_at_of`, `link_alt_alt_bis`, `theoremS_transport_link_zero`, `closed_linkIRR_of_join_count`, `rowNGP_of_siegelWeil`, `rowMainEqP_of_siegelWeil`, `siegelWeilUnitaryConvergent`, `adm13_of_kernel_join`, `wuDecayOfGap`, `besselMapUniqGap` and `siegel_mass_ratio` are. So are, by a later run of the same computation, `LinkIRR`, `TheoremF`, `FrameLA1`, `frame_linear_algebra_first`, `Frame_at`, `FrameF`, `frameF_of`, `frame_at_of`, `matRealF_of`, `matSF_of`, `ChCode_at`, `CodeUpF`, `RamRank_at`, `ramShare_at_of`, `delp_of_allRows`,

`Green`, `matKF_of`, `unfolding_constant_pos_of_diagonal`, `TheoremFSharp`, `RRV_finrank`, `jEE_basis` (the field of `D7`) and `CompressTraceSquareBound`. The final theorem does not depend on `CorBounds`, `TraceSquareAndDiagBounds` or `MeanBoundSecondForm`. Several declarations of the archive are named `offDiag`. The one that concerns this section is `GlobalTheory.CarrierGap.offDiag`, which proves the off-diagonal bound `TheoremSSelected` of (i) from the Siegel–Weil statements and `LoopRowS`. The final theorem depends on it, and also on `each_unit_bound_statement` and on `Data.SlopeIneq` (§2).

What remains.

(i) The definition of $\mathcal{H}$, $\mathcal{C}_{\mathsf{w}}$ and P_{w} in classical terms (the group, the level, the discrete series, the Poincaré series), and of the terms of the expansion (R1).

(ii) A proof of the first bound of (III.246): the main term with constant 1 from (R2), and the bounds for the other pairs. This is the centre of (S).

(iii) A proof of (P): why the pairing of the chain is an inner product with $P_{\mathsf{w}}\varphi$.

(iv) The count (R3) with its rate, and the distribution of the labels in (R4).

(v) A run of the dependency computation for the names that were not queried: `Win_at`, `SizeKRel`, `SizeM`, `SizeN`, `SizeK`, `SizesAt`, `CompressDiagDefectBound`.

(vi) Whether the number $\operatorname{tr}(Q^2)D/(\operatorname{tr} Q)^2 - 1$ of the family of the source is of the size x^{-C_H} or smaller; only the upper bound is known (the paragraph on the error term at the end of §2).

III.9 The bound for Galois cohomology

This section concerns the bound (II.46) of Part II, §6.8 (Part I, §I.7, "Denominators"): a bound for the lengths of certain Galois cohomology groups of the field E, which the announcement [Alp26] calls a weakening of Vandiver's conjecture. It is the part of the proof that uses the most mathematics from outside analytic number theory, and no proof of it is given here. The section states it exactly, shows how it is used, describes the structure of its proof in the Lean development with the published theorems that correspond to each step, and works out the simplest case.

Proved outright. Lemma G1 (lengths, orders and indices); Proposition G2 (how the bound gives the estimate for the denominators, given the inequality `LinkN_at` of the source).

Proved from theorems of books that we consulted. Proposition G3: for the field $\mathbb{Q}$ the groups vanish when a class group vanishes; with its corollaries for odd twists (Herbrand's theorem), for the twists with $(\ell - 1) \mid 2e$, and under Vandiver's conjecture. The theorems used are in Washington's book [Was82]; their statements and numbers were checked in a copy of the first edition. Lemma G4: the class number of $\mathbb{Q}(\sqrt{\ell})$ is less than ℓ; it is proved from Minkowski's theorem on lattice points [Neu99, Ch. I, (4.4)]. The numerical statements of §5 are computed exactly.

Proved from the standard statements of class field theory. The vanishing of $\mathrm{Sha}_5(\mathbb{Q}(i), 4)$, at the end of §5. The proof is for this one field and this one twist.

Taken from the source without proof. Theorem G (the bound itself, Lean: `VWinAt` for the trivial character); the inequality `LinkN_at`; and every statement of §4.

Citations. They are of three kinds. (1) The books [Was82], [Neu99] and [NSW00] were read in local copies ([Was82] and [NSW00] in the first edition); the test group G19 compared each statement cited from them with the page. (2) The articles [Sou03], [Kat04], [Kur92], [BC04], [Car12], [Car14], [CH13], [BS23] and the book [Rub00] were not available to us. Where a theorem or a section of

one of them is named, the wording was compared with the publication by an outside reviewer, and we say no more than his report says. (3) Everything else is cited from memory and is marked so: [Sou79], the theorem of Voevodsky and Rost, [MW84], [FW79], the work of Harris and Taylor and of Shin, and the last column of the table of §4.

Letters introduced in this section: ℓ (a prime, as in §III.5), e (the twist; Part II, §6.8 writes a), t (a number of twists; in the construction it is the parameter written t_0 in §III.5, and t in §III.4; Part II, §6.8 does not name it and writes the bound $\lfloor x^{400} \rfloor$), V_0, S_ℓ, $G_{E,\ell}$, G_v, W_e, $\mathrm{Sha}_\ell(E,e)$, $h_{\ell,e}$, $h^{\pm}_{\ell,e}$, $\jmath$ (complex conjugation), ψ (a Dirichlet character), $\mathfrak{B}$, $\eta_j(w)$, H_j, $y = x^\varepsilon$, V, e_O, f_O, κ, c_ψ, x'; in §4: $\mathcal{K}$, $\mathcal{R}$, R, R', c_0, ϑ, ϱ; and in §5: A, $A^{(i)}$, X_∞, H_∞, Γ, Π, ω, F, h, $\mathfrak{d}$, $\mathfrak{a}$, α, T, L, j, N, c, δ, λ, z, ϕ, a_1, a_2, rec, u, ϵ, β. The letters $x, \varepsilon, b, m, f, q, d, n$ have their meaning from Parts I and II; in particular $b = \exp(x^\varepsilon)$.

The letters are local. The letters of this section have the meanings given here in this section only. The clashes most likely to confuse are these. n is the power of 2 with $[E:\mathbb{Q}] = 4n$; §III.5 takes its n from the record of the construction (`ord`), and §1 says how the order of the frame is related to the n of Theorem G; Soulé's twist, which he writes n, is written n' in §4. t is a number of twists, the t_0 of §III.5; in §III.5 the letter t is an index. e is the twist; $e(w)$ and e_O are ramification indices (the e_w of §III.6); e is never the base of the natural logarithm here. L_0 is the base field of a tower, in §4 only; the coefficient field, which is L_0 in §III.6, is called K here, and $\mathcal{K}$ is an imaginary quadratic field. ψ is a Dirichlet character and $\Psi = \psi\chi_\ell^e$ a character; in §III.2 $\psi(u,\chi)$ is a sum over prime powers, in §III.4 and §III.7 ψ and Ψ are ranks, and in §III.8 they are weights and unitary maps. C is the constant of (III.257), or an exponent in the table of §4; it is not the C of another section, and it is not one of c (a homomorphism, in §5), c_0, c_ψ. Inside the section four letters have two or more uses. T is a finite group in Lemma G1 and a real number in Lemma G4. N is a level or a conductor in §3 and §4, the norm $N(w)$, $N(\mathfrak{a})$ of a prime or an ideal, and the norm map of class groups in the proof of the corollaries. j numbers the maps of the chain in §2 and is the map that extends ideals in §5. $\epsilon = \pm 1$ is a sign in Step 4 of §5; it is not ε.

1. The statement in the source

The groups. Let E be a number field, ℓ a prime and e an integer. Let S_ℓ be the set of primes of E above ℓ, and $G_{E,\ell}$ the Galois group over E of the largest extension of E in which no finite prime outside S_ℓ ramifies. Let $\chi_\ell\colon G_{E,\ell} \to \mathbb{Z}_\ell^\times$ be the cyclotomic character, and let $W_e = \mathbb{Q}_\ell/\mathbb{Z}_\ell(-e)$: the discrete group $\mathbb{Q}_\ell/\mathbb{Z}_\ell$ on which σ acts by multiplication by $\chi_\ell(\sigma)^{-e}$. For $v \in S_\ell$ let $G_v \subset G_E$ be a decomposition group at v, mapped to $G_{E,\ell}$. Put

$$\begin{aligned} \mathrm{Sha}_\ell(E,e) &= \ker\Big(H^1(G_{E,\ell}, W_e) \longrightarrow \prod_{v\in S_\ell} H^1(G_v, W_e)\Big), \\ h_{\ell,e} &= \mathrm{length}_{\mathbb{Z}_\ell} \mathrm{Sha}_\ell(E,e) \ \in\ \{0,1,2,\dots\} \cup \{\infty\}. \end{aligned} \tag{III.251}$$

Cohomology is that of continuous cochains: H^1 is the group of continuous maps ξ with $\xi(\sigma\sigma') = \xi(\sigma) + \sigma\cdot\xi(\sigma')$, modulo the maps $\sigma \mapsto \sigma\cdot w - w$. So $\mathrm{Sha}_\ell(E,e)$ consists of the classes that are unramified outside ℓ and trivial at every prime above ℓ, as in Part II, §6.8. The length of a $\mathbb{Z}_\ell$-module is the length of a longest chain of submodules; a finite module of length h has ℓ^h elements, and the length is ∞ if there is no finite bound. In the Lean: `LData` (a choice of a place of $\overline{E}$ above each $v \in S_\ell$ and of the character χ_ℓ on $G_{E,\ell}$), `Sha` and `shaLen` (`Formalism/GaloisCohomology/Statement.lean`, lines 54, 75, 78); the value ∞ is written $\top$. The length does not depend on the choices; this is proved in the source (`shaLen_eq_shaInv`).

The statement VVle(E,t,V_0), for an integer $t\ge 1$ and a real number V_0 (Lean: `VVle`, in the namespace `UpperTheory.Top`; `ExistsPageN.lean`:12). *For every algebraic closure of E and every choice of the data above, for every prime ℓ:*

$$h_{\ell,e}<\infty\quad(1\le e<t),\qquad\text{and}\qquad\sum_{\ell\in S}\log\ell\sum_{1\le e<t}h_{\ell,e}\ \le\ V_0\quad\text{for every finite set } S \text{ of primes.}\tag{III.252}$$

(The first clause is needed: in the Lean sum an infinite length is counted as 0.) If VVle(E,t,V_0) holds, $t'\le t$ and $V_0\le V_0'$, then VVle(E,t',V_0') holds (`VVle.mono`).

The field. For $n\ge 1$ a power of 2 and primes $q\equiv 3 \pmod 8$, $d\equiv 7 \pmod{16}$, let $E=\mathbb{Q}(\zeta_{4n}+\zeta_{4n}^{-1},\sqrt{-q},\sqrt{-d})$. It is the subfield of a cyclotomic field fixed by the automorphisms $\zeta\mapsto\zeta^\alpha$ with $\alpha\equiv\pm1 \pmod{4n}$ and $(\frac{\alpha}{q})=(\frac{\alpha}{d})=1$ (Lean: `IsChiField` for the trivial character, line 19). It is a field with complex multiplication, of degree $4n$ over $\mathbb{Q}$, and its Galois group over $\mathbb{Q}$ is $\mathbb{Z}/n\times(\mathbb{Z}/2)^2$.

Theorem G [quoted]. (Lean: `VWinAt` for the trivial character and the exponent 401, lines 44–49 of the same file; proved in the source as `vwin_of_row`.) *Let $0<\varepsilon<1/52$, and for $x\ge1$ let $n(x)$ be a power of 2 with $x^{2+8\varepsilon}\le n(x)\le 2x^{2+8\varepsilon}$. There are $J\ge1$ and X_0 with the following property. Let $X\ge X_0$, put $Y_0=X^{1/J}$, and let $q\equiv3 \pmod 8$ and $d\equiv7 \pmod{16}$ be primes with*

$$\tfrac12 Y_0^{\varepsilon/5}<\log q\le Y_0^{\varepsilon/5},\qquad \tfrac12 Y_0^{\varepsilon/5}\log^2 Y_0<\log d\le Y_0^{\varepsilon/5}\log^2 Y_0,\qquad q\mid d+1,\tag{III.253}$$

and such that every odd prime $p\le Y_0^{\varepsilon/5}$ divides $d+1$. Then there is Y with $Y_0\le Y$ and $Y^2\le X$ such that, for every $x\in[Y,Y^2]$ and for the field E with $n=n(x)$,

$$\text{VVle}(E,\ \lfloor x^{401}\rfloor,\ \sqrt{b})\text{ holds},\qquad b=\exp(x^\varepsilon).\tag{III.254}$$

The congruences and sizes required of (q,d) are those of the primes of Estimate B (Lean: `WinPair`). The interval $[Y,Y^2]$ is the "window"; its origin is explained in §4.

From Theorem G to the bound of Part II. The bound (II.46) of Part II has $\lfloor x^{400}\rfloor$ where (III.254) has $\lfloor x^{401}\rfloor$. The source passes from one to the other by a change of scale (Lean: `winPairs`, `WinPairs.lean`:5–81). Let $c_\psi\ge1$ be the constant of the source such that the order of the frame of the construction is c_ψ times the power of 2 of Theorem G (the constant of `fieldChi`; we did not evaluate it for the trivial character). Let $\kappa\ge1$ be the number with $\kappa^{2+8\varepsilon}=c_\psi$. Theorem G is applied at a point x', and the construction works at the scale $x=\kappa x'$. Then $x^{2+8\varepsilon}=c_\psi\,x'^{\,2+8\varepsilon}$, so the bounds for $n(x')$ become $x^{2+8\varepsilon}\le c_\psi\,n(x')\le 2x^{2+8\varepsilon}$. The proof takes x' so large that $\kappa^{400}\le x'$. Hence

$$\lfloor x^{400}\rfloor=\lfloor\kappa^{400}x'^{\,400}\rfloor\le\lfloor x'^{\,401}\rfloor,\qquad \exp(x'^{\,\varepsilon})\le\exp(x^\varepsilon),\tag{III.255}$$

and (III.254) at x' gives VVle$(E,\lfloor x^{400}\rfloor,\sqrt{b})$ with $b=\exp(x^\varepsilon)$, by `VVle.mono`. This is the field `bound` of the record `WinDatum` (`TwistForm.lean`:25). The window $[Y,Y^2]$ for x' becomes a window $[Z,Z^{3/2}]$ for x, with $Z=\kappa Y$; this uses $\kappa^3\le Y$. So the remark of Part II, §6.8, that the source states the bound for twists below x^{401} at a slightly different scale, is accurate.

The number of twists of the construction (Lean: `R.t`) is not a free parameter. The condition `RankVal` (`PairRowIffQdRows.lean`:28–32) fixes it by an equation in the other parameters, and `eventually_rank_le_rpow_of_lt_one_div_fiftytwo` proves that it is at most x^{400} for $x\ge 2^{100}$. The theorem `supplyOfAdmOf` then applies `VVle.mono` once more. The family of the construction receives the result in the form `HypV`: for large x, VVle$(E,t,\sqrt{b})$ holds for its fields E and its number of twists t (`DivinNOfHypV.lean`:10). The declarations named in this paragraph and in the preceding one are among those the final theorem depends on.

2. From lengths to indices of lattices

Lemma G1. (i) *Let T be a finite abelian group and h_ℓ the length of its ℓ-primary part as a $\mathbb{Z}_\ell$-module. Then* $\log|T| = \sum_\ell h_\ell \log \ell$. (ii) *Let O be the ring of integers of a finite extension of $\mathbb{Q}_\ell$, with ramification index e_O and residue degree f_O. If T is an O-module of finite length, then* $\operatorname{length}_{\mathbb{Z}_\ell} T = f_O \operatorname{length}_O T$ *and* $|T| = \ell^{f_O \operatorname{length}_O T}$. *If T_0 is a finite $\mathbb{Z}_\ell$-module, then* $\operatorname{length}_O(O \otimes_{\mathbb{Z}_\ell} T_0) = e_O \operatorname{length}_{\mathbb{Z}_\ell} T_0$. (iii) *Let $M' \subset M$ be lattices of the same rank over the ring of integers O_K of a number field K. Then* $\log[M : M'] = \sum_w \operatorname{length}_{O_{K,w}} ((M/M')_w) \log N(w)$, *the sum over the primes w of K; here $N(w)$ is the norm and the subscript w is the completion.*

Proof. (i) A finite $\mathbb{Z}_\ell$-module of length h has a chain of submodules with h successive quotients $\mathbb{Z}/\ell$, so it has ℓ^h elements; and T is the product of its primary parts. (ii) The simple O-module is the residue field, which has ℓ^{f_O} elements and length f_O over $\mathbb{Z}_\ell$. For $T_0 = \mathbb{Z}/\ell^s$ the module $O/\ell^s O$ has length $e_O s$, because ℓO is the e_Oth power of the maximal ideal; a general T_0 is a product of such groups. (iii) M/M' is the product of its completions, by the Chinese remainder theorem, and a module of length h over $O_{K,w}$ has $N(w)^h$ elements. □

By (ii), $\operatorname{length}_O(O \otimes T_0)/e_O$ does not depend on O. The Lean uses this normalisation when it enlarges the coefficients to treat a group character by character (`Chr.nrm`: the length divided by the length of $O/\ell O$, which is e_O; Lean: `eO`).

The inequality of the source. The construction has a coefficient field, a number field which we call K here (Lean: `LL`), and for each map of the chain, numbered $j = 1, \dots, m$, and each prime w of K a rational number $\eta_j(w) \geq 0$ (Lean: `etaJ`). A prime w is called *twisted* if the character twist chosen at w is not the base one (Lean: $\mu_w \neq \mu_0$). Put

$$H_j = \sum_{w \text{ twisted}} \frac{\log N(w)}{[K : \mathbb{Q}]}\, e(w)\, \eta_j(w), \tag{III.256}$$

where $e(w)$ is the ramification index of w over $\mathbb{Q}$ and the sum is over a finite set (Lean: `etaSum`, `GlobalTheoremInputs.lean`:36). The sum runs over twisted w only.

What is known about the meaning of $\eta_j(w)$. The number $\eta_j(w)$ of (III.256) is the field `etaJ` of the record of data of the final theorem (`DataRows.lean`:162). The source also has a number with the same name for a family of lattices (`DefFam.etaJ`, `LiftLat.lean`:147). For that number the meaning is clear. Times $e(w)$ it is the length of a quotient of two lattices at w: the definition is `BlockDat.etaW` (`LiftLat.lean`:116), the length over the localisation at w of $A/(A \cap A')$ for two lattices A, A' (`locLen`), and the relation $\eta \cdot e(w) =$ `etaW` is `eta_units` (`LiftLatLemmas.lean`:46). The final theorem depends on `etaW`, `locLen` and `eta_units`. *It is not established here that the field `etaJ` of the record equals the number of the family at the twisted primes.* An earlier version of this section cited the declarations `TieEta` (`Statement.lean`:90) and `etaN_eq_etaJ` (`CoverAct.lean`:108) for this. By the dependency computation of the test group G19, both are in the archive and the final theorem depends on neither. So they establish nothing about the $\eta_j(w)$ that the final theorem uses. The statement that is missing is this: for the record of data of the final theorem and every twisted w, $e(w)\,\eta_j(w)$ equals `etaW` at w, a length of $A/(A \cap A')$. The archive contains statements of this form (the field `etaJ` of the record `NumTie`, `RowSideTie.lean`:244, and the theorem `etaJ_join`); whether the final theorem depends on them was not computed. The record itself only says that $\eta_j(w)$ is a sum of local numbers (its field `etaJ_tie`). To decide the question one has to read the proof of `LinkN` (`linkN_of_twGood_galoisPreconditions`, which the final theorem depends on). Until this is done, the following is an expectation and not a statement of this paper: that $e(w)\eta_j(w)$ is the length at w of a quotient M/M' of two lattices over O_K; that, by Lemma G1(iii), $[K : \mathbb{Q}]\, H_j$ is the part of $\log[M : M']$ that comes from the twisted primes; and that $\eta_j(w)$ is the defect of the

index of lattices of Part II, §6.8. No proof in this paper uses it. Proposition G2 and the chain use only the inequality `LinkN_at`.

`LinkN_at` (line 40) [quoted]. *If* $\sum_\ell \log \ell \sum_{1\le e<t} h_{\ell,e}$ *converges, then*

$$H_j \;\le\; d_j\Big(C(f\,m\,b\,x^{2\varepsilon} + (\log x)^2 N_c) + \sum_\ell \log \ell \sum_{1\le e<t} h_{\ell,e}\Big). \tag{III.257}$$

Here C is a constant, d_j is a dimension (Lean: `dZO`; the statement `SmallLevelRankEq` says that d_j equals a number D_0 not depending on j), $f \ge 1$ and $m \ge 1$ are integers of the construction (m is the number of maps), and $N_c \ge 0$ (Lean: `budgetNc`). The statement `LinkN` asserts (III.257) for all members of the family, for large x, with one C, assuming the first clause of (III.252). It is proved in the source; see §6(ii).

Proposition G2 (Lean: `divinN_at`, `DivinNOfHypV.lean`:76). *Let* $x > 1$, $y = x^\varepsilon$, $b = \exp(y)$, $f, m \ge 1$, $N_c \ge 0$, $D_0 \ge 0$, C *real,* $B_0 = f\,m\,b\,y$. *Suppose* $(\log x)^2 \le y$, *that* (III.257) *holds with* $d_j = D_0$, *and that* $V = \sum_\ell \log \ell \sum_{1\le e<t} h_{\ell,e} \le b$. *Then*

$$H_j \;\le\; D_0\,(|C|+1)\,x^\varepsilon\,(B_0 + N_c). \tag{III.258}$$

Proof. Since $f\,m\,b\,y^2 = yB_0$ and $(\log x)^2 N_c \le yN_c$, the first term in the bracket of (III.257) is at most $|C|\,y\,(B_0 + N_c)$. Since $f, m, y \ge 1$, we have $V \le b \le B_0 \le y(B_0 + N_c)$. □

(III.258) is the statement `DIVinN_at`; it is the bound for the denominators that the rest of the chain uses. Theorem G gives $V \le \sqrt{b}$, and $\sqrt{b} \le b$. So at this step a bound by b would suffice. The square root is not used anywhere else. In the source the statement `HypV` is used by one theorem only, `divinN_of_hypV` (its single use is in `DesWin.lean`:29), and that theorem replaces $\sqrt{b}$ by b at once, by $\sqrt{b} \le b$ for $b \ge 1$. Besides the bound it takes from `HypV` the finiteness of the lengths and the convergence of the sum (`vvRecord`). So a bound by b in Theorem G would suffice for the deduction.

3. Why the statement is of Vandiver type

Let ℓ be odd and let $\jmath$ be complex conjugation on E. It acts on $\mathrm{Sha}_\ell(E, e)$ as an involution, and the Lean splits the group into the part where $\jmath = (-1)^e$ and the part where $\jmath = (-1)^{e+1}$, of lengths $h^+_{\ell,e}$ and $h^-_{\ell,e}$; $h_{\ell,e} = h^+_{\ell,e} + h^-_{\ell,e}$ (`shaLen_eq_add`, proved there by linear algebra). Let E be abelian over $\mathbb{Q}$ and $\ell \nmid [E:\mathbb{Q}]$. In the classical theory the group then splits, after enlarging the coefficients to a ring O, into parts indexed by the Dirichlet characters ψ of $\mathrm{Gal}(E/\mathbb{Q})$; the part of ψ lies in the plus part if $\psi(-1) = (-1)^e$ and in the minus part otherwise. The source does not state a decomposition. It proves an inequality (`HV.redPrime`): for $\ell \ne 2$ and $\ell \nmid [E:\mathbb{Q}]$,

$$e_O\, h^+_{\ell,e} \;\le \sum_{\psi(-1)=(-1)^e} \mathrm{length}_O\,(\text{the group of } \psi), \tag{III.259}$$

the sum over the characters ψ of $\mathrm{Gal}(E/\mathbb{Q})$ with values in O. The group of ψ is defined over $\mathbb{Q}$, with coefficients twisted by ψ (Lean: `Vlen`, the length of `Chr.V'` at the level ℓN, N the conductor of E). For the other ℓ there is a similar inequality with a loss term (`HV.redAll`). For the field E of Theorem G every odd ℓ is prime to $[E:\mathbb{Q}] = 4n$.

By the duality theorem of Poitou and Tate, $\mathrm{Sha}_\ell(E, e)$ has the same length as the kernel of $H^2(G_{E,\ell}, \mathbb{Z}_\ell(e+1)) \to \prod_{v|\ell} H^2(G_v, \mathbb{Z}_\ell(e+1))$. The duality is [NSW00, §8.6, Thm. 8.6.8], and the long exact sequence that contains it is Theorem 8.6.13 there; these numbers are those of the

first edition, which was read. Theorem 8.6.8 is stated there for finitely generated modules and Theorem 8.6.13 for finite modules. For the statement used here, with the coefficients $\mathbb{Q}_\ell/\mathbb{Z}_\ell(-e)$ and $\mathbb{Z}_\ell(e+1)$, one passes to limits, and we did not write this out. In the Lean this equality of lengths is the statement `PTShaAt` (`Statement.lean`:101); it is stated under the condition that $\ell \neq 2$ or E has no real place. For $e \geq 1$ the group $H^2(G_{E,\ell}, \mathbb{Z}_\ell(e+1))$ is finite (Soulé [Sou79], from memory) and, for odd ℓ, isomorphic to $K_{2e}(O_E) \otimes \mathbb{Z}_\ell$, where K_{2e} is Quillen's K-group (the conjecture of Quillen and Lichtenbaum, a consequence of the theorem of Voevodsky and Rost; from memory). Take $E = \mathbb{Q}$. For odd e the order of $K_{2e}(\mathbb{Z}) \otimes \mathbb{Z}_\ell$ is the power of ℓ in the numerator of $B_{e+1}/(e+1)$: this is the minus part, and it is governed by a value of ζ (Mazur and Wiles [MW84], from memory). For even e the group is the plus part. Vandiver's conjecture says that ℓ does not divide the class number of $\mathbb{Q}(\zeta_\ell + \zeta_\ell^{-1})$. Kurihara [Kur92] relates it to the vanishing of the groups $K_{4s}(\mathbb{Z})$, $s \geq 1$: his Proposition 3.7 proves one implication, and his Remark 3.11 gives the converse under the comparison of K-theory with cohomology that is used there (compared with the article by a reviewer, not by us). An equivalence for the integral groups also needs that comparison and a treatment of the 2-primary component. We do not use an equivalence. Corollary (c) of §5 proves the sufficient implication that is used here: under Vandiver's conjecture, $h_{\ell,e}(\mathbb{Q}) = 0$ for all odd ℓ and all even $e \geq 2$.

Theorem G is weaker than a statement of vanishing in two ways: it bounds a sum weighted by $\log \ell$, and the bound $\sqrt{b} = \exp(x^\varepsilon/2)$ is very large. It is stronger in one way: the field is E, of degree about $4x^{2+8\varepsilon}$, and the twists go up to x^{401}. Theorem G asserts no vanishing for the plus part of E. For the field $\mathbb{Q}$ some plus parts do vanish: those with $(\ell-1) \mid 2e$ (Corollary (b) of §5; this is the statement `PropZ` of the source under the identification discussed in §5; for example $\ell = 5$, $e = 2$), and all of them under Vandiver's conjecture (Corollary (c)).

4. The structure of the proof in the source

What follows was obtained by reading statements, and the proof terms that connect them, downward from `vwin_of_row`. The long proofs were not read. Sources in the literature are cited as said in the paragraph "Citations" at the head of the section: from memory, unless a theorem or a section is named.

Two dependency computations were made for this section, by the test groups G12 and G19 (files `outputs/depcheckG12.out` and `outputs/depcheckG19.out` in `zeta-paper_tests/displays/G12` and `G19`). They give the same number of declarations that the final theorem depends on (1 478 198) and the same three axioms. The version of this section that was tested prints 133 Lean names. G12's run covers 89 of them: 84 are among the declarations the final theorem depends on, and five are not. G19's run covers 35 of the other 44: 33 are among them, and two are not. The remaining nine are not names of declarations (local names, fields of records and the three axioms). The seven exceptions are these. `OneChar`, `Adm3`, `Adm2`, `Sp4In` and `KatoUse` are in the archive but the final theorem does not depend on them. It uses variants: `OneCharU_s` (`NumberU3US.lean`:49) in place of `OneChar`, with the conditions `Adm3U_s` and `Adm2A` in place of `Adm3` and `Adm2`; and `SymplecticFourImageDense` in place of `Sp4In`. That the variants are used is in G19's run and in a run of the authors, not in G12's. The other two exceptions are `TieEta` and `etaN_eq_etaJ`; see §2. Names first printed in the present version (`NumTie`, `etaJ_join`, `etaJ_tie`, `DefFam.etaJ`) were not tested.

The top. `vwin_of_row` takes one hypothesis, `RowOb` (`WeightsConjBis.lean`:69): *for every field E with complex multiplication, every $s \geq 1$, every regular algebraic cuspidal automorphic representation π of GL_s over E that is conjugate self-dual up to a twist, every ℓ and every*

isomorphism $\iota\colon \overline{\mathbb{Q}}_\ell \simeq \mathbb{C}$*, there is a continuous semisimple representation of* $\mathrm{Gal}(\overline{E}/E)$ *on* $\overline{\mathbb{Q}}_\ell^s$*, with predicted weights and the local comparisons in the record* `PropsB` (lines 54–67). The record says the following. The representation is continuous and semisimple. At the primes not above ℓ it is compared with π in a weakened sense (`AwayB`). For each embedding τ of E a multiset of integers is given, and it equals the multiset computed from the infinitesimal character of π; this equality, and a symmetry under complex conjugation, are stated without a condition. They prescribe the multiset. They do not say that it is the multiset of the jumps of the representation. The assertions about the actual jumps are in the part `AboveB` (lines 39–52). They are made with respect to a local comparison datum (Lean: `LocBridge`, `Pst.Induces`), and in two cases only: where the local component of π is spherical (the representation is then crystalline and de Rham there, with the predicted jumps), and in a specified local case (the conditions `TmW` and `Sqf` with `IsQpBig`, or `Tor`, together with `mono0`; we did not transcribe them). The classical comparison is with the theorem stated in the introduction of Chenevier and Harris [CH13] and with the compatibility of local and global correspondences of Caraiani [Car12], [Car14, Thm. 1.1] (these by the reviewer's report); they build on the work of Harris and Taylor and of Shin (from memory). The normalisations and the hypotheses on the two sides have to be specified before an exact dictionary can be claimed. We have not done this. In the Lean `Row0b` is a theorem proved in the archive (`galoisLocalGlobalStrong_recParameter`). The theorem `vwin_of_row` depends only on the three standard axioms of Lean (`propext`, `Classical.choice`, `Quot.sound`), by the dependency computation. So no statement on this route is left as an assumption other than `Row0b`.

All other inputs named "literature" on this route are theorems proved in the archive. They are passed through eight structures whose fields are hypotheses. Each structure is built by a term, as follows.

(1) `OpenImageCondition`: one field, `Row0b`, filled in `vwin_of_row`. The field is never read: the declarations the final theorem depends on contain no projection and no case analysis of this structure. In its file 62 theorems take it as an argument and ignore it.

(2) `LiteratureRankNHypothesis`: four fields (local, automorphic, cohomological, complex multiplication), filled by the next four structures. The last field is not read.

(3) `LitLocAlt`: 18 fields, filled by the term `local_inputs_from_literature`, each field by a theorem of the archive; 11 are read.

(4) `LiteratureAutomorphyHypothesis`: 15 fields; 13 are filled by theorems of the archive, one by the theorem `baseChangeN`, and one by the hypothesis `Row0b`; 10 are read.

(5) `LitCoh`: 12 fields, filled by `galois_cohomology_from_literature`; all are read.

(6) `LiteratureCMHypothesis`: 11 fields, filled by `cm_abelian_inputs_from_literature`; none is read.

(7) `HypothesesH'`: 15 fields with those of its parent; all but one are read. It is filled by the term `published_hypotheses_from_patching_alt`.

(8) `U3.CohInp`: 5 fields, filled by `Core0.cohInp_alt`; all are read.

The propositions passed as arguments further down (`LemmaK`, `CountI` to `CountV`, `CountZ`, `KatoTwoW`, `ThmBab`, `ThmBcFree`, `PsiBound`, `FerreroWashington`, `SymplecticFourImageDense` and others) are each supplied by a theorem that the final theorem depends on.

The structure `OpenLiteratureStatements` is not on this route. It belongs to the other half of the final theorem, which proves `LinkN`. It has three fields, and all three are read. Two of them are filled with the hypothesis `Row0b`; their names are `caraianiLocalGlobalDual` and

`galoisRepOfPolarizedPairLit`. The third is filled with the theorem `recTExists`; its name is `localLanglandsExists`.

The decomposition (`vvle_of_parts`, `vwin_of_counts`). The sum in (III.252) is split into the prime 2, the minus parts and the plus parts for odd ℓ. Let $\mathfrak{B}$ be a finite set of odd primes depending only on the character (Lean: `Bchi`); for ζ the definitions give $\mathfrak{B} = \{3, 7\}$. In the table, N is the conductor of E.

Case	Statement in the Lean, in classical terms	Related published work (from memory)
$\ell = 2$; and $\ell \in \mathfrak{B}$, plus part	`ThmBab`: on the window, $\sum_{e<t} h_{\ell,e} \le \exp(x^{4\varepsilon/5})$ for $\ell = 2$ over E, and for $\ell \in \mathfrak{B}$ over bounded layers above a fixed field (layers in the sense of "The window" below)	Ferrero and Washington [FW79] ($\mu = 0$); Iwasawa's theory of $\mathbb{Z}_\ell$-extensions; Soulé [Sou03]
Minus part, ℓ odd	`ThmBcFree`, for $n \ge 4$ and $t \ge 2$: $\sum_\ell \log \ell \sum_{e<t} h^-_{\ell,e} \le 3\,[E:\mathbb{Q}]\, t^2 \log(tN)$, plus a term for the odd primes dividing the degree of the field of the character ψ of the source (Lean: `M.deg`; none for ζ). From `PsiBound`: the length of the part of ψ is at most the valuation of $N \prod_{p \le e+2} p \cdot B_{e+1,\psi}$, if $\ell \nmid [E:\mathbb{Q}]$ and $\psi(-1) = (-1)^{e+1}$; the hypotheses are stated after the table	Stickelberger; Mazur and Wiles [MW84]; Euler systems of Gauss sums (Kolyvagin; Rubin [Rub00, §2.2])
Plus part, $\ell \notin \mathfrak{B}$, $(\ell-1) \nmid 2e$ (Case I)	`CountI`, from `OneCharU_s`: under the hypotheses stated after the table, the length is at most $2 \sum_{\rho \in \mathcal{R}} v_\ell\big(\operatorname{tr}\rho(\sigma) - \Psi(\sigma) - 1 - \Psi(\sigma)^{-1}\big)$, $\Psi = \psi\chi_\ell^e$, for a finite set $\mathcal{R}$ of three-dimensional representations ρ of the Galois group of an imaginary quadratic field	congruences with Eisenstein series (Ribet; Mazur and Wiles; Bellaïche and Chenevier [BC04]); modularity lifting (Taylor and Wiles); Kato [Kat04]
Plus part, $\ell \notin \mathfrak{B}$, $(\ell-1) \mid 2e$, $\psi = 1$ (Case Z)	`PropZ`: for $\ell \ge 5$ and even $e \ge 2$ the group over $\mathbb{Q}$ vanishes	class numbers of $\mathbb{Q}$ and $\mathbb{Q}(\sqrt{\ell})$; see §5, Corollary (b), and the identification discussed there
Plus part, the remaining ψ with $(\ell-1) \mid 2e$ (Cases II, III)	`CountII`, `CountIII`: the length times $\log \ell$ is at most $(\ell N e)^C$; from a statement on automorphic forms on GSp_4 (`SymplecticFourImageDense`). Case III is empty for ζ	not identified
Plus part, $\ell \in \mathfrak{B}$ (Cases II, IV, V)	`SumIn`: reduction to the layers of the first row	as in the first row

The last column names related work. It does not say that the statement of the Lean is a published theorem; for that question see the three lists at the end of the section. `CountV` has the shape of `CountII`, for Case V with a further condition that we did not transcribe.

The row for the minus part, with its hypotheses (`PsiBound`, `ZeroMemUnrHom.lean`:97) [quoted]. Let E be an abelian field with complex multiplication, and $N \ge 1$ a level with a homomorphism $(\mathbb{Z}/N)^\times \to \mathrm{Gal}(E/\mathbb{Q})$ compatible with the action on roots of unity (Lean: `IsCycAct`). Let ℓ be an odd prime with $\ell \nmid [E:\mathbb{Q}]$. Let O be a discrete valuation ring, finite over $\mathbb{Z}_\ell$, inside an algebraic closure of $\mathbb{Q}_\ell$, and let v_O be its additive valuation, which is 1 on a generator of the maximal ideal. Let $e \ge 1$ and let ψ be a character of $\mathrm{Gal}(E/\mathbb{Q})$ with values in $O^\times$ and $\psi(-1) = (-1)^{e+1}$.

Then

$$\operatorname{length}_O\big(\text{the group of } \psi\big) \;\le\; v_O\Big(N \prod_{p \le e+2} p \cdot B_{e+1,\psi}\Big), \tag{III.260}$$

with the Bernoulli number of the primitive character. Both sides are in units of length over O; on $\mathbb{Z}_\ell$ we have $v_O = e_O\, v_\ell$. So `PsiBound` is an upper bound for a cohomology length in terms of the valuation of a Bernoulli number with explicit integral factors. It is not a bound for the valuation. For $\psi = 1$ the field E is still a field with complex multiplication, so the statement does not apply to $E = \mathbb{Q}$. To use it over $\mathbb{Q}$ one needs a descent, for example from the part of the trivial character over $\mathbb{Q}(i)$; we did not carry this out. Rubin [Rub00, §2.2] gives general bounds from Euler systems under explicit hypotheses (by the reviewer's report). The specialisation to the bound by a Bernoulli number is not supplied here.

The row for Case I, with its hypotheses (`OneCharU_s`, `NumberU3US.lean`:49) [quoted]. Let $\mathcal{K}$ be an imaginary quadratic field, ℓ a prime, e an integer, N a level and ψ a Dirichlet character modulo N with values in the coefficient ring. Assume the twelve conditions of the record `GD` (`Counts.lean`:9); among them: $\ell \neq 2$; $e \ge 1$; N is even; ℓ does not divide the order of ψ; ℓ does not divide the discriminant of $\mathcal{K}$ times its class number; the discriminant of $\mathcal{K}$ is prime to N; $\psi(-1) = (-1)^e$; and $(\ell - 1) \nmid 2e$. Fix an isomorphism $\iota\colon \overline{\mathbb{Q}}_\ell \simeq \mathbb{C}$. Call a three-dimensional representation ρ of $\mathrm{Gal}(\overline{\mathcal{K}}/\mathcal{K})$ over $\overline{\mathbb{Q}}_\ell$ *admissible* if it satisfies

$$\texttt{Adm3U_s}(\iota, e, N, \rho) \qquad \text{or} \qquad \texttt{Adm2A}\big(\iota, e, N\,|\mathrm{disc}\,\mathcal{K}|, \rho\big). \tag{III.261}$$

The same ι, e and levels are used in the hypothesis and in the conclusion. Let $\sigma \in \mathrm{Gal}(\overline{\mathcal{K}}/\mathcal{K})$ be such that $\operatorname{tr}\rho(\sigma) \neq \Psi(\sigma) + 1 + \Psi(\sigma)^{-1}$ for *every* admissible ρ. Then the length of the group of ψ is finite, and there is a finite set $\mathcal{R}$ of admissible representations, no two of them isomorphic, such that the normalised length is at most $2\sum_{\rho\in\mathcal{R}} v_\ell\big(\operatorname{tr}\rho(\sigma) - \Psi(\sigma) - 1 - \Psi(\sigma)^{-1}\big)$. The condition on σ makes every term of the sum finite. For the statement `OneChar`, which has the same shape, the conditions `Adm3` and `Adm2` say in classical terms: ρ is irreducible, automorphic, conjugate self-dual, of weights $(e, 0, -e)$; or $\rho = W \oplus 1$ with W attached to a newform of weight $2e + 1$. Whether `Adm3U_s` and `Adm2A` differ from them in substance was not examined. Bellaïche and Chenevier [BC04, Thm. 1.1] construct an extension in a Selmer group for their characters (by the reviewer's report). This illustrates the method of congruences. A uniform upper bound for a length is a further statement.

Further inputs, all proved in the archive. `LemmaK` (`LemmaK.lean`:13): for every prime ℓ, every totally complex field L containing the ℓth roots of unity (the fourth roots if $\ell = 2$) and every $e \ge 1$, the group $H^2(G_{L,\ell}, \mathbb{Z}_\ell(e+1))$ is finite and the logarithm of its order is at most $\exp\big((2e[L:\mathbb{Q}])^C \log(2|\mathrm{disc}\,L|)\big)$. So `LemmaK` bounds the whole group, for every ℓ, and it includes $\ell = 2$. Finiteness is due to Soulé ([Sou79], from memory). A quantitative bound away from small primes is in [Sou03]: Theorem 2 there discards the torsion at small primes, and Theorem 3 bounds the groups $H^2(O_L[1/p], \mathbb{Z}_p(n'))$ in a product over the primes $p \ge 4n' - 1$, $p \ge [L:\mathbb{Q}] + 2$ only. Here n' is the twist, which Soulé writes n; it is not the n of this section. These two statements of [Sou03] were compared with the article by a reviewer, not by us. We have not identified a published estimate for `LemmaK` in its full small-prime range. The proof of `LemmaK` in the Lean names a statement on the rank of the homology of $\mathrm{GL}_s(O_L)$ (`BOdd`; Borel's theorem). `FerreroWashington`: for an abelian field L containing those roots of unity, the ℓ-rank of the class group of $L(\zeta_{\ell^r})$ is bounded independently of r; it is proved in the archive. In the cyclotomic $\mathbb{Z}_\ell$-tower, a bounded ℓ-rank of the class groups is equivalent to $\mu = 0$ [Was82, Prop. 13.23] (first edition, p. 284); so this is the theorem of [FW79] (from memory) in its form with bounded rank. The theorems of Poitou and Tate, Tate's local duality and Euler characteristic, the formula of Greenberg and Wiles, and the local theorems

of Bloch and Kato are collected in the structure `LitCoh.KatoTwoW` (`KatoTwoW.lean`:10–14): let W be a two-dimensional representation of $\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$ over $\overline{\mathbb{Q}}_\ell$ and θ a character, such that the twist of W by χ_ℓ^{-e} is attached to a newform of weight $2e+1$ (the attachment hypothesis), and assume the hypotheses `Used` and `WitAQ` of the source, which we did not transcribe. Then the two rational Selmer spaces of Bloch and Kato, of W and of $W \otimes \theta$, vanish: every cocycle with values in the vector space over $\overline{\mathbb{Q}}_\ell$ that satisfies the local conditions is a coboundary. Kato [Kat04, §14.1 and Thm. 14.2(1)] proves that the discrete Selmer groups are finite at the critical twists other than the central one (by the reviewer's report). The two statements are of different kinds: the Lean statement is about vector spaces, and it bounds no integral length. In our reading the relevant point is $s = e+1$, on the edge of the critical strip.

The window. The statement `ABound` (`SecondBound.lean`:19–25; line 25 is its second clause) explains it. Let $\ell = 2$ or $\ell \in \mathfrak{B}$, let L_0 be a fixed base field and $R \geq 0$ an integer. Put

$$j_0 = 3 + \lfloor \log_\ell \max(1, R) \rfloor \tag{III.262}$$

(Lean: `jZero`; the logarithm of the Lean is 0 at $R = 0$). The first clause of `ABound` says: there are an integer $c_0 \geq 0$ such that L_0 contains a primitive root of unity of order ℓ^{c_0}, an integer R' with $0 \leq R' \leq R$, and points $s_1, \ldots, s_{R'}$ of a field containing $\mathbb{Q}_\ell$, with a non-archimedean absolute value $|\cdot|_\ell$ extending that of $\mathbb{Q}_\ell$, with this property. We call the s_i the *exceptional centres*. Let $e \geq 1$, let $\vartheta \geq 0$ be an integer, and let L_1 be a *layer of degree* ℓ^ϑ *over* L_0 in the sense of the Lean (`IsLayer`, `StmtFields.lean`:47): a number field $L_1 = L_0(\zeta_{\ell^k})$ for some $k \geq 0$, with $[L_1 : L_0] = \ell^\vartheta$. Let $h_{\ell,e}$ be the length (III.251) for the field L_1. Then for every integer $\varrho \geq 0$

$$h_{\ell,e} \;\leq\; R' \left(\max(\vartheta,\, j_0) + c_0 + \varrho \right) \qquad \text{whenever } |s_i - e|_\ell \geq \ell^{-\varrho} \text{ for all } i. \tag{III.263}$$

If L_0 contains the ℓth roots of unity (the fourth roots if $\ell = 2$), a layer in this sense is the field of degree ℓ^ϑ over L_0 in the cyclotomic $\mathbb{Z}_\ell$-extension of L_0; this is the case in the uses `TwoIn` and `FixedIn`. The statement `ABound` itself does not assume it, since c_0 may be 0. Without it a layer need not lie in the cyclotomic $\mathbb{Z}_\ell$-extension: for $\ell = 2$ and $L_0 = \mathbb{Q}$ the field $\mathbb{Q}(i)$ is a layer with $\vartheta = 1$, and it is not in the cyclotomic $\mathbb{Z}_2$-extension of $\mathbb{Q}$. A second clause (`SecondBound`) compares a layer of degree ℓ^ϑ with a layer of degree ℓ^{j_0}, with a loss $R \cdot \max(\vartheta - j_0,\, 0)$; we do not transcribe it. By (III.263) the lengths are small unless e is ℓ-adically very close to one of at most R exceptional centres. The statement defines its conclusion by the distances from e to the centres. It does not say what the centres are. Iwasawa's theory gives a statement of this shape when $\mu = 0$: a module with $\lambda \leq R$, and the zeros of its characteristic power series in the place of the centres. The Lean statement does not identify the centres with such zeros in the coordinate e, and we did not check this in the proof. For the exceptional e only `LemmaK` is available, and it is useful only if e is at most a small power of x. The window is an interval of x in which no exceptional e lies between that small power and x^{401}; the number J of Theorem G grows with R. This paragraph is our reading of the statements `ABound`, `FirstBound` and `ThmBab`; the proof was not read.

Three claims of earlier notes, compared with the Lean statements. *(1) "The minus part is a Stickelberger index bound."* The statement is a bound of a length by the valuation of a generalised Bernoulli number, as in the table; summed over ℓ it is bounded by $[E : \mathbb{Q}] \log\left(16 N^{e+2} (e+1)!\right)$ (`PrimeBound`). Its proof passes through a reciprocity statement (`Rec`) deduced from the duality of Poitou and Tate, and the files that prove it are named after Kolyvagin's derivative construction and Jacobi sums. A cruder bound of Stickelberger type (`StickBound`) is used only for $\ell \mid [E : \mathbb{Q}]$. *(2) "The plus part uses congruences on $U(3)$ in the manner of Ribet."* Confirmed for Case I at the level of the statement `OneCharU_s`. The inequality bounds a cohomology group from above by a congruence number. Ribet's method produces classes; an upper bound of this kind needs in

addition a comparison of a deformation ring with a Hecke algebra, and the Lean names such inputs (`adjoint_lifting_of_cohomology`, `published_hypotheses_from_patching_alt`). *(3) "Prismatic F-crystals are needed because the local fields at 2 have unbounded degree."* The archive contains the statement `BhattScholzeIntegral`: every lattice in a crystalline representation of the Galois group of any finite extension of $\mathbb{Q}_\ell$ comes from a prismatic F-crystal. Bhatt and Scholze [BS23, Thms. 1.2 and 5.6] give the equivalence of prismatic F-crystals with the finite free $\mathbb{Z}_p$-lattices whose rationalisations are crystalline (by the reviewer's report). By the dependency computation it belongs to the half that proves `LinkN`, not to the proof of Theorem G. The completion of E at a prime above 2 has ramification index n and residue degree 2, so its degree $2n$ is unbounded; the theory of Fontaine and Laffaille needs an unramified field and small weights. So the claim is consistent with the statements. That the result is *needed* cannot be read off from statements.

5. The simplest case: the field $\mathbb{Q}$, and $\mathbb{Q}(i)$

This subsection is an illustration. It is not used by the source. Let ℓ be an odd prime. Let A be the ℓ-primary part of the class group of $\mathbb{Q}(\zeta_\ell)$, let $\Pi = \mathrm{Gal}(\mathbb{Q}(\zeta_\ell)/\mathbb{Q})$, let $\omega\colon \Pi \to \mathbb{Z}_\ell^\times$ be the character with $\omega(\varpi) \equiv \alpha \pmod{\ell}$ when $\varpi\zeta_\ell = \zeta_\ell^\alpha$, and let $A^{(i)}$ be the subgroup of A on which Π acts by ω^i. References to [Was82] are to the first edition, which we consulted; there $A^{(i)}$ is written A_i.

Proposition G3. *Let $e \neq 0$. If $A^{(i)} = 0$ for $i \equiv -e \pmod{\ell-1}$, then $\mathrm{Sha}_\ell(\mathbb{Q}, e) = 0$.*

Proof. Let F_∞ be the field generated over $\mathbb{Q}$ by all roots of unity of ℓ-power order, and let $H_\infty \subset G_{\mathbb{Q},\ell}$ be the subgroup fixing F_∞; then χ_ℓ identifies $G_{\mathbb{Q},\ell}/H_\infty$ with $\mathbb{Z}_\ell^\times$, and H_∞ acts trivially on W_e. The sequence of inflation and restriction reads

$$0 \to H^1(\mathbb{Z}_\ell^\times, W_e) \to H^1(G_{\mathbb{Q},\ell}, W_e) \to \mathrm{Hom}(H_\infty, W_e). \tag{III.264}$$

The group $\mathbb{Z}_\ell^\times$ has a topological generator γ_0, so $H^1(\mathbb{Z}_\ell^\times, W_e) = W_e/(\gamma_0^{-e} - 1)W_e$. This is 0: $\gamma_0^{-e} - 1$ is a nonzero element of $\mathbb{Z}_\ell$ and $\mathbb{Q}_\ell/\mathbb{Z}_\ell$ is divisible. So a class z in $\mathrm{Sha}_\ell(\mathbb{Q}, e)$ is determined by its restriction $\phi\colon H_\infty \to W_e$, a continuous homomorphism with $\phi(\sigma\sigma'\sigma^{-1}) = \chi_\ell(\sigma)^{-e}\phi(\sigma')$ for $\sigma \in G_{\mathbb{Q},\ell}$, $\sigma' \in H_\infty$. Since z is trivial on a decomposition group G_ℓ at ℓ, ϕ vanishes on $G_\ell \cap H_\infty$, and by the last formula on its conjugates. In particular ϕ is unramified at the primes above ℓ; it is unramified elsewhere by definition of $G_{\mathbb{Q},\ell}$. So ϕ factors through X_∞, the Galois group over F_∞ of its largest unramified abelian extension of ℓ-power order. Restricted to $\Pi \subset \mathbb{Z}_\ell^\times$ the formula says $\phi(\varpi \cdot z') = \omega(\varpi)^{-e}\phi(z')$, so ϕ factors through the part $X_\infty^{(i)}$ on which Π acts by ω^i, $i \equiv -e$. Let $\Gamma = \mathrm{Gal}(F_\infty/\mathbb{Q}(\zeta_\ell))$, with generator γ. One prime of $\mathbb{Q}(\zeta_\ell)$ ramifies in F_∞, and it ramifies totally; hence $X_\infty/(\gamma - 1)X_\infty \cong A$ [Was82, Prop. 13.22, at the bottom level]. The proposition does not state that the isomorphism commutes with Π. It does. Let H be the largest unramified abelian extension of $\mathbb{Q}(\zeta_\ell)$ of ℓ-power degree. The isomorphism is the restriction $X_\infty \to \mathrm{Gal}(H/\mathbb{Q}(\zeta_\ell))$, which has kernel $(\gamma - 1)X_\infty$, followed by the inverse of the Artin map $A \to \mathrm{Gal}(H/\mathbb{Q}(\zeta_\ell))$. An element ϖ of $\mathrm{Gal}(F_\infty/\mathbb{Q}) = \mathbb{Z}_\ell^\times$ acts on both Galois groups by conjugation with a lift of ϖ, and restriction commutes with this. The Artin map sends $\varpi\mathfrak{a}$ to the conjugate by ϖ of the image of $\mathfrak{a}$. So $X_\infty^{(i)}/(\gamma - 1) = A^{(i)} = 0$, and $X_\infty^{(i)} = 0$ by Nakayama's lemma for compact modules over $\mathbb{Z}_\ell[[\Gamma]]$ [Was82, Lemma 13.16]. Hence $\phi = 0$. □

Lemma G4. *For every prime ℓ the class number of $\mathbb{Q}(\sqrt{\ell})$ is less than ℓ.*

Proof. Put $F = \mathbb{Q}(\sqrt{\ell})$ and write h for its class number. Let $\mathfrak{d}$ be the discriminant of F: $\mathfrak{d} = \ell$ if $\ell \equiv 1 \pmod 4$, $\mathfrak{d} = 4\ell$ if $\ell \equiv 3 \pmod 4$, and $\mathfrak{d} = 8$ if $\ell = 2$. In each case $\frac{1}{2}\sqrt{\mathfrak{d}} \leq \sqrt{\ell}$. Write α' for the conjugate of $\alpha \in F$, and $N(\mathfrak{a})$ for the norm of an ideal.

Minkowski's bound: every ideal class of F contains an integral ideal of norm at most $\frac{1}{2}\sqrt{\mathfrak{d}}$. This is [Neu99, Ch. I, §6, Exercise 3] for a field of degree 2 with two real places. It is an exercise there,

so we give the proof from two statements of the text, (5.2) and (4.4). Let $\mathfrak{a}$ be a nonzero integral ideal. Its image under $\alpha \mapsto (\alpha, \alpha')$ is a lattice in $\mathbb{R}^2$ of covolume $\sqrt{\mathfrak{d}}\,N(\mathfrak{a})$ [Neu99, Ch. I, (5.2)]. For $T > 0$ the set $|x| + |y| \le T$ is convex and symmetric, of area $2T^2$. Minkowski's lattice point theorem [Neu99, Ch. I, (4.4)] says that it contains a nonzero point of the lattice if $2T^2 > 4\sqrt{\mathfrak{d}}\,N(\mathfrak{a})$. The inequality is strict there. (The case of equality for compact sets is the last sentence of Exercise 2 of §4 there, without proof; we do not use it.) Take $T^2 = 2\sqrt{\mathfrak{d}}\,N(\mathfrak{a})$. For every $T' > T$ the theorem gives a nonzero lattice point with $|x| + |y| \le T'$. The lattice has only finitely many points with $|x| + |y| \le 2T$. So one of them, nonzero, serves for a sequence of numbers T' decreasing to T. We get $\alpha \in \mathfrak{a}$, $\alpha \neq 0$, with $|\alpha| + |\alpha'| \le T$, hence

$$|\alpha\alpha'| \;\le\; \tfrac{1}{4}\bigl(|\alpha| + |\alpha'|\bigr)^2 \;\le\; \tfrac{1}{4}T^2 = \tfrac{1}{2}\sqrt{\mathfrak{d}}\,N(\mathfrak{a}). \tag{III.265}$$

Now let an ideal class be given, and let $\mathfrak{a}$ be an integral ideal in the inverse class. The ideal $(\alpha)\mathfrak{a}^{-1}$ is integral, because $\alpha \in \mathfrak{a}$. It lies in the given class, and its norm is $|\alpha\alpha'|/N(\mathfrak{a}) \le \tfrac{1}{2}\sqrt{\mathfrak{d}}$.

Counting: for $j \ge 1$, the number of integral ideals of F of norm j is at most j. An ideal of norm j is in one way a product, over the primes p dividing j, of ideals whose norms are the powers of p in j. So let $j = p^k$. If p splits, $(p) = \mathfrak{p}\mathfrak{p}'$, the ideals of norm p^k are the $k+1$ ideals $\mathfrak{p}^{k_1}\mathfrak{p}'^{k_2}$ with $k_1 + k_2 = k$. If p is inert, there is one such ideal for even k and none for odd k. If p ramifies, $(p) = \mathfrak{p}^2$, there is one, $\mathfrak{p}^k$. In each case the number is at most $k+1$, the number of divisors of p^k. Hence the number of ideals of norm j is at most the number of positive divisors of j, which is at most j.

Conclusion. Every ideal class contains an integral ideal of norm $j \le \tfrac{1}{2}\sqrt{\mathfrak{d}} \le \sqrt{\ell}$. The sum of the integers j with $1 \le j \le \sqrt{\ell}$ is at most $\tfrac{1}{2}\sqrt{\ell}\,(\sqrt{\ell}+1)$. So

$$h \;\le\; \sum_{1 \le j \le \sqrt{\ell}} j \;\le\; \frac{\sqrt{\ell}\,(\sqrt{\ell}+1)}{2} = \frac{\ell + \sqrt{\ell}}{2} \;<\; \ell, \tag{III.266}$$

because $\sqrt{\ell} < \ell$ for $\ell > 1$. This holds for every prime ℓ, including 2 and 3. □

Numerical check. The test group G12 computed the class numbers of $\mathbb{Q}(\sqrt{\ell})$ for all 17 984 primes $\ell < 2 \cdot 10^5$ by cycles of reduced forms (script `t04_classno.py` in `zeta-paper_tests/displays/G12`): $h < \sqrt{\ell}$ in every case. The test group G19 reproduced this with PARI (the largest class number is 153, at $\ell = 176401$; PARI's values were not certified), and for the 9592 primes $\ell < 10^5$ it checked every inequality of the proof.

Corollaries. (a) *Let $e \ge 1$ be odd. If ℓ does not divide the numerator of $B_{e+1}/(e+1)$, then* $\mathrm{Sha}_\ell(\mathbb{Q}, e) = 0$. (b) *Let $e \ge 2$ be even and $(\ell - 1) \mid 2e$. Then* $\mathrm{Sha}_\ell(\mathbb{Q}, e) = 0$. (c) *Let $e \ge 2$ be even. If ℓ does not divide the class number of $\mathbb{Q}(\zeta_\ell + \zeta_\ell^{-1})$, then* $\mathrm{Sha}_\ell(\mathbb{Q}, e) = 0$.

Proof. We use this fact: if $F \subset \mathbb{Q}(\zeta_\ell)$ is a subfield, then the ℓ-part of the class group of F is the part of A fixed by $\mathrm{Gal}(\mathbb{Q}(\zeta_\ell)/F)$. Indeed, write $L = \mathbb{Q}(\zeta_\ell)$ in this proof, let $j\colon \mathrm{Cl}(F) \to \mathrm{Cl}(L)$ be the map that extends ideals, and $N\colon \mathrm{Cl}(L) \to \mathrm{Cl}(F)$ the norm. Then

$$N \circ j = [L : F] \;\text{ on } \mathrm{Cl}(F), \qquad j \circ N = \sum_{\sigma \in \mathrm{Gal}(L/F)} \sigma \;\text{ on } \mathrm{Cl}(L). \tag{III.267}$$

The degree $[L : F]$ is prime to ℓ. By the first identity j is injective on the ℓ-part. By the second, a class z of A fixed by $\mathrm{Gal}(L/F)$ satisfies $j(N(z)) = [L : F]\,z$, so it lies in the image of j; and the image of j is fixed by $\mathrm{Gal}(L/F)$. (a) Here $i \equiv -e$ is odd. If $(\ell - 1) \mid (e+1)$ then $i \equiv 1$ and $A^{(1)} = 0$ [Was82, Prop. 6.16]. Otherwise take $3 \le i \le \ell - 2$. By Herbrand's theorem [Was82, Thm. 6.17], $A^{(i)} = 0$ unless $\ell \mid B_{\ell - i}$; and $B_{\ell-i}/(\ell - i) \equiv B_{e+1}/(e+1) \pmod{\ell}$ by Kummer's congruence [Was82,

Cor. 5.14], since $\ell - i \equiv e + 1 \not\equiv 0 \pmod{\ell - 1}$. (b) If $(\ell - 1) \mid e$ then $i \equiv 0$ and $A^{(0)}$ is the ℓ-part of the class group of $\mathbb{Q}$, by the fact above. Otherwise $(\ell - 1)/2$ divides e and is even, so $\ell \equiv 1 \pmod 4$ and $i \equiv (\ell - 1)/2$. The part of A fixed by the squares in Π is $A^{(0)} \oplus A^{((\ell-1)/2)}$, and the field fixed by the squares is $\mathbb{Q}(\sqrt{\ell})$. So $A^{((\ell-1)/2)}$ is the ℓ-part of the class group of $\mathbb{Q}(\sqrt{\ell})$. It is 0, because the class number is less than ℓ (Lemma G4). (c) For even e, $A^{(i)}$ lies in the part of A fixed by complex conjugation, which is the ℓ-part of the class group of the real subfield. □

Kummer's congruence, as used in (a), was also tested numerically by the test group G12: $B_r/r \equiv B_{r'}/r' \pmod{\ell}$ for the 2543 triples (ℓ, r, r') with ℓ prime, $5 \le \ell < 200$, r and r' even, $2 \le r \le \ell - 3$, $r < r' \le 220$ and $r \equiv r' \pmod{\ell - 1}$; in 14 of them ℓ divides B_r (script `t03_bernoulli.py`, part G, in `zeta-paper_tests/displays/G12`). The test group G19 dropped the condition $r \le \ell - 3$ and required $r \not\equiv 0 \pmod{\ell - 1}$ instead: the congruence holds for all 9738 such triples.

What is proved here, and what the Lean states. Everything proved in this subsection is about the groups $\mathrm{Sha}_\ell(E, e)$ of (III.251). Two identifications with the Lean are assumed and not proved. The first is the transcription of §1: that `Sha` and `shaLen` of the Lean are the group and the length of (III.251). It was compared at the level of the definitions (the same set S_ℓ, the same twist χ_ℓ^{-e}, discrete coefficients $\mathbb{Q}_\ell/\mathbb{Z}_\ell$, continuous cocycles, the kernel of the restrictions to chosen decomposition groups; `shaLen_eq_shaInv` removes the choices). Under it, the vanishing of $\mathrm{Sha}_5(\mathbb{Q}(i), 4)$ proved below says that `shaLen` is 0 for $\mathbb{Q}(i)$, $\ell = 5$, $e = 4$. The second concerns `PropZ`, which says: for $\ell \ge 5$ and even $e \ge 2$ with $(\ell - 1) \mid 2e$, the number `Chr.V` of the trivial character is 0. Corollary (b) is, for $\ell \ge 5$, this statement, if the group `Chr.V` of the Lean is identified with $\mathrm{Sha}_\ell(\mathbb{Q}, e)$ with enlarged coefficients. At the level of the definitions this is so: `Chr.V` is the length of the group of classes that vanish on the decomposition groups above ℓ and on all inertia groups (`ShaFamMono.lean`:30–32). The cohomology of the Lean is continuous cohomology: `H1U` is defined by restricting the representation to the subgroup (`RestrictionMemInvariantH1.lean`:35), and the group `H1` is the quotient of the continuous cocycles by the coboundaries (`SeparatedDvrSecondCohomologyMem.lean`:42–66). A full identification of the coefficients and of the local conditions of the Lean with those of (III.251) was not checked. So Corollary (b) is a proof of the classical statement, and it is a proof of `PropZ` only under this identification. The source proves `PropZ` independently (`X.propZ`). Corollary (a) is proved from Herbrand's theorem and does not use `PsiBound`. As said in §4, `PsiBound` with $\psi = 1$ needs a field with complex multiplication, so it gives a statement over $\mathbb{Q}$ only after a descent. The precise length over $\mathbb{Q}$, $v_\ell(B_{e+1}/(e+1))$ for $(\ell - 1) \nmid (e + 1)$, is the theorem of Mazur and Wiles ([MW84], from memory).

The field $\mathbb{Q}(i)$. For odd ℓ, $\mathrm{Sha}_\ell(\mathbb{Q}(i), e)$ is the sum of the part on which complex conjugation $\jmath$ acts as $+1$ and the part on which it acts as -1. The first is $\mathrm{Sha}_\ell(\mathbb{Q}, e)$, and the second is the part of the character χ_4 modulo 4 (restriction and corestriction; not proved here). For odd e the first is the minus part and the second the plus part; for even e it is the other way round. The minus part is governed by B_{e+1} (e odd) and by B_{e+1,χ_4} (e even), where $B_{r,\chi} = f_\chi^{r-1} \sum_{\alpha=1}^{f_\chi} \chi(\alpha) B_r(\alpha/f_\chi)$, f_χ the conductor and $B_r(X)$ the Bernoulli polynomial (Lean: `genB`). With this range of summation $B_{1,1} = +\frac{1}{2}$, as in the Lean. Exact computation gives the odd primes dividing the numerator of $B_{e+1,\chi}/(e+1)$:

e	χ	odd prime factors	e	χ	odd prime factors
1, 3, 5, 7, 9	1	none	2	χ_4	none
11	1	691	4	χ_4	5
13	1	none	6	χ_4	61
15	1	3617	8	χ_4	$5 \cdot 277$
17	1	43867	10	χ_4	$19 \cdot 2659$
19	1	$283 \cdot 617$	12	χ_4	$5 \cdot 13 \cdot 43 \cdot 967$
21	1	$131 \cdot 593$	14	χ_4	$47 \cdot 4241723$
23	1	$103 \cdot 2294797$			
25	1	657931			
27	1	$9349 \cdot 362903$			
29	1	$1721 \cdot 1001259881$			
31	1	$37 \cdot 683 \cdot 305065927$			

By Corollary (a), $\mathrm{Sha}_\ell(\mathbb{Q}, e) = 0$ for every odd ℓ when $e \in \{1, 3, 5, 7, 9, 13\}$, and for every odd $\ell \neq 691$ when $e = 11$. Grant the description of the two parts given above. Then for $E = \mathbb{Q}(i)$ and $t = 10$ the minus part of the sum in (III.252) has no contribution from odd e; for even $e \leq 8$ and odd ℓ the statement `PsiBound` of the source bounds the contribution of ℓ by $v_\ell(4 \prod_{p \leq e+2} p \cdot B_{e+1,\chi_4})$.

A prime that divides the Bernoulli number while the group is zero. Take $e = 4$ and $\ell = 5$. The prime 5 divides the numerator of $B_{5,\chi_4}/5 = -\frac{5}{2}$. The reason is an Euler factor. Let $L_5(s, \chi_4\omega)$ be the 5-adic L-function. For $r \equiv 1 \pmod 4$ its value at $1-r$ is $-(1-\chi_4(5)5^{r-1})B_{r,\chi_4}/r$ [Was82, Thm. 5.11]. At $r = 1$ this is 0, because $\chi_4(5) = 1$. All its values at integers are congruent modulo 5 [Was82, Cor. 5.13]. So 5 divides the numerator of $B_{e+1,\chi_4}/(e+1)$ for every $e \equiv 0 \pmod 4$, $e \geq 4$. Consistently, the computation finds the factor 5 for every $e \equiv 0 \pmod 4$ with $e \leq 24$ and for no $e \equiv 2 \pmod 4$ in that range.

The group $\mathrm{Sha}_5(\mathbb{Q}(i), 4)$ is nevertheless 0; in particular its part on which $\jmath$ acts as -1 is 0. We prove this. The proof uses local and global class field theory in their standard form and one computation modulo 25. It is for this field and this twist only. Put $E = \mathbb{Q}(i)$, $G = G_{E,5}$, $W = W_4 = \mathbb{Q}_5/\mathbb{Z}_5(-4)$ and $\chi = \chi_5$. For a subgroup H write $\mathrm{Hom}(H, \mathbb{F}_5)$ for the continuous homomorphisms.

Step 1: the invariants. The prime 5 splits in E. So $S_5 = \{v, v'\}$ and $E_v = E_{v'} = \mathbb{Q}_5$. The character χ maps G onto $\mathbb{Z}_5^\times$, because E is ramified only at 2 and every subfield $\neq \mathbb{Q}$ of the field of 5-power roots of unity is ramified at 5. It maps each of G_v, $G_{v'}$ onto $\mathbb{Z}_5^\times$, because the cyclotomic character of $\mathbb{Q}_5$ is onto. An element $w \in W$ is fixed if and only if $(u^{-4} - 1)w = 0$ for all $u \in \mathbb{Z}_5^\times$. Now $u^4 \equiv 1 \pmod 5$ for all u, and $2^4 - 1 = 15$ is not divisible by 25. So, for $H = G$, G_v or $G_{v'}$,

$$H^0(H, W) = W[5], \tag{III.268}$$

the subgroup of order 5, and H acts trivially on it. We identify $W[5]$ with $\mathbb{F}_5$ by $\frac{1}{5} \mapsto 1$.

Step 2: the connecting homomorphism. The sequence $0 \to W[5] \to W \xrightarrow{5} W \to 0$ gives, for the same H, the exact sequence

$$H^0(H, W) \xrightarrow{5} H^0(H, W) \xrightarrow{\delta} \mathrm{Hom}(H, \mathbb{F}_5) \longrightarrow H^1(H, W). \tag{III.269}$$

The first map is 0 by Step 1. So the kernel of $\mathrm{Hom}(H, \mathbb{F}_5) \to H^1(H, W)$ is $\delta(W[5])$, a group of order 5 generated by $c = \delta(\frac{1}{5})$. To compute c, lift $\frac{1}{5}$ to $\frac{1}{25} \in W$: $c(g)$ is the class of $g \cdot \frac{1}{25} - \frac{1}{25} = (\chi(g)^{-4} - 1)/25$, that is,

$$c(g) = \frac{\chi(g)^{-4} - 1}{5} \bmod 5. \tag{III.270}$$

The same formula holds on G and on G_v, $G_{v'}$, so the restriction of c is the generator there. For $u \in \mathbb{Z}_5^\times$ put $\lambda(u) = (u^4 - 1)/5 \bmod 5$. It depends only on u modulo 25. It is a homomorphism $\mathbb{Z}_5^\times \to \mathbb{F}_5$, because $(1 + 5x)(1 + 5y) \equiv 1 + 5(x + y) \pmod{25}$. And $c(g) = -\lambda(\chi(g))$, because $u^{-4} - 1 = -(u^4 - 1)u^{-4}$ and $u^{-4} \equiv 1 \pmod 5$.

Step 3: a class killed by 5. Let $z \in \mathrm{Sha}_5(E, 4)$ with $5z = 0$. By the sequence of Step 2 continued one term, z is the image of some $\phi \in \mathrm{Hom}(G, \mathbb{F}_5)$. The restriction of z to G_v is 0, so the restriction of ϕ to G_v lies in the kernel described in Step 2. The same holds at v'. So there are $a_1, a_2 \in \mathbb{F}_5$ with

$$\phi|_{G_v} = a_1\, c|_{G_v}, \qquad \phi|_{G_{v'}} = a_2\, c|_{G_{v'}}. \tag{III.271}$$

Step 4: reciprocity gives $a_1 = a_2$. Let rec be the reciprocity map of local class field theory for $\mathbb{Q}_5$, from $\mathbb{Q}_5^\times$ to the Galois group of the largest abelian extension. In its action on the roots of unity of 5-power order, rec(5) is trivial, and rec(u) acts as u^ϵ for $u \in \mathbb{Z}_5^\times$, where $\epsilon = \pm 1$ depends only on the normalisation of rec. So at each of the two places, for $u \in \mathbb{Z}_5^\times$ and an integer k,

$$c(\mathrm{rec}(5^k u)) = -\lambda(u^\epsilon) = -\epsilon\, \lambda(u). \tag{III.272}$$

Global class field theory says: the homomorphism ϕ factors through the largest abelian quotient of G, and for every $\beta \in E^\times$ the sum over all places w of E of the values of ϕ on the local reciprocity image of β at w is 0. Take $\beta = 2 + i$, of norm 5. At a finite place $w \nmid 5$, β is a unit and ϕ is unramified, so the term is 0. At the infinite place the term is 0, because $\mathbb{C}^\times$ is connected. There remain v and v'. They correspond to the two square roots of -1 in $\mathbb{Z}_5$. Let v be the place at which $i \equiv 7 \pmod{25}$; indeed $7^2 = 49 \equiv -1$. At v, $\beta \equiv 9 \pmod{25}$ is a unit, and

$$9^4 = 6561 \equiv 11 \pmod{25}, \qquad \lambda(9) = 2. \tag{III.273}$$

At v', $i \equiv -7$ and $\beta \equiv -5 \pmod{25}$ is not a unit. Write $\beta = 5/\bar\beta$ with $\bar\beta = 2 - i$. At v', $\bar\beta \equiv 9 \pmod{25}$, so $\beta = 5u$ with $\lambda(u) = -\lambda(9) = -2$. The relation is

$$0 = a_1 \cdot (-\epsilon) \cdot 2 + a_2 \cdot (-\epsilon) \cdot (-2) = -2\epsilon\,(a_1 - a_2) \quad \text{in } \mathbb{F}_5. \tag{III.274}$$

Hence $a_1 = a_2$. (The unit i gives no condition, as it must: $\lambda(7) = 0$, because $7^4 \equiv 1 \pmod{25}$.)

Step 5: $z = 0$. The homomorphism $\phi - a_1 c \colon G \to \mathbb{F}_5$ is trivial on G_v and on $G_{v'}$. So it is unramified at the primes above 5. It is unramified at all other primes, by the definition of G. The field E has no real place. So the extension of E cut out by $\phi - a_1 c$ is abelian and unramified everywhere. The class number of $\mathbb{Q}(i)$ is 1, so this extension is E and $\phi = a_1 c$. By Step 2 the image of c in $H^1(G, W)$ is 0. Hence $z = 0$.

Step 6: the whole group. A continuous cocycle $G \to W$ has finite image, because G is compact and W is discrete. So its values lie in the subgroup killed by some power of 5, and that power kills its class. If $z \neq 0$ were a class in $\mathrm{Sha}_5(E, 4)$ of order 5^k, then $5^{k-1}z$ would be a nonzero class in $\mathrm{Sha}_5(E, 4)$ killed by 5. Step 5 excludes this. So $\mathrm{Sha}_5(\mathbb{Q}(i), 4) = 0$. □

The congruences of Step 4 were recomputed by a script; the element $1 + 2i$ in place of $2 + i$ gives the same relation. The result agrees with Corollary (b): if the description of the two parts given above is granted, the part with $\jmath = +1$ is $\mathrm{Sha}_5(\mathbb{Q}, 4)$, which is 0 because $(5 - 1) \mid 2 \cdot 4$.

Why the Bernoulli number is divisible by 5 *and the group is* 0 (from memory; not proved here and not used). By the duality of Poitou and Tate, $\mathrm{Sha}_5(E, 4)$ is dual to the kernel of $H^2(G, \mathbb{Z}_5(5)) \to \prod_{v|5} H^2(G_v, \mathbb{Z}_5(5))$, and by Tate's local duality and Step 1 each local group has order 5. A consequence of the main conjecture [MW84] should give the order $5^{\,v_5(B_{5,\chi_4}/5)} = 5$ for

the part of $H^2(G, \mathbb{Z}_5(5))$ with $\jmath = -1$. So the Bernoulli number would measure a global group that maps injectively to the local groups. We have not compared this specialisation of the main conjecture with [MW84]. The proof above does not use it. The class number of $\mathbb{Q}(\zeta_{20})$, which is 1 (from memory), does not give the result by the method of Proposition G3: two primes of $\mathbb{Q}(\zeta_{20})$ lie above 5 and both ramify in the cyclotomic $\mathbb{Z}_5$-tower, so [Was82, Prop. 13.22] does not apply.

Three tests of the normalisations of the source, for the characters of conductor 1, 4, 8 and $r \le 42$: $B_{r,\chi} \neq 0$ exactly when $\chi(-1) = (-1)^r$ (except $\chi = 1$, $r = 1$, where $B_{1,1} = \frac{1}{2}$ in the normalisation above); the number $N \prod_{p \le r+1} p \cdot B_{r,\chi}$ is an integer for $N \in \{f_\chi, 8\}$ (and $N\,B_{r,\chi}$ alone is not, in 42 cases); and its absolute value is at most $16N^{r+1}r!$ (the largest ratio found is $1/32$). For $1 \le e \le 40$ the sum $\sum_\ell v_\ell \log \ell$ over the odd primes in the numerators above is at most $2\log\left(16 \cdot 4^{e+2}(e+1)!\right)$, the bound that `PrimeBound` gives for $\mathbb{Q}(i)$.

Sizes. For the field of Theorem G the bound for the minus part in the table is $3 \cdot 4n \cdot t^2 \log(tN)$ with $n \le 2x^{2+8\varepsilon}$ and $t \le x^{401}$: a power of x, about x^{804}. This upper bound is below $\sqrt{b} = \exp(x^\varepsilon/2)$ only for very large x: for $\varepsilon = 1/53$, from about $\log x = 742$ on (script). The crossing illustrates a sufficient scale, obtained from an upper bound. It gives no lower bound on the scale at which the actual cohomology satisfies Theorem G.

6. Assumed here, from the source

(i) *Theorem G.* The statement `VWin` (`ExistsPageN.lean`:49). It is proved by `vwin_of_row` from `RowOb`, and `RowOb` by `galoisLocalGlobalStrong_recParameter`. Both are among the declarations the final theorem depends on, and `vwin_of_row` depends only on the three standard axioms.

(ii) *The inequality* (III.257). The statement `LinkN`. Its proof in the source is in the theorems `final_alt_alt` and `linkN_of_twGood_galoisPreconditions`. Only the statements were read. The proof names bounds at the primes 2 and 3 (`boundAtTwoThree_linked_of_recT`) and a family of "Galois preconditions".

(iii) *The fact that $d_j = D_0$* (`SmallLevelRankEq`). The meaning of $\eta_j(w)$ is *not* in this list of things taken from the source, because it is not established: the lattice definition (`etaW`, `locLen`, `eta_units`) is for the number of a family, and the declarations `TieEta` and `etaN_eq_etaJ`, cited earlier for the tie with the field `etaJ` of the record, are not among those the final theorem depends on (§2). The identification of $\eta_j(w)$ with the defect of the index of lattices of Part II, §6.8 is open. Nothing proved here uses it.

(iv) *The identification of the data.* That the numbers `shaLen` in the record of the construction are the lengths $h_{\ell,e}$ of the field E (`nd.sha_eq`); and the passage from Theorem G to `HypV` described in §1: the change of scale (`winPairs`), the bound for the number of twists (`RankVal` and the theorem `eventually_rank_le_rpow_of_lt_one_div_fiftytwo`) and `supplyOfAdmOf`. The statements and the proof of `winPairs` were read; the other proofs were not.

(v) *Every row of the table of §4,* the hypotheses of the row for Case I, and the inputs listed after it.

What this section proves, and what it assumes. *Proved here outright:* Lemma G1 and Proposition G2. *Proved here from theorems of [Was82], checked in the first edition* (Propositions 6.16 and 13.22, Lemma 13.16, Theorem 6.17, Corollary 5.14): Proposition G3 and the corollaries (a), (b), (c); corollary (b) also uses Lemma G4. Proposition 13.22 does not state the compatibility with II; the proof of Proposition G3 supplies it. The divisibility of $B_{e+1,\chi_4}/(e+1)$ by 5 for $e \equiv 0 \pmod 4$

follows from Theorem 5.11 and Corollary 5.13 of the same book. *Proved here from Minkowski's lattice point theorem* ([Neu99], Ch. I, (4.4) with its strict inequality, and (5.2); checked in a copy): Lemma G4. *Proved here from the standard statements of local and global class field theory* (the reciprocity law for principal ideles, the action of the local reciprocity map of $\mathbb{Q}_5$ on roots of unity, and the triviality of the Hilbert class field of $\mathbb{Q}(i)$), for this single field and twist: $\mathrm{Sha}_5(\mathbb{Q}(i), 4) = 0$. *Not proved here:* the description of the two parts of $\mathrm{Sha}_\ell(\mathbb{Q}(i), e)$ by characters; the remark on the order of $H^2(G_{E,5}, \mathbb{Z}_5(5))$ after the proof for $\mathbb{Q}(i)$; the two identifications with the Lean stated in §5; and the meaning of $\eta_j(w)$ as a length of a quotient of lattices (§2). *Computed exactly:* the table of §5 and the three tests. *Taken from the source:* Theorem G, (III.257), items (i)–(v). *Cited as published theorems:* the correspondences in the last column of the table of §4, and the theorems of Soulé, Voevodsky and Rost, Mazur and Wiles, and Kurihara quoted in §3. They are cited from memory, except where a theorem or a section is named; a named place in an article was compared by a reviewer, not by us (see "Citations" at the head of the section).

What is in the literature and what is not. We sort the statements of the source named in this section into three lists. The source proves all of them again.

(A) Published, and cited with the place. `FerreroWashington`: [FW79] (from memory), in the form with bounded rank by [Was82, Prop. 13.23] (read). The duality theorems in `LitCoh`: Poitou and Tate [NSW00, §8.6]; for Tate's local theorems, the formula of Greenberg and Wiles and the local theorems of Bloch and Kato the attribution is from memory. `BhattScholzeIntegral`: [BS23, Thms. 1.2 and 5.6] (compared by a reviewer). We did not compare the Lean statements with the published ones word by word. `PropZ` does not belong to this list: it is not cited from a publication. Its classical form is Corollary (b), proved in §5 from class numbers; that Corollary (b) is `PropZ` depends on the identification of `Chr.V` stated in §5, which was not checked in full.

(B) A published counterpart exists, and it covers less or has not been matched. `PsiBound`: an upper bound for a cohomology length in terms of the valuation of a Bernoulli number with explicit integral factors. It corresponds to a consequence of [MW84] (from memory); general bounds from Euler systems are in [Rub00, §2.2], and the specialisation is not supplied. `RowOb`: a continuous semisimple representation, with predicted weights and the local comparisons in `PropsB`; compare [CH13], [Car12], [Car14, Thm. 1.1], with normalisations and hypotheses still to be matched. `KatoTwoW`: the two rational Selmer spaces of Bloch and Kato vanish under the attachment, `Used` and `WitAQ` hypotheses; Kato [Kat04, §14.1, Thm. 14.2(1)] proves finiteness of the discrete Selmer groups at the critical twists other than the central one. `LemmaK`: Soulé gives finiteness ([Sou79], from memory) and a quantitative bound away from small primes ([Sou03, Thms. 2 and 3], compared by a reviewer). We have not identified a published estimate for `LemmaK` in its full small-prime range. `OneCharU_s`: the method of congruences is illustrated by [BC04, Thm. 1.1], which constructs a Selmer extension; the upper bound is in list (C).

(C) Appears to be new, or not identified in the literature. Theorem G itself (`VWin`); the bounds for the plus part, `CountI`, `OneCharU_s`, `PropGK`, `CountII`, `CountIII`, `CountV`, `SumIn`, `SumOut`; the uniform statement `ABound` with the window, `ThmBab`; `LinkN`; the statement `SymplecticFourImageDense` on GSp_4; and `LemmaK` in its range of small primes. Each of them is a theorem of the archive that the final theorem depends on. Three of these names are not explained in the body. `SumOut` says: for the odd primes $\ell \notin \mathfrak{B}$ together, the sum of the plus parts over E, weighted by $\log \ell$, for the twists $e < t$, is at most $(Nt)^C$, with the level N of the source. `PropGK` says: there is a nonzero integer, of logarithm bounded by a power of $N(e+1)$ times $\log p$ for an auxiliary prime p (and a factor that depends on the orbit), such that for every $\ell \neq p$ twice its ℓ-adic valuation bounds the sum of the normalised lengths over an orbit of characters. `CountV` is described after the table of §4. For `ThmBab` and `SumIn` the table of §4 names related work (Ferrero

and Washington, Iwasawa, Soulé); that work is not a counterpart of the statements. An upper bound for the plus part of the groups $\mathrm{Sha}_\ell(E, e)$, $e \geq 1$ (groups of twisted Galois cohomology; they are not ideal class groups), by a congruence number of automorphic forms, uniform in the field and the twist, is not a published theorem as far as we know. Every statement of novelty in this list is provisional: it is a judgment made without a search of the literature, and it was not tested.

What remains.

(i) To read the proof of one row of the table completely. The minus part (`PsiBound`) is the natural first choice: it concerns only abelian fields and Bernoulli numbers.

(ii) To state `OneCharU_s` with its hypotheses on ρ (`Adm3U_s`, `Adm2A`) in classical terms, to compare them with `Adm3` and `Adm2`, and to find how the nonvanishing integer of `PropGK` is produced.

(iii) To identify the published counterpart of the rows for GSp_4.

(iv) To establish the meaning of $\eta_j(w)$: to find in the proof of `LinkN` (`linkN_of_twGood_galoisPreconditions`) the statement that ties the field `etaJ` of the record to the lattice length `etaW` at the twisted primes, and to check that the final theorem depends on it. The declarations `TieEta` and `etaN_eq_etaJ` do not serve: the final theorem does not depend on them. Until then the identification of $\eta_j(w)$ with the defect of the index of lattices of Part II, §6.8 is not established.

(v) To check the constants in the passage from the three parts to $\sqrt{b}$ (`vwin_of_joint`), and to evaluate the constant c_ψ of §1 for the trivial character.

(vi) To prove the description of the parts of $\mathrm{Sha}_\ell(\mathbb{Q}(i), e)$ by characters, and to compare the remark on $H^2(G_{E,5}, \mathbb{Z}_5(5))$ in §5 with [MW84].

(vii) To state `PsiBound` over $\mathbb{Q}$ by descent, and to match the normalisations of `Row0b` with [CH13, Car12, Car14].

Computations

The numerical statements in this paper were computed with the following scripts (Python, with `mpmath` and `sympy`); each prints the numbers quoted.

`e1_running_example.py`	G, G_χ, $R_{q,d}$, the table for $q = 3$, $d = 23$, P and M (Part I, §I.1, I.5, I.6; Part II, §4.1–4.3, §5.2–5.3, §6.7)
`e1b_check_cells.py`	the eight cell constants, by a second method (Part II, §5.2)
`e2_lemniscate.py`, `e2b_size.py`	the numbers A_n: divisibility and size (Part I, §I.3; Part II, §6.3)
`e4_sigma.py`	the Weierstrass σ-function: Gaussian size without the divisibility (Part II, §6.5)
`e5_kronecker.py`	the two-variable function $\sigma(z+w)/(\sigma(z)\sigma(w))$ of the lemniscatic lattice: divisibility at inert primes and size of its mixed derivatives (an experiment; its numbers are not quoted in the text)
`e6_hecke_theta.py`	T_n, the period, the rational numbers r_k (Part I, §I.4.1)

References

[Alp26] L. Alpöge, *[Remark on a proof of the Riemann hypothesis]*, with a formal proof in Lean 4, 20 November 2026. https://alpo.ge/remark.pdf.

[Apo76] T. M. Apostol, *Introduction to Analytic Number Theory*, Springer, 1976.

[Bar61] V. Bargmann, On a Hilbert space of analytic functions and an associated integral transform, *Comm. Pure Appl. Math.* 14 (1961), 187–214.

[BC04] J. Bellaïche and G. Chenevier, Formes non tempérées pour $U(3)$ et conjectures de Bloch–Kato, *Ann. Sci. École Norm. Sup.* 37 (2004), 611–662.

[Beh46] F. A. Behrend, On sets of integers which contain no three terms in arithmetical progression, *Proc. Nat. Acad. Sci. U.S.A.* 32 (1946), 331–332.

[Bom65] E. Bombieri, On the large sieve, *Mathematika* 12 (1965), 201–225.

[Bom74] E. Bombieri, Counting points on curves over finite fields (d'après S. A. Stepanov), Séminaire Bourbaki, exp. 430 (1973), Lecture Notes in Math. 383, Springer, 1974.

[Bom87] E. Bombieri, *Le grand crible dans la théorie analytique des nombres*, 2nd ed., Astérisque 18, Soc. Math. France, 1987.

[Bos96] 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.

[BS23] B. Bhatt and P. Scholze, Prismatic F-crystals and crystalline Galois representations, *Camb. J. Math.* 11 (2023), 507–562.

[Car12] A. Caraiani, Local-global compatibility and the action of monodromy on nearby cycles, *Duke Math. J.* 161 (2012), 2311–2413.

[Car14] A. Caraiani, Monodromy and local-global compatibility for $l = p$, *Algebra Number Theory* 8 (2014), 1597–1646.

[CH13] G. Chenevier and M. Harris, Construction of automorphic Galois representations, II, *Camb. J. Math.* 1 (2013), 53–73.

[Che52] P. L. Tchebichef, Mémoire sur les nombres premiers, *J. Math. Pures Appl.* 17 (1852), 366–390.

[CS67] S. Chowla and A. Selberg, On Epstein's zeta-function, *J. reine angew. Math.* 227 (1967), 86–110.

[Dav00] H. Davenport, *Multiplicative Number Theory*, 3rd ed., revised by H. L. Montgomery, Springer, 2000.

[Dam70] R. M. Damerell, L-functions of elliptic curves with complex multiplication. I, *Acta Arith.* 17 (1970), 287–301.

[Deu41] M. Deuring, Die Typen der Multiplikatorenringe elliptischer Funktionenkörper, *Abh. Math. Sem. Hansischen Univ.* 14 (1941), 197–272.

[DH66] H. Davenport and H. Halberstam, The values of a trigonometrical polynomial at well spaced points, *Mathematika* 13 (1966), 91–96.

[Erd32] P. Erdős, Beweis eines Satzes von Tschebyschef, *Acta Litt. Sci. Szeged* 5 (1932), 194–198.

[FW79] B. Ferrero and L. C. Washington, The Iwasawa invariant μ_p vanishes for abelian number fields, *Ann. of Math.* 109 (1979), 377–395.

[Gal70] P. X. Gallagher, A large sieve density estimate near $\sigma = 1$, *Invent. Math.* 11 (1970), 329–339.

[GGP12] W. T. Gan, B. H. Gross and D. Prasad, Symplectic local root numbers, central critical L-values, and restriction problems in the representation theory of classical groups, *Astérisque* 346 (2012), 1–109.

[GHY01] W. T. Gan, J. P. Hanke and J.-K. Yu, On an exact mass formula of Shimura, *Duke Math. J.* 107 (2001), 103–133.

[Gra84] D. R. Grayson, Reduction theory using semistability, *Comment. Math. Helv.* 59 (1984), 600–634.

[HB95] D. R. Heath-Brown, A mean value estimate for real character sums, *Acta Arith.* 72 (1995), 235–275.

[Hon70] T. Honda, On the theory of commutative formal groups, *J. Math. Soc. Japan* 22 (1970), 213–246.

[How79] R. Howe, θ-series and invariant theory, in *Automorphic Forms, Representations and L-functions*, Proc. Sympos. Pure Math. 33, part 1, Amer. Math. Soc., 1979, 275–285.

[Hur99] A. Hurwitz, Über die Entwicklungskoeffizienten der lemniskatischen Funktionen, *Math. Ann.* 51 (1899), 196–226.

[Jac86] H. Jacquet, Sur un résultat de Waldspurger, *Ann. Sci. École Norm. Sup.* 19 (1986), 185–229.

[Kat04] K. Kato, p-adic Hodge theory and values of zeta functions of modular forms, *Astérisque* 295 (2004), 117–290.

[Kud84] S. S. Kudla, Seesaw dual reductive pairs, in *Automorphic Forms of Several Variables* (Katata, 1983), Progr. Math. 46, Birkhäuser, 1984.

[Kur92] M. Kurihara, Some remarks on conjectures about cyclotomic fields and K-groups of $\mathbb{Z}$, *Compositio Math.* 81 (1992), 223–236.

[Lan05] E. Landau, Über einen Satz von Tschebyschef, *Math. Ann.* 61 (1905), 527–550.

[Leg08] A.-M. Legendre, *Essai sur la théorie des nombres*, 2nd ed., Paris, 1808.

[LT65] J. Lubin and J. Tate, Formal complex multiplication in local fields, *Ann. of Math.* 81 (1965), 380–387.

[Mon71] H. L. Montgomery, *Topics in Multiplicative Number Theory*, Lecture Notes in Math. 227, Springer, 1971.

[MV07] H. L. Montgomery and R. C. Vaughan, *Multiplicative Number Theory I. Classical Theory*, Cambridge Univ. Press, 2007.

[MW84] B. Mazur and A. Wiles, Class fields of abelian extensions of $\mathbb{Q}$, *Invent. Math.* 76 (1984), 179–330.

[Neu99] J. Neukirch, *Algebraic Number Theory*, Grundlehren Math. Wiss. 322, Springer, 1999.

[NSW00] J. Neukirch, A. Schmidt and K. Wingberg, *Cohomology of Number Fields*, 1st ed., Grundlehren Math. Wiss. 323, Springer, 2000.

[Pag35] A. Page, On the number of primes in an arithmetic progression, *Proc. London Math. Soc.* 39 (1935), 116–141.

[Ral87] S. Rallis, *L-functions and the Oscillator Representation*, Lecture Notes in Math. 1245, Springer, 1987.

[Rub00] K. Rubin, *Euler Systems*, Ann. of Math. Studies 147, Princeton Univ. Press, 2000.

[Sie35] C. L. Siegel, Über die analytische Theorie der quadratischen Formen, *Ann. of Math.* 36 (1935), 527–606.

[Sou03] C. Soulé, A bound for the torsion in the K-theory of algebraic integers, *Doc. Math.* Extra Vol. Kato (2003), 761–788.

[Sou79] C. Soulé, K-théorie des anneaux d'entiers de corps de nombres et cohomologie étale, *Invent. Math.* 55 (1979), 251–295.

[Ste69] S. A. 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.

[Stu76] U. Stuhler, Eine Bemerkung zur Reduktionstheorie quadratischer Formen, *Arch. Math.* 27 (1976), 604–610.

[Tit86] E. C. Titchmarsh, *The Theory of the Riemann Zeta-function*, 2nd ed., revised by D. R. Heath-Brown, Oxford University Press, 1986.

[TZ21] J. Thorner and A. Zaman, Refinements to the prime number theorem for arithmetic progressions, *Math. Z.* 306 (2024), article 54; arXiv:2108.10878.

[Was82] L. C. Washington, *Introduction to Cyclotomic Fields*, Graduate Texts in Math. 83, Springer, 1982.

[Wei52] A. Weil, Sur les "formules explicites" de la théorie des nombres premiers, *Comm. Sém. Math. Univ. Lund*, tome supplémentaire (1952), 252–265.

[Wei64] A. Weil, Sur certains groupes d'opérateurs unitaires, *Acta Math.* 111 (1964), 143–211.

[Wei76] A. Weil, *Elliptic Functions according to Eisenstein and Kronecker*, Springer, 1976.

[Wey25] H. Weyl, Theorie der Darstellung kontinuierlicher halb-einfacher Gruppen durch lineare Transformationen. I, *Math. Z.* 23 (1925), 271–309.