hexagon:2611.01087v1

READING ALPÖGE'S PROOF OF THE RIEMANN HYPOTHESIS

TOKUHIRO ETO

Abstract. On 20 November 2026, Levent Alpöge released a Lean formalization of an announced proof of the Riemann hypothesis. This paper provides an illustrated introduction to its mathematical ideas for mathematicians outside number theory. We interpret the proof through Chebyshev's classical principle, organizing its main ideas around three mechanisms: divisibility, amplification and cost control. These are realized through height-two formal groups, a tower of theta lifts, and bounds for Tate–Shafarevich groups. Elementary examples and conceptual diagrams illustrate how these ingredients interact to achieve square-root accuracy in prime sums. We also present a conditional reduction checked in Lean and raise several questions concerning possible simplifications of the proof and applications beyond the Riemann hypothesis. This AI-assisted exposition aims at mathematical understanding rather than completeness or independent verification.

1. Introduction

1.1. **What this paper is.** This is an introduction for mathematicians outside number theory, written by readers of the proof and not by its author. Its aim is to explain why each theory that the proof uses is needed at the point where it enters, and how the pieces fit together. It does not reproduce the proof, and it gives motivation and structure priority over completeness. Three kinds of statement occur, and we keep them apart: classical results, cited to their authors and collected in §8; the results of the formalization, all due to its author; and our own exposition and reconstruction, marked as such. Our own contributions are of the last kind:

- a reconstruction of the proof around Chebyshev's principle and three mechanisms (divisibility, amplification, cost control), with elementary toy models (§2);
- an account of how these mechanisms interact, with the conceptual error model behind the length $m \approx \sqrt{x}$ of the tower (§§2.4, 6.4);
- a small separate Lean project that checks the conditional reduction of the Riemann hypothesis to six statements (Appendix A); it reuses parts of the development's code (§1.3);
- four numbered questions (Questions 1–4, in §§3.5, 5.3, 6.4, 7.2), with a smaller Question 0 on Landau's criterion, collected in §8.3, where a simplification or further work may be possible. They are questions, not claims; the proof or existing work may already answer some of them.

Readers new to the subject can start at §2 and return to §§1.2–1.4 later.

1.2. **What was released.** On 20 November 2026 L. Alpöge released a Lean 4 development that proves Mathlib's statement of the Riemann hypothesis, together with a four-page note [4]. The development also proves that Dirichlet L-functions do not vanish for $\operatorname{Re} s > 1/2$; "GRH" in the release means GRH for Dirichlet L-functions. It has about 17 million lines. The note says that the arguments were found and formalized by the author working with Claude, and that the author chose to present only the formalization. The note contains a timeline of the discovery; the author says

Date: November 25, 2026 (Version 1).

2020 *Mathematics Subject Classification.* Primary 11M26; Secondary 11-02, 11N05, 11M06, 14L05, 11G15, 11F27, 14G40, 11F70, 11F80, 11R34, 68V20.

Key words and phrases. Riemann hypothesis, Chebyshev's method, Landau's oscillation theorem, Lubin–Tate formal groups, theta correspondence, slope inequality, Arakelov geometry, Tate–Shafarevich groups, Vandiver's conjecture, residually reducible Galois representations, Lean, formalized mathematics.

it was provided by Claude and that, outside the author's expertise, it paraphrases a summary that "might make zero sense". We quote it with this caveat.

All mathematics described here is due to the author of the formalization. Our description is based on our reading of the released code, in the mirror `jsmorph/rh` at commit `ecd09ff`, whose README gives the SHA-256 of the release archive[1], and of the author's note. "New" always means "appears to be new": we did not find it in the literature we checked. Several other accounts of the proof have already appeared [2, 12, 24, 45–48, 55, 66].

1.3. **What was checked, and how we mark it.** We could not build the whole development, so the correctness of its deep parts rests on the author's build. A text search of the source (for `sorry`, `admit`, line-initial `axiom`, `native_decide` and similar escape hatches) finds none. This is consistent with every hypothesis used along the way being supplied by a proved statement of the development, provided the author's build succeeds; it is not a check of axiom dependencies and not a formal verification, and whether those statements are non-vacuous we have checked only where we say so. We use four labels.

- LEAN-CHECKED: proved in our separate, small Lean project, which derives the Riemann hypothesis from six statements that we formulated to match theorems of the development (Appendix A). The match was checked in Lean against a partial build of the original for (A), (T), (V), (G), and only by reading for (V1)–(V3); the class number $h(\mathbb{Q}(\sqrt{-7})) = 1$ used in (V3) is not evaluated in Lean. The project is not independent of the development: its proof of Landau's criterion is the development's, taken verbatim or instantiated for ζ (a few lemmas re-proved along the original lines), and most of the lemmas of the (V1)–(V3) $\Rightarrow$ (V) argument are ported. The proofs of the prime-side and supply layers were written afresh, but their interface definitions, the definitions of the remainders, and the lemma that a zero-free half-plane gives RH are the development's.
- CODE-CHECKED: the relevant place of the original code was inspected, by AI agents and AI reviewers, not by Lean. We give the depth: *statement* (the statement was read), *proof read* (the proof was read), *traced* (the dependency on the path to the main theorem was followed).
- RECONSTRUCTED: our mathematical reconstruction or interpretation, from the code and the documents.
- UNCHECKED: details or dependencies that we, or our agents, have not checked. This does not mean that they are mathematically open.

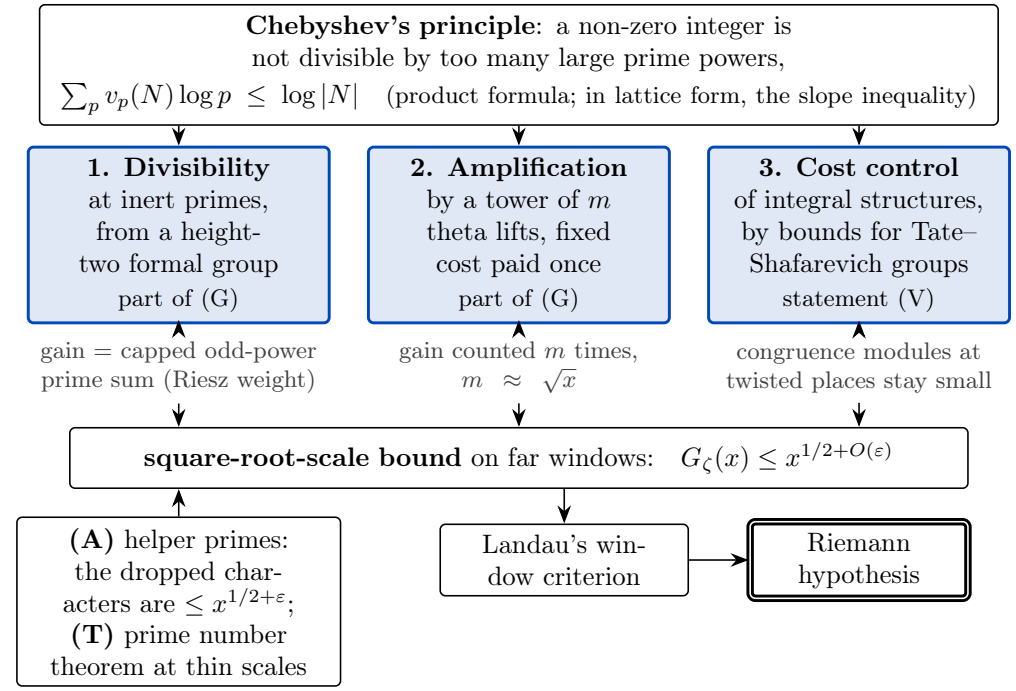
Each Act ends with short verification notes; a more detailed status table, with the depth of each code inspection, is available from the author on request. Inside the text we keep a hedge where a reader could otherwise take an interpretation for a fact. Where a statement rests on Adriazola's manuscript [2] rather than on our reading, we cite it. References were checked where marked in our records; not all have been checked against the originals (§9).

1.4. **The argument in one page.** Write Λ for the von Mangoldt function and

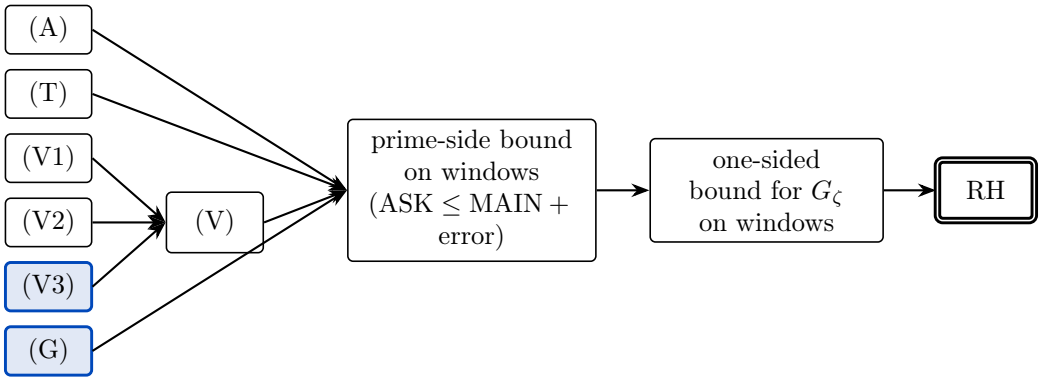
$$G_\zeta(y) = \sum_{n \le y} \Lambda(n)\Big(\frac{y}{n} - 1\Big) - y \log y + (1+\gamma)y$$

for the Riesz-weighted remainder in the prime number theorem. Let x denote the point at which $G_\zeta(x)$ is to be bounded, let $\varepsilon > 0$ be small and C a fixed constant. The proof has the following shape; the terms in it are explained in §2 and Act 0, and a first-time reader may go there directly

[1] `9631b4f020199e580ba86178cb56c6e9b09f690c2369a80dc4334aa93f49c7fc`; quoted from that README, not recomputed by us.



(a) Conceptual flow.



(b) Logical dependency: implications checked in our Lean project.

FIGURE 1. Architecture of the proof. (a) A conceptual diagram, our reading of how the ideas fit together, not a dependency graph. (b) The reduction checked in our Lean project: Mathlib's Riemann hypothesis follows from six statements formulated by us to match theorems of the development, which we have not built; the bridge to (A), (T), (V), (G) was checked against a partial build, those to (V1)–(V3) only read (Appendix A, §7.3). Blue boxes: where the parts that appear to be new lie.

and come back.

$$
\begin{aligned}
&\mathrm{RH} \Longleftarrow G_\zeta(x) \le x^\theta \text{ on far windows } [Y, Y^{3/2}], \text{ for every } \theta > \tfrac{1}{2} && (\S 3.1)\\
&\quad \Longleftarrow \textstyle\sum_{\text{tower}} (\mathrm{ASK} - \mathrm{MAIN}) \le x^{1/2+C\varepsilon} \cdot (\text{thick mass}) \text{ at each such } x && (\text{App. A})\\
&\quad \Longleftarrow (\mathrm{A}) + (\mathrm{T}) \ + \ (\mathrm{V}) \ + \ (\mathrm{G}).
\end{aligned} \tag{1}
$$

Here ASK is a non-negative sum of Riesz-weighted $\log p$ over odd powers of primes p inert in an auxiliary imaginary quadratic field, MAIN is its main term, the sum is over the m steps of a tower,

and the "thick mass" is the total weight of the classes in which the prime sums have length about x (§3.3). The four inputs:

- **(A), (T): analytic number theory.** Helper primes q, d for which the twisted remainders have square-root size on a range, and a prime number theorem with saving at very large scales. Both are classical in substance (§3.4).
- **(G): the geometric bound**, which is the inequality in the second line itself. It is proved by Chebyshev's principle in Arakelov form (Act 1), fed by divisibility of CM periods at inert primes (Act 2) and amplified by a tower of theta lifts (Act 3).
- **(V): a Galois bound.** The Tate–Shafarevich groups $Ш^1(E, \mathbb{Q}_\ell/\mathbb{Z}_\ell(-e'))$ of the auxiliary CM fields E, summed over all ℓ with weight $\log \ell$, are at most $\exp(x^\varepsilon/2)$ on suitable windows. The minus parts follow from known consequences of the main conjecture; for the plus parts the development gives a Vandiver-type argument that appears to be new (Act 4).

The link between divisibility and the prime sum is a dictionary (§§2.3 and 3.3): the capped divisibility produced by a height-two formal group at inert primes has the shape of the Riesz-weighted sum ASK, up to a logarithmic loss. The product formula then bounds the total gain by an archimedean size, computed from Fock-space norms and the Chowla–Selberg formula; that size is MAIN. Figure 1 shows the conceptual flow and, separately, the logical reduction that we checked in Lean.

1.5. **A reading guide.**

- *Readers from analysis, geometry or PDE:* §2; Landau's criterion (§3.1) and the simplified sum (§3.2); the openings of the Acts (§§4, 5, 6.1, 7.1); the Questions and §8.3. Sections 7.3–7.5 and the appendices can be skipped on a first reading. (Sections 3–7 are called Acts 0–4.)
- *For the big idea only:* §2, then the first pages of each Act and the Finale (§8).
- *Analytic number theorists:* §2, Act 0 (§3), the ledger of Act 3 (§§6.3–6.4) and Appendix A.
- *Formal groups and arithmetic geometry:* Acts 1 and 2 (§§4–5), in particular Theorem 5.3 and the size side.
- *Automorphic forms and Galois representations:* Act 3 (§6) for the theta tower, and Act 4 (§7) for the Galois bound, Lemma M and residually reducible automorphy.
- *Readers who want to check the Lean development:* the verification notes at the end of each Act; a status table and a companion file recording the code locations behind each claim are available from the author on request.

1.6. **Notation.** Several letters would naturally carry more than one meaning; we fix them as follows. Nobody needs to memorize this list on a first reading; each symbol is introduced again where it is used.

symbol	meaning
x; ρ	the point where $G_\zeta(x)$ is bounded; zeros of ζ. $\mathcal{L}g$ is a Laplace transform.
q, d	the helper primes, $q \equiv 3 \bmod 8$, $d \equiv 7 \bmod 16$.
κ, H_κ	the 2-power level, $H_\kappa = (\mathbb{Z}/2^{\kappa+2})^\times/\{\pm 1\}$, with $\lvert H_\kappa\rvert = 2^\kappa$.
E, $\mathcal{G}$, Φ	the host field, its Galois group, and a CM type of it.
K	$\mathbb{Q}(\sqrt{-d})$ in Acts 1–2 (a general CM field in the slope inequality); $\mathbb{Q}(\sqrt{-p_K})$ in the (V3) argument.
a, $\beta(\tau)$, b	a length (equivalently a total jet order); a weight (equivalently a cap) at $\tau \in \Phi$; the base value of the thick weights.
t, TT; u; j, m	top places $t \in TT \subset H_\kappa$; averaging translates $u \in H_\kappa$; steps $j < m$ of the tower.
X, $\ell(X)$, D	the formal-group coordinate at an inert prime p, the Lubin–Tate logarithm, and $d/d\ell(X)$.
e; e'; t_V	an exponent (odd in the gain); the Tate twist; the cutoff $\lfloor x^{401} \rfloor$ for e'.
$T_{\mathfrak{w}}$, g_0, g_1	the level prime of a word $\mathfrak{w}$ in a finite family $\mathfrak{W}$ (written T when $\mathfrak{w}$ is fixed); the two type primes.

symbol	meaning
$r_1 < r_2$, N_w	the auxiliary primes in the plus-part argument; norms of places above them.
k, ν, k_W	the weight budget; the odd rank step of the tower; the rank $k+m-1$ of the Shimura-side spaces.
γ_p; Ω	$p/(p^2-1) = \sum_{e \text{ odd}} p^{-e}$ (not Euler's γ); the CM-period constant of §5.3.
(A), (T), (V), (G)	the six statements of Appendix A, with (V1)–(V3); the letter T alone is the level prime.

2. The proof without the technical machinery

This section can be read on its own. It explains why each of the three mechanisms is needed, before any of them is built. Nothing in it is used later as a proof.

2.1. **Why the square root matters.** The Riemann hypothesis is equivalent to $\psi(x) - x = O(x^{1/2+\varepsilon})$ for every $\varepsilon > 0$. An error $O(x^\theta)$ with a fixed $\theta < 1$ would already exclude zeros with real part above θ. The exponent $1/2$ cannot be lowered, because ζ has zeros on the critical line; a proof must therefore reach the square-root scale.

The proof works with the Riesz-weighted remainder G_ζ above. The weight $(y/n-1)$ makes the explicit formula absolutely convergent, and under RH it gives $|G_\zeta(y)| \ll \sqrt{y}$ (§3.1), so the smoothing loses nothing. Two further simplifications are available. First, only an *upper* bound for G_ζ is needed: a classical argument of Landau shows that a one-sided bound already excludes zeros. (Heuristically, a real zero would push the remainder in one direction and a complex zero would make it oscillate with large positive values; the proof in §3.1 works differently.) Second, the bound is needed only on a sequence of far windows $[Y, Y^{3/2}]$, not at every x. Both simplifications matter later: upper bounds are what divisibility arguments give, and windows allow auxiliary choices that depend on x.

2.2. **What an integer knows about primes.** If a non-zero integer N is divisible by p^{v_p} for each prime p, then

$$\sum_p v_p \log p \le \log |N|.$$

This is Chebyshev's principle. It turns a lower bound for divisibility into an upper bound for a prime sum. The classical example is the factorial: by Legendre's formula $v_p(a!) = \sum_{e\ge 1} \lfloor a/p^e \rfloor$, so $\log a! = \sum_{n\le a} \Lambda(n)\lfloor a/n \rfloor$, and comparing $a!$ with binomial coefficients gives Chebyshev's bounds $\psi(x) \asymp x$.

There is an arithmetic variant that is closer to the proof. Let $K = \mathbb{Q}(\sqrt{-d})$. Replacing $a!$ by the ratio of $(a!)^2$ and Bhargava's factorial [8] for $\mathcal{O}_K$ removes exactly the contributions of the split primes and of the even powers of the inert primes (Remark 4.1):

$$\log \frac{(a!)^2}{N(a!_{\mathcal{O}_K})} = \sum_{n\le a} \Lambda(n)\big(1 - \chi_{-d}(n)\big)\left\lfloor \frac{a}{n} \right\rfloor.$$

The right side is a prime sum over odd powers of inert primes, and the left side is the logarithm of an integral object. But this does not reach the square root. The size of the left side is not known independently of the prime sum, and one application of the principle gives an error of size about x instead of $\sqrt{x}$: Chebyshev quality, not more. The proof has to find integral objects with the same divisibility whose archimedean size can be computed, and it has to use them many times.

2.3. **Manufacturing the right divisibility.** Take $a = 1000$ and $p = 5$. Legendre's formula gives $v_5(1000!) = 200+40+8+1 = 249$. The contributions of the even exponents $e = 2, 4$ are $40+1 = 41$; what remains, $200+8 = 208$, is the odd-exponent part $\sum_{e \text{ odd}} \lfloor a/p^e \rfloor$ (Figure 2). At a prime p inert in K the residue field has p^2 elements, and it absorbs the even exponents (Remark 4.1). The proof needs a natural source of divisibility by the odd part.

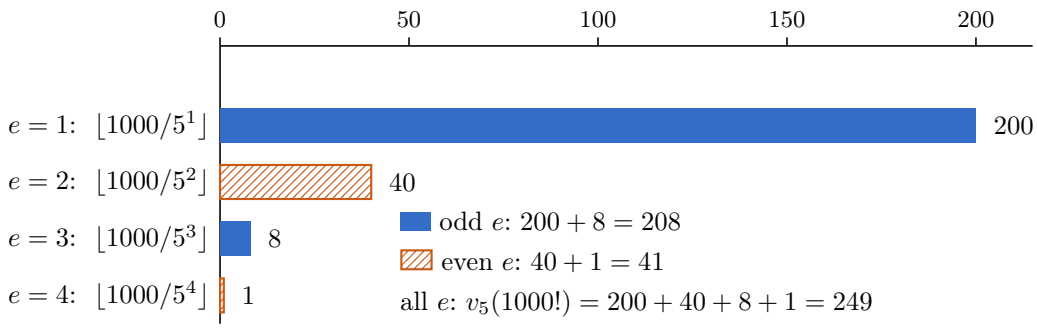


FIGURE 2. From factorials to the inert prime sum (toy model, $a = 1000$, $p = 5$). Legendre's formula counts $\lfloor a/p^e \rfloor$ for each e; for p inert, Bhargava's factorial of $\mathcal{O}_K$ removes the even powers, leaving the odd part 208 (Remark 4.1). For a height-two formal group this count is only a *lower bound* for valuations of derivatives of integral functions, and the proof uses the weaker $a\,p/(p^2-1) - \lg_p(a) = 203.3$ (Proposition 5.1).

A height-two formal group provides it. At a prime p inert in K, a CM elliptic curve has supersingular reduction, and its formal group has height two. Its logarithm $\ell(X) = \sum_j p^{-j} X^{p^{2j}}$ has one denominator p for each step $X^{p^{2j}}$, that is, one for each even power of p. Taking a invariant derivatives of an integral function therefore produces $a!$ divided by these denominators: the divisibility that survives is, roughly, the odd part of $a!$, about $a\gamma_p$ with $\gamma_p = p/(p^2-1) = \sum_{e\ \mathrm{odd}} p^{-e}$ (Proposition 5.1). For height one the rate would be $1/(p-1)$ and nothing would remain; this is why the primes must be inert. The sharp form, divisibility by exactly $p^{\sum_{e\ \mathrm{odd}} \lfloor a/p^e \rfloor}$, is known (Remark 5.2); the proof uses only a weaker real-valued bound, and it does not claim equality for every function.

The divisibility used is capped. Each odd shift e contributes not a/p^e but $(a/p^e - \beta)_+$, where the cap β is a weight attached to a position in the construction; a heuristic for where the cap comes from is in §5.2. Later a is very large, $a \approx x\,e^{x^\varepsilon}$, and the cap $\beta \approx e^{x^\varepsilon}$ brings the length a/β back to about x. Summing over inert primes and odd exponents gives

$$\sum_{p\ \mathrm{inert}} \sum_{e\ \mathrm{odd}} 2\log p \left(\frac{a}{p^e} - \beta\right)_+, \qquad \left(\frac{a}{n} - \beta\right)_+ = \beta\left(\frac{a/\beta}{n} - 1\right)_+.$$

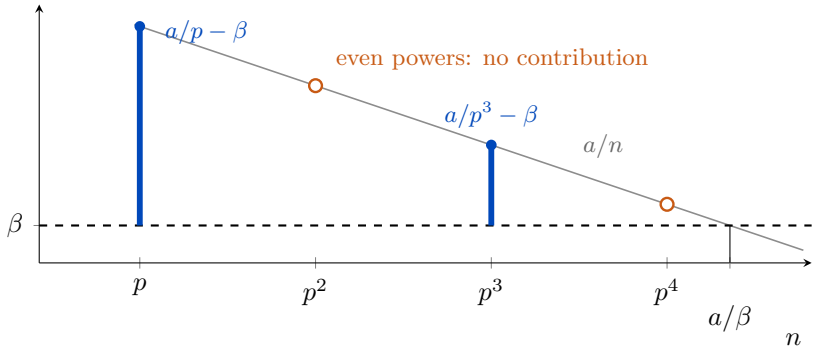
The identity on the right is the central dictionary (Figure 3): the capped Legendre count is exactly the Riesz weight, for a prime sum of length a/β. The cap has two effects: it makes the sum finite (without it, $\sum_p a\gamma_p \log p$ diverges), and it turns divisibility into the Riesz-weighted prime sum that Landau's criterion needs.

2.4. **Amplifying the gain.** A single application of Chebyshev's principle compares one prime sum with one archimedean size. Schematically, the balance identity of Act 0 gives

$$\mathrm{ASK} - \mathrm{MAIN} \ \approx\ \beta\,(G_\zeta - G_d)\left(\frac{a}{\beta}\right) \ +\ (\text{other characters}) \qquad (\text{schematic}),$$

so with $a \approx x\beta$ an additive error of size a in the comparison is an error of size x in G_ζ near $a/\beta \approx x$. In our reading of the ledger, the comparison carries a fixed cost of that size, a normalization that does not shrink with the sum. One application therefore gives an error of size x, the Chebyshev quality of §2.2. Nor does a single step with a larger a help: the fixed cost and the signal grow together, and their ratio stays of order x up to logarithms (§6.3).

The proof repeats the gain without repeating the fixed cost. It builds a tower of m steps, each step a theta lift. At each step the same divisibility, and hence the same prime sum, appears again. The fixed cost is attached to one choice made once, at the bottom of the tower (a "level prime"), and



$$\sum_{e \text{ odd}} \Big(\frac{a}{p^e}-\beta\Big)_+, \qquad \Big(\frac{a}{n}-\beta\Big)_+ = \beta\Big(\frac{a/\beta}{n}-1\Big)_+$$

FIGURE 3. From capped divisibility to the Riesz weight (schematic: inert $p = 3$, $a/\beta = 120$, logarithmic scales, so stem lengths are not to scale; the gray curve and the dashed cap are guides). Each odd power $p^e < a/\beta$ contributes its stem $(a/p^e-\beta)_+$ and even powers contribute nothing: a sum of discrete terms, not an area. By the identity, this is β times the Riesz weight of a prime sum of length a/β (§3.3).

is paid once. So the gain is counted m times and the fixed cost once: after dividing by the signal, which is now m times larger, the fixed cost contributes x/m instead of x.

The square root itself does not come from the tower. The remainders twisted by the helper characters can be made of square-root size: this is the classical input (A), which chooses helper primes for which they are at most $x^{1/2+\varepsilon}$ (in our reading by a mean-square argument over the conductor). G_ζ itself cannot be treated in this way. The geometry transfers the problem from G_ζ to these twisted remainders, at the cost x/m, and the tower's job is to bring x/m down to the level $\sqrt{x}$, which needs $m \gtrsim \sqrt{x}$. A weight budget of the tower also sits at that level. A longer tower gains nothing, since the target is already at $\sqrt{x}$ (see Question 3). A conceptual error model is

$$\frac{x}{m} + m + \sqrt{x} \qquad \text{(up to factors } x^{O(\varepsilon)}\text{)},$$

which is of size $\sqrt{x}$ for $m \approx \sqrt{x}$ (Figure 4). This is a conceptual model, not a proof: it omits density factors and constants, and parts of its derivation are unverified (§6.4).

2.5. **Paying for the amplification.** Building many divisible objects is not enough. The objects of the tower are lattices of automorphic forms, and the slope inequality of Act 1 charges for how far each lattice is from splitting integrally along its Hecke eigenspaces. This cost, measured by congruence modules, is not a priori small.

Congruence modules are classically bounded by Galois cohomology, through Selmer-type groups of the associated Galois representations. For the lifts in the tower these Selmer groups reduce, in our reading, to Tate–Shafarevich groups of twists of $\mathbb{Q}_\ell/\mathbb{Z}_\ell$ over the auxiliary CM fields. So the proof needs a bound for these Tate–Shafarevich groups. It asks for a bound of size $\exp(x^\varepsilon/2)$ on windows; how much of this is actually consumed is discussed in §7.2. The minus half follows from the main conjecture of Iwasawa theory; the plus half, a Vandiver-type bound polynomial in the conductor and uniform in ℓ up to a term from the primes 3 and 7, is the subject of Act 4.

3. Act 0: putting the problem in the right shape

This Act makes precise the target of §2.1: an upper bound, on far windows, for a non-negative inert prime sum. It first states Landau's criterion, then studies a simplified sum, and only then introduces the field and the averaging that the proof actually uses.

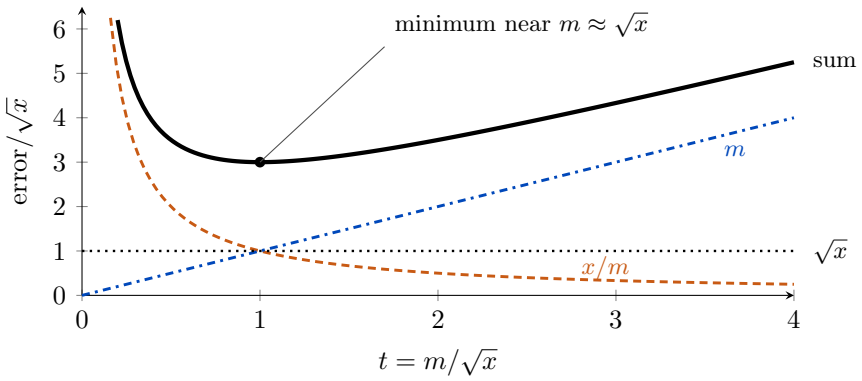


FIGURE 4. Why the square root appears (conceptual model, in units of $\sqrt{x}$ after division by the signal; $t = m/\sqrt{x}$). The level $\sqrt{x}$ comes from the dropped characters through (A); the tower brings the fixed cost x/m down to it. The rising branch comes only from a buffer term ($B_0 \propto m$ in the bounds as written) whose necessity we do not know, so the position of the minimum is not claimed to be optimal (§6.4). Density factors, constants and $x^{O(\varepsilon)}$ are omitted.

3.1. **Landau's criterion in window form.** The Riesz weight has Mellin kernel $1/(s(s-1))$:

$$\int_1^\infty G_\zeta(y)\, y^{-s-1}\, dy = -\frac{\zeta'(s)/\zeta(s) + 1/(s-1) + 1 - (1+\gamma)s}{s(s-1)} \qquad (\operatorname{Re} s > 1).$$

So the Laplace transform $\mathcal{L}g$ of $g(u) = G_\zeta(e^u)$ continues to $\operatorname{Re} s > 0$ with a simple pole at each zero ρ of ζ (the logarithmic derivative has simple poles whatever the multiplicity), with residue $-m_\rho/(\rho(\rho-1))$, m_ρ the multiplicity; for real $\rho \in (0,1)$ this residue is positive, and $\mathcal{L}g$ has no other real singularity in $(0,\infty)$.

Theorem 3.1 (window criterion; LEAN-CHECKED)**.** *Suppose that for every $\theta > 1/2$ and every Y_0 there is $Y \geq Y_0$ with $G_\zeta(y) \leq y^\theta$ for all $y \in [Y, Y^{3/2}]$. Then the Riemann hypothesis holds.*

This is Landau's classical argument for functions of one sign (see [42]), localized to windows. Write $T_k(s) = \int_0^\infty g(u)\, u^k e^{-su}\, du/k!$, the k-th Taylor coefficient of $\mathcal{L}g$ at s up to the sign $(-1)^k$. Fix a real σ, depending only on the window exponent, and take k so that the kernel $u^k e^{-\sigma u}/k!$ peaks inside a window $[U, (1+\delta)U]$, $U = \log Y$.

Upper bound. Since $|g| + g = 2g_+$ (integrate it against the kernel), the one-sided hypothesis bounds $|T_k(\sigma + i\gamma)| + T_k(\sigma)$, for every real γ, by $e^{o(U)}(\sigma - \theta)^{-k}$. The tails are bounded through the a priori bound $|G_\zeta(y)| \ll y^2$.

Lower bound. Suppose a zero has real part $> \theta$, and let $r \leq \sigma - \theta$ be the radius of a pole-free disc around σ, as large as possible up to $\sigma - \theta$. If some pole ρ has $\sigma - \operatorname{Re}\rho < r$, move to $z = \sigma + i\gamma$ with γ near $\operatorname{Im}\rho$, chosen so that a single pole ρ_1 is nearest to z, at distance $< r$. Then $|T_k(z)| \approx |c|\,|z - \rho_1|^{-k-1}$, while $T_k(\sigma)$ is smaller by Cauchy's estimate on the pole-free disc. Otherwise every pole has $\operatorname{Re}\rho \leq \sigma - r$ (then $r < \sigma - \theta$, so r is the distance to the nearest pole); the nearest pole is a single real point $\rho_0 = \sigma - r$, with $\rho_0 \geq \operatorname{Re}\rho$ for every pole, hence $\rho_0 > \theta$; its residue is positive, and $T_k(\sigma)$ is about $c\,(\sigma - \rho_0)^{-k-1}$ with $c > 0$. Either way the lower bound beats $(\sigma - \theta)^{-k}$, a contradiction as $k \to \infty$.

Arbitrarily far windows are needed only to make k arbitrarily large. No sign change of the moments is used: complex poles are handled by moving the expansion point off the real axis. The non-trivial inputs are the continuation with simple poles and positive real residues, and the identity $|g| + g = 2g_+$. The window exponent $3/2$ may be any number > 1, and the pointwise bound may be weakened to $\int_{\text{window}} (G_\zeta)_+(y)\, y^{-\theta-1}\, dy \leq Y^\eta$ for every $\eta > 0$. The development's Lean proof,

which our project reuses for ζ, contains every step, the tails included. We regard the window form as classical in substance; Rubinstein [47] gives a different proof, through an oscillation theorem of Kaczorowski and Pintz.

Question 0 (How weak can the window criterion be?). The Lean version (the development's, reused in our project) accepts windows $[Y, Y^{1+\delta}]$ for any $\delta > 0$ and an integrated one-sided bound in place of a pointwise one. Would a sparse set of points, or a weaker integrated bound, suffice?

Why this shape. The window bound is equivalent to RH: under RH the explicit formula $G_\zeta(y) = \log 2\pi - \sum_\rho y^\rho/(\rho(\rho-1)) + O(y^{-2})$ converges absolutely and $|\rho(\rho-1)| = |\rho|^2$, so $|G_\zeta(y)| \le (2+\gamma-\log 4\pi)\sqrt{y} + O(1) \approx 0.046\sqrt{y}$. Nothing is lost. This shape gives:

(i) only an *upper* bound is needed, which is what Chebyshev's principle (§2.2) gives;
(ii) only for a *non-negative* sum, so primes that cannot be controlled may be dropped;
(iii) only on sparse, far windows, so auxiliary primes can be chosen per window, and the Galois input is needed only there;
(iv) the *Riesz* weight: it gives absolute convergence, and it is the shape that capped divisibility produces (§3.3). With the floor weight the transform is $-\zeta'(s)/s$, the zeros disappear, and one is left with Chebyshev-type information: the ζ-part of the floor-weight sum is $\log\lfloor a\rfloor!$, known exactly by Stirling, so the floor-weight inert sum carries information only on $L(s, \chi_{-d})$, never on ζ (Remark 4.1).

3.2. **A simplified inert sum.** Before the full construction, consider, for $a > 0$ and a single weight $\beta > 0$,

$$S(a,\beta) = 2 \sum_{p \text{ inert in } \mathbb{Q}(\sqrt{-d})} \sum_{e \text{ odd} \ge 1} \log p \left(\frac{a}{p^e} - \beta\right)_+ .$$

- *Inert primes:* they are the primes at which the CM curve of Act 2 has supersingular reduction, so that the formal group has height two.
- *Odd exponents:* the height-two formal group absorbs the even powers of p (§2.3); what is left is the odd part.
- *The Riesz weight:* the cap turns the floor weight into the Riesz weight of Landau's criterion (§2.3).
- *The length a/β:* only prime powers $n = p^e < a/\beta$ contribute, so $S(a,\beta)$ is a prime sum of length a/β.

The sum S omits the ramified prime d and uses one weight for all positions. Writing $y = a/\beta$, one has $\sum_n \Lambda(n)(1-(\frac{n}{d}))(\frac{a}{n}-\beta)_+ = \beta\sum_{n\le y}\Lambda(n)(1-(\frac{n}{d}))(\frac{y}{n}-1)$, which is $S(a,\beta)$ plus the contribution of the powers of d, and the prime number theorem in Riesz form gives

$$S(a,\beta) + (\text{powers of } d) = a\big(\log a - 1 - \gamma + \ell_d - \log\beta\big) + \beta\,(G_\zeta - G_d)\Big(\frac{a}{\beta}\Big),$$

with G_d the remainder twisted by $(\cdot/d)$ and ℓ_d as in Definition 3.2 below. In the application $d > a/\beta$, so the powers of d do not occur, and an upper bound for S of square-root quality would bound $G_\zeta - G_d$ from above near a/β.

Three features of the real proof are absent from S. (1) The divisibility comes from many positions, embeddings τ in a CM type of a larger field, and each position has its own weight $\beta(\tau)$. (2) The prime sums must therefore be split according to the Frobenius class of n in that field. (3) The twisted remainders that appear must themselves be of square-root size; this needs averaging over the classes and auxiliary primes chosen for the purpose. This leads to the host field and to ASK.

3.3. **The host field and the inert prime sum.** In the geometry, $\tau \in \Phi$ will be an embedding in a CM type of the host field, a a total jet order, $\beta(\tau)$ the cap at τ, and the shift $\mathrm{sh}(t)$ defined below the Frobenius offset of a partner position (RECONSTRUCTED).

Fix $\kappa \geq 0$, the prime $d \equiv 7 \bmod 16$, and $\xi = (\cdot/q)$. Since $d \equiv 3 \bmod 4$, $(\cdot/d)$ agrees with χ_{-d} on odd integers. Put

$$\mathcal{G} = (\mathbb{Z}/2 \times \mathbb{Z}/2) \times H_\kappa, \qquad \Phi = \{((0,\tau_1),\tau_2) : \tau_1 \in \mathbb{Z}/2,\ \tau_2 \in H_\kappa\} \cong \mathbb{Z}/2 \times H_\kappa.$$

Then $\mathcal{G}$ is the Galois group of the *host field* $E = \mathbb{Q}(\zeta_{2^{\kappa+2}})^+(\sqrt{-d}, \sqrt{-q})$, a CM field, and we write elements of Φ as pairs (τ_1, τ_2). For odd n prime to d let $\mathrm{Frob}(n) \in \mathcal{G}$ have first bit 1 iff $(n/d) = -1$, second bit 1 iff $(n/q) = -1$, and cyclotomic part $\pm n \bmod 2^{\kappa+2}$; powers of d go to the inert class. Since $d \equiv 7 \bmod 8$ and $q \mid d+1$, the primes 2 and q split in $\mathbb{Q}(\sqrt{-d})$, so even n and powers of q have inert weight 0 below; the values of Frob there play no role. Let χ_C, χ_Q be the characters reading the two bits, and $\mathrm{sh}(t) = ((1,1), t)$ for $t \in H_\kappa$.

Why this shape. The cyclotomic part supplies many positions and a group to average over. Its size $|H_\kappa| \asymp x^{2+8\varepsilon}$ fits the weight budget of the tower with slack $x^{5\varepsilon}$ ($k \leq |H_\kappa|^2 x^{1/2-\varepsilon}$ with $k \asymp x^{9/2+10\varepsilon}$, §6.4). The second quadratic field $\mathbb{Q}(\sqrt{-q})$ separates thick from thin classes (the thick classes τ have second bit 1; this refers to τ, not to n, for which $\mathrm{Frob}(n) = \tau - \mathrm{sh}(t)$ then has second bit 0, that is, $(n/q) = +1$), and after averaging it leaves exactly the characters $\chi_Q, \chi_C\chi_Q$ whose remainders (A) controls. We do not know a deeper reason for either choice.

Definition 3.2. For $a > 0$, $t \in H_\kappa$ and weights $\beta : \Phi \to \mathbb{R}_{>0}$ put

$$\mathrm{ASK}(a,t,\beta) = \sum_{\tau\in\Phi} \sum_{\substack{n \leq a/\beta(\tau) \\ \mathrm{Frob}(n)+\mathrm{sh}(t)=\tau}} \Lambda(n)\big(1 - (\tfrac{n}{d})\big)\Big(\frac{a}{n} - \beta(\tau)\Big),$$

$$\mathrm{MAIN}(a,\beta) = a\Big(\log a - 1 - \gamma + \ell_d - \frac{1}{|\Phi|}\sum_{\tau\in\Phi}\log\beta(\tau)\Big), \qquad \ell_d = \mathrm{Re}\,\frac{L'}{L}\big(1, (\tfrac{\cdot}{d})\big),$$

and, for a character ψ of $\mathcal{G}$ and a constant ℓ_ψ, $G_d(y) = \sum_{n\leq y}\Lambda(n)(\frac{n}{d})(\frac{y}{n} - 1) + y\ell_d$ and $\mathcal{G}_\psi(y) = \sum_{n\leq y}\Lambda(n)(1 - (\frac{n}{d}))\psi(\mathrm{Frob}n)(\frac{y}{n} - 1) + y\ell_\psi$.

The weight $\Lambda(n)(1 - (n/d))$ is $2\log p$ on odd powers of primes inert in $\mathbb{Q}(\sqrt{-d})$, $\log d$ on powers of d, and 0 otherwise; so $\mathrm{ASK} \geq 0$.

The cap and the Riesz weight. In the notation of ASK, the dictionary of §2.3 reads as follows. The contribution of each odd shift e is capped by a weight β attached to a partner position, so the local gain at an inert p is

$$\mathrm{gain}_p = \sum_{e \text{ odd}} \Big(\frac{a}{p^e} - \beta\Big)_+ \quad \text{(up to a logarithmic loss)},$$

$$2\log p \cdot \mathrm{gain}_p = \sum_{\substack{n = p^e \\ e \text{ odd}}} \Lambda(n)\big(1 - (\tfrac{n}{d})\big)\Big(\frac{a}{n} - \beta\Big)_+.$$

In our reading the cap is the weight $\beta(\tau)$ of Definition 3.2, up to bounded offsets [2, Lemma 12.23]. With one weight and t fixed, ASK restricted to the inert primes is the simplified sum S split by Frobenius classes. In the application $\beta \approx b = e^{x^\varepsilon}$ and $a \approx xb$, so the sums have length $a/\beta \approx x$.

Theorem 3.3 (balance identity; LEAN-CHECKED)**.** *For $a > 0$, $t \in H_\kappa$, $\beta > 0$ and any constants (ℓ_ψ),*

$$\mathrm{ASK}(a,t,\beta) - \mathrm{MAIN}(a,\beta) = \sum_{\tau\in\Phi}\frac{\beta(\tau)}{|\mathcal{G}|}\Big[2(G_\zeta - G_d)\Big(\frac{a}{\beta(\tau)}\Big) + \sum_{\psi\neq 1,\chi_C}\psi\big(\mathrm{sh}(t) - \tau\big)\,\mathcal{G}_\psi\Big(\frac{a}{\beta(\tau)}\Big)\Big].$$

Proof. Orthogonality of characters on $\mathcal{G}$. The character χ_C is -1 on the support of the inert weight and contributes as the trivial character does; the trivial character gives $2(G_\zeta - G_d)$ plus the main term, using $2|\Phi| = |\mathcal{G}|$. The constants ℓ_ψ contribute $a\ell_\psi\sum_{\tau\in\Phi}\psi(\mathrm{sh}(t) - \tau)/|\mathcal{G}|$, which vanishes for $\psi \neq 1, \chi_C$, since $\Phi = \ker\chi_C$ is a subgroup of index 2. □

Thick and thin classes: why two kinds. The identity expresses ASK−MAIN through remainders at the points $a/\beta(\tau)$, one for each class τ. For the final bound one wants only G_ζ, near x. Two different mechanisms can make the other remainders harmless, and each needs its own range of lengths. On a set of classes with weights near b, the lengths a/β are about x; averaging over translates kills most characters there, and the few that survive need square-root bounds, which the helper primes supply. On all other classes the weights are chosen so small that the lengths are enormous, at least $e^{x^{\varepsilon/3}}/3$; there a prime number theorem with saving makes every twisted remainder negligible relative to its length. The first kind of class carries the signal; the second only has to be harmless. Precisely: call the classes $((0,1), s+u)$, written $(1, s+u)$ in the pair notation, with s in a set $T_2 \subset H_\kappa$, *thick*, with weights in $[b, b+\eta+1]$; all other classes are *thin*. Here the lengths $a_t \approx xb$ are indexed by top places $t \in TT$, $s \in T_2$ labels the thick classes, and $u \in H_\kappa$ runs over translates, over which the inequality is averaged (the weights may depend on u; Appendix A).

classes	weights β	length a/β of the prime sums; what controls them
thick, $(1, s+u)$ with $s \in T_2$	in $[b, b+\eta+1]$, $b = e^{x^\varepsilon}$	about x; averaging and the helper primes (A)
thin, all others	small	at least $e^{x^{\varepsilon/3}}/3$; prime number theorem with saving (T)

Averaging over $u \in H_\kappa$ kills the characters non-trivial on H_κ, and, with all thick weights equal to b, the summands Q_τ of Theorem 3.3 satisfy

$$\frac{1}{|H_\kappa|}\sum_u \sum_{s\in T_2} Q_{(1,s+u)} = \frac{b\,|T_2|}{|\mathcal{G}|}\Big(2(G_\zeta - G_d) + \mathcal{G}_{\chi_Q} - \mathcal{G}_{\chi_C\chi_Q}\Big)\Big(\frac{a}{b}\Big).$$

Since $\chi_C = -1$ on the support of the inert weight, $\mathcal{G}_{\chi_C\chi_Q} = -\mathcal{G}_{\chi_Q}$ up to the constants, and $\mathcal{G}_{\chi_Q}$ is the difference of the remainders twisted by $(\cdot/q)$ and by $(\cdot/q)(\cdot/d)$. Rewriting, the thick average is $b|T_2|/|\mathcal{G}|$ times $2(G_\zeta - G_d + G_q - G_{qd})(a/b)$, up to terms linear in a/b (a rewriting of the display, not the identity used in the proof): on thick classes only the three twisted remainders G_d, G_q, G_{qd} survive besides G_ζ. One visible reason for the size $b = e^{x^\varepsilon}$, assuming the weights are bounded below by a fixed constant (we have not checked which lower bound the code uses): with $a \approx xb$, thin weights must be at most about $xb\,e^{-x^{\varepsilon/3}}$, which needs b well above $e^{x^{\varepsilon/3}}$; we have not checked whether this is the binding constraint.

3.4. **The analytic input.**

Statement 3.4 ((A), helper primes)**.** For $0 < \varepsilon < 1/52$ and $A \geq 1$, every large Y has primes $q \equiv 3 \bmod 8$, $d \equiv 7 \bmod 16$ with $\log q \asymp Y^{\varepsilon/5}$, $\log d \asymp Y^{\varepsilon/5}\log^2 Y$, $q \mid d+1$, and $d+1$ divisible by every odd prime $p \leq Y^{\varepsilon/5}$, such that: every primitive χ inducing $(\cdot/q)^i(\cdot/d)^j$, $i,j \in \{0,1\}$, $(i,j) \neq (0,0)$, has $|G_\chi(y)| \leq y^{1/2+\varepsilon/2}$ and $|L'/L(1,\chi)| \leq 4\log y$ for $y \in [Y, Y^A]$, where $G_\chi(y) = \sum_{n\leq y}\Lambda(n)\chi(n)(y/n-1) + yL'/L(1,\chi)$; and the twists of these characters by real characters of conductor dividing 8 have no zeros with $\operatorname{Re} s > 1 - 1/400$ up to height $8e^{Y^{\varepsilon/5}}(e^{Y^{\varepsilon/5}\log^2 Y})^j$.

Statement 3.5 ((T), thin-scale prime number theorem)**.** For $\varepsilon, A, c, c_0 > 0$ and large x, every primitive χ of conductor N with $\log N \leq Ax^{\varepsilon/5}\log^2 x$, without real zeros in $(1-c_0, 1)$ if it is real and non-trivial, satisfies $|G_\chi(y)| \leq y\,e^{-cx^{\varepsilon/8}}$ for $y \geq e^{x^{\varepsilon/3}}/3$. For the trivial character G_χ means G_ζ, so (T) also covers the ζ-part of the thin classes.

Both are classical analytic number theory: a zero-density estimate for the family of quadratic characters, in boxes; the Landau–Page theorem; Siegel's theorem; and the prime number theorem for characters without exceptional zeros, from the classical zero-free region (in (T), $\log y/\log N \gg x^{2\varepsilon/15}/\log^2 x$, so the saving has the form $\exp(-c\log y/\log N)$, outside the Siegel–Walfisz range).

The idea behind (A) appears to be that q is huge compared with the range $y \le Y^A$, so a mean-square argument over such q yields square-root cancellation for a suitable choice; the divisibility condition makes every small odd prime split in $\mathbb{Q}(\sqrt{-d})$, since $-d \equiv 1 \bmod p$. The third clause of (A) is what lets (T) apply to the real host characters.

3.5. **What the geometry must supply.** Summing the balance identity over the m steps of the tower, over top places $t \in TT$ and over translates, the proof needs an upper bound of the form

$$\frac{1}{|H_\kappa|}\sum_u \sum_{t\in TT}\sum_{j<m} (\mathrm{ASK} - \mathrm{MAIN}) \le x^{1/2+C\varepsilon}\cdot(\text{thick mass}),$$

the second line of (1). Dividing by the thick mass, the thick average leaves G_ζ near x plus three twisted remainders bounded by (A), and the thin classes are negligible by (T). The exact statements, the Galois hypothesis (V) and the geometric statement (G), and the deduction of RH are in Appendix A. Under the Riemann hypothesis the prime-side inequality is easy, and it is not what the geometry contributes. With a one-step tower and the smallest host ($\kappa = 0$, $d = 7$, $\xi = \chi_{-4}$ in place of $(\cdot/q)$, one thick class), the balance identity writes ASK − MAIN through G_ζ, G_7 and two further quadratic remainders, all of square-root size under RH for ζ and GRH for three quadratic L-functions. GRH is not even needed: with $\kappa = 0$, $m = 1$ and the helper primes of (A), RH bounds G_ζ, (A) bounds the twisted remainders of the thick class, and (T) handles the thin class; so RH, (A) and (T) already give the prime-side inequality (a short hand computation, not checked in Lean). Together with the reduction, the prime-side inequality is therefore equivalent to RH, given the classical (A) and (T). This says nothing about (G). The statement (G) assumes neither RH nor GRH; its inputs are (A) and the Galois bound (V); its host field is fixed by (A) and by $2^\kappa \asymp x^{2+8\varepsilon}$; and it must bound the G_ζ-term by the geometry and the tower alone.

Question 1 (Can the averaging mechanism be simplified?). The host field of degree about $x^{2+8\varepsilon}$, the thick and thin classes, the helper primes and the averaging over characters serve two purposes: they isolate G_ζ from the twisted remainders, and they provide many positions $\tau \in \Phi$ for the geometry. How much of this is essential? For RH alone one might start from the simplified sum $S(a,\beta)$ of §3.2, which already isolates $G_\zeta - G_d$. But the target is not easier for the simple sum: by §3.2, a square-root upper bound for $S(a,\beta)$ on far windows is equivalent, given the bound for G_d in (A), to the window bound for G_ζ, hence to RH. What might be simplified is the machinery that produces the bound (many positions with their own weights, the host field, the thick and thin classes), not the bound itself. How many positions τ, and which parts of the host field (the cyclotomic part, the second quadratic field), does the geometry itself need?

Verification notes for Act 0. Not read: the proofs of (A) and (T) in the development (they build axiom-clean locally). Reconstructed: the mean-square description of (A), the role of its third clause, and the dictionary between positions and embeddings. The Lean-checked window criterion allows windows $[Y, Y^{1+\delta}]$ for any $\delta > 0$.

4. Act 1: Chebyshev's principle and the product formula

Act 0 reduced the problem to an upper bound for a non-negative prime sum. This Act explains the principle that gives such bounds, and why it has to be used in lattice form.

Remark 4.1 (the toy model of §2.2, prime by prime). Let $K = \mathbb{Q}(\sqrt{-d})$ and let $a!_{\mathcal{O}_K} = \prod_{\mathfrak{P}} \mathfrak{P}^{\sum_{j\ge1}\lfloor a/N\mathfrak{P}^j\rfloor}$ be Bhargava's factorial [7]. In the identity of §2.2, a split p contributes 0 on both sides; an inert p, with residue field of size p^2, contributes $(2\sum_e \lfloor a/p^e\rfloor - 2\sum_j \lfloor a/p^{2j}\rfloor)\log p = 2\sum_{e \text{ odd}} \lfloor a/p^e\rfloor \log p$; and $p = d$ contributes $\log d \sum_j \lfloor a/d^j\rfloor$.

So the inert odd-power prime sum, with the floor weight, is the logarithmic size of an integral object. For an algebraic integer $P \neq 0$ in a number field K, Chebyshev's principle is the product

formula:

$$\sum_{w\nmid\infty} v_w(P)\log Nw \;\le\; \sum_{v\mid\infty}\log|P|_v.$$

As noted in §2.2, the toy model gives only Chebyshev quality; the proof replaces the ratio of factorials by integral periods whose archimedean size can be computed, and applies the principle in the lattice form of Arakelov geometry.

Why lattices. A single algebraic number records one divisibility. The proof needs to record, at once, the divisibility of a whole space of periods: the jets of order a at many positions, for all forms in a space of automorphic forms. The natural object is a lattice of such forms, with Hermitian norms at the infinite places. Divisibility at a finite place then becomes a twist of the lattice by an ideal, and the archimedean size becomes the covolume. The quantity that relates the two is the slope; this is Bost's slope method [11].

The lattice form. A Hermitian lattice over a CM field K is an $\mathcal{O}_K$-lattice $L\subset K^D$ with Hermitian metrics at the infinite places; its slope is $\mu(L) = -\frac{2}{[K:\mathbb{Q}]D}\log\operatorname{covol}(L)$. The slope inequality: if $\varphi: L\to M$ is injective with $\|\varphi x\|\le C\|x\|$, then $\mu(L)\le\mu_{\max}(M)+2\log C$ in this normalization of μ (a weak form of Bost [11, Prop. 4.3(1), (4.7)], used with $C=1$). If an ideal $I=\prod_w \mathfrak{p}_w^{-s_w}$ ($s_w\ge 0$) twists L to IL, the slope rises by $\frac{2}{[K:\mathbb{Q}]}\sum_w s_w\log Nw$; so divisibility of φ by I turns into a slope gain, which the slope inequality caps. This is the form used: a lattice $\mathcal{N}$ of automorphic forms, twisted by the divisibility of its periods at inert places, is mapped by a period map B into the dual of a lattice RR spanned by test vectors (integral automorphic forms of prescribed local type), and

$$\mu(\mathcal{N}) + \frac{2}{[K:\mathbb{Q}]}\sum_w \log Nw\cdot\operatorname{sg}(w) \;\lesssim\; \mu_{\max}(RR^\vee) + h(B), \tag{2}$$

where $\operatorname{sg}(w)$ is the divisibility exponent at w and $h(B)$ the archimedean size of B (the logarithm of its norms, averaged over the infinite places). Three things must be supplied: the divisibility (Act 2), the archimedean sizes $h(B)$ and $\mu_{\max}(RR^\vee)$, and a lower bound for $\mu(\mathcal{N})$ (Act 2, size side, and Act 3).

One map or a family. The code treats a vanishing period map separately: for a word $\mathfrak{w}$ of Act 3 with $B_{\mathfrak{w}}=0$ the integrality and height clauses hold trivially, with the same budget. What is finally applied is not the inequality above for one map but a joint version for the family $(B_{\mathfrak{w}})_{\mathfrak{w}\in\mathfrak{W}}$. It bounds the mean of $\mu(\mathcal{N}_{\mathfrak{w}})-h(B_{\mathfrak{w}})$ by the *mean* slope of the common target, up to $\log|\mathfrak{W}|$ and error terms that grow with the excess of the maximal slopes of the $\mathcal{N}_{\mathfrak{w}}$ over that mean. Its non-degeneracy hypothesis is not injectivity of each $B_{\mathfrak{w}}$. It is a Gram-matrix condition: the diagonal blocks of the combined period matrices are close to the identity in Hilbert–Schmidt norm, and the off-diagonal blocks are small (see Question 2). We have read the statement of this joint estimate, not its proof. In the code the base side is bounded through the mean slope of $RR^\vee$. In Adriazola's reading [2, §13], passing from $\mu_{\max}$ to the mean slope uses the whole family of period maps $B_{\mathfrak{w}}$, one for each level prime $T_{\mathfrak{w}}$ of Act 3, and an approximate orthogonality of the corresponding Poincaré series.

Verification notes for Act 1. Read in the statement: the passage from $\mu_{\max}$ to the mean slope is a joint estimate over words under a Gram-matrix condition. Unchecked: its proof; where the Gram condition is supplied; how a word with $B_{\mathfrak{w}}=0$ is accounted for inside the joint proof.

5. Act 2: divisibility at inert primes

The toy model has the right divisibility but no computable size. This Act produces the same divisibility from objects whose size can be computed. The path is: the ordinary factorial (Legendre), Bhargava's factorial for $\mathcal{O}_K$ (the inert odd-power part), and finally the formal group of a CM elliptic curve at an inert prime, which realizes that divisibility on jets of functions.

5.1. **The height-two formal group.** Let $p \geq 5$ be inert in $K = \mathbb{Q}(\sqrt{-d})$ and W the completion of the maximal unramified extension of $\mathbb{Z}_p$. A CM elliptic curve with CM by $\mathcal{O}_K$ has, after a choice of model, good supersingular reduction over W (Deuring [22], Serre–Tate [51]; in the code the good model is obtained through global class field theory, which the development proves rather than assumes). Its formal group has height 2 and, as a formal $\mathbb{Z}_{p^2}$-module, height 1; so over W it is the Lubin–Tate group with logarithm $\ell(X) = \sum_{j\geq 0} p^{-j} X^{p^{2j}}$ [40], [29, Thm. 2, (5.2)]. Let $D = d/d\ell(X)$ and $\lg_p(a) = \lfloor \log_p a \rfloor + 1$, the number of base-$p$ digits of a. Put $\gamma_p = p/(p^2-1)$.

Proposition 5.1. *For $\theta \in W[[X]]$ and $a \geq 1$,*

$$v_p\big(D^a\theta(0)\big) = v_p\big(a!\,[z^a]\theta(\ell^{-1}(z))\big) \;\geq\; v_p(a!) - \frac{a}{p^2-1} \;\geq\; \gamma_p a - \lg_p(a).$$

Proof. Let $\varpi^{p^2-1} = p$. The series $\ell(\varpi Y)/\varpi = \sum_j \varpi^{p^{2j}-1} p^{-j} Y^{p^{2j}}$ has coefficients of valuation $(p^{2j}-1)/(p^2-1) - j \geq 0$, so it and its inverse μ are integral. From $\ell^{-1}(z) = \varpi\mu(z/\varpi)$, $v([z^a]\ell^{-1}) \geq -(a-1)/(p^2-1)$, which gives the first inequality. Legendre's formula $v_p(a!) = (a - s_p(a))/(p-1)$, with digit sum $s_p(a) \leq (p-1)\lg_p(a)$, and $\frac{1}{p-1} - \frac{1}{p^2-1} = \gamma_p$ give the second. □

This is the toy model in real form: the height-two logarithm puts one denominator on each $X^{p^{2j}}$, so it eats only the even powers of p in $a!$; for height one the rate would be $1/(p-1)$ and nothing would remain. It matches the note's phrase "$D^p \equiv 0$ on a CM curve for p inert": since $\ell' \equiv 1 \bmod p$, $D \equiv d/dX$ and $D^p \equiv 0 \bmod p$, a vanishing Hasse invariant. The one-variable statement is our reconstruction of the mechanism; the code proves a several-variable form for the theta kernel.

Remark 5.2 (the sharp form; known, not used). In fact $v([z^a]\theta(\ell^{-1}(z))) \geq -\sum_{j\geq 1}\lfloor a/p^{2j}\rfloor$, hence $v_p(D^a\theta(0)) \geq \sum_{e \text{ odd}} \lfloor a/p^e \rfloor$: the toy model is realized exactly, as a lower bound. This is known: the coefficients of $[c](Z) = \ell^{-1}(c\,\ell(Z))$ are integer-valued polynomials in $c \in \mathbb{Z}_{p^2}$ of degree at most a (de Shalit–Iceland [21]; see [20, (1.9)]); the c^a-coefficient of the coefficient of Z^a is $[z^a]\ell^{-1}$ (zero unless $a \equiv 1 \bmod p^2-1$); and the c^a-coefficients of integer-valued polynomials of degree at most a form $p^{-w(a)}\mathbb{Z}_{p^2}$, $w(a) = \sum_i \lfloor a/p^{2i} \rfloor$, which is Bhargava's factorial [7] (Cahen–Chabert [15, Prop. 1.3], as quoted in [21, Prop. 2.2]); superadditivity of w passes to $\theta \circ \ell^{-1}$.

5.2. **Theorem K: the theta kernel and the odd shifts.** *What height two gives, and what goes wrong.* Proposition 5.1 gives the odd-part divisibility for jets of *integral* functions. The functions that the proof needs are theta kernels, and the theta kernel is not integral at p: in logarithmic coordinates it has the form $\exp(-\mathcal{E})U$ with U integral and $\mathcal{E}$ a quadratic expression with denominators. A priori these denominators could cancel the gain. What is needed is a statement that the non-integral part is small, and that what remains of it is aligned with the odd shifts, so that the gain survives in the capped form of §3.3. Theorem K says this: up to an integral factor, the whole non-integral part is one exponential $\exp(-\varpi^{p+1} hh/p)$, where hh pairs positions at odd distance and each power of hh costs at most one p^{-1}.

The setting is as follows. The theta kernel is a series in several matrix-valued variables z_s indexed by $s \in \mathbb{Z}/n'$ (n' even); in our reading these are coordinates at the Frobenius conjugates. Put $S_s = \sum_l (-1)^l z_{s-2l}^{\,p^{2l}}$ and $hh = \sum_s u_s \sum_{a,b} S^p_{s,ab} S_{s+1,ab}$, a sum over matrix entries with unit coefficients u_s. Call a series of the *class* if it is $\sum_i p^{-i} hh^i U_i$ with U_i integral. For such series the code proves, by a routine count (Legendre's formula, a min/max optimization, and vanishing of Hasse derivatives on products of minors), a jet bound of the shape $\sum_s \sum_{e \text{ odd}} (\deg_s/p^e - \mathrm{cap})_+$ minus a logarithmic loss: the same capped odd-shift expression as in §3.3. The substance is that the theta kernel lies in the class.

Theorem 5.3 (Theorem K; proved in the development)**.** *Let $p \geq 5$ be inert, $\varpi^{p^2-1} = p$, and let $\Theta = \exp(-\mathcal{E})\,U$ be the theta kernel in logarithmic coordinates, with $U \circ \ell$ integral and $\mathcal{E} = \sum_i \mathrm{be}(z_i, z_i')$*

for a biadditive σ-semilinear series be *in pairs of variables* (z_i, z_i'). *For series* f, g *in the formal-group coordinates write* $f \simeq g$ *if* $\exp(f-g)$ *is integral, and put* $A(X,Y) = \sum_{0\le i\le j} p^{-(i+j+1)} X^{p^{2i}} Y^{p^{2j+1}}$, $A'(X,Y) = A(Y,X)$.

(a) *There is* $c \in \mathbb{Z}_{p^2}$, $|c| \le 1$, *with* $\mathrm{be} \circ \ell \simeq G_c = c\,A - \sigma(c)\,A'$.

(b) *After rescaling the variables by* ϖ *and passing to Frobenius-twisted coordinates (in which* hh *is written),* $\Theta = \exp(-\varpi^{p+1} hh/p) \cdot V$ *with* V *integral over* $W[\varpi]$, *where* $u_s = c$ *for odd* s *and* $u_s = -\sigma(c)$ *for even* s.

(c) *Hence* $\Theta = \sum_i p^{-i} hh^i U_i$ *with* $U_i = (-\varpi^{p+1})^i V/i!$ *integral; each* hh *costs at most one* p^{-1}, *the net valuation of the coefficient of* hh^i *being* $-i + s_p(i)/(p-1)$.

Outline. (a) classifies biadditive quasi-logarithms on the height-two Lubin–Tate group up to integral exponentials, using the Dieudonné–Dwork integrality criterion for exp. (b) In $G_c(\varpi X, \varpi Y)$ the (i,j) term has coefficient $\varpi^{M_{ij}-(p^2-1)(i+j+1)}$ with $M_{ij} = p^{2i} + p^{2j+1}$: exactly ϖ^{p+1}/p for $(i,j) = (0,0)$, and ϖ^{p+1} times an integral series otherwise, whose exponential is integral by the estimate in (c). So only $p^{-1}(c\,XY^p - \sigma c\,YX^p)$ breaks integrality, and modulo p it becomes hh, by the orthogonality of a trace-dual basis of $\mathbb{F}_{p^{n'}}$ over $\mathbb{F}_{p^2}$. (c) follows from $v(\varpi^{(p+1)i}/i!) = s_p(i)/(p-1) \ge 0$. We checked the identities behind (a) and (b) numerically for $p = 3, 5, 7$, with negative controls that do fail.

Why the shifts are odd. In S_s^p every exponent is p^{2l+1}. So a term of hh linear in the lower variables pairs a variable at position k_0, raised to p^{2l+1}, with a variable at position $k_0 + 2l + 1$. This is the source of the odd exponents in the gain and in ASK.

Where the cap comes from (heuristic; RECONSTRUCTED*).* In the jet count for the class, a factor hh used at the odd shift e consumes degree p^e at a position s and one "hit" of a lower minor at the partner position $s+e$; once the hits at the partner are used up, the Hasse derivatives vanish. So the number of uses at shift e is at most $\min(a_s/p^e, \mathrm{cap}(s+e))$, where $\mathrm{cap}(s+e)$ is the number of hits available at the partner, and the denominators eat at most $\sum_{e\ \mathrm{odd}} \min(a_s/p^e, \mathrm{cap}(s+e))$ of the gain $\gamma_p a_s = \sum_{e\ \mathrm{odd}} a_s/p^e$. What survives is $\sum_{e\ \mathrm{odd}} (a_s/p^e - \mathrm{cap}(s+e))_+$. That this cap is the weight $\beta(\tau)$ up to bounded offsets is Adriazola's reading [2, Lemma 12.23].

Known and new. The classification (a) fits the theory of biextensions of formal groups and its Dieudonné-theoretic description (Mumford [44]; in display language Zink [68]), though we found no statement in this explicit form; theta functions on formal groups (relevant here because at an inert prime the whole p-divisible group is formal), modulo quadratic exponentials, are the subject of Barsotti–Cristante theory [18]. What appears to be new is the explicit triangular normal form, and the collapse of the whole non-integral part to $\exp(-\varpi^{p+1} hh/p)$, which forces one denominator per odd shift. That the theta kernel has the form $\exp(-\mathcal{E})U$ at all is proved in the code from the good model of the CM curve; we have not read that proof.

5.3. **The size side: Fock norms, CM periods, covolumes.** *Fock norms.* Polynomial spaces carry the Bargmann–Fock inner product $\|z^\alpha\|^2 = \alpha!$. Integral Weyl-module lattices and their duals are compared using Cauchy–Binet, multinomial coefficients and the Capelli identity [14,30,31]. This appears to produce the term $a \log a - a$ of MAIN, by Stirling applied to hook-length products. Two slope estimates on this side, a lower bound for the slope of an integral Weyl module and an exact slope of a column lattice, appear to be new and elementary; we found no source.

CM periods. The Chowla–Selberg formula [50], in Kronecker-limit form (Kronecker's limit formula: [38,52]): for K of discriminant $-d$, $d > 4$, and $\chi = \chi_{-d}$,

$$\frac{1}{h_K} \sum_{c \in \mathrm{Cl}_K} \log\bigl(4\pi\,\mathrm{Im}\,\tau_c\,|\eta(\tau_c)|^4\bigr) = \gamma + \log\frac{2\pi}{\sqrt{d}} - \frac{L'}{L}(1,\chi).$$

For a CM lattice $\Lambda = \omega(\mathbb{Z}\tau + \mathbb{Z})$ normalized by $\Delta(\Lambda) = 1$ one has $\mathrm{area}(\Lambda)/\pi = 4\pi\,\mathrm{Im}\,\tau |\eta(\tau)|^4$, and the Galois conjugates of Λ run over the ideal classes equally often; so the mean of $\log(\mathrm{area}\,/\pi)$ over

conjugates is the left side. The constant that enters each step of the tower is $\log\Omega = \gamma + \log(\pi/v) - L'/L(1,\chi_{-d})$ for a normalizing constant v attached to the host field (in the code $v = \mathrm{rd}_E/2$, half the root discriminant of E; the discrepancy $\log(\mathrm{rd}_E/\sqrt{d})$ with the Chowla–Selberg constant comes from the relative different of $E/\mathbb{Q}(\sqrt{-d})$ [2, Lemma 15.4, Prop. 15.6]). The term $\log(\pi/v)$ cancels against the level prime (Act 3), and what remains, $-\gamma + L'/L$, is the $-\gamma + \ell_d$ of MAIN.

Covolumes. For the j-th lift lattice M_j (of rank D_j), with $(k+j-1, m-j)$ the signature of the j-th Shimura-side space and $\beta_{j\tau}$ the weights,

$$\frac{\log \mathrm{covol}(M_j)}{D_j} \lesssim \sum_{\tau} \big((k+j-1)-(m-j)\big)\,\beta_{j\tau}\log\Omega + (\text{per-step error}),$$

where $\lesssim$ also absorbs lower-order terms that we do not reproduce. This is proved by arithmetic Hilbert–Samuel applied to nef models of Hodge bundles on unitary Shimura varieties, with resolution of singularities behind it. Only the upper bound $\mathrm{vol}(\bar{L}) \le \bar{L}^{D+1}$ for the number of small sections is used; for nef $\bar{L}$ equality holds (Moriwaki [43, Cor. 5.5]; the volume is a limit by Chen [16, Thm. 5.2]; see also [1, 26, 64, 67]). These covolumes feed the bound for the base side in (2) through the product formula along the tower. The constant in the covolume bound must be uniform in the level, which grows with x. We know of no published arithmetic Hilbert–Samuel theorem with such uniformity. The development appears to pass to the limit through a tower of models, using an exact identity at good primes [2, §15.14.4].

Where the terms of MAIN *come from.* The following table records the role of each source (RECONSTRUCTED from the code; it is a map of roles, not a list of theorems).

term of MAIN or of the ledger	source	role
$a\log a - a$	Fock norms, Capelli identity, Stirling	archimedean size of the jets of order a, once per step
$-\gamma + \ell_d$	Chowla–Selberg, through $\log\Omega$	size of the CM periods, once per step
$-\overline{\log\beta}$	choice of the level prime	enters through the lower bound for $\log T$ (§6.3)
$\log(\pi/v)$, $\log(g_0 g_1)$	$\log\Omega$; degree count at the type primes	once per step; canceled by the level prime
(no term of MAIN)	covolume estimates	bound the base side of (2), so that only MAIN and errors remain; in our reading (unchecked) the $\log\Omega$ of MAIN is the one in the height of the period map, and the covolumes' $\log\Omega$ terms go into the base side

Question 2 (Can the covolume estimate be obtained more directly?). The covolume bound uses arithmetic Hilbert–Samuel for nef models of Hodge bundles, with resolution of singularities behind it, and its constant must be uniform in the level. A covolume of a Hermitian lattice combines an archimedean part and an integral part (the lattice at the finite places). Here the archimedean part is an L^2-type determinant, the covolume of spaces of sections of a Hodge bundle on unitary Shimura varieties whose level grows with x. Bergman-kernel asymptotics already enter a proof of arithmetic Hilbert–Samuel: Abbes and Bouche [1, Prop. 3.2 and the proof of Lemma 3.9] use the asymptotics of the Bergman density $\sum_j |s_j(x)|^2$, uniform in the point (their Prop. 3.2, due to Bouche and to Tian), together with spectral asymptotics for the $\bar{\partial}$-Laplacian. The sharpened analytic question is therefore whether such estimates (Bergman or reproducing kernels, spectral, L^2 or determinant bounds, potential theory) can be made uniform in the level, the property that the proof needs and for which we know no published statement. Such estimates would not by themselves control the

integral structure at the finite places, which is where the arithmetic enters; what might plausibly be replaced analytically is the archimedean part. The passage from the maximal to the mean slope (§4) is already a Gram-matrix and Hilbert–Schmidt statement in the proof; what remains there is the supply of the Gram condition and the proof of the joint estimate.

Verification notes for Act 2. Unchecked: the proof that the theta kernel has the form $\exp(-\mathcal{E})U$; uniformity of the covolume constant in the level. Reconstructed: Proposition 5.1 in one variable, the heuristic for the cap, the attribution of $a\log a - a$ to the Fock side, and the table of the origins of MAIN.

6. Act 3: multiplicity

6.1. **What the tower has to do.** One application of Acts 1 and 2 gives a slope inequality whose gain is one copy of the prime sum ASK and whose cost contains a fixed normalization. As in §2.4, in our reading of the ledger one step gives only Chebyshev quality. The rest of this Act answers the following questions.

- *What is passed along the tower?* Automorphic forms, lifted from one unitary group to the next by theta kernels; at each step, in our reading, their periods are again jets of theta kernels at CM points, with the divisibility of Act 2.
- *What is amplified?* The prime sum: the lattice gain, and with it ASK and MAIN, appear once per step, m times.
- *Why is the fixed cost paid once?* Because it is attached to a level prime chosen once, at the bottom of the tower, and the identity of §6.3 cancels the order-m constants against it.

The length is then fixed by §6.4: $m \approx \sqrt{x}$ is where the fixed cost x/m reaches the level $\sqrt{x}$ set by the dropped characters (Question 3).

6.2. **Stepanov's method and the tower.** Stepanov's method for curves over finite fields uses an auxiliary polynomial vanishing to high multiplicity at the rational points; over $\mathbb{Z}$ there is no such polynomial, and multiplicity has no obvious meaning. In the author's description, iterated theta lifts play its role [4]. The code contains no vanishing-order argument. What plays the role of multiplicity is the length m of a tower.

The spaces. There are integers k, m even and ν odd with $m \le k$, $3m < \nu$, $m^2 \le k + m(\nu+1)$; put $k_W = k + m - 1$. Over the host field there are four families of Hermitian spaces: a definite chain X_j of rank $k + j(\nu+1)$, $0 \le j \le m$, along which theta lifts are iterated; spaces Y_j of rank $k_W + j\nu$ on the "Bessel side", definite at $j = m$; spaces W_j $(1 \le j \le m)$ of rank k_W and signature $(k + j - 1, m - j)$ on the "Shimura side", which do not form a chain; and definite spaces V_j $(1 \le j \le m)$ of rank $k_W + \nu$, targets of lifts from W_j. On $W_j \to V_j$ the Langlands parameter acquires a block $\mu_j[\nu]$, matching the note's "$\pi \mapsto \pi \boxplus \mu_j[t]$". The construction is made for each element $\mathfrak{w}$ of a finite family $\mathfrak{W}$ of "words". For each word there is a prime $T_{\mathfrak{w}}$, the *level prime*, and the space Y_0 of the tower of $\mathfrak{w}$ contains a vector of norm $-T_{\mathfrak{w}}$. Each level prime is fixed once, at the bottom of its tower (Figure 5; the family is described in [2, Prop. 3.1]).

The objects. A "column" of the tower is the period obtained by lifting an automorphic form φ through the m steps (a composite of m kernel maps, finite sums over classes of lattices weighted by inverse automorphism counts) and pairing the result with a test vector; the base lattice RR is built place by place from integral test vectors attached to the Hecke characters assigned to each place. The lattice $\mathcal{N}$ of (2) is the lattice of forms that are "ramified at T", twisted by the expected divisibility of their periods. Three facts about these objects are used.

- *Transport.* The transport of Bessel periods through the tower by seesaw identities (in the style of Kudla's seesaw, with Bessel and Fourier–Jacobi periods of Gan–Gross–Prasad type) appears to be used to carry finitely many linear relations among periods, with constants uniform over the Hecke twists.

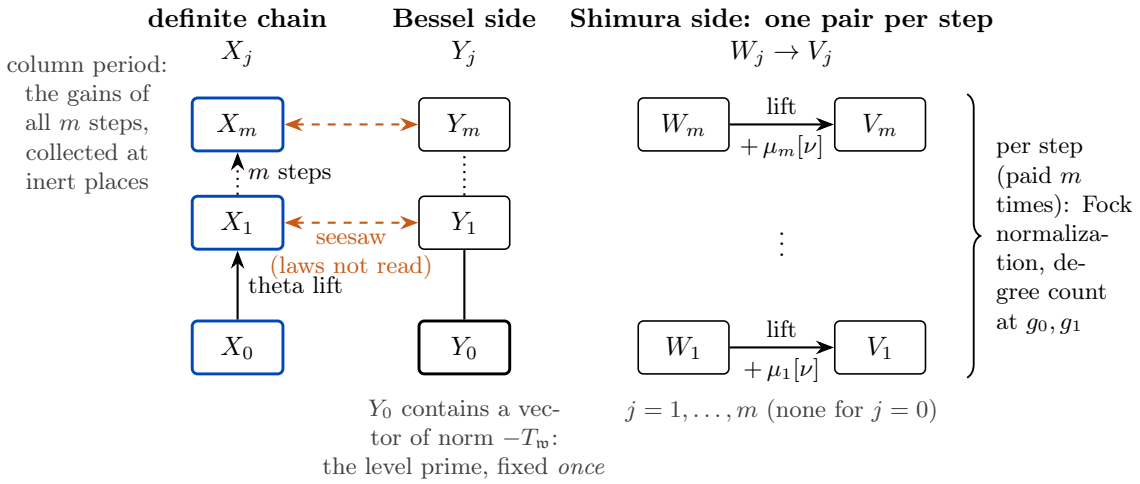


FIGURE 5. Anatomy of the theta-lift tower for one word $\mathfrak{w}$ (schematic). Lifts are iterated along the definite chain $X_0 \to \cdots \to X_m$, and, in our reading, the gains of the m steps are collected in the column period at the top. The Y_j (Bessel side) form a second chain; Y_0 carries the level prime $T_{\mathfrak{w}}$, fixed once (§6.3). For $j = 1, \dots, m$ a space W_j, not a chain, lifts to V_j, adding a block $\mu_j[\nu]$. Solid links: relations our agents read in the code, not a formal verification. Dashed: the seesaw transport of periods, whose laws we have not read. Links between the W_j and the chains are not drawn (not checked).

- *Non-vanishing,* in two forms: a bottom-rung non-vanishing of a Fourier–Jacobi period, and an existence statement (some ramified form has a non-zero projected Poincaré series), obtained by bounding the kernel dimension by the Hilbert–Schmidt deviation of a Gram matrix from the identity. This existence statement does not give $B \neq 0$ (that case is split off, §4). It is used instead to normalize the Bessel data, which makes the period map well defined on the base lattice and equal to the column pairing, and, it appears, to obtain a non-zero constant in a Bessel identity.
- *Gluing* of the base lattice across Hecke characters. It uses two auxiliary primes $\equiv 1 \bmod 15$, each carrying an auxiliary character ramified only there, with tame inertia of order exactly 15; of the two, one is chosen different from the coefficient prime ℓ.[2]

The tools appear to be standard: Rallis inner products, Siegel–Weil, the Siegel mass formula and seesaw identities.

6.3. **How the gain enters, and why the fixed cost is paid once.** Fix a word $\mathfrak{w}$ and write $T = T_{\mathfrak{w}}$ and $\mathcal{N} = \mathcal{N}_{\mathfrak{w}}$.

The gain as a twist. The slope of $\mathcal{N}$ is computed from its definition: $\mathcal{N}$ is the T-ramified part of the bottom lattice M_0, twisted by the ideal of expected divisibilities, so

$$\mu(\mathcal{N}) = \mu(M_0) - (\text{a level correction at } T) + \frac{2}{[K:\mathbb{Q}]} \sum_w \log Nw \cdot \Big\lfloor \sum_{j<m} (\mathrm{gain}_j + F_j - E_j)(w) \Big\rfloor,$$

with the gains of §3.3 at inert places, and correction terms F_j, E_j; at the two type primes the gain vanishes and E_j absorbs the possible poles of the periods there. The rounding loss from the floor is bounded separately. The divisibility of Act 2 is what makes the period map integral on this

[2]The order-15 character supplies, for each ℓ, a prime $p \mid 15$ with $p \notin \{2, \ell\}$ (5 if $\ell = 3$, otherwise 3), and hence a tame inertia element without inertia invariants. In our reading this is why 15 appears: it is the smallest odd number with a prime factor other than ℓ for every ℓ.

twisted lattice. So the gain enters (2) m times. The gain is not free: there is a Hecke loss at the inert primes [2, §12]. It is made negligible by the choice of the weights, through a non-degeneracy statement for integral representations of GL_k.

The ledger. The right side of the resulting divisor inequality contains three kinds of term:

- $m\big(a\log a - a - a\log\Omega\big)$, from the Fock norms and CM periods, once per step;
- $-a\log T$, from the level prime, once;
- $m\,a\log(g_0g_1)$, from a degree count of each lift at two auxiliary "type" primes $g_0 \neq g_1$, once per step (the periods may have poles at the type primes, and the degree count measures them).

The level prime is chosen, in a short interval, split in the CM field and with a congruence condition, so that $\log T$ matches the order-m constants. Its existence appears to need a prime number theorem in progressions; we have not read where the development proves it. With $a_\Sigma = \sum_t a_t$, $\overline{\log\beta_j}$ the mean of $\log\beta_{j\tau}$ over $\tau\in\Phi$, and $L_a = \log a_\Sigma + x^{\varepsilon/3}$, the choice of T gives

$$\sum_{j<m}\Big(\overline{\log\beta_j} + \log\frac{v}{\pi}\Big) + m\log(g_0g_1) - c\Big(L_a + \frac{m^2L_a}{k}\Big) \le \log T \le c\,m\,x^{\varepsilon}.$$

Substituting this lower bound for $\log T$ into the term $-a_\Sigma\log T$, one checks the identity

$$\sum_t m\big(a_t\log a_t - a_t - a_t\log\Omega\big) - a_\Sigma\Big[\sum_{j<m}\Big(\overline{\log\beta_j} + \log\frac{v}{\pi}\Big) + m\log(g_0g_1)\Big] + m\,a_\Sigma\log(g_0g_1)$$
$$= \sum_{j<m}\sum_t a_t\Big(\log a_t - 1 - \gamma + \frac{L'}{L}(1,\chi_{-d}) - \overline{\log\beta_j}\Big),$$

which is m copies of MAIN: the terms $m\,a_\Sigma(\log v - \log\pi)$ and $m\,a_\Sigma\log(g_0g_1)$ have canceled. The only cost of the choice of T is $c\,a_\Sigma(L_a + m^2L_a/k)$, and it is *not* multiplied by m. Had it been paid once per step, the gain would be lost. So the fixed normalization cost is paid once, while the gap ASK − MAIN is counted m times. This holds for each word. The divisor inequalities for the words are then combined by the joint estimate of Act 1 [2, §13]. This arrangement appears to be new.

In the notation of §2.4: a naive repetition would pay $mF + \sum_j C_j$, the tower pays $F + \sum_j C_j$, and after division by the signal, which is proportional to m, the fixed cost contributes F/m (Figure 6). The identity above is the precise statement; the F, C_j picture only summarizes it.

6.4. **Why the error is $x^{1/2+O(\varepsilon)}$.** The sizes on the main path are

$$b = e^{x^{\varepsilon}},\quad a \approx xb,\quad |H_\kappa| \asymp x^{2+8\varepsilon},\quad m \asymp x^{1/2-\varepsilon/16},\quad k \asymp x^{9/2+10\varepsilon},$$

and the per-step error budget is $B_0 = 2|H_\kappa|\,m\,b\,x^{\varepsilon}$. In the part we checked, its factor m is used only to absorb a buffer term at the type primes; we have not checked whether the level-prime budget also needs it. The original's final estimate has the form

$$G_\zeta(x) \le x^{1/2+(C+3)\varepsilon} + \sum_{\psi\ \text{dropped}} \|G_\psi(x)\|,$$

where the dropped characters are the twisted helper characters that the thick average leaves, each at most $x^{1/2+\varepsilon}$ by (A) (for RH there are three, the twists by $(\cdot/d)$, $(\cdot/q)$ and $(\cdot/qd)$, which occur in the thick average $2(G_\zeta - G_d + G_q - G_{qd})$ of §3.3; for GRH there are $3o$, o the order of the character, cf. [2, proof of Thm. 6.3]). This is equivalent to the bound in the proof of Theorem A.4. The first term is the tower's. Its error has three sources: the fixed cost of the level prime and of the base side; the per-step errors, m times B_0; and the weight budget k. Divided by the signal $|TT|\,m\,b\,\delta_2$,

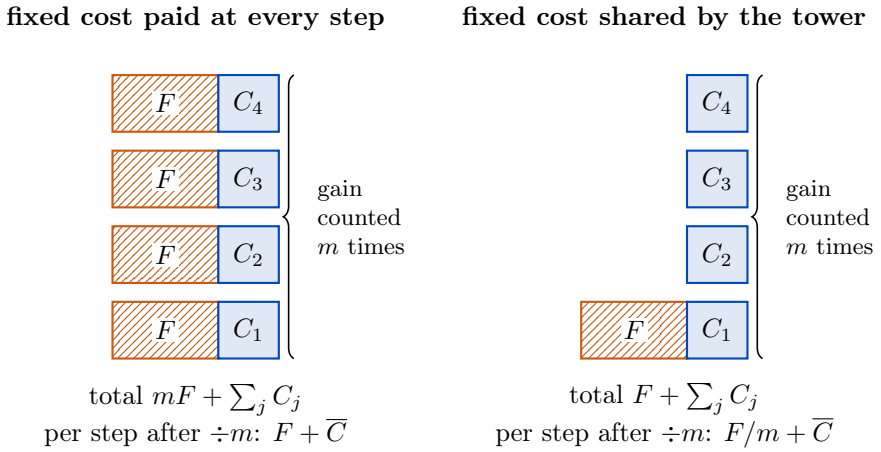


FIGURE 6. One fixed cost, many gains (conceptual model, $m = 4$; F a fixed cost, C_j the cost of step j, $\overline{C}$ their mean). Paid at every step, F survives averaging; shared by the m steps, it becomes F/m. In the proof the mechanism is an exact identity at the level prime $T_{\mathfrak{w}}$ (§6.3); the model is not the proof.

with $\delta_2 = |T_2|/|H_\kappa|$, they give, each times $x^{O(\varepsilon)}$,

$$\underbrace{\frac{x}{m\,\delta_2}}_{\text{fixed cost}} + \underbrace{\frac{|H_\kappa|}{|TT|}\cdot\frac{m}{\delta_2}}_{\text{per-step cost}} + \underbrace{\frac{|H_\kappa|}{|TT|}\cdot\frac{k}{|H_\kappa|^2}}_{\text{weight budget}}.$$

item	size before normalization	paid	after dividing by the signal (up to $x^{O(\varepsilon)}$)
signal	thick mass $\lvert TT\rvert\, m\, b\, \delta_2$	m times	1
fixed cost	level prime and base side, $\approx a_\Sigma(L_a + m^2 L_a/k)$	once	$x/(m\delta_2)$
per-step cost	$B_0 = 2\lvert H_\kappa\rvert m b x^{\varepsilon}$ per step	m times	$(\lvert H_\kappa\rvert/\lvert TT\rvert)\, m/\delta_2$
weight budget	k	—	$(\lvert H_\kappa\rvert/\lvert TT\rvert)\, k/\lvert H_\kappa\rvert^2$
dropped characters	$\le x^{1/2+\varepsilon}$ each, by (A)	independent of the tower	$x^{1/2+\varepsilon}$

In the size data we have read, δ_2 and $|TT|/|H_\kappa|$ are $x^{-O(\varepsilon)}$ (the thick classes are chosen with a Behrend-type density; the author's footnote 7 calls Behrend inessential, and that it refers to this choice is our inference); we have not checked this on the main path, and Definition A.1 alone gives only $\delta_2 \ge x^{-1/2}$. Granting it, the three terms are about x/m, m and $k/|H_\kappa|^2 \le x^{1/2-\varepsilon}$, all $x^{1/2+O(\varepsilon)}$ for m near $\sqrt{x}$.

Why $m \approx \sqrt{x}$. Granting the density conditions, the bracket reduces to $x/m + m + \sqrt{x}$, up to $x^{O(\varepsilon)}$. In our reading the term m is the buffer (the factor m in B_0); the level $\sqrt{x}$ is set by the dropped characters, each at most $x^{1/2+\varepsilon}$ by (A), which do not depend on the tower, and by the weight budget; and $m \approx \sqrt{x}$ is where x/m reaches that level. This is the note's "$\sqrt{X}$" for the length of the chain. The model $x/m + m + \sqrt{x}$ of §2.4 is a conceptual summary of this reading, not a proof: it omits the density factors and constants above, and the density conditions and the role of the buffer term are unverified.

Question 3 (Is the cost of amplification intrinsic?). The model $x/m + m + \sqrt{x}$ describes the structure of the upper bounds in the existing proof; it is not a lower bound for anything. A longer tower gains nothing, since the target and the dropped characters are already at $\sqrt{x}$; in the bounds

as written the buffer term proportional to m would also penalize it, and we do not know whether that term is intrinsic. Is the term x/m intrinsic to the construction, or only to the way the fixed cost is paid? Is the buffer, the factor m in the per-step budget B_0 at the type primes, a technical loss? Each lift's error is bounded separately in the proof; could the tower, the composite of m kernel maps, be estimated as one operator, for instance by telescoping, determinant, spectral or variational arguments? Could the amplification be realized by another construction? Removing the term m alone would not give a short tower: the term x/m remains, and still needs $m \gtrsim \sqrt{x}$.

Verification notes for Act 3. Unchecked: the seesaw laws; the gluing of the base lattice; the inputs to the slope of $\mathcal{N}$ other than the rounding bound; the Hecke non-degeneracy; the combination over $\mathfrak{W}$ (Act 1); the density conditions; whether the level-prime budget needs the factor m. Reconstructed: the explanation of the number 15 and the reading of the bracket.

7. Act 4: paying with Ш

7.1. **The idea of Act 4.** Acts 1–3 produce, at every step of the tower, a lattice of automorphic forms whose periods are highly divisible. The slope inequality charges for the integral structure of these lattices, and this charge has to be bounded. Act 4 can be read on its own. Its argument has eight steps; the starred steps are our interpretation, and not every step is an unconditional theorem.

(1) The integral structures of the theta lifts cost something at each step: their lattices need not split integrally along Hecke eigenspaces.
(2) This cost appears as the size of congruence modules, at certain twisted places (§7.2).
(3)* Congruence modules are bounded by Galois cohomology, through Selmer-type groups; for the lifts of the tower these reduce to the Tate–Shafarevich groups $Ш^1(E, \mathbb{Q}_\ell/\mathbb{Z}_\ell(-e'))$ of the host fields.
(4) These groups split into a minus part and a plus part. The minus part is controlled by known theory, the main conjecture of Iwasawa theory.
(5)* The plus part sits at a trivial zero of the L-function. The main conjecture expresses its order as an index of cyclotomic elements (cf. [32]), not as an L-value, and gives no bound polynomial in the conductor; in our reading this is why it needs another method.
(6)* Deformation theory bounds the plus part by the cotangent space of a deformation ring at a residually reducible point (in the code, of a model ring whose identification with the deformation ring we have not read).
(7)* A commutative-algebra lemma, which appears to be new (Lemma M), bounds that cotangent space by congruences between automorphic points and an Eisenstein point; these congruences are valuations of one non-zero algebraic integer, or of its norm, for all ℓ at once.
(8) The logarithm of the archimedean size of that integer is polynomial in the conductor, and this gives the uniform bound.

At step (8) Chebyshev's principle is used a second time, now for the primes ℓ: a non-zero integer NN satisfies $\sum_\ell \log \ell \cdot v_\ell(NN) \le \log |NN|$, and this is what makes the bound uniform in ℓ (Figure 7).

7.2. **Where Ш enters.** The host fields are required to satisfy

$$\sum_{\ell} \log \ell \sum_{1 \le e' < t_V} \operatorname{length} Ш^1\big(G_{E,S_\ell}, \mathbb{Q}_\ell/\mathbb{Z}_\ell(-e')\big) \le \sqrt{b} = e^{x^{\varepsilon}/2},$$

which is the bound $\mathrm{VV}(E, t_V, \sqrt{b})$ of Statement A.2, with $t_V = \lfloor x^{401} \rfloor$. This is more than the geometry needs: there only twists e' below the odd rank step ν occur, and the window data ask for $\lfloor x^{400} \rfloor$. The bound is used once per lift $j < m$. Each lift carries an ℓ-adic Hecke lattice, the ℓ-adic completion of its lattice of theta lifts. At each twisted place one takes the congruence module of the relevant Hecke eigenspace, the image of the lattice under the projection to the eigenspace modulo the intersection of the lattice with it, in the sense of Ribet and Hida. The code bounds its length by

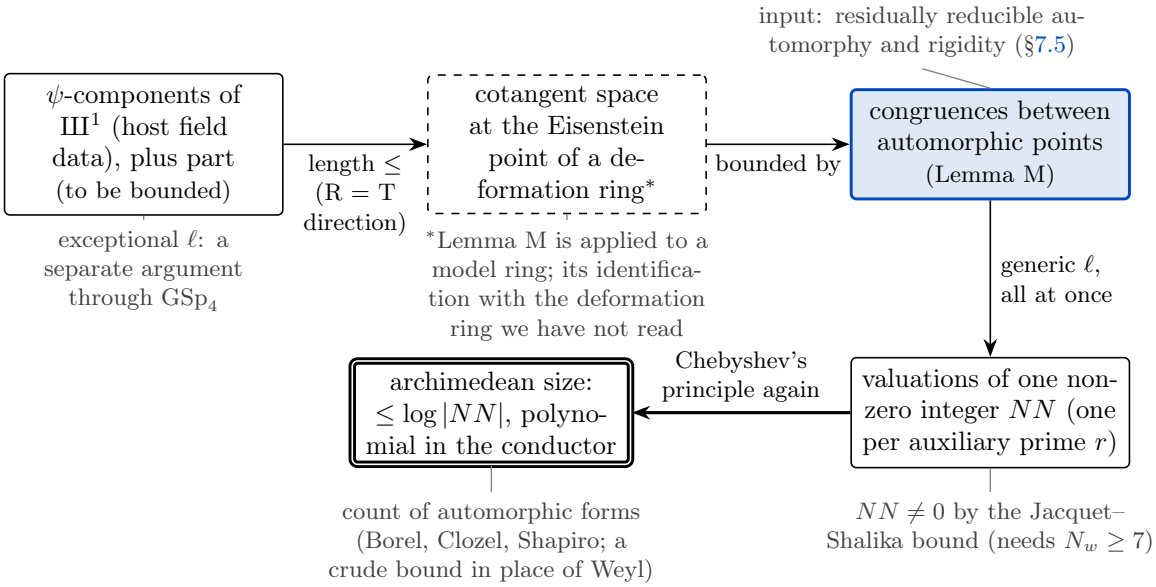


FIGURE 7. The Galois-theoretic cost (conceptual diagram of the argument for (V3), §7.4; our interpretation, not a chain of unconditional theorems). The length of $Ш^1$ is bounded by a cotangent space (dashed: Lemma M is applied to a model ring whose identification with the deformation ring we have not read), that by congruences between automorphic points, and these, uniformly in the generic ℓ and with one NN per auxiliary prime, by $\log|NN|$, polynomial in the conductor. Exceptional ℓ go through GSp_4.

an explicit polynomial term, local terms, and the Ш sum above, in which ℓ runs only over the two type primes g_0, g_1. These lengths enter the bound for the slope of the base lattice, the right side of (2), in our reading through its discriminant, as part of an error of size $C\,m\,(B_0+N_c)\,x^{O(\varepsilon)}$ already present in the ledger of §6.4; here $N_c = |T_2|\,k\,b\,x^{2\varepsilon}/|H_\kappa|^2$ [2, §17.1] (his n is our $|H_\kappa|$). The size $\sqrt{b}$ of the Ш bound reaches the geometry only through this term, and not through the gain or the level correction; finiteness of the groups $Ш^1$ is used in addition inside the proof of the congruence bound itself. The development proves (V) for all ℓ, but in our reading the geometry consumes Ш terms only at $\ell \in \{g_0, g_1\}$; we do not know whether a weaker statement would suffice (Question 4).

The tie to lattices. In our reading the congruence lengths recorded in the ledger are actual lattice invariants, not free numbers: two terms in proofs on the main path identify them with the length of the projection image modulo the intersection, and with the corresponding local length on the lift side. Two points we have not checked: whether these lattices are exactly the theta-side lattices of the tower, and where one hypothesis of the tie is supplied.

Question 4 (Is the full Galois bound necessary?). (V) bounds Ш for all primes ℓ and all twists $e' < \lfloor x^{401} \rfloor$, with size $e^{x^\varepsilon/2}$. In our reading the geometry consumes the size only at the two type primes g_0, g_1 (the auxiliary primes at which the Hecke lattices of the lifts are compared), only for twists below the odd rank step ν (its window data ask for $\lfloor x^{400} \rfloor$), and separately needs the finiteness of the groups. Is a bound only at the primes and twists actually used enough? Is uniformity in ℓ forced by the order in which the construction chooses g_0, g_1, or by how that choice interacts with its other conditions? A positive answer might greatly reduce what Act 4 has to supply; we state this only as a possibility.

Remark 7.1 (the size of the Galois bound). In the ledger as written, the code uses the bound only through $\sqrt{b} \le b \le B_0$, where $B_0 = 2|H_\kappa|\,m\,b\,x^\varepsilon$ is the per-step budget, that is, b times a

polynomial factor in x. The step it feeds appears to need only a bound $\le Kx^{\varepsilon}(B_0 + N_c)$, in particular $\le K\,b\,x^{2\varepsilon}$. So $\exp(cx^{\varepsilon})$ for any fixed $c \le 1$ appears to suffice, while $c > 1$ would not fit the ledger as written; (V1)–(V3) in fact deliver about $\exp(x^{4\varepsilon/5})$. That finiteness is used inside the proof of the congruence bound is our reading. We have not checked that every downstream consumer of the resulting divisibility bound ignores the size of the congruence term.

Why Ш *(interpretation).* By the numerical criterion of Wiles, Ribet and Hida, congruence modules are bounded by adjoint Selmer groups. The Galois representations of the lifts are sums of CM characters times powers of the cyclotomic character, so the adjoint splits into twists of $\mathbb{Q}_\ell/\mathbb{Z}_\ell(-e')$ by characters of finite order. For $e' \ge 1$ the local conditions contribute only small terms, and the Bloch–Kato Selmer groups of these twists are essentially $Ш^1$. This is why a Vandiver-type statement appears.

By Poitou–Tate duality, $Ш^1(E, \ell, e')$ is essentially dual to $H^2(\mathcal{O}_E[1/\ell], \mathbb{Z}_\ell(1+e'))$. By the Quillen–Lichtenbaum conjecture (a theorem of Rost and Voevodsky) this is $K_{2e'}(\mathcal{O}_E) \otimes \mathbb{Z}_\ell$. So, up to local terms and the prime 2, (V) bounds odd parts of K-groups of the host fields. This is how (V) enters (G), as the hypothesis of Statement A.3.

7.3. **Three parts.** Write $Ш^{\pm}$ for the parts of $Ш^1(E, \ell, e')$ on which complex conjugation acts by $\pm(-1)^{e'}$; for odd ℓ the length of Ш is the sum of their lengths. Write $\mathrm{VV}_\ell(L, t_V, B)$ for the bound at a single prime ℓ, and $\mathrm{VV}^{\pm}(E, t_V, B)$ for the bound on $\sum_{\ell \text{ odd}} \log \ell \sum_{e' < t_V} \operatorname{length} Ш^{\pm}$.

Statement 7.2 ((V1), bad primes and windows)**.** For $0 < \varepsilon < 1/52$ and $2^{\kappa(x)} \asymp x^{2+8\varepsilon}$ there are $J \ge 1$ and layer depths $j(\ell)$ such that for all large X and every pair (q, d) at scale $X^{1/J}$ there is a window $[Y, Y^2] \subset [X^{1/J}, X]$ on which, with $t_V = \lfloor x^{401} \rfloor$: the host fields satisfy $\mathrm{VV}_2(E, t_V, \exp(x^{4\varepsilon/5}))$; and for $\ell \in \{3, 7\}$ the layer of degree $\ell^{j(\ell)}$ of the cyclotomic $\mathbb{Z}_\ell$-tower over $\mathbb{Q}(\zeta_\ell)$ satisfies $\mathrm{VV}_\ell(\cdot, t_V, \exp(x^{4\varepsilon/5}))$.

Statement 7.3 ((V2), minus part)**.** For $\kappa, t_V \ge 2$ and primes $q \equiv 3 \bmod 8$, $d \equiv 7 \bmod 16$, every host field satisfies $\mathrm{VV}^{-}\big(E, t_V, 3[E : \mathbb{Q}] t_V^2 \log(t_V \operatorname{cond} E)\big)$.

Statement 7.4 ((V3), plus part)**.** For every choice of layer depths there are C, R such that for $\kappa, t_V \ge 2$ and primes $q \equiv 3 \bmod 8$, $d \equiv 7 \bmod 16$ with $q, d > \max(2t_V + 1, R)$, and $B \ge 0$ bounding the groups at $\ell \in \{3, 7\}$ on those layers, every host field satisfies $\mathrm{VV}^{+}\big(E, t_V, (N_E t_V)^C + CB\big)$, where $N_E = 2^{\kappa+2} qd$.

Theorem 7.5 (Lean-checked)**.** *(V1), (V2) and (V3) imply (V).*

Sketch. On the windows of (V1) the 2-part is at most $\exp(x^{4\varepsilon/5})$. The minus part is polynomial in $[E : \mathbb{Q}] \asymp x^{2+8\varepsilon}$, t_V and $\log \operatorname{cond} E$. The plus part is at most $(N_E t_V)^C + C \exp(x^{4\varepsilon/5})$. Since $\log(N_E t_V) \ll x^{3\varepsilon/5}$, all three are below $\exp(x^{\varepsilon}/2)$ for large x. This is why (V3) need only be polynomial in the conductor: N_E is itself only $\exp(O(x^{\varepsilon/5} \log^2 x))$. □

(V1) is classical in its ingredients: Ferrero–Washington [25] (see also [53]) in the form "the ℓ-rank of class groups is bounded along the cyclotomic $\mathbb{Z}_\ell$-tower", an effective Soulé-type bound [56], which is doubly exponential, and Poitou–Tate and local duality. The windows are then selected among the ranges $[Y, Y^2]$; we have not read how the selection reaches $\exp(x^{4\varepsilon/5})$ (dividing a doubly exponential total by J would not suffice). In our reading, this is the only source of windows in (V).

(V2) is a weak form of known results. For an abelian CM field $E \subset \mathbb{Q}(\zeta_N)$, odd $\ell \nmid [E : \mathbb{Q}]$, $e' \ge 1$ and a character ψ whose Dirichlet character χ has $\chi(-1) = (-1)^{e'+1}$, the code proves $\operatorname{length} Ш^1(E, (\mathbb{Q}_\ell/\mathbb{Z}_\ell \otimes \mathcal{O})(-e'))^{\psi} \le v(N \cdot \mathrm{primorial}(e' + 2) \cdot B_{e'+1,\chi})$, with Ш here trivial on all inertia groups. This is the parity in which $L(\chi, -e') = -B_{e'+1,\chi}/(e' + 1) \ne 0$; the bound follows from the main conjecture (Mazur–Wiles [41], Wiles [63]) in the finite-level form of Kolster–Nguyen Quang Do–Fleckinger [35] and Huber–Kings [32, Thm. 1.3.1]. The code proves it by an Euler-system

argument in the style of Kolyvagin and Rubin [36,39]. Whether χ or χ^{-1} governs the eigenspace we have checked only modulo ℓ.

(V3) appears to be new. It has no window and no analytic input, and is polynomial in the conductor and uniform in ℓ. Vandiver's conjecture is its class-group analog ($e' = 0$) and is open [61]. The note calls this bound "a significant weakening of Vandiver" and adds that it is "still pretty strong".

7.4. **The argument for (V3).**

(1) *Parity.* The plus part decomposes into ψ-twisted groups over $\mathbb{Q}$ for the characters ψ of $\mathrm{Gal}(E/\mathbb{Q})$ with $\psi(-1) = (-1)^{e'}$; this is the only parity condition. We believe this is why only the plus part needs the argument: in the other parity $L(\chi, -e') \neq 0$ and (V2) applies, while here $L(\chi, -e')$ has a trivial zero, and the main conjecture gives an index of cyclotomic elements rather than an L-value, with no bound polynomial in the conductor. Primes ℓ are sorted into a uniform case (ℓ outside the bad set and $(\ell - 1) \nmid 2e'$) and finitely many exceptional cases with per-ℓ polynomial bounds. Two of the exceptional cases (Cases II and III in the code) use deformations and automorphy for the symplectic group GSp_4 [2, §§17.3, 17.4.1]. Their counts come from a separate GSp_4 part of the development, which we have not read.

(2) *The auxiliary field.* The argument works over $K = \mathbb{Q}(\sqrt{-p_K})$, where, for a character of conductor f, p_K is the least prime $\equiv 3 \bmod 4$ with $p_K \geq 7$ and $p_K \nmid f$. The congruence makes K imaginary quadratic with discriminant $-p_K$; we do not know why $p_K \geq 7$ is required. For the trivial character $p_K = 7$, and this is the 7 of the bad set $\{3, 7\}$ (primes dividing the discriminant and class number of K must be excluded). The 3 of the bad set enters as an explicitly added prime, and also as the value at the trivial character of a second auxiliary prime; we have not identified its role. The residual representation is $\bar{r} = \Psi \oplus 1 \oplus \Psi^{-1}$, $\Psi = \chi_\ell^{e'} \psi$, into the rank-3 unitary group of $K/\mathbb{Q}$, ordinary at ℓ.

(3) *Ш is bounded by a cotangent space* (standard). A class in Ш gives a non-split extension, hence a first-order deformation of $\bar{r}$ at the Eisenstein point, hence a weight -1 tangent vector of the ordinary deformation ring, so that the length of Ш is at most that of the weight -1 cotangent module. This is Mazur's identification of tangent spaces with Selmer groups, used to bound Ш *from above*: the $R = T$ direction, as in Skinner–Wiles [54], not the direction of Ribet or Bellaïche–Chenevier, which gives lower bounds.

(4) *The cotangent space is bounded by congruences* (Lemma M below; appears to be new). The dictionary is given after the lemma. The outcome is $\mathrm{length}\, Ш \leq 2 \sum_\varrho v_\ell(\mathrm{tr}\, \varrho(\sigma) - t_{\mathrm{Eis}}(\sigma))$, a sum over the automorphic points ϱ.

(5) *One integer for all ℓ.* Let NN be the norm of $\prod$(Hecke eigenvalue − Eisenstein value) over the relevant forms and the Frobenius elements at the places w of K above an auxiliary prime r. Then $\sum_\ell \log \ell \cdot v_\ell(NN) \leq \log |NN|$, uniformly in ℓ. Two auxiliary primes $r_1 < r_2$ are used, so that $\ell = r_1$ is covered by r_2. They are chosen with $r_1 \geq 30\, N_\psi\, p_K + 1$, N_ψ the conductor of the character, with $r \nmid 30\, N_\psi \,\mathrm{disc}\, K$, and with $q, d > r_2$. So $r \geq 7$, whence $N_w \geq 7$ below, and at the trivial character $r_1 \geq 211$; and r_1, r_2 give the threshold R of (V3). The norms are $N_w \in \{r, r^2\}$.

(6) $NN \neq 0$, from Jacquet–Shalika alone. The Eisenstein value $N_w^{-e'} z + 1 + N_w^{e'} z^{-1}$ (z a root of unity) has absolute value at least $N_w^{e'} - 1 - N_w^{-e'}$. After the unitary twist a cuspidal trace λ is a sum of three Satake parameters, each of absolute value in $[N_w^{-1/2}, N_w^{1/2}]$ (Jacquet–Shalika [34], valid over any number field), with product of absolute values 1; the code deduces $|\lambda| \leq \sqrt{N_w} + 2$. The two ranges are disjoint for $N_w \geq 7$ and $e' \geq 1$. The hypothesis is needed for the bound as used: at $N_w = 5$, $e' = 1$, $z = -1$ the Eisenstein value is $-21/5$, of absolute value below $\sqrt{5} + 2$ (the sharp bound $\sqrt{N_w} + 1 + N_w^{-1/2}$ would separate even there). No

Ramanujan-type bound is used. A purity bound $|\lambda| \leq 3$, which is also valid, would separate the ranges already for $N_w \geq 5$ [2, Prop. 17.21, Rem. 17.23]. On the main path, however, the code uses Jacquet–Shalika and so needs $N_w \geq 7$.

(7) *Polynomial size.* The number of automorphic forms of bounded level and weight is polynomial: cuspidal cohomology injects into $H^*(\Gamma, E)$ (Borel [10]), the relevant forms are cohomological (Clozel [17]), $H^*(\Gamma, E)$ is bounded by Shapiro's lemma and the index of the level group, and the coefficient system by a crude polynomial bound (in place of Weyl's dimension formula [62]). Hence $\log|NN|$ is polynomial in N_E and t_V.

What Lemma M does, before its statement. Its input is a graded complete local ring, in practice the ordinary deformation ring at the residually reducible point, presented with one fewer equation than coordinates, together with a function f that vanishes at the residual point and at no "two-signed" point. Its output is a bound for the weight -1 cotangent space at the residual point, where Ш lives, by the valuations of f at finitely many points. Ordinary congruence-module theory is not enough here. The numerical criterion of Wiles and Lenstra gives the inequality "cotangent $\geq$ congruence" in general, with equality exactly for complete intersections; the inequality needed here, cotangent bounded by a congruence-type quantity, is the complete-intersection direction, and no complete-intersection hypothesis is assumed (but see the calibration below). Moreover the right side of Lemma M is a sum of valuations of one function at points, not the congruence ideal of a Hecke algebra.

Proposition 7.6 (Lemma M; proved in the development)**.** *Let $\mathcal{O}$ be a complete discrete valuation ring of characteristic 0, $\mathbb{L} \supset \mathcal{O}$ algebraically closed, and v a valuation on $\mathbb{L}$ extending that of $\mathcal{O}$. Let $A = \bigoplus_{j\in\mathbb{Z}} A_j$ be a graded $\mathcal{O}$-algebra with A_0 complete local, generated over A_0 by finitely many elements of weights $\pm1, \pm2$, with a point $o : A \to \mathcal{O}$ killing A_j for $j \neq 0$. Let I be generated by $r_- + 1$ equations of weight -1, s_- of weight -2, r_+ of weight 1, s_+ of weight 2 and m_0 of weight 0 vanishing at o, and let $f \in A_0$. Call a point $\xi : A \to \mathbb{L}$ killing I* two-signed *if it kills neither all positive nor all negative weights. Assume:*

(M1) $r_- \geq 1$; there is a smooth graded chart at o with $r_-, s_-, r_+ + 1, s_+ + 1, m_0$ coordinates in weights $-1, -2, 1, 2, 0$; and the equations of weights $-1, -2, 0$ span the cotangent lattices of their weights up to non-zero scalars;

(M2) the image of A_0 in A/I is a finitely generated $\mathcal{O}$-module;

(M3) at every torus-fixed point $o_1 \neq o$ of A/I, points killing only the negative weights and points killing only the positive weights, with the same restriction to A_0, do not both occur;

(M4) at every two-signed point the tangent space of A/I has dimension at most 1;

(M5) $o(f) = 0$, and $\xi(f) \neq 0$ at every two-signed point ξ.

Then the two-signed points are covered by a finite set $\mathcal{T}$, the weight -1 cotangent length $\cot_{-1}$ of A/I at o is finite, and $\cot_{-1} \leq 2\sum_{\xi\in\mathcal{T}} v(\xi(f))$.

Dictionary. The lemma is applied as follows (Reconstructed). A/I is (a model of) the ordinary deformation ring of $\bar{r}$, graded by conjugation by the diagonal torus: the entry in position (i, j) of the universal representation has weight $w_i - w_j \in \{0, \pm1, \pm2\}$ with $(w_0, w_1, w_2) = (1, 0, -1)$. The point o is $\bar{r}$, and Ш embeds in the weight -1 tangent space, the dual of the cotangent module, so $\operatorname{length} Ш \leq \cot_{-1}$. The function is $f = E_{00} + E_{11} + E_{22} - \operatorname{tr}\bar{r}(\sigma)$, the trace of the universal representation at a Frobenius σ (the E_{ii} being its diagonal entries) minus its residual value. A point is two-signed when it is neither upper- nor lower-triangular; such points are either irreducible over G_K or of block type $2 + 1$. Irreducible ones are irreducible on every open subgroup (a step proved in the development), hence automorphic (§7.5) and the $2 + 1$ ones come from modular forms, so $\xi(f) = \operatorname{tr}\varrho(\sigma) - t_{\mathrm{Eis}}(\sigma)$.

Calibration. No Gorenstein or complete-intersection hypothesis appears. But the presentation has exactly one fewer equation than coordinates, and A/I should have Krull dimension 2; this is the

shape of a complete intersection, and experts should check how much of that property is built in. For comparison, the Wiles–Lenstra numerical criterion gives the opposite inequality between cotangent and congruence modules, with equality exactly for complete intersections; non-complete-intersection cases are studied by Böckle–Khare–Manning [9] and Iyengar–Khare–Manning [33], and a residually reducible variant appears in Wake–Wang-Erickson [59]. An elementary analog: for $\mathcal{O}[Z]/(Zg(Z))$ with g monic, at $Z = 0$, the cotangent length is $v(g(0)) = \sum_{g(\theta)=0} v(\theta)$. We found no published analog of Lemma M, but have not searched exhaustively.

A Vandiver-direction bound over $\mathbb{Q}$. The development also proves a Vandiver-direction bound over $\mathbb{Q}$, as a theorem, whose proof we have not read. There are C, A such that for every even $e' \geq 2$ and every finite set S of primes $\ell \geq 5$ with $(\ell-1) \nmid 2e'$ the groups $Ш^1(\mathbb{Q}, \mathbb{Q}_\ell/\mathbb{Z}_\ell(-e'))$ are finite and

$$\sum_{\ell \in S} \log \ell \cdot \frac{\operatorname{length}_{\mathcal{O}_\ell} Ш^1(\mathbb{Q}, \mathbb{Q}_\ell/\mathbb{Z}_\ell(-e') \otimes \mathcal{O}_\ell)}{e(\mathcal{O}_\ell)} \leq C(e'+1)^A,$$

for any choice of coefficient rings $\mathcal{O}_\ell$, $e(\mathcal{O}_\ell)$ being the ramification index; we checked this form of the statement. With $\mathcal{O}_\ell = \mathbb{Z}_\ell$ it follows that $Ш^1(\mathbb{Q}, \mathbb{Q}_\ell/\mathbb{Z}_\ell(-e')) = 0$ once $\log \ell > C(e'+1)^A$ [2, Cor. 17.15]. It comes from the plus-part machinery applied over $\mathbb{Q}(\sqrt{-3})$ with auxiliary primes 7 and 11. It is not on the path to RH (we checked this in the code; see also [2, Prop. 17.19]).[3]

7.5. **Residually reducible automorphy.** *Why the proof needs it.* Lemma M bounds Ш by values of f at the two-signed points. For these values to be differences between Hecke eigenvalues and Eisenstein values of automorphic forms, so that they enter the single integer NN whose size can be counted, every two-signed point must come from an automorphic form. The irreducible two-signed points are deformations of a reducible residual representation, so this is a residually reducible automorphy theorem, and it is needed for all such points, not only for a chosen family.

Step 4 needs the following theorem of the development, which it proves for imaginary quadratic K (written for CM fields). Let $\bar{r}$ be the residual representation of §7.4, $\operatorname{diag}(\bar{\Psi}, 1, \bar{\Psi}^{-1})$ on G_K, extended to $G_\mathbb{Q}$ with values in $\mathrm{GL}_3 \rtimes \{\pm 1\}$. There is no similitude factor; this is not the group $\mathcal{G}_3$ of Clozel–Harris–Taylor, and comparing with it costs a twist by ϵ^{-1}. Consider its ordinary deformations with Hodge–Tate weights fixed at $(e', 0, -e')$ (not a Hida family), minimal ramification away from ℓ and a normalization. Then:

(a) the weight-zero part of the universal ring (for the action of the diagonal torus) is finite over $\mathcal{O}$; in characteristic ℓ all deformations have the residual characteristic polynomial on a subgroup of finite index;
(b) every characteristic-zero point that is irreducible on every open subgroup of G_K is automorphic: it comes from a regular algebraic conjugate self-dual cuspidal representation of $\mathrm{GL}_3(\mathbb{A}_K)$;
(c) at the points of type $2+1$ the two-dimensional block comes from a newform of weight $2e'+1$.

A sibling statement over $\mathbb{Q}(i)$ serves an exceptional case.

The proof. It follows the Skinner–Wiles strategy [54].

- A Selmer rank is bounded by a constant C_0 by Ferrero–Washington [25], proved in the development. Then an auxiliary CM field is taken high in a cyclotomic $\mathbb{Z}_\ell$-tower, so that the sum $\delta = \sum_v \delta_v$ of its local degrees at the places above ℓ satisfies $\delta > 3C_0 + 7$ and each $\delta_v > 5$.

[3] Separately, the windowed bound (V) alone gives $\log\log P(U) \leq \frac{1}{2}U^{\varepsilon/401}$ along a sequence $U \to \infty$, where $P(U)$ is the product of the orders of these groups over $e' < U$ and ℓ outside a finite set; according to [2, Prop. 17.22], the published bounds of Soulé [56] and de la Torre Durán [19], through $K_{2e'}(\mathbb{Z}) \otimes \mathbb{Z}_\ell$, give only $\log\log P(U) \ll U^2 \log U$. We checked the deduction from (V), but not those sources, and (V) itself is not verified by us. Vandiver's conjecture for all primes is related to the vanishing of $K_{4k}(\mathbb{Z})$ for all $k \geq 1$ [37]; we have not checked the exact statement.

- Over that field, the ingredients are Hida theory, Eisenstein seed points (the ad^0 of Eisenstein-congruent elliptic newforms), and Taylor–Wiles-type estimates at $k[[X]]$-valued points, k a finite field.
- The triangular locus has dimension at most $3\delta - 1$. This gives connectedness through "thick" points, hence automorphy of all irreducible points.
- A solvable descent returns to K.

Rigidity of the automorphic points uses a theorem of Kato and the vanishing of adjoint Bloch–Kato Selmer groups, a special case of Thorne [58, Thm. A] (Thm. 6.1 in the arXiv version). That theorem says: for a polarizable regular algebraic cuspidal π on GL_n over a CM field F with $r_{\pi,\iota}|_{G_{F(\zeta_{p^\infty})}}$ irreducible, $H^1_f(F^+, \mathrm{ad}\, r_{\pi,\iota}) = 0$. The code proves its own version, with $n \leq 4$ and ordinary. Two inputs differ from Skinner–Wiles:

- *The rank bound.* For the analogous step, Skinner–Wiles use Washington's theorem on the non-p-part of class numbers in cyclotomic $\mathbb{Z}_p$-extensions [60] after an abelian base change (as recalled in [23, §1.3.2]). The development instead uses a bound for the ℓ-rank of the ℓ-part. This is a different theorem in the same role: keeping a Selmer dimension small compared with the degree.
- *No Schur hypothesis.* In our reading the Schur hypothesis is avoided as follows. The deformations are lifts of the diagonal residual representation on the nose, normalized by $r(c) = 1$ and by the vanishing of two off-diagonal coordinates at an element $g_1 \in G_K$ with $\bar{\Psi}(g_1)^2 \neq 1$. For the two root directions of the remaining orthogonal group these conditions give a transversal slice, because $\bar{\Psi}(g_1)^2 - 1$ is a unit. The remaining torus is handled by passing to its invariant (weight-zero) subring, which contains all traces and plays the role of the determinant ring. So what is used is that the constituents are distinct, not Schur.

Known and new. The theorems of Allen–Newton–Thorne [3, Thms. 1.1, 5.1] and Thorne [57, Thm. 7.1] assume that the residual representation is Schur [57, Def. 3.2] (Thorne's theorem assumes two adequate constituents with conditions that imply this) and that there is a Steinberg place, which controls the reducible locus.[4] Neither holds here, so the theorem is not a special case of theirs [2, §17.4]:

- $\bar{r}$ is not Schur, since the quotient $\bar{\Psi}$ and the dual of the subobject $\bar{\Psi}^{-1}$ are isomorphic after conjugation: the polarization exchanges $\bar{\Psi}$ and $\bar{\Psi}^{-1}$ instead of preserving each;
- there is no Steinberg place; the reducible locus is controlled instead by the rank bound and the tower.

The closest result we found is the theorem of Berger–Klosin [5, Thm. 10.9]. It treats the same non-Schur residual type, again without a Steinberg place, through pseudodeformations and Hida theory for $U(2,1)$. Its conclusion is conditional p-adic modularity: it assumes among other things reducedness of Hecke algebras and p split in K, and it gives specialization of a Λ-adic Hecke algebra, not classical automorphy in a fixed weight. So the unconditional statement appears to be new. The method combines known parts (the Skinner–Wiles big field and connectedness, Hida theory, Taylor–Wiles, Thorne's adjoint vanishing [58], Bellaïche–Chenevier-type rigidifications) with a normalization that replaces the Schur hypothesis. A single web search found no other non-Schur unitary lifting theorem; this is not exhaustive.

Galois representations. Galois representations for the relevant forms are constructed in the code through definite theta lifts, a route that differs from the published constructions, which go through the cohomology of Shimura varieties. The local correspondence used is a weak fragment of local Langlands [13, 27, 28, 49], together with Bernstein–Zelevinsky theory [6, 65].

[4] In Thm. 1.1 of [3] these are hypothesis (4), that the constituents be individually conjugate self-dual and mutually non-isomorphic, and hypotheses (5) and (6c), the Steinberg place.

Verification notes for Act 4. Unchecked: the proofs of Lemma M and of the automorphy theorem; the identification of the model ring with the deformation ring; the window selection of (V1); the GSp_4 part; the two points of the tie to lattices; the roles of the 3 in $\{3, 7\}$ and of $p_K \geq 7$; non-circularity beyond a name-based check. Reconstructed: steps (3), (5), (7) of §7.1, the "Why III" paragraph, the dictionary for Lemma M, and the account of how Schur is avoided.

8. Finale

8.1. **Classical inputs.** Each is used or proved inside the development, often in a special or non-standard form; the table gives the form used. A theorem number is given where it has been checked.

statement	reference	form used
one-sided bound on windows ⇒ zero-free half-plane	Landau; see Montgomery–Vaughan [42]	window version, Riesz weight
zero-density boxes, Landau–Page, Siegel, PNT for characters	classical	helper primes (A), thin scales (T)
slope inequality	Bost [11, Prop. 4.3(1)]	Euclidean, $C = 1$
good reduction of CM curves; supersingularity at inert primes	Serre–Tate [51]; Deuring [22]	over W, via class field theory
height-2 Lubin–Tate groups	Lubin–Tate [40]; Honda [29, Thm. 2, (5.2)]	explicit logarithm
integrality of $\exp(P)$	Dieudonné–Dwork (classical)	several variables
Capelli identity	[14, 30, 31]	as is
Chowla–Selberg, Kronecker limit formula	[38, 50, 52]	mean over conjugates
arithmetic Hilbert–Samuel	Moriwaki [43, Cor. 5.5]; Chen [16, Thm. 5.2]; [1,26,64,67]	upper bound only
resolution of singularities	Hironaka (classical)	nef models
theta correspondence: Rallis, Siegel–Weil, seesaw	classical	tower
dense sets without short relations	Behrend (classical)	thick classes; inessential by the author's footnote 7
bounded ℓ-rank along cyclotomic towers	Ferrero–Washington [25]; [53]	class groups mod ℓ
effective finiteness of K-groups	Soulé [56]	effective, doubly exponential
minus part ≤ valuation of $B_{e'+1,\chi}$	[35, 41, 63], [32, Thm. 1.3.1]; [36, 39]	inequality, finite level
Satake parameters in $[N_v^{-1/2}, N_v^{1/2}]$	Jacquet–Shalika [34]	stated in the code for totally complex fields
GSp_4 deformations, automorphy, rigidity	proved in the development; described in [2, §§17.3, 17.4.1]	Cases II, III of (V3)
injectivity of cuspidal cohomology; cohomological forms	Borel [10]; Clozel [17]; [62]	counting forms
residually reducible automorphy	[3, 54, 57], Bellaïche–Chenevier; closest: Berger–Klosin [5]	parts; the unconditional statement appears new (limited search)
vanishing of adjoint Bloch–Kato Selmer groups	Thorne [58, Thm. A]	special case ($n \leq 4$, ordinary)
local Langlands, depth zero; fixed vectors	[13, 27, 28, 49]; [6, 65]	weak fragment

8.2. **What appears new.** "Appears new" here means: not found in a limited, non-exhaustive search of the literature.

(1) The tower as multiplicity, with a level prime for each word absorbing the order-m constants so that the fixed cost is paid once (§6.3).
(2) Theorem 5.3: the explicit normal form and the collapse of the non-integral part of the theta kernel to one denominator per odd shift, which makes the gain match the odd-power terms of ASK (§5.2). The one-variable mechanism and even its sharp form are known.
(3) (V3) (§7.4): unconditional residually reducible automorphy for all irreducible points, for a non-Schur residual representation without a Steinberg place (closest: [5], conditional); Lemma M; all ℓ at once through one integer per auxiliary prime.
(4) Secondary: the seesaw transport of linear relations among periods; the dimension argument for non-vanishing; the gluing of the base lattice; the covolume bound for each lift; Galois representations through theta lifts; the slope estimates on the Fock side.

8.3. **Possible simplifications of the proof.** Questions 1–4 point to places where the proof might be simplified, each reducing its dependence on one theory.

- *The averaging* (Question 1): fewer helper primes and a smaller host field might reduce the analytic input (A) and the size of the construction, but not the difficulty of the target bound.
- *The covolumes* (Question 2): a direct analytic bound for the archimedean part, uniform in the level, might reduce the use of arithmetic Hilbert–Samuel (whether resolution serves the archimedean or the integral side we have not checked).
- *The amplification* (Question 3): a bound for the tower as a whole, or another amplifying construction, might simplify the per-step bookkeeping; fewer lifts would need a smaller fixed cost.
- *The Galois bound* (Question 4): a bound only at the primes and twists actually consumed might reduce what Act 4, with its Vandiver-type bound and residually reducible automorphy, must supply.

A natural analytic question belongs here too: how weak can Landau's window criterion be (Question 0)?

8.4. **Possible uses beyond the Riemann hypothesis.** We record questions, not claims.

- Does the normal form of Theorem 5.3 have analogues for other formal groups, or for theta functions of other CM types, that give other "odd-part" divisibilities?
- Can the amplification by a tower of theta lifts, with one fixed cost, be used for other one-sided bounds of Chebyshev type?
- Is Lemma M an instance of a general inequality, cotangent space bounded by congruences, without a complete-intersection hypothesis?
- *Vandiver-type bounds.* Off the RH path the development also proves a Vandiver-direction bound over $\mathbb{Q}$, polynomial in e' and uniform in ℓ (§7.4). The note mentions consequences for K-groups of $\mathbb{Z}$; what do these bounds say about K-groups?
- Residually reducible automorphy for non-Schur residual representations without a Steinberg place: how far does the method extend?

8.5. **What experts should examine next.** This is the single list of points that we have not checked; the verification notes refer to it. "Not checked" means not checked by us or by our agents. It does not mean mathematically open: the development or the literature may settle any of these points. We order them by importance for the structure of the proof, roughly in the order in which later steps depend on them; each item points to the section where it arises.

(1) *The theta kernel*: the proof that the kernel has the form $\exp(-\mathcal{E})U$, on which Theorem 5.3 and all of Act 2 rest.

(2) *The tower*: the seesaw laws and the gluing; the inputs to the slope of $\mathcal{N}$ that we have not read; the Hecke non-degeneracy at inert primes; the proof of the joint mean-slope estimate over words and the supply of its Gram-matrix condition (Act 1; [2, §13]); which parameter block each step of the definite chain carries.
(3) *Uniformity in the level*: the covolume bound needs a constant uniform in the level, and we know of no published uniform arithmetic Hilbert–Samuel theorem (cf. [2, §15.14.4]).
(4) *Where* Ш *enters*: the proof of the code's bound for the congruence modules of each lift's Hecke lattice in terms of Ш, whether its local terms are as small as we suggest, and the two unchecked points of the tie (§7.2); uniformity in ℓ (Question 4).
(5) *Residually reducible automorphy*: the proof beyond its outline, in particular that the deformation space is represented and that the classification of reducible points (upper or lower triangular, or $2+1$) is exhaustive, and that the model ring to which Lemma M is applied is the deformation ring; and a check of non-circularity finer than the name-based one.
(6) *Lemma M*: its proof, and whether its hypotheses amount to a complete-intersection condition.
(7) *The sizes*: whether the buffer proportional to m is needed (Question 3); the size assumptions on δ_2 and $|TT|/|H_\kappa|$.
(8) *(V1)–(V3)*: the window selection; the effective Soulé-type bound; χ versus χ^{-1}; the GSp_4 part of the exceptional cases; the roles of the 3 in $\{3,7\}$ and of the condition $p_K \geq 7$.
(9) *Galois representations through theta lifts*: non-circularity of the route.
(10) *Arakelov definitions*: whether the axiomatic height used in the code is the Gillet–Soulé height, and whether a non-normal base is covered by Moriwaki's set-up.

Separate from these are bibliographic details: several theorem numbers were checked only against arXiv or preprint versions; the exact statement behind the Kurihara reference; the bounds quoted from Soulé and de la Torre Durán; and the version of Rubinstein's text.

8.6. **A personal note.** As a mathematician working in analysis rather than number theory, I never expected to find myself reading a proof of the Riemann hypothesis in such a way. Much of the mathematics involved remains beyond my expertise, but exploring its underlying ideas has been a genuinely rewarding experience.

There may be different views on how mathematics should be discovered, proved, and communicated. Whatever one's position on these questions, I find the achievement of constructing a formal proof of such a longstanding problem truly remarkable. I would like to extend my warmest congratulations to Levent Alpöge and express my deep admiration and gratitude to the Lean and Mathlib communities, whose collective efforts have made this possible.

Mathematical discovery, formal verification, and mathematical understanding are distinct endeavors. I wrote this paper simply as one of the many mathematicians who wish to share in the excitement of this moment. I hope that this modest exposition will allow others, regardless of their mathematical background, to experience some of the same fascination.

9. Use of AI tools

The reading of the Lean development, the numerical checks, the literature searches and the drafting of this paper were carried out with AI agents: Anthropic's Claude (Fable 5.2 and Opus 5.5), running in Claude Code as many worker and reviewer agents. Worker agents read parts of the code and wrote reports. Independent reviewer agents, including ChatGPT (OpenAI), re-checked each report against the source, and several claims were corrected or withdrawn as a result. The small Lean project that checks that the Riemann hypothesis follows from the six statements of Appendix A and §7.3 was also written with these agents. "Code-checked" in this paper therefore means checked by these agents, not by a human and not by Lean.

Bibliographic data were taken only from sources that the agents actually retrieved. The human author checked the bibliographic data of part of the references against zbMATH, as marked in our records; theorem numbers were checked by the agents against the sources where we give them, in a few cases only against arXiv or preprint versions. We became aware of the other accounts of the proof only after a first draft of this paper existed; statements that rest on Adriazola's manuscript are cited as such. The human author directed the work, chose its scope and presentation, and is responsible for the content. Errors are likely to remain, and corrections are welcome.

Appendix A. The formal reduction checked in Lean

This appendix records the exact statements of the reduction, as checked in our Lean project. It can be skipped on a first reading.

Definition A.1 (amplified host)**.** A host $(\kappa, (\cdot/q), d)$ is *amplified at* x if there are an integer $m \geq 1$ (the number of steps of the tower), a non-empty set $TT \subset H_\kappa$ of top places, $T_2 \subset H_\kappa$, lengths a_t ($t \in TT$), weights $\beta(u, j, \tau) > 0$ ($u \in H_\kappa$, $j < m$, $\tau \in \Phi$) and reals $b > 0$, η such that: $|H_\kappa| \leq \sqrt{x}\,|T_2|$ and $x^{5/2}(\eta+1) \leq b$; $|a_t/b - x| \leq 1$; thick weights lie in $[b, b+\eta+1]$ and thin ones satisfy $a_t/\beta \geq X'$ with $X' \geq e^{x^{\varepsilon/3}}/3$; and

$$\frac{1}{|H_\kappa|}\sum_{u}\sum_{t\in TT}\sum_{j<m}\Big(\mathrm{ASK}\big(a_t, t, \beta(u,j,\cdot)\big) - \mathrm{MAIN}\big(a_t, \beta(u,j,\cdot)\big)\Big) \leq x^{1/2+C\varepsilon}\,|TT|\,m\,b\,\frac{|T_2|}{|H_\kappa|}.$$

The right side is $x^{1/2+C\varepsilon}$ times the thick mass.

For a number field E, a prime ℓ and $e' \geq 1$, let $Ш^1(E, \ell, e')$ be the kernel of restriction from $H^1(G_{E,S_\ell}, \mathbb{Q}_\ell/\mathbb{Z}_\ell(-e'))$ to the decomposition groups above ℓ, S_ℓ being the places above ℓ. Say $\mathrm{VV}(E, t_V, B)$ holds if these groups are finite for $1 \leq e' < t_V$ and $\sum_\ell \log \ell \sum_{e'<t_V} \text{length}\, Ш^1(E, \ell, e') \leq B$.

Statement A.2 ((V), Galois bound on windows)**.** For every $0 < \varepsilon < 1/52$ and every function $\kappa(x)$ with $x^{2+8\varepsilon} \leq 2^{\kappa(x)} \leq 2x^{2+8\varepsilon}$ there are $J \geq 1$ and X_0 such that for every $X \geq X_0$ and every pair (q, d) of the shape in (A) at scale $X^{1/J}$ there is Y with $X^{1/J} \leq Y$ and $Y^2 \leq X$ such that for every $x \in [Y, Y^2]$ every host field E for $(\kappa(x), q, d)$ satisfies $\mathrm{VV}(E, \lfloor x^{401}\rfloor, \sqrt{\exp(x^\varepsilon)})$.

Statement A.3 ((G), geometric bound)**.** Suppose that for every $0 < \varepsilon < 1/52$ a number $J(\varepsilon) \geq 1$ is given such that there are windows $[Z, Z^{3/2}]$, with Z arbitrarily large, at every point x of which there are $Y \in [x^{1/J(\varepsilon)}, x]$, helper primes (q, d) satisfying the conclusion of (A) on $[Y, x]$, and κ with $x^{2+8\varepsilon} \leq 2^\kappa \leq 2x^{2+8\varepsilon}$, such that every host field E for (κ, q, d) satisfies $\mathrm{VV}(E, \lfloor x^{400}\rfloor, \sqrt{\exp(x^\varepsilon)})$. Then there are $C, \varepsilon_1 > 0$ such that for every $0 < \varepsilon < \varepsilon_1$ and every Y_0 there is $W \geq Y_0$ such that every $x \in [W, W^{3/2}]$ carries $Y \in [x^{1/J(\varepsilon)}, x]$, helper primes (q, d) satisfying the conclusion of (A) on $[Y, x]$, κ with $x^{2+8\varepsilon} \leq 2^\kappa \leq 2x^{2+8\varepsilon}$, and a threshold $X' \geq e^{x^{\varepsilon/3}}/3$, for which the host $(\kappa, (\cdot/q), d)$ is amplified at x with the constant C.

Theorem A.4 (Lean-checked)**.** *(A), (T), (V) and (G) imply the Riemann hypothesis.*

Sketch. At each x of the windows, replace the thick weights by b (the class terms are monotone in the weight, at a cost $(\eta+1)\cdot 2N\log N$, $N = a/b \leq x+1$, which $x^{5/2}(\eta+1) \leq b$ makes negligible) and divide the core inequality by the thick mass. The thick average leaves $2(G_\zeta - G_d) + \mathcal{G}_{\chi_Q} - \mathcal{G}_{\chi_C\chi_Q}$ near x, and the three twisted terms are $O(x^{1/2+\varepsilon})$ by (A). The thin classes are negligible by (T): the host characters are primitive Dirichlet characters of conductor dividing $2^{\kappa+2}qd$, and the zero-free clause of (A) excludes the exceptional real zeros that (T) must avoid. This gives $G_\zeta(x) \leq 2x^{1/2+C\varepsilon} + 6(x+1)^{1/2+\varepsilon} + \log(x+1)(1+\log(x+1)) + 14$, and Theorem 3.1 applies. □

This final form is the one proved in our Lean reduction; the original reaches the equivalent form of §6.4. Together with Theorem 7.5, the six statements (A), (T), (V1), (V2), (V3) and (G) imply the Riemann hypothesis.

References

[1] A. Abbes and T. Bouche, *Théorème de Hilbert–Samuel "arithmétique"*, Ann. Inst. Fourier (Grenoble) **45** (1995), no. 2, 375–401. doi:10.5802/aif.1458 (Cited on pp. 16, 28.)

[2] Jimmie Adriazola, *The Alpöge proof of the Riemann Hypothesis*, manuscript, 287 pp., revised 24 November 2026, https://github.com/jimjamzola/RH-Alpoge-Reconstruction, 2026. (Cited on pp. 2, 10, 13, 15, 16, 17, 19, 22, 24, 25, 26, 27, 28, 30.)

[3] Patrick B. Allen, James Newton, and Jack A. Thorne, *Automorphy lifting for residually reducible l-adic Galois representations, II*, Compositio Math. **156** (2020), no. 11, 2399–2422. doi:10.1112/S0010437X20007484 arXiv:1912.11269 (Cited on pp. 27, 28.)

[4] Levent Alpöge, *Untitled four-page note accompanying the Lean release*, https://alpo.ge/remark.pdf, 2026, 4 pp., no title. (Cited on pp. 1, 17.)

[5] Tobias Berger and Krzysztof Klosin, *On p-adic modularity of abelian threefolds with imaginary multiplication and rational p-isogeny*, 2026. arXiv:2611.19007 (Cited on pp. 27, 28, 29.)

[6] I. N. Bernstein and A. V. Zelevinsky, *Induced representations of reductive* $\mathfrak{p}$*-adic groups. I*, Ann. Sci. École Norm. Sup. **10** (1977), 441–472. doi:10.24033/asens.1333 (Cited on pp. 27, 28.)

[7] Manjul Bhargava, *P-orderings and polynomial functions on arbitrary subsets of Dedekind rings*, Journal für die reine und angewandte Mathematik (1997), 101–128. doi:10.1515/crll.1997.490.101 (Cited on pp. 12, 14.)

[8] ———, *The factorial function and generalizations*, The American Mathematical Monthly **107** (2000), no. 9, 783–799. doi:10.1080/00029890.2000.12005273 (Cited on p. 5.)

[9] Gebhard Böckle, Chandrashekhar B. Khare, and Jeffrey Manning, *Wiles defect for Hecke algebras that are not complete intersections*, Compos. Math. **157** (2021), no. 9, 2046–2088. doi:10.1112/S0010437X21007454 (Cited on p. 26.)

[10] Armand Borel, *Regularization theorems in Lie algebra cohomology. Applications*, Duke Math. J. **50** (1983), 605–623. doi:10.1215/S0012-7094-83-05028-7 (Cited on pp. 25, 28.)

[11] Jean-Benoît Bost, *Périodes et isogénies des variétés abéliennes sur les corps de nombres (d'après D. Masser et G. Wüstholz)*, Séminaire Bourbaki, Vol. 1994/95, Exp. 795, Astérisque, vol. 237, 1996, pp. 115–161. (Cited on pp. 13, 28.)

[12] José Brox, *Distilling the Generalized Riemann Hypothesis*, https://github.com/JoseBrox/GRH_distillation, 2026. (Cited on p. 2.)

[13] Colin J. Bushnell and Guy Henniart, *The essentially tame local Langlands correspondence, I*, J. Amer. Math. Soc. **18** (2005), 685–710. doi:10.1090/S0894-0347-05-00487-X (Cited on pp. 27, 28.)

[14] Alfredo Capelli, *Sur les opérations dans la théorie des formes algébriques*, Math. Ann. **37** (1890), 1–37. doi:10.1007/BF01206702 (Cited on pp. 15, 28.)

[15] Jean-Luc Chabert and Paul-Jean Cahen, *Old problems and new questions around integer-valued polynomials and factorial sequences*, Multiplicative ideal theory in commutative algebra. A tribute to the work of Robert Gilmer, New York, NY: Springer, 2006, pp. 89–108. (Cited on p. 14.)

[16] Huayi Chen, *Positive degree and arithmetic bigness*, 2008. arXiv:0803.2583 (Cited on pp. 16, 28.)

[17] Laurent Clozel, *Motifs et formes automorphes: applications du principe de fonctorialité*, Automorphic forms, Shimura varieties, and L-functions, Vol. I (Ann Arbor, MI, 1988), Perspect. Math., vol. 10, Academic Press, Boston, MA, 1990, pp. 77–159. (Cited on pp. 25, 28.)

[18] Valentino Cristante, *Theta functions and Barsotti–Tate groups*, Ann. Scuola Norm. Sup. Pisa Cl. Sci. (4) **7** (1980), no. 2, 181–215. (Cited on p. 15.)

[19] Alejandro de la Torre Durán, *Perfect forms and torsion bounds in algebraic K-theory over totally real and CM fields*, 2026, v2, 19 November 2026. arXiv:2611.06300 (Cited on p. 26.)

[20] Ehud De Shalit, *Mahler bases and elementary p-adic analysis*, J. Théor. Nombres Bordx. **28** (2016), no. 3, 597–620. doi:10.5802/jtnb.955 (Cited on p. 14.)

[21] Ehud de Shalit and Eran Iceland, *Integer valued polynomials and Lubin-Tate formal groups*, J. Number Theory **129** (2009), no. 3, 632–639. doi:10.1016/j.jnt.2008.10.014 (Cited on p. 14.)

[22] Max Deuring, *Die Typen der Multiplikatorenringe elliptischer Funktionenkörper.*, Abh. Math. Semin. Univ. Hamb. **14** (1941), 197–272 (German). doi:10.1007/BF02940746 (Cited on pp. 14, 28.)

[23] Andrea Dotto and Jack A. Thorne, *On the modularity of residually reducible Galois representations of rank 2*, 2026. arXiv:2610.41764 (Cited on p. 27.)

[24] Jonathan Epstein, *RH (Alpöge) Blueprint*, https://github.com/jae-63/rh-alpoge-blueprint, 2026. (Cited on p. 2.)

[25] Bruce Ferrero and Lawrence C. Washington, *The Iwasawa invariant* μ_p *vanishes for abelian number fields*, Ann. Math. (2) **109** (1979), 377–395. doi:10.2307/1971116 (Cited on pp. 23, 26, 28.)

[26] H. Gillet and C. Soulé, *An arithmetic Riemann–Roch theorem*, Invent. Math. **110** (1992), no. 3, 473–543. doi:10.1007/BF01231343 (Cited on pp. 16, 28.)

[27] Michael Harris and Richard Taylor, *The geometry and cohomology of some simple Shimura varieties*, Annals of Mathematics Studies, vol. 151, Princeton University Press, 2001. doi:10.1515/9781400837205 (Cited on pp. 27, 28.)

[28] Guy Henniart, *Une preuve simple des conjectures de Langlands pour* $\mathrm{GL}(n)$ *sur un corps p-adique*, Invent. Math. **139** (2000), 439–455. doi:10.1007/s002220050012 (Cited on pp. 27, 28.)

[29] T. Honda, *On the theory of commutative formal groups*, J. Math. Soc. Japan **22** (1970), 213–246. doi:10.2969/jmsj/02220213 (Cited on pp. 14, 28.)

[30] Roger Howe, *Remarks on classical invariant theory*, Trans. Amer. Math. Soc. **313** (1989), 539–570. doi:10.1090/S0002-9947-1989-0986027-X (Cited on pp. 15, 28.)

[31] Roger Howe and Tōru Umeda, *The Capelli identity, the double commutant theorem, and multiplicity-free actions*, Math. Ann. **290** (1991), 565–619. doi:10.1007/BF01459261 (Cited on pp. 15, 28.)

[32] Annette Huber and Guido Kings, *Bloch-Kato conjecture and Main Conjecture of Iwasawa theory for Dirichlet characters*, Duke Math. J. **119** (2003), no. 3, 393–464. doi:10.1215/S0012-7094-03-11931-6 (Cited on pp. 21, 23, 28.)

[33] Srikanth B. Iyengar, Chandrashekhar B. Khare, and Jeffrey Manning, *Congruence modules and the Wiles–Lenstra–Diamond numerical criterion in higher codimensions*, Invent. Math. **238** (2024), 769–864. doi:10.1007/s00222-024-01292-y arXiv:2206.08212 (Cited on p. 26.)

[34] H. Jacquet and J. A. Shalika, *On Euler products and the classification of automorphic representations. I*, Am. J. Math. **103** (1981), 499–558. doi:10.2307/2374103 (Cited on pp. 24, 28.)

[35] Manfred Kolster, Thong Nguyen Quang Do, and Vincent Fleckinger, *Twisted S-units, p-adic class number formulas, and the Lichtenbaum conjectures*, Duke Math. J. **84** (1996), no. 3, 679–717. doi:10.1215/S0012-7094-96-08421-5 (Cited on pp. 23, 28.)

[36] V. A. Kolyvagin, *Euler systems*, The Grothendieck Festschrift, Collect. Artic. in Honor of the 60th Birthday of A. Grothendieck. Vol. II, Prog. Math. 87, 435-483 (1990)., 1990. (Cited on pp. 24, 28.)

[37] Masato Kurihara, *Some remarks on conjectures about cyclotomic fields and K-groups of* $\mathbb{Z}$, Compos. Math. **81** (1992), no. 2, 223–236. (Cited on p. 26.)

[38] Serge Lang, *Elliptic functions*, 2nd ed., Graduate Texts in Math., vol. 112, Springer, 1987. doi:10.1007/978-1-4612-4752-4 (Cited on pp. 15, 28.)

[39] ———, *Cyclotomic fields. I and II. With an appendix by Karl Rubin: The main conjecture*, combined 2nd edition ed., Grad. Texts Math., vol. 121, New York etc.: Springer-Verlag, 1990. (Cited on pp. 24, 28.)

[40] Jonathan Lubin and John Tate, *Formal complex multiplication in local fields*, Ann. Math. (2) **81** (1965), 380–387. doi:10.2307/1970622 (Cited on pp. 14, 28.)

[41] B. Mazur and A. Wiles, *Class fields of abelian extensions of* $\mathbb{Q}$, Invent. Math. **76** (1984), 179–330. doi:10.1007/BF01388599 (Cited on pp. 23, 28.)

[42] Hugh L. Montgomery and Robert C. Vaughan, *Multiplicative number theory i: Classical theory*, Cambridge Studies in Advanced Mathematics, vol. 97, Cambridge University Press, 2007. doi:10.1017/CBO9780511618314 (Cited on pp. 8, 28.)

[43] Atsushi Moriwaki, *Continuity of volumes on arithmetic varieties*, J. Algebr. Geom. **18** (2009), no. 3, 407–457. doi:10.1090/S1056-3911-08-00500-6 (Cited on pp. 16, 28.)

[44] David Mumford, *Bi-extensions of formal groups*, Algebraic Geometry (Internat. Colloq., Tata Inst. Fund. Res., Bombay, 1968), Oxford Univ. Press, 1969, pp. 307–322. (Cited on p. 15.)

[45] Bartosz Naskręcki, *The announced proof of the Riemann hypothesis*, https://github.com/nasqret/riemann-rev, https://nasqret.github.io/riemann-rev/atlas/, 2026, Status dashboard, guide and map (riemann-rev), with the atlas "The architecture of the announced proof of the Riemann hypothesis — an illustrated atlas". (Cited on p. 2.)

[46] *The Riemann Hypothesis in Lean: a human-readable blueprint of Alpöge's proof*, draft of 24 November 2026, 2026. (Cited on p. 2.)

[47] Michael Rubinstein, *The Alpöge–Claude proof of the Riemann hypothesis explained for the Riemann zeta function*, Working draft, https://www.math.uwaterloo.ca/~mrubinst/alpoge_claude_grh/zeta-paper.pdf, 2026, November 21, 2026; author line: "Claude Fable 5.2, Codex Astra 6.2, and Gemini Carbon at the behest of Michael O. Rubinstein". (Cited on pp. 2, 9.)

[48] ———, *Alpöge/Claude proof of the Generalized Riemann Hypothesis*, https://www.math.uwaterloo.ca/~mrubinst/alpoge_claude_grh/, 2026. (Cited on p. 2.)

[49] Peter Scholze, *The local Langlands correspondence for* GL_n *over p-adic fields*, Invent. Math. **192** (2013), 663–715. doi:10.1007/s00222-012-0420-5 (Cited on pp. 27, 28.)

[50] Atle Selberg and Sarvadaman Chowla, *On Epstein's zeta-function*, J. Reine Angew. Math. **227** (1967), 86–110. doi:10.1515/crll.1967.227.86 (Cited on pp. 15, 28.)

[51] Jean-Pierre Serre and J. Tate, *Good reduction of abelian varieties*, Ann. Math. (2) **88** (1968), 492–517. doi:10.2307/1970722 (Cited on pp. 14, 28.)

[52] Carl Ludwig Siegel, *Lectures on advanced analytic number theory*, Tata Institute of Fundamental Research, Bombay, 1961. (Cited on pp. 15, 28.)
[53] W. Sinnott, *On the μ-invariant of the Γ-transform of a rational function*, Invent. Math. **75** (1984), 273–282. doi:10.1007/BF01388565 (Cited on pp. 23, 28.)
[54] C. M. Skinner and A. J. Wiles, *Residually reducible representations and modular forms*, Inst. Hautes Études Sci. Publ. Math. **89** (1999), 5–126. doi:10.1007/BF02698855 (Cited on pp. 24, 26, 28.)
[55] Siniša Slijepčević, *Outline of Alpöge's proof of the generalised Riemann hypothesis*, Preprint, hexagon:2611.00812v1, https://hexagonmath.org/abs/2611.00812, 2026, November 21, 2026. (Cited on p. 2.)
[56] Christophe Soulé, *Perfect forms and the Vandiver conjecture*, J. Reine Angew. Math. **517** (1999), 209–221. doi:10.1515/crll.1999.095 (Cited on pp. 23, 26, 28.)
[57] Jack A. Thorne, *Automorphy lifting for residually reducible l-adic Galois representations*, J. Amer. Math. Soc. **28** (2015), no. 3, 785–870. doi:10.1090/S0894-0347-2014-00812-2 (Cited on pp. 27, 28.)
[58] ———, *On the vanishing of adjoint Bloch-Kato Selmer groups of irreducible automorphic Galois representations*, Pure Appl. Math. Q. **18** (2022), no. 5, 2159–2202. doi:10.4310/PAMQ.2022.v18.n5.a5 arXiv:2207.04925 (Cited on pp. 27, 28.)
[59] Preston Wake and Carl Wang-Erickson, *The rank of Mazur's Eisenstein ideal*, Duke Math. J. **169** (2020), no. 1, 31–115. doi:10.1215/00127094-2019-0039 (Cited on p. 26.)
[60] Lawrence C. Washington, *The non-p-part of the class number in a cyclotomic $\mathbf{Z}_p$-extension*, Invent. Math. **49** (1978), no. 1, 87–97. doi:10.1007/BF01399512 (Cited on p. 27.)
[61] ———, *Introduction to cyclotomic fields*, 2nd ed., Graduate Texts in Math., vol. 83, Springer, 1997. doi:10.1007/978-1-4612-1934-7 (Cited on p. 24.)
[62] Hermann Weyl, *Theorie der Darstellung kontinuierlicher halb-einfacher Gruppen durch lineare Transformationen. I*, Math. Z. **23** (1925), 271–309. doi:10.1007/BF01506234 (Cited on pp. 25, 28.)
[63] A. Wiles, *The Iwasawa conjecture for totally real fields*, Ann. Math. (2) **131** (1990), no. 3, 493–540. doi:10.2307/1971468 (Cited on pp. 23, 28.)
[64] Xinyi Yuan, *Big line bundles over arithmetic varieties*, Invent. Math. **173** (2008), no. 3, 603–649. doi:10.1007/s00222-008-0127-9 (Cited on pp. 16, 28.)
[65] A. V. Zelevinsky, *Induced representations of reductive $\mathfrak{p}$-adic groups. II. On irreducible representations of* $\mathrm{GL}(n)$, Ann. Sci. École Norm. Sup. **13** (1980), 165–210. doi:10.24033/asens.1379 (Cited on pp. 27, 28.)
[66] Robin Zhang, *A preliminary note on Alpöge's proposed proof of the generalized Riemann hypothesis*, Preprint, hexagon:2611.00809v2, https://hexagonmath.org/abs/2611.00809v2, 2026, v2, November 22, 2026 (v1, November 21, 2026). (Cited on p. 2.)
[67] Shouwu Zhang, *Positive line bundles on arithmetic varieties*, J. Amer. Math. Soc. **8** (1995), no. 1, 187–221. doi:10.1090/S0894-0347-1995-1254133-7 (Cited on pp. 16, 28.)
[68] Thomas Zink, *The display of a formal p-divisible group*, Astérisque **278** (2002), 127–248. (Cited on p. 15.)

Université Claude Bernard Lyon 1, CNRS, Centrale Lyon, INSA Lyon, Université Jean Monnet, ICJ UMR5208, 69622 Villeurbanne, France
Email address: eto@math.univ-lyon1.fr