hexagon:2611.01245v1

Improved Lower Bounds for the Pair Correlation of Riemann Zeta Zeros via Adaptive Denominator Families

Analytic Number Theory

Tina Rezvanian Ehsan Shah Hosseini

November 26, 2026

We establish an unconditional lower bound for Selberg's prime-variance integral at polynomial lengths, with no assumption on the zeros of the Riemann zeta-function. The additional arithmetic mass is $a_* \approx 4.4548 \times 10^{-5}$, about 23.11 times Aquila's first-range coefficient, and the admissible window exponent is 0.000054 in place of 0.000039 [1]. On the Riemann hypothesis, the same mass keeps Montgomery's pair-correlation function strictly above 1/2 through $\alpha \le 1.0000445503$; with no assumption on the zeros, the corresponding lower bound holds for the complex-difference pair-correlation statistic $F_{T,c}(\alpha)$ uniformly on $1 \le \alpha \le 500000/499973$. The gain comes from partitioning the auxiliary denominators past $\sqrt{X}$ into eight adaptive-roughness bands whose prime-factor floor increases with the band, together with a one-sided structured decomposition $G = I - P + R + E$ of the finite-interval Ramanujan Gram operator. The proof combines the equidistribution estimates of Polymath [10], the shifted divisor bound of Walker [14], the classical-band smooth covariance and mixed zero-shift factorization (Lemma 3) of Aquila [1], and the explicit-formula identities on the line $\mathrm{Re}(s) = 3/2$ of Baluyot, Goldston, Suriajaya, and Turnage-Butterbaugh [2].

Keywords. Riemann zeta-function; pair correlation; Selberg integral; Riemann hypothesis; sieve theory.

1. Introduction

The distribution of the nontrivial zeros of the Riemann zeta-function is tied to the distribution of the primes. Montgomery's pair-correlation function makes that tie quantitative. For $T \ge 2$ and real α it is

$$F_T(\alpha) := \frac{2\pi}{T \log T} \sum_{0<\gamma,\gamma' \le T} T^{i\alpha(\gamma-\gamma')} \frac{4}{4 + (\gamma - \gamma')^2}, \tag{1.1}$$

where γ runs over ordinates of zeros. The same weighted statistic is the point at which pair correlation of zeros is compared with the eigenvalue statistics of random matrices.

Under the Riemann hypothesis the shape of $F_T(\alpha)$ is known on a long initial range: it follows the diagonal prediction up to the threshold $\alpha = 1$. Past that threshold the conjecture is that the function levels off,

$$F_T(\alpha) \sim 1 \qquad (\alpha \ge 1).$$

A lower bound of comparable strength is a different matter. In the RH-only form used here (and in the unconditional complex-difference form), the estimate follows from Aquila [1]; earlier versions in [5, 14] assume respectively GRH for Dirichlet L-functions and the distribution hypothesis AP(27/53):

$$F_T(\alpha) \geq \tfrac{3}{2} - \alpha - o(1), \tag{1.2}$$

meets the value $1/2$ at $\alpha = 1$ and drops below it at once. Keeping a lower bound strictly above $1/2$ on any interval $\alpha > 1$ is therefore its own problem.

The same threshold governs short-interval fluctuations of the primes. Let $\psi(x) := \sum_{n \leq x} \Lambda(n)$ be the Chebyshev function, and for $X \geq 2$ and $1 \leq h \leq X$ write

$$J(X, h) := \int_X^{2X} \big(\psi(x + h) - \psi(x) - h\big)^2 \, dx. \tag{1.3}$$

A lower bound for $J(X, h)$ at lengths $h = X^\kappa$, with $\kappa > 0$ small, transfers to a lower bound for pair correlation on a corresponding interval past $\alpha = 1$. On the Riemann hypothesis the transfer lands on $F_T(\alpha)$. With no hypothesis on the zeros it lands on the complex-difference statistic of Baluyot, Goldston, Suriajaya, and Turnage-Butterbaugh [2],

$$F_{T,\mathrm{c}}(\alpha) := \frac{2\pi}{T \log T} \sum_{0 < \gamma, \gamma' \leq T} T^{\alpha(\rho - \rho')} \frac{4}{4 - (\rho - \rho')^2}, \tag{1.4}$$

which agrees identically with $F_T(\alpha)$ whenever every nontrivial zero has real part $1/2$.

The passage from prime variance to pair correlation is variational. One looks for an auxiliary arithmetic function whose correlation with $\psi(x + h) - \psi(x) - h$ is large compared with its own squared norm. Cauchy–Schwarz then returns a lower bound for $J(X, h)$. Constructing that test function is the problem.

Ramanujan sums give a basis adapted to the arithmetic. A mode of denominator q is orthogonal to the modes of every other denominator, when the orthogonality is taken over a complete set of periods. Correlations can therefore be sorted by q. Orthogonality over complete periods does not control the same correlations on a window $[X, 2X]$ once q is large relative to the length of the window. The denominators that improve (1.2) are large enough for this distinction to matter.

Aquila [1] showed that restricted families of extra denominators, set beside the classical range of small moduli, produce a positive increment of harmonic mass and push the lower bound strictly past $\alpha = 1$. Aquila's first range has

$$(\theta_{\mathrm{Aq},1}, a_{\mathrm{Aq},1}) = \left(0.000039, \ \frac{74}{100000} \frac{2605072188928623487}{10^{21}}\right), \qquad a_{\mathrm{Aq},1} \approx 1.9277534 \times 10^{-6}.$$

A second range in that preprint has $(\theta_2, a_{\mathrm{Aq},2}) = (0.001, 1.8 \times 10^{-15})$, reaching $\alpha \leq 1/(1 - 0.001)$, with a much smaller coefficient.

The construction here changes the shape of that family. The extra denominators are partitioned into eight bands, and the restriction on their prime factors is allowed to depend on the size of the band. A larger denominator may carry a larger smooth part, provided its core is made of correspondingly larger primes. One constraint, applied uniformly, leaves mass on the table. A constraint adapted to the band can collect that mass and still leave the mode under analytic control.

Any such family has to meet three requirements at once. It has to contribute a positive amount of arithmetic mass, or the lower bound does not move. On the finite interval where $J(X, h)$ is measured, its squared norm has to stay comparable to that mass. Its correlations against shifted copies of the prime sequence have to be negligible beside the main term. The three requirements are met by three different arguments: an adaptive choice of the bands, a bound for their Gram

operator, and a dispersion estimate for the shifted correlations. Conductor subtraction belongs only to the third of these. It is a device for proving the dispersion bound, and it is not needed in order to see why the bound should be true.

The resulting improvement is still small on the scale of the conjecture. On the scale of the previous extra mass it is not. The eight bands produce

$$a_* := \frac{4454837383235106954244 0}{10^{27}} \approx 4.4548373832 \times 10^{-5}, \tag{1.5}$$

which is about 23.109 times $a_{\mathrm{Aq},1}$. Its window exponent 0.000054 is $54/39 \approx 1.385$ times Aquila's first-range exponent. Inserted into the classical shape, this is the lower bound

$$F_T(\alpha) \geq \tfrac{3}{2} - (1 - a_*)\alpha - o(1).$$

The right-hand side stays strictly above $1/2$ through $\alpha \leq 1.0000445503$. The theorem that carries the inequality is stated uniformly up to $\alpha_{\max} = 1/(1-\theta_1) \approx 1.000054$, with $\theta_1 = 27/500000$ the longest prime-variance window the argument reaches. Figure 1 draws the leading terms in coordinates that make this gap visible. The raw graph of $F(\alpha)$ near $\alpha = 1$ does not: the whole improvement sits inside an interval of length about 5×10^{-5}.

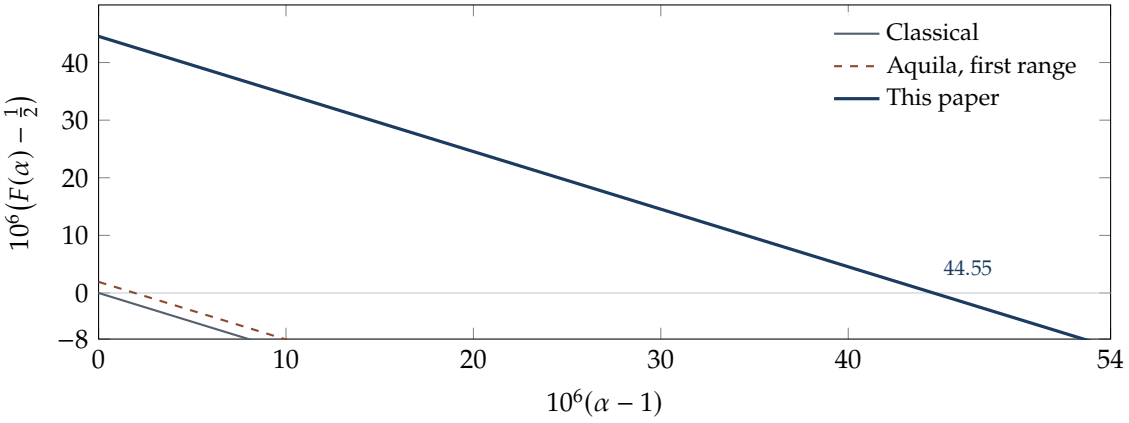

Figure 1: The quantitative improvement, in coordinates $x = 10^6(\alpha - 1)$ and $y = 10^6\big(F(\alpha) - 1/2\big)$. Each curve is the leading term of a lower bound, with the $o(1)$ omitted. Aquila's first range ends at $x = 39$; the present range ends at $x = 54$. The new bound meets $1/2$ at $x \approx 44.55$, that is at $\alpha \approx 1.00004455$.

The mass (1.5) is not a fitted constant. It is the sum of eight integers produced by the certificate `certificates/typeIII_core_certificate.cpp`. Band i contributes $(b_i - l_i) \cdot m_i$ in units of 10^{-27}, where m_i is the certified core mass of that band and $[l_i, b_i]$ is its denominator window in units of 10^{-6}. Dividing by 10^{21} puts the same integers in the units of Figure 2. The eight heights are 0.400697, 2.391894, 4.277787, 5.743036, 6.884742, 7.772775, 8.457916, and 8.619526. Their sum is $a_* \times 10^6$.

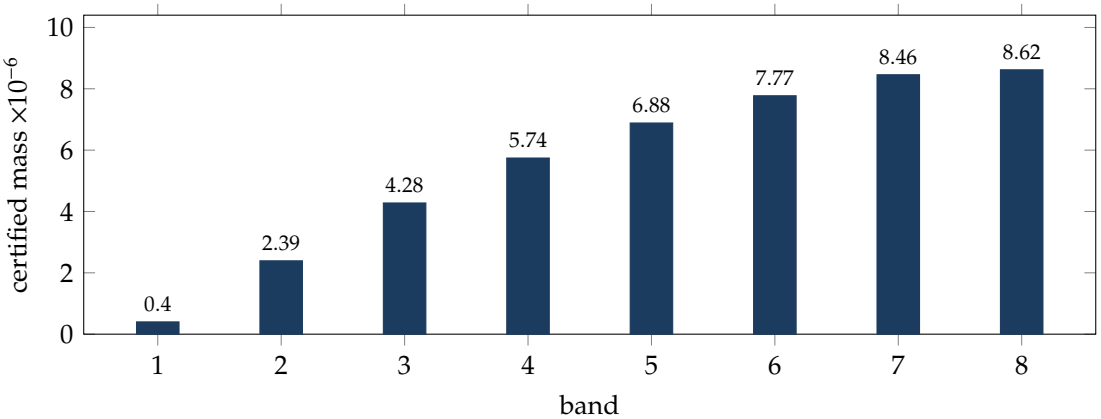

Figure 2: Certified arithmetic mass of the eight bands, in units of 10^{-6}. Bar i is the integer $(b_i - l_i) \cdot m_i$ from the recorded run of `typeIII_core_certificate.cpp`, divided by 10^{21}. Labels are rounded to two decimals; the bar heights are the six-decimal values listed in the text.

The rest of the argument is the verification that this mass survives. Write A_1 for the harmonic mass of the extra bands, G for their Gram operator in the normalized Ramanujan basis, and $\mathcal{E}$ for the aggregate of their correlations against shifted prime sequences. The three requirements become three comparisons. The first is the certificate:

$$A_1 \geq \big(a_* - o(1)\big) \log X. \tag{1.6}$$

The second is the Gram bound, unconditional at the frozen parameters:

$$G \preceq \big(1 + o(1)\big) I. \tag{1.7}$$

The third is the dispersion bound. With $\varepsilon_{\mathrm{d}} = 11/100000$ the fixed completion exponent,

$$\mathcal{E} \ll X^{1-\varepsilon_{\mathrm{d}}/2+o(1)}. \tag{1.8}$$

This comparison is proved in Appendix C and Appendix D, which are propositions of this manuscript. Every hypothesis, intermediate estimate, and source reference is kept. Theorems 1.1 and 1.2 use those propositions together with the auxiliary-norm propositions of Appendix F. They are not conditional on an external obligation. The mass (1.6) and the operator bound (1.7) do not depend on them.

1.1. Main results

The shift window and the pair-correlation window are fixed together with a_*:

$$\theta_1 := \frac{27}{500000} = 0.000054, \qquad \alpha_{\max} := \frac{1}{1-\theta_1} = \frac{500000}{499973} \approx 1.000054002916.$$

Theorem 1.1 (Prime variance at polynomial lengths). *With no assumption of the Riemann hypothesis, uniformly for $1 \leq h \leq X^{\theta_1}$ as $X \to \infty$,*

$$J(X,h) \geq Xh \left\{ \tfrac{1}{2} \log\left(\frac{X}{h^3}\right) + a_* \log X - o(\log X) \right\}. \tag{1.9}$$

In particular, writing $h = X^{\kappa}$ with $0 \leq \kappa \leq \theta_1$,

$$\frac{J(X,h)}{Xh \log X} \geq \tfrac{1}{2} - \tfrac{3}{2}\kappa + 0.0000445483738 - o(1).$$

Proof. The argument is Section 13. It uses the smooth sliding auxiliary $D_{h,\xi}$, so h is real, and invokes Propositions 13.1–13.5, including the classical numerator covariance of Proposition 13.4. The adaptive numerator's principal characters, proper prime powers, low and high conductors, and noncoprime shifts are assembled in Appendix E. Theorem 6.4 is used only for the adaptive square at shift zero. The collar in (3.11) stays fixed until after $X \to \infty$, and is sent to zero only in that later limit. On the enlarged range $0 < |j| \leq H(\log X)^2$ the ratio of the high-conductor error to the principal term is $X^{\theta_1 - \varepsilon_{\mathrm{d}}/2+o(1)} = X^{-10^{-6}+o(1)}$. □

Theorem 1.2 (Pair correlation beyond one). *Uniformly for $1 \leq \alpha \leq \alpha_{\max}$ as $T \to \infty$:*

(i) *Assuming the Riemann hypothesis for* $\zeta(s)$,

$$F_T(\alpha) \geq \tfrac{3}{2} - (1 - a_*)\alpha - o(1). \tag{1.10}$$

The right-hand side stays strictly above $1/2$ *for* $1 \leq \alpha \leq 1.0000445503$, *since* $(1 - a_*)^{-1} \approx 1.00004455$.

(ii) *With no assumption of the Riemann hypothesis,*

$$F_{T,\mathrm{c}}(\alpha) \geq \tfrac{3}{2} - (1 - a_*)\alpha - o(1). \tag{1.11}$$

Proof. The argument is Section 14. It invokes Propositions 14.1, 14.2, and 13.4, together with the five-piece adaptive numerator assembly in Appendix E; Theorem 12.1 supplies only its high-conductor piece. It does not divide the Selberg integral by $XH \log T$. The factor X^{-1} is the square of the amplitude $X^{-1/2}$ in (14.4). The identities taken from [2] are Lemmas 1–3 of that article. The collar stays fixed until after $T \to \infty$. The endpoint $\alpha = \alpha_{\max}$ is included, and the enlarged shifts retain the ratio $X^{-10^{-6}+o(1)}$. □

Remark 1.3 (What is conditional). The $o(\log X)$ term in (1.9) and the $o(1)$ terms in (1.10) and (1.11) are ineffective only because the classical base range uses the Siegel–Walfisz theorem. The extra mass a_* and the Gram bound are effective. On the Riemann hypothesis, $F_{T,\mathrm{c}}(\alpha) = F_T(\alpha)$, so Theorem 1.2(i) is Theorem 1.2(ii) specialized to that case. No hypothesis on Dirichlet L-functions is used.

1.2. Roadmap

The paper is organized around (1.6), (1.7), and (1.8). Section 2 reduces Selberg's integral to the construction of an auxiliary arithmetic sequence. That reduction is elementary and does not use the dispersion estimates. Section 3 builds the sequence from Ramanujan sums and records why a single cutoff past the square root is not enough. Section 4 replaces that cutoff by eight bands whose roughness depends on the size of the denominator. Section 5 turns the mass of those bands into the integer certificate a_*. Section 6 proves that this mass does not inflate the squared norm: the Gram operator stays at most $1 + O(a_{\min}^{-2} X^{-a_{\min}})$. Section 7 builds the compactly supported weight whose Fourier transform decays fast enough that this smoothing does not spend that saving. Section 9 reduces the remaining prime discrepancy to four convolution types. Section 10 assembles the Type I and Type II bounds from the large sieve and two exponential-sum propositions of Polymath. Section 11 prepares the Type III bound through the reduction to a Kloosterman correlation. Section 12 collects the dispersion estimates. Section 13 deduces Theorem 1.1 from the auxiliary-norm propositions. Section 14 deduces Theorem 1.2. Section 15 interprets the construction. The proofs of the shifted correlations, the mixed zero shift, the periodic main terms, the two norm bounds, and the Mellin bridge are Appendix F. The arithmetic mass is the sum drawn in Figure 2. Appendix A collects notation and the frozen parameters.

2. Prime Fluctuations and the Variational Principle

This section reduces Selberg's integral to the construction of one auxiliary arithmetic sequence. The argument uses only the prime number theorem, Cauchy–Schwarz, and counting at the integers. No estimate of prime pairs is assumed.

2.1. The von Mangoldt function

The weight that matches the prime number theorem is logarithmic. The von Mangoldt function is

$$\Lambda(n) = \begin{cases} \log p, & n = p^k \text{ for some prime } p \text{ and some integer } k \geq 1, \\ 0, & \text{otherwise.} \end{cases} \tag{2.1}$$

Chebyshev's function is its partial-sum function,

$$\psi(x) = \sum_{n \leq x} \Lambda(n). \tag{2.2}$$

A prime power p^k with $k \geq 2$ contributes $\log p$, the same amount as the underlying prime. There are $O(\sqrt{x})$ such powers up to x, so they do not change the main term of $\psi(x)$. Keeping them makes the identities below exact rather than exact up to an error of size $\sqrt{x}$.

The prime number theorem asserts that these weights have mean one:

$$\psi(x) \sim x. \tag{2.3}$$

The centered sequence

$$s(n) = \Lambda(n) - 1 \tag{2.4}$$

therefore has a small partial sum,

$$\sum_{n \leq x} s(n) = \psi(x) - \lfloor x \rfloor = o(x).$$

The average of s over $[1, x]$ is understood; the arrangement of s inside short pieces of that interval is the object the later estimates have to control.

2.2. Prime fluctuations in short intervals

For $h > 0$ the centered fluctuation at scale h is

$$\Delta_h \psi(x) = \psi(x + h) - \psi(x) - h. \tag{2.5}$$

The parameter h is the length of the interval on which the primes are being compared with their mean. A short interval is usually empty of primes, so $\Delta_h \psi(x)$ spends most of its time near $-h$ and jumps by $\log p$ when a prime power falls inside. A longer interval adds many such jumps, and the path of $\Delta_h \psi$ varies more slowly.

On the integers the fluctuation is exactly that moving sum. If n and h are positive integers and $x \in [n, n + 1)$, then ψ is constant on $[n, n + 1)$ and on $[n + h, n + h + 1)$, and

$$\psi(n + h) - \psi(n) = \sum_{k=1}^{h} \Lambda(n + k).$$

Subtracting h gives

$$\Delta_h \psi(x) = \sum_{k=1}^{h} s(n + k), \qquad x \in [n, n + 1). \tag{2.6}$$

Selberg's integral measures the square of this fluctuation on a long range of x:

$$J(X, h) = \int_X^{2X} \big(\Delta_h \psi(x)\big)^2 \, dx. \tag{2.7}$$

This is (1.3). For integer endpoints it collapses to a finite sum.

Proposition 2.1 (Discrete form of $J(X,h)$)**.** *Let X and h be positive integers. Then*

$$J(X,h)=\sum_{n=X}^{2X-1}\left(\sum_{k=1}^{h}s(n+k)\right)^2. \tag{2.8}$$

Proof. Split the integral at the integers $n=X,\dots,2X-1$. On each $[n,n+1)$ the integrand equals the square of the inner sum in (2.6), and the piece has length 1. □

The size suggested by the pair-correlation conjecture is $J(X,h)\asymp Xh\log(X/h)$ while h stays away from X. The lower bounds that have been proved are smaller. The classical bound contributing to the introduction has main term $\frac{1}{2}Xh\log(X/h^3)$, and Theorem 1.1 improves a secondary term in that shape. Both statements are asymptotic.

2.3. Arithmetic correlations

Expand the square in (2.8). With the abbreviation $S(n)=\sum_{k=1}^{h}s(n+k)$,

$$S(n)^2=\sum_{1\le a,b\le h}s(n+a)\,s(n+b).$$

The diagonal terms $a=b$ produce $\sum_{a=1}^{h}s(n+a)^2$. The cross terms pair a value of s with a shift of itself. Set $j=b-a$. For each j with $|j|<h$ there are exactly $h-|j|$ pairs (a,b) in $\{1,\dots,h\}^2$.

Normalize the shifted sums on the outer window by its length:

$$C_X(j)=\frac{1}{X}\sum_{X<n\le 2X}s(n)\,s(n+j). \tag{2.9}$$

The sum runs over $n=X+1,\dots,2X$. The variance sum runs over $n=X,\dots,2X-1$ and then reads s at $n+a$ with $1\le a\le h$, so the two windows differ by at most h endpoints. That discrepancy is the remainder in the next proposition.

Proposition 2.2 (Triangular decomposition)**.** *Let $X\ge 2$ and $h\ge 1$ be integers. Then*

$$\frac{J(X,h)}{X}=h\,C_X(0)+2\sum_{j=1}^{h-1}(h-j)\,C_X(j)+R, \tag{2.10}$$

where the boundary remainder satisfies

$$|R|\le 3h^3\frac{\bigl(\log(2X+h)\bigr)^2}{X}. \tag{2.11}$$

In particular $R=0$ when $h=1$.

Proof. Write $L=\log(2X+h)$ and $A_j=\{a:1\le a\le h,\ 1\le a+j\le h\}$, so $|A_j|=h-|j|$. Then

$$J(X,h)=\sum_{|j|<h}\sum_{a\in A_j}\sum_{n=X}^{2X-1}s(n+a)\,s(n+a+j).$$

The inner sum is over the X integers $m\in[X+a,\ 2X+a-1]$. The sum defining $XC_X(j)$ runs over $[X+1,\ 2X]$. These intervals differ by $a-1$ integers at each end. For $X\ge 2$ every argument that appears is at least 2, hence $|s|\le L$ and each product is at most L^2. The comparison of one pair (a,j) therefore costs at most $2(a-1)L^2\le 2(h-1)L^2$. Summing over $a\in A_j$ and over $|j|<h$, and using $\sum_{|j|<h}(h-|j|)=h^2$, bounds the total comparison by $2h^3L^2$.

Dividing by X, this comparison contributes at most $2h^3L^2/X$ to R. The terms j and $-j$ are not yet equal: shifting the sum for $C_X(-j)$ by j places shows that $X|C_X(j) - C_X(-j)| \le 2jL^2$. Replacing $C_X(j) + C_X(-j)$ by $2C_X(j)$ therefore changes J/X by at most

$$\sum_{j=1}^{h-1}(h-j)\cdot\frac{2jL^2}{X} = \frac{h(h^2-1)}{3}\cdot\frac{L^2}{X} \le \frac{h^3L^2}{X}.$$

The two contributions sum to at most $3h^3L^2/X$. If $h = 1$, then $J(X,1) = \sum_{n=X+1}^{2X} s(n)^2 = XC_X(0)$. □

The term $hC_X(0)$ is the diagonal. Since $(\Lambda(n)-1)^2 = \Lambda(n)^2 - 2\Lambda(n) + 1$ and $\sum_{X<p\le 2X}(\log p)^2 \sim X\log X$ by the prime number theorem,

$$C_X(0) \sim \log X.$$

At the finite value $X = 10000$ the same sum is $C_X(0) = 8.540$, while $\log X = 9.210$. The diagonal piece of (2.10) is then about $h\log X$.

The sum over $j \ge 1$ is the off-diagonal. There are $h-1$ shifts, and the weight $h-j$ is largest for the small shifts. A single shift already asks for the joint behavior of $s(n)$ and $s(n+j)$, which is a prime-pair question when both arguments are prime. These correlations are structured and not small: for $X = 10000$ and $1 \le j \le 40$, every odd shift lies near -1 (with mean -0.993), and every even shift is positive (with mean 0.857).

The odd shifts have a direct explanation. If j is odd, one of n and $n+j$ is even. Away from the powers of 2, the even value of s equals -1. The product is then minus the value of s on the odd argument. The odd integers carry essentially all of the mass of Λ, and they are half of the integers, so the average of s on odds is near $+1$. The correlation of an odd shift is therefore near -1. Even shifts do not factor in this way: both arguments can be odd primes at once, and the product $(\log p - 1)(\log q - 1)$ is large and positive. The spikes at multiples of 6 are the finite trace of that coincidence. They are a computation at this X, not a theorem.

A lower bound that required the numerical value of every $C_X(j)$ would be as hard as the original problem. The next subsection replaces that demand by the correlation of s against one sequence that we are free to choose.

2.4. The variational principle

Let W be a nonnegative measurable weight, and define

$$\langle f, g\rangle_W = \int_0^\infty W(x/X)\, f(x)\,\overline{g(x)}\,dx, \qquad \|f\|_W^2 = \langle f, f\rangle_W. \tag{2.12}$$

For an arbitrary auxiliary function T with $\|T\|_W > 0$, Cauchy–Schwarz gives

$$\int_0^\infty W(x/X)\left|\Delta_h\psi(x)\right|^2 dx \ge \frac{\left|\langle \Delta_h\psi,\ T\rangle_W\right|^2}{\|T\|_W^2}. \tag{2.13}$$

Equality holds if and only if T and $\Delta_h\psi$ are linearly dependent on the set where $W > 0$: there is a scalar c such that $T(x) = c\,\Delta_h\psi(x)$ for almost every such x. The optimal choice is the fluctuation itself. It cannot be used as a method, because the numerator and the denominator then contain $J(X,h)$. Any other T produces a lower bound from the correlation of T against $\Delta_h\psi$ and from the norm of T.

The two factors compete. The numerator is large when T tracks the fluctuation. The denominator is large when T is large, whether or not the tracking is good. An auxiliary is useful

when its covariance with the primes is large and its own norm remains comparable to that covariance.

For the integral $J(X,h)$ one takes W to be the indicator of the interval $[1,2]$. On the integers of Proposition 2.1 the same comparison is Cauchy–Schwarz for a pair of vectors. Set

$$S(n)=\sum_{k=1}^{h} s(n+k), \qquad U(n)=\sum_{k=1}^{h} B(n+k)$$

for a real sequence B. If U is not identically zero on $n=X,\dots,2X-1$, then

$$J(X,h)\ge \frac{\left(\sum_{n=X}^{2X-1} S(n)\,U(n)\right)^2}{\sum_{n=X}^{2X-1} U(n)^2}, \tag{2.14}$$

with equality if and only if $S=c\,U$ on that range.

2.5. Reduction to an arithmetic test function

The auxiliary used above is a moving sum of a sequence B. In general one allows a fixed profile w, nonnegative and supported on an interval of bounded length, and sets

$$T(x)=\sum_{n\ge 1} B(n)\,w\!\left(\frac{n-x}{h}\right). \tag{2.15}$$

If w is the indicator of $(0,1]$, then $T(x)=U(x)$ at positive integers x, and (2.14) is (2.13) for that choice. A smoother w changes the weights inside the window and does not change the bookkeeping.

Substitute (2.15) into the covariance $\langle \Delta_h\psi, T\rangle_W$. After the change of variables from the continuous variable x to the integer n at which B is evaluated, the covariance becomes a linear combination, over shifts j, of the sums

$$\sum_{n} W(n/X)\,\Lambda(n)\,B(n+j) \tag{2.16}$$

and of the same sums with Λ replaced by 1. The replacement accounts for the centering in $s=\Lambda-1$. The shift $j=0$ is the correlation of B with Λ along the long interval. Each $j\neq 0$ is the correlation of B with a translate of the prime sequence.

The squared norm $\|T\|_W^2$ expands in the same way. It is a linear combination of

$$\sum_{n} W(n/X)\,B(n)\,B(n+j). \tag{2.17}$$

The term $j=0$ is the sum of squares of B. The ratio in (2.13) stays large only if this sum of squares remains comparable to the square of the diagonal covariance. Each term $j\neq 0$ enlarges the norm without contributing to that diagonal, unless the corresponding covariance happens to be large as well.

The sequence B is therefore required to do three things, and only these three.

(1) Produce a large positive correlation against Λ at shift $j=0$.

(2) Keep $\sum_n W(n/X)\,B(n)^2$ from overwhelming that correlation.

(3) Keep the shifted sums (2.16) and (2.17), for $j\neq 0$, from cancelling the diagonal contribution.

The dispersion estimates of the paper are one approach to the third item. They are not used in this section.

2.6. From the variational principle to Ramanujan sums

What remains is the choice of $B(n)$. The correlations (2.16) and (2.17) are ordinary sums, so B has to be arithmetical enough that those sums can be evaluated. The evaluation has to be easier than a direct prime-pair sum, or the reduction has not helped. On a complete set of periods the summands should be orthogonal, so that the off-diagonal terms in the norm are smaller than the diagonal; on the finite window $[X, 2X]$ the same orthogonality can hold only approximately. Finally, the family of available sequences should be large enough that a stronger lower bound can be reached by admitting more of them.

Ramanujan sums meet this list. The sum of denominator q is periodic of period q, orthogonal over a full period to the sum of every other denominator, and its correlation against Λ is an elementary function of q. The classical construction takes B to be a linear combination of these sums over a restricted set of denominators. Denominators that are too smooth recover only the diagonal main term already visible in $C_X(0)$. Denominators that are too large lose the approximate orthogonality on $[X, 2X]$. The next section makes that classical family explicit. The remainder of the argument changes the restriction, one band of denominators at a time.

The notation for those sums, and the frozen parameters of the later estimates, are collected in Appendix A. None of that appendix is required for the identities of this section.

3. Ramanujan Sums and the Classical Denominator Construction

The variational principle of Section 2 asks for a sequence $B(n)$ with a large diagonal correlation against Λ, a controlled square sum, and small shifted correlations. This section builds that sequence from Ramanujan sums, reconstructs the classical range of denominators, and isolates the obstruction that appears when the denominators approach $\sqrt{X}$.

3.1. From Fourier expansions to Ramanujan sums

On the cyclic group $\mathbb{Z}/q\mathbb{Z}$ the functions

$$e_q(an) = \exp\Big(\frac{2\pi i\, an}{q}\Big), \qquad a \in \mathbb{Z}, \tag{3.1}$$

are periodic of period q in both a and n. They are orthogonal on a complete set of residues: for integers a, b,

$$\frac{1}{q}\sum_{n=1}^{q} e_q\big((a-b)n\big) = \mathbf{1}_{q\mid(a-b)}.$$

An arithmetic problem that is unchanged when n is shifted by a fixed modulus sees these frequencies through rational numbers a/q. Two fractions determine the same frequency precisely when they differ by an integer, so the natural labels are the reduced classes: denominators $q \geq 1$ and numerators a coprime to q. Grouping those classes by the denominator produces the Ramanujan sum

$$c_q(n) = \sum_{\substack{1\leq a\leq q\\(a,q)=1}} e_q(an). \tag{3.2}$$

The sum is real. If $(a, q) = 1$, then $(q - a, q) = 1$, and $e_q((q-a)n)$ is the complex conjugate of $e_q(an)$. The terms pair, except for the real term $a = q$ when it is coprime to q, which occurs only for $q = 1$.

The first few sums are periodic and take only integer values. Directly from (3.2),

$$c_2(n) = \begin{cases} 1, & n \text{ even}, \\ -1, & n \text{ odd}, \end{cases} \qquad c_3(n) = \begin{cases} 2, & 3 \mid n, \\ -1, & 3 \nmid n, \end{cases}$$

$$c_4(n) = \begin{cases} 2, & 4 \mid n, \\ -2, & n \equiv 2 \pmod 4, \\ 0, & n \text{ odd}, \end{cases} \qquad c_6(n) = \begin{cases} 2, & 6 \mid n, \\ -2, & n \equiv 3 \pmod 6, \\ -1, & n \equiv 2, 4 \pmod 6, \\ 1, & n \equiv 1, 5 \pmod 6. \end{cases}$$

The integrality is not special to these moduli: it follows from the divisor formula in the next subsection.

3.2. Elementary arithmetic properties

Recall that $\mu(q)$ vanishes when a square divides q, equals $(-1)^k$ when q is a product of k distinct primes, and equals 1 at $q = 1$. Euler's totient $\varphi(q)$ counts the residues coprime to q.

Proposition 3.1 (Divisor formula). *For every pair of positive integers q and n,*

$$c_q(n) = \sum_{d \mid (q,n)} d\, \mu(q/d). \tag{3.3}$$

In particular $c_q(n)$ is an integer.

Proof. Insert $\sum_{d \mid (a,q)} \mu(d) = \mathbf{1}_{(a,q)=1}$ into (3.2) and write $a = bd$:

$$c_q(n) = \sum_{d \mid q} \mu(d) \sum_{b=1}^{q/d} e_{q/d}(bn).$$

The inner geometric sum equals q/d when $(q/d) \mid n$, and equals 0 otherwise. The change of variables $e = q/d$ therefore retains only the divisors e of $\gcd(q, n)$, with coefficient $e\, \mu(q/e)$. □

Proposition 3.2 (Multiplicativity). *If $(q, r) = 1$, then $c_{qr}(n) = c_q(n)\, c_r(n)$ for every n.*

Proof. Since $(q, r) = 1$, the Chinese remainder theorem identifies $(\mathbb{Z}/qr\mathbb{Z})^\times$ with $(\mathbb{Z}/q\mathbb{Z})^\times \times (\mathbb{Z}/r\mathbb{Z})^\times$. If $a \bmod qr$ corresponds to (a_1, a_2), then $e_{qr}(an) = e_q(a_1 n)\, e_r(a_2 n)$. The sum over the product of the two unit groups factors. □

Two evaluations are immediate. At $n = 0$, or at any multiple of q, every term in (3.2) equals 1, so $c_q(0) = \varphi(q)$. If instead $(n, q) = 1$, the only divisor of $\gcd(q, n)$ is 1, and (3.3) collapses to $c_q(n) = \mu(q)$.

A square factor simplifies the formula still further, by killing it. If a square divides q, then $\mu(q) = 0$, so $c_q(n) = 0$ whenever $(n, q) = 1$. In particular $c_q(p) = 0$ for every prime p that does not divide q. Such a denominator contributes nothing to the value of the auxiliary sequence at large primes. From here on, every denominator family is taken to consist of squarefree integers, unless a single example asks for a non-squarefree modulus such as $q = 4$.

3.3. Periodic orthogonality and shifted correlations

Write $\mathbb{E}_{\mathrm{per}}$ for the average over any complete set of periods. If f is periodic of period M, then $\mathbb{E}_{\mathrm{per}}[f] = M^{-1} \sum_{n=1}^{M} f(n)$, and the value does not depend on which period is used. For $q > 1$ the Ramanujan sum has mean zero: the terms $e_q(an)$ for $a \not\equiv 0 \pmod q$ each average to zero, and $a = q$ is not coprime to q. Equivalently, $c_q(0) = \varphi(q)$ is the sum of the values, not their average.

Proposition 3.3 (Shifted orthogonality). *Let $q, r \geq 1$ and let j be an integer. If M is a common multiple of q and r, then*

$$\frac{1}{M}\sum_{n=1}^{M} c_q(n)\, c_r(n+j) = \mathbf{1}_{q=r}\, c_q(j). \tag{3.4}$$

In particular, the unshifted average is $\mathbb{E}_{\mathrm{per}}[c_q c_r] = \mathbf{1}_{q=r}\, \varphi(q)$.

Proof. Expand both sums by (3.2). The average of $e(n(a/q + b/r))$ over n mod M equals 1 when $a/q + b/r$ is an integer, and equals 0 otherwise, because $M(a/q + b/r)$ is an integer. Thus the left side of (3.4) is

$$\sum_{\substack{(a,q)=1\\(b,r)=1\\ a/q+b/r\in\mathbb{Z}}} e_r(bj).$$

Write $d = \gcd(q, r)$, $q = dq'$, and $r = dr'$, so $\gcd(q', r') = 1$. The condition $a/q + b/r \in \mathbb{Z}$ becomes $ar' + bq' \equiv 0 \pmod{dq'r'}$. Hence $q' \mid a$. But $\gcd(a, q) = 1$, so $q' = 1$. The symmetric argument gives $r' = 1$. Therefore $q = r$, and there is no contribution when $q \neq r$.

When $q = r$, the condition is $a + b \equiv 0 \pmod{q}$. As a runs through the units, so does $b \equiv -a$, and the sum becomes $\sum_{(a,q)=1} e_q(-aj) = c_q(-j)$. The divisor formula shows $c_q(-j) = c_q(j)$. □

The normalized functions

$$f_q(n) = \frac{c_q(n)}{\sqrt{\varphi(q)}} \tag{3.5}$$

therefore satisfy $\mathbb{E}_{\mathrm{per}}[f_q f_r] = \mathbf{1}_{q=r}$. They are an orthonormal set for the periodic average. The average is taken over a whole number of periods. On an interval $(X, 2X]$ whose length is not a multiple of the joint period $\mathrm{lcm}(q, r)$, the same identity is only approximate. The error is the content of Section 3.7.

3.4. Constructing the auxiliary arithmetic function

Return to the sequence B of Section 2. Let $\mathcal{R}$ be a finite set of squarefree positive integers, and define

$$B_{\mathcal{R}}(n) = \sum_{q\in\mathcal{R}} \frac{\mu(q)}{\varphi(q)}\, c_q(n). \tag{3.6}$$

The coefficient is fixed by the value on large primes. If p is a prime that divides no element of $\mathcal{R}$, then $(p, q) = 1$ and $c_q(p) = \mu(q)$, so

$$\frac{\mu(q)\, c_q(p)}{\varphi(q)} = \frac{\mu(q)^2}{\varphi(q)} = \frac{1}{\varphi(q)},$$

where the last step uses that q is squarefree. Every denominator therefore contributes a positive amount at such a prime, and the contributions add:

$$B_{\mathcal{R}}(p) = \sum_{q\in\mathcal{R}} \frac{1}{\varphi(q)}. \tag{3.7}$$

Had the coefficient been 1 rather than $\mu(q)$, the sign of $c_q(p) = \mu(q)$ would have alternated with the number of prime factors of q, and the sum over q would have cancelled. The factor $\mu(q)$ removes that alternation. The factor $1/\varphi(q)$ is the normalization that makes the prime response equal to the periodic square, as the next two displays show.

The arithmetic mass of the family is

$$A(\mathcal{R}) = \sum_{q\in\mathcal{R}} \frac{1}{\varphi(q)}. \tag{3.8}$$

Proposition 3.3 gives the periodic norm exactly. Expanding the square and retaining only the diagonal $q = r$,

$$\mathbb{E}_{\mathrm{per}}\big[B_{\mathcal{R}}^2\big] = \sum_{q\in\mathcal{R}} \frac{\mu(q)^2}{\varphi(q)^2}\,\varphi(q) = A(\mathcal{R}), \tag{3.9}$$

since $\mu(q)^2 = 1$ on $\mathcal{R}$. Combined with (3.7), this is the structural fact of the construction: at every prime outside the denominators, $B_{\mathcal{R}}$ takes the constant value $A(\mathcal{R})$, and the mean square of $B_{\mathcal{R}}$ over a full set of periods is the same constant.

For instance, when $\mathcal{R} = \{q \le 10 : \mu(q) \neq 0\}$, one has $A(\mathcal{R}) = 11/3$ and every prime $p > 10$ satisfies $B_{\mathcal{R}}(p) = 11/3$, whereas $p \in \{2, 3, 5, 7\}$ divides a denominator in the family so (3.7) does not apply.

3.5. The equivalent divisor representation

Substitute (3.3) into (3.6) and collect the coefficient of each divisor of n:

$$B_{\mathcal{R}}(n) = \sum_{q\in\mathcal{R}} \frac{\mu(q)}{\varphi(q)} \sum_{d\mid(q,n)} d\,\mu(q/d) = \sum_{d\mid n} \lambda_d,$$

where

$$\lambda_d = d \sum_{\substack{q\in\mathcal{R}\\ d\mid q}} \frac{\mu(q)\,\mu(q/d)}{\varphi(q)}. \tag{3.10}$$

The two formulas are identities, not approximations. The Ramanujan form (3.6) makes the periodic norm (3.9) immediate, because the summands are orthogonal. The divisor form $B_{\mathcal{R}}(n) = \sum_{d\mid n} \lambda_d$ makes sums of $B_{\mathcal{R}}$ along arithmetic progressions into sums over the moduli d, which is the shape required by a large sieve or a dispersion estimate. Neither form is a substitute for the other.

A complete check at a small family: take $\mathcal{R} = \{2, 3\}$. Then $\varphi(2) = 1$, $\varphi(3) = 2$, and

$$\begin{aligned}
\lambda_1 &= \frac{\mu(2)^2}{\varphi(2)} + \frac{\mu(3)^2}{\varphi(3)} = \frac{3}{2},\\
\lambda_2 &= 2\cdot\frac{\mu(2)\mu(1)}{\varphi(2)} = -2,\\
\lambda_3 &= 3\cdot\frac{\mu(3)\mu(1)}{\varphi(3)} = -\frac{3}{2}.
\end{aligned}$$

No element of $\mathcal{R}$ is divisible by 6, so $\lambda_6 = 0$. The Ramanujan form at $n = 6$ is

$$B(6) = \frac{\mu(2)}{\varphi(2)}c_2(6) + \frac{\mu(3)}{\varphi(3)}c_3(6) = (-1)(1) + \Big(-\frac{1}{2}\Big)(2) = -2,$$

while the divisor form is $\lambda_1 + \lambda_2 + \lambda_3 = 3/2 - 2 - 3/2 = -2$.

3.6. The classical denominator construction

The base band used in this paper is the one in Aquila [1, §2, §7]. Fix a window length H, equal to h for the prime-variance integral and to X/T for the pair-correlation transfer, and fix $\eta > 0$. The denominators are the squarefree integers in the interval $(U, R]$,

$$U = HX^{2\eta}, \qquad R = \sqrt{X/H}\, X^{-10\eta}, \tag{3.11}$$

and the sequence is

$$B_0(n) = \sum_{U<r\le R} \frac{\mu(r)\, c_r(n)}{\varphi(r)}. \tag{3.12}$$

The factor $\mu(r)$ already vanishes when r is not squarefree, so the restriction to squarefree denominators does not change the sum. Its mass is

$$A_0 = \sum_{U<r\le R} \frac{\mu(r)^2}{\varphi(r)}. \tag{3.13}$$

The ratio of the endpoints is elementary:

$$\log(R/U) = \tfrac{1}{2}\log(X/H^3) - 12\eta \log X.$$

Aquila's evaluation of the harmonic sum, from the Dirichlet series

$$\sum_{r\ge1} \frac{\mu(r)^2}{\varphi(r)}\, r^{-s} = \zeta(1+s) \prod_p \big(1 - p^{-1-s}\big)\big(1 + p^{-s}/(p-1)\big)$$

for $\mathrm{Re}(s) > 0$, gives $A_0 = \log(R/U) + O(1)$. The product converges for $\mathrm{Re}(s) > -1/2$ and equals 1 at $s = 0$, so the pole of $\zeta(1+s)$ produces a pure logarithm. That asymptotic is imported from [1, §2]. The identity $\mathbb{E}_{\mathrm{per}}[B_0^2] = A_0$ is (3.9) and is proved here.

The cutoff R has to stay a power of X below $\sqrt{X}$. Then $R^2 \le X^{1-20\eta}$. The finite-interval square of B_0 is the diagonal term of Proposition 13.5: the main term is the periodic mass, and the discrepancy is a Gevrey tail. It is not a count of one remainder per divisor pair. The cutoff U has to stay a power of X above H. The shifts $|j| \le H$ are then long compared with the small conductors that appear when $c_r(n + j)$ is opened, which is where the argument calls on the Siegel–Walfisz theorem. Both constraints are visible in (3.11). Their ratio is what produces the factor $\frac{1}{2}\log(X/H^3)$: once H is as large as $X^{1/3}$, that logarithm ceases to be positive.

The passage from these identities to a lower bound for $J(X, h)$ is Section 13. In [1, §2] the covariance of B_0 against Λ, after the shifts $|j| \le H$ are summed against a smooth window, equals XA_0 up to an error $o(X \log X)$. The error is ineffective, because Siegel–Walfisz is applied on conductors $f \le (\log X)^E$. The finite-interval norm of B_0 is the diagonal term of Proposition 13.5. Cauchy–Schwarz in the form (2.13), followed by sending the collar to zero only after $X \to \infty$, is then the classical bound

$$J(X, h) \ge Xh\Big\{\tfrac{1}{2}\log(X/h^3) - o(\log X)\Big\}$$

for $1 \le h \le X^{1/3-\varepsilon}$. The shape of the logarithm is the computation of $\log(R/U)$ above. The identification of the covariance and of the finite norm with XA_0 is cited from Aquila, not derived here.

Aquila's additional family is not a longer interval of the same kind. Section 6.1 of [1] takes every squarefree core c that is X^δ-densely divisible, lies in the interval $[X^{1/2+\alpha_0}, X^{1/2+\alpha_1}]$, and has all prime factors greater than X^a, and multiplies it by an independent squarefree factor $u \le X^v$. The numerical exponents in that section are

$$a = 0.003, \quad \alpha_0 = 0.00001, \quad \alpha_1 = 0.00073, \quad v = 0.00074, \quad \delta = 0.008. \tag{3.14}$$

The factorization $q = cu$ is unique at these exponents, because every prime factor of u is at most $X^v < X^a$ and every prime factor of c is greater than X^a. The u-cutoff does not depend on c: the family is a rectangle in the (c, u) coordinates. Each such q still contributes $1/\varphi(q)$ at primes $p \sim X$, by (3.7), and

$$\sum_{u \le X^v} \frac{\mu(u)^2}{\varphi(u)} = v \log X + O(1).$$

The core sum is a constant. Aquila's lower bound [1, (6.2)] for that constant is the ratio $2605072188928623487/10^{21}$ displayed next. Multiplied by v, this rectangle contributes

$$a_{\mathrm{Aq},1} = \frac{74}{100000} \frac{2605072188928623487}{10^{21}} \approx 1.9277534 \times 10^{-6},$$

which is the coefficient in Aquila's first theorem, with $\theta_{\mathrm{Aq},1} = 0.000039$. The rectangle gain over the classical band is obtained by stepping past $\sqrt{X}$ with one rough core and one short squarefree factor, not by raising the single cutoff R.

Extending an interval of denominators increases the cumulative mass $\sum_{q \le Q} \mu(q)^2/\varphi(q)$ logarithmically in Q. The next subsection explains why that increase is not free.

3.7. The finite-interval obstruction

Proposition 3.3 averages over a whole number of joint periods. The period of the product $f_q f_r$ divides $\mathrm{lcm}(q, r)$, and $\mathrm{lcm}(q, r) = qr/(q, r) \le qr$. If q and r both lie near $\sqrt{X}$, the product qr is near X or larger. An interval of length X then contains only a fraction of a period, or less than one period, and the average of $f_q(n) f_r(n)$ over $(X, 2X]$ need not be close to $\mathbf{1}_{q=r}$.

The finite-interval Gram matrix records the difference. For squarefree q, r and $X \ge 1$,

$$G_{qr}(X) = \frac{1}{X} \sum_{X < n \le 2X} f_q(n)\, f_r(n). \tag{3.15}$$

The periodic Gram matrix is the identity. The matrix $G(X)$ is what enters the norm of a test function on $(X, 2X]$: if $B = \sum_q \alpha_q f_q$, then

$$\frac{1}{X} \sum_{X < n \le 2X} B(n)^2 = \alpha^{\mathsf{T}} G(X)\, \alpha.$$

For the coefficient vector of (3.6), namely $\alpha_q = \mu(q)/\sqrt{\varphi(q)}$, the periodic value of this quadratic form is $A(\mathcal{R})$. A variational lower bound improves when $A(\mathcal{R})$ grows, but only while $G(X)$ remains close enough to the identity that the quadratic form does not grow faster than $A(\mathcal{R})$. Adding denominators increases the mass and, at the same time, admits pairs whose joint period is long. The second effect can cancel the first.

When q and r approach $\sqrt{X}$, the least common multiple can be comparable with X, so periodic orthogonality alone gives no uniform operator bound for $G(X)$. Controlling the single quadratic form attached to one coefficient vector would be weaker than controlling the operator norm required below. The structured decomposition in Section 6, rather than a finite numerical plot, supplies that operator estimate.

3.8. The design problem

The base band B_0 produces a mass $\frac{1}{2} \log(X/H^3) + O(\eta \log X)$ and stays inside the range where the periodic main term and the Gevrey tail control the finite norm. Aquila's first-range rectangle contributes about $1.9278 \times 10^{-6} \log X$, with one roughness threshold and one small-factor cutoff

applied uniformly. Raising a single cutoff through $\sqrt{X}$ is not enough, because long joint periods must be controlled in operator norm.

The construction problem is to find a second family $\mathcal{R}_1$, disjoint from $(U, R]$, for which all four of the following hold.

(1) The mass $A(\mathcal{R}_1)$ is at least $(a - o(1)) \log X$ for some absolute $a > 0$.

(2) The Gram matrix of the normalized sums f_q, $q \in \mathcal{R}_1$, remains bounded by $1 + o(1)$ in operator norm on the interval $(X, 2X]$.

(3) The shifted correlations of the resulting sequence against $\Lambda(\cdot + j)$ are smaller than the diagonal by a positive power of X.

(4) The arithmetic shape of the family is one to which Walker's Proposition 2.1 applies at a nonzero shift.

Write $B_1 = B_{\mathcal{R}_1}$ and $B = B_0 + B_1$. Disjointness of the two families gives $\mathbb{E}_{\mathrm{per}}[B_0 B_1] = 0$ by Proposition 3.3, so the periodic norms add. The rectangle of [1, §6.1] meets the list with a very small a, because every core is tied to the same cutoff $u \le X^v$. An adaptive family lets that cutoff depend on the size of the core. That is the construction taken up next. The frozen numerical parameters it uses are collected in Appendix A. The construction itself is the next section.

4. Adaptive Denominator Families

Section 3 produced a mass of size $\frac{1}{2}\log(X/H^3)$ from denominators below $\sqrt{X}$, and a much smaller additional mass from one rectangle of denominators past $\sqrt{X}$. This section builds a larger family past the square root. The constraints are introduced before any frozen number, and each one is tied to a later estimate. The eight-band schedule appears only after those constraints are in place.

4.1. Extending beyond the classical range

A squarefree denominator q contributes $1/\varphi(q)$ to the arithmetic mass, and it contributes the same amount at every prime that does not divide q. Denominators larger than $\sqrt{X}$ are attractive for that reason alone: the cumulative mass keeps growing logarithmically as the cutoff increases. The classical band stops near $\sqrt{X/H}$ because that is where the finite-interval Gram matrix is still close to the identity for every coefficient vector, and where the divisor form of B_0 can be counted in arithmetic progressions.

Three difficulties appear together once q exceeds $\sqrt{X}$. The joint period of two such denominators is at most qr and is often as large as X or larger, so periodic orthogonality no longer controls the average on $(X, 2X]$. The same long moduli make a shifted correlation $\sum_n \Lambda(n)B(n + j)$ into a discrepancy over a progression whose modulus is past the square root, which is outside the range of the classical large sieve. And an unrestricted set of all squarefree $q \le Q$ with $Q > \sqrt{X}$ buys logarithmic mass while allowing the Gram spectrum to spread. A larger cutoff is not by itself a construction. The denominators that are kept have to be factored in a way that restores a rough piece, a short piece, and a divisor in every intermediate scale.

4.2. Core and small-factor decomposition

Every denominator in the new family is written

$$q = cu, \tag{4.1}$$

where c is the core and u is a squarefree factor. The symbolic size constraints, for a parameter $b > 0$ and a small $\eta_c > 0$, are

$$X^{1/2+\eta_c} < c \le X^{1/2+b}, \qquad cu \le X^{1/2+b}. \tag{4.2}$$

The second inequality and the lower bound on c give

$$u < X^{b-\eta_c}. \tag{4.3}$$

The core carries the size past $\sqrt{X}$. The factor u is short. One core therefore generates many denominators, namely c times each admissible u, and only the core has to carry the roughness and the dense divisibility imposed below. The short factor is not asked to be rough.

The following integers are an illustration of the bookkeeping, not a model of the production range. Suppose one only required $c > 100$ and $cu \le 1000$. The core $c = 210 = 2 \cdot 3 \cdot 5 \cdot 7$ then admits the squarefree factors $u = 1, 2, 3$, and produces the denominators 210, 420, and 630. The core $c = 330 = 2 \cdot 3 \cdot 5 \cdot 11$ admits only $u = 1, 2, 3$ as well. None of these integers is a core at the scale $X^{1/2}$.

4.3. Roughness

Write $P^-(c)$ for the least prime factor of c, with the convention $P^-(1) = \infty$. The roughness condition is

$$P^-(c) > X^a \tag{4.4}$$

for a parameter $a > 0$. Every prime factor of the core then exceeds X^a. If $a > b$, the bound (4.3) is stronger than needed: $u < X^{b-\eta_c} < X^b < X^a$, so every prime factor of u is strictly smaller than every prime factor of c. In particular

$$(c, u) = 1. \tag{4.5}$$

The factorization $q = cu$ is then unique. Given q, the core is the product of the prime factors of q that exceed X^a, and $u = q/c$.

The same separation is what later estimates spend. A shared prime between two cores is a prime larger than X^a, and there are few such primes dividing a single integer. A divisor of q that omits one prime of c loses a factor larger than X^a and falls below the square root, which is the range where a large sieve applies. Both of these conclusions get weaker if a is decreased, and both get more expensive if a is increased: every core prime must exceed a higher threshold, so fewer cores remain. Roughness is a tradeoff between the number of available cores and the strength of the structural bounds, not a normalization that can be chosen after the fact.

4.4. Dense divisibility

A dispersion estimate factors a modulus near a scale chosen by the bilinear lengths, not near a scale chosen by the integer. Ordinary unique factorization supplies the prime factors. It does not supply a divisor in a short interval of scales. That is a separate property of the integer.

An integer $d \ge 1$ is *one-fold y-densely divisible*, for a real $y \ge 1$, when for every real R with $1 \le R \le d$ there is a divisor $r \mid d$ satisfying $R/y \le r \le R$. In logarithmic coordinates the divisors of d come within $\log y$ of every point of $[0, \log d]$.

Proposition 4.1 (Consecutive divisors). *Let $1 = d_1 < d_2 < \cdots < d_m = d$ be the positive divisors of d in increasing order. Then d is y-densely divisible if and only if $d_{i+1}/d_i \le y$ for every i.*

Proof. Suppose some ratio exceeds y, and set $R = d_{i+1} \cdot y/(y+1)$ or, more simply, choose R with $yd_i < R < d_{i+1}$, which is possible when $d_{i+1} > yd_i$ and one adjusts to keep $R \le d$. Every divisor at most d_i is then $< R/y$, and every divisor at least d_{i+1} is $> R$. Conversely, let r be the largest

divisor that is $\le R$. The next divisor is at most yr. It cannot lie strictly below R, so the next divisor exceeds R and $r \ge R/y$. If there is no next divisor then $r = d \le R$, and $r < R/y$ would force $d < R/y \le d/y$, which is impossible. □

Thus $210 = 2 \cdot 3 \cdot 5 \cdot 7$ is 2-densely divisible: its ordered divisors are

$$1, 2, 3, 5, 6, 7, 10, 14, 15, 21, 30, 35, 42, 70, 105, 210,$$

and the largest consecutive ratio is 2. The integer $46 = 2 \cdot 23$ is not: the step from 2 to 23 has ratio 11.5.

For a squarefree integer the same condition can be read off the primes. Write $c = p_1 p_2 \cdots p_k$ with $p_1 < p_2 < \cdots < p_k$, and write $P_j = p_1 \cdots p_j$.

Proposition 4.2 (Ordered primes). *A squarefree integer $c = p_1 \cdots p_k$, written with increasing primes, is y-densely divisible if and only if $p_1 \le y$ and $p_{j+1} \le y\, P_j$ for every $j = 1, \dots, k-1$.*

Proof. If $p_{j+1} > yP_j$, every divisor that is not composed entirely of $\{p_1, \dots, p_j\}$ is at least p_{j+1}. The largest divisor of P_j is P_j, and the next divisor of c is at least p_{j+1}, so the consecutive ratio exceeds y. Conversely, the hypothesis says that each new prime is at most y times the product of the smaller ones. Any integer R falls between two partial products in the divisor chain, and the prime condition prevents a gap larger than y at the moment a new prime is introduced. The remaining ratios, between divisors that use the same set of primes, are at most the corresponding ratio for a smaller product, and the claim follows by induction on k. □

The certificate uses this test in logarithmic coordinates. If the primes are measured in units of $\log X$, the condition $p_{j+1} \le X^{\delta} P_j$ says that the next exponent is at most δ plus the sum of the preceding exponents. The parameter in the construction is

$$y = X^{\delta}, \tag{4.6}$$

with $\delta > 0$ fixed. A larger δ makes dense divisibility easier to satisfy and makes the divisor r land in a longer interval $[R/y, R]$, which is a weaker input to a dispersion estimate. A smaller δ does the opposite. Roughness pushes prime factors up. Dense divisibility forbids prime factors from jumping too far ahead of the product of the smaller ones. The two conditions pull in opposite directions, and a single pair (a, δ) has to serve every core in the family.

4.5. The single-band construction

Fix one pair (a, b) with $a > b > \eta_c > 0$, and fix $\delta > 0$. The admissible cores are

$$C(a,b) = \Big\{ c \ge 1 : \mu(c)^2 = 1,\ X^{1/2+\eta_c} < c \le X^{1/2+b},\ P^-(c) > X^a,\ c \text{ is } X^{\delta}\text{-densely divisible} \Big\}. \tag{4.7}$$

For each such core the short factors and the denominators are

$$\mathcal{U}(c) = \big\{ u \ge 1 : \mu(u)^2 = 1,\ cu \le X^{1/2+b} \big\}, \tag{4.8}$$

$$\mathcal{R}(a,b) = \big\{ cu : c \in C(a,b),\ u \in \mathcal{U}(c) \big\}. \tag{4.9}$$

Since $a > b$, every such product is squarefree and the factorization $q = cu$ is unique, by (4.5). The arithmetic mass factors over the two pieces:

$$A(a,b) = \sum_{q \in \mathcal{R}(a,b)} \frac{1}{\varphi(q)} = \sum_{c \in C(a,b)} \frac{1}{\varphi(c)} \sum_{u \in \mathcal{U}(c)} \frac{1}{\varphi(u)}. \tag{4.10}$$

The inner sum is a short harmonic sum. Its length depends on c: a core near the top of the range $X^{1/2+b}$ has almost no room for u, while a smaller core has more. That dependence is already an improvement on a rectangle, in which every core is given the same cutoff $u \le X^{v}$.

One pair (a, b) is still rigid. The roughness X^{a} is imposed on every core up to $X^{1/2+b}$, including cores much smaller than that upper bound. A core of size $X^{1/2+\eta_c}$ only needs its primes to exceed the primes of its own short factor, which is shorter than $X^{b-\eta_c}$. Forcing it to meet the roughness required by the largest cores throws those smaller cores away. The mass (4.10) pays for a constraint that only the top of the range requires.

4.6. From fixed roughness to adaptive roughness

Let the denominator range be split at increasing endpoints

$$0 < b_1 < b_2 < \cdots < b_m = b_{\max},$$

and attach to the ith piece a roughness exponent a_i. The structural constraints, still without numerical values, are

$$a_i > b_i + b_{\max} \qquad (1 \le i \le m), \qquad a_{\min} > b_{\max}, \tag{4.11}$$

where $a_{\min} = \min_i a_i$. They are used as follows.

The inequality $a_{\min} > b_{\max}$ is the global separation of primes. Every short factor arising from any band satisfies $u < X^{b_{\max}}$, while every core prime exceeds $X^{a_{\min}}$. The two sets of primes are disjoint, so a short factor from one band is coprime not only to its own core but to every core in every band. Unique recovery of the core, as the $X^{a_{\min}}$-rough part of q, uses this and nothing weaker. The Gram decomposition groups denominators by their cores, and that grouping is well-defined only if the core is a function of q.

The inequality $a_i > b_i + b_{\max}$ is the cross-band constraint. If a core c admissible at level a_i were multiplied by one further prime larger than X^{a_i}, the product would exceed

$$X^{1/2+\eta_c} \cdot X^{a_i} > X^{1/2+b_i+b_{\max}} \ge X^{1/2+b_{\max}}.$$

No such product can itself be a core. Consequently no core divides a different core. The omitted-core bound is the weaker consequence $a_i - b_i > b_{\max}$: deleting one core prime from a denominator of size at most $X^{1/2+b_i}$ leaves a divisor of size at most $X^{1/2+b_i-a_i} < X^{1/2-b_{\max}}$. Divisors that have lost a core prime fall a fixed power below $\sqrt{X}$, which is the input the large sieve needs. Divisors that keep the whole core remain available for the dispersion argument.

Imposing a_m on every band would satisfy (4.11), since $a_m > b_m + b_{\max} \ge b_i + b_{\max}$. It would also delete every core whose least prime factor lies between X^{a_i} and X^{a_m}. Those cores are legal in band i. An adaptive assignment keeps them, and raises the roughness only for the bands whose endpoint b_i is large enough to require it.

4.7. The eight-band architecture

The production schedule uses eight bands. The exact values are

$$\begin{aligned}(b_i, a_i)_{i=1}^{8} = \big(&\tfrac{1}{2000}, \tfrac{9}{2000};\ \tfrac{1}{1000}, \tfrac{1}{200};\ \tfrac{3}{2000}, \tfrac{11}{2000};\ \tfrac{1}{500}, \tfrac{3}{500};\\ &\tfrac{1}{400}, \tfrac{13}{2000};\ \tfrac{3}{1000}, \tfrac{7}{1000};\ \tfrac{7}{2000}, \tfrac{3}{400};\ \tfrac{199}{50000}, \tfrac{1}{125}\big).\end{aligned} \tag{4.12}$$

In decimals, b_i runs through 0.00050, 0.00100, . . . , 0.00350, 0.00398, and a_i runs through 0.00450, 0.00500, . . . , 0.00750, 0.00800. The remaining symbolic parameters are fixed at the same time:

$$\eta_c = \frac{1}{100000}, \qquad \delta = \frac{253}{10000}, \qquad b_{\max} = b_8, \qquad a_{\min} = a_1.$$

A core is assigned to the largest index i for which it meets the roughness, the size, and the dense-divisibility conditions of band i. Explicitly, C_i is the set of squarefree c such that

(i) $X^{1/2+\eta_c} < c \le X^{1/2+b_i}$ and $P^-(c) > X^{a_i}$,

(ii) c is X^{δ}-densely divisible,

(iii) i is maximal with these properties.

The short factors and the denominators are then (4.8) and (4.9) inside band i, and

$$\mathcal{R}_1^{\mathrm{tot}} = \bigcup_{i=1}^{8} \mathcal{R}_i, \qquad \mathcal{R}_i = \mathcal{R}(a_i, b_i) \text{ restricted to cores in } C_i.$$

Maximality makes the sets C_i disjoint. Combined with $a_{\min} > b_{\max}$, it makes the representation $q = cu$ unique across all eight bands: the core is the product of the prime factors of q that exceed $X^{a_{\min}}$.

The same set admits a description by the size of q rather than by the size of c. Set $l_1 = \eta_c$ and $l_i = b_{i-1}$ for $i \ge 2$. The eight intervals $(X^{1/2+l_i}, X^{1/2+b_i}]$ are disjoint and cover $(X^{1/2+\eta_c}, X^{1/2+b_{\max}}]$. A core of logarithmic excess t, meaning $c = X^{1/2+t}$, is rough enough for band i precisely while t is small enough that $P^-(c) > X^{a_i}$ and $c \le X^{1/2+b_i}$. Placing the short factors so that $q = cu$ falls in $(X^{1/2+l_i}, X^{1/2+b_i}]$ reproduces $\mathcal{R}_1^{\mathrm{tot}}$ without repetition. Figure 3 draws the roughness a_i on each of those denominator intervals, against the line $a = t + b_{\max}$.

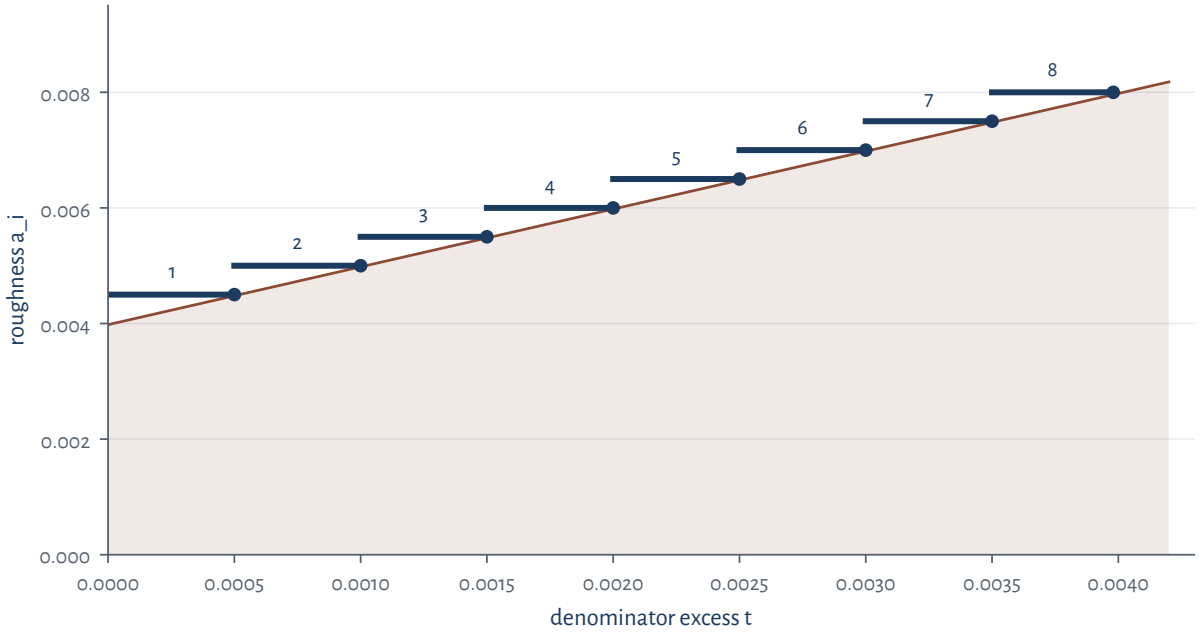

Figure 3: The eight production bands. Segment i is the roughness a_i on the denominator excess interval $(l_i, b_i]$, with $l_1 = \eta_c$ and $l_i = b_{i-1}$. The copper region is forbidden for the right endpoint by $a_i > b_i + b_{\max}$.

The numerical certificate does not sum over this entire set. It keeps only cores up to $X^{1/2+l_i^{\mathrm{cert}}}$, with $l_1^{\mathrm{cert}} = 0.00010$ and $l_i^{\mathrm{cert}} = b_{i-1}$ for $i \ge 2$, and then attaches short factors that place q in $(X^{1/2+l_i^{\mathrm{cert}}}, X^{1/2+b_i}]$. A certified core $c \le X^{1/2+l_i^{\mathrm{cert}}}$ with $P^-(c) > X^{a_i}$ lies in its unique maximal core band C_k with $k \ge i$; hence every attached $u \in \mathcal{U}_i(c)$ gives $cu \in \mathcal{R}_k \subset \mathcal{R}_1^{\mathrm{tot}}$. Thus every counted q lies in $\mathcal{R}_1^{\mathrm{tot}}$, and the counted set is a subset. Because $1/\varphi(q) > 0$, the sum over the subset is a lower bound for A_1. It is not a claim that the omitted denominators are empty.

The inequalities (4.11) hold in exact rational arithmetic. For each i,

$$a_i - b_i - b_{\max} = \begin{cases} 1/50000, & 1 \le i \le 7, \\ 1/25000, & i = 8, \end{cases} \qquad a_i - b_i = \begin{cases} 1/250, & 1 \le i \le 7, \\ 201/50000, & i = 8, \end{cases}$$

and $a_{\min} - b_{\max} = 13/25000$. These identities are exact rational checks on the fractions in (4.12) and do not depend on X.

4.8. Structural lemmas

The lemmas below use only the definitions of this section and the inequalities (4.11), specialized to the eight bands. Each one names the estimate that will call on it.

Lemma 4.3 (Unique decomposition). *Every $q \in \mathcal{R}_1^{\mathrm{tot}}$ has a unique writing $q = cu$ with $c \in \bigcup_i C_i$ and $u \in \mathcal{U}_i(c)$. The integer c is the product of the prime factors of q that exceed $X^{a_{\min}}$. In particular the bands $\mathcal{R}_i$ are pairwise disjoint, and every such q is squarefree.*

Proof. Every short factor is $< X^{b_{\max}} < X^{a_{\min}}$, and every core prime exceeds $X^{a_{\min}}$, so $(c, u) = 1$ and q is squarefree. The core is therefore determined by q. Disjointness of the sets C_i prevents a second writing. □

This is the splitting used in every later sum over q.

Lemma 4.4 (Core antichain). *If $c \in C_i$ and $c' \in C_j$ are distinct, then c does not divide c'. A divisor of a denominator $q \in \mathcal{R}_1^{\mathrm{tot}}$ contains at most one full core, and if it contains one, that core is the $X^{a_{\min}}$-rough part of the divisor.*

Proof. Suppose $c \mid c'$ and $c \neq c'$. Then c'/c is an integer greater than 1, composed of primes dividing c', hence $c'/c > X^{a_j}$. Therefore

$$c' > X^{1/2+\eta_c+a_j} > X^{1/2+b_j+b_{\max}} \geq X^{1/2+b_j},$$

which contradicts $c' \leq X^{1/2+b_j}$. □

The Gram theorem decomposes the matrix by cores. The antichain says that a core appearing in a divisor identifies itself.

Lemma 4.5 (Coefficient bound). *In the divisor form $B_1(n) = \sum_{d \mid n} \lambda_d$ of Section 4.9 below, the coefficients are supported on squarefree $d \leq X^{1/2+b_{\max}}$ and satisfy $|\lambda_d| \ll (\log X)^2$.*

Proof. The support bound is $q \leq X^{1/2+b_{\max}}$. For squarefree coprime d and m, the coefficient (3.10) reduces to

$$\lambda_d = \frac{\mu(d)\, d}{\varphi(d)} \sum_{\substack{dm \in \mathcal{R}_1^{\mathrm{tot}} \\ (m,d)=1}} \frac{\mu(m)^2}{\varphi(m)},$$

because $\mu(dm)\mu(m) = \mu(d)\mu(m)^2$. The sum is at most $\sum_{m \leq X} \mu(m)^2/\varphi(m) \ll \log X$, by the same Dirichlet series used for A_0. The prefactor satisfies $d/\varphi(d) \ll \log\log(3d) \ll \log X$. □

Dispersion estimates sum $|\lambda_d|$ against a discrepancy. The logarithmic bound is the coefficient input. It is not a bound on the discrepancy.

Lemma 4.6 (Omitted core). *Let $q = cu \in \mathcal{R}_i$ and let $d \mid q$. If d is not divisible by every prime factor of c, then*

$$d \leq X^{1/2+b_i-a_i} \leq X^{1/2-1/250}.$$

Proof. Some prime $p \mid c$ does not divide d, and $p > X^{a_i}$, so $d \leq q/p < X^{1/2+b_i-a_i}$. The numerical step is $a_i - b_i \geq 1/250$. □

These divisors are the ones passed to a large sieve. The divisors that keep the whole core stay above $X^{1/2+\eta_c}$ and are the ones passed to dispersion.

Lemma 4.7 (Cross-band separation). *The product of two distinct cores exceeds $X^{1+2\eta_c}$. Hence the least common multiple of two divisors, each containing a full core, exceeds $X^{1+2\eta_c}$.*

Proof. Each core exceeds $X^{1/2+\eta_c}$. The product of two distinct cores is therefore larger than $X^{1+2\eta_c}$, and it divides the least common multiple of any two divisors that contain those cores. □

On the interval $(X, 2X]$ no integer is divisible by two distinct cores. Off-diagonal Gram entries that would require both cores are empty at the scale of X.

Lemma 4.8 (Dense factorization). *Let $d = cu$ with $c \in \mathcal{C}_i$ and $u \in \mathcal{U}_i(c)$. For every real T with $u \le T \le d$ there is a factorization $c = c_1c_2$ such that*

$$r = uc_1, \qquad q = c_2 \tag{4.13}$$

satisfy $T/X^\delta \le r \le T$ and $d = qr$. If $q > 1$, then $P^-(q) > X^{a_i}$. The factor r need not be free of large primes: c_1 may contain primes of the core. For $1 \le T < u$, the product $d = cu$ is still one-fold X^δ-densely divisible, although the rough-complement factorization (4.13) is not asserted at those scales.

Proof. Dense divisibility of c, applied at the scale $T/u \in [1, c]$, supplies a divisor $c_1 \mid c$ with $(T/u)/X^\delta \le c_1 \le T/u$. Set $c_2 = c/c_1$. Then $r = uc_1$ lies in $[T/X^\delta, T]$, and $qr = uc_1c_2 = d$. Every prime factor of c_2 divides c, so it exceeds X^{a_i}. For $T < u$, the integer u is X^{b_i}-smooth and hence X^{b_i}-densely divisible. Since $b_i < \delta$, it is X^δ-densely divisible; Polymath Lemma 2.10(ii)–(iii) then shows that cu is one-fold X^δ-densely divisible at those remaining scales. □

This is the factorization a Type I or Type II dispersion estimate requires of a full-core divisor: a factor r in a prescribed scale, and a complementary factor q that remains rough. Roughness is a property of q, not of r.

4.9. The additional auxiliary sequence

The additional sequence and its mass are

$$B_1(n) = \sum_{q \in \mathcal{R}_1^{\mathrm{tot}}} \frac{\mu(q)}{\varphi(q)} c_q(n), \qquad A_1 = \sum_{q \in \mathcal{R}_1^{\mathrm{tot}}} \frac{1}{\varphi(q)}. \tag{4.14}$$

By the general divisor formula (3.10), $B_1(n) = \sum_{d \mid n} \lambda_d$ with the coefficients of Lemma 4.5. The classical sequence B_0 of (3.12) uses denominators in $(U, R]$. Every new denominator exceeds $X^{1/2+\eta_c}$, while $R < \sqrt{X}$, so the two families are disjoint. Proposition 3.3 gives $\mathbb{E}_{\mathrm{per}}[B_0B_1] = 0$, and the sum

$$B(n) = B_0(n) + B_1(n) \tag{4.15}$$

has periodic mean square $A_0 + A_1$.

What this section establishes is the shape of $\mathcal{R}_1^{\mathrm{tot}}$: unique cores, an antichain, a logarithmic coefficient bound, an omitted-core size bound, separation of distinct cores, and a rough complementary factor at every intermediate scale. What it does not establish is the size of A_1, the operator norm of the finite-interval Gram matrix, or the shifted correlations against Λ. Those are the mass, the Gram bound, and the dispersion bound. The first of them is a lower bound of the shape $(a - o(1)) \log X$. The construction was built so that such a lower bound can be larger than the single rectangle of Section 3. How large it is, for this schedule, is the next question.

5. Arithmetic Mass and the Computational Certificate

The family $\mathcal{R}_1^{\mathrm{tot}}$ of Section 4 is large enough that its mass cannot be read off by listing denominators. This section turns the mass into an integral over prime exponents, replaces that integral by a smaller integer sum, and records the integer the resulting recurrence produces. The coefficient below is the output of that recurrence. It defines the frozen constant a_*. The comparison of the integral with the prime sums is classical, and the independent unbounded-integer replay recorded in `certificates/replay_v1.1.txt` reproduces every production band.

5.1. The quantity to be estimated

Recall A_1 from (4.14). Written as a function of the level X,

$$A_1(X) = \sum_{q \in \mathcal{R}_1^{\mathrm{tot}}} \frac{1}{\varphi(q)}. \tag{5.1}$$

On the periodic average, $\mathbb{E}_{\mathrm{per}}[B_1^2] = A_1(X)$. At every prime p larger than every denominator in the family, $B_1(p) = A_1(X)$. Both facts are Section 3. Once the finite-interval norm of B_1 is known to be at most $(1 + o(1))XA_1(X)$, the variational quotient of Section 2 therefore improves by a multiple of $A_1(X)$. The mass is the potential size of the improvement. It is not yet the improvement.

The goal is a positive constant a_* with

$$\liminf_{X \to \infty} \frac{A_1(X)}{\log X} \geq a_* > 0. \tag{5.2}$$

Listing the denominators at one accessible X cannot prove this. A core is an integer near $X^{1/2}$. The number of candidates up to $X^{1/2+b_{\max}}$ grows like a power of X, and a computation at $X = 10^{12}$ says nothing about the limit. The argument has to be asymptotic, and the asymptotic has to be bounded from below by a calculation that does not depend on X.

5.2. Cores and short factors

By Lemma 4.3, every $q \in \mathcal{R}_1^{\mathrm{tot}}$ is squarefree and has a unique writing $q = cu$ with $(c, u) = 1$. Multiplicativity of φ gives $\varphi(cu) = \varphi(c)\varphi(u)$, and

$$A_1(X) = \sum_{i=1}^{8} \sum_{c \in C_i} \frac{1}{\varphi(c)} \sum_{u \in \mathcal{U}_i(c)} \frac{1}{\varphi(u)}. \tag{5.3}$$

Fix a band and a core of logarithmic excess t, so $c = X^{1/2+t}$ with $\eta_c < t \leq l$, where l is the top of the core window used below. The short factors that place q in a denominator interval of logarithmic length $b_i - l$ run over a range whose logarithmic length is exactly $b_i - l$, independently of t. Mertens' theorem supplies

$$\sum \frac{\mu(u)^2}{\varphi(u)} = (b_i - l) \log X + O(1), \tag{5.4}$$

with the $O(1)$ uniform for cores in the window. Uniformity uses two facts. The logarithmic length $b_i - l$ is a positive constant, the same for every such core. And every short factor that occurs satisfies $1 < u < X^{b_{\max}}$, a range on which the secondary term in Mertens' theorem stays bounded as $X \to \infty$. The bound $O(1)$ is then small beside $(b_i - l) \log X$. If the logarithmic length were allowed to shrink with X, this comparison would fail. It does not shrink here.

A core so large that no short factor fits contributes nothing. Omitting it lowers the sum. After (5.4), the liminf of $A_1(X)/\log X$ is at least $\sum_i (b_i - l_i)$ times a liminf of weighted core masses $\sum_{c \in C_i} 1/\varphi(c)$. The rest of the section estimates those core masses from below.

5.3. Logarithmic coordinates

Write a squarefree core as $c = p_1 \cdots p_m$ with $p_1 < \cdots < p_m$, and set

$$t_j = \frac{\log p_j}{\log X}. \tag{5.5}$$

The size restriction $X^{1/2+\eta_c} < c \le X^{1/2+b_i}$ becomes

$$\tfrac{1}{2} + \eta_c < \sum_{j=1}^{m} t_j \le \tfrac{1}{2} + b_i. \tag{5.6}$$

Roughness $P^-(c) > X^{a_i}$ becomes $t_j > a_i$ for every j. One-fold X^{δ}-dense divisibility, in the ordered-prime form of Proposition 4.2, becomes $t_1 \le \delta$ and

$$t_{j+1} \le \delta + \sum_{\nu=1}^{j} t_\nu. \tag{5.7}$$

An integer factorization problem has become a question about points $(t_1, \dots, t_m)$ in a region of exponent space.

5.4. The integral that lower-bounds the core mass

Mertens' theorem, in the form $\sum_{p \le y} 1/p = \log\log y + C + o(1)$, gives for $0 < \alpha < \beta$

$$\sum_{X^\alpha < p \le X^\beta} \frac{1}{p} = \log\frac{\beta}{\alpha} + o(1) = \int_\alpha^\beta \frac{dt}{t} + o(1). \tag{5.8}$$

On any set of exponents bounded away from 0, the same formula holds with $1/(p-1)$ in place of $1/p$, up to an error $O(\sum 1/p^2)$ that remains bounded. Every prime factor admitted by the construction has exponent at least $a_{\min} = 0.0045$, so the bins used below stay in that range. The $o(1)$ is uniform across those bins as $X \to \infty$.

A squarefree core contributes $\prod_j 1/(p_j - 1)$. Summing over distinct increasing primes and passing to the limit, the core mass is at least the integral of $\prod dt_j/t_j$ over the region cut out by (5.6)–(5.7), up to an error that tends to zero. The integral is over ordered tuples. The same integral over unordered tuples in a single bin of logarithmic length λ is $\lambda^j/j!$, the factor $j!$ undoing the $j!$ orderings of j coordinates. That factorial is the combinatorial multiplicity, not an extra weight.

The inequality points the right way for a lower bound. Every configuration the integral keeps is a limit of actual prime tuples, the passage (5.8) contributes $o(1)$, and every configuration the argument refuses to integrate is a positive quantity left out of the sum. The integral does not upper-bound the mass. A smaller integral, obtained by shrinking the region or by replacing λ with a smaller number, is still a lower bound for the liminf.

5.5. Building the cores one prime at a time

Let $S_j = t_1 + \cdots + t_j$, with $S_0 = 0$. The next prime may be appended only when its exponent t satisfies $t > a_i$, $t > t_j$, and $t \le \delta + S_j$, and only when $S_j + t$ does not already exceed the top of the size window. The state that decides this is the pair (j, S_j), together with the bin that contained t_j, which enforces $t > t_j$. No other information about the tuple is required. Configurations can therefore be summed by a recursion on that state, instead of by listing integers near $X^{1/2}$.

The last prime is treated differently. If the prefix sum τ lies in an interval $[t_0, t_1]$, the final exponent must lie in

$$\big(\max(a_i,\ A-\tau,\ \rho),\ \min(B-\tau,\ \delta+\tau)\big],$$

where ρ is the upper end of the previous bin, A is the bottom of the core window, and B is its top. The integral of dt/t over that interval is at least its length divided by any upper bound for the right endpoint. The certificate uses a length that is positive throughout the whole prefix cell, and a denominator that is at least the right endpoint throughout the cell. Both choices are sufficient for a lower bound. They are not necessary: a core whose final prime only fits for some prefixes inside the cell is discarded.

5.6. Bins, and what is thrown away

Exponent space is partitioned into bins of width $\Delta = 10^{-4}$. In the units 10^{-6} used by the program, the width is the integer $h = 100$. A prime exponent that falls in the bin $[kh,\ (k+1)h)$ is recorded only by the integer k. A prefix of n primes whose bin indices sum to s then has true exponent sum τ somewhere in $[sh,\ (s+n)h]$, in the same units. A transition that appends j primes from bin k is accepted only when the dense-divisibility and size constraints hold for every τ in that range. The integral weight attached to an accepted cell is a number that lower-bounds dt/t everywhere in the cell.

A cell that fails the test contributes zero. So does every tuple with more than 65 prime factors, and every bin that is asked to contribute more than 8 primes at once. Each of those omissions removes a positive part of the integral. None of them can increase the certified value.

5.7. The recurrence

The state is a pair of integers (n, s). Here n is the number of prime factors already chosen, and s is the sum of their bin indices, so the exponent sum in units of 10^{-6} lies in $[sh,\ (s+n)h]$. Let $w(n, s)$ be a lower bound for the total weight of such prefixes, scaled by $Z = 10^{21}$.

Initialization is one empty prefix:

$$w(0,0) = Z, \qquad w(n,s) = 0 \text{ otherwise.}$$

The bins are visited in increasing order of k, starting at $k = a/h$. From a state (n, s) whose bin indices are all strictly less than k, and which meets the dense-divisibility test for bin k, the recurrence appends $j = 1, \dots, 8$ primes of that bin, provided $n + j \le 65$. The new state is $(n + j,\ s + jk)$. If $\lambda = \log((k+1)/k)$, the volume of j unordered coordinates in the bin is $\lambda^j/j!$. The inequality $\log((k+1)/k) > 2/(2k+1)$ gives

$$\frac{\lambda^j}{j!} > \frac{1}{j!}\Big(\frac{2}{2k+1}\Big)^j.$$

The integer weight actually used is the downward recurrence $\gamma_0 = C$ and

$$\gamma_j = \Big\lfloor \gamma_{j-1} \cdot \frac{2}{(2k+1)\,j} \Big\rfloor, \qquad C = 10^{15}, \tag{5.9}$$

so $\gamma_j/C \le (2/(2k+1))^j/j! < \lambda^j/j!$. The updated weight is $\lfloor w(n,s)\,\gamma_j/C \rfloor$, added into $w(n+j,\ s+jk)$.

Within one bin the states are read from those with fewer factors toward those with more, and a newly written state is not read again until a later bin. A prefix is therefore extended only by primes from a strictly higher bin. The same tuple of bin indices cannot be built twice.

When the new prefix still has room for a final prime, for a prefix box $[t_0, t_1]$ and a bin $[l, r)$, the recurrence takes the terminal interval length in units of 10^{-6} to be the minimum of six affine lower bounds valid at every point of the box,

$$\text{width} = \min\{B - A,\ B - t_1 - r,\ B - t_1 - a,\ \delta + 2t_0 - A,\ \delta + t_0 - r,\ \delta + t_0 - a\},$$

and takes the denominator to be the upper bound $\text{den} = \min(B - t_0,\ \delta + t_1)$ for the right endpoint. When $\text{width} > 0$, the recurrence adds

$$\left\lfloor w \cdot \frac{\text{width}}{\text{den}} \right\rfloor$$

to the output, and contributes zero otherwise. Then $\text{width}/\text{den} \le \int dt/t$ throughout the prefix cell.

The same rules, with ordinary integers, are `certificates/mass_certificate.py`. The production program `certificates/typeIII_core_certificate.cpp` is the same recurrence in 128-bit integers. On the cell $(h, A, B, a) = (1000, 10000, 14000, 2000)$, both return

$$460599464028034591918.$$

5.8. Rounding

A floating-point quadrature would not be a certificate: a rounding error of unknown sign can raise the reported mass. Every division in the recurrence above is an integer division that discards the remainder, and every comparison that fails contributes zero. If w lower-bounds the scaled weight of a prefix, then $\lfloor w\gamma_j/C \rfloor$ lower-bounds the scaled weight of the extended prefix, because γ_j/C already lower-bounds the simplex volume. The terminal quotient is rounded in the same direction. Truncating the state space at 65 factors and 8 primes per bin removes terms and does not add them.

The production array for one band has $(65 + 1)(B/h + 2)$ entries. For the largest band, B/h is about $5 \cdot 10^3$, so the array is a few hundred thousand entries of 128-bit integers, a few megabytes. The Python transcription uses unbounded integers. There is also an a priori fixed-width bound. Since rounding down only decreases coefficients, the state generating function is coefficientwise bounded by

$$\sum_{n,s} \frac{\mathrm{dp}[n,s]}{Z} x^s \le \prod_{k\ge1} \exp\left(\frac{2}{2k+1} x^k\right) \le \prod_{k\ge1} \exp(x^k/k) = \frac{1}{1-x}.$$

Hence $\sum_n \mathrm{dp}[n,s] \le Z$ for every s, and in particular each state is at most $Z = 10^{21}$. Every transition product is therefore at most $ZC = 10^{36} < 2^{127} - 1$. There are fewer than $7 \cdot 10^9$ candidate transitions in a production band, $B < 5.1 \cdot 10^5$, every terminal denominator satisfies $\text{den} \ge a \ge 4500$, and the eight band widths sum to 3880. Thus the weighted terminal accumulator is less than

$$\frac{(7 \cdot 10^9)(ZB)(3880)}{4500} < 3.1 \cdot 10^{36} < 2^{127} - 1.$$

Thus signed `__int128` cannot overflow. An independent replay of all eight production bands in unbounded integers is recorded in `certificates/replay_v1.1.txt`. Each core integer and the weighted sum match the frozen table. That replay independently executes the recurrence in arbitrary-precision arithmetic.

5.9. The eight integers

The structural bands of Section 4 and the bands in the certificate are not the same cut. The program uses the endpoints

$$(l_i, b_i) \in \{10^{-6}\} \cdot \{100, 500, 1000, 1500, 2000, 2500, 3000, 3500, 3980\},$$

so $l_1^{\mathrm{cert}} = 0.00010$ and $l_i^{\mathrm{cert}} = b_{i-1}$ for $i \geq 2$, while the structural family begins at $\eta_c = 0.00001$. Each call sums cores whose exponent sum lies in $(0.500010,\ 0.500000 + l_i^{\mathrm{cert}}]$, with roughness a_i on the adaptive run. Every denominator generated from those cores lies in $\mathcal{R}_1^{\mathrm{tot}}$. The cores between $X^{1/2+\eta_c}$ and $X^{1/2+l_1^{\mathrm{cert}}}$, and the cores between $X^{1/2+l_i^{\mathrm{cert}}}$ and $X^{1/2+b_i}$, are not in the sum. They would add a nonnegative amount.

The adaptive run, with bin width 10^{-4}, produced the eight core-mass integers below. The weighted contribution of band i is $(b_i - l_i) \cdot 10^6$ times the core integer. Their sum is exact.

band	$[l_i, b_i] \times 10^6$	core integer	weighted contribution
1	$[100, 500]$	1001742904340234889	400697161736093955600
2	$[500, 1000]$	4783788310407153729	2391894155203576864500
3	$[1000, 1500]$	8555573897250838838	4277786948625419419000
4	$[1500, 2000]$	11486072609517088464	5743036304758544232000
5	$[2000, 2500]$	13769484377416183316	6884742188708091658000
6	$[2500, 3000]$	15545549211557282283	7772774605778641141500
7	$[3000, 3500]$	16915832031587637880	8457916015793818940000
8	$[3500, 3980]$	17957346774472673608	8619526451746883331840
		sum	44548373832351069542440

Dividing by 10^{27} gives the frozen constant

$$a_* = \frac{44548373832351069542440}{10^{27}} \approx 4.4548373832351 \times 10^{-5}. \tag{5.10}$$

Figure 2 shows the eight weighted contributions in units of 10^{-6}. They increase with the band: the largest single contribution is the eighth, 8.620×10^{-6} after rounding, and the first is 0.401×10^{-6}.

5.10. Checks that are not the certificate

Three comparisons were made, and none of them is a substitute for the frozen run.

On the reduced cell $(h, A, B, a) = (1000, 10000, 14000, 2000)$, the Python integers and the C++ program return the same value, 460599464028034591918. The same cell with bin widths $h = 1000$, 500, and 200 returns the integers 460599464028034591918, 535909786793817208375, and 585545392914716797273. Finer bins raise the lower bound. The production certificate does not use those finer bins. It freezes $h = 100$.

The same C++ program with every band's roughness replaced by the single value 0.00810, and with no other change, returns

$$\frac{35145332681619737654860}{10^{27}} \approx 3.514533268 \times 10^{-5}.$$

The adaptive schedule returns the larger integer (5.10). The comparison shows that, inside this recurrence, lowering the roughness on the smaller bands increases the certified mass. Figure 4 illustrates the admissible prime-exponent region that the recurrence discretizes.

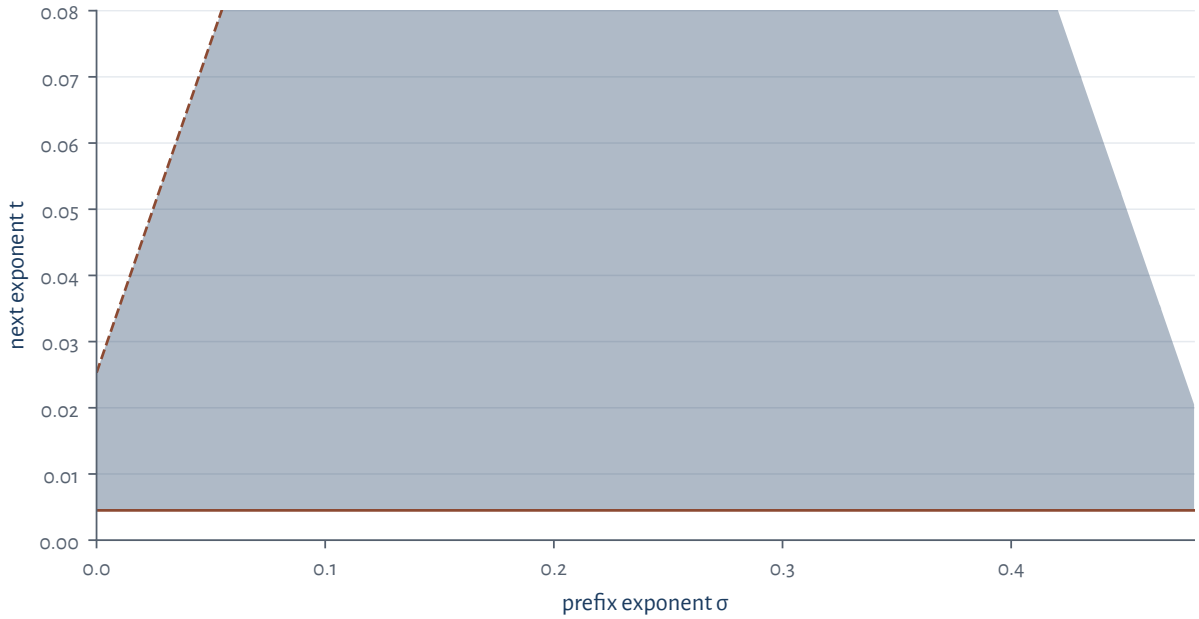

Figure 4: Where the next prime exponent t may sit, given a prefix sum σ. The floor is the production roughness $a = 0.0045$. The rising edge is dense divisibility, $t \le \delta + \sigma$ with $\delta = 0.0253$. The right-hand cliff is the size cap $\sigma + t \le 1/2 + b$ with $b = 0.00050$.

5.11. The mass statement

Theorem 5.1 (Certified adaptive mass). *Let a_* be the integer* (5.10)*. Then*

$$A_1(X) \ge \big(a_* - o(1)\big) \log X. \tag{5.11}$$

The chain is the one this section has followed. The denominator family is the eight-band set of Section 4. Logarithmic coordinates turn a core into a point of exponent space. Mertens' theorem turns the sum over those cores into an integral of $\prod dt_j/t_j$, up to an error $o(1)$. The certificate integrates a smaller region, on which every constraint holds throughout a bin, and rounds every weight downward. The integer it returns is $a_* \times 10^{27}$. Multiplying by the logarithmic length of each short-factor window and sending $X \to \infty$ gives (5.11).

Two cautions belong with the statement. The integer is the output of the frozen program, reproduced by the C++ run of this section and by the unbounded-integer Python replay of all eight production bands in `certificates/replay_v1.1.txt`. And (5.11) is a lower bound for a periodic mean square. It does not by itself improve the variance integral. The squared norm of B_1 on the finite interval $(X, 2X]$ has to stay comparable to $XA_1(X)$. That comparison is the next section.

6. Finite-Interval Orthogonality and the Structured Gram Operator

Section 5 produces a large periodic mean square. The variational principle needs the same control on $(X, 2X]$, where the Ramanujan sums are no longer integrated over complete periods. The argument below does not estimate every off-diagonal entry and then sum them. It splits the Gram matrix into a positive piece, which can be dropped, and two pieces whose operator norms are small for a structural reason.

6.1. Periodic orthogonality on a finite interval

The normalized sums of Section 3 are

$$f_q(n) = \frac{c_q(n)}{\sqrt{\varphi(q)}}.$$

Proposition 3.3 says that their periodic average is exactly the identity:

$$\mathbb{E}_{\mathrm{per}}[f_q f_r] = \mathbf{1}_{q=r}.$$

An observation window of length X is a complete set of periods only when X is a multiple of $\mathrm{lcm}(q, r)$. For the denominators past $\sqrt{X}$ that multiple is typically larger than X, so the identity fails. The failure is not uniform in the pair (q, r): short joint periods are still sampled many times, and long ones may be sampled once or not at all.

Let W be a nonnegative weight on $(0, \infty)$, not identically zero, and write $\widehat{W}(0) = \int_0^\infty W(u)\,du$. The normalized weighted Gram matrix on $\mathcal{R}_1^{\mathrm{tot}}$ is

$$G_{qr} = \frac{1}{X\widehat{W}(0)} \sum_{n\geq 1} W\Big(\frac{n}{X}\Big)\, f_q(n)\, f_r(n). \tag{6.1}$$

The matrix is real and symmetric because $f_q(n)$ is real. It is positive semidefinite because, for any coefficient vector α,

$$\alpha^{\mathsf{T}} G \alpha = \frac{1}{X\widehat{W}(0)} \sum_{n\geq 1} W\Big(\frac{n}{X}\Big) \Big|\sum_q \alpha_q f_q(n)\Big|^2 \geq 0.$$

A bound on each off-diagonal entry does not bound the operator norm. There are about $X^{1/2+b_{\max}}$ denominators. An error of size X^{-c} in every entry can still produce an operator norm as large as $X^{1/2+b_{\max}-c}$, which is useless if c is smaller than $1/2$. The argument has to use cancellation or a block structure, not the crudest entrywise estimate.

6.2. The norm as a quadratic form

The additional sequence is the same linear combination as in Section 3,

$$B_1(n) = \sum_{q\in\mathcal{R}_1^{\mathrm{tot}}} v_q\, f_q(n), \qquad v_q = \frac{\mu(q)}{\sqrt{\varphi(q)}}. \tag{6.2}$$

Its weighted square is the quadratic form of G:

$$\frac{1}{X\widehat{W}(0)} \sum_{n\geq 1} W\Big(\frac{n}{X}\Big)\, B_1(n)^2 = v^{\mathsf{T}} G v. \tag{6.3}$$

The Euclidean norm of this particular vector is the mass. Since $\mu(q)^2 = 1$ on the squarefree family,

$$|v|_2^2 = \sum_q \frac{1}{\varphi(q)} = A_1.$$

If $G \preceq (1 + o(1))I$ in the Loewner order, then $v^{\mathsf{T}} G v \leq (1 + o(1))|v|_2^2$, which is

$$\sum_n W(n/X)\, B_1(n)^2 \leq X\widehat{W}(0)\,(1 + o(1))\,A_1.$$

That is the norm control the variational principle asks for. The largest eigenvalue of G is the operator norm, since G is symmetric and positive semidefinite, and it is the inflation suffered by the worst coefficient vector. An upper bound is enough. A matching lower bound $G \geq (1 - o(1))I$ would say that no vector loses mass. The variance argument never needs that direction.

6.3. Divisor expansion

Substitute $c_q(n) = \sum_{d|(q,n)} d\,\mu(q/d)$ into the product and change variables $n = m[d_1, d_2]$. For $q_1, q_2 \in \mathcal{R}_1^{\mathrm{tot}}$,

$$\sum_{n\ge 1} W\Big(\frac{n}{X}\Big)\, c_{q_1}(n)\, c_{q_2}(n) = \sum_{d_1|q_1} \sum_{d_2|q_2} d_1 d_2\, \mu(q_1/d_1)\, \mu(q_2/d_2)\, S_W([d_1, d_2]), \tag{6.4}$$

where the lattice sum is

$$S_W(L) = \sum_{m\ge 1} W\Big(\frac{mL}{X}\Big). \tag{6.5}$$

This identity is algebraic. No estimate has been used. The periodic mean of the same product is the case of a constant weight, and Proposition 3.3 evaluates it:

$$\sum_{d_1|q_1} \sum_{d_2|q_2} \frac{d_1 d_2\, \mu(q_1/d_1)\, \mu(q_2/d_2)}{[d_1, d_2]} = \mathbb{E}_{\mathrm{per}}[c_{q_1} c_{q_2}] = \mathbf{1}_{q_1=q_2}\, \varphi(q_1). \tag{6.6}$$

Dividing by $\sqrt{\varphi(q_1)\varphi(q_2)}$ produces the identity matrix. The finite-interval matrix G differs from that identity by the extent to which $S_W(L)$ differs from $X\widehat{W}(0)/L$.

6.4. Which pairs of denominators interact

Let $q_1 = c_1 u_1$ and $q_2 = c_2 u_2$, and write $l = [d_1, d_2]$. Section 4 splits the pairs (d_1, d_2) into two regimes.

If d_1 omits a prime factor of c_1, that prime exceeds X^{a_i} and

$$l \le \frac{q_1 q_2}{X^{a_i}} \le X^{1+b_i+b_j-a_i}.$$

The constraint $a_i > b_i + b_{\max}$ is exactly the statement that the exponent on the right is at most $1 - (a_i - b_i - b_{\max}) \le 1 - 1/50000$. The symmetric omission of a prime of c_2, and the case of a prime shared by c_1 and c_2, give the same power of X. In all three situations

$$l \le X^{1-1/50000}. \tag{6.7}$$

This is where the cross-band constraint enters the Gram argument. It is not used to make an entry small by hand. It puts every unwanted least common multiple a fixed power below X, which is the input the Poisson tail needs.

If instead both divisors keep their full cores and the cores are coprime, then

$$l = c_1 c_2 [e_1, e_2] > X^{1+2\eta_c},$$

with $e_s \mid u_s$. A single interval of length X cannot contain a positive multiple of such an l inside the support of a weight that lives on a bounded set of n/X. Adaptive roughness is what makes the cores coprime to every short factor, so this factorization of l is legitimate across bands and not only inside one band. An unrestricted family of denominators has no such split: a pair near $\sqrt{X}$ can have an lcm of size X that falls into neither regime, and the argument stops.

6.5. The decomposition

Theorem 6.1 (Structured decomposition). *Index the matrix by denominators $q = cu$. Define*

$$P_{c_1u_1,\,c_2u_2} = \frac{\mathbf{1}_{u_1=u_2}}{\sqrt{\varphi(c_1)\varphi(c_2)}}, \tag{6.8}$$

$$R_{c_1u_1,\,c_2u_2} = \frac{\mathbf{1}_{u_1=u_2}\,\mathbf{1}_{(c_1,c_2)>1}}{\sqrt{\varphi(c_1)\varphi(c_2)}}. \tag{6.9}$$

There is a symmetric matrix E, whose entries are the Poisson tails written in the next subsection, such that

$$G = I - P + R + E. \tag{6.10}$$

Proof. Start from (6.4) and add and subtract $X\widehat{W}(0)/l$ on every pair (d_1, d_2). The subtracted terms reproduce (6.6), hence the identity matrix after normalization.

The pairs in which both divisors contain the full core and the cores are coprime do not contribute their periodic main term to the actual sum $S_W(l)$, because $l > X^{1+2\eta_c}$ forces every argument ml/X out of the effective support of W. Subtracting a main term that is absent must be compensated. For those pairs the coefficient sum is

$$\mathbf{1}_{(c_1,c_2)=1} \sum_{e_1|u_1} \sum_{e_2|u_2} (e_1, e_2)\, \mu(u_1/e_1)\, \mu(u_2/e_2). \tag{6.11}$$

The sum is multiplicative over the primes dividing u_1u_2. A prime that divides exactly one of u_1, u_2 contributes $-1+1 = 0$. A prime that divides both contributes $1-1-1+p = \varphi(p)$. Therefore (6.11) equals $\mathbf{1}_{u_1=u_2}\, \varphi(u_1)$. Dividing by $\sqrt{\varphi(q_1)\varphi(q_2)} = \sqrt{\varphi(c_1)\varphi(c_2)\varphi(u_1)\varphi(u_2)}$ produces

$$\frac{\mathbf{1}_{u_1=u_2}\, \mathbf{1}_{(c_1,c_2)=1}}{\sqrt{\varphi(c_1)\varphi(c_2)}} = P_{q_1q_2} - R_{q_1q_2}.$$

Removing this contribution subtracts P and adds R. Everything that remains — the difference between $S_W(l)$ and $X\widehat{W}(0)/l$ on the pairs with $l \le X^{1-1/50000}$, and the actual sum on the long pairs, which is not exactly zero for every admissible weight — is the matrix E. □

Bounding $G - I$ in one stroke would treat $-P$ as an error. The next subsection shows that $-P$ points the right way, so its size does not have to be small.

6.6. The missing term is positive

Proposition 6.2 (Positivity of P)**.** *The matrix P is positive semidefinite. Consequently $I - P \preceq I$.*

Proof. For a vector $v = (v_{cu})$,

$$v^{\mathsf{T}} P v = \sum_{u} \left| \sum_{\substack{c \\ cu \in \mathcal{R}_1^{\mathrm{tot}}}} \frac{v_{cu}}{\sqrt{\varphi(c)}} \right|^2 \ge 0.$$

The sum is one square for each short factor. Dropping a positive semidefinite matrix cannot raise a quadratic form, which is the relation $I - P \preceq I$. □

On the six-denominator illustration of Section 6.11, P has eigenvalues 0 with multiplicity four and $1/12$ with multiplicity two. The positive eigenvalues are not small in every normalization one might invent, and they do not need to be. Their sign is the property the argument uses.

6.7. Shared primes

The matrix R is zero unless the two denominators carry the same short factor and their cores share a prime. It is symmetric and entrywise nonnegative, and it is block diagonal in the short factor u.

Proposition 6.3 (Operator norm of R)**.**

$$\|R\|_{\mathrm{op}} \ll a_{\min}^{-2}\, X^{-a_{\min}}.$$

Proof. Apply the Schur test on each block, with positive weights $w(c) = 1/\sqrt{\varphi(c)}$. The resulting bound is

$$\|R\|_{\mathrm{op}} \le \max_{u} \max_{c_1} \sum_{\substack{c_2 u \in \mathcal{R}_1^{\mathrm{tot}} \\ (c_1,c_2)>1}} \frac{1}{\varphi(c_2)}.$$

A core is at most $X^{1/2+b_{\max}}$ and is $X^{a_{\min}}$-rough, so it has at most $(1/2 + b_{\max})/a_{\min} < 112$ prime factors. Fix one such prime $p > X^{a_{\min}}$ and sum over cores divisible by p. Writing $c_2 = pc'$ with c' squarefree and $X^{a_{\min}}$-rough,

$$\sum_{p \mid c_2} \frac{1}{\varphi(c_2)} \le \frac{1}{p-1} \prod_{X^{a_{\min}} < \ell \le X} \Big(1 + \frac{1}{\ell - 1}\Big).$$

The product is $\ll a_{\min}^{-1}$ by Mertens' theorem. Since $p > X^{a_{\min}}$,

$$\frac{1}{p-1} \le \frac{X^{-a_{\min}}}{1 - X^{-a_{\min}}} \ll X^{-a_{\min}}.$$

Summing over the primes of c_1 multiplies by another factor $\ll a_{\min}^{-1}$. □

The separation $a_{\min} > b_{\max}$ is already built into the indexing: short-factor primes and core primes are different, so the condition $(c_1, c_2) > 1$ really is a large prime. The cross-band constraint is not used here. It was used to place the complementary least common multiples below $X^{1-1/50000}$.

6.8. The Poisson tail

On a lattice of step L, the smooth sum and its periodic main term differ by a Fourier series. With $\widehat{W}$ the Fourier transform of W,

$$S_W(L) = \frac{X\widehat{W}(0)}{L} + \frac{X}{L} \sum_{k \ne 0} \widehat{W}\Big(\frac{kX}{L}\Big), \tag{6.12}$$

whenever the formula applies. Two hypotheses cover the weights used in this paper.

(a) W is smooth and supported on a fixed compact subset of $(0, \infty)$.

(b) $W(t) = \exp\big(-(\log t)^2/(2\sigma_0^2)\big)$ for a fixed $\sigma_0 > 0$.

In both cases $\widehat{W}(\xi)$ decays faster than any negative power of $|\xi|$. The explicit Gevrey-2 kernel that produces a weight of type (a), with a stated decay, is postponed to the next section. The Gram bound needs the decay, not the construction of the kernel.

If $L \le X^{1-1/50000}$, every nonzero frequency satisfies $|kX/L| \ge X^{1/50000}$. Rapid decay makes each tail $O_A(X^{-A})$ for every fixed A. If $L > X^{1+2\eta_c}$, a compactly supported weight vanishes identically on the lattice, and the log-Gaussian weight is $O_A(X^{-A})$ because $\log(L/X) \ge 2\eta_c \log X$. Each Gram entry sums $O(X^{o(1)})$ divisor pairs. The matrix has at most $X^{1/2+b_{\max}}$ rows. The Frobenius norm of the resulting error matrix E is therefore $O_A(X^{-A})$ for every A, after adjusting A to absorb the number of rows. Hence $\|E\|_{\mathrm{op}} \ll_A X^{-A}$.

6.9. The operator bound

Theorem 6.4 (Gram bound). *Assume the eight-band constraints of Section 4, and assume W satisfies* (a) *or* (b) *above. Then*

$$G \le \big(1 + O(a_{\min}^{-2} X^{-a_{\min}})\big) I. \tag{6.13}$$

The implied constant depends on W and on the fixed exponents, and it is uniform across the eight bands.

Proof. By Theorem 6.1 and Proposition 6.2,

$$G = I - P + R + E \preceq I + R + E,$$

since subtracting P cannot increase a quadratic form. Proposition 6.3 bounds $\|R\|_{\mathrm{op}}$, and the previous subsection bounds $\|E\|_{\mathrm{op}}$ by a quantity smaller than $X^{-a_{\min}}$ for large X. A symmetric matrix is at most its operator norm times the identity, so

$$R + E \preceq \big(\|R\|_{\mathrm{op}} + \|E\|_{\mathrm{op}}\big) I.$$

The rate that survives is $a_{\min}^{-2} X^{-a_{\min}}$. The bound is one-sided. It does not say that G is within that distance of the identity from below. □

Smoothing at a fixed scale does not change the shape of the bound. If W_+ is a fixed smooth majorant of the indicator of $[1, 2]$, with integral at most $1 + \varepsilon_W$, nonnegativity of the square gives the same upper bound on the unsmoothed interval $[X, 2X]$, multiplied by $1 + \varepsilon_W$. Sending $\varepsilon_W \to 0$ after $X \to \infty$ removes the smoothing loss.

6.10. The norm of B_1, and what is not included

Apply Theorem 6.4 to the coefficient vector (6.2). Then $|v|_2^2 = A_1$ and

$$\sum_{n \geq 1} W\Big(\frac{n}{X}\Big) B_1(n)^2 \leq X \widehat{W}(0)\, A_1 \big(1 + O(a_{\min}^{-2} X^{-a_{\min}})\big). \tag{6.14}$$

Combined with the lower bound $A_1 \geq (a_* - o(1)) \log X$ of Section 5, the additional family contributes a periodic mass of size $a_* \log X$ and a finite-interval square at most $(1 + o(1))$ times that mass. The power $X^{-a_{\min}} = X^{-0.0045}$ is absorbed in the $o(1)$.

The cross term with the classical band is not part of this quadratic form. The families are disjoint and $R < \sqrt{X} < X^{1/2+\eta_c}$, so $\mathbb{E}_{\mathrm{per}}[B_0 B_1] = 0$. The mixed sum and the sums at $j \neq 0$ are not consequences of Theorem 6.4, because that theorem uses positivity at shift zero. The mixed sum is Proposition 13.2, cited from Aquila's Lemma 3 with the hypotheses matched in Appendix F. The nonzero shifts are Proposition 13.1, cited from Walker's Proposition 2.1 with the same matching, together with the periodic main terms of Proposition 13.3.

6.11. A finite illustration

The identities and the operator bound are asymptotic. The following computation checks the algebra on six denominators and shows the scale at which the finite-interval matrix settles. It is not a substitute for Theorem 6.4.

Take the cores $35 = 5 \cdot 7$, $55 = 5 \cdot 11$, and $77 = 7 \cdot 11$, each with short factors 1 and 2. The six denominators are $35, 70, 55, 110, 77, 154$. They share core primes, so R is not zero, and they are coprime to their short factors. At $X = 4000$, with W the indicator of $(X, 2X]$, the largest entry of $G - (I - P + R)$ is 0.0093. The eigenvalues of P are 0, with multiplicity four, and $1/12$, with multiplicity two. As X runs through $250, 500, 1000, 2000, 4000, 8000, 16000$, the largest eigenvalue of G is

$$1.606,\ 1.161,\ 1.165,\ 1.044,\ 1.018,\ 1.016,\ 1.004.$$

The drift toward 1 is not monotone at the two smallest lengths. By $X = 16000$ the excess over 1 is about 0.004.

The same computation with every squarefree denominator up to a cutoff Q, at the single length $X = 400$, separates the mass from the worst-case inflation. At $Q = 6, 10, 15, 20, 30, 40$ the masses are 3.25, 3.67, 4.14, 4.26, 4.73, and 5.05. The excesses of the largest eigenvalue over 1 are 0.016, 0.027, 0.049, 0.073, 0.144, and 0.236. The mass grows steadily. The operator excess grows faster once Q passes $\sqrt{X} = 20$, and it is already moving before that point. This is the same phenomenon as in Section 3.7, read off the largest eigenvalue rather than off one quadratic form.

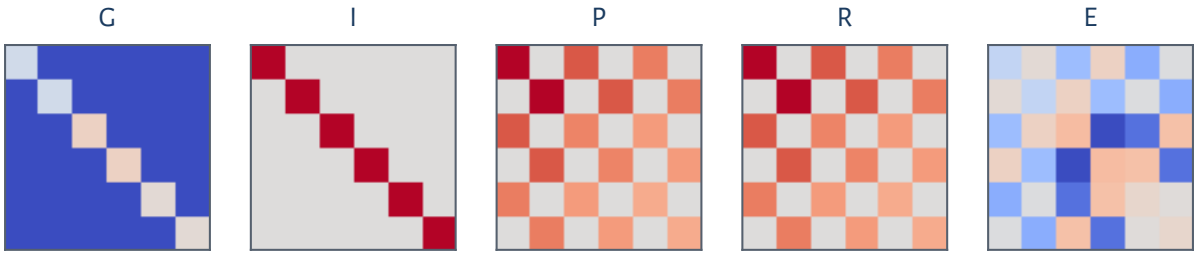

Figure 5: The six-denominator Gram matrix at $X = 4000$, the identity, and the three pieces P, R, and $E = G - (I - P + R)$. Each panel has its own color scale. The denominators are $35, 70, 55, 110, 77, 154$.

Together with Section 5, this section supplies two of the three inputs to the variational principle: a lower bound for the additional mass, and an upper bound for the diagonal square of B_1. The shifted correlations of B_1 against the primes are not controlled by G. They are the dispersion problem. The weight W was assumed to have a rapidly decaying Fourier transform. The next section constructs such a weight.

7. Smooth Observation Weights and the Gevrey-2 Fourier Construction

The Gram bound treats W as given. A sharp cutoff does not qualify: its Fourier transform decays only like $1/|t|$, and the derivatives of a smoothed cutoff of width $X^{-\varepsilon}$ grow like $X^{m\varepsilon}$. The dispersion margins in this argument are powers of X on the scale of 10^{-5}. A fixed loss X^{ε} with ε larger than that margin would erase them. This section builds one explicit weight that is compactly supported, Gevrey-2, and stretched-exponentially small in frequency, with no such loss.

7.1. Why the weight has to be smooth

The sums in Theorem 6.4 are

$$\sum_{n\geq 1} W(n/X)\, f_q(n)\, f_r(n).$$

If $W = \mathbf{1}_{[1,2]}$, the sum is a sharp interval. Opening it by Poisson summation produces a Fourier transform

$$\int_1^2 e(-tu)\, du = e(-3t/2)\, \frac{\sin(\pi t)}{\pi t}$$

when the frequency is normalized in the usual way, and in every normalization the transform decays only like $1/|t|$. Truncating the frequency integral at T then costs $1/T$, and removing that cost forces T to be a power of X.

A smooth compactly supported approximation repairs the decay, but it creates a second cost. If the transition at each endpoint has width $X^{-\varepsilon}$, the mth derivative of the weight is of size $X^{m\varepsilon}$. Dispersion estimates differentiate or integrate by parts against the weight. Each derivative can spend a factor X^{ε}. The margins that remain after the Type I, Type II, and Type III bookkeeping are smaller than a fixed positive ε chosen for convenience. The weight therefore has to meet three conditions at once: compact support, so that there is no infinite tail to estimate; Fourier decay fast enough to truncate at a polylogarithmic frequency; and derivatives that grow only like $(m!)^2$, not like $X^{m\varepsilon}$.

A real-analytic function cannot do this. The identity theorem says that a function analytic on $\mathbb{R}$ and vanishing on an interval vanishes everywhere. Compact support and analyticity are

incompatible unless the function is zero. Ordinary C^∞ compact support is possible, but the usual constructions do not record the derivative growth. The Gevrey class of order 2 is the smallest standard class that allows both.

7.2. Logarithmic coordinates

Set $y = \log u$ and $W(u) = \phi(\log u)$. Fourier inversion in the logarithmic coordinate, for a Schwartz function ϕ, reads

$$W(n/X) = \frac{1}{2\pi} \int_{\mathbb{R}} \widehat{\phi}(t)\, X^{-it}\, n^{it}\, dt, \tag{7.1}$$

up to the normalization $\widehat{\phi}(t) = \int \phi(y)\, e^{-ity}\, dy$. The identity $(mn)^{it} = m^{it} n^{it}$ says that a multiplicative convolution becomes a product of the same phase. A weight inserted into a Dirichlet convolution can therefore be split across the factors, one Mellin frequency at a time. The integral over t still has to be cut. If the cut is a power of X, the phase n^{it} produces derivatives of size $X^{\varepsilon m}$. If the cut is only a power of $\log X$, those derivatives are powers of $\log X$. The kernel below is built so that the tail beyond $C(\log X)^2$ is $O(X^{-A})$ for every fixed A.

7.3. Gevrey regularity of order two

A function f on $\mathbb{R}$ is Gevrey of order $s \geq 1$ when there are constants A, B such that

$$\|f^{(m)}\|_\infty \leq AB^m (m!)^s$$

for every $m \geq 0$. Order $s = 1$ is the real-analytic class, up to the radius encoded by B. Every C^∞ function is not of this form: the right-hand side may grow faster than every exponential of $m \log m$. Order $s = 2$ sits between the two. It is still strong enough that the Fourier transform decays like $\exp(-c\sqrt{|t|})$, and it is weak enough that a nonzero compactly supported example exists. The construction gives the example. The decay is the reason the class is useful; the formula comes next.

7.4. The infinite convolution

The width of the kernel is a fixed $\ell > 0$. It is not the prime-interval length h. Set

$$r_j = \frac{6\ell}{\pi^2 j^2}, \qquad j = 1, 2, \ldots. \tag{7.2}$$

Then $\sum_{j\geq 1} j^{-2} = \pi^2/6$, so $\sum_{j\geq 1} r_j = \ell$. Let

$$\chi_r(y) = \frac{1}{2r}\, \mathbf{1}_{[-r,r]}(y)$$

be the rectangular probability density of half-width r, and write

$$K_{\ell,M} = \chi_{r_1} * \cdots * \chi_{r_M}$$

for the convolution of the first M densities. Each factor is nonnegative and integrates to 1, so the same is true of $K_{\ell,M}$. The support is contained in $[-\sum_{j\leq M} r_j,\ \sum_{j\leq M} r_j] \subseteq [-\ell, \ell]$. The functions are even. As M increases, one more integration against a bounded density raises the regularity by one, while the support stays inside $[-\ell, \ell]$.

The infinite convolution K_ℓ is the limit. Its Fourier transform is the infinite product

$$\widehat{K}_\ell(t) = \prod_{j=1}^{\infty} \frac{\sin(r_j t)}{r_j t}, \tag{7.3}$$

with the value 1 at $t = 0$. The product converges uniformly on compact sets because $\sum r_j^2 < \infty$ and $|\sin u/u| \le 1$. The corresponding distribution is a probability measure supported on $[-\ell, \ell]$. Super-polynomial decay of the product, proved next, implies that the density K_ℓ is C^∞.

7.5. Fourier decay

Proposition 7.1 (Stretched-exponential decay). *Let $c_* = \sqrt{3}\log 2/\pi$. For every real t,*

$$\left|\widehat{K}_\ell(t)\right| \le 2\exp\left(-c_*\sqrt{\ell|t|}\right). \tag{7.4}$$

Proof. The product is even, so take $t > 0$. Set $J = \left\lfloor \sqrt{3\ell t}/\pi \right\rfloor$. If $J = 0$, then $c_*\sqrt{\ell t} < \log 2$, and the right-hand side of (7.4) is at least 1, which dominates the product. If $J \ge 1$, then for every $j \le J$

$$r_j t = \frac{6\ell t}{\pi^2 j^2} \ge \frac{6\ell t}{\pi^2 \cdot 3\ell t/\pi^2} = 2,$$

so each of the first J factors is at most $1/2$. The remaining factors are at most 1. Therefore

$$\left|\widehat{K}_\ell(t)\right| \le 2^{-J} \le 2\cdot 2^{-\sqrt{3\ell t}/\pi} = 2\exp\left(-c_*\sqrt{\ell t}\right),$$

where the middle step uses $J \ge \sqrt{3\ell t}/\pi - 1$. □

The constant is absolute. Once ℓ is fixed, every later truncation is a matter of choosing the frequency large compared with $(\log X)^2$, not large compared with a power of X.

7.6. Derivatives and the frequency tail

Proposition 7.2 (Gevrey-2 bound and polylogarithmic tail). *There are absolute constants $B_0 = 8/c_*^2$ and $C_0 = 8/(\pi c_*^2)$ such that, for every integer $m \ge 0$,*

$$\|K_\ell^{(m)}\|_\infty \le \frac{C_0}{\ell}\left(\frac{B_0}{\ell}\right)^m (m!)^2. \tag{7.5}$$

For each fixed $A > 0$, the choice $C_A = (A+2)^2/(c_*^2\ell)$ gives

$$\int_{|t|>C_A(\log X)^2} \left|\widehat{K}_\ell(t)\right| dt \ll_{A,\ell} X^{-A}. \tag{7.6}$$

Proof. Fourier inversion and Proposition 7.1 give

$$\|K_\ell^{(m)}\|_\infty \le \frac{1}{2\pi}\int_{\mathbb{R}} |t|^m \left|\widehat{K}_\ell(t)\right| dt \le \frac{2}{\pi}\int_0^\infty t^m \exp(-c_*\sqrt{\ell t})\, dt.$$

The substitution $u = c_*\sqrt{\ell t}$ produces

$$\|K_\ell^{(m)}\|_\infty \le \frac{4}{\pi c_*^2 \ell}\left(\frac{1}{c_*^2\ell}\right)^m (2m+1)!.$$

The inequality $(2m+1)! = (2m+1)\binom{2m}{m}(m!)^2 \le 2\cdot 8^m (m!)^2$, which uses $\binom{2m}{m} \le 4^m$ and $2m+1 \le 2^{m+1}$, yields (7.5).

For the tail, set $T_0 = C_A(\log X)^2$, so that $c_*\sqrt{\ell T_0} = (A+2)\log X$. The same substitution gives

$$\int_{|t|>T_0} \left|\widehat{K}_\ell(t)\right| dt \le \frac{8}{c_*^2\ell}\left((A+2)\log X + 1\right) X^{-(A+2)} \ll_{A,\ell} X^{-A}.$$

□

The width ℓ is fixed. It does not tend to zero with X. A width that shrank like $X^{-\varepsilon}$ would put X^ε into B_0/ℓ and would recreate the derivative penalty the construction avoids. The frequency cutoff grows, but only like $(\log X)^2$.

7.7. The aperture

Fix an interval $[y_1, y_2]$ and a collar $\ell > 0$ small enough that $y_2 - y_1 > 2\ell$. Define

$$\phi = \mathbf{1}_{[y_1+\ell,\ y_2-\ell]} * K_\ell. \tag{7.7}$$

Then $0 \le \phi \le 1$, the support of ϕ is contained in $[y_1, y_2]$, and $\phi \equiv 1$ on $[y_1 + 2\ell,\ y_2 - 2\ell]$. Indeed K_ℓ is a probability density supported on $[-\ell, \ell]$, so convolution against it preserves the bounds 0 and 1 and spreads the support by at most ℓ. The Fourier transform satisfies $|\widehat{\phi}(t)| \le (y_2 - y_1)\,|\widehat{K}_\ell(t)|$.

Returning to the multiplicative coordinate, $W(u) = \phi(\log u)$ is smooth and supported on $[e^{y_1}, e^{y_2}]$. Proposition 7.2 truncates the inversion:

$$W(n/X) = \frac{1}{2\pi} \int_{|t| \le C_A (\log X)^2} \widehat{\phi}(t)\, X^{-it}\, n^{it}\, dt + O_{A,\ell}(X^{-A}), \tag{7.8}$$

uniformly in $n \ge 1$. The same construction, with the indicator of any fixed compact subinterval of $(0, \infty)$, produces a weight supported on a slight enlargement of that interval.

The order of limits is fixed. The collar ℓ is chosen first and then held fixed while $X \to \infty$. The tail $O(X^{-A})$ and the polylogarithmic derivative bounds are valid for that fixed collar. After the limit in X, the collar may be sent to zero. Sending it to zero first, or letting it depend on X, is a different argument and is not the one used here.

7.8. Twisted coefficients

Let w be smooth and supported on a fixed compact subinterval of $(0, \infty)$, and set $\widetilde{w}_t(u) = w(u)\, u^{it}$. Leibniz's rule gives

$$\left|\widetilde{w}_t^{(m)}(u)\right| \ll_{m,w} (1 + |t|)^m,$$

because each derivative may fall on w or on the power u^{it}, and a derivative of u^{it} brings a factor of size $|t|/u$. On the support of w, the factor $1/u$ is bounded. Restricting to the frequencies that survive in (7.8),

$$\left|\widetilde{w}_t^{(m)}(u)\right| \ll_{m,w,A} (\log X)^{2m}. \tag{7.9}$$

Polymath's smooth coefficient class asks for bounds of this shape: derivatives controlled by a fixed power of a parameter that is here only $\log X$. A dyadic decomposition of a convolution factor can then use ordinary intervals of ratio 2. The smooth weight has already removed the sharp edge, so there is no boundary layer of relative width $(\log X)^{-A}$ to be cut away and estimated separately. The factors that remain smooth are the ones carrying w. An arbitrary bounded arithmetic coefficient is not claimed to satisfy (7.9). Only the factor that has been multiplied by the aperture is.

7.9. The two-variable kernel

The passage from prime variance to pair correlation integrates a shift $u = j/H$ against a frequency window g and a spatial cutoff b. The resulting kernel is

$$K(\beta, u) = a(\beta)\, b(\beta)\, \widehat{g}\left(-\frac{u}{2\pi\beta}\right), \tag{7.10}$$

where $\widehat{g}(\xi) = \int g(v)\, e^{-2\pi i v\xi}\, dv$, the functions b and g are compactly supported and Gevrey-2, and a is Gevrey-2 on a neighborhood of the support of b. The logarithmic transform in the multiplicative variable is

$$\widehat{K}_{\log}(t, u) = \int_{\mathbb{R}} K(e^s, u)\, e^{-its}\, ds.$$

The phase u/β mixes the two variables. The substitution $v = \beta x$ removes the mixture:

$$\widehat{g}\left(-\frac{u}{2\pi\beta}\right) = \beta \int g(\beta x)\, e^{iux}\, dx.$$

Integration by parts in x now sees a Gevrey-2 function of x, and derivatives in $s = \log\beta$ fall on a, b, and the Jacobian. The resulting joint bound has the shape

$$|u|^N \left|\partial_s^m K(e^s, u)\right| \le A_0\, B_u^N (N!)^2\, B_t^m (m!)^2$$

for constants depending only on a, b, g, and the fixed supports. Choosing N and m near $\sqrt{|u|}$ and $\sqrt{|t|}$, and using $k! \le k^k$, converts each factorial into a factor 4^{-N} or 4^{-m}. Therefore there are constants $c_1, c_2, C > 0$, depending only on those fixed data, such that

$$\left|\widehat{K}_{\log}(t,u)\right| \le C \exp\left(-c_1\sqrt{|u|} - c_2\sqrt{|t|}\right). \tag{7.11}$$

Both truncations are polylogarithmic. For each fixed $A > 0$ there is a constant C_A such that the integral of the kernel over $|u| > C_A(\log X)^2$, and the integral of its logarithmic transform over $|t| > C_A(\log X)^2$, are $O(X^{-A})$ up to a factor $(1+|u|)^{-2}$ in the remaining variable. Relative to the natural scales H for the shift and $\log X$ for the Mellin frequency, the number of terms that survive is a power of $\log X$.

7.10. What the construction provides

Proposition 7.3 (Smoothing package). *Fix $\ell > 0$ and a compact interval in the logarithmic coordinate, with collar ℓ. The weight $W(u) = \phi(\log u)$ built above satisfies:*

(i) *W is nonnegative, at most 1, and supported on a fixed compact subinterval of $(0,\infty)$.*

(ii) *On the logarithmic line, ϕ obeys the Gevrey-2 bound (7.5).*

(iii) *$\widehat{K}_\ell$ obeys the decay (7.4), and the Mellin integral of W may be cut at $C_A(\log X)^2$ with tail $O(X^{-A})$.*

(iv) *A smooth factor twisted by n^{it} inside that frequency range has derivatives $O((\log X)^{2m})$.*

(v) *A two-variable kernel of the form (7.10), with Gevrey-2 inputs, obeys (7.11).*

None of these statements uses a dispersion estimate. They are the input a weighted dispersion theorem is allowed to quote. The power savings budgeted in the arithmetic block estimates are not spent on the smoothing. The next section turns the arithmetic-progression sums that remain into character sums, sorted by conductor.

8. Dirichlet Characters and Conductor-Subtracted Discrepancy

The divisor form of B_1 converts a correlation against the primes into a family of sums in arithmetic progressions. This section splits those sums by the conductor of a Dirichlet character. The split is algebraic. It does not estimate the sums. It decides which part of each sum is an outer residue condition and which part is a character sum of conductor larger than a fixed power of X.

8.1. Progressions

A progression $n \equiv a \pmod d$ is a reduced class when $(a, d) = 1$. Only reduced classes carry primes larger than d. For an arithmetic function g, the ordinary discrepancy against the reduced classes is

$$\Delta(g; a \bmod d) = \sum_{n \equiv a \pmod d} g(n) - \frac{1}{\varphi(d)} \sum_{(n,d)=1} g(n). \tag{8.1}$$

The second sum is the average of g over integers coprime to d. If g were equidistributed among the $\varphi(d)$ reduced classes, every discrepancy would vanish.

Because $B_1(n) = \sum_{d \mid n} \lambda_d$, a sum $\sum_n \Lambda(n) B_1(n + j)$ opens into sums of Λ over the progressions $n \equiv -j \pmod d$. The moduli d run through the divisors of the eight-band denominators. There are too many of them to estimate one at a time. The useful bounds are averages over the family, weighted by $|\lambda_d|$.

8.2. Characters

A Dirichlet character modulo d is a homomorphism

$$\chi \colon (\mathbb{Z}/d\mathbb{Z})^\times \to \mathbb{C}^\times,$$

extended by $\chi(n) = 0$ when $(n, d) > 1$. The extension is completely multiplicative on all of $\mathbb{Z}$: $\chi(mn) = \chi(m)\chi(n)$ for every pair of integers, since both sides vanish as soon as mn shares a factor with d. The principal character is $\chi_0(n) = \mathbf{1}_{(n,d)=1}$.

For $(a, d) = 1$ and every integer n,

$$\mathbf{1}_{n \equiv a \pmod d} = \frac{1}{\varphi(d)} \sum_{\chi \bmod d} \overline{\chi}(a)\, \chi(n). \tag{8.2}$$

The identity is the orthogonality of the irreducible characters of a finite abelian group, together with the convention that both sides are zero when $(n, d) > 1$. It converts a residue condition into a linear combination of character sums.

Modulo 5, the group of units is cyclic of order 4, generated by 2. The quadratic character is the Legendre symbol:

$n \bmod 5$	0	1	2	3	4
$\left(\frac{n}{5}\right)$	0	1	-1	-1	1.

It is completely multiplicative, and it is orthogonal to the principal character on the units. Multiplicativity and orthogonality together characterize the table; orthogonality alone does not determine the values.

8.3. Conductors

A character χ modulo d is induced by a character χ^* modulo f when $f \mid d$ and

$$\chi(n) = \chi^*(n)\, \mathbf{1}_{(n,d)=1}. \tag{8.3}$$

The conductor $f = \operatorname{cond}(\chi)$ is the smallest such modulus, and χ^* is then primitive. The modulus d is the level at which the progression is written. The conductor is the level at which the oscillation actually lives. They differ precisely when χ is imprimitive.

The factor $\mathbf{1}_{(n,d)=1}$ cannot be omitted. A primitive character modulo f may be nonzero at an integer that shares a factor with d/f. The induced character modulo d must vanish there. For squarefree d, the Chinese remainder theorem splits $(\mathbb{Z}/d\mathbb{Z})^\times$ into the product of the groups modulo the prime factors of d, and the conductor of χ is the product of the conductors of the local factors.

The Legendre symbol modulo 5 induces a character modulo 15 by (8.3). On residues coprime to 15 it repeats the symbol; on multiples of 3 or 5 it is zero, even when the underlying residue modulo 5 was a unit.

8.4. Discrepancy as a character sum

Substitute (8.2) into (8.1). The principal character reproduces $\mathbf{1}_{(n,d)=1}$, so its term cancels the average over reduced classes. For $(a,d)=1$,

$$\Delta(g;a \bmod d) = \frac{1}{\varphi(d)} \sum_{\substack{\chi \bmod d \\ \mathrm{cond}(\chi)>1}} \overline{\chi}(a) \sum_n g(n)\,\chi(n). \tag{8.4}$$

Each non-principal character carries its own irregularity. The analytically difficult ones are those of small conductor. A real character of conductor f may have an exceptional zero, and the classical way of removing it on very small f is the Siegel–Walfisz theorem, whose implied constant is ineffective. Conductor subtraction does not delete that phenomenon from the original discrepancy. It moves every character of conductor at most K into a separate sum, which has to be estimated or restored on its own terms. What remains is a sum over conductors larger than K, where exceptional zeros of very small modulus no longer sit inside the character sum being estimated.

8.5. Subtracting small conductors

Following Drappeau, fix a cutoff $K \geq 1$ and define

$$u_K(n,a;d) = \mathbf{1}_{n\equiv a \pmod d} - \frac{1}{\varphi(d)} \sum_{\substack{\chi \bmod d \\ \mathrm{cond}(\chi)\leq K}} \overline{\chi}(a)\,\chi(n). \tag{8.5}$$

The conductor-subtracted discrepancy is $\Delta_K(g;a \bmod d) = \sum_n g(n)\,u_K(n,a;d)$. The same cancellation as in (8.4), stopped at conductor K, gives

$$\Delta_K(g;a \bmod d) = \frac{1}{\varphi(d)} \sum_{\substack{\chi \bmod d \\ \mathrm{cond}(\chi)>K}} \overline{\chi}(a) \sum_n g(n)\,\chi(n). \tag{8.6}$$

When $K=1$, the sum on the right of (8.5) contains only the principal character, so u_1 is the kernel of (8.1) and $\Delta_1 = \Delta$.

In the production argument the cutoff is the fixed power

$$K = X^k, \qquad k = \frac{1}{10000}. \tag{8.7}$$

Characters of conductor at most K are few enough, and of small enough modulus, to be handled by a large sieve or by a trivial bound. They are not free. The original numerator of the variational principle is an ordinary correlation, which sees Δ and not only Δ_K. The difference $\Delta_K - \Delta$ is estimated later and added back. Subtraction is a reorganization, not a deletion.

8.6. The outer and inner pieces

Let $d = qr$ with $(q,r)=1$, and assume every prime factor of q exceeds K. This is the roughness that Lemma 4.8 arranges for the complementary factor called q there: $P^-(q) > X^{a_i}$, and $a_i \geq a_{\min} > k$, so $P^-(q) > K$ throughout the production range. The factor r is not required to be free of large primes. A prime of the core may sit in r. The conductor argument uses only the roughness of q.

Lemma 8.1 (Exact splitting). *Let $(q,r)=1$, $(a,qr)=1$, and $P^-(q)>K$. Then*

$$\Delta_K(g;a \bmod qr) = E_1 + \frac{E_2}{\varphi(q)}, \tag{8.8}$$

where

$$E_1 = \sum_{n\equiv a \pmod{qr}} g(n) - \frac{1}{\varphi(q)} \sum_{\substack{n\equiv a \pmod r \\ (n,q)=1}} g(n), \tag{8.9}$$

$$E_2 = \Delta_K\big(g\,\mathbf{1}_{(\cdot,q)=1};\ a \bmod r\big). \tag{8.10}$$

Proof. By the Chinese remainder theorem, every character modulo qr is a product $\chi_q \otimes \chi_r$. If χ_q is non-principal, the conductor of the product is divisible by a prime factor of q, hence exceeds K. Every character modulo qr of conductor at most K is therefore principal on $(\mathbb{Z}/q\mathbb{Z})^\times$: it has the form $\chi(n) = \mathbf{1}_{(n,q)=1}\,\psi(n)$ with ψ modulo r and $\operatorname{cond}(\psi) \le K$. Orthogonality over those characters produces

$$\frac{1}{\varphi(qr)} \sum_{\operatorname{cond}(\chi)\le K} \overline{\chi}(a)\,\chi(n) = \mathbf{1}_{(n,q)=1}\,\frac{1}{\varphi(q)}\big(\mathbf{1}_{n\equiv a \pmod r} - u_K(n,a;r)\big).$$

Subtract this from $\mathbf{1}_{n\equiv a \pmod{qr}}$ and sum against $g(n)$. The first two resulting sums are E_1. The sum against $u_K(\cdot\,,a;r)$, restricted to $(n,q)=1$, is $E_2/\varphi(q)$. □

The identity is false as soon as a prime factor of q is at most K: a non-principal character modulo that prime then has conductor at most K and is invisible to the right-hand side. The two pieces ask for different estimates. The outer discrepancy E_1 is a comparison of two residue conditions, one modulo qr and one modulo r. The inner discrepancy E_2 is a conductor-subtracted character sum on the smaller modulus r, with the coprimality mask $(n,q)=1$ left in place.

8.7. Convolutions

If χ is completely multiplicative, a Dirichlet convolution factors:

$$\sum_n (\alpha * \beta)(n)\,\chi(n) = \Big(\sum_m \alpha(m)\,\chi(m)\Big)\Big(\sum_n \beta(n)\,\chi(n)\Big), \tag{8.11}$$

whenever the sums converge absolutely. The same identity holds for any finite number of factors. The extension of χ by zero is already completely multiplicative, so a coprimality mask $\mathbf{1}_{(n,d)=1}$ built into χ is preserved: it is not an extra condition one may drop after the factorization. A factor that has been multiplied by a smooth weight from Section 7 may be twisted by n^{it} inside the polylogarithmic frequency range, and its derivatives remain $O((\log X)^{2m})$. A factor that is an arbitrary bounded arithmetic function, such as a von Mangoldt weight or a divisor coefficient, is not smooth. The Heath–Brown identity, which replaces Λ by convolutions of the divisor function and a short von Mangoldt weight, is the device that makes those factors few enough to estimate. It is the subject of the next section.

8.8. Individual sums and averages

For a primitive non-principal character χ modulo q, the Pólya–Vinogradov inequality gives

$$\Big|\sum_{n\le N} \chi(n)\Big| \ll \sqrt{q}\,\log q,$$

uniformly in N. Partial summation carries the same square-root cancellation to a smooth weight of bounded variation, with an extra factor depending only on that variation. The inequality is an individual bound. It does not see a family of moduli.

The multiplicative large sieve bounds an average. In the form used for the inner sum, one groups primitive characters by their conductor f and weights the family by $1/\varphi(f)$. The weight is the one produced by writing an imprimitive character sum in terms of its primitive inducing character and summing the geometric factor over the cofactor; without it, the same character would be counted once for every modulus it induces. An individual bound and an averaged square are not interchangeable. The inner estimate of a Type I or Type II block is an average. A single shifted correlation against one character is not.

The applications of these two theorems to the Type I, Type II, and Type III blocks are not given here. Each block has its own range of conductors and its own smooth factor.

8.9. Putting the low conductors back

The numerator of the variational principle is a sum of $\Lambda(n)B_1(n+j)$, which opens into ordinary discrepancies Δ, not into Δ_K. The difference is exactly the characters that subtraction removed. For $(a,d)=1$,

$$\Delta_K(g;a \bmod d)-\Delta(g;a \bmod d)=-\frac{1}{\varphi(d)}\sum_{\substack{\chi \bmod d\\ 1<\mathrm{cond}(\chi)\le K}}\overline{\chi}(a)\sum_n g(n)\,\chi(n). \tag{8.12}$$

The conductor-one term is absent: it was already cancelled in both Δ and Δ_K. Grouping the remaining characters by their primitive conductor f and the imprimitive cofactor d/f is the shape in which a large sieve, or a trivial bound, can be applied.

Three uses of character sums occur in the argument, and they are not one estimate.

(1) The inner sum E_2 in a Type I or Type II block is an average over primitive conductors, weighted by $1/\varphi(f)$, and is the place the large sieve enters.

(2) The Type III block is estimated after the characters of conductor at most K have been removed; that removal is a different sum, taken over small f directly.

(3) The identity (8.12) restores those small conductors in the shifted prime correlation. It is estimated on its own, after the blocks, and it is not absorbed into either of the other two.

8.10. The target

The analytical objective for the additional family, uniformly for $0<|j|\le J$, is

$$\sum_{\substack{d\in\mathcal{D}_1\\(d,j)=1}}|\lambda_d|\left|\Delta_K\big(W(\cdot/X)\,\Lambda;\,-j \bmod d\big)\right|\ll X^{1-\varepsilon_{\mathrm{d}}/2+o(1)}. \tag{8.13}$$

Here $\mathcal{D}_1$ is the set of squarefree moduli that appear as divisors in the form of B_1, and

$$\varepsilon_{\mathrm{d}}=\frac{11}{100000}=0.00011. \tag{8.14}$$

Uniformity in j matters because the variational numerator sums the shifts $|j|\le H$. A bound that lost a factor of the number of shifts would be empty. The divisors in $\mathcal{D}_1$ are of two kinds. Those that keep a full core remain above $X^{1/2+\eta_c}$ and are the ones to which Lemma 8.1 applies. Those that omit a core prime are at most $X^{1/2-1/250}$, by Lemma 4.6, and are estimated by a large sieve rather than by dispersion.

The saving $\varepsilon_{\mathrm{d}}/2$ is larger than the longest shift exponent the argument uses. With $\theta_1 = 27/500000$,

$$\frac{\varepsilon_{\mathrm{d}}}{2} - \theta_1 = \frac{1}{1000000} = 10^{-6} > 0.$$

Summing $O(X^{\theta_1})$ shifts therefore still leaves a power of X to spare. The estimate (8.13) is not proved in this section. The von Mangoldt function is not a convolution. The next section replaces it by convolutions to which the outer and inner bounds can be applied.

9. Heath–Brown's Identity and the Reduction

The target (8.13) is a discrepancy of Λ. Heath–Brown's identity replaces Λ, on the support of the smooth weight, by a finite sum of convolutions. A dyadic partition then localizes each factor, and a combinatorial fact about their scales says that every tuple is of one of four types. Two of those types are estimated in this section. The other two are the remaining dispersion theorems. Nothing here assumes those theorems.

9.1. Why the von Mangoldt function is decomposed

A character sum of Λ does not factor. A character sum of a convolution does: Section 8 records

$$\sum_n (\alpha * \beta)(n)\,\chi(n) = \Big(\sum_m \alpha(m)\,\chi(m)\Big)\Big(\sum_n \beta(n)\,\chi(n)\Big)$$

whenever the sums converge absolutely. The same holds for several factors. The point of a multilinear decomposition is to trade one prime weight for several weights, each localized at its own scale, so that the character sums and the residue conditions can be estimated by the shape of those scales.

9.2. Convolution

Dirichlet convolution is

$$(f * g)(n) = \sum_{ab=n} f(a)\,g(b).$$

It is associative and commutative. Write $\mathbf{1}$ for the constant function 1, and δ_1 for the function that is 1 at $n = 1$ and 0 elsewhere. Möbius inversion is $\mu * \mathbf{1} = \delta_1$. Since $-\zeta'/\zeta = (\zeta^{-1}) \cdot (-\zeta')$ and ζ^{-1} is the Dirichlet series of μ, one has the elementary identity

$$\Lambda = \mu * \log. \tag{9.1}$$

The identity is exact and useless by itself for a discrepancy: μ is as irregular as Λ, and it is supported on all squarefree integers. Truncating μ and correcting the truncation by a finite inclusion-exclusion is Heath–Brown's identity.

For a concrete convolution, $(\mu * \mathbf{1})(n) = \delta_1(n)$ and $(\mathbf{1} * \mathbf{1})(n) = \tau(n)$. The second says that a smooth weight on each factor of a convolution of two constant functions becomes a smooth weight on the divisors, which is the model for a Type 0 sum.

9.3. The identity at depth five

Let $K_{\mathrm{HB}} = 5$ and let $\mu_{\le}(n) = \mu(n)\,\mathbf{1}_{n \le (2X)^{1/5}}$. This depth is not the conductor cutoff $K = X^{1/10000}$. The two letters are unrelated. On the interval $[X, 2X]$, Heath–Brown's identity at depth 5, in the form recorded by Polymath as Lemma 3.3, is

$$\Lambda(n) = \sum_{r=1}^{5} (-1)^{r-1} \binom{5}{r} \big(\mu_{\le}^{*r} * \mathbf{1}^{*(r-1)} * \log\big)(n). \tag{9.2}$$

The identity is the truncation of $\Lambda = \mu * \log$ at $n \le (2X)^{1/5}$ inside each of $r \le 5$ Möbius factors, with the binomial theorem supplying the terms in which one or more factors of μ have been replaced by the tail. Because $n \le 2X$ and five factors each larger than $(2X)^{1/5}$ cannot multiply to n, the tail vanishes on $[X, 2X]$. Each summand is a convolution of at most $r + (r-1) + 1 = 2r \le 10$ functions: r copies of the truncated Möbius function, $r-1$ copies of $\mathbf{1}$, and one logarithm.

Depth five is enough because the frozen separation exponent satisfies

$$\frac{1}{K_{\mathrm{HB}}} = \frac{1}{5} = 0.200 < 0.202 = 2\sigma, \qquad \sigma = \frac{101}{1000}.$$

The margin is $2\sigma - 1/5 = 1/500$. Every convolution factor that is larger than $X^{2\sigma}$ is therefore too large to be a truncated Möbius factor, and it is one of the smooth factors $\mathbf{1}$ or log. That is the only place the numerical value of the depth is used. It is not a substitute for the classification below.

9.4. Dyadic localization

Each convolution factor is still summed over a long range. Let $1 = \sum_{N \in 2^{\mathbb{Z}_{\ge 0}}} w(n/N)$ be a smooth partition of unity, with w supported on $[1/2, 2]$ and with $\|w^{(m)}\|_\infty$ bounded for each fixed m, independently of N. Inserting the partition into every factor of (9.2) produces tuples $(N_1, \dots, N_{2r})$. A tuple whose product of supports misses $[X, 2X]$ contributes nothing, once the sum is multiplied by the observation weight supported in a fixed subinterval of $[X, 2X]$. Every retained tuple therefore satisfies

$$N_1 \cdots N_{2r} \asymp X, \tag{9.3}$$

up to a factor $2^{O(1)}$ that depends only on the number of factors. The product of the scales is not the support of the convolution. Each factor $w(n/N_\nu)$ is supported on $[N_\nu/2,\ 2N_\nu]$, so the convolution is supported on a product of those intervals. The ratio between that product and $N_1 \cdots N_{2r}$ is at most $2^{O(1)}$, with the implied constant at most 2^{20} for ten factors. A fixed factor is not a power of X.

The number of retained tuples is $O((\log X)^{10})$. There are at most ten factors, and each scale N_ν has $O(\log X)$ dyadic choices up to $O(X)$. This count is $X^{o(1)}$. It does not introduce a fixed power X^{ε}.

9.5. The smooth weight on each block

The truncated Mellin formula (7.8) writes $W(n/X)$ as an integral of n^{it} over $|t| \le C_A(\log X)^2$, plus $O(X^{-A})$. On a convolution,

$$(n_1 \cdots n_m)^{it} = n_1^{it} \cdots n_m^{it}.$$

Each smooth factor, multiplied by n^{it} in that range, keeps derivative bounds $O_m((\log X)^{2m})$ by (7.9). It remains a smooth coefficient sequence in Polymath's sense. A truncated Möbius factor, multiplied by the same power, stays a bounded arithmetic coefficient sequence. No derivative of $\mu_{\le}$ is required. The integral of the tail $|t| > C_A(\log X)^2$ contributes $O(X^{-A})$ uniformly in n, and for $A = 10$ its contribution to the discrepancy target is $O(X^{-8})$. The same truncation applies on any fixed compact observation interval in place of a subinterval of $[X, 2X]$, with constants allowed to depend on that interval.

9.6. Scales

Write $N_\nu = X^{t_\nu}$. The product constraint becomes

$$\sum_\nu t_\nu = 1 + O\big(\tfrac{1}{\log X}\big).$$

Grouping two factors into one bilinear piece adds their exponents. The frozen separation parameter is $\sigma = 101/1000$. The three numbers that organize the tuples are

$$2\sigma = 0.202, \qquad \tfrac{1}{2} - \sigma = 0.399, \qquad \tfrac{1}{2} + \sigma = 0.601.$$

A single smooth factor larger than $X^{1/2+\sigma}$ is Type 0. A subset of factors whose product lies between $X^{1/2-\sigma}$ and $X^{1/2+\sigma}$ is a bilinear block, Type I or Type II according to which side of a slightly smaller splitting point it occupies. What remains is three smooth factors, each of medium size, and that is Type III.

Lemma 9.1 (Classification). *Let $(N_1, \dots, N_{2r})$ be a retained tuple, $1 \le r \le 5$, and write $N_\nu = X^{t_\nu}$. The first r exponents, those of the truncated Möbius factors, satisfy $t_\nu \le 1/5 + O(1/\log X) < 2\sigma$. Up to fixed factors $2^{\pm 10}$, the convolution is of at least one of the following types.*

(i) ***Type 0.*** *Some smooth factor has scale $\gg X^{1/2+\sigma} = X^{0.601}$.*

(ii) ***Type I.*** *The factors partition into two blocks of scales M and N, with $MN \asymp X$ and*

$$X^{1/2-\sigma} = X^{0.399} \ll N \le X^{n_{\mathrm{split}}}, \qquad n_{\mathrm{split}} = \tfrac{1}{2} - 2\varpi - 4\varepsilon_{\mathrm{d}} = 0.49556.$$

(iii) ***Type II.*** *The same bilinear shape, with $X^{0.49556} < N \ll X^{1/2}$.*

(iv) ***Type III.*** *Three distinct smooth factors satisfy $X^{2\sigma} \ll N_\nu \ll X^{1/2-\sigma}$ and $N_i N_{i'} \gg X^{1/2+\sigma}$ for $i \ne i'$.*

Proof. This is Polymath's Lemma 3.1 at $\sigma = 0.101$. The inequality $1/5 < 2\sigma$ says that every factor larger than $X^{2\sigma}$ is smooth. Suppose no smooth factor exceeds $X^{1/2+\sigma}$, and no subsum of the exponents lies in $[1/2 - \sigma,\ 1/2 + \sigma]$. Let $\mathcal{B}$ be the set of indices with $t_\nu > 2\sigma$, and write S for the sum of the remaining exponents, each of which is at most 2σ.

The interval $[1/2 - \sigma,\ 1/2 + \sigma]$ has width 2σ. Adding a term of size at most 2σ cannot jump across it. Starting from 0 and adjoining the small exponents one at a time, the partial sums therefore cannot pass from below $1/2 - \sigma$ to above $1/2 + \sigma$. They also cannot land inside the interval, by the assumption that no subsum does. Hence $S < 1/2 - \sigma$. In particular $\mathcal{B}$ is nonempty, and the sum of the exponents in $\mathcal{B}$ exceeds $1/2 + \sigma$.

If $|\mathcal{B}| = 1$, that one exponent exceeds $1/2 + \sigma$, which is Type 0. If $|\mathcal{B}| = 2$, neither exponent lies in $[1/2 - \sigma,\ 1/2 + \sigma]$ and neither exceeds $1/2 + \sigma$, so both are at most $1/2 - \sigma$. Adjoin the small exponents to the larger of the two. The same no-jump rule says that this running sum stays at most $1/2 - \sigma$, and the second exponent is at most $1/2 - \sigma$, so the total is at most $1 - 2\sigma < 1$. If $|\mathcal{B}| \ge 4$, no pair from $\mathcal{B}$ has sum in $[1/2 - \sigma,\ 1/2 + \sigma]$. Combined with the lower bound 2σ on each element, every pair sums to more than $1/2 + \sigma$. Two disjoint pairs then sum to more than $1 + 2\sigma$. The only remaining possibility is $|\mathcal{B}| = 3$. Every pair from $\mathcal{B}$ sums to more than $1/2 + \sigma$, which is Type III. □

The fixed factors $2^{\pm 10}$ enlarge the implied constants and do not change the exponents. The strict inequality $2\sigma - 1/5 = 1/500 > 0$ absorbs the $O(1/\log X)$ in the scale constraint for large X. The classification is this argument. It is not a count of sample tuples.

9.7. Type 0

A Type 0 block has the shape $g = \alpha_S * \psi_0$, where ψ_0 is smooth of scale $N_0 \gg X^{1/2+\sigma}$ and the complementary coefficient α_S is supported at scale $N_S \ll X^{1/2-\sigma}$. The long factor is smooth, with derivative bounds $O((\log X)^{2m})$ after the Mellin twist. Poisson summation on a progression modulo d cancels the zero frequency in the discrepancy and gives

$$\bigl|\Delta(\psi_0; b \bmod d)\bigr| \ll (\log X)^{O(1)} \frac{d}{N_0}.$$

Consequently, with $D = X^{1/2+b_{\max}}$,

$$\sum_d |\lambda_d| \left|\Delta(g;\, -j \bmod d)\right| \ll X^{o(1)} \frac{N_S D^2}{N_0} = X \left(\frac{D}{N_0}\right)^2 X^{o(1)} \ll X^{1-2(\sigma-b_{\max})+o(1)} = X^{20149/25000+o(1)}.$$

The difference $\Delta_K - \Delta$ on this block is smaller. Pólya–Vinogradov on the primitive characters of conductor $f \le K$, followed by partial summation, contributes $O(X^{o(1)} N_S K^{3/2})$. With $N_S \ll X^{0.399}$ and $K^{3/2} = X^{0.00015}$, this is $O(X^{0.39915+o(1)})$. Type 0 does not use a dispersion estimate. The Poisson estimate saves $2(\sigma - b_{\max}) = 4851/25000$ in the exponent.

9.8. Type I and Type II

A bilinear block is $g = \alpha * \beta$ with scales M and N, $MN \asymp X$. Write $n = \log N / \log X$ for the shorter scale, so $n \le 1/2$. Type I is the range

$$n_0 \le n \le n_{\text{split}}, \qquad n_0 = \tfrac{1}{2} - \sigma, \qquad n_{\text{split}} = \tfrac{1}{2} - 2\varpi - 4\varepsilon_{\mathrm{d}}.$$

The two scales are separated by a fixed power of X. Exponential-sum estimates that complete the longer variable, and that use the roughness of the outer modulus, fit this range. Type II is the complementary range $n_{\text{split}} \le n \le 1/2$, where the scales are nearly equal and the Cauchy–Schwarz step has to be arranged differently. Both ranges are required to return the same saving $X^{-\varepsilon_{\mathrm{d}}/2+o(1)}$. The proofs are the Type I and Type II arguments of the next section. They are not given here.

9.9. Type III

A Type III block has the shape

$$g = \alpha * \psi_1 * \psi_2 * \psi_3,$$

where ψ_1, ψ_2, ψ_3 are smooth and α is only a coefficient sequence. The smooth scales satisfy

$$X^{2\sigma} \ll N_\nu \ll X^{1/2-\sigma}$$

and, for $i \ne j$,

$$N_i N_j \gg X^{1/2+\sigma}.$$

The pairwise lower bounds are not implied by the individual ones. Three factors of size $X^{2\sigma}$ have product $X^{6\sigma} = X^{0.606}$, and the complementary exponent $1 - 6\sigma = 0.394$ is much larger than the bound obtained from the pairs. Multiplying the three pairwise inequalities gives

$$(N_1 N_2 N_3)^2 \gg X^{3(1/2+\sigma)},$$

hence

$$N_1 N_2 N_3 \gg X^{3/4+3\sigma/2} = X^{0.9015}$$

and, since $M N_1 N_2 N_3 \asymp X$,

$$M \ll X^{1/4-3\sigma/2} = X^{0.0985}.$$

At $\sigma = 101/1000$ the two exponents are $1803/2000$ and $197/2000$. The pairwise constraints are what force M down to $X^{0.0985}$. The individual floors do not. Completion of the three smooth variables produces hyper-Kloosterman sums. That analysis is the Type III argument, deferred to the section after the bilinear estimates. The verification that the Mellin-twisted profiles meet Polymath's definition of a smooth coefficient sequence, while α is only a coefficient sequence, is Lemma 11.2. It is a derivative check, not the dispersion bound.

9.10. Depth five covers every tuple

Lemma 9.1 is the statement that depth five is sufficient. Its proof uses $1/K_{\mathrm{HB}} < 2\sigma$ once, to guarantee that every large factor is smooth, and then only the width 2σ of the forbidden interval for bilinear sums. At the frozen values,

$$\frac{1}{5} = 0.200 < 0.202 = 2\sigma.$$

The $O(1/\log X)$ uncertainty in each exponent is smaller than the margin $1/500$ for large X, so a tuple that sits $O(1/\log X)$ away from a boundary falls on the side the inequality predicts, up to the fixed dyadic factors already allowed. No sample of tuples is used.

9.11. Full cores, omitted cores, and prime powers

Section 4 splits the moduli that appear in B_1 into those that contain a whole core and those that omit at least one core prime. The omitted ones satisfy $d \le X^{1/2-1/250} = X^{0.496}$. Type 0 on this range is already covered by the bound above, which is stronger and applies up to $X^{1/2+b_{\max}}$. For a Type I, Type II, or Type III block, group the convolution as $\beta_1 * \beta_2$ at scales $L_1, L_2 \le X^{1-2\sigma} = X^{0.798}$. Cauchy–Schwarz and the classical multiplicative large sieve, in the form of Iwaniec–Kowalski, Theorem 7.13, then give

$$\sum_{\substack{d \le X^{0.496} \\ (d,j)=1}} |\lambda_d| \left|\Delta_K(g; -j \bmod d)\right| \ll X^{o(1)} \max_{K/2 < F \le X^{0.496}} \left(F\sqrt{X} + \sqrt{XL_1} + \sqrt{XL_2} + X/F\right).$$

The four terms are at most $X^{0.996}$, $X^{0.899}$, $X^{0.899}$, and $X^{1-k} = X^{0.99990}$. The largest is $X^{1-k+o(1)}$. Since $k = 1/10000 > \varepsilon_{\mathrm{d}}/2$, this is stronger than the target (8.13). Only full-core moduli are passed to the Type I, Type II, and Type III arguments.

Heath–Brown's identity is applied directly to Λ, including prime powers. Proper prime powers p^m, $m \ge 2$, are separated only in the later numerator assembly where a coprimality mask is removed. Reversing the order of summation in d and in p^m, their progression contribution is $O(X^{1/2+o(1)})$. The characters of conductor at most K contribute an extra factor K, and the total is $O(X^{1/2+k+o(1)}) = O(X^{0.5001+o(1)})$. This is smaller than every block bound above.

9.12. The reduction

Proposition 9.2 (Reduction). *Assume the Type 0 bound of this section, the omitted-core bound of this section, and the prime-power bound of this section. Assume in addition a Type I/II bound of strength $X^{1-\varepsilon_{\mathrm{d}}/2+o(1)}$ for every full-core modulus, and a Type III bound of strength $X^{1-\varepsilon_{\mathrm{III}}+o(1)}$ with $\varepsilon_{\mathrm{III}} = 1/1000$. Then the target* (8.13) *holds.*

The losses are as follows. There are $O((\log X)^{10})$ dyadic tuples and $O(\log X)$ values of the depth index r, which is $X^{o(1)}$. The Mellin integral is cut at $(\log X)^2$ with a negligible tail. Type 0 saves $X^{4851/25000}$, which is stronger than $X^{-\varepsilon_{\mathrm{d}}/2}$. The omitted-core sum saves X^k. Prime powers save about $X^{1/2}$. Type I and Type II are required to save $X^{\varepsilon_{\mathrm{d}}/2}$. Type III is required to save $X^{\varepsilon_{\mathrm{III}}}$, and $1/1000 > \varepsilon_{\mathrm{d}}/2$, so the Type III saving is the stronger of the two. The bottleneck among the estimates that are not proved in this section is therefore the Type I/II saving.

The reduction is complete as a classification: every retained tuple is one of the four types, and the three elementary contributions have been estimated. The Type I/II and Type III estimates are proved in the next two sections and their appendices.

10. Conductor-Subtracted Type I and Type II Dispersion

This section estimates the bilinear blocks. The propositions below are the proof of Theorem 10.1. Appendix C records the same chain and does not replace it. The inner bound uses the classical large sieve. The nonzero frequencies use Polymath's Propositions 5.12 and 5.10, cited with their hypotheses matched. Their exponential-sum proofs are not reproduced.

10.1. The bilinear theorem

Let $g = \alpha * \beta$ be a Type I or Type II block from Lemma 9.1, with scales M and N satisfying $MN \asymp X$ and $N \le M$. The coefficients obey $|\alpha(m)|, |\beta(n)| \ll X^{o(1)}$ on their supports. A Mellin factor n^{it} with $|t| \le C(\log X)^2$ has modulus one, so it does not change these bounds. The sum runs over full-core moduli $d \in \mathcal{D}_1^{\mathrm{full}}$ coprime to the shift, with $0 < |j| \le J \le X^{0.01}$ and $|\lambda_d| \ll (\log X)^2$. The shorter exponent $n = \log N/\log X$ lies in one of the two ranges of Section 9:

$$n_0 \le n \le n_{\mathrm{split}} \qquad \text{or} \qquad n_{\mathrm{split}} \le n \le \tfrac{1}{2},$$

where $n_0 = 1/2 - \sigma = 399/1000$ and $n_{\mathrm{split}} = 1/2 - 2\varpi - 4\varepsilon_{\mathrm{d}} = 12389/25000$.

Theorem 10.1 (Bilinear dispersion). *Under the hypotheses above,*

$$\sum_{\substack{d \in \mathcal{D}_1^{\mathrm{full}} \\ (d,j)=1}} |\lambda_d| \left|\Delta_K(g;\, -j \bmod d)\right| \ll X^{1-\varepsilon_{\mathrm{d}}/2+o(1)}. \tag{10.1}$$

The bound is uniform in the residue class, in the shift, and in the Mellin frequency.

Polymath's Theorem 5.1 estimates an ordinary discrepancy Δ on a modulus that has been factored by a preliminary sieve, and it allows an exceptional set of moduli. The sum here is the conductor-subtracted discrepancy Δ_K. The factorization is supplied by dense divisibility of the core, for every full-core modulus. The coefficient sequences are the Heath–Brown factors of Section 9, including the smooth Mellin twist. Siegel–Walfisz is not a hypothesis of the exponential-sum propositions used below. It remains the source of the ineffective error in the base band B_0, which is a different sum.

10.2. Proof strategy

A single residue class modulo $d \asymp X^{1/2}$ is too long, and too entangled with the conductor cutoff, to be estimated as one progression. Every full-core modulus factors as $d = qr$ with $(q, r) = 1$ and with every prime of q larger than K. Lemma 8.1 then splits the discrepancy exactly:

$$\Delta_K(g;\, -j \bmod qr) = E_1 + \frac{\mathcal{E}_2}{\varphi(q)}.$$

The outer piece E_1 compares the progression modulo qr with the coprime progression modulo r. The inner piece $\mathcal{E}_2$ is a character sum on r, taken over conductors $f > K$.

The inner piece is an average of primitive character sums, weighted by $1/\varphi(f)$. Cauchy–Schwarz and the large sieve bound it. The outer piece is expanded by a second Cauchy–Schwarz, in the variables (r, m). The resulting squared form Σ splits into three contributions, which are estimated separately: the diagonal $n_1 = n_2$, the constant term of the off-diagonal, and the nonzero frequencies after Poisson summation. The first two are elementary and use the roughness of q. The third is the exponential-sum input.

The argument proceeds in ten steps. Lemma 10.2 factors the modulus. Equations (10.4)–(10.5) define the two unnormalized sums S_1 and S_2. Lemma 10.3 bounds S_2. Lemma 10.4 expands

the outer square before any size estimate. Lemma 10.5 bounds the diagonal of Σ. Lemma 10.6 bounds the constant term. Lemma 10.7 completes the nonzero frequencies and matches them with Polymath (5.27)–(5.30). Lemmas 10.8 and 10.9 feed that expression to Propositions 5.12 and 5.10, and the assembly takes the square root. Throughout, S_1 and S_2 are unnormalized sums and Σ is the squared form. A bound for Σ becomes a bound for S_1 only after the square root, and the factor $X^{\varepsilon_{\mathrm{d}}}$ inside a dual length is not part of that normalization.

10.3. Factorization of the full-core moduli

Every $d \in \mathcal{D}_1^{\mathrm{full}}$ with $(d, j) = 1$ has a unique core c in one of the eight bands of Section 4, and writes as $d = cu$ with $(c, u) = 1$. The core is one-fold X^{δ}-densely divisible, every prime of c exceeds X^{a_i}, and the small factor satisfies $u < X^{b_{\max}}$. Write e_s for the part of u that is kept with the inner modulus. Then $e_s < X^{b_{\max}} = X^{199/50000}$.

Lemma 10.2 (Full-core modulus factorization). *There is a factorization $c = c_1c_2$ such that $r = uc_1$ and $q = c_2$ satisfy $d = qr$, $(q, r) = 1$, and*

$$NX^{-3\varepsilon_{\mathrm{d}}-\delta} \le r \le NX^{-3\varepsilon_{\mathrm{d}}}. \tag{10.2}$$

Every prime of q divides c, so $P^-(q) > X^{a_i} \ge X^{a_{\min}} > K$. On a dyadic box $q \sim Q$, $r \sim R$,

$$RQ^2 \ll X^{63463/100000} = X^{0.63463} \ll X, \qquad Q \ll X^{13063/100000} = X^{0.13063} \le N. \tag{10.3}$$

Proof. Set $R_0 = NX^{-3\varepsilon_{\mathrm{d}}}$. The lower end of the Type I range gives $R_0 \ge X^{n_0-3\varepsilon_{\mathrm{d}}} = X^{0.39867}$, which is larger than $X^{b_{\max}}$. Dense divisibility of c, recalled in Section 4, supplies a divisor $r_1 \mid c$ with $R_0/(e_sX^{\delta}) \le r_1 \le R_0/e_s$. The choice $c_1 = r_1$ and $c_2 = c/r_1$ is the factorization in the statement, and $r = e_sr_1$ lies in the interval (10.2). Coprimality $(q, r) = 1$ follows from $(c, u) = 1$ and from c_1 and c_2 being complementary divisors of the squarefree core. The factor r carries every prime of e_s and every prime of c_1. It may contain large primes. The conductor argument uses only $P^-(q) > K$.

Since $qr < X^{1/2+2\omega}$ and $r \ge NX^{-3\varepsilon_{\mathrm{d}}-\delta}$,

$$Q \le \frac{X^{1/2+2\omega}}{R} \le X^{1/2+2\omega+3\varepsilon_{\mathrm{d}}+\delta-n}.$$

The exponent is largest at $n = n_0$, where it equals 13063/100000. That is strictly below n_0, with margin 26837/100000, so $Q \le N$ throughout both ranges. The same bounds give

$$RQ^2 \le \frac{(QR)^2}{R} \le X^{1+4\omega+3\varepsilon_{\mathrm{d}}+\delta-n} \le X^{63463/100000}$$

at $n = n_0$. The margin under X is 36537/100000. These are the constraints Polymath labels (5.12), (5.13), (5.15), and (5.34). □

Dense divisibility produces a divisor in every scale window of logarithmic width δ. The factorization is therefore available for every full-core modulus. Polymath's Lemma 5.2 removes an exceptional set before the same scale constraints can be imposed. That removal is not used here.

10.4. The outer and inner sums

Lemma 8.1 applies to every pair produced by Lemma 10.2, because $P^-(q) > K$. Write E_1 for the first two sums in that identity and $\mathcal{E}_2$ for the conductor-subtracted sum on the modulus r, so

that $\Delta_K = E_1 + \mathcal{E}_2/\varphi(q)$. Both pieces are unnormalized. Define

$$S_1 = \sum_{d=qr} |\lambda_d|\,|E_1(q,r)|, \tag{10.4}$$

$$S_2 = \sum_{d=qr} |\lambda_d|\,\frac{|\mathcal{E}_2(q,r)|}{\varphi(q)}. \tag{10.5}$$

The sums run over the full-core moduli coprime to j, in one band and one dyadic box. The triangle inequality gives $|\sum |\lambda_d|\,|\Delta_K|| \le S_1 + S_2$ on that box. There are eight bands and $O((\log X)^2)$ boxes.

The absolute values $|\lambda_d|$ and the signs coming from a global residue class are absorbed into coefficients $c_{q,r}$ with $|c_{q,r}| \le 1$, supported on the canonical splits of one band. After dividing out $(\log X)^2$, the sum S_1 is majorized by the maximum of

$$\sum_{r\sim R}\sum_{q\sim Q} c_{q,r}\,\Delta_0(\alpha * \beta; a, b_1, b_2; q, r)$$

over those coefficients and the primitive classes in Lemma 10.4. A complex Mellin factor stays inside β. From here on, S_1 and S_2 denote these unnormalized sums, and Σ denotes the squared form constructed in the next subsection but one.

10.5. The inner large sieve

Lemma 10.3 (Inner large-sieve estimate). *On the range of Lemma 10.2,*

$$|S_2| \ll X^{o(1)} \max_{K<F\le R}\left(\frac{X}{F} + \frac{X}{\sqrt{M}} + \frac{X}{\sqrt{N}} + F\sqrt{X}\right). \tag{10.6}$$

The maximum is $X^{1-k+o(1)}$.

Proof. The modulus r is squarefree. Each character of conductor $f \mid r$ with $f > K$ is induced by a unique primitive character χ^* mod f. Writing $r = fz$ with $(f,z) = 1$, the induced character is $\chi^*(n)\mathbf{1}_{(n,z)=1}$. Therefore

$$\mathcal{E}_2(q,r) = \sum_{\substack{f\mid r\\ f>K}} \sum_{\chi^* \bmod f} \frac{\overline{\chi^*}(a)}{\varphi(r)} \sum_n g(n)\,\mathbf{1}_{(n,qz)=1}\,\chi^*(n).$$

The mask $\mathbf{1}_{(n,qz)=1}$ is absorbed into the coefficient sequences of the convolution $g = \alpha * \beta$, which factors by (8.11). Using $\varphi(r) = \varphi(f)\varphi(z)$ and $|\lambda_d| \ll (\log X)^2$,

$$S_2 \ll (\log X)^2 \sum_{q,z} \frac{1}{\varphi(q)\varphi(z)} \sum_{K<f\le R} \frac{1}{\varphi(f)} \sum_{\chi^* \bmod f} |A(\chi^*)|\,|B(\chi^*)|.$$

Extending the sum to all q and z only enlarges it, and $\sum_{n\le Y} 1/\varphi(n) \ll \log Y$. Roughness of q is not used as cancellation here. It was used in Lemma 8.1, to place every small conductor on r.

Since $\varphi(f) \gg f/\log\log(f+3)$, the weight $1/\varphi(f)$ contributes a factor $X^{o(1)}/F$ on the dyadic segment $f \sim F$. Cauchy–Schwarz in the primitive characters, followed by the large sieve

$$\sum_{f\le 2F} \sum_{\chi^* \bmod f} \Big|\sum_{m\in\mathcal{I}} a_m \chi^*(m)\Big|^2 \ll \big(F^2 + |\mathcal{I}|\big) \sum |a_m|^2 \tag{10.7}$$

for an interval $\mathcal{I}$, produces

$$\sum_{f\sim F}\sum_{\chi^*}\frac{1}{\varphi(f)}|A|^2 \ll X^{o(1)}\Big(FM+\frac{M^2}{F}\Big),$$

and the same bound with N in place of M. The product of the square roots is at most $F\sqrt{X}+X/\sqrt{M}+X/\sqrt{N}+X/F$, because $MN\asymp X$. This is (10.6).

The inequality (10.7) is Iwaniec–Kowalski, Theorem 7.13. The hypotheses used here are primitive characters, coefficients supported on intervals, and $\sum|\alpha|^2\ll X^{o(1)}M$, which follows from $|\alpha|\ll X^{o(1)}$. The inequality is not reproved.

Label $M\ge X^{1/2}$ and $N\ge X^{n_0}$. On $K<F\le R\le NX^{-3\varepsilon_{\mathrm{d}}}$ the four exponents are at most

$$\begin{aligned} X/F &\qquad 1-k=0.999900,\\ X/\sqrt{M} &\qquad 3/4=0.750000,\\ X/\sqrt{N} &\qquad 3/4+\sigma/2=0.800500,\\ F\sqrt{X} &\qquad 1-3\varepsilon_{\mathrm{d}}=0.999670. \end{aligned}$$

The largest is X/F. The outer target has exponent $1-\varepsilon_{\mathrm{d}}/2=0.999945$, and the margin is $k-\varepsilon_{\mathrm{d}}/2=9/200000$. □

The weight $1/\varphi(f)$ is part of the estimate. Without it, the same Cauchy–Schwarz produces $F^2\sqrt{X}$ in place of $F\sqrt{X}$, and at the top of the conductor range that exponent exceeds 1. The variable F is the primitive conductor. The outer scale $Q\ll X^{0.13063}$ is a different quantity, and it is not substituted into the slot $F\sqrt{X}$.

The conclusion passed to the assembly is $|S_2|\ll X^{1-k+o(1)}$.

10.6. The outer expansion

Lemma 10.4 (Outer residue averaging and Cauchy–Schwarz). *Let P_I be Polymath's product of allowed primes, and for primitive classes $a,b_1,b_2 \pmod{P_I}$ define*

$$\Delta_0(g;a,b_1,b_2;q,r)=\sum_{\substack{s\equiv a\ (r)\\ s\equiv b_1\ (q)}}g(s)-\sum_{\substack{s\equiv a\ (r)\\ s\equiv b_2\ (q)}}g(s).$$

The Chinese remainder theorem and the triangle inequality give

$$\sum_{q,r}|E_1(q,r)|\le\frac{1}{\varphi(P_I)}\sum_{\substack{b\ (P_I)\\ (b,P_I)=1}}\sum_{q,r}|\Delta_0(g;a,a,b;q,r)|.$$

Thus it suffices to fix primitive a,b_1,b_2. After linearizing the absolute values with $|c_{q,r}|\le1$, expanding $g=\alpha*\beta$, and choosing a smooth majorant $\psi_M\ge1$ on the support of α, Cauchy–Schwarz gives

$$\Big(\sum_{q,r}|\Delta_0(g;a,b_1,b_2;q,r)|\Big)^2\ll X^{o(1)}MR\,\Sigma, \tag{10.8}$$

where

$$\Sigma=\sum_r\sum_m\psi_M(m)\Big|\sum_{\substack{q,n\\ mn\equiv a\ (r)}}c_{q,r}\beta(n)\big(\mathbf{1}_{mn\equiv b_1\ (q)}-\mathbf{1}_{mn\equiv b_2\ (q)}\big)\Big|^2. \tag{10.9}$$

Proof. The first display is Polymath (5.18), where the average over $b \pmod{P_I}$ is taken before Cauchy–Schwarz. The remaining displays are Polymath (5.19). There are $O(R)$ values of r and $\sum_m |\alpha(m)|^2 \ll X^{o(1)}M$. The factor n^{it} remains inside β_n, while ψ_M is independent of t. □

The target for the squared form, the one that returns $X^{1-\varepsilon_{\mathrm{d}}/2}$ after the square root, is

$$|\Sigma| \ll X^{-\varepsilon_{\mathrm{d}}+o(1)}\frac{MN^2}{R}. \tag{10.10}$$

Indeed $MR \cdot MN^2/R = (MN)^2$, and $MN \asymp X$.

10.7. The diagonal

Lemma 10.5 (Diagonal contribution). *Expanding* (10.9) *gives* $\Sigma(b_1,b_1) - \Sigma(b_1,b_2) - \Sigma(b_2,b_1) + \Sigma(b_2,b_2)$*, as in Polymath (5.20). For each ordered pair* (u,v)*, the terms* $n_1 = n_2$ *in* $\Sigma(u,v)$ *satisfy*

$$|\Sigma_{\mathrm{diag}}(u,v)| \ll X^{1+o(1)} \ll X^{-3\varepsilon_{\mathrm{d}}+o(1)}\frac{MN^2}{R}. \tag{10.11}$$

Proof. Cauchy–Schwarz has removed α, so the diagonal weight is $|\beta(n)|^2$. For a fixed ordered pair (u,v), the three congruences depend on (m,n) through $s = mn$ and, when compatible, define one class modulo $[q_1,q_2]r$. Grouping factorizations of s costs $X^{o(1)}$. Therefore

$$\sum_{r,q_1,q_2}\left(\frac{X}{[q_1,q_2]r}+1\right) \le X(1+\log(2Q))^3 + RQ^2.$$

Since $RQ^2 \ll X$, this proves the first bound. The estimate is applied separately to the four ordered pairs, before their signed sum is formed. Finally $X = (R/N)(MN^2/R)$ and $R/N \le X^{-3\varepsilon_{\mathrm{d}}}$. □

10.8. Zero-frequency cancellation

Lemma 10.6 (Zero-frequency cancellation). *For fixed primitive* u,v*, let* $\Sigma_0(u,v)$ *be the zero-frequency term obtained by completing the* m*-sum in* $\Sigma(u,v)$*, and put* $q_0 = (q_1,q_2)$*. The sub-sum with* $q_0 = 1$ *is a quantity* $\mathcal{X}$ *independent of* u,v*. The remaining part satisfies*

$$|\Sigma_0(u,v) - \mathcal{X}| \ll X^{o(1)}\left(MN + \frac{MN^2}{RX^{a_i}}\right) \ll X^{-\varepsilon_{\mathrm{d}}+o(1)}\frac{MN^2}{R}. \tag{10.12}$$

Thus the four copies of $\mathcal{X}$ *cancel in* $\Sigma(b_1,b_1) - \Sigma(b_1,b_2) - \Sigma(b_2,b_1) + \Sigma(b_2,b_2)$*.*

Proof. This is the decomposition following Polymath (5.24). When $q_0 = 1$ the compatibility indicator is identically one, so the zero frequency contains neither u nor v. If $q_0 > 1$, roughness gives $q_0 \ge P^-(q_i) > X^{a_i}$. Summing first over $\ell = (n_2-n_1)/r$, the compatibility congruence modulo q_0 costs

$$\left(1+\frac{N}{q_0R}\right)X^{o(1)}.$$

The reciprocal-modulus factors then give the first estimate in (10.12), exactly as for $\Sigma_0'(b_1,b_2)$ between Polymath (5.25) and (5.26). The second uses $MN = (R/N)(MN^2/R)$, $R/N \le X^{-3\varepsilon_{\mathrm{d}}}$, and $a_i \ge a_{\min} > \varepsilon_{\mathrm{d}}$. □

10.9. Poisson completion

Lemma 10.7 (Poisson completion and CRT phase). *For an off-diagonal pair put $n_2 = n + \ell r$, $\ell \neq 0$, let $q_0 = (q_1, q_2)$, and rename the original moduli $q_i = q_0 q_i'$. The three congruences for m are*

$$mn \equiv u \pmod{q_0 q_1'}, \qquad m(n+\ell r) \equiv v \pmod{q_0 q_2'}, \qquad mn \equiv a \pmod{r}.$$

They are compatible exactly when

$$C(n) := \mathbf{1}_{u/n \equiv v/(n+\ell r)\ (q_0)} = 1. \tag{10.13}$$

Poisson completion of the smooth majorant ψ_M modulo $L = q_0 q_1' q_2' r$, truncated at

$$H = X^{\varepsilon_{\mathrm{d}}} \frac{RQ^2}{q_0 M} \tag{10.14}$$

up to an absolute factor, produces

$$\Upsilon_{\ell,r}(u,v;q_0) = \sum_{\substack{q_1', q_2' \asymp Q/q_0 \\ (q_1', q_2')=1}} \mathbf{1}_{\mathcal{D}}(q_0 q_1', q_0 q_2', r) \sum_{1 \le |h| \ll H} \left| \sum_n C(n) \beta(n) \overline{\beta(n+\ell r)} \Phi_\ell \right|, \tag{10.15}$$

$$\Phi_\ell = e_r\left(\frac{ah}{n q_0 q_1' q_2'}\right) e_{q_0 q_1'}\left(\frac{uh}{n r q_2'}\right) e_{q_2'}\left(\frac{vh}{(n+\ell r) r q_0 q_1'}\right). \tag{10.16}$$

Here $\mathbf{1}_{\mathcal{D}}$ records the original support conditions, and

$$|\widehat{\Sigma}_1(u,v)| \ll X^{-\varepsilon_{\mathrm{d}}} \frac{M}{RQ^2} \sum_{1 \le |\ell| \ll N/R} \sum_{q_0 \ll Q} q_0 \sum_r \Upsilon_{\ell,r}(u,v;q_0). \tag{10.17}$$

Proof. Poisson summation on the unique compatible class modulo L is exact, and rapid decay of $\widehat{\psi_M}$ truncates at $X^{\varepsilon_{\mathrm{d}}} L/M$. On the dyadic box this is $\ll H$ up to an absolute factor. Both original moduli retain q_0, so the first two congruences give (10.13). The Chinese remainder theorem gives (10.16). The second congruence is imposed modulo $q_0 q_2'$, while the pairwise coprime CRT components used in the phase are $r, q_0 q_1', q_2'$. The coefficients $c_{q,r}$ are independent of n, so their absolute values are taken outside the n-sum while $C(n)$ remains inside. This is Polymath (5.27)–(5.30), including the factor $X^{-\varepsilon_{\mathrm{d}}} M q_0/(RQ^2)$. □

The dictionary is now literal: Polymath's x, N, M, ε are $X, N, M, \varepsilon_{\mathrm{d}}$, and its parameters $q_0, q_1, q_2, r, \ell, h, C, \Phi_\ell$ are those above. The smooth majorant ψ_M is independent of t; the factor n^{it} stays in β_n, where its modulus is one. The moduli are squarefree because they lie in $\mathcal{D}_1$.

10.10. Type I exponential sums

Lemma 10.8 (Type I exponential-sum estimate). *Let $n_0 \le n \le n_{\mathrm{split}}$. Polymath's Proposition 5.12, applied to the sum produced by Lemma 10.7, gives*

$$\sum_{r \sim R} \Upsilon_{\ell,r}(b_1, b_2; q_0) \ll X^{-\varepsilon_{\mathrm{d}} + o(1)} Q^2 R N \frac{(q_0, \ell)}{q_0^2}.$$

The prefactors of Lemma 10.7 convert this into $|\Sigma_1| \ll X^{-\varepsilon_{\mathrm{d}} + o(1)} MN^2/R$.

Proof. Proposition 5.12 takes the phase of Lemma 10.7, requires $(q_0q_1q_2s_2, r) = 1$, and requires the products q_0q_ir and q_0s_2r to lie in the one-fold X^{δ}-densely divisible set. Lemma 10.2 supplies the coprimality with r. The products are the original moduli qr, which are one-fold densely divisible by the support of the family. Since $(j, k) = (0, 0)$, the separate factors q and r are only 0-fold densely divisible. In the proof of the proposition, the least common multiple of those products is still one-fold densely divisible, by one application of Polymath's Lemma 2.10(ii). That is the hypothesis under which the first bound of Corollary 4.16 applies. The corollary is the output of one q-van der Corput estimate, Proposition 4.12(ii). The estimate is cited.

The Cauchy–Schwarz step that precedes the proposition leaves q_1 outside the square and duplicates only q_2. It uses $Q \le N$, which is (10.3), and which holds on the whole Type I interval. The margin is positive at n_0 and would change sign below n_0, so the same appeal is not made on a longer interval. Substituting H from (10.14) produces three braces in $|\Upsilon|^2$. Compared with the squared ceiling $X^{-2\varepsilon_{\mathrm{d}}}$, their exponents are

$$\begin{aligned} \tfrac{1}{4} + 9\varpi + \tfrac{5}{2}\delta + 9\varepsilon_{\mathrm{d}} - \tfrac{5}{6}n &\le -\tfrac{13}{50000}, \\ 8\varpi + 3\delta + 11\varepsilon_{\mathrm{d}} - n &\le -\tfrac{30589}{100000}, \\ \varepsilon_{\mathrm{d}} - \left(\tfrac{1}{2} - 2\varpi - n\right) &\le -3\varepsilon_{\mathrm{d}} = -\tfrac{33}{100000}. \end{aligned}$$

Each expression is monotone in n. The first two maxima lie at $n = n_0$ and the third at $n = n_{\mathrm{split}}$. All three lie strictly below $-2\varepsilon_{\mathrm{d}} = -11/50000$. The square root is the per-modulus bound in the statement, and the sum over $r \sim R$ is Polymath (5.31) on this range. The front factor $X^{\varepsilon_{\mathrm{d}}}$ in the completion cancels the factor $X^{-\varepsilon_{\mathrm{d}}}$ coming from $1/H$. It does not cancel the factor $X^{\varepsilon_{\mathrm{d}}}$ inside the length of the dual sum, which has already been accounted for in the three braces. □

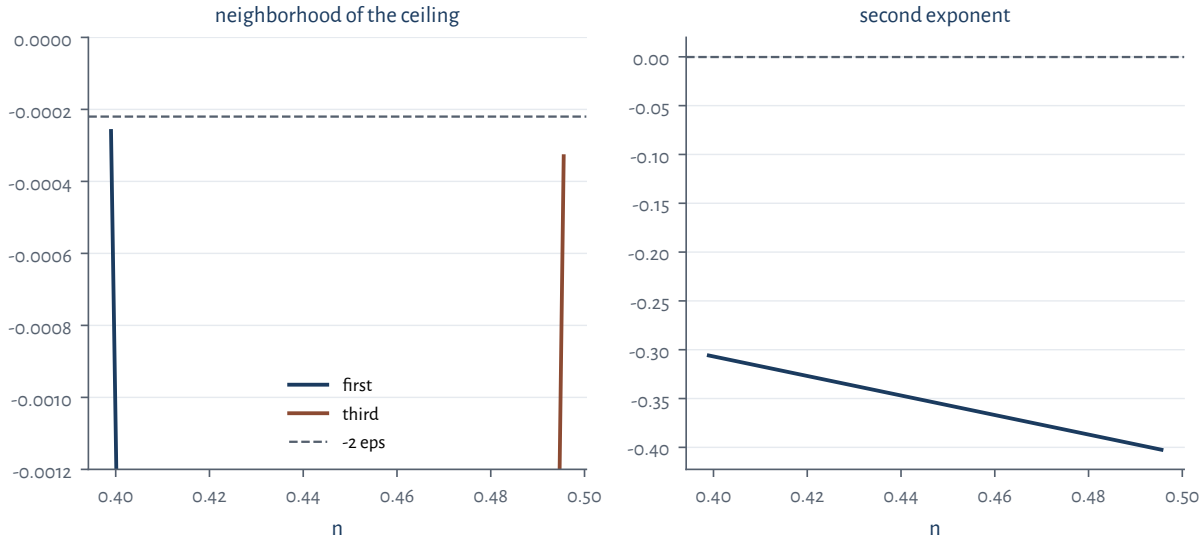

Figure 6: The three Type I braces on $[n_0,\ n_{\mathrm{split}}]$. The left panel is clipped at -0.0012, so only the portions nearest the dashed ceiling $-2\varepsilon_{\mathrm{d}}$ remain visible: the first brace at $n = n_0$ and the third at $n = n_{\mathrm{split}}$. The second brace stays near -0.306.

10.11. Type II exponential sums

Lemma 10.9 (Type II exponential-sum estimate)**.** *Let $n_{\mathrm{split}} \le n \le 1/2$. Polymath's Proposition 5.10, applied to the same completed sum, gives the same bound on $\sum_{r\sim R} \Upsilon_{\ell,r}$, and hence the same bound $|\Sigma_1| \ll X^{-\varepsilon_{\mathrm{d}}+o(1)}MN^2/R$.*

Proof. Proposition 5.10 takes the same phase and the coprimality $(q_0q_1q_2s_1s_2, r) = 1$, now with both q_1 and q_2 duplicated. Its proof uses the second bound of Corollary 4.16, whose input

is the completion bound in Proposition 4.12(i). That bound needs squarefree moduli and a phase of the shape produced by Lemma 10.7. Both are present. The corollary is stated under a dense-divisibility hypothesis, which the one-fold moduli satisfy, and the argument does not factor the modulus further. Theorems 5.8(ii) and 5.8(iii), which require two-fold and four-fold dense divisibility, are not used. The estimate is cited.

The Cauchy–Schwarz step is the one in Polymath (5.34)–(5.35). Before it, the sum Υ is linear in the phase sum. After it, the combinatorial prefactor is exactly the square of $Q^2N(q_0,\ell)q_0^{-2}$. No square root has been taken at that stage. Substituting the full length H, including the factor $X^{\varepsilon_{\mathrm{d}}}$ inside H and inside H^2, produces three braces. Their exponents are

$$\begin{aligned} 1+12\varpi+\tfrac{7}{2}\delta+\tfrac{25}{2}\varepsilon_{\mathrm{d}}-\tfrac{5}{2}n &\le -\tfrac{4999}{40000},\\ 8\varpi+3\delta+11\varepsilon_{\mathrm{d}}-n &\le -\tfrac{8049}{20000},\\ -2\varepsilon_{\mathrm{d}}+2n-1 &\le -2\varepsilon_{\mathrm{d}}. \end{aligned}$$

The first two maxima lie at $n=n_{\mathrm{split}}$ and are strictly below $-2\varepsilon_{\mathrm{d}}$. The third equals $-2\varepsilon_{\mathrm{d}}$ at $n=1/2$ and $R=NX^{-3\varepsilon_{\mathrm{d}}}$, and is strictly smaller for every smaller n. That endpoint is $N=M$. The third brace is the normalized form of $X^{\varepsilon_{\mathrm{d}}}R/M$. Its bound is $X^{-2\varepsilon_{\mathrm{d}}+o(1)}$, which is exactly the required squared target (with its allowed $o(1)$); after the square root it is $X^{-\varepsilon_{\mathrm{d}}+o(1)}$. We do not replace the squared ceiling by $-\varepsilon_{\mathrm{d}}$.

The same first brace is positive at $n=n_0$. Proposition 5.10 is applied only on the Type II interval. □

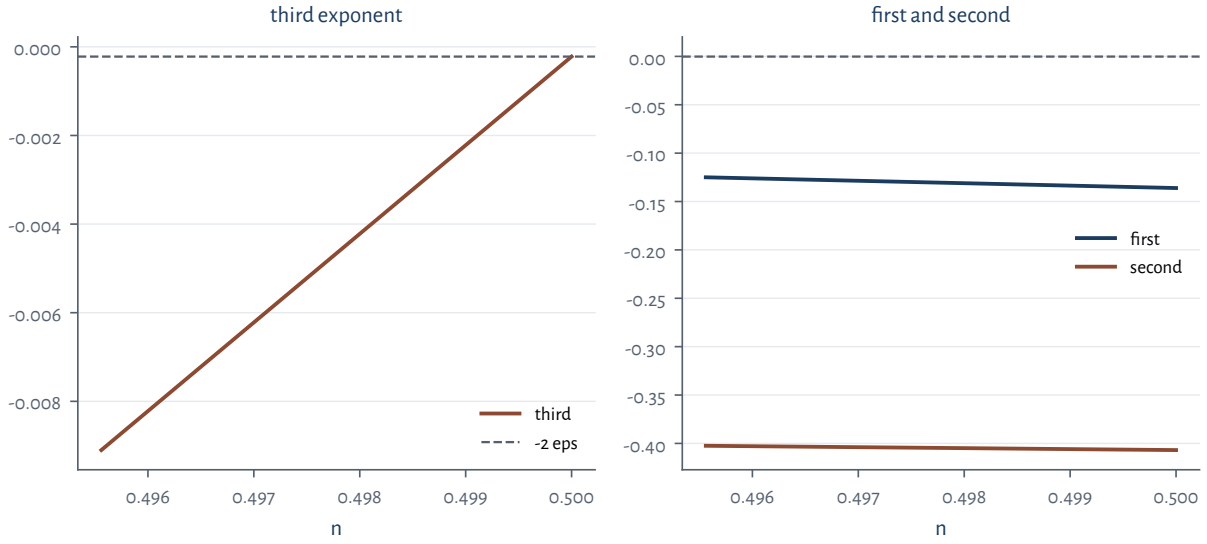

Figure 7: The three Type II braces on $[n_{\mathrm{split}},\ 1/2]$. The third meets the dashed ceiling $-2\varepsilon_{\mathrm{d}}$ at $n=1/2$. The first and second stay below -0.12 and -0.40.

10.12. The exponent ledger

The six comparisons above are monotone, so each maximum is an endpoint value. Written as exact rationals at the frozen parameters, they are

	Type I	Type II
first	$-13/50000$	$-4999/40000$
second	$-30589/100000$	$-8049/20000$
third	$-33/100000$	$-11/50000$
ceiling	$-11/50000$	$-11/50000$.

In decimals the first row is -0.000260 and -0.124975, the second is -0.305890 and -0.402450, and the third is -0.000330 and -0.000220. The ceiling is -0.000220 in both columns.

Five of the six inequalities are strict. The Type I margins under the ceiling are $1/25000$, $30567/100000$, and ε_{d}. The first two Type II margins are $4991/40000$ and $8047/20000$, up to the common denominator of the ceiling. The third Type II comparison is an equality at $N = M$, which is $n = 1/2$, with $R = NX^{-3\varepsilon_{\mathrm{d}}}$. The nominal margin on that one brace is zero: its exponent equals the squared ceiling $-2\varepsilon_{\mathrm{d}}$ and is strictly smaller for every smaller n. Fixed dyadic comparability factors are part of the $X^{o(1)}$ already permitted in the squared target $X^{-2\varepsilon_{\mathrm{d}}+o(1)}$. Taking the square root gives $X^{-\varepsilon_{\mathrm{d}}+o(1)}$ at the Υ-level and then the stated dispersion exponent $1-\varepsilon_{\mathrm{d}}/2$. Subtracting the shift count θ_1 leaves the global saving

$$\frac{\varepsilon_{\mathrm{d}}}{2}-\theta_1=\frac{1}{10^6}.$$

The zero nominal margin is not subtracted from that saving. The dyadic factor is not absorbed by enlarging ε_{d}.

These identities show that the numerical values sit on the correct side of the ceiling. They do not replace the hypothesis match in Lemmas 10.8 and 10.9. A brace below $-2\varepsilon_{\mathrm{d}}$ is the conclusion of Proposition 5.12 or Proposition 5.10 on the range where that proposition's hypotheses have been checked. Outside those ranges the same formal expression can lie above the ceiling, as it does for the Type II first brace at $n = n_0$.

10.13. Assembly

Proof of Theorem 10.1. The squared form of Lemma 10.4 splits into the diagonal, the off-diagonal constant term, and the nonzero frequencies. Lemmas 10.5, 10.6, and 10.8–10.9 bound them by

$$X^{-3\varepsilon_{\mathrm{d}}+o(1)}\frac{MN^2}{R},\qquad X^{-\varepsilon_{\mathrm{d}}+o(1)}\frac{MN^2}{R},\qquad X^{-\varepsilon_{\mathrm{d}}+o(1)}\frac{MN^2}{R}$$

respectively, up to an error $O_A(X^{-A})$ from the completion. The diagonal is the smaller contribution. Adding them gives

$$\Sigma\ll X^{-\varepsilon_{\mathrm{d}}+o(1)}\frac{MN^2}{R}.$$

The outer Cauchy–Schwarz is $|S_1|^2\ll X^{o(1)}MR\,\Sigma$. Substitution produces

$$|S_1|^2\ll X^{o(1)}MR\cdot X^{-\varepsilon_{\mathrm{d}}+o(1)}\frac{MN^2}{R}=X^{-\varepsilon_{\mathrm{d}}+o(1)}(MN)^2.$$

The cancellation of R is exact. With $MN\asymp X$,

$$|S_1|\ll X^{1-\varepsilon_{\mathrm{d}}/2+o(1)}.$$

Lemma 10.3 supplies $|S_2|\ll X^{1-k+o(1)}$. Since $k>\varepsilon_{\mathrm{d}}/2$, the exponent $1-k$ is strictly smaller than $1-\varepsilon_{\mathrm{d}}/2$, and

$$|S_1|+|S_2|\ll X^{1-\varepsilon_{\mathrm{d}}/2+o(1)}.$$

The remaining factors are absolute or logarithmic. Eight bands and $O((\log X)^2)$ dyadic boxes contribute $X^{o(1)}$. The same bound covers $|\lambda_d|\ll(\log X)^{O(1)}$ and the coefficient sequences. The factor 8 in the dual length is absolute. None of these is a fixed positive power of X.

The prefactor $\sum_m|\alpha(m)|^2$ does not depend on the shift or on t. Every bound used for Σ is uniform in the residue class. Lemma 10.9 records that n^{it} remains inside Definition 2.5 for $|t|\le C(\log X)^2$. The condition $(d,j)=1$ is the range of the original sum. The estimate is therefore uniform in $0<|j|\le J$ and in that frequency range. Summing the boxes gives (10.1). □

The three mechanisms in the bound are the factorization and the cancellation, the averaged character sums, and the completed exponential sums. The first is special to the eight-band family: dense divisibility supplies the scale of r, and roughness supplies both $P^-(q) > K$ and the saving X^{a_i} on a shared prime. The second and third are published theorems, applied on the ranges where their hypotheses hold. The Type III blocks are not included in Theorem 10.1. They are the subject of the next section.

11. Type III Dispersion

A Type III block has three smooth factors at once. That is the feature the bilinear argument does not have, and it is what permits a completion in three variables. The complete argument is Proposition material in Appendix D. This section records the bound. The combination with the Type I/II theorem and the remaining blocks is the assembly in Section 12 and Appendix E.

11.1. The trilinear target

Let $g = \alpha * \psi_1 * \psi_2 * \psi_3$ be a Type III block from Lemma 9.1, and let $0 < |j| \le J$. The sum runs over full-core moduli coprime to the shift. The target, with $\varepsilon_{\mathrm{III}} = 1/1000$, is

$$\sum_{\substack{d \in \mathcal{D}_1^{\mathrm{full}} \\ (d,j)=1}} |\lambda_d| \left|\Delta(g;\, -j \bmod d)\right| \ll X^{1-\varepsilon_{\mathrm{III}}+o(1)}. \tag{11.1}$$

The coprimality $(d, j) = 1$ is part of the statement: without it the residue $-j \pmod d$ need not be invertible. The sum in (11.1) is an ordinary discrepancy. The passage from Δ_K to Δ is Lemma 11.3 below, whose size is $X^{0.09875+o(1)}$.

Theorem 11.1 (Type III dispersion)**.** *The bound* (11.1) *holds, uniformly in the Mellin frequency, the shift, and the eight bands. The proof is Appendix D, which is the proof of that theorem.*

The estimate of $T(\mathbf{b})$ and the restoration of the Fourier and Taylor errors are part of that proof. They are not left as a hypothesis of this theorem.

11.2. Proof strategy

The three smooth factors are completed by Poisson summation. The resulting complete exponential sums factor, by the Chinese remainder theorem, into a degenerate sum on the radical of the frequencies and a hyper-Kloosterman sum on the complementary modulus. Averaging those Kloosterman sums is the step that has to produce the power saving. A pointwise bound, even one as strong as Deligne's, does not by itself control the average over the family.

The argument proceeds as follows. Lemma 11.2 checks that the twisted profiles are smooth and that the complementary coefficient satisfies $M \ll X^{0.0985}$. Lemma 11.3 removes conductors at most K. Lemma 11.4 partitions the moduli into short intervals at the cost of one factor $X^{\varepsilon_{\mathrm{III}}}$. Lemma 11.5 completes each smooth factor. Lemma 11.6 expands the completion about a fixed reference modulus so that the dependence on the varying modulus factors out of the frequency sum. Lemma 11.7 reduces the nondegenerate part to a correlation $T(\mathbf{b})$. The correlation is then bounded via Polymath's Corollary 6.26 (which requires $(r_1 r_2, s) = 1$ and does not require $(r_1, r_2) = 1$) and Lemma 7.4, after which the assembly sums the intervals once and restores the remainders.

11.3. Coefficient admissibility

Lemma 11.2 (Twisted smooth profiles and complementary scale). *Let $|t| \le C(\log X)^2$, and let*

$$\psi_\nu(n) = w_\nu(n/N_\nu)\, L_\nu(n/N_\nu)\, n^{it},$$

where w_ν is smooth and supported on a fixed interval $[c, C] \subset (0, \infty)$, and L_ν is either 1 or $\log(N_\nu u)/\log(2N_\nu)$. Then the profile

$$\varphi_\nu(u) = N_\nu^{it}\, w_\nu(u)\, L_\nu(u)\, u^{it}$$

satisfies $|\varphi_\nu^{(m)}(u)| \ll_m (\log X)^{2m}$ on $[c, C]$. In particular each ψ_ν is smooth at scale N_ν in Polymath's Definition 2.5. The factor α is a coefficient sequence at scale M. No derivative of α is required.

If $X^{2\sigma} \ll N_\nu \ll X^{1/2-\sigma}$ and $N_i N_{i'} \gg X^{1/2+\sigma}$ for $i \neq i'$, and if $MN_1N_2N_3 \asymp X$, then

$$N_1N_2N_3 \gg X^{1803/2000} = X^{0.9015}, \qquad M \ll X^{197/2000} = X^{0.0985}. \tag{11.2}$$

A fixed dyadic factor $2^{\pm A}$ on each scale changes the implied constants and not the exponents.

Proof. Write $n^{it} = N_\nu^{it}(n/N_\nu)^{it}$. The factor N_ν^{it} has modulus one and does not depend on the variable of the profile, so it factors out of every derivative. On $[c, C]$,

$$\partial_u^m u^{it} = u^{it-m}\, it(it-1)\cdots(it-m+1),$$

and the falling factorial is at most $(|t| + m - 1)^m$. Leibniz's rule, together with $\|w_\nu^{(m)}\|_\infty \ll_m 1$, bounds the derivatives of $w_\nu(u)u^{it}$ by a constant times $(1 + |t|)^m$. The logarithmic weight equals $1 + \log u/\log(2N_\nu)$. Its derivatives of positive order are $O_m(1/\log N_\nu)$ on $[c, C]$. Under $|t| \le C(\log X)^2$ one obtains $|\varphi_\nu^{(m)}| \ll_m (\log X)^{2m}$.

The factor α is a convolution of the remaining Heath–Brown factors, each bounded by $\tau(n)^{O(1)}(\log X)^{O(1)}$. Convolution preserves that shape. Definition 2.5 therefore makes α a coefficient sequence at the product of those scales. A sequence that alternates between 1 and -1 meets the same pointwise bound. Its forward differences have size 2, while a profile obeying the smooth bound changes by $O((\log X)^{O(1)}/M)$ from one integer to the next. The alternating sequence is an admissible coefficient and is not a smooth profile.

The pairwise lower bounds give

$$N_1N_2N_3 = \sqrt{(N_1N_2)(N_1N_3)(N_2N_3)} \gg X^{3(1/2+\sigma)/2} = X^{3/4+3\sigma/2}.$$

At $\sigma = 101/1000$ the exponent is 1803/2000. Combined with $MN_1N_2N_3 \asymp X$, this is $M \ll X^{1/4-3\sigma/2} = X^{197/2000}$. The individual bounds $X^{2\sigma} \ll N_\nu \ll X^{1/2-\sigma}$ do not imply the same estimate: three factors of size $X^{2\sigma}$ have product $X^{6\sigma} = X^{303/500}$, and the complementary exponent $1 - 6\sigma = 197/500$ is strictly larger than 197/2000.

Multiplying each smooth scale by a fixed factor in $[2^{-A}, 2^A]$ reparameterizes the profile on an interval whose endpoints depend only on c, C, and A. The product of the three scales changes by at most 2^{3A}, which is absorbed in the implied constant. □

The scale parameter also satisfies the inequality Polymath labels (7.4),

$$\sigma > \tfrac{1}{18} + \tfrac{28\omega}{9} + \tfrac{2\delta}{9},$$

since the right-hand side is 337/5000 = 0.0674 and $\sigma = 0.101$. This is a constraint on the frozen parameters. It is not the dispersion bound.

11.4. Small-conductor subtraction

Lemma 11.3 (Small-conductor subtraction). *For squarefree d and $(a,d)=1$,*

$$\Delta_K(g;a \bmod d)-\Delta(g;a \bmod d)=-\frac{1}{\varphi(d)}\sum_{\substack{\chi \bmod d\\ 1<\mathrm{cond}(\chi)\le K}}\overline{\chi}(a)\sum_n g(n)\,\chi(n). \tag{11.3}$$

If g is a Type III block as in Lemma 11.2, the sum of these differences against $|\lambda_d|$ satisfies

$$E_{\le K}(g)\ll X^{o(1)}MK^{5/2}\ll X^{79/800+o(1)}=X^{0.09875+o(1)}. \tag{11.4}$$

Proof. The conductor-one term in u_K is the principal term that defines Δ, so the difference retains only $1<\mathrm{cond}(\chi)\le K$. Since d is squarefree, group the characters by their primitive conductor f, write $d=fd'$ with $(f,d')=1$, and let χ^* be the primitive character modulo f. The induced character is $\chi(n)=\chi^*(n)\mathbf{1}_{(n,d)=1}$. The mask is part of the induction. Dropping it changes the sum.

Complete multiplicativity factors the character sum of the convolution:

$$\sum_n g(n)\,\chi(n)=\Big(\sum_m \alpha(m)\,\chi(m)\Big)\prod_{\nu=1}^{3}\Big(\sum_n \psi_\nu(n)\,\chi(n)\Big).$$

The sum over α is $O(M(\log X)^{O(1)})$. No cancellation in α is used.

For each smooth factor, Möbius inversion restores the mask $(n,d')=1$. There are $\tau(d')$ terms. On each of them, partial summation against the partial sums of χ^*, combined with the Pólya–Vinogradov bound $\left|\sum_{m\le u}\chi^*(m)\right|\ll\sqrt{f}\,\log f$, produces a factor $\sqrt{f}\,\log f$ times the total variation of the weight. That variation is $\ll 1+|t|$, because the change of variables $v=ru/N_\nu$ makes the variation of w_ν absolute and the Mellin factor contributes $|t|$. Thus each smooth factor contributes its own square root, and the product of the three is $\ll\tau(d')^3(\log X)^{O(1)}f^{3/2}$. The factor $\sqrt{f}$ is kept until after the product is formed.

There are at most $\varphi(f)$ primitive characters modulo f. The weight $1/\varphi(d)=1/(\varphi(f)\varphi(d'))$ cancels $\varphi(f)$. What remains is

$$E_{\le K}(g)\ll(\log X)^{O(1)}M\sum_{1<f\le K}f^{3/2}\sum_{d'}\frac{\tau(d')^3}{\varphi(d')}.$$

The inner sum over $d'\le X$ is $(\log X)^{O(1)}$. The outer sum has at most K terms, each at most $K^{3/2}$, hence is $\ll K^{5/2}$.

At the frozen values, $M\ll X^{197/2000}$ and $K^{5/2}=X^{5k/2}=X^{1/4000}$. The sum of the exponents is $395/4000=79/800$. Since $79/800<1-\varepsilon_{\mathrm{III}}$, this contribution is smaller than the target (11.1) by a fixed power of X. □

The bound uses Pólya–Vinogradov and does not use the trilinear completion. It is the reason the main argument may be written for Δ rather than for Δ_K.

11.5. Short modulus intervals

Lemma 11.4 (Short modulus intervals). *Let $\delta=X^{-\varepsilon_{\mathrm{III}}}$, $A=X^{1/2+\eta_c}$, and $B=X^{1/2+b_{\max}}$. The half-open intervals*

$$\mathcal{I}(Q_k)=[Q_k,\ (1+\delta)Q_k),\qquad Q_0=A,\quad Q_{k+1}=(1+\delta)Q_k,$$

continued until the first right endpoint exceeds B, are disjoint and cover $(A,B]$. Every integer d with $A<d\le B$ lies in exactly one of them. Their number is $\ll X^{\varepsilon_{\mathrm{III}}}\log X$. Squarefreeness, one-fold dense divisibility, and $(d,j)=1$ pass to each interval. If each interval contributes at most $X^{1-2\varepsilon_{\mathrm{III}}+o(1)}$, the sum of the intervals is $X^{1-\varepsilon_{\mathrm{III}}+o(1)}$.

Proof. The sequence Q_k is strictly increasing. The intervals $[Q_k, Q_{k+1})$ partition $[A, Q_N)$, which contains $(A, B]$. An integer equal to a right endpoint Q_{k+1} belongs to the next interval. If B itself equals some Q_k, the stopping test $Q_k > B$ fails, so the construction takes one further step and B is the left endpoint of the last interval. The endpoint A is not included.

The last left endpoint satisfies $Q_{N-1} \le B < Q_N$, so

$$(N-1)\log(1+\delta) \le (b_{\max} - \eta_c)\log X.$$

For $0 < \delta < 1$, $\log(1+\delta) > \delta/2$. Hence

$$N < 1 + 2(b_{\max} - \eta_c)\, X^{\varepsilon_{\mathrm{III}}} \log X.$$

The coefficient of the logarithm is an absolute constant. No modulus in $(A, B]$ is omitted, and none is removed by an exceptional set. Squarefreeness, dense divisibility, and coprimality to j are properties of the integer d, so they pass to every subcollection.

The weights $|\lambda_d| \ll (\log X)^{O(1)}$ are absorbed into complex coefficients of modulus at most 1, by dividing by a constant multiple of $(\log X)^{O(1)}$ and multiplying by a complex sign. Summing N interval bounds of size $X^{1-2\varepsilon_{\mathrm{III}}+o(1)}$ multiplies by $X^{\varepsilon_{\mathrm{III}}+o(1)}$ and produces the exponent $1-\varepsilon_{\mathrm{III}}$. This identity is the cost of the partition. The per-interval estimate is the content of Appendix D. □

The per-interval target passed to the completion is $X^{1-2\varepsilon_{\mathrm{III}}+o(1)}$.

11.6. Fourier completion

Lemma 11.5 (Periodic Fourier completion)**.** *Let $\Psi_\nu(y) = \sum_n \psi_\nu(n)\, e(-ny)$ and let $\widehat{\varphi}_\nu$ be the continuous Fourier transform of the profile φ_ν. Then*

$$\Psi_\nu(y) = N_\nu \sum_{k\in\mathbb{Z}} \widehat{\varphi}_\nu\big(N_\nu(k+y)\big), \tag{11.5}$$

so Ψ_ν is periodic of period 1, while $\widehat{\varphi}_\nu$ is not. For every fixed $r \ge 0$ and $A \ge 0$,

$$\big|\Psi_\nu^{(r)}(y)\big| \ll_{r,A} \frac{N_\nu^{1+r}(\log X)^{O_{r,A}(1)}}{(1+N_\nu\|y\|)^A}, \tag{11.6}$$

where $\|y\|$ is the distance from y to the nearest integer. On a short interval $q \in \mathcal{I}(Q)$ with $N_\nu \ll X^{-\sigma}Q$, discrete inversion gives

$$\psi_\nu(n) = \frac{1}{q} \sum_{-q/2<h\le q/2} \Psi_\nu(h/q)\, e(hn/q),$$

and the product of the three smooth factors carries the normalization $N_1N_2N_3/q^3$. Frequencies with $N_\nu\|h/q\| \ge X^{\varepsilon_{\mathrm{III}}/2}$ may be omitted. The choice $A = 10001$ makes the summed tail $O(X^{-1})$.

Proof. Poisson summation applied to $x \mapsto \varphi_\nu(x/N_\nu)$ produces the factor N_ν and the shifted frequencies $k+y$. Omitting N_ν changes the value. Periodicity of Ψ_ν is the relation $e(-n) = 1$.

Lemma 11.2 supplies $|\varphi_\nu^{(m)}| \ll_m (\log X)^{2m}$. Integrating the continuous transform by parts A times, with vanishing boundary terms, gives $|\widehat{\varphi}_\nu(\xi)| \ll_{r,A} (\log X)^{O(1)}(1+|\xi|)^{-A}$. The dual sum is dominated by the integer closest to $-y$, at distance $\|y\|$. A bound with $|y|$ in place of $\|y\|$, asserted for every real y, is false: the model sum $\sum_{n=1}^{N} e(-ny)$ has period 1 and size about N at every integer, and at $y = 0.98$, $N = 30$, it is still large, while $N/(1+N|y|)$ is about 1.

The support restriction $N_\nu \ll X^{1/2-\sigma} \ll X^{-\sigma}Q$, together with $q \asymp Q$, puts ψ_ν inside an interval of length less than q for large X. Discrete inversion on that interval contributes a factor $1/q$, while Ψ_ν itself carries N_ν, so each smooth factor is normalized by N_ν/q. The product of the three is $N_1N_2N_3/q^3$.

Retaining only frequencies with $N_\nu\|h/q\| < X^{\varepsilon_{\mathrm{III}}/2}$, the discarded tail of one factor has exponent at most $1 + r - A\varepsilon_{\mathrm{III}}/2$. Four further summations, coming from the moduli in one interval, the coefficient m, and the other two smooth variables, replace this by $1 + r + 4 - A\varepsilon_{\mathrm{III}}/2$. For $r = 0$ and $\varepsilon_{\mathrm{III}} = 1/1000$, the choice $A = 10001$ makes the exponent negative, while $A = 1000$ does not. The same count applies to each dual variable. □

Write $H_\nu = Q/N_\nu$ and $H = H_1H_2H_3$. The truncation retains $|h| \le X^{\varepsilon_{\mathrm{III}}/2}H_\nu$ up to the fixed comparison of q with Q.

11.7. Taylor separation

Lemma 11.6 (Taylor separation at order $J = 3230$)**.** *Define $F_\nu(z) = N_\nu^{-1}\Psi_\nu(z/N_\nu)$. Then*

$$\frac{\Psi_\nu(h/q)}{q} = \frac{N_\nu}{q}\,F_\nu\Big(\frac{N_\nu h}{q}\Big).$$

On the truncated range $|h| \le X^{\varepsilon_{\mathrm{III}}/2}Q/N_\nu$ and the interval $|q - Q| \le X^{-\varepsilon_{\mathrm{III}}}Q$,

$$\eta = N_\nu h\Big(\frac{1}{q} - \frac{1}{Q}\Big) \qquad \textit{satisfies} \qquad |\eta| \ll X^{-\varepsilon_{\mathrm{III}}/2}.$$

Writing $z_0 = N_\nu h/Q$,

$$F_\nu(z_0 + \eta) = \sum_{r=0}^{J} \frac{F_\nu^{(r)}(z_0)}{r!}\,\eta^r + \mathcal{R}_J,$$

with $|F_\nu^{(r)}(z)| \ll_r 1$ and $|\mathcal{R}_J| \ll_J X^{-(J+1)\varepsilon_{\mathrm{III}}/2}$. The phase N_ν^{it} factors out of F_ν and has modulus one. The derivative $F_\nu^{(r)}(z_0)$ depends on h, Q, N_ν, and t, and not on q. The displacement η carries the dependence on the varying modulus. With the sign

$$\Big(\frac{h}{q} - \frac{h}{Q}\Big)^r = (-1)^r\Big(\frac{h}{Q}\Big)^r\Big(\frac{q - Q}{q}\Big)^r,$$

the expansion is the shape of Polymath (7.13).

Proof. The identity for F_ν is the chain rule $\Psi_\nu(y) = N_\nu F_\nu(N_\nu y)$. The trivial bound $|\Psi_\nu^{(r)}(y)| \ll_r N_\nu^{1+r}$ uses only $|\psi_\nu| \ll 1$ and the length of the support, so $|F_\nu^{(r)}| \ll_r 1$. The logarithmic powers of Lemma 11.5 appear when integrations are used to obtain decay. They are absent from this trivial bound.

On the short interval, $|1/q - 1/Q| \ll X^{-\varepsilon_{\mathrm{III}}}/Q$. Combined with the frequency truncation,

$$|\eta| \le N_\nu \cdot X^{\varepsilon_{\mathrm{III}}/2}\frac{Q}{N_\nu} \cdot O\Big(\frac{X^{-\varepsilon_{\mathrm{III}}}}{Q}\Big) \ll X^{-\varepsilon_{\mathrm{III}}/2}.$$

Taylor's theorem with the Lagrange remainder then gives $|\mathcal{R}_J| \ll_J |\eta|^{J+1}$. The implied constant depends on J and not on q. Once J is fixed, the constant contributes no power of X.

One completed factor, summed over its frequencies and the moduli of one interval, has model error exponent $\frac{1}{2} + 2\varpi - (J+2)\varepsilon_{\mathrm{III}}/2$. At the frozen values this is -0.002 for $J = 1010$. That count does not include the other two factors or the coefficient of length M.

The product of the three completed factors has size $N/Q^3 = 1/H$ on the main terms. Replacing one factor by its remainder multiplies by $X^{-(J+1)\varepsilon_{\mathrm{III}}/2}$. Summing the three frequency ranges contributes $X^{3\varepsilon_{\mathrm{III}}/2}H$, which cancels H. For an error term no cancellation may be assumed in the inner complete congruence sum: two unit variables determine the third, so its trivial size is

at most $\varphi(q)^2 \le q^2 \ll Q^2$. The remaining sums are the $X^{-\varepsilon_{\mathrm{III}}}Q$ moduli in the interval and the coefficient length $M \ll X^{1/4-3\sigma/2}$. The resulting exponent is

$$\tfrac{3}{4} + 2\varpi - \tfrac{3}{2}\sigma + 1 + 2b_{\max} - \frac{J\varepsilon_{\mathrm{III}}}{2} = \frac{80523 - 25J}{50000}.$$

At the former order $J = 1210$ this exponent is $50273/50000 > 1$, so that order does not control the error. We take $J = 3230$, for which the per-interval exponent is $-227/50000 = -0.00454$. The three choices of which factor carries the remainder, and the $(J+1)^3$ monomials, are constants independent of X. The same order is used in the Type III assembly of Appendix D. □

The expansion is an algebraic separation. The monomials in $(q-Q)/q$ carry the modulus, and the derivatives at z_0 carry the frequency. Nothing in the construction approximates F_ν by a polynomial for the sake of numerical evaluation.

11.8. Hyper-Kloosterman sums

The classical Kloosterman sum pairs one multiplicative constraint $xy \equiv a \pmod q$ with two additive oscillations. The sum used here has three variables:

$$\mathrm{Kl}_3(a;q) = \frac{1}{q} \sum_{\substack{x_1x_2x_3 \equiv a \pmod q \\ (x_1x_2x_3,q)=1}} e_q(x_1+x_2+x_3). \tag{11.7}$$

The condition $x_1x_2x_3 \equiv a \pmod q$ is the multiplicative constraint. The additive character oscillates as the variables run through the units. The factor $1/q$ is the normalization of this paper. The sum is complete: every admissible triple modulo q appears. An incomplete sum would restrict the variables to proper subintervals, and the completion of Section 11 is arranged so that the sums which remain are complete.

Deligne's Weil II bound, in the form used by Polymath, supplies square-root cancellation for this complete sum and yields $|\mathrm{Kl}_3(a;q)| \ll 1$ in the normalization (11.7). The bound is cited. A pointwise bound of this strength, summed trivially over the moduli in an interval and over the frequencies, does not produce the saving $X^{-2\varepsilon_{\mathrm{III}}}$ required of one interval. The estimate has to see the average of products of these sums. That average is Polymath's Corollary 6.26, depending on Remark 6.10. The corollary is cited with its hypotheses checked below, and it is not reproved here.

11.9. Reduction to a correlation

Lemma 11.7 (Reduction to hyper-Kloosterman correlations). *Let $H_\nu = Q/N_\nu$ and $H = H_1H_2H_3$. The completed sum carries the factor $Q/H = N/Q^2$. If a coordinate of $\mathbf{h}$ vanishes modulo q, the complete sum*

$$F(\mathbf{h},a;q) = \frac{1}{q} \sum_{\substack{n_1n_2n_3 \equiv a \pmod q \\ (n_1n_2n_3,q)=1}} e_q(\mathbf{h}\cdot\mathbf{n})$$

is independent of the invertible residue a, and that contribution is removed by the discrepancy. For the remaining frequencies, write $h_i = b_i\ell_i$ with $(\ell_i,q)=1$, and set $b = \mathrm{rad}(b_1b_2b_3)$ and $d = q/b$. If q is one-fold X^δ-densely divisible, then d is one-fold bX^δ-densely divisible. The sum factors as

$$F(\mathbf{h},a;q) = F(\overline{d}\,\mathbf{h},a;b)\,F(\overline{b}\,\mathbf{h},a;d),$$

and the second factor equals $\mathrm{Kl}_3(a h_1h_2h_3\overline{b^3};d)$. The first factor is independent of d and satisfies

$$\left|F(\overline{d}\,\mathbf{h},a;b)\right| \le \frac{\mathrm{rad}(b_1)\,\mathrm{rad}(b_2)\,\mathrm{rad}(b_3)}{b^2}.$$

Grouping the frequencies by the product $\ell = \ell_1\ell_2\ell_3$, with multiplicity at most $4\tau_3(|\ell|)$, produces the shape of Polymath (7.19), a sum of correlations $T(\mathbf{b})$. The size of $T(\mathbf{b})$ is not estimated.

Proof. The identity $Q/H = N/Q^2$ is $H = Q^3/N$. In the completed sum the inner sum over $\mathbf{n}$ is qF, and

$$\frac{1}{H} \cdot qF = \frac{Q}{H} \cdot \frac{q}{Q} F.$$

The ratio q/Q is absorbed into the bounded coefficient of the modulus. The displayed coefficient is Q/H.

If $h_3 \equiv 0 \pmod q$ and $(a, q) = 1$, then for each pair of units n_1, n_2 there is exactly one unit $n_3 \equiv a\overline{n_1 n_2}$, and the phase in n_3 equals 1. The resulting sum does not contain a. The same holds if another coordinate vanishes. These frequencies therefore contribute the same quantity at every invertible residue, and the discrepancy removes them. No absolute-value bound is required for that cancellation.

Now assume no coordinate vanishes. Let b_i be the largest divisor of h_i whose prime factors all divide q, and set $\ell_i = h_i/b_i$. For squarefree q, the radical b equals $(h_1h_2h_3, q)$, and $q = bd$ with $(b, d) = 1$. If q is y-densely divisible and $m \mid q$, the quotient q/m is my-densely divisible: a divisor of q in a window of ratio y meets q/m in a divisor in a window of ratio my. Taking $m = b$ and $y = X^\delta$ gives the claim for d.

The unit group modulo q is the product of the unit groups modulo b and modulo d, and

$$\frac{x}{q} \equiv \frac{x\overline{d}}{b} + \frac{x\overline{b}}{d} \pmod 1.$$

The additive character and the product condition both split, and $1/q = (1/b)(1/d)$. On the modulus d, each h_i is invertible. The change of variables $u_i \equiv \overline{b}\, h_i n_i \pmod d$ converts the second factor into $\mathrm{Kl}_3(ah_1h_2h_3\overline{b^3}; d)$.

Since b is squarefree, the factor modulo b is a product of prime sums. At a prime $p \mid b$ at least one coordinate of the frequency vanishes. The sum over the units collapses to a Ramanujan sum, whose absolute value is at most $(c_1, p)(c_2, p)(c_3, p)/p^2$. The product of these bounds is $\mathrm{rad}(b_1)\,\mathrm{rad}(b_2)\,\mathrm{rad}(b_3)/b^2$. The value depends only on which coordinates vanish, hence only on the b_i, and it factors out of the sums over d.

A product $\ell = \ell_1\ell_2\ell_3$ arises from at most $4\tau_3(|\ell|)$ ordered signed triples. Extending the range of each ℓ_i to a range for the product only enlarges the sum. The resulting expression is Polymath (7.19). Its estimate is the application of Polymath Lemma 7.4, Corollary 6.26, and Proposition 6.11 recorded in the next subsection. □

11.10. The correlation estimate

The sum $T(\mathbf{b})$ is an average of products of normalized hyper-Kloosterman sums. Corollary 6.26 of [10] requires each of r_1, r_2, and s to be squarefree and requires $(s, r_1) = (s, r_2) = 1$. It does not require $(r_1, r_2) = 1$. The retained factor $(r_1, r_2, m_2 - m_1)$ is the shared-prime contribution, and the modulus in the corollary is $[r_1, r_2]$. Polymath Lemma 7.4 applies that corollary and Proposition 6.11 to exactly this sum; those results, including Deligne's bound used in Proposition 6.11, are cited rather than reproved.

The three raw margins at $\varepsilon_{\mathrm{III}} = 0$ are $63/2500$, $2429/40000$, and $291/4000$. The square-root losses are $\varepsilon_{\mathrm{III}}$, $5\varepsilon_{\mathrm{III}}/4$, and $3\varepsilon_{\mathrm{III}}/4$. Together with the per-interval target $2\varepsilon_{\mathrm{III}}$, the costs are 0.003, 0.00325, and 0.00275. The allowance $13/4000$ is the second cost only. The remaining margins are 0.0222, 0.057475, and 0.070.

11.11. The exponent ledger

The comparisons used in the preceding cited estimate are as follows.

The product of the smooth scales is $X^{1803/2000}$, and $M \ll X^{197/2000}$, by the pairwise constraints. The individual floors do not give this bound.

The small-conductor contribution has exponent $79/800 = 0.09875$. The gap below the trivial exponent 1 is $721/800$, which is larger than $\varepsilon_{\mathrm{III}}$.

The partition multiplies a per-interval bound $X^{1-2\varepsilon_{\mathrm{III}}}$ by $X^{\varepsilon_{\mathrm{III}}}$ once. The product is exactly $X^{1-\varepsilon_{\mathrm{III}}}$.

The Taylor remainder of the three-factor product, including the inner complete congruence sum bounded trivially by Q^2, has exponent $(80523 - 25J)/50000$. The expansion is taken at $J = 3230$, where this is $-227/50000$. Appendix D uses the same exponent and sums the modulus intervals once.

The three remaining margins, after the actual costs rather than after a common allowance, are positive. The first is the smallest.

11.12. Assembly

The per-interval main term is $X^{1-2\varepsilon_{\mathrm{III}}+o(1)}$. Lemma 11.4 sums the intervals once:

$$X^{1-2\varepsilon_{\mathrm{III}}+o(1)} \cdot X^{\varepsilon_{\mathrm{III}}+o(1)} = X^{1-\varepsilon_{\mathrm{III}}+o(1)}.$$

Lemma 11.3 restores the conductors up to K at size $X^{0.09875+o(1)}$. The Fourier tail at $A = 10001$, after its external budget, has per-interval exponent $-1/2000$, and the corrected Taylor remainder at $J = 3230$ has per-interval exponent $-227/50000$. After the same single interval count, both remain below $X^{1-\varepsilon_{\mathrm{III}}}$. The details are the assembly in Appendix D. The factor $X^{o(1)}$ from the dyadic tuples and the eight bands is the same shape used in Section 10.

The Type I and Type II bounds do not supply this estimate: a Type III block has no bilinear splitting inside the ranges of those theorems. The combination is Section 12.

12. Assembly of the Weighted Dispersion Theorem

This section adds no exponential-sum estimate. It invokes Theorem 10.1, proved in Appendix C, and Theorem 11.1, proved in Appendix D. The classification of Heath–Brown blocks, the omitted-core sieve, and the prime-power count are Appendix E.

Theorem 12.1 (Weighted dispersion). *Let W be a fixed Gevrey-2 weight on a fixed compact subinterval of $(0,\infty)$, and let $0 < |j| \le J$ with $J \le X^{1/100}$. Then, uniformly in the invertible residue, in $|t| \le C(\log X)^2$, and in the eight bands,*

$$\sum_{\substack{d\in\mathcal{D}_1\\(d,j)=1}} |\lambda_d| \left|\Delta_K\big(W(\cdot/X)(\cdot/X)^{it}\Lambda;\, -j \bmod d\big)\right| \ll X^{1-\varepsilon_{\mathrm{d}}/2+o(1)}. \tag{12.1}$$

Proof. Appendix E assigns every Heath–Brown tuple to Type 0, to the bilinear class, or to Type III. These three predicates are mutually exclusive. Inside the bilinear class the shorter exponent is sent to Theorem 10.1, once. A Type III tuple is sent to Theorem 11.1. A Type 0 tuple is the Poisson estimate of that appendix, of exponent $20149/25000$. An omitted-core divisor lies under $X^{1/2-1/250}$ and is estimated by the large sieve in the same appendix, of exponent $9999/10000$. Both are strictly smaller than $1 - \varepsilon_{\mathrm{d}}/2$. Full-core and omitted-core divisors partition the support, so λ_d is not counted twice. The enlarged shift range $J \le X^{1/100}$ is the hypothesis of Theorem 10.1. Theorem 11.1 imposes no further restriction on $|j|$ beyond $(d, j) = 1$. □

The passage from this one-shift bound to the variance is the next section. It is not left open.

13. Shifted Correlations and the Variance Bound

The dispersion input is Theorem 12.1. The norm inputs are the four propositions of this section. Theorem 6.4 is used only for the adaptive square at shift zero. The proofs are Appendix F, and they are not repeated here.

Proposition 13.1 (Shifted correlations). *Let $B = B_0 + B_1$ and let W be a fixed Gevrey-2 weight supported in $[c_1, c_2] \subset (0, \infty)$. For $0 < |j| \le J \le c_1 X/2$,*

$$\sum_n W(n/X)\, B(n)B(n+j) = X\widehat{W}(0)\, \mathbb{E}_{\mathrm{per}}[B(n)B(n+j)] + O\big(X^{1-14453/3700000+o(1)}\big).$$

Summing the shifts $0 < |j| \le CH(\log X)^2$ with $H \le X^{\theta_1}$ produces

$$O\big(X^{1-35633/9250000+o(1)}\big).$$

The shift $j = 0$ is excluded.

Proposition 13.2 (Mixed zero shift).

$$\sum_n W(n/X)\, B_0(n)B_1(n) \ll X \exp\big(-c\sqrt{\log X}\big).$$

Proposition 13.3 (Aggregate periodic main terms). *Let $U = HX^{2\alpha_{\mathrm{col}}}$ be the lower cutoff in (3.11), with the collar $\alpha_{\mathrm{col}} > 0$ identified with the exponent η of that display, and let $J \le CH(\log X)^2$. Then*

$$\sum_{0<|j|\le J} \big|\mathbb{E}_{\mathrm{per}}[B(n)B(n+j)]\big| \ll X^{-2\alpha_{\mathrm{col}}+o(1)}.$$

The collar stays fixed.

Proposition 13.4 (Classical numerator covariance). *Let B_0 have the cutoffs (3.11), with the collar fixed. Put $\Lambda_{\mathrm{pr}}(n) = (\log n)\mathbf{1}_{n \text{ prime}}$. For the smooth numerator kernel*

$$\widetilde{\kappa}_{h,\xi}(j) = h \int_0^1 \omega_\xi(v + j/h)\, dv,$$

uniformly for every real $1 \le h \le X^{\theta_1}$, one has

$$\sum_{0<|j|\le h} \widetilde{\kappa}_{h,\xi}(j) \sum_p W(p/X)(\log p) B_0(p+j) = o_\xi\big(hX \log X\big).$$

For the Mellin kernel in Section 14, with $J \ll H(\log X)^2$, the corresponding aggregate is

$$o\big(TX \log X\big).$$

Proof. The variance statement is the smooth signed-shift estimate of [1, §2, eq. (2.1)], after writing $m = p + j$; its kernel is $\widetilde{\kappa}_{h,\xi} = h(\mathbf{1}_{(0,1)} * \widetilde{\omega}_\xi)(j/h)$, not the denominator autocorrelation kernel. The spectral statement is [1, §3.1, eqs. (3.2)–(3.4)]. Those statements are naturally prime-supported; the passage from Λ_{pr} to Λ is the separate proper-prime-power estimate in Appendix E. The parameter dictionary is $U = hX^{2\eta}$, $R = (X/h)^{1/2}X^{-10\eta}$ in the variance problem and $h = H = X/T$ in the spectral problem. The shift weights are fixed smooth functions (or their Mellin superpositions), so the Poisson completion in those cited estimates applies. The collar is held fixed while $X \to \infty$. This is signed cancellation over the shifts, not a per-shift large-sieve estimate. □

Proposition 13.5 (Variance norm). *Let $1 \le h \le X^{\theta_1}$ and let κ_h be supported on $|j| \le h$ with $|\kappa_h(j)| \le h$. Then*

$$\sum_{|j|\le h} \kappa_h(j) \sum_n W(n/X)\, B(n)B(n+j) \le hX\widehat{W}(0)\,(A_0 + A_1)\,(1 + o(1)).$$

The collar in (3.11) *stays fixed through this limit.*

13.1. Smooth sliding windows for real h

Fix $0 < \xi < 1/10$ and choose $\omega_\xi \in C_c^\infty((0,1))$, with $0 \le \omega_\xi \le 1$, equal to 1 on $[2\xi, 1-2\xi]$, and with

$$\int \omega_\xi = 1 + O(\xi), \qquad \int \omega_\xi^2 = 1 + O(\xi).$$

For every real $h \ge 1$, set

$$D_{h,\xi}(t) = \sum_{m\ge 1} B(m)\omega_\xi\left(\frac{m-t}{h}\right).$$

This definition has no integer-h restriction. If $\vartheta(x) = \sum_{p\le x} \log p$, then the change of variables $t = p - hv$ gives the exact prime-supported identity

$$\frac{1}{h}\int W(t/X)\{\vartheta(t+h)-\vartheta(t)\}D_{h,\xi}(t)\,dt = \int_0^1 \sum_{p,m} (\log p)B(m)W((p-hv)/X)\omega_\xi((m-p)/h+v)\,dv.$$

Expanding the denominator gives, with $\kappa_{h,\xi}(j) = h\int \omega_\xi(v)\omega_\xi(v+j/h)\,dv$,

$$\int W(t/X)|D_{h,\xi}(t)|^2\,dt = \sum_{|j|\le h} \kappa_{h,\xi}(j) \sum_n W(n/X)B(n)B(n+j) + O_W(h^3X^{o(1)}).$$

Writing $m = p + j$ in the preceding numerator identity and replacing $W((p-hv)/X)$ by $W(p/X)$ gives

$$\int W(t/X)\{\vartheta(t+h)-\vartheta(t)\}D_{h,\xi}(t)\,dt = \sum_{|j|\le h} \widetilde{\kappa}_{h,\xi}(j) \sum_p W(p/X)(\log p)B(p+j) + O_W(h^3X^{o(1)}).$$

The shift orientation is therefore $\Lambda_{\mathrm{pr}}(p)B(p+j)$. Thus the numerator coefficient is $\widetilde{\kappa}_{h,\xi}(0) = h\int \omega_\xi$, while the diagonal denominator coefficient is $\kappa_{h,\xi}(0) = h\int \omega_\xi^2$; their Cauchy–Schwarz ratio is

$$\frac{(\int \omega_\xi)^2}{\int \omega_\xi^2} = 1 + O(\xi).$$

Replacing ϑ by ψ is precisely the proper-prime-power term estimated separately in Appendix E. Proposition 13.4 applies to the classical part; the high-conductor, low-conductor, principal-character, proper-prime-power, and noncoprime adaptive pieces are those assembled in Appendix E. After $X \to \infty$, first the classical collar and then ξ are sent to zero. Thus no floor $\lfloor h \rfloor$ enters the coefficient of the theorem.

13.2. Six correlations and the translation error

Write $C_{ab}(j) = \sum_n W(n/X)\, B_a(n)\, B_b(n+j)$. Proposition 13.5 is the sum of the six correlations against κ_h. The Gram theorem supplies only $C_{11}(0)$. The classical square $C_{00}(0)$ is the mass A_0, up to the Gevrey tail in the proof of that proposition. The mixed sums at shift zero are Proposition 13.2. Nonzero shifts are Proposition 13.1 and Proposition 13.3. Since $\theta_1 < a_{\min}$, no core prime divides a nonzero admissible shift.

The short-interval square is not that triangular sum. The comparison is Appendix E: $E_{\mathrm{tr}}(W) \ll_W h^3 X^{o(1)}$, and the ratio to $hX(A_0 + A_1)$ is $O(X^{2\theta_1 - 1 + o(1)})$ throughout $1 \le h \le X^{\theta_1}$. The derivative of W changes sign, so the comparison is not an application of Theorem 6.4.

13.3. Numerator

The centering term in the continuous numerator is $-h \int W(t/X) D_{h,\xi}(t)\, dt$; it is rapidly decaying by the zero mean of every Ramanujan mode and the Gevrey Poisson estimate in Appendix E. It does not produce a second copy of A_0 or A_1. The B_0-against-Λ nonzero shifts are Proposition 13.4. For B_1, Appendix E assembles, in this order: proper prime powers; the principal-character progression main term; conductors $1 < f \le K$ after the imprimitive mask is removed on primes only; conductors $f > K$ through Theorem 12.1; and shifts with $(d, j) > 1$. Thus Theorem 12.1 is only the high-conductor component, not the whole adaptive covariance.

On the enlarged range $0 < |j| \le H(\log X)^2$ with $H \le X^{\theta_1}$, one has $H(\log X)^2 \le X^{1/100}$ for large X, which is the shift hypothesis of Theorem 10.1. The ratio of the high-conductor error to the principal term $HX(A_0 + A_1)$ retains the shift count,

$$X^{\theta_1 - \varepsilon_{\mathrm{d}}/2 + o(1)} = X^{-10^{-6} + o(1)}.$$

The ratio $X^{-\varepsilon_{\mathrm{d}}/2}$ is not claimed.

13.4. The variance inequality

Cauchy–Schwarz with the real-h sliding auxiliary $D_{h,\xi}$, Proposition 13.5 for its denominator, and the complete numerator assembly above proves Theorem 1.1. The Gram bound keeps the square of B_1 from exceeding its mass, so one factor of A_1 survives. The limits are $X \to \infty$, then the collar tends to zero, then $\xi \to 0$. The $o(\log X)$ is ineffective only through the Siegel–Walfisz theorem on B_0; the coefficient a_* is effective.

14. Spectral Transfer to Pair Correlation

This section is the spectral argument. It is not a division of the Selberg integral $J(X, h)$ by $XH \log T$. The Dirichlet polynomial and the Selberg integral are different objects. The factor X^{-1} is the square of the amplitude $X^{-1/2}$ in the explicit formula.

14.1. The two statistics, and what is cited

Montgomery's function F_T is (1.1). The complex-difference statistic $F_{T,\mathrm{c}}$ is (1.4). If every nontrivial zero has real part 1/2, then $\rho - \rho' = i(\gamma - \gamma')$, so the sums agree term by term. Theorem 1.2(i) is Theorem 1.2(ii) under that hypothesis.

The published theorem of Baluyot, Goldston, Suriajaya, and Turnage-Butterbaugh [2] is the unconditional Montgomery asymptotic for $0 \le \alpha \le 1$. It is not the bound past $\alpha = 1$. What this section uses from that article, and not that theorem, are three unconditional identities: the explicit formula on the line $\mathrm{Re}(s) = 3/2$ (their Lemma 1), the zero-count bound (their Lemma 2),

and the integral identity for the complex-difference sum (their Lemma 3). The argument past $\alpha = 1$ is the weighted estimate below. It is not attributed to them.

Aquila's Lemma 3, cited as Proposition 13.2, is the mixed zero-shift bound. The contour step inside that lemma is not reproved. Walker's Proposition 2.1, cited as Proposition 13.1, is the nonzero-shift discrepancy for the single divisor function $B = B_0 + B_1$. The periodic main terms are Proposition 13.3. The norm $\|R_b\|_g^2$ is Proposition 14.1. The passage from the joint kernel to Theorem 12.1 is Proposition 14.2.

Proposition 14.1 (Spectral norm). *With the collar of* (3.11) *held fixed,*

$$\|R_b\|_g^2 \le TX\, g_0(A_0 + A_1) \int b^2\,(1 + o(1)).$$

The linearization of the kernel contributes $O(TX \cdot X^{2\theta_1 - 1 + o(1)})$.

Proposition 14.2 (Mellin bridge). *On the logarithmic coordinate the joint kernel is the inverse Fourier integral of an absolutely integrable transform. Truncation at* $|t| > C_A(\log X)^2$ *costs* $O(X^{-A})$ *for every fixed* A. *On the truncated integral, Theorem 12.1 applies uniformly for* $|t| \le C(\log X)^2$ *and for* $0 < |j| \le X^{1/100}$.

14.2. Dirichlet polynomials

Let $X = T^\alpha$ with $1 \le \alpha \le \alpha_{\max}$, so $H = X/T$ satisfies $1 \le H \le X^{\theta_1}$. Define

$$a_1(u) = \mathbf{1}_{(0,1]}(u)\sqrt{u}, \qquad a_2(u) = \mathbf{1}_{(1,\infty)}(u)\, u^{-3/2}, \qquad a = a_1 + a_2.$$

Then $\int_0^\infty a(u)^2\, du = 1$. Let $0 \le g \le 1$ be a Gevrey-2 minorant supported in $(0,1)$, with $g_0 = \int_0^1 g > 0$, and let b be a Gevrey-2 approximant of a, supported away from 0, 1, and infinity, with $\int ab > 0$. Set

$$S_a(t) = \sum_{n \ge 1} a(n/X)\Lambda(n)\, n^{-it}, \qquad R_b(t) = \sum_{m \ge 1} b(m/X)B(m)\, m^{-it}.$$

Termwise integration against $g(t/T)\, dt$ produces the kernel $T\widehat{g}\big(-(T/2\pi)\log(m/n)\big)$. Proposition 14.2 justifies the truncation. Shifts $|j| > J$, with $J = \lceil HC(\log X)^2 \rceil$, and the linearization of $b((n+j)/X)$ and of $\log(1 + j/n)$ contribute $O(TX \cdot X^{2\theta_1 - 1 + o(1)})$. The diagonal $j = 0$ is $TX(A_0 + A_1)g_0 \int ab$, up to the Siegel–Walfisz error on B_0.

For $j \ne 0$, Proposition 13.4 supplies the classical B_0-against-Λ covariance. The adaptive covariance is the sum of five pieces from Appendix E: the principal-character progression main term, proper prime powers, low conductors $1 < f \le K$, high conductors $f > K$ (the only piece to which Theorem 12.1 is applied), and noncoprime shifts. The same estimates are uniform for the Mellin-inverted smooth kernel and $J \le X^{\theta_1 + o(1)} \le X^{1/100}$. Their sum is $o(TX \log X)$; the high-conductor part has relative size $X^{-10^{-6} + o(1)}$. Finally, $\operatorname{supp} b$ has a fixed gap around 1. Since $J = o(X)$, a pair with $|m - n| \le J$ cannot straddle that gap, so the two definitions of a create no unaccounted cross-gap pair. Thus

$$\langle S_a, R_b \rangle_g = TX\, g_0(A_0 + A_1) \int ab + o(TX \log X). \tag{14.1}$$

Proposition 14.1 gives

$$\|R_b\|_g^2 \le TX\, g_0(A_0 + A_1) \int b^2 + o(TX \log X), \tag{14.2}$$

Cauchy–Schwarz yields the Dirichlet-polynomial lower bound

$$\int_{\mathbb{R}} g(t/T)\,|S_a(t)|^2\,dt \geq \big(1-o(1)\big)\,TX(A_0+A_1)\,g_0\,\frac{\big(\int ab\big)^2}{\int b^2}. \tag{14.3}$$

The ratio $\gamma_{a,b} = (\int ab)^2/\int b^2$ lies in $(0,1]$ because $\int a^2 = 1$, and it can be made arbitrarily close to 1 before $T \to \infty$.

14.3. The factor $X^{-1/2}$

Write $\rho = \frac{1}{2} + \delta_\rho + i\gamma$ and

$$M_T(t) = \sum_{0<\gamma\leq T} \frac{2X^{\delta_\rho + i(\gamma - t)}}{1 + ((t-\gamma) + i\delta_\rho)^2},$$

and let M_∞ be the same sum over all nontrivial zeros. By [2, Lemma 3],

$$2\pi \cdot \frac{T\log T}{2\pi} F_{T,c}(\alpha) = \int_{\mathbb{R}} |M_T(t)|^2\,dt \geq \int_{\mathbb{R}} g(t/T)\,|M_T(t)|^2\,dt.$$

On the support of $g(t/T)$, the tails $\gamma > T$ and $\gamma < 0$ satisfy $|M_\infty - M_T| \ll_g \sqrt{X}\,\log(2T)/T$ by [2, Lemma 2]. The squared $L^2(g(t/T)\,dt)$ norm of this difference is $O(T^{\alpha-1}\log^2 T) = o(T)$ for $\alpha \leq \alpha_{\max} < 2$.

The amplitude identity is

$$X^{-1/2}a(n/X) = n^{-1/2}\min\{n/X,\ X/n\}.$$

[2, Lemma 1], the explicit formula on $\mathrm{Re}(s) = 3/2$, therefore gives, on the support of $g(t/T)$,

$$M_\infty(t) = -X^{-1/2}S_a(t) + O_g\Big(\frac{\log T}{X} + \frac{\sqrt{X}}{T^2} + \frac{X^{-5/2}}{T}\Big). \tag{14.4}$$

Squaring produces the factor X^{-1} in front of $|S_a(t)|^2$. The squared $L^2(g(t/T)\,dt)$ norm of the error is $o(T)$, uniformly for $1 \leq \alpha \leq \alpha_{\max}$. No upper bound for $\|X^{-1/2}S_a\|_g$ is used. If $M_T = -X^{-1/2}S_a + E$, the reverse triangle inequality and (14.3) give

$$\|M_T\|_g \geq \|X^{-1/2}S_a\|_g - \|E\|_g = \big(1-o(1)\big)\|X^{-1/2}S_a\|_g,$$

because the proved lower bound has $\|X^{-1/2}S_a\|_g \gg \sqrt{T\log T}$, whereas $\|E\|_g = o(\sqrt{T})$. Combining this norm inequality with (14.3) gives

$$\int_{\mathbb{R}} g(t/T)\,|M_T(t)|^2\,dt \geq \big(1-o(1)\big)\,T(A_0+A_1)\,g_0\gamma_{a,b} - o(T).$$

Dividing by $T\log T$ is the last step, and only this step:

$$F_{T,c}(\alpha) \geq \big(g_0\gamma_{a,b} - o(1)\big)\frac{A_0+A_1}{\log T} - o(1).$$

14.4. Normalization ledger

Step	Factor
Dirichlet-polynomial norm $\int g\lvert S_a\rvert^2$	$TX(A_0+A_1)\,g_0\gamma_{a,b}$
Explicit formula, amplitude $X^{-1/2}$	multiplies S_a
Square of that amplitude	X^{-1}
Spectral norm $\int g\lvert M_T\rvert^2$	$T(A_0+A_1)\,g_0\gamma_{a,b}$
Division by $T\log T$	$(A_0+A_1)/\log T$
$\frac{1}{2}\log(X/H^3)/\log T$	$\frac{3}{2}-\alpha$
$A_1/\log T$	$a_*\alpha$, from $\log X=\alpha\log T$

The X^{-1} is the square of the coefficient in (14.4). It is not inserted in order to match dimensions, and it is not the quotient $J(X,H)/(XH\log T)$.

With $\log X=\alpha\log T$ and $\log H=(\alpha-1)\log T$,

$$\frac{\frac{1}{2}\log(X/H^3)}{\log T}=\frac{3}{2}-\alpha,\qquad \frac{A_1}{\log T}\ge(a_*-o(1))\alpha.$$

The collar η in (3.11) is removed after $X\to\infty$, as in Appendix E. The order is: choose $\eta>0$ small enough that its contribution is less than any prescribed ε, hold it fixed, then choose g and b, then let $T\to\infty$, and only afterwards send η to zero. This is uniform on the closed interval $1\le\alpha\le\alpha_{\max}=500000/499973$, because $1\le H\le X^{\theta_1}$ holds at every point, including the endpoint. Under the Riemann hypothesis, $F_T=F_{T,\mathrm{c}}$. Without it, the bound is the one for $F_{T,\mathrm{c}}$.

15. Discussion

The construction is one way of designing an arithmetic test function under three constraints at once: a large harmonic mass, a Gram operator that does not exceed the identity by more than $o(1)$, and shifted correlations small enough to leave a power of X. The frozen parameters are one feasible point. They are not asserted to be the best point. The dispersion arguments are Appendices C and D. The variance inequality is Section 13. The spectral argument is Section 14. The independent integer replay of the eight production bands is recorded in `certificates/replay_v1.1.txt` and matches the frozen numerator.

15.1. The four components

The denominator family is cut into eight bands, and the least prime factor allowed in a core rises with the band. A single roughness, set high enough for the largest band, would delete cores that the smaller bands can carry. The certificate turns the mass of the bands that survive into the integer a_*. The Gram decomposition $G=I-P+R+E$ shows that this mass does not inflate the squared norm: P is positive semidefinite, R is small because a shared prime is rough, and E is negligible on the pairs the cross-band constraint permits. Conductor subtraction then splits the shifted prime sums into an outer congruence and an inner character average, on the factorization $d=qr$ that dense divisibility supplies.

None of the four steps replaces another. Mass without the Gram bound does not control $J(X,h)$. The Gram bound without dispersion does not control the shifts. Dispersion without the adaptive family has nothing past the classical range of moduli to estimate. The improvement is the simultaneous satisfaction of the three constraints.

15.2. What is new

The classical auxiliary sequence stops a fixed power below $X^{1/2}$. Aquila develops the two-family architecture, conductor-subtracted dispersion, the core/smooth factorization with a rough complementary factor, the cofactor Ramanujan identity, and the BGSTB transfer past $\alpha = 1$. Those ingredients are cited here. The additions of this paper are the eight-band adaptive roughness schedule, the one-sided structured Gram decomposition $G = I - P + R + E$, and the integer recurrence certificate. The strictest roughness in the schedule is $a_8 = 1/125$. Imposing it on every band would exclude cores that are legal in the lower bands. The certified coefficient is about 23.109 times Aquila's first-range coefficient, and the window exponent is 54/39 times as large. Aquila's second range, $(\theta_2, a_2) = (0.001, 1.8 \times 10^{-15})$, is much wider but carries a much smaller coefficient.

The exponential-sum engines are inherited: Heath–Brown's identity, the large sieve, Pólya–Vinogradov, Polymath's Propositions 5.10 and 5.12, and, for the trilinear sums, Deligne's bound and Polymath's Corollary 6.26. The structural arguments that are particular to this family are the band-dependent roughness, the certificate, the splitting $G = I - P + R + E$, and the factorization in which r may contain large primes while $P^-(q) > K$.

15.3. Mass against dispersion

The design problem is to maximize $A(\mathcal{R})$ over denominator sets $\mathcal{R}$, subject to $G_{\mathcal{R}} \preceq (1 + o(1))I$ and to the dispersion bounds that the shifts require. A larger mass obtained by admitting smaller primes, or by pushing the denominators far past $X^{1/2}$, changes the factorization $d = qr$ and the size of the off-diagonal Gram terms. Lowering a roughness threshold admits more cores and increases the chance that two denominators share a small prime, which is the coupling R measures. The shift range is limited by the dispersion saving, not by the mass. A family with a larger a_* and a weaker dispersion bound can move the crossing α_{cross} and still be forced to stop at a smaller $\alpha_{\max}$.

15.4. Which inequality limits the range

The power savings at the frozen parameters are $\varepsilon_{\text{d}}/2 = 0.000055$ for Type I and Type II, and $\varepsilon_{\text{III}} = 0.001$ for Type III. The shift exponent in use is $\theta_1 = 0.000054$. The difference is 10^{-6}. Type I/II is the limiting mechanism. Inside it, the third Type II brace meets the squared ceiling $-2\varepsilon_{\text{d}}$ exactly when $N = M$. The Type III margins are larger by orders of magnitude. Raising the Type III saving further would not move θ_1, because the shifts are already paid for by the smaller Type I/II saving. The inequalities worth improving are the balanced Type II endpoint and any argument that sums the shifts with cancellation instead of by the triangle inequality. The present sum over $|j| \le X^{\theta_1}$ spends one full factor X^{θ_1}.

15.5. Two different optimizations

The frozen vector is $(\varpi, \delta, \sigma, \varepsilon_{\text{d}}, k)$. Some constraints are structural: $a_i > b_i + b_{\max}$, $1/5 < 2\sigma$, and $k > \varepsilon_{\text{d}}/2$. Some are analytical: the Type II endpoint $N = M$, and the requirement that the three Type I braces stay under $-2\varepsilon_{\text{d}}$. At the frozen point the cross-band margin is $1/50000$ on the first seven bands, the Heath–Brown margin $2\sigma - 1/5$ is $1/500$, and the shift margin is 10^{-6}. The Type II comparison at $N = M$ has no nominal slack.

No derivative of a_* with respect to these parameters has been computed, and no second feasible point has been certified. Two objectives are nevertheless distinct:

$$\max_{\mathbf{p}} a_*(\mathbf{p}) \qquad \text{and} \qquad \max_{\mathbf{p}} \theta_{\max}(\mathbf{p}).$$

The first moves α_{cross}. The second moves $\alpha_{\max}$. Nothing in the feasible region forces them to be largest at the same $\mathbf{p}$.

15.6. Finer bands

Eight equal steps are a discretization. More bands, unequal widths, or a roughness $a(b)$ that varies with the denominator excess b, would try to follow the mass instead of a step. The picture of that limit is an integral $\int \mathcal{M}(b, a(b))\, db$, where $\mathcal{M}$ would be a mass density for cores of excess b and roughness $a(b)$. That integral is a proposal. It is not an identity, and it has not been evaluated.

Any finer partition still has to keep $a(b) > b + b_{\max}$, or the cross-band Gram estimate fails. It also has to supply a dispersion factorization at every roughness the profile uses. Figure 3 draws the eight steps that were certified and the boundary $a = t + b_{\max}$ of the constraint.

15.7. Other coefficient vectors

The decomposition $G = I - P + R + E$ uses the shape of the Ramanujan coefficients. The same splitting may exist for other families, with a different positive correction in place of P. The variational problem that the Gram bound serves is

$$\max_{v \neq 0} \frac{|v^* p|^2}{v^* G v},$$

where p is the vector of arithmetic masses. The present v is the vector of normalized Ramanujan coefficients. A different v could raise the ratio. It would also change the shifted sums, which are controlled for this v and are not controlled for a general v. That pair of questions, a better Rayleigh quotient together with a dispersion estimate for the new coefficients, is a larger change than a retuning of (ϖ, δ, σ).

15.8. Stronger dispersion

Two-fold or higher dense divisibility would open Polymath's estimates that this paper does not use. A different factorization of d, or a Cauchy–Schwarz arrangement that keeps more cancellation before the absolute values, would attack the Type I/II braces directly. Cancellation across the eight bands, if it exists, is discarded when the bands are summed by the triangle inequality. The same is true of the shifts. Replacing $\sum_{|j| \le X^{\theta_1}} E_j \le X^{\theta_1} \max E_j$ by an estimate that sees oscillation in j would spend less than the full factor X^{θ_1}, and the spectral endpoint would move even if every individual E_j stayed the same.

15.9. What the computations are

The rational parameters are checked as exact fractions. The mass a_* is an integer produced by downward rounding: the C++ certificate and the arbitrary-precision Python transcription agree on the reduced cross-check cell, and the production integer is the output of the C++ run. That integer is a certificate for a lower bound. It is not a floating-point estimate.

Finite Gram matrices, character tables, Fourier tails, Taylor polynomials, and Kloosterman factorizations appear in the notebooks as checks of identities on small inputs, or as pictures of a decay that a theorem already states. An exponent ledger is a comparison of rational numbers. It becomes a proof only when the corresponding hypothesis match has been written. A notebook that marks a lemma closed is not a substitute for the lemma.

15.10. Limitations

The spectral interval past $\alpha = 1$ has length about 5×10^{-5}. The lower bound is stronger than $\frac{3}{2} - \alpha$ on that interval and is not the pair-correlation conjecture. It does not imply the Riemann hypothesis. The margin 10^{-6} is large enough for an asymptotic $o(1)$ and small enough that a practical check at moderate height is not informative. The frozen point has not been compared with other feasible points. Finite plots in the paper are illustrations or certificates of the kind just described. They do not replace the limits $X \to \infty$ and $T \to \infty$.

The chain from the dispersion estimates to Theorems 1.1 and 1.2 is Sections 13 and 14, invoking Appendix F. Theorem 1.1 does not assume the Riemann hypothesis. Theorem 1.2 is stated for $F_{T,c}$ with no hypothesis on the zeros, and for F_T only under the Riemann hypothesis. The external theorems are cited with their hypotheses matched and are not reproved. In particular, Aquila's Lemma 3 and the smooth covariance estimates of that preprint are inputs, and that preprint had not been refereed when this manuscript was prepared. The integer replay checks the mass recurrence. It is not a check of those citations.

15.11. Reproducibility

The eight core integers are produced by the C++ program in the `certificates` directory and transcribed in the Python file beside it. The inequality checker is the Python program in that same directory. The parameters in Appendix A are the ones used by those programs. Compiling the manuscript requires a standard TEX Live installation providing `newpxtext` and `newpxmath`.

15.12. Assistance

Language-model tools were used to draft, check, and revise text and code. They were not a source of the cited theorems. Every numerical constant that enters a theorem was recomputed by the certificate or the inequality checker, not accepted from a model.

A. Notation, Frozen Parameters, and Verification Certificate

This appendix collects the arithmetic notation, the conductor-subtracted discrepancy conventions, the frozen rational parameter schedule, the exact rational inequality ledger checked by `certificates/verify_inequalities.py`, and the computational verification summary for the mass recurrence in Section 5. The Gevrey estimates are proved in Appendix B, the Type I/II dispersion derivations in Appendix C, the Type III derivations in Appendix D, the global numerator and translation assembly in Appendix E, and the auxiliary-norm propositions in Appendix F.

A.1. Arithmetic and Asymptotic Notation

Throughout the paper, X denotes a sufficiently large real parameter and $L := \log X$. We write $e(z) := \exp(2\pi i z)$ and $e_q(a) := e(a/q)$. For integers $d \geq 1$, $\mu(d)$ is the Möbius function, $\varphi(d)$ is Euler's totient, $\tau(d)$ is the divisor function, $\omega(d)$ is the number of distinct prime factors, and

$$c_q(n) := \sum_{\substack{a \bmod q \\ (a,q)=1}} e_q(an) = \sum_{d \mid (q,n)} d\,\mu(q/d)$$

is the Ramanujan sum. We write $n \sim N$ for $N \leq n < 2N$ (or $cN \leq n \leq CN$ with fixed $0 < c < C$), and $A \lessapprox B$ for $|A| \ll X^{o(1)}B$ (i.e., bounded by $C_\varepsilon X^\varepsilon B$ for every fixed $\varepsilon > 0$, or $(\log X)^{O(1)}B$ where specified).

A.2. Densely Divisible Integers and Conductor Subtraction

Following Polymath [10, Def. 2.1], for a bounded subset $I \subset (0,\infty)$ let $P_I := \prod_{p\in I} p$ and let $\mathcal{S}_I$ be the set of squarefree positive integers whose prime factors lie in I. For $y \geq 1$, an integer d is *y-densely divisible* (1-fold densely divisible) if for every $1 \leq R \leq d$ there exists a divisor $r \mid d$ with $R/y \leq r \leq R$. We denote

$$\mathcal{D}_1^{\mathrm{dd}}(I;y) := \{d \in \mathcal{S}_I : d \text{ is } y\text{-densely divisible}\}.$$

For an arithmetic sequence $g:\mathbb{N}\to\mathbb{C}$, a modulus $d \geq 1$, and a reduced residue class $a \pmod d$ (so $(a,d)=1$), we define the ordinary discrepancy and Drappeau's K-conductor-subtracted discrepancy [4] by

$$\Delta(g;a \bmod d) := \sum_{n\equiv a \pmod d} g(n) - \frac{1}{\varphi(d)} \sum_{(n,d)=1} g(n), \tag{A.1}$$

$$u_K(n,a;d) := \mathbf{1}_{n\equiv a \pmod d} - \frac{1}{\varphi(d)} \sum_{\substack{\chi \bmod d \\ \mathrm{cond}(\chi)\leq K}} \overline{\chi}(a)\chi(n), \tag{A.2}$$

$$\Delta_K(g;a \bmod d) := \sum_n g(n)\, u_K(n,a;d). \tag{A.3}$$

When $K=1$, only the principal character contributes, so $\Delta_1(g;a \bmod d) = \Delta(g;a \bmod d)$. For general $K \geq 1$, the two discrepancies differ by the non-principal small-conductor characters:

$$\Delta_K(g;a \bmod d) - \Delta(g;a \bmod d) = -\frac{1}{\varphi(d)} \sum_{\substack{\chi \bmod d \\ 1<\mathrm{cond}(\chi)\leq K}} \overline{\chi}(a) \sum_n g(n)\chi(n). \tag{A.4}$$

A.3. Frozen Global Parameters and Eight-Band Schedule

All analytic exponents in the construction are held strictly fixed at the rational values in Table 1. The derived scale parameters are

$$\begin{aligned} n_0 &= \tfrac{1}{2} - \sigma = \tfrac{399}{1000} = 0.399, \\ n_{\mathrm{split}} &= \tfrac{1}{2} - 2\varpi - 4\varepsilon_{\mathrm{d}} = \tfrac{12389}{25000} = 0.49556, \\ \kappa_{\mathrm{disp}} &= \varepsilon_{\mathrm{d}}/2 = \tfrac{11}{200000} = 0.000055, \\ \alpha_{\max} &= 1/(1-\theta_1) = \tfrac{500000}{499973} \approx 1.000054002916, \\ \alpha_{\mathrm{cross}} &= 1/(1-a_*) \approx 1.000044550358. \end{aligned}$$

The extra denominator set is partitioned into eight adaptive-roughness bands ($i = 1,\dots,8$) with exponents (b_i, a_i) listed in Table 2. Every band satisfies the three structural invariants

$$a_i - b_i - b_{\max} \geq \frac{1}{50000} = 0.000020 \qquad (1 \leq i \leq 8), \tag{A.5}$$

$$a_{\min} - b_{\max} = \frac{13}{25000} = 0.000520 > 0, \tag{A.6}$$

$$a_i - b_i \geq \frac{1}{250} = 0.004000 \qquad (1 \leq i \leq 8). \tag{A.7}$$

Table 1: Frozen global rational parameters.

Symbol	Exact Value	Decimal	Role in Proof
ϖ	1/500	0.00200	Polymath modulus excess ceiling $Q \le X^{1/2+2\varpi} = X^{0.504}$
δ	253/10000	0.02530	Dense-divisibility parameter $y = X^{\delta}$
σ	101/1000	0.10100	Bilinear/trilinear split parameter ($n_0 = \frac{1}{2} - \sigma = 0.399$)
ε_{d}	11/100000	0.00011	Polymath §5 completion parameter (Polymath's ε)
c_0	$4\varepsilon_{\mathrm{d}} = 11/25000$	0.00044	Type I / Type II split offset ($n_{\mathrm{split}} = 0.49556$)
k	1/10000	0.00010	Conductor-subtraction cutoff $K = X^k$
$b_{\max}$	199/50000	0.00398	Top extra-band denominator exponent ($< 2\varpi = 0.00400$)
$a_{\min}$	9/2000	0.00450	Minimum core-prime roughness exponent across all 8 bands
η_c	1/100000	0.00001	Core lower-size floor $c > X^{1/2+\eta_c}$
K_{HB}	5	5	Heath–Brown identity depth ($1/K_{\mathrm{HB}} = 0.200 < 2\sigma$)
$\varepsilon_{\mathrm{III}}$	1/1000	0.00100	Polymath §7 Type III finer-than-dyadic q-interval parameter
θ_1	27/500000	0.000054	Frozen shift exponent ceiling ($< \varepsilon_{\mathrm{d}}/2 = 0.000055$)

Table 2: Frozen eight-band adaptive roughness schedule and exact rational invariants.

Band i	b_i (exact)	a_i (exact)	b_i (dec.)	a_i (dec.)	$a_i - b_i - b_{\max}$	$a_i - b_i$
1	1/2000	9/2000	0.00050	0.00450	0.000020 (1/50000)	0.004000 (1/250)
2	1/1000	1/200	0.00100	0.00500	0.000020 (1/50000)	0.004000 (1/250)
3	3/2000	11/2000	0.00150	0.00550	0.000020 (1/50000)	0.004000 (1/250)
4	1/500	3/500	0.00200	0.00600	0.000020 (1/50000)	0.004000 (1/250)
5	1/400	13/2000	0.00250	0.00650	0.000020 (1/50000)	0.004000 (1/250)
6	3/1000	7/1000	0.00300	0.00700	0.000020 (1/50000)	0.004000 (1/250)
7	7/2000	3/400	0.00350	0.00750	0.000020 (1/50000)	0.004000 (1/250)
8	199/50000	1/125	0.00398	0.00800	0.000040 (1/25000)	0.004020 (201/50000)

A.4. Exact Rational Inequality Ledger

The analytic inequalities required by the Polymath Type I/II/III dispersion estimates, Drappeau conductor subtraction, inner large sieves, Walker's off-diagonal estimate, Heath–Brown's identity at depth $K_{\mathrm{HB}} = 5$, and the roughness invariants (A.5)–(A.7) are checked in exact rational `Fraction` arithmetic by `certificates/verify_inequalities.py`:

(1) **Type I and Type II braces.** The six braces of Lemmas 10.8 and 10.9 are compared against the squared ceiling $-2\varepsilon_{\mathrm{d}} = -11/50000$ on $[n_0, n_{\mathrm{split}}]$ and $[n_{\mathrm{split}}, 1/2]$. At n_0 the first two Type I braces equal $-13/50000$ (margin $1/25000$) and $-30589/100000$; at n_{split} the third equals $-33/100000$. At n_{split} the first two Type II braces lie strictly below the ceiling with margins $24951/200000$ and $40223/100000$. At $n = 1/2$ with $R = NX^{-3\varepsilon_{\mathrm{d}}}$, the third Type II brace meets the squared ceiling $-2\varepsilon_{\mathrm{d}}$ with zero nominal margin; fixed dyadic factors remain inside the permitted $X^{-2\varepsilon_{\mathrm{d}}+o(1)}$ squared-brace target, yielding $1 - \varepsilon_{\mathrm{d}}/2 = 199989/200000 = 0.999945$ after the two square roots.

(2) **Type III margins and Taylor order.** The raw savings at $\varepsilon_{\mathrm{III}} = 0$ are $63/2500 = 0.025200$, $2429/40000 = 0.060725$, and $291/4000 = 0.072750$. Subtracting the three actual costs $3\varepsilon_{\mathrm{III}} = 0.003000$, $13\varepsilon_{\mathrm{III}}/4 = 0.003250$, and $11\varepsilon_{\mathrm{III}}/4 = 0.002750$ leaves positive margins $111/5000 = 0.022200$, $2299/40000 = 0.057475$, and $7/100 = 0.070000$. The complementary coefficient exponent is $1/4 - 3\sigma/2 = 197/2000 = 0.0985$, the small-conductor exponent is $79/800 = 0.09875$, the Fourier tail exponent at $A = 10001$ is $-1/2000$, and the per-interval Taylor remainder exponent (including the trivial Q^2 complete-sum bound) is $(80523 - 25J)/50000 = -227/50000$ at $J = 3230$. Summing the $O(X^{\varepsilon_{\mathrm{III}}} \log X)$ intervals once gives the global Type III exponent $1 - \varepsilon_{\mathrm{III}} = 999/1000 = 0.999$.

(3) **Global error budget.** The block and integration exponents are summarized below.

Contribution	Value	Kind
Type 0	$20149/25000 = 0.805960$	absolute exponent
Type 0, small conductors	$7983/20000 = 0.399150$	absolute exponent
Type I/II	$199989/200000 = 0.999945$	absolute exponent
Type III	$999/1000 = 0.999000$	absolute exponent
Omitted core	$9999/10000 = 0.999900$	absolute exponent
Low conductors	$125061/250000 = 0.500244$	absolute exponent
High conductors, summed over shifts	$999999/1000000 = 0.999999$	absolute exponent
Shift margin $\varepsilon_{\mathrm{d}}/2 - \theta_1$	$1/1000000 = 0.000001$	relative saving
Spectral exponent θ_1	$27/500000 = 0.000054$	length exponent

A.5. Certificate Checksums

The production grid in Section 5 uses bin width $h = 100$ in units of 10^{-6} ($\Delta = 10^{-4}$), prefix scaling $Z = 10^{21}$, and simplex scaling $C = 10^{15}$. The parameter block together with the eight certified core integers of Section 5.9 has SHA-256 digest

`f9fde41fe80aad92394108d57a2c9211bdedc453a3c9855e6d789591ab5f3063,`

and the production source `certificates/typeIII_core_certificate.cpp` has SHA-256 digest

`ff9c40838b5b757965419179c49b5b1cebdaa3d40269110fa67d6a0099bfc6cd.`

B. Gevrey Weights and Fourier Decay

The one-variable estimates in this appendix are the proofs behind Section 7. The joint derivative bound used by Proposition 14.2 is the one recorded here and in Section 7.

B.1. Conventions

The Fourier transform on $\mathbb{R}$ is $\widehat{\phi}(t) = \int_{\mathbb{R}} \phi(y)\, e^{-ity}\, dy$, and the inversion formula is

$$\phi(y) = \frac{1}{2\pi} \int_{\mathbb{R}} \widehat{\phi}(t)\, e^{ity}\, dt,$$

when ϕ is Schwartz, and by density when $\widehat{\phi}$ is integrable. Differentiation under the integral multiplies by $(it)^m$. Translation by a multiplies $\widehat{\phi}$ by e^{-ita}. Scaling $\phi(y/\ell)$ multiplies the transform by ℓ and replaces t by ℓt. Convolution becomes a product. The additive character on $\mathbb{Z}/q\mathbb{Z}$ is $e_q(a) = \exp(2\pi i a/q)$. The periodic sum $\sum_{n=1}^{q} f(n)\, e_q(-hn)$ is the discrete transform used for Ramanujan sums. It is not the same normalization as the continuous transform: the factor 2π sits in the exponential, and there is no $1/(2\pi)$ in front of the sum over residues.

The logarithmic transform of a weight $W(u) = \phi(\log u)$ is $\widehat{\phi}$. Equation (7.1) is the inversion formula in the coordinate $y = \log u$, evaluated at $u = n/X$. The kernel (7.10) writes $\widehat{g}(\xi) = \int g(v)\, e^{-2\pi i v\xi}\, dv$. The factor 2π is in the exponential, so $\xi = -u/(2\pi\beta)$ matches the angular convention used for $\widehat{\phi}$: the substitution $v = \beta x$ turns $e^{-2\pi i v\xi}$ into e^{iux}. Both normalizations are used as written. They are not interchangeable without that factor.

B.2. The kernel

Fix $\ell > 0$ and set $r_j = 6\ell/(\pi^2 j^2)$. Then $\sum r_j = \ell$. Let χ_j be the probability density of the interval $[-r_j, r_j]$. The partial convolution $K_{\ell,M} = \chi_1 * \cdots * \chi_M$ is nonnegative, even, integrates to 1, and is supported in $[-\sum_{j\le M} r_j,\ \sum_{j\le M} r_j] \subset [-\ell, \ell]$. The sequence is tight in the space of probability measures on $[-\ell, \ell]$, so it has a weak limit K_ℓ, still a probability measure on that interval. Its Fourier transform is the infinite product (7.3). The product converges uniformly on compact sets because $|\sin u/u| \le 1$ and $\sum r_j^2 < \infty$. The decay proved next makes $\widehat{K}_\ell$ integrable, so the limit has a C^∞ density, still denoted K_ℓ.

B.3. The envelope

Proposition B.1 (Stretched-exponential envelope). *For every real t, with $c_* = \sqrt{3}\,\log 2/\pi$,*

$$\boxed{\left|\widehat{K}_\ell(t)\right| \le 2\exp\left(-c_*\sqrt{\ell|t|}\right)}.$$

Proof. The product is even. Let $t > 0$ and $J = \left\lfloor \sqrt{3\ell t}/\pi \right\rfloor$. If $J = 0$, then $c_*\sqrt{\ell t} < \log 2$, so the right-hand side is at least 1 and dominates the product. If $J \ge 1$, then $r_j t \ge 2$ for every $j \le J$, hence each of those factors is at most $1/2$. The others are at most 1. Therefore the product is at most 2^{-J}. The inequality $J \ge \sqrt{3\ell t}/\pi - 1$ gives $2^{-J} \le 2\exp(-c_*\sqrt{\ell t})$. □

Integrability follows by comparison with $\int_0^\infty \exp(-c_*\sqrt{\ell t})\, dt < \infty$.

B.4. Derivatives

Fourier inversion and the envelope give

$$\|K_\ell^{(m)}\|_\infty \le \frac{1}{2\pi} \int_{\mathbb{R}} |t|^m \left|\widehat{K}_\ell(t)\right| dt \le \frac{2}{\pi} \int_0^\infty t^m \exp(-c_*\sqrt{\ell t})\, dt.$$

The substitution $u = c_*\sqrt{\ell t}$ produces

$$\|K_\ell^{(m)}\|_\infty \le \frac{4}{\pi c_*^2 \ell} \left(\frac{1}{c_*^2 \ell}\right)^m (2m+1)!.$$

The bound $(2m+1)! = (2m+1)\binom{2m}{m}(m!)^2 \le 2\cdot 8^m (m!)^2$ uses $\binom{2m}{m} \le 4^m$ and $2m+1 \le 2^{m+1}$. Therefore, with $B_0 = 8/c_*^2$ and $C_0 = 8/(\pi c_*^2)$,

$$\boxed{\|K_\ell^{(m)}\|_\infty \le \frac{C_0}{\ell}\left(\frac{B_0}{\ell}\right)^m (m!)^2}.$$

This is Gevrey regularity of order 2.

B.5. Closure

If f and g obey $\|f^{(m)}\|_\infty \le AB^m(m!)^2$ and the same bound for g, Leibniz's rule gives

$$\|(fg)^{(m)}\|_\infty \le \sum_{k=0}^{m} \binom{m}{k} AB^k (k!)^2 \, AB^{m-k}((m-k)!)^2.$$

The summand equals $A^2B^m(m!)^2/\binom{m}{k}$, so the product is at most $A^2(m+1)B^m(m!)^2$. The factor $m+1$ is absorbed by replacing B with $2B$, since $m+1 \le 2^m$ for $m \ge 1$. Convolution against a probability density supported on a fixed interval preserves the class, with a constant depending on that interval, because it is integration against translates. Translation and scaling replace B by $B/|\lambda|$ when the argument is multiplied by $\lambda \ne 0$.

Composition with a function analytic on a neighborhood of a compact set uses Faà di Bruno's formula. The terms are indexed by the set partitions of the derivative order. An analytic outer function contributes a Cauchy factor $MR^{-k}k!$, with k the number of blocks and R fixed by the neighborhood. Grouping by the number of blocks, the series is majorized by a geometric series in B/R. The logarithm on an interval bounded away from 0 and ∞, and the reciprocal on an interval bounded away from 0, are analytic on a neighborhood of that interval, so both preserve the class. A nonsmooth arithmetic coefficient is not given a derivative bound. Only the factor that carries the aperture is.

B.6. The aperture

Let $\phi = \mathbf{1}_{[y_1+\ell,\ y_2-\ell]} * K_\ell$, with $y_2 - y_1 > 2\ell$. Since K_ℓ is a probability density on $[-\ell, \ell]$, one has $0 \le \phi \le 1$, the support of ϕ lies in $[y_1, y_2]$, and $\phi = 1$ on $[y_1 + 2\ell,\ y_2 - 2\ell]$. The plateau of height 1 has width $y_2 - y_1 - 2\ell$, so the displayed hypothesis ensures that a plateau remains. As $\ell \to 0$, the measures K_ℓ converge weakly to a point mass, and ϕ converges pointwise to the indicator off the two endpoints. The weight $W(u) = \phi(\log u)$ is supported on $[e^{y_1}, e^{y_2}]$.

B.7. The Mellin tail

Inversion gives

$$W(n/X) = \frac{1}{2\pi}\int_{\mathbb{R}} \widehat{\phi}(t)\, X^{-it}\, n^{it}\, dt.$$

Since $|\widehat{\phi}(t)| \le (y_2 - y_1)\,|\widehat{K}_\ell(t)|$, the tail estimate of Proposition 7.2 applies with the same cutoff $C_A = (A+2)^2/(c_*^2\ell)$. The integral of $|\widehat{\phi}|$ over the whole line is finite and depends on ℓ and on the length $y_2 - y_1$, not on X. The identity $|(n/X)^{it}| = 1$ lets the tail pass outside the arithmetic sum. For the discrepancy, a progression of modulus d contains $\ll X/d + 1$ integers of size X. With $|\lambda_d| \ll (\log X)^2$ and $\sum_{d \le X^{1/2+b_{\max}}} 1/d \ll \log X$,

$$\sum_d |\lambda_d| \Big| \sum_{n \equiv -j \pmod d} r_A(n)\,\Lambda(n) \Big| \ll X^{1-A+o(1)},$$

where $|r_A(n)| \ll_A X^{-A}$. At $A = 10$ this is $O(X^{-9+o(1)})$, the majorant used by Theorem 12.1, and it is smaller than $X^{199989/200000}$.

B.8. Twisted profiles

Let $\phi_t(u) = w(u)\,L(u)\,u^{it}$, with w smooth and supported on a fixed compact subinterval of $(0,\infty)$ and L either 1 or a logarithm of u normalized by $\log N$. The derivative of u^{it} is $it\,u^{it-1}$, and the mth derivative is the falling factorial $it(it-1)\cdots(it-m+1)\,u^{it-m}$, of size at most $(|t|+m)^m u^{-m}$. On the support of w, the powers of u^{-1} are bounded. Leibniz's rule therefore yields $|\phi_t^{(m)}| \ll_{m,w} (1+|t|)^m$. On the frequencies $|t| \le C_A(\log X)^2$ this is $O_m((\log X)^{2m})$. Rescaling $u = n/N$ factors out N^{it}, which has modulus one and does not depend on u. A coefficient that has not been multiplied by w is not differentiated.

B.9. The joint kernel

The kernel of the spectral transfer is (7.10). The following lemma supplies the mixed estimate that is needed below.

Lemma B.2 (Joint Gevrey bridge). *Let $F, g \in G_c^2(\mathbb{R})$, and suppose the supports are contained in fixed compact sets. For*

$$H(s,x) = F(s)g(e^s x)$$

there are constants $A, B > 0$, depending only on the functions and their supports, such that for all $m, N \ge 0$,

$$\|\partial_s^m \partial_x^N H\|_{L^1(\mathbb{R}^2)} \le AB^{m+N}(m!)^2(N!)^2. \tag{B.1}$$

Consequently its two-dimensional Fourier transform satisfies

$$|\widehat{H}(t,u)| \le C\exp\Big(-c_1\sqrt{|t|} - c_2\sqrt{|u|}\Big). \tag{B.2}$$

The same conclusions hold after multiplication by finitely many fixed compactly supported Gevrey–2 factors in s.

Proof. Put $y = e^s x$. First

$$\partial_x^N H(s,x) = F(s)e^{Ns}g^{(N)}(y).$$

For $c \ge 0$, the Euler operator identity gives

$$\partial_s^c g^{(N)}(e^s x) = \sum_{r=0}^{c} \left\{ {c \atop r} \right\} y^r g^{(N+r)}(y),$$

where $\left\{ {c \atop r} \right\}$ is a Stirling number of the second kind. On the fixed support, powers of y and e^s are bounded exponentially in the derivative order. We use

$$\left\{ {c \atop r} \right\} \le \frac{r^c}{r!}, \qquad (N+r)! \le 2^{N+r}N!\,r!, \qquad \sum_{r=0}^{c} \left\{ {c \atop r} \right\} (r!)^2 \le C_0^c (c!)^2.$$

The last inequality follows from $r^c r! \le c^c c! \le e^c (c!)^2$, followed by summation over $r \le c$.

Apply Leibniz's rule, assigning a, b, c derivatives respectively to $F(s)$, e^{Ns}, and $g^{(N)}(e^s x)$, with $a+b+c = m$. Since $N^b \le e^N b!$, the Gevrey bounds and the preceding estimates make the corresponding term at most a constant to the power $m+N$, times

$$\frac{m!}{a!\,b!\,c!}(a!)^2 b!(c!)^2(N!)^2 = m!\,a!\,c!\,(N!)^2 \le (m!)^2(N!)^2.$$

There are at most 3^m triples. The joint support of H has fixed finite measure, so the same estimate holds in L^1, proving (B.1). An additional finite product of Gevrey–2 factors is handled by the same multinomial estimate.

Integrating by parts m times in s and N times in x gives

$$|\widehat{H}(t,u)| \le \frac{AB^{m+N}(m!)^2(N!)^2}{|t|^m|u|^N}.$$

For $v \ge 4B$, choose $k = \lfloor\sqrt{v/(4B)}\rfloor$. Since $k! \le k^k$,

$$\frac{B^k(k!)^2}{v^k} \le 4^{-k} \le Ce^{-c\sqrt{v}}.$$

Choose m from $v = |t|$ and N from $v = |u|$; bounded frequencies are absorbed into C. This proves (B.2). □

In (7.10), the substitution $v = \beta x$ writes

$$\widehat{g}\Big(-\frac{u}{2\pi\beta}\Big) = \beta \int g(\beta x)e^{iux}\,dx.$$

With $s = \log\beta$, every resulting amplitude is a finite product of fixed Gevrey–2 factors and a factor of the form covered by Lemma B.2. Hence

$$\boxed{\big|\widehat{K}_{\log}(t,u)\big| \le C\exp\big(-c_1\sqrt{|u|} - c_2\sqrt{|t|}\big).}$$

The arithmetic comparison with the dispersion margin is the one already used for a single weight. At most $X^{\theta_1}(\log X)^{O(1)}$ shifts are retained, each contributing at most $X^{1-\varepsilon_{\mathrm{d}}/2+o(1)}$, and $\varepsilon_{\mathrm{d}}/2 - \theta_1 = 10^{-6}$. A tail $O(X^{-A})$ per shift, summed over that range, remains smaller.

B.10. The order of limits

The collar ℓ, the support of ϕ, and the functions a, b, and g are fixed before $X \to \infty$. Every implied constant in the one-variable bounds depends on those choices and on A, and not on X. After the limit in X, the collar may be sent to zero. If ℓ were allowed to shrink like a negative power of X before that limit, the factor B_0/ℓ in the derivative bound would become a power of X, and the frequency cutoff C_A would grow by that same power. That is a different argument.

C. Type I and Type II Derivations

This appendix records the derivation of Theorem 10.1. No parameter is changed, and the exponential-sum arguments inside Polymath's Propositions 5.10 and 5.12 are cited rather than reproduced. The classical large sieve is cited in the same way. The corrected outer expansion is Lemmas 10.4–10.7; this appendix does not introduce a second version of it.

Throughout, S_1 and S_2 are unnormalized sums and Σ is a squared form. A bound for Σ becomes a bound for S_1 only after the square root in Section C.9. The factor $X^{\varepsilon_{\mathrm{d}}}$ inside a dual length is not part of that square root.

C.1. Notation

Polymath's x, N, M, and ε are X, N, M, and ε_{d}, with $MN \asymp X$ and $N \le M$. After the completion, Polymath's q_i is our q_0q_i'.

Quantity	Polymath	Parallax	Reference
X	x	arithmetic scale	Theorem 10.1
M, N	M, N	coefficient scales	$n = \log N/\log X$
Q, R	Q, R	dyadic labels of q, r	(10.2), (10.3)
q_0	q_0	(q_1, q_2)	Lemma 10.6
L	L	$[q_1, q_2]r$	(10.14), (5.29)
H	$x^{\varepsilon}L/M$	dual cutoff	(10.14), (5.26)
K	—	conductor cutoff X^k	(8.7)
ε_{d}	ε	dispersion parameter	11/100000

The constraints Polymath labels (5.12), (5.13), (5.14), (5.15), and (5.34) are the comparisons in Lemma 10.2. The conductor K is not one of those scales. Siegel–Walfisz is not a hypothesis of Propositions 5.10 or 5.12. It is the source of the ineffective error in the base band B_0, which is a different sum.

C.2. Modulus factorization

Lemma C.1 (Full-core modulus factorization, restated). *Let $d = cu \in \mathcal{D}_1^{\mathrm{full}}$ with $(c, u) = 1$, and let c be one-fold X^{δ}-densely divisible. Set $R_0 = NX^{-3\varepsilon_{\mathrm{d}}}$ and write e_s for the part of u kept with the inner modulus, so $e_s < X^{b_{\max}}$. There is a divisor $c_1 \mid c$ such that $r = uc_1$ and $q = c/c_1$ satisfy $d = qr$, $(q, r) = 1$, both factors squarefree, and*

$$NX^{-3\varepsilon_{\mathrm{d}}-\delta} \le r \le NX^{-3\varepsilon_{\mathrm{d}}}.$$

Every prime of q divides c, so $P^-(q) > X^{a_i} \ge X^{a_{\min}} > K$. On a dyadic box $q \sim Q$, $r \sim R$,

$$Q \le X^{13063/100000}, \qquad RQ^2 \le X^{63463/100000}.$$

Proof. Dense divisibility, Definition 2.1 of [10] as recalled in Section A, supplies $r_1 \mid c$ with $R_0/(e_sX^{\delta}) \le r_1 \le R_0/e_s$, because $R_0 > X^{b_{\max}} > e_s$ at the lower end of the Type I range: $R_0 \ge X^{n_0-3\varepsilon_{\mathrm{d}}} = X^{0.39867}$. Set $c_1 = r_1$ and $q = c/r_1$. The inner modulus is $r = e_sr_1$. This is the product $r = uc_1$ of the statement once u denotes the small factor kept with r. Since d is squarefree and $(c, u) = 1$, the factors q and r are squarefree and coprime. The factor r carries every prime of e_s and every prime of c_1. It may contain large primes. The conductor argument uses only $P^-(q) > K$.

The support constraint $qr < X^{1/2+2\omega}$ and the lower bound on r give

$$Q \le X^{1/2+2\omega+3\varepsilon_{\mathrm{d}}+\delta-n}.$$

The exponent decreases in n, so its maximum on both ranges is at $n = n_0$, where it equals 13063/100000. The margin under n_0 is 26837/100000, so $Q \le N$. The same bounds give

$$RQ^2 \le X^{1+4\omega+3\varepsilon_{\mathrm{d}}+\delta-n} \le X^{63463/100000}$$

at $n = n_0$. The margin under X is 36537/100000. A dyadic box $R \le r < 2R$, $Q \le q < 2Q$ multiplies these comparisons by a fixed factor $2^{O(1)}$. That factor is $X^{o(1)}$ and is not absorbed into δ.

Polymath's Lemma 5.2 removes an exceptional set of moduli before the same scale constraints. That removal is not used. Dense divisibility supplies a divisor in every window, for every full-core modulus. □

A finite witness that r may contain a core prime is $c = 210$, $u = 11$, $c_1 = 15$, hence $q = 14$ and $r = 165$. The prime 3 divides r. In this toy $P^-(q) = 2$; the condition $P^-(q) > K$ is imposed only on the actual cores.

C.3. Conductor subtraction

Lemma 8.1 is the identity

$$\Delta_K(g; a \bmod qr) = E_1 + \frac{E_2}{\varphi(q)}, \tag{C.1}$$

with E_1 and E_2 as in (8.9) and (8.10). The sum E_2 is unnormalized. The inner contribution to the discrepancy is the quotient $E_2/\varphi(q)$, and S_2 in (10.5) sums $|\lambda_d|$ times that quotient. Section 10 writes $\mathcal{E}_2$ for the same unnormalized sum.

Proof of the identity. Since $(q, r) = 1$, the Chinese remainder theorem identifies every character modulo qr with a product $\chi_q \otimes \chi_r$. If χ_q is non-principal, then $\operatorname{cond}(\chi_q \otimes \chi_r)$ is divisible by a prime factor of q. The hypothesis $P^-(q) > K$ makes that conductor larger than K. Every character modulo qr of conductor at most K is therefore principal on $(\mathbb{Z}/q\mathbb{Z})^\times$:

$$\chi(n) = \mathbf{1}_{(n,q)=1}\, \psi(n),$$

with ψ modulo r and $\operatorname{cond}(\psi) = \operatorname{cond}(\chi) \le K$. The coprimality mask is part of the character. It is not restored later.

Orthogonality (8.2) over this subclass, using $\varphi(qr) = \varphi(q)\varphi(r)$, gives

$$\frac{1}{\varphi(qr)} \sum_{\operatorname{cond}(\chi)\le K} \overline{\chi}(a)\,\chi(n) = \mathbf{1}_{(n,q)=1}\,\frac{1}{\varphi(q)}\big(\mathbf{1}_{n\equiv a \pmod r} - u_K(n, a; r)\big).$$

The sum over all characters modulo r reproduces the progression modulo r. Removing the characters of conductor greater than K leaves u_K, defined in (8.5). Subtract the displayed average from $\mathbf{1}_{n\equiv a \pmod{qr}}$ and sum against $g(n)$. The first two sums are E_1. The sum against u_K, restricted by $\mathbf{1}_{(n,q)=1}$, is $E_2/\varphi(q)$.

The identity is linear in g and the kernel is pointwise, so it holds for complex g, including a factor n^{it}. It is false as soon as a prime factor of q is at most K: a non-principal character modulo that prime then has conductor at most K and is absent from the right-hand side. □

The pointwise kernel was checked on $(q, r, K, a) = (5, 3, 2, 1)$, $(7, 4, 3, 1)$, $(7, 4, 3, 3)$, $(35, 8, 3, 1)$, and $(11, 6, 4, 1)$. The tuple $(q, r, K) = (3, 5, 4)$ fails, as the criterion predicts. These checks do not depend on g.

C.4. The inner large sieve

On the range of Lemma C.1, the modulus r is squarefree. Each character of conductor $f \mid r$ with $f > K$ is induced by a unique primitive character $\chi^* \bmod f$. Writing $r = fz$ with $(f, z) = 1$, the induced character is $\chi^*(n)\mathbf{1}_{(n,z)=1}$, not $\chi^*(n)$ alone. Therefore

$$E_2(q, r) = \sum_{\substack{f\mid r\\ f>K}} \sum_{\chi^* \bmod f} \frac{\overline{\chi^*}(a)}{\varphi(r)} \sum_n g(n)\,\mathbf{1}_{(n,qz)=1}\,\chi^*(n). \tag{C.2}$$

The mask $\mathbf{1}_{(n,qz)=1}$ is the imprimitive cofactor together with $(n, q) = 1$. For $g = \alpha * \beta$ it is absorbed into the two coefficient sequences by (8.11). A twist n^{it} has modulus one and is absorbed in the same way. Using $\varphi(r) = \varphi(f)\varphi(z)$ and $|\lambda_d| \ll (\log X)^2$,

$$S_2 \ll (\log X)^2 \sum_{q,z} \frac{1}{\varphi(q)\varphi(z)} \sum_{K<f\le R} \frac{1}{\varphi(f)} \sum_{\chi^* \bmod f} |A(\chi^*)|\,|B(\chi^*)|.$$

Extending the sum to all q and z only enlarges it. The bound $\sum_{n\le Y} 1/\varphi(n) \ll \log Y$ contributes $(\log X)^2 = X^{o(1)}$. Roughness of q is not used as cancellation in this upper bound. It was used to place every small conductor on r.

On a dyadic segment $f \sim F$, one has $1/\varphi(f) \ll X^{o(1)}/F$ because $\varphi(f) \gg f/\log\log(f+3)$. Cauchy–Schwarz in the primitive characters, followed by the large sieve

$$\sum_{f\le 2F}\sum_{\chi^* \bmod f}\Big|\sum_{m\in\mathcal{I}} a_m\chi^*(m)\Big|^2 \ll \big(F^2+|\mathcal{I}|\big)\sum|a_m|^2 \tag{C.3}$$

for an interval $\mathcal{I}$, gives

$$\sum_{f\sim F}\sum_{\chi^*}\frac{1}{\varphi(f)}|A|^2 \ll X^{o(1)}\Big(FM+\frac{M^2}{F}\Big),$$

and the same bound with N in place of M. The hypotheses used from Iwaniec–Kowalski, Theorem 7.13, are primitive characters, coefficients supported on intervals, and $\sum|\alpha|^2 \ll X^{o(1)}M$, which follows from $|\alpha| \ll X^{o(1)}$. The inequality is not reproved. A numerical check for a few coefficients and $Q \le 5$ is a sanity test of the implementation, not a proof of the theorem.

The product of the square roots is the radical

$$\Big(FM+\frac{M^2}{F}\Big)\Big(FN+\frac{N^2}{F}\Big) = F^2MN+MN^2+M^2N+\frac{M^2N^2}{F^2}.$$

With $MN \asymp X$ this is $F^2X+XN+XM+X^2/F^2$, and

$$\sqrt{F^2X+XN+XM+\frac{X^2}{F^2}} \le F\sqrt{X}+\sqrt{XN}+\sqrt{XM}+\frac{X}{F}.$$

The middle two terms are $X/\sqrt{M}$ and $X/\sqrt{N}$. Summing the $O(\log X)$ dyadic segments produces (10.6).

Label $M \ge X^{1/2}$ and $N \ge X^{n_0}$. On $K < F \le R \le NX^{-3\varepsilon_{\mathrm{d}}}$ the four exponents are at most

X/F	$1-k$	$9999/10000$,
$X/\sqrt{M}$	$3/4$	$3/4$,
$X/\sqrt{N}$	$3/4+\sigma/2$	$1601/2000$,
$F\sqrt{X}$	$1-3\varepsilon_{\mathrm{d}}$	$99967/100000$.

The largest is X/F. The outer target is $1-\varepsilon_{\mathrm{d}}/2 = 199989/200000$, and the margin is $k-\varepsilon_{\mathrm{d}}/2 = 9/200000$.

The weight $1/\varphi(f)$ is part of the estimate. Without it, the same Cauchy–Schwarz produces $F^2\sqrt{X}$ in place of $F\sqrt{X}$. At the top of the conductor range that exponent is $3/2-6\varepsilon_{\mathrm{d}} > 1$. The variable F is the primitive conductor. The outer scale $Q \le X^{13063/100000}$ is not substituted into the slot $F\sqrt{X}$.

C.5. Outer expansion, diagonal, zero frequency, and completion

The derivation is Lemmas 10.4–10.7 in Section 10 and is not duplicated here. The average over primitive $b \bmod P_I$ precedes Cauchy–Schwarz; the diagonal bound is applied to each ordered pair before the signed four-form is assembled; the common $q_0 = 1$ zero frequency cancels only in that signed form; and the compatibility indicator remains inside the n-sum. The second congruence has modulus q_0q_2', while its pairwise-coprime CRT character has modulus q_2'.

C.6. Proposition 5.12

The Type I range is $n_0 \le n \le n_{\mathrm{split}}$. Proposition 5.12 of [10] is applied to the sum produced above. Its hypotheses, and the checks that put them in force, are the following.

Hypothesis	Where it is checked	Inequality
Phase (10.16)	Lemma 10.7	exact identity
$(q_0q_1q_2s_2, r) = 1$	Lemma C.1	$(q, r) = 1$
q_0q_ir, q_0s_2r one-fold X^δ-densely divisible	support of $\mathcal{D}_1^{\mathrm{full}}$	one fold, by construction
$(j, k) = (0, 0)$	Theorem 5.8(i) of [10]	separate factors are 0-fold
Lemma 2.10(ii), once	inside the proof of Prop. 5.12	the lcm remains one-fold
Corollary 4.16, first bound	from Prop. 4.12(ii), one q-van der Corput	cited, not reproved
$Q \le N$, (5.14)	(10.3)	margin 26837/100000 at n_0
Definition 2.5	n^{it} remains in β_n	$\lvert t\rvert \le C(\log X)^2$
$54\varpi + 15\delta + 5\sigma < 1$	frozen rationals	$397/400 < 1$

One-fold dense divisibility is the input. Since $(j, k) = (0, 0)$, the separate factors q and r are only 0-fold. In the proof of the proposition, Lemma 2.10(ii) is applied once, to the least common multiple of the three products, and that least common multiple remains one-fold. That is the hypothesis under which the first bound of Corollary 4.16 applies. No two-fold or four-fold statement is assumed. Theorems 5.8(ii) and 5.8(iii), which require those stronger folds, are not used on this range.

The Cauchy–Schwarz step that precedes the proposition leaves q_1 outside the square and duplicates only q_2. It uses $Q \le N$. The exponent of N/Q is positive on the whole Type I interval and is smallest at n_0. Just below n_0 it changes sign, so the same appeal is not made on a longer interval. Substituting H and discarding powers of q_0 in the denominator, which enlarges the sum because the first term carries $q_0^{-29/6}$ and $-29/6 < -4$, produces three braces. Their exponents, compared with the squared ceiling $X^{-2\varepsilon_{\mathrm{d}}}$, are recorded in the next section but one. The square root is the per-modulus bound, and the sum over $r \sim R$ is (5.31) on this range. The prefactors convert it into $|\Sigma_1| \ll X^{-\varepsilon_{\mathrm{d}}+o(1)}MN^2/R$. The front factor $X^{\varepsilon_{\mathrm{d}}}$ in the completion cancels the factor $X^{-\varepsilon_{\mathrm{d}}}$ coming from $1/H$. It does not cancel the factor $X^{\varepsilon_{\mathrm{d}}}$ inside the length of the dual sum, which has already been accounted for in the braces.

C.7. Proposition 5.10

The Type II range is $n_{\mathrm{split}} \le n \le 1/2$. Proposition 5.10 takes the same phase and the coprimality $(q_0q_1q_2s_1s_2, r) = 1$, now with both q_1 and q_2 duplicated. Its proof uses the second bound of Corollary 4.16, whose input is the completion bound in Proposition 4.12(i). That bound needs squarefree moduli and a phase of the shape $e_{d_1}(c_1/n)\, e_{d_2}(c_2/(n + \ell r))$. Both are present: the moduli lie in $\mathcal{D}_1$, and (10.16) is that shape. The corollary is stated under a dense-divisibility hypothesis, which the one-fold moduli satisfy. The argument does not factor the modulus further, so it does not consume a second fold. Theorems 5.8(ii) and 5.8(iii) are not invoked. The estimate is cited.

The Cauchy–Schwarz step is (5.34)–(5.35). Before it, Υ is linear in the phase sum. After it, the combinatorial prefactor is exactly

$$\frac{N^2Q^4(q_0, \ell)^2}{q_0^4} = \Big(Q^2N\frac{(q_0, \ell)}{q_0^2}\Big)^2.$$

No square root has been taken at that stage. Substituting the full length H, including the factor $X^{\varepsilon_{\mathrm{d}}}$ inside H and inside H^2, produces the three Type II braces. The same first brace is positive at $n = n_0$. Proposition 5.10 is applied only on the Type II interval.

The balanced endpoint is $N = M$, hence $n = 1/2$, together with $R = NX^{-3\varepsilon_{\mathrm{d}}}$. The third brace is the normalized form of $X^{\varepsilon_{\mathrm{d}}}R/M$, and at that endpoint its exponent equals $-2\varepsilon_{\mathrm{d}}$. For every smaller n it is strictly smaller. A fixed dyadic factor 2^L contributes $X^{o(1)}$. The brace occurs inside $|\Upsilon|^2$, so its allowed squared target is $X^{-2\varepsilon_{\mathrm{d}}+o(1)}$. Taking the square root gives $X^{-\varepsilon_{\mathrm{d}}+o(1)}$, which the

prefactors convert into the target $X^{-\varepsilon_d+o(1)}MN^2/R$ for Σ. The nominal margin on this brace is zero; no fixed inequality is asserted between $-2\varepsilon_d + o(1)$ and $-\varepsilon_d$.

The structural inequality for this range is $68\varpi + 14\delta = 2451/5000 < 1$, the second of the two inequalities recorded with (5.31).

C.8. Terminal exponents

The braces are exponents inside $|\Upsilon|^2$. They are compared with the squared ceiling $-2\varepsilon_d = -11/50000$. Appendix A recomputes the endpoint values from the primitive rationals. The symbolic forms, and the endpoint at which each is largest, are as follows. Each coefficient of n has a fixed sign, so each maximum on a closed interval is an endpoint.

Type I, on $[n_0, n_{\mathrm{split}}]$:

$$\begin{aligned} \tfrac{1}{4} + 9\varpi + \tfrac{5}{2}\delta + 9\varepsilon_d - \tfrac{5}{6}n &\le -\tfrac{13}{50000}, \\ 8\varpi + 3\delta + 11\varepsilon_d - n &\le -\tfrac{30589}{100000}, \\ \varepsilon_d - \left(\tfrac{1}{2} - 2\varpi - n\right) &\le -3\varepsilon_d = -\tfrac{33}{100000}. \end{aligned}$$

The first two decrease in n, so their maxima are at n_0. The third increases in n, so its maximum is at n_{split}. All three lie strictly below the ceiling. The rooms under the ceiling are $1/25000$, $30567/100000$, and ε_d. At $n = n_{\mathrm{split}} + 2\varepsilon_d$ the third exponent equals $-\varepsilon_d$, which is above the ceiling. That point is outside the hypothesis $N \ll X^{1/2-2\varpi-4\varepsilon_d}$.

Type II, on $[n_{\mathrm{split}}, 1/2]$:

$$\begin{aligned} 1 + 12\varpi + \tfrac{7}{2}\delta + \tfrac{25}{2}\varepsilon_d - \tfrac{5}{2}n &\le -\tfrac{4999}{40000}, \\ 8\varpi + 3\delta + 11\varepsilon_d - n &\le -\tfrac{8049}{20000}, \\ -2\varepsilon_d + 2n - 1 &\le -2\varepsilon_d. \end{aligned}$$

The first two decrease in n, so their maxima are at n_{split} and are strictly below the ceiling. The third increases in n and equals the ceiling at $n = 1/2$. At $n = n_0$ the first exponent is positive. At $n = 1/2 + \varepsilon_d$ the third exponent is 0. Both points are outside the Type II hypothesis.

A brace below $-2\varepsilon_d$ is the conclusion of Proposition 5.12 or Proposition 5.10 on the range where that proposition's hypotheses have been checked. The same formal expression can lie above the ceiling outside those ranges. The ceiling $-2\varepsilon_d$ applies inside $|\Upsilon|^2$; its square root supplies $X^{-\varepsilon_d+o(1)}$ in the target for Σ, and the final outer Cauchy–Schwarz supplies the dispersion saving $X^{-\varepsilon_d/2+o(1)}$. These are three different levels. Appendix A records that the first Type I room of $1/25000$ survives a displacement 2^{11} once $X > 2^{229167}$. That comparison is not repeated here.

C.9. Assembly

The squared form splits into the diagonal, the off-diagonal constant term, and the nonzero frequencies. Their bounds relative to MN^2/R are

$$X^{-3\varepsilon_d+o(1)}, \qquad X^{-\varepsilon_d+o(1)}, \qquad X^{-\varepsilon_d+o(1)},$$

up to an error $O_A(X^{-A})$ from the completion. Adding them gives

$$\Sigma \ll X^{-\varepsilon_d+o(1)}\frac{MN^2}{R}.$$

The diagonal is the smaller contribution. The outer Cauchy–Schwarz is $|S_1|^2 \ll X^{o(1)}MR\,\Sigma$. Substitution produces

$$|S_1|^2 \ll X^{o(1)}MR \cdot X^{-\varepsilon_d+o(1)}\frac{MN^2}{R} = X^{-\varepsilon_d+o(1)}(MN)^2,$$

because $MR \cdot MN^2/R = (MN)^2$. With $MN \asymp X$,

$$|S_1| \ll X^{1-\varepsilon_{\mathrm{d}}/2+o(1)}.$$

The inner sum contributes $|S_2| \ll X^{1-k+o(1)}$. Since $k > \varepsilon_{\mathrm{d}}/2$, the exponent $1-k$ is strictly smaller than $1-\varepsilon_{\mathrm{d}}/2$, and

$$\boxed{|S_1| + |S_2| \ll X^{199989/200000+o(1)} = X^{0.999945+o(1)}.}$$

The remaining factors are absolute or logarithmic. Eight bands and $O((\log X)^2)$ dyadic boxes contribute $X^{o(1)}$. The same bound covers $|\lambda_d| \ll (\log X)^{O(1)}$ and the coefficient sequences. The four signed pieces of Σ and the factor 8 in the dual length are absolute. None of them is a fixed positive power of X. The prefactor $\sum_m |\alpha(m)|^2$ does not depend on the shift or on t. Every bound used for Σ is uniform in the residue class, and n^{it} remains inside Definition 2.5 for $|t| \le C(\log X)^2$. The condition $(d, j) = 1$ is the range of the original sum. Summing the boxes gives (10.1).

D. Type III Derivations

This appendix records the proof of Theorem 11.1. No parameter is changed. Pólya–Vinogradov, Polymath's Corollary 6.26, and the complete-sum bound of Proposition 6.11, including Deligne's theorem, are cited with their hypotheses matched. They are not reproved. The combination with the remaining blocks is Appendix E.

D.1. Notation

Polymath's x, N, M, and the three smooth scales are X, $N_1N_2N_3$, M, and N_ν. The reference modulus of one short interval is Q.

Symbol	Meaning	Reference
M	nonsmooth coefficient scale	(11.2), (2.13)
N_ν	three smooth scales	Definition 2.5 of [10]
Q	reference modulus	Lemma 11.4
$H_\nu = Q/N_\nu$	one dual scale	after (11.6)
$H = H_1H_2H_3$	combined dual scale	$Q/H = N/Q^2$
$\mathbf{b}$	noncoprime frequency factors	Lemma 11.7
$H_{\mathbf{b}}$	reduced frequency length	$X^{3\varepsilon_{\mathrm{III}}/2}H/(b_1b_2b_3)$
S	auxiliary factorization scale	§7.3 of [10]
$T(\mathbf{b})$	terminal correlation	(7.19)
$\varepsilon_{\mathrm{III}}$	Type III parameter	1/1000

The structural constraint Polymath labels (7.4) is $\sigma > 1/18 + 28\varpi/9 + 2\delta/9$. The right-hand side equals $337/5000 = 0.0674$, and $\sigma = 101/1000$. Siegel–Walfisz is not a hypothesis of this argument.

D.2. Admissible scales

Lemma 11.2 is the coefficient statement. Each smooth factor is

$$\psi_\nu(n) = w_\nu(n/N_\nu)\, L_\nu(n/N_\nu)\, n^{it},$$

with w_ν smooth and supported on a fixed interval $[c, C] \subset (0, \infty)$, L_ν either 1 or the ratio $\log(N_\nu u)/\log(2N_\nu)$, and $|t| \le C(\log X)^2$. Write $n^{it} = N_\nu^{it}(n/N_\nu)^{it}$. The phase N_ν^{it} has modulus one and does not depend on the profile variable, so it factors out of every derivative. On $[c, C]$,

$$\partial_u^m u^{it} = u^{it-m}\, it(it-1)\cdots(it-m+1),$$

and the falling factorial is at most $(|t|+m-1)^m$. Leibniz's rule gives $|\varphi_\nu^{(m)}| \ll_m (1+|t|)^m$, hence $|\varphi_\nu^{(m)}| \ll_m (\log X)^{2m}$. This is Definition 2.5(iii). The factor α is a coefficient sequence at scale M. No derivative of α is estimated. A sequence that alternates between 1 and -1 meets the coefficient bound and has forward differences of size 2, while a smooth profile changes by $O((\log X)^{O(1)}/M)$ from one integer to the next.

The pairwise bounds $N_i N_{i'} \gg X^{1/2+\sigma}$ give

$$N_1N_2N_3 = \sqrt{(N_1N_2)(N_1N_3)(N_2N_3)} \gg X^{3/4+3\sigma/2} = X^{1803/2000}.$$

With $MN_1N_2N_3 \asymp X$,

$$M \ll X^{1/4-3\sigma/2} = X^{197/2000}.$$

The individual floors $X^{2\sigma} \ll N_\nu \ll X^{1/2-\sigma}$ do not imply this bound. Three factors of size $X^{2\sigma}$ have product $X^{6\sigma} = X^{303/500}$, and the complementary exponent $1-6\sigma = 197/500$ is strictly larger than 197/2000. Multiplying each smooth scale by a fixed factor in $[2^{-A}, 2^A]$ changes the product by at most 2^{3A} and does not change the exponent of X.

D.3. Small conductors

The difference between the two discrepancies is (11.3). The conductor-one term in u_K is the principal term that defines Δ, so the difference retains only $1 < \mathrm{cond}(\chi) \le K$.

Since d is squarefree, group those characters by their primitive conductor f. Write $d = fd'$ with $(f, d') = 1$. Then $\varphi(d) = \varphi(f)\varphi(d')$, and the induced character is $\chi(n) = \chi^*(n)\mathbf{1}_{(n,d)=1}$. The mask is part of the induction. Dropping it changes the sum. This identity $\varphi(fd') = \varphi(f)\varphi(d')$ uses the coprimality. It fails without that hypothesis.

Complete multiplicativity factors the convolution. The sum over α is $O(M(\log X)^{O(1)})$. No cancellation in α is used. For each smooth factor, Möbius inversion restores the mask $(n, d') = 1$. There are $\tau(d')$ terms. On each of them, partial summation against the partial sums of the non-principal primitive character χ^*, combined with the Pólya–Vinogradov bound $\left|\sum_{m\le u} \chi^*(m)\right| \ll \sqrt{f}\log f$, produces one factor $\sqrt{f}\log f$ times the total variation of the weight. That variation is $\ll 1+|t|$. The product of the three smooth bounds is $\ll \tau(d')^3(\log X)^{O(1)} f^{3/2}$. The square root is kept until after the product is formed.

There are at most $\varphi(f)$ primitive characters modulo f. The weight $1/\varphi(d)$ cancels $\varphi(f)$. The sum over $d' \le X$ of $\tau(d')^3/\varphi(d')$ is $(\log X)^{O(1)}$. The sum over $1 < f \le K$ of $f^{3/2}$ is $\ll K^{5/2}$. Therefore

$$E_{\le K}(g) \ll X^{o(1)} M K^{5/2}.$$

At the frozen values, $M \ll X^{197/2000}$ and $K^{5/2} = X^{5k/2} = X^{1/4000}$. The sum of the exponents is $395/4000 = 79/800 = 0.09875$. Since $79/800 < 999/1000$, this contribution is smaller than the target by a fixed power of X. The bound is uniform for $|t| \le C(\log X)^2$. It does not use the trilinear completion.

D.4. Short intervals

Let $\delta = X^{-\varepsilon_{\mathrm{III}}}$, $A = X^{1/2+\eta_c}$, and $B = X^{1/2+b_{\max}}$. The half-open intervals of Lemma 11.4 are disjoint and cover $(A, B]$. An integer equal to a right endpoint belongs to the next interval. If B itself equals some node Q_k, the stopping test fails, so the construction takes one further step and B is the left endpoint of the last interval. The endpoint A is not included. No modulus in $(A, B]$ is omitted, and none is removed by an exceptional set.

The last left endpoint satisfies $Q_{N-1} \le B < Q_N$, so $(N-1)\log(1+\delta) \le (b_{\max} - \eta_c)\log X$. For $0 < \delta < 1$, $\log(1+\delta) > \delta/2$, hence

$$N < 1 + 2(b_{\max} - \eta_c)\, X^{\varepsilon_{\mathrm{III}}} \log X.$$

The comparison $b_{\max} > \eta_c$ makes the range nonempty. Squarefreeness, one-fold dense divisibility, and $(d, j) = 1$ are properties of d, so they pass to every subcollection. The weights $|\lambda_d| \ll (\log X)^{O(1)}$ are absorbed into complex coefficients of modulus at most 1.

If each interval contributes at most $X^{1-2\varepsilon_{\mathrm{III}}+o(1)}$, the sum of the intervals is

$$X^{1-2\varepsilon_{\mathrm{III}}+o(1)} \cdot X^{\varepsilon_{\mathrm{III}}+o(1)} = X^{1-\varepsilon_{\mathrm{III}}+o(1)}.$$

The factor $X^{\varepsilon_{\mathrm{III}}}$ is applied once, in the assembly. A second application would replace the exponent by 1. The per-interval bound is not proved by the partition.

D.5. Fourier completion

The periodic sum and the continuous transform are related by (11.5). Poisson summation applied to $x \mapsto \varphi_\nu(x/N_\nu)$ produces the factor N_ν and the shifted frequencies $k + y$. Omitting N_ν changes the value. Periodicity $\Psi_\nu(y + 1) = \Psi_\nu(y)$ is the relation $e(-n) = 1$. The continuous transform is not periodic.

Lemma 11.2 supplies $|\varphi_\nu^{(m)}| \ll_m (\log X)^{2m}$. Integrating by parts A times, with vanishing boundary terms, gives $|\widehat{\varphi}_\nu(\xi)| \ll_{r,A} (\log X)^{O(1)}(1 + |\xi|)^{-A}$. The dual sum is dominated by the integer closest to $-y$, at distance $\|y\|$. A bound with $|y|$ in place of $\|y\|$, asserted for every real y, is false. The model sum $\sum_{n=1}^{N} e(-ny)$ has period 1 and size about N at every integer. At $y = 0.98$ and $N = 30$ it is still large, while $N/(1 + N|y|)$ is about 1.

The support restriction $N_\nu \ll X^{1/2-\sigma} \ll X^{-\sigma}Q$, together with $q \asymp Q$, puts ψ_ν inside an interval of length less than q for large X. Discrete inversion on that interval contributes a factor $1/q$, while Ψ_ν itself carries N_ν, so each smooth factor is normalized by N_ν/q. The product of the three is $N_1N_2N_3/q^3$.

Frequencies with $N_\nu\|h/q\| \geq X^{\varepsilon_{\mathrm{III}}/2}$ may be omitted. The discarded tail of one factor has exponent at most $1 + r - A\varepsilon_{\mathrm{III}}/2$. Four further summations, from the moduli in one interval, the coefficient m, and the other two smooth variables, replace this by $1 + r + 4 - A\varepsilon_{\mathrm{III}}/2$. For $r = 0$ and $\varepsilon_{\mathrm{III}} = 1/1000$,

$$1 + 4 - \frac{A}{2000}$$

equals $9/2$ at $A = 1000$ and equals $-1/2000$ at $A = 10001$. The same count applies to each dual variable. The truncation produces one negative power of X. It does not produce an arbitrarily strong saving.

D.6. Taylor separation

Define $F_\nu(z) = N_\nu^{-1}\Psi_\nu(z/N_\nu)$. The chain rule is $\Psi_\nu(y) = N_\nu F_\nu(N_\nu y)$, so

$$\frac{\Psi_\nu(h/q)}{q} = \frac{N_\nu}{q} F_\nu\Big(\frac{N_\nu h}{q}\Big).$$

The phase N_ν^{it} factors out of F_ν and has modulus one. Omitting it changes the value of an identity. The trivial bound $|\Psi_\nu^{(r)}(y)| \ll_r N_\nu^{1+r}$ uses only $|\psi_\nu| \ll 1$ and the length of the support, so $|F_\nu^{(r)}| \ll_r 1$. The logarithmic powers of the decay estimate appear only when integrations are used. They are absent from this trivial bound.

On $|q - Q| \leq X^{-\varepsilon_{\mathrm{III}}}Q$ and $|h| \leq X^{\varepsilon_{\mathrm{III}}/2}Q/N_\nu$,

$$\eta = N_\nu h\Big(\frac{1}{q} - \frac{1}{Q}\Big) \qquad \text{satisfies} \qquad |\eta| \ll X^{-\varepsilon_{\mathrm{III}}/2}.$$

Taylor's theorem with the Lagrange remainder gives $|\mathcal{R}_J| \ll_J |\eta|^{J+1} \ll_J X^{-(J+1)\varepsilon_{\mathrm{III}}/2}$. The implied constant depends on J and not on q. Once J is fixed, it contributes no power of X.

One completed factor, summed over its own frequencies and the moduli of one interval, has model exponent $\frac{1}{2} + 2\varpi - (J+2)\varepsilon_{\mathrm{III}}/2$. At the frozen values this is -0.002 for $J = 1010$. That count does not include the other two factors or the coefficient of length M.

The product of the three completed factors has size $N/Q^3 = 1/H$ on the main terms. Replacing one factor by its remainder multiplies by $X^{-(J+1)\varepsilon_{\mathrm{III}}/2}$. Summing the three frequency ranges contributes $X^{3\varepsilon_{\mathrm{III}}/2}H$, which cancels H. Before any Kloosterman cancellation is invoked, the inner congruence sum has $\varphi(q)^2 \le q^2 \ll Q^2$ terms; this factor must be retained in the Taylor error. The remaining sums are the $X^{-\varepsilon_{\mathrm{III}}}Q$ moduli and $M \ll X^{1/4-3\sigma/2}$. The resulting exponent is

$$\tfrac{3}{4} + 2\varpi - \tfrac{3}{2}\sigma + 1 + 2b_{\max} - \frac{J\varepsilon_{\mathrm{III}}}{2} = \frac{80523 - 25J}{50000}.$$

It equals $50273/50000 > 1$ at the former order $J = 1210$. The expansion is therefore taken at $J = 3230$, where the exponent is $-227/50000 = -0.00454$. The three choices of which factor carries the remainder, and the $(J+1)^3$ monomials, are constants independent of X.

The displacement in Polymath (7.13) is $\eta_{\mathrm{P}} = h/q - h/Q$, so $\eta = N_\nu \eta_{\mathrm{P}}$ and

$$\frac{N_\nu}{q}\frac{F_\nu^{(r)}(z_0)}{r!}\eta^r = \frac{1}{q}\frac{\Psi_\nu^{(r)}(h/Q)}{r!}\eta_{\mathrm{P}}^r.$$

Since $\eta_{\mathrm{P}} = -(h/Q)(q-Q)/q$,

$$\Big(\frac{h}{q} - \frac{h}{Q}\Big)^r = (-1)^r\Big(\frac{h}{Q}\Big)^r\Big(\frac{q-Q}{q}\Big)^r.$$

With that sign, the coefficient of the monomial is independent of q, and the factors $(Q/q)((Q-q)/q)^r$ carry the dependence on the varying modulus. The additive character in the inversion formula is the remaining arithmetic dependence on q.

D.7. The complete sum

Define

$$F(\mathbf{h}, a; q) = \frac{1}{q}\sum_{\substack{n_1 n_2 n_3 \equiv a \pmod q \\ (n_1 n_2 n_3, q) = 1}} e_q(\mathbf{h}\cdot\mathbf{n}).$$

Since $H = Q^3/N$, the identity $Q/H = N/Q^2$ is immediate. In the passage from (7.14) to the line after (7.15), the sum over $\mathbf{n}$ is qF, and

$$\frac{1}{H}\cdot qF = \frac{Q}{H}\cdot\frac{q}{Q}F.$$

The ratio q/Q is absorbed into the bounded coefficient of the modulus. The displayed coefficient is Q/H.

Suppose one coordinate vanishes, say $h_3 \equiv 0 \pmod q$, and $(a, q) = 1$. For each pair of units n_1, n_2 there is exactly one unit $n_3 \equiv a\overline{n_1 n_2}$, and the phase in n_3 equals 1. The resulting sum does not contain a. The same holds if another coordinate vanishes. These frequencies contribute the same quantity at every invertible residue, and the discrepancy removes them. No absolute value is taken before that cancellation.

D.8. Noncoprime frequencies

For the remaining frequencies, let b_i be the largest divisor of h_i whose prime factors all divide q, and set $\ell_i = h_i/b_i$, so $(\ell_i, q) = 1$. The integer $b = \mathrm{rad}(b_1 b_2 b_3)$ is squarefree. For squarefree q it equals $(h_1 h_2 h_3, q)$, and $q = bd$ with $(b, d) = 1$.

If q is y-densely divisible and $m \mid q$, then q/m is my-densely divisible. A divisor of q in a window of ratio y meets q/m in a divisor in a window of ratio my. Taking $m = b$ and $y = X^{\delta}$ gives $d \in \mathcal{D}_I(bX^{\delta})$.

The unit group modulo q is the product of the unit groups modulo b and modulo d, and

$$\frac{x}{q} \equiv \frac{x\overline{d}}{b} + \frac{x\overline{b}}{d} \pmod 1.$$

The additive character and the product condition both split, and $1/q = (1/b)(1/d)$. On the modulus d, each h_i is invertible. The change of variables $u_i \equiv \overline{b}\, h_i n_i \pmod d$ converts the second factor into $\mathrm{Kl}_3(a h_1 h_2 h_3 \overline{m b^3}; d)$, in the normalization (11.7).

Since b is squarefree, the factor modulo b is a product of prime sums. At a prime $p \mid b$ at least one coordinate vanishes. The sum over the units collapses to a Ramanujan sum, whose absolute value is at most $(c_1, p)(c_2, p)(c_3, p)/p^2$. The product of these bounds is

$$\left|F(\overline{d}\,\mathbf{h}, a\overline{m}; b)\right| \le \frac{\mathrm{rad}(b_1)\,\mathrm{rad}(b_2)\,\mathrm{rad}(b_3)}{b^2}.$$

The value depends only on which coordinates vanish, hence only on the b_i. It factors out of the sums over d and m.

A product $\ell = \ell_1\ell_2\ell_3$ arises from at most $4\tau_3(|\ell|)$ ordered signed triples: $\tau_3(|\ell|)$ counts the positive ordered triples, and four sign patterns produce a fixed sign of the product. Extending each range $|\ell_i| \le X^{\varepsilon_{\mathrm{III}}/2} H_i/b_i$ to a range for the product only enlarges the sum. Inserting the bound leaves a constant multiple of

$$\frac{Q}{H} \sum_{\mathbf{b}} \frac{\mathrm{rad}(b_1)\,\mathrm{rad}(b_2)\,\mathrm{rad}(b_3)}{b^2}\, T(\mathbf{b}).$$

This is (7.19). The size of $T(\mathbf{b})$ is not estimated in this reduction.

D.9. The auxiliary factorization

Let $y = bX^{\delta}$ and $1 \le S \le yQ/(2b)$. Every modulus d in $T(\mathbf{b})$ lies in an interval of relative length $X^{-\varepsilon_{\mathrm{III}}}$ about Q/b and is y-densely divisible. The constraint on S implies $S \le dy$ for large X. Density supplies $s \mid d$ with $S/y \le s \le S$. Set $r = d/s$. Squarefreeness gives $(r, s) = 1$, and $d \asymp Q/b$ rearranges to $Q/(bS) \ll r \ll yQ/(bS)$. If $d < S$, the factorization $s = d$, $r = 1$ is the one used in the source. The condition $(b\ell, d) = 1$ splits into $(b\ell, r) = (b\ell, s) = 1$.

The subsequent estimates depend on S through the diagonal term, which grows with S, and through the third off-diagonal term, which decreases with S. The choice that balances those two terms and obeys the upper constraint is

$$S = \min\left(\left(\frac{Q}{b}\right)^{4/3} \frac{y^{1/3} M^{2/3}}{H_{\mathbf{b}}^{2/3}},\ \frac{yQ}{2b} \right).$$

At the first argument of the minimum, the balanced monomial is $y^{1/3} H_{\mathbf{b}}^{1/3} M^{5/3} Q^{7/3}/b^{7/3}$.

D.10. Corollary 6.26

Cauchy–Schwarz in the pairs (s, ℓ), with weights $s^{-1/2}\tau_3(|\ell|)$ and $s^{1/2}$ times the inner sum, gives $|T(\mathbf{b})|^2 \le T_1 T_2$. The product of the weights is 1. The first factor satisfies $T_1 \ll X^{o(1)} H_{\mathbf{b}}$. A new smooth cutoff, equal to 1 on the frequency range and carrying no Mellin frequency, only enlarges T_2. Its derivatives of fixed order are $O(1)$.

The expanded correlation is the sum in Corollary 6.26 of [10]. The hypotheses, and the checks that put them in force, are the following.

Hypothesis	Where it is checked	What is required
r_1, r_2, s squarefree	each $r_i s$ divides a squarefree modulus	squarefree support
$(r_1 r_2, s) = 1$	complementary factors of d	r_1 and r_2 need not be coprime
modulus $[r_1, r_2]$	the corollary, not the product $r_1 r_2$	shared primes are allowed
a_i a unit modulo $r_i s$	$(m_i, b r_i s) = 1$ and $(a b_1 b_2 b_3, r_1 r_2 s) = 1$	invertibility
smooth cutoff on ℓ	the new majorant of the frequency range	derivatives $O(1)$
$H_{\mathbf{b}} \ll (s[r_1, r_2])^{O(1)}$	the frequency length and modulus are polynomially related on each dyadic box	the corollary's actual length hypothesis
both gcd factors kept	$(r_1, r_2, m_2 - m_1)$ and $(m_1 r_1^3 - m_2 r_2^3, s)$	neither is omitted

The input of the corollary is Proposition 6.11, including Deligne's theorem on the non-diagonal prime sums. That bound is cited. If $s[r_1, r_2] = 1$, then $\mathrm{Kl}_3(\cdot\,; 1) = 1$ and the required estimate is trivial; otherwise the displayed polynomial relation verifies the length condition. Squarefreeness of r_1 and r_2 does not force $(r_1, r_2) = 1$, because the two factors are not divisors of one common integer. The shared-prime factor is meaningful for that reason. The example $(r_1, r_2, s) = (6, 15, 31)$ has $(r_1, r_2) = 3$, $(r_1 r_2, s) = 1$, and $[r_1, r_2] = 30$, which is strictly smaller than the product 90.

D.11. Diagonal and off-diagonal terms

The diagonal of the correlation bound is the vanishing of

$$a_2 r_1^3 - a_1 r_2^3 = \frac{a b_1 b_2 b_3}{b^3 m_1 m_2}(m_1 r_1^3 - m_2 r_2^3).$$

The prefactor is nonzero, so the condition is $m_1 r_1^3 = m_2 r_2^3$. It is not the condition $r_1 = r_2$. The pair $r_1 = 2$, $m_1 = 1$, $r_2 = 1$, $m_2 = 8$ satisfies it with unequal factors. On this set, Remark 6.10 gives $|\mathrm{Kl}_3| \ll X^{o(1)}$, and the sum over ℓ contributes $H_{\mathbf{b}}$. For each fixed (r_1, m_1) the equation has $X^{o(1)}$ solutions. The values of s attached to a fixed r_1 satisfy $\sum s \ll (Q/(b r_1))^2$. Therefore

$$T_2' \ll X^{o(1)} H_{\mathbf{b}} M \sum_{r_1 \gg Q/(bS)} \left(\frac{Q}{b r_1}\right)^2,$$

and the sum over r_1 is $\ll QS/b$. This is (7.22).

Off the diagonal, Corollary 6.26 contributes the two square roots

$$(r_1, r_2, m_2 - m_1)^{1/2} (m_1 r_1^3 - m_2 r_2^3, s)^{1/2}.$$

Neither factor is omitted. The inequality $(n, s)^{1/2} \le \sum_{d \mid (n,s)} d^{1/2}$ gives, on an interval of length K,

$$\sum_{s \le K} (n, s)^{1/2} \le K\, \tau(|n|),$$

and the same expansion gives $\sum_{m_1, m_2 \asymp M} (r_1, r_2, m_1 - m_2)^{1/2} \ll \tau((r_1, r_2))\, M^2$. Substituting $r_i = r t_i$ and summing the geometric series, with $r_1 \asymp r_2$ kept, produces the shape of Lemma 7.4 before any square root:

$$T_2'' \ll X^{o(1)} \frac{M^2 Q^2}{b^2} \left(\frac{H_{\mathbf{b}} b^{1/2}}{Q^{1/2}} \left(\frac{bS}{Q}\right)^{1/2} + \frac{Q^{1/2}}{b^{1/2}} \left(\frac{X^{\delta} Q}{S}\right)^{1/2} \right).$$

Adding the diagonal gives the bound for $T_2 = T_2' + T_2''$ before the square root. The square root is taken only after that sum.

Since $T_1 \ll X^{o(1)}H_{\mathbf{b}}$ and $H_{\mathbf{b}} \ll X^{3\varepsilon_{\mathrm{III}}/2}H/(b_1b_2b_3)$, the three contributions to $|T|$ carry the powers $H_{\mathbf{b}}^{2/3}$, $H_{\mathbf{b}}^{5/6}$, and $H_{\mathbf{b}}^{1/2}$. The losses are therefore $\varepsilon_{\mathrm{III}}$, $5\varepsilon_{\mathrm{III}}/4$, and $3\varepsilon_{\mathrm{III}}/4$. The published factor $X^{5\varepsilon_{\mathrm{III}}/4}$ is the largest of the three and is exact for the second term only. Using it for all three terms is a valid majorant and is not the loss of the first or the third term. The series of Lemma 7.5 converges: after $b_i = t_iu_i$ with t_i squarefree, the general term is $\ll \tau(b)^{O(1)}b^{-3/2}$. Summing it removes $\mathbf{b}$.

D.12. Terminal exponents

Dividing by MN and inserting $H = Q^3/N$, $M \asymp X/N$, $N \gg X^{3/4+3\sigma/2}$, and $Q \ll X^{1/2+2\omega}$ produces three ratios. At $\varepsilon_{\mathrm{III}} = 0$ their savings are 63/2500, 2429/40000, and 291/4000. The per-interval requirement (7.16) asks each ratio to be $\ll X^{-2\varepsilon_{\mathrm{III}}}$. The costs of the square root and of this target are

$$3\varepsilon_{\mathrm{III}} = \frac{3}{1000}, \qquad \tfrac{13}{4}\varepsilon_{\mathrm{III}} = \frac{13}{4000}, \qquad \tfrac{11}{4}\varepsilon_{\mathrm{III}} = \frac{11}{4000}.$$

Subtracting them from the raw margins leaves 111/5000, 2299/40000, and 7/100. All three differences are positive, and the first is the smallest. The allowance 13/4000 is the second cost only. Subtracting it from the first or the third raw margin overstates that loss. The number of intervals is not a fourth subtraction inside each margin.

D.13. Assembly

On one modulus interval, the correlation estimate supplies a main contribution $X^{1-2\varepsilon_{\mathrm{III}}+o(1)} = X^{0.998+o(1)}$. Lemma 11.4 multiplies by the number of intervals once:

$$X^{\varepsilon_{\mathrm{III}}+o(1)}\,X^{1-2\varepsilon_{\mathrm{III}}+o(1)} = X^{1-\varepsilon_{\mathrm{III}}+o(1)} = X^{0.999+o(1)}.$$

The Fourier tail has per-interval exponent $-1/2000$. Summing the intervals once raises it to $1/2000 = 0.0005$. The Taylor remainder at $J = 3230$ has per-interval exponent $-227/50000 = -0.00454$. Summing the intervals once raises it to $-177/50000 = -0.00354$. The small-conductor term has exponent $79/800 = 0.09875$ and is not multiplied by the interval count. Each of $1/2000$, $-177/50000$, and $79/800$ is strictly smaller than $999/1000$.

Contribution	Per interval	After one interval count
Main estimate	$1 - 2\varepsilon_{\mathrm{III}} = 0.998$	0.999
Fourier tail	$-1/2000 = -0.0005$	$1/2000 = 0.0005$
Taylor remainder	$-227/50000 = -0.00454$	$-177/50000 = -0.00354$
Small conductors	—	$79/800 = 0.09875$

Therefore

$$\boxed{E_{\mathrm{III}} \ll X^{999/1000+o(1)} = X^{0.999+o(1)}.}$$

Heath–Brown's identity at depth five produces $O((\log X)^{10})$ dyadic tuples, and there are five choices of the depth index. The eight bands are disjoint, so a modulus is counted once; the crude majorant that estimates each band by the full sum contributes the absolute factor 8. None of these factors is a fixed positive power of X. The bounds are uniform for $|t| \le C(\log X)^2$, because $|n^{it}| = 1$ and N_ν^{it} does not change the modulus of a Taylor coefficient, and they are uniform in the shift, because the shift enters only as the residue $-j \bmod d$ under the condition $(d, j) = 1$. The resulting exponent 999/1000 is strictly smaller than the Type I/II exponent 199989/200000. The combination with the remaining blocks is Appendix E.

E. Global Integration

This appendix is the assembly of the variance inequality. The dispersion propositions are Appendices C and D. The classical collar is removed by the Gevrey tail after $X \to \infty$, and the coefficient that remains is the full frozen value a_*. The weight-translation error is included. The cited exponential-sum theorems and the complex-difference explicit-formula identities are applied with their hypotheses matched.

E.1. Classification

On $n \le BX$, with $\mu_{\le}(n) = \mu(n)\mathbf{1}_{n\le(BX)^{1/5}}$,

$$\Lambda(n) = \sum_{r=1}^{5}(-1)^{r-1}\binom{5}{r}\big(\mu_{\le}^{*r} * \mathbf{1}^{*(r-1)} * \log\big)(n).$$

Five integers strictly larger than $(BX)^{1/5}$ have product larger than BX, so the tail $\mu_{>}^{*5}$ vanishes on that range. The binomial expansion of $\mu = \mu_{\le} + \mu_{>}$ and the relation $\mu * \mathbf{1} = \delta_1$ produce the displayed sum. Each summand has $2r \le 10$ factors. The Möbius factors are supported on $n \le (BX)^{1/5}$. The copies of $\mathbf{1}$ and the logarithm are the smooth factors.

A smooth dyadic cutoff on $[N/2, 2N]$, inserted in each of $m \le 10$ factors, supports the convolution on $[2^{-m}P, 2^{m}P]$. Against a weight supported on $[AX, BX]$, a retained tuple satisfies

$$A\,2^{-m}X \le P \le B\,2^{m}X.$$

Let L be large enough that 2^L dominates both $B\,2^m$ and $2^m/A$. For $[X, 2X]$ one has $L = 11$, hence $2^{-10}X \le P \le 2^{11}X$. For $[X/4, 8X]$ one has $L = 13$. Dividing the exponents by their sum produces a tuple summing to 1 and moves each exponent by at most $L\log 2/\log X$. The classification is applied to the renormalized tuple. The original scales differ from it by the fixed factor 2^L.

Polymath's Lemma 3.1 splits a nonnegative tuple summing to 1 into three mutually exclusive classes. A set of indices is large when its sum is at least $1/2 + \sigma$ and small when its sum is at most $1/2 - \sigma$.

If some coordinate is at least $1/2 + \sigma$, the tuple is Type 0. That factor exceeds the Möbius ceiling: a truncated Möbius factor has exponent at most $1/5 + O(1/\log X)$, and $2\sigma - 1/5 = 1/500$. For $L = 11$ the gap exceeds the distortion once $X > 2^{5500}$. The factor is then smooth, and the complementary convolution is a coefficient sequence of exponent at most $1/2 - \sigma$.

If no coordinate is that large and some subsum lies in $(1/2 - \sigma, 1/2 + \sigma)$, the complement lies in the same interval. The shorter exponent n satisfies $n_0 < n \le 1/2$. It is sent to the frozen Type I theorem when $n < n_{\text{split}}$, to the frozen Type II theorem when $n > n_{\text{split}}$, and to both when $n = n_{\text{split}}$. The tuple is estimated once. Both sides are convolutions of coefficient sequences. A convolution of two sequences supported in fixed multiples of their scales is a coefficient sequence at the product of the scales, up to the fixed factor 4. No derivative of a Möbius factor is taken.

If neither of those occurs, every set is large or small. The counting argument of Section 9 produces exactly three powerful indices, each at least 2σ and at most $1/2 - \sigma$, with each pair summing to at least $1/2 + \sigma$. Four such indices would give $1 \ge 2(1/2 + \sigma) > 1$, and the gap in that comparison is $5\sigma - 1/2 = 1/200$. Each of the three exceeds the Möbius ceiling, so each is smooth. The remaining convolution, including every truncated Möbius factor, is the coefficient sequence of the frozen Type III theorem. The fixed factor $2^{O(1)}$ is the comparability already allowed there.

Multiplication by n^{it} with $|t| \le C(\log X)^2$ has modulus one, so a coefficient bound is unchanged. On a smooth factor the derivative bounds are those of Lemma 11.2. The first Type I brace has derivative $-5/6$ in n and lies $1/25000$ below its ceiling at n_0. A displacement of size $\eta = L\log 2/\log X$ stays inside that margin when $\eta < 6/125000$. For $L = 11$ this is $X > 2^{229167}$. At

$n = 1/2$ the third Type II brace has no nominal margin, and a fixed factor 2^L is the factor $X^{o(1)}$ already present in that theorem.

E.2. Type 0

Let $g = \alpha * \psi$, with ψ smooth at scale $N \gg X^{1/2+\sigma}$ and α a coefficient sequence at scale $M \ll X^{1/2-\sigma}$. On a coprime progression the convolution opens as a sum over units ℓ mod d of a coefficient sum times $\Delta(\psi; a\overline{\ell} \bmod d)$. The mask $(n, d) = 1$ remains. Since $|\alpha|$ sums to $\ll X^{o(1)}M$,

$$\sum_{d \le D} |\lambda_d| \, |\Delta(g)| \ll X^{o(1)} M \sum_{d \le D} \sup |\Delta(\psi; b \bmod d)|.$$

Poisson summation gives

$$\sum_{n \equiv b \pmod d} \psi(n) = \frac{N}{d} \sum_{h \in \mathbb{Z}} e_d(hb) \, \widehat{\varphi}\Big(\frac{hN}{d}\Big).$$

The frequency $h = 0$ does not depend on b. The discrepancy subtracts the average over the units, so that term cancels. The remaining transform is $\ll (\log X)^{O(1)}(1 + |\xi|)^{-2}$, and on $d \le D \ll N$ one has $|\Delta(\psi)| \ll (\log X)^{O(1)} d/N$. The sum $\sum_{d \le D} d \le D^2$ then produces

$$\sum_{d \le D} |\lambda_d| \, |\Delta(g)| \ll X^{o(1)} \frac{MD^2}{N} = X \Big(\frac{D}{N}\Big)^2 X^{o(1)}.$$

The worst scales $D = X^{1/2+b_{\max}}$ and $N = X^{1/2+\sigma}$ give

$$1 - 2(\sigma - b_{\max}) = \frac{20149}{25000} = 0.80596.$$

A bound that kept only one factor D/N would have the larger exponent $1 - (\sigma - b_{\max})$. The second factor is the sum of the moduli.

The difference $\Delta_K - \Delta$ retains one smooth factor, so Pólya–Vinogradov contributes one square root $\sqrt{f}$, not three. The sum over $f \le K$ is $\ll K^{3/2}$. The resulting exponent is

$$\tfrac{1}{2} - \sigma + \tfrac{3}{2}k = \frac{7983}{20000} = 0.39915,$$

which is strictly smaller than $20149/25000$.

E.3. Omitted cores

The short-factor definition in band i is $cu \le X^{1/2+b_i}$. The parent ceiling is $q \le X^{1/2+b_i}$. It already contains the short factor. It is not $X^{1/2+b_i+b_{\max}}$. If $d \mid q$ omits a prime $p \mid c$, then $p > X^{a_i}$ and

$$d \le q/p < X^{1/2+b_i-a_i}.$$

The differences $a_i - b_i$ equal $1/250$ for the first seven bands and $201/50000$ for the eighth. The largest omitted-core exponent is $1/2 - 1/250 = 0.496$. Adding $b_{\max}$ again would make the conductor term $F\sqrt{X}$ larger than $1 - \varepsilon_{\mathrm{d}}/2$, and the sieve would not close. The coefficients on the subset still satisfy $|\lambda_d| \ll (\log X)^2$.

The condition $(d, j) = 1$ makes $-j$ a unit. Characters of conductor $f \mid d$ with $f > K$ are induced by primitive characters, and the cofactor mask $\mathbf{1}_{(n,z)=1}$ is kept. A Type I or Type II block is already a bipartition into two coefficient sequences of length at most $X^{1/2+\sigma} < X^{1-2\sigma}$. For a Type III block, two smooth factors go on one side and the third goes with α on the other.

Both sides are then at most $X^{1-2\sigma}$. A grouping that puts all three smooth factors on one side can exceed that length, and the sieve is not applied to it. A Type 0 block is not sent here: its omitted-core divisors are inside the Type 0 sum.

Cauchy–Schwarz and the multiplicative large sieve produce the three shapes $F\sqrt{X}$, $\sqrt{XL}$, and X/F. On $K < F \le X^{1/2-1/250}$ their exponents are 249/250, 899/1000, and 9999/10000. The largest is $1-k$, so

$$E_{\mathrm{omit}} \ll X^{0.9999+o(1)}.$$

The gap under $1-\varepsilon_{\mathrm{d}}/2$ is $k-\varepsilon_{\mathrm{d}}/2 = 9/200000$.

E.4. One shift, and the sum of the shifts

Let W be a fixed Gevrey-2 weight on a fixed interval $[A,B] \subset (0,\infty)$. Fourier inversion cuts the Mellin integral at $C_A(\log X)^2$. The integral of $|\widehat{\phi}|$ over that range is $O(1)$, with a constant depending on the fixed kernel. On the support of W, Heath–Brown's identity at depth five produces $O((\log X)^{10})$ tuples. A factor 2^L and the eight bands contribute $X^{o(1)}$. Each tuple is assigned by the classification above and estimated once.

The full-core contributions are the Type 0 exponent 20149/25000, the Type I/II exponent 199989/200000, and the Type III exponent 999/1000. The omitted-core contributions are $1-k$ for Type I, Type II, and Type III, and the Type 0 bound for a Type 0 block. The largest is 199989/200000. Therefore

$$E_j(W) \ll X^{0.999945+o(1)},$$

uniformly in the residue and in the fixed weight. The discarded Mellin tail, passed through $|\lambda_d| \ll (\log X)^2$ and $\sum_{d\le X^{1/2+b_{\max}}} 1/d \ll \log X$, is $O(X^{1-A+o(1)})$. At $A=10$ this is $O(X^{-9+o(1)})$, which is smaller than $X^{199989/200000}$.

There are at most $2X^{\theta_1}$ shifts in $0<|j|\le X^{\theta_1}$. Hence

$$\sum_{0<|j|\le X^{\theta_1}} E_j(W) \ll X^{1-\varepsilon_{\mathrm{d}}/2+\theta_1+o(1)}.$$

The identity $\varepsilon_{\mathrm{d}}/2-\theta_1 = 1/1000000$ gives

$$\boxed{\sum_{0<|j|\le X^{\theta_1}} E_j(W) \ll X^{999999/1000000+o(1)} = X^{0.999999+o(1)}.}$$

Divisors with $(d,j)>1$ are not included. Characters of conductor at most K are not restored in this sum.

E.5. Low conductors

First split

$$\Lambda = \Lambda_{\mathrm{pr}} + \Lambda_{\mathrm{pp}}, \qquad \Lambda_{\mathrm{pp}}(n) = \Lambda(n)\mathbf{1}_{n=p^\nu,\ \nu\ge 2}.$$

On the fixed X-scale support, $\sum_n |\Lambda_{\mathrm{pp}}(n)| \ll X^{1/2}\log^2 X$, so its aggregate over $0<|j|\le J$ is $O(JX^{1/2+o(1)})$. This separation is made before an imprimitive character is replaced by its primitive inducing character.

For $(a,d)=1$,

$$\Delta(g;a \bmod d) - \Delta_K(g;a \bmod d) = \frac{1}{\varphi(d)} \sum_{\substack{\chi \bmod d \\ 1<\mathrm{cond}(\chi)\le K}} \overline{\chi}(a) \sum_n g(n)\,\chi(n).$$

The conductor of the principal character is 1, so it is excluded from this low-conductor difference (but its nonzero-shift progression main term is estimated separately below). Since d is squarefree, write $d = fz$ with $(f, z) = 1$. The induced character is $\chi^*(n)\mathbf{1}_{(n,z)=1}$. For the prime part, $n = p \asymp X$, whereas $z \le d \le X^{1/2+b_{\max}}$; hence $(p, z) = 1$ for all sufficiently large X. The mask can therefore be dropped only after the proper prime powers have been separated.

For squarefree d, the coefficient $\lambda_d/\varphi(d)$ is a sum over cofactors m with dm in the adaptive family. Substituting $d = fz$ and summing the unitary divisors of a squarefree cofactor m' produces

$$\boxed{\sum_{\substack{z \mid m' \\ (z,j)=1}} \frac{\mu(z)\, z}{\varphi(z)^2\, \varphi(m'/z)} = \frac{c_{m'}(j)}{\varphi(m')^2}.} \tag{E.1}$$

Both sides are multiplicative. On a prime $p \nmid j$ the two divisors $z = 1$ and $z = p$ reproduce $c_p(j)/\varphi(p)^2$, using $\mu(p) = -1$. On a prime $p \mid j$ only $z = 1$ is admitted, and $c_p(j) = \varphi(p)$. The sum over the cofactor is an identity. It is taken before any absolute value. An absolute value inside the sum would replace $c_{m'}(j)$ by a larger majorant and would lose the cancellation on the primes that do not divide j.

The coefficient of each primitive character is then a factor $\mu(f)f/\varphi(f)^2$ times a sum of $c_{m'}(j)/\varphi(m')^2$. For squarefree q and $n \ne 0$, $|c_q(n)| = \varphi(\gcd(q, n))$, so

$$\sum_{0<|j|\le J} |c_q(j)| \ll J \sum_{g \mid q} \frac{\varphi(g)}{g} \ll J\, \tau(q).$$

The term $j = 0$ is excluded. Including it adds $\varphi(q)$. For $q = 30$ and $J = 1$, the sum over $0 < |j| \le 1$ equals 2, while the sum that includes $j = 0$ equals 10, and $\tau(30) = 8$.

The cofactors that occur satisfy $m' \gg X^{1/2+\eta_c}/f$. The tail

$$\sum_{m' \gg X^{1/2+\eta_c}/f} \frac{\tau(m')}{\varphi(m')^2} \ll f X^{-1/2-\eta_c+o(1)}$$

(note the negative 1/2) together with $f/\varphi(f) \ll X^{o(1)}$, and at most $\varphi(f)$ primitive characters gives

$$\sum_{0<|j|\le J} |L_j| \ll X^{o(1)}\, J\, K^2\, X^{-1/2-\eta_c} \sup_{\chi} \Big| \sum_n W(n/X) \Lambda_{\mathrm{pr}}(n) \chi(n) \Big|.$$

The supremum is $\ll X$, because $\sum_{n \asymp X} \Lambda_{\mathrm{pr}}(n) \le \sum_{n \asymp X} \Lambda(n) \ll X$. It is the length of the sum. With $J = X^{\theta_1}$ and $K = X^k$,

$$\sum_{0<|j|\le X^{\theta_1}} |L_j| \ll X^{1/2-\eta_c+2k+\theta_1+o(1)} = X^{125061/250000+o(1)} = X^{0.500244+o(1)}.$$

An exceptional real character is inside the same supremum. The bound does not require $L(s, \chi)$ to have no real zero near $s = 1$. A square-root bound for $\sum \Lambda(n)\chi(n)$ would be a theorem about the zeros of $L(s, \chi)$. It requires the generalized Riemann hypothesis for that L-function, not the Riemann hypothesis for ζ. That conditional bound is not used. Pólya–Vinogradov does not apply to Λ, which is not of bounded variation, and Siegel–Walfisz is not called: the cutoff $K = X^k$ is larger than every polylogarithmic conductor range. The proper prime powers were separated before the imprimitive mask was removed.

A kernel supported on $|j| \le h \le X^{\theta_1}$ and bounded by a constant multiple of h multiplies the aggregate by one further factor h. The exponent becomes $1/2 - \eta_c + 2k + 2\theta_1 = 250149/500000$. The principal term of the numerator has size $hX(A_0 + A_1)$, of exponent $1 + \theta_1$ up to a logarithm. The difference of the exponents is $249878/500000 > 0$, so the contribution is $o(hX)$.

E.6. Principal-character and prime-power aggregates

For $j \neq 0$ the ordinary progression sum is its principal-character main term plus its discrepancy; the principal term is not removed by the discrepancy. After the prime-power separation above, its contribution from the adaptive family is

$$P_j = \left(\sum_p W(p/X)\log p\right) \sum_{q \in \mathcal{R}_1^{\mathrm{tot}}} \frac{c_q(j)}{\varphi(q)^2}.$$

Using $q > X^{1/2+\eta_c}$, the squarefree divisor identity and $\sum_{0<|j|\le J} |c_q(j)| \ll J\tau(q)$ gives

$$\sum_{0<|j|\le J} |P_j| \ll JX^{1/2-\eta_c+o(1)} = o(X) \qquad (J \le X^{\theta_1+o(1)}).$$

The ordinary proper-prime-power aggregate, separated before the imprimitive mask was dropped, contributes

$$O(JX^{1/2+o(1)}) = o(X).$$

Accordingly, the high-conductor discrepancy is applied to $\Lambda_{\mathrm{pr}} = \Lambda - \Lambda_{\mathrm{pp}}$ by linearity: Theorem 12.1 bounds the Λ term. In the subtracted $\Delta_K(\Lambda_{\mathrm{pp}})$ term, the character part with conductor at most K introduces one further factor K, so its shift aggregate is

$$O(JKX^{1/2+o(1)}) = O(X^{1/2+k+\theta_1+o(1)}) = O(X^{0.500154+o(1)}) = o(X).$$

In the final assembly the ordinary prime-power aggregate and this conductor-subtracted term are kept distinct; the same prime powers are not counted twice. After multiplication by a variance kernel of size $O(h)$, with $J \ll h(\log X)^2$, the latter is $O(X^{1/2+k+2\theta_1+o(1)}) = o(hX)$.

E.7. Noncoprime shifts

Let $0 < |j| \le X^{\theta_1}$, let $AX \le n \le BX$, and suppose $\Lambda(n) \neq 0$, $d \mid (n+j)$, and $(d,j) > 1$. Then $n = p^\nu$. A prime r dividing both d and j divides $n+j$, hence divides n, so $r = p$. Since d is squarefree, $\gcd(d,j) = p$.

If $\nu = 1$, then $p \ge AX$ and $p \le |j| \le X^{\theta_1}$. For $X > (1/A)^{1/(1-\theta_1)}$ one has $AX > X^{\theta_1}$. No such prime exists. If $\nu \ge 2$, each prime $p \mid j$ contributes at most one power in $[AX, BX]$, and there are $O(\log|j|)$ such primes. The divisor sum at that power is $\ll (\log X)^2\tau(n+j) \ll X^{o(1)}$. Thus $N_j \ll X^{o(1)}$, and

$$\sum_{0<|j|\le X^{\theta_1}} |N_j| \ll X^{\theta_1+o(1)}.$$

A kernel bounded by a constant multiple of h produces an error $h^2X^{o(1)}$. The ratio to the principal term hX is $\ll h/X$. The inequality $h \le X^{\theta_1}$ is inserted only after that ratio is written, and the ratio is $O(X^{\theta_1-1+o(1)})$.

E.8. The six correlations

The six correlations and the resulting denominator bound are Proposition 13.5. The proof is Appendix F. It does not estimate a shifted lattice by one remainder of size $O(1)$ for each divisor pair. Theorem 6.4 is used only for $C_{11}(0)$. The classical zero-shift square is the periodic mass A_0 plus the Gevrey tail recorded in that proof. The mixed zero shift is Proposition 13.2. The nonzero shifts are Propositions 13.1 and 13.3.

E.9. Translation of the weight

For real h, expand the continuous auxiliary directly. With $m = n + j$ and $t = n - hv$,

$$\int W(t/X)|D_{h,\xi}(t)|^2\,dt = h\sum_j \int_{\mathbb{R}} \omega_\xi(v)\omega_\xi(v + j/h)\sum_n W((n - hv)/X)B(n)B(n + j)\,dv.$$

The supports restrict $|j| \le h$ and v to a fixed compact set. The mean-value theorem gives $W((n - hv)/X) - W(n/X) \ll_W h/X$. There are $O(h)$ shifts, the outside factor is h, and $|B(n)B(n + j)| \le X^{o(1)}$ on a support of length $O(X)$. Thus the replacement costs

$$E_{\mathrm{tr}}(W) \ll_W h^3 X^{o(1)}.$$

Likewise, writing $m = p + j$ in the exact numerator identity gives

$$\int W(t/X)\{\vartheta(t + h) - \vartheta(t)\}D_{h,\xi}(t)\,dt = h\int_0^1 \sum_{p,j}(\log p)B(p + j)W((p - hv)/X)\omega_\xi(v + j/h)\,dv.$$

Replacing the shifted weight by $W(p/X)$ has the same $O_W(h^3X^{o(1)})$ cost. Consequently

$$\frac{E_{\mathrm{tr}}(W)}{hX(A_0 + A_1)} \ll_W h^2X^{-1+o(1)} \le X^{2\theta_1-1+o(1)}$$

uniformly for $1 \le h \le X^{\theta_1}$. The exponent $2\theta_1 - 1$ is negative, so the error is absorbed on the whole admissible interval. For the sharp indicator of $[X, 2X)$, the windows $[X, 2X)$ and $[X - r, 2X - r)$ differ by $|r|$ integers at each endpoint. A smooth minorant, chosen before $X \to \infty$, replaces that layer by the Lipschitz estimate.

E.10. Centering

The centering contribution in the continuous variance numerator is

$$-h\int W(t/X)D_{h,\xi}(t)\,dt.$$

Every denominator is greater than 1, so every Ramanujan mode in B has mean zero. Poisson summation with the compactly supported Gevrey weight makes the displayed term $O_A(X^{-A})$ for each fixed A, uniformly for $h \le X^{\theta_1}$, after choosing the decay order. It is not a second copy of A_0 or A_1. For the Λ term at nonzero shifts, the classical piece is Proposition 13.4, with kernel $\widetilde{\kappa}_{h,\xi}$ and orientation $\Lambda_{\mathrm{pr}}(p)B_0(p + j)$. The adaptive principal-character progression main term remains; it is the aggregate estimated in the preceding subsection, not part of the discrepancy. Proper prime powers inside Λ were likewise separated there. The ordinary aggregate contributes $O(JX^{1/2+o(1)})$, and the conductor-subtracted term contributes $O(JKX^{1/2+o(1)})$, over the nonzero shifts.

E.11. The variational inequality

For real $h \ge 1$, put

$$\Delta_h\psi(t) = \psi(t + h) - \psi(t) - h, \qquad D_{h,\xi}(t) = \sum_m B(m)\omega_\xi((m - t)/h),$$

with ω_ξ as in Section 13. If W is a smooth minorant of $\mathbf{1}_{[1,2]}$, continuous Cauchy–Schwarz gives

$$J(X, h) \ge \int_{\mathbb{R}} W(t/X)|\Delta_h\psi(t)|^2\,dt \ge \frac{\left|\int_{\mathbb{R}} W(t/X)\Delta_h\psi(t)D_{h,\xi}(t)\,dt\right|^2}{\int_{\mathbb{R}} W(t/X)|D_{h,\xi}(t)|^2\,dt}.$$

The numerator at zero shift is $h(1+O(\xi))X\widehat{W}(0)(A_0+A_1)$, up to the errors already estimated. The classical piece carries the collar loss $-12\eta\log X$ from (3.11), and the adaptive piece satisfies $A_1 \geq (a_* - o(1))\log X$. The classical nonzero shifts are Proposition 13.4. The adaptive nonzero shifts, weighted by a smooth kernel of size at most $O_\xi(h)$, are the principal-character term, proper prime powers, the high- and low-conductor sums, and the noncoprime count. At the enlarged endpoint the high-conductor ratio retains the shift count,

$$X^{\theta_1-\varepsilon_{\mathrm{d}}/2+o(1)} = X^{-10^{-6}+o(1)}.$$

The translation error is absorbed uniformly on $1 \leq h \leq X^{\theta_1}$, as above. The denominator is Proposition 13.5.

One factor of A_0+A_1 survives the square and the division. The collar η and the smooth-window collar ξ are fixed during $X\to\infty$; then $\eta\to 0$ and $\xi\to 0$. They may be arbitrarily small, because the Gevrey tail does not impose a positive lower bound proportional to θ_1. Uniformly for real $1 \leq h \leq X^{\theta_1}$,

$$J(X,h) \geq Xh\left\{\tfrac{1}{2}\log\frac{X}{h^3} + a_*\log X - o(\log X)\right\}.$$

The $o(\log X)$ is ineffective only through the Siegel–Walfisz theorem on the base band B_0. The coefficient a_* and the Gram bound are effective. The smoothing minorant is chosen before $X\to\infty$, and $\widehat{W}(0)$ may be taken arbitrarily close to 1 without changing the coefficient of either logarithm.

E.12. The classical collar

The cutoffs are those of (3.11), with $H = h$: $U = hX^{2\eta}$ and $R = \sqrt{X/h}\,X^{-10\eta}$. A bounded-variation count of one remainder per divisor pair would require $20\eta > \theta_1$ and would retain a loss $12\eta\log X$. Keeping that requirement while sending η to zero would leave a loss at least $\frac{3}{5}\theta_1\log X$, a positive fraction of $a_*\log X$. That count is not used. The estimate used here is the Gevrey tail in the proof of Proposition 13.5, with separation $X/M \geq X^{20\eta}$. Every fixed $\eta>0$ is admissible during $X\to\infty$. The second limit sends η to zero, and the coefficient that remains is the frozen a_*. The stronger ratio $X^{-\varepsilon_{\mathrm{d}}/2}$ for the enlarged high-conductor error is not claimed.

E.13. The ledger

The comparisons that enter the ratio to the principal term $hX(A_0+A_1)$ are the following. The classical row is the Gevrey discrepancy, not the bounded-variation majorant.

Contribution	Size
High-conductor sum of the shifts	$X^{\theta_1-\varepsilon_{\mathrm{d}}/2+o(1)} = X^{-10^{-6}+o(1)}$, relative
Low-conductor aggregate	$X^{0.500244+o(1)}$, absolute
Low-conductor aggregate, after the kernel h	$X^{250149/500000+o(1)}$, absolute
Adaptive principal characters	$JX^{1/2-\eta_c+o(1)}$, absolute
Adaptive proper prime powers	$JX^{1/2+o(1)}$, absolute
Noncoprime aggregate	$X^{\theta_1+o(1)}$, absolute
Classical numerator covariance	$o(hX\log X)$, Aquila §2
Weight translation	$h^2X^{-1+o(1)}$, relative
Gevrey tail at $A=10$	$X^{-9+o(1)}$, absolute
Walker shifted discrepancy, shifts summed	$X^{-35633/9250000+o(1)}$, relative
Periodic nonzero-shift main terms	$X^{-2\eta+o(1)}$, relative, collar fixed
Mixed zero shift	$\exp(-c\sqrt{\log X})/\log X$, relative
Adaptive Gram error	$O(a_{\min}^{-2}X^{-a_{\min}})$, relative

An absolute error is a bound for a sum. A relative error is that sum divided by the principal term. The high-conductor row is the exponent 999999/1000000 of the summed shifts, written as a saving against X. The two low-conductor rows are different: the first is $\sum |L_j|$, and the second is the same sum multiplied by the kernel bound h.

F. Auxiliary Norms

This appendix proves the six propositions stated in Sections 13 and 14. The main text invokes them and does not repeat these arguments. No frozen parameter is changed. Theorem 6.4 is used only at shift zero, for the adaptive square.

F.1. Shifted correlations

Proof of Proposition 13.1. The statement matched here is Proposition 2.1 of Walker [14] (version 2 of that preprint). For $1 \le |k| \le X$ and $2 \le z \le X$, if $\rho(d) = 0$ for $d > z$ and $|\rho(d)| \le B$, and if $\lambda(n) = \sum_{d|n} \rho(d)$, that proposition states

$$\sum_{X_1 < n \le X_2} \lambda(n)\lambda(n+k) = (X - |k|) \sum_{\substack{d_1, d_2 \le z \\ (d_1, d_2) | k}} \frac{\rho(d_1)}{d_1} \frac{\rho(d_2)}{d_2} (d_1, d_2) + O_\varepsilon\big(B^2 \min(z^2, X^{47/74+\varepsilon} z^{53/74})\big),$$

with $X_1 = \max(0, -k)$ and $X_2 = \min(X, X - k)$. The shift $k = 0$ is outside the proposition.

For $B = B_0 + B_1$ the coefficients vanish off squarefree moduli and off every modulus larger than $z = X^{1/2+b_{\max}}$. The classical denominators are at most $R \le X^{1/2-\beta}$ for any fixed $\beta > 0$, hence inside the same z. Thus B is one divisor function, and the numbered statement applies to $B(n)B(n+j)$. A product of two different coefficient functions would be an extension of the opening bilinear form; that extension is not a numbered statement of the paper, and it is not required for this single function. Both orientations are the cases j and $-j$.

The pointwise bound is $|\lambda_d| \ll (\log X)^2$, so B in the proposition may be taken to be $(\log X)^2$, and $B^2 = X^{o(1)}$. The level satisfies

$$\frac{\log z}{\log X} = \frac{1}{2} + b_{\max} = \frac{25199}{50000} < \frac{27}{53},$$

so the optimized error is the one in force. It is smaller than $z^2 = X^{1+2b_{\max}}$:

$$\frac{47}{74} + \frac{53}{74}\Big(\frac{1}{2} + b_{\max}\Big) = \frac{3685547}{3700000} < 1 + 2b_{\max}.$$

One shifted correlation, compared with length X, therefore has error $O(X^{1-14453/3700000+o(1)})$. The range $0 < |j| \le CH(\log X)^2$ with $H \le X^{\theta_1}$ lies inside $1 \le |j| \le X$. Partial summation against a weight of total variation $V \ll 1$ multiplies the error by a constant. Summing the $O(X^{\theta_1+o(1)})$ shifts, and keeping θ_1, produces

$$X^{1-14453/3700000+\theta_1+o(1)} = X^{1-35633/9250000+o(1)}.$$

The main term of the proposition is X times the periodic mean, which is the term written in the statement. □

F.2. Mixed zero shift

Proof of Proposition 13.2. The periodic mean $\mathbb{E}_{\mathrm{per}}[B_0 B_1]$ vanishes because every classical denominator is at most $R < X^{1/2}$ and every adaptive denominator exceeds $X^{1/2+\eta_c}$. The eight bands are disjoint, so a modulus carries the coefficient of at most one band, and $|\lambda_d|$ is not multiplied by 8.

Aquila's Lemma 3 is applied to each band separately, with classical level $Z = R \le X$ and band level $Y_i = X^{1/2+b_i} \le X^{1/2+b_{\max}} = X^{1-24801/50000}$. The coefficient bound $(\log X)^2$ and the vanishing mean are the same for every band. The lemma is not applied once to a single rectangle and then copied onto the union. The sum of the eight bounds is

$$\sum_n W(n/X)\, B_0(n) B_1(n) \ll X \exp\bigl(-c\sqrt{\log X}\bigr)$$

The contour step inside the lemma is cited, not reproved. The shift is zero, so Walker's proposition is not used. □

F.3. Periodic main terms

Proof of Proposition 13.3. Equation (3.11) sets $U = HX^{2\eta}$ and $R = \sqrt{X/H}\, X^{-10\eta}$. The factor 2 belongs to the lower cutoff. The exponent 10η belongs to R and does not enter U. Every adaptive denominator exceeds $X^{1/2+\eta_c}$, hence exceeds U. The identification $\alpha_{\mathrm{col}} = \eta$ is the minimum denominator used below. The collar stays fixed.

For squarefree r and $n \neq 0$, $|c_r(n)| = \varphi(\gcd(r,n))$. Group the nonzero shifts by $g \mid r$. The number of j with $0 < |j| \le J$ and $\gcd(r,j) = g$ is at most $2J/g$. Since $\varphi(g) \le g$,

$$\sum_{0<|j|\le J} |c_r(j)| \le 2J \sum_{g\mid r} \frac{\varphi(g)}{g} \le 2J\, \tau(r).$$

The value $c_r(0) = \varphi(r)$ is not in this sum. Periodic orthogonality gives $\mathbb{E}_{\mathrm{per}}[B(n)B(n+j)] = \sum_r c_r(j)/\varphi(r)^2$, so

$$\sum_{0<|j|\le J} \bigl|\mathbb{E}_{\mathrm{per}}[B(n)B(n+j)]\bigr| \le 2J \sum_{r>U} \frac{\tau(r)}{\varphi(r)^2}.$$

The bounds $\varphi(r) \gg_\varepsilon r^{1-\varepsilon}$ and $\tau(r) \ll_\varepsilon r^{\varepsilon}$ give a series $\ll_\varepsilon U^{-1+\varepsilon}$. With $J \le CH(\log X)^2$ and $U = HX^{2\alpha_{\mathrm{col}}}$,

$$\frac{J}{U} \ll (\log X)^2 X^{-2\alpha_{\mathrm{col}}} = X^{-2\alpha_{\mathrm{col}}+o(1)}.$$

The factor U^{ε} is $X^{o(1)}$ once ε is chosen after the fixed collar. The same bound with H replaced by h is the form used for the variance denominator. □

F.4. The two norm bounds

Proof of Proposition 13.5. The zero-shift adaptive square is Theorem 6.4; with $a_{\min} = 9/2000$, it contributes

$$X\widehat{W}(0)\, A_1\bigl(1 + O(a_{\min}^{-2} X^{-a_{\min}})\bigr).$$

The zero-shift mixed term is Proposition 13.2.

The zero-shift classical square has periodic main term $X\widehat{W}(0)\, A_0$. A nonzero frequency m/M of a pair of classical denominators lies at least $1/M$ from every integer, so the argument of $\widehat{W}$ is at least X/M. Here $M \le R^2$ and $R = \sqrt{X/h}\, X^{-10\eta}$, hence $X/M \ge X^{20\eta}$. Since $W \in C_c^\infty(\mathbb{R})$, its additive Fourier transform is Schwartz; repeated integration by parts makes the summed tail $O_A(X^{-A})$ for every fixed A, after the $X^{o(1)}$ divisor weights are absorbed by enlarging A. This is not a count of one remainder of size $O(1)$ per divisor pair.

For $0 < |j| \le h$, Proposition 13.1 supplies the discrepancy and Proposition 13.3 supplies the summed periodic means, with $U = hX^{2\eta}$. Multiplication by a kernel bounded by h produces an error $O(hX^{1-2\eta+o(1)})$ from the periodic means and $O(hX^{1-35633/9250000+o(1)})$ from the summed Walker discrepancy. Both are $o(hX\log X)$ while $\eta > 0$ stays fixed. The Gram factor $O(X^{-9/2000})$

and the mixed factor $\exp(-c\sqrt{\log X})/\log X$ are of the same kind. Therefore the triangular sum is at most $hX\widehat{W}(0)\,(A_0+A_1)\,(1+o(1))$.

The passage from this triangular sum to the short-interval square is the translation estimate of Appendix E. Its ratio to $hX(A_0+A_1)$ is $O(X^{2\theta_1-1+o(1)}) = O(X^{-249973/250000+o(1)})$. The collar is not sent to zero inside this estimate. □

Proof of Proposition 14.1. The factor T in $\|R_b\|_g^2$ is the change of variables in $\int g(t/T)\,dt$. The weight $K_{\mathrm{norm}}(\beta,u) = b(\beta)^2\widehat{g}(-u/(2\pi\beta))$ has, by Lemma B.2 at order $N=0$ and $m=1$, a first derivative in the logarithmic observation variable bounded independently of u. Partial summation is legitimate for $|u| \le C(\log X)^2$.

The diagonal $j=0$ contributes at most $TXg_0(A_0+A_1)\int b^2$, by Theorem 6.4 on B_1, the classical Gevrey diagonal of the previous proof, and Proposition 13.2. The nonzero shifts contribute $T\cdot O(X^{1-35633/9250000+o(1)})$ by Proposition 13.1, and the periodic means contribute $O(TX^{1-2\alpha_{\mathrm{col}}+o(1)})$ by Proposition 13.3. The linearization of the kernel contributes $O(TX\cdot X^{2\theta_1-1+o(1)})$. Lemma B.2 gives a truncation cost $O(TX\cdot X^{-A})$ at $|u| > C_A(\log X)^2$, for every fixed A.

Relative to $TX\log X$, with the collar fixed, these are

Contribution	Relative saving
Walker shifted discrepancy, shifts summed	$X^{-35633/9250000+o(1)}$
Periodic nonzero-shift main terms	$X^{-2\alpha_{\mathrm{col}}+o(1)}$
Kernel linearization	$X^{-249973/250000+o(1)}$
Adaptive diagonal Gram error	$O(X^{-9/2000})$
Mixed zero-shift correlation	$\exp(-c\sqrt{\log X})/\log X$
Gevrey truncation	X^{-A} for every fixed A

Each tends to zero. Hence $\|R_b\|_g^2 \le TX\,g_0(A_0+A_1)\int b^2\,(1+o(1))$. □

F.5. The Mellin bridge

Proof of Proposition 14.2. On the logarithmic coordinate $s=\log\beta$,

$$K(e^s,u) = \frac{1}{2\pi}\int_{\mathbb{R}} \widehat{K}_{\log}(t,u)\,e^{its}\,dt.$$

The integrand is absolutely integrable because

$$\bigl|\widehat{K}_{\log}(t,u)\bigr| \le C\exp\bigl(-c_1\sqrt{|u|}-c_2\sqrt{|t|}\bigr),$$

by Lemma B.2. Truncation at $|t| > C_A(\log X)^2$ costs $O(X^{-A})$. The square on the logarithm is the one matched to $\sqrt{|t|}$. The approximant b is supported on finitely many compact components, all bounded away from 0, 1, and infinity, so the integral splits into a finite sum over those components and the gap around $\beta=1$ contributes nothing. On each component the truncated integral is a finite measure in t. Theorem 12.1 is uniform for $|t| \le C(\log X)^2$ and for $0<|j|\le X^{1/100}$. Fubini applies to the truncated integral. The tail in t, multiplied by the trivial bound on the discrepancy and summed over the shifts, is $O(X^{-A+O(1)})$ and is made $O(X^{-10})$ by the choice of A. □

Data and code availability

The accompanying verification package contains the production C++ arithmetic certificate `certificates/typeIII_core_certificate.cpp`, the independent arbitrary-precision Python transcription `certificates/mass_certificate.py`, the full eight-band unbounded-integer replay log `certificates/replay_v1.1.txt`, and the exact rational inequality checker `certificates/`

`verify_inequalities.py`. The two SHA-256 digests recorded in Appendix A uniquely identify the frozen parameter block, the eight core integers, and the C++ certificate source.

From the package root, the exact rational inequality ledger is verified by running

```
python3 certificates/verify_inequalities.py,
```

and the reduced-cell cross-check $(h, A, B, a) = (1000, 10000, 14000, 2000)$, which returns the integer 460599464028034591918 in both implementations, is executed by running

```
python3 certificates/mass_certificate.py 1000 10000 14000 2000.
```

Acknowledgments

The two-family architecture, conductor subtraction, cofactor identity, smooth classical covariance, and BGSTB transfer past $\alpha = 1$ build on the framework of Aquila [1]; Lemma 3 of that preprint supplies the mixed zero-shift bound in Proposition 13.2. We also rely on Heath–Brown's identity [6], Montgomery's pair-correlation method [8], the multiplicative large sieve [7, Thm. 7.13], the Pólya–Vinogradov inequality [9, 12], the Siegel–Walfisz theorem [11, 13], Polymath's equidistribution estimates [10], Deligne's bound [3], Drappeau's conductor-subtracted dispersion method [4], Walker's shifted sieve-weight correlation bound [14], and the unconditional complex-difference explicit-formula identities of Baluyot, Goldston, Suriajaya, and Turnage-Butterbaugh [2].

References

[1] W. Aquila, *The Selberg integral at polynomial lengths and pair correlation beyond one*, preprint (2026), https://hexagonmath.org/pdf/2611.01021v1.

[2] S. Baluyot, D. A. Goldston, A. I. Suriajaya, and C. L. Turnage-Butterbaugh, *An unconditional Montgomery theorem for pair correlation of zeros of the Riemann zeta-function*, Acta Arith. **214** (2024), 357–376, doi:10.4064/aa230612-20-3, arXiv:2306.04799.

[3] P. Deligne, *La conjecture de Weil. II*, Inst. Hautes Études Sci. Publ. Math. **52** (1980), 137–252.

[4] S. Drappeau, *Sums of Kloosterman sums in arithmetic progressions, and the error term in the dispersion method*, Proc. Lond. Math. Soc. (3) **114** (2017), no. 4, 684–732, arXiv:1504.05549.

[5] D. A. Goldston, S. M. Gonek, A. E. Özlük, and C. Snyder, *On the pair correlation of zeros of the Riemann zeta-function*, Proc. Lond. Math. Soc. (3) **80** (2000), no. 1, 31–49.

[6] D. R. Heath-Brown, *Prime numbers in short intervals and a generalized Vaughan identity*, Canad. J. Math. **34** (1982), no. 6, 1365–1377.

[7] H. Iwaniec and E. Kowalski, *Analytic Number Theory*, Amer. Math. Soc. Colloquium Publications **53**, American Mathematical Society, Providence, RI, 2004.

[8] H. L. Montgomery, *The pair correlation of zeros of the zeta function*, in *Analytic Number Theory* (St. Louis, 1972), Proc. Sympos. Pure Math. **24**, American Mathematical Society, Providence, RI, 1973, 181–193.

[9] G. Pólya, *Über die Verteilung der quadratischen Reste und Nichtreste*, Nachr. Ges. Wiss. Göttingen, Math.-Phys. Kl. (1918), 21–29.

[10] D. H. J. Polymath, *New equidistribution estimates of Zhang type*, Algebra & Number Theory **8** (2014), no. 9, 2067–2199, arXiv:1402.0811.

[11] C. L. Siegel, *Über die Classenzahl quadratischer Zahlkörper*, Acta Arith. **1** (1935), 83–86.

[12] I. M. Vinogradov, *Sur la distribution des résidus et des non-résidus des puissances*, J. Soc. Phys.-Math. Univ. Perm **1** (1918), 94–98.

[13] A. Walfisz, *Zur additiven Zahlentheorie. II*, Math. Z. **40** (1936), 592–607.

[14] A. Walker, *Correlations of sieve weights and distributions of zeros*, Int. J. Number Theory **19** (2023), no. 1, 139–167, arXiv:2101.04418.