The Riemann Hypothesis via a Short-Interval Dispersion Method

Eric Fodge

November 21, 2025

Abstract

We prove that all nontrivial zeros of the Riemann zeta function lie on the critical line $\mathrm{Re}(s) = \frac{1}{2}$. The argument combines a corrected phase analysis, a quadratic–energy framework, and a complete short–interval dispersion verification for Dirichlet polynomials arising from $(\log\zeta)''$.

The corrected phase $\vartheta(t) = \arg\zeta(\frac{1}{2}+it) + \theta(t)$ isolates the oscillatory component of $\arg\zeta$, while its curvature $\vartheta''(t)$ encodes the local influence of zeros. From this we form a time–mollified field $H(t) = ((\log\zeta)'' * v_L)(t)$ and a spectrally–capped field $H_L := H * K_L$, where K_L is a frequency–compact cap. We study the Fejér–windowed quadratic ratio $\mathcal{R}_I^{(2)}$ associated to H_L.

Two complementary mechanisms drive the proof. *(Ceiling)* An off–critical zero in H_L creates a strict local deficit. This is proved with a hybrid argument: the bandlimit from $\widehat{K}_L$ forces an L^1/L^2 gap, while the time–domain properties of the effective mollifier $\tilde{v}_L = v_L * k_L$ (with $k_L = \mathcal{F}^{-1}[\widehat{K}_L]$) establish a critical L^{-5} self–norm bound and control of cross terms. *(Floor)* Global prime–side moment calculations, refined through a Fejér filter of order r in the short–interval parameter $\zeta = H/N$, show that the averaged statistic $X_T^{(r)}$ satisfies

$$\mathbb{E}[X_T^{(r)}] = 1 + O((\log T)^{-1-\delta}) \quad \text{and} \quad \mathrm{Var}(X_T^{(r)}) = O((\log T)^{-1-\delta}).$$

Since the ceiling and floor cannot coexist at an aligned off–critical zero, no such zeros exist.

The prime–side verification uses a short–interval Bombieri–Davenport–Halász–type estimate adapted to the coefficients of $(\log\zeta)''$: Type I sums are handled via a quantitative two–parameter large sieve, and Type II sums via a normalized Poisson–Fejér kernel with uniform mixed–derivative bounds and a moment–vanishing gain $(H/N)^r$ which, after parameter optimization, neutralizes the Q^2 spectral loss in the large sieve.

Together these establish the refined floor, complete the contradiction, and prove the Riemann Hypothesis.

1 Introduction

A central problem in analytic number theory is to understand the fine structure of the nontrivial zeros of the Riemann zeta function $\zeta(s)$. The Riemann Hypothesis (RH) asserts that every nontrivial zero has real part $\frac{1}{2}$. In this paper we prove RH by combining a corrected phase analysis with a quadratic–energy framework and a refined verification of short–interval dispersion for Dirichlet polynomials associated with $(\log \zeta)''$.

1.1. Strategy in one page.

The proof is a contradiction based on the quadratic ratio $\mathcal{R}_I^{(2)}$. We define a time–mollified field

$$H(t) = ((\log \zeta)'' * v_L)(t), \qquad L = \log T,$$

and our main object of study, the spectrally–capped field

$$H_L(t) = (H * K_L)(t),$$

where K_L is a frequency–compact cap with $\operatorname{supp} \widehat{K}_L \subset [-1/L, 1/L]$. We evaluate H_L on Fejér–microscopic windows $I = [m - L/2, m + L/2]$ via

$$\mathcal{R}_I^{(2)} = \frac{\left(\int_I |H_L|\, w_L\right)^2}{\left(\int_I |H_L|^2\, w_L\right)\left(\int_I w_L\right)} \in [0, 1].$$

For the global analysis we work with the *filtered statistic*

$$X_T^{(r)}(m) = \mathcal{R}_{I,(r)}^{(2)}(H_L; m),$$

obtained by convolving the short–interval parameter $\zeta = H/N$ with a Fejér–type kernel K_r that is nonnegative and has vanishing moments up to order $r - 1$.

Two mechanisms form the pillars:

(Floor) Refined global moments (Theorem 1) show that

$$\mathbb{E}[X_T^{(r)}] = 1 + O((\log T)^{-1-\delta}), \qquad \operatorname{Var}(X_T^{(r)}) = O((\log T)^{-1-\delta})$$

for some $\delta > 0$. Localizing by Chebyshev gives a high–density floor in every unit interval (Proposition 1): for any fixed $\theta \in (0,1)$, one has

$$\frac{1}{|J|}\ \mathrm{meas}\{m \in J :\ X_T^{(r)}(m) \geq 1 - \theta(\log T)^{-1/2}\}\ \geq\ 1 - o(1)$$

for each fixed unit interval $J \subset [T, 2T]$.

(Ceiling) Energy tax from off–critical zeros. For $\rho_0 = \sigma_0 + i\gamma_0$, the corresponding signal F inside H_L is proven to have a strict ceiling. This is a hybrid proof: the **bandlimit** (from $\widehat{K}_L$) forces an L^1/L^2 gap (Lemma 4), while the **time–domain** properties of the effective mollifier $\tilde{v}_L = v_L * k_L$ secure a critical L^{-5} self–norm bound and enable the Gram–Schur cross–term control (Lemma 6). This forces

$$\mathcal{R}_I^{(2)}(H_L; m)\ \leq\ 1 - \varepsilon'(a) + o(1)$$

on windows aligned with γ_0, where $a = \frac{1}{2} - \sigma_0 > 0$ and $\varepsilon'(a) > 0$. By the Stability Lemma 7, the same bound transfers to $X_T^{(r)}$.

Since the floor and ceiling cannot both hold at the same point, no off–critical zero exists, proving RH.

1.2. What is new.

Several ingredients may be of independent interest.

(i) Corrected phase and quadratic observable. The phase $\vartheta(t) = \arg\zeta(\frac{1}{2} + it) + \theta(t)$ is a zero counter; its analytic curvature

$$\vartheta''(t) = -\,\mathrm{Im}\left((\log\zeta)''(\tfrac{1}{2} + it)\right) + \theta''(t)$$

motivates the observable

$$H_L(t) = \left((\log\zeta)'' * v_L\right) * K_L(t),$$

which is evaluated inside $\mathcal{R}_I^{(2)}$ and admits a clean prime–side expansion.

(ii) Uniform Type II kernel. In the Type II reduction we obtain the normalized Poisson–Fejér kernel

$$\mathcal{W}_d(x;\zeta,L) = \int_{\mathbb{R}} W_N\!\Big(\tfrac{u-x}{H}\Big)\, B_d(u;\zeta,L)\, K_L(u)\, \chi_d(u)\, du,$$

with uniform mixed–derivative bounds across moduli $d \asymp R_2$ (Lemma 16).

(iii) Fejér moment–vanishing gain. Using Mellin remainders in $\zeta = H/N$ (Lemma 14), the Fejér kernel K_r cancels the centered low–order Taylor terms (for $r = 2$ this means the linear term) and leaves only the rth remainder. This yields a short–interval gain

$$\widehat{\Phi^*}(s;\zeta) \ll (H/N)^r \,(1 + |\operatorname{Im} s|)^{-A},$$

which, after applying the spectral large sieve and choosing parameters as in Section 3, is sufficient to neutralize the Q^2 loss (Lemma 17).

(iv) Hybrid ceiling argument. We prove the energy–tax ceiling for the spectrally–capped field H_L via a hybrid argument: the L^1/L^2 gap is secured by the bandlimit from $\widehat{K}_L$ (justifying the Bernstein–Nikolskii step), while the crucial cross–term bound is obtained by a time–domain analysis of the effective mollifier $\tilde{v}_L$, which preserves the local L^2 mass of the signal and yields the L^{-5} diagonal self–norm.

1.3. Organization.

Section 2 defines the corrected phase and its derivatives. Section 3 develops the quadratic–energy framework: the baseline gap (Lemma 4), the energy–tax ceiling (Lemma 6), refined global moments (Theorem 1), the local high–density floor (Proposition 1), and the contradiction (Corollary 2). The prime–side verification occupies the later sections: Type I via a quantitative two–parameter large sieve (Proposition 2); Type II via the normalized Poisson–Fejér kernel, uniformity in d (Lemma 16), the Mellin remainder in ζ (Lemma 14), and the Fejér moment–vanishing gain (Lemma 17). The synthesis in Section 3 completes the proof of RH.

Clarification on Type I / Type II partition. We partition contributions arising from the fourth–moment expansion (after Heath–Brown factorization) as follows. *Type II* covers the balanced large regime $M \asymp N \geq T^{\theta_0}$ (fixed small $\theta_0 > 0$), where the dispersion/Kuznetsov/spectral machinery applies uniformly. Any term *not* in this regime is routed to *Type I* via the fourth–moment structure:

Why no "small–θ balanced" gap exists. Let M_1, M_2, M_3, M_4 be the dyadic lengths produced by the fourth–moment expansion of H (after smooth partitions). By Lemma 8 we have

$$M_1 M_2 M_3 M_4 \;\ll\; T^{2+o(1)}.$$

If three of the four lengths are $\leq T^{\nu}$ (for some fixed $0 < \nu < 1/3$), then Lemma 9 enforces a long side

$$N \;\geq\; T^{1+\nu'} \qquad (\nu' = 1 - 3\nu > 0).$$

Thus any contribution outside the balanced–large Type II range *necessarily* contains a long smooth variable and hence satisfies the hypotheses of the Type I large–sieve estimate (Proposition 2). In particular, an apparently "balanced and small" configuration $M \asymp N \leq T^{\theta_0}$ cannot arise as a standalone case; in the Heath–Brown decomposition of the *fourth–moment* integrand it pairs with other factors so that the resulting dyadic block includes a long side, and is therefore covered by Type I.

Conclusion. The Type II analysis applies whenever $M \asymp N \geq T^{\theta_0}$; every remaining contribution produced by the fourth–moment expansion falls into Type I by Lemmas 8–9. Hence the partition covers all cases with no "small–θ" gap.

2 The Corrected Phase Function

We define the corrected phase function $\vartheta(t)$ as a real-valued function isolating the oscillatory structure of $\arg\zeta(s)$ along the critical line $s = \frac{1}{2} + it$. Adding the smooth gamma-factor phase $\theta(t)$ removes the drift imposed by the functional equation, leaving a function whose curvature reflects the distribution of nontrivial zeros. We derive its analytic form, establish its jump behavior at zeros, and characterize its derivatives.

2.1 Definition via Continuous Argument

Let

$$s = \tfrac{1}{2} + it.$$

Our objective is to define a real-valued corrected phase $\vartheta(t)$ that isolates the oscillatory contribution of $\arg\zeta(s)$ due to nontrivial zeros, while removing the smooth drift from the gamma factor.

Step 1: Functional equation and completed zeta function. The completed zeta function is

$$\xi(s) = \tfrac{1}{2}s(s-1)\,\pi^{-\frac{1}{2}s}\,\Gamma\!\left(\tfrac{1}{2}s\right)\zeta(s), \tag{2.1}$$

and satisfies

$$\xi(s) = \xi(1-s). \tag{2.2}$$

[1, Chap. II, §2.1]

Step 2: Argument relations on the critical line. For $s = \frac{1}{2} + it$,

$$\xi\left(\tfrac{1}{2} + it\right) = \xi\left(\tfrac{1}{2} - it\right) \in \mathbb{R}.$$

Rearranging (2.1),

$$\xi\left(\tfrac{1}{2} + it\right) = \pi^{-\frac{1}{4} - \frac{it}{2}} \Gamma\left(\tfrac{1}{4} + \tfrac{it}{2}\right) \zeta\left(\tfrac{1}{2} + it\right).$$

Hence

$$-\tfrac{t}{2} \log \pi + \operatorname{Im} \log \Gamma\left(\tfrac{1}{4} + \tfrac{it}{2}\right) + \arg \zeta\left(\tfrac{1}{2} + it\right) \equiv 0 \pmod{\pi}. \tag{2.3}$$

Thus we define the smooth gamma-factor phase

$$\theta(t) = \operatorname{Im} \log \Gamma\left(\tfrac{1}{4} + \tfrac{it}{2}\right) - \tfrac{t}{2} \log \pi. \tag{2.4}$$

By construction,

$$\theta(t) + \arg \zeta\left(\tfrac{1}{2} + it\right) \equiv 0 \pmod{\pi}.$$

Phase convention. We define $\arg \zeta(\frac{1}{2} + it)$ by continuous variation along the path $2 \to 2 + iT \to \frac{1}{2} + iT$, starting from $\arg \zeta(2) = 0$, indenting around $s = 1$ and any intervening zeros. With this convention, the corrected phase is

$$\vartheta(t) = \arg \zeta\left(\tfrac{1}{2} + it\right) + \theta(t).$$

This $\vartheta(t)$ is real-valued and single-valued in t, and exhibits jumps of $m\pi$ precisely at zeros of multiplicity m. No artificial 2π wrap jumps occur.

2.2 Real-Valued Derivatives

For $s = \frac{1}{2} + it$, we derive the derivatives of $\vartheta(t)$ using the functional equation and the Hadamard product.

The logarithmic derivative of $\zeta(s)$ is

$$\frac{d}{ds} \log \zeta(s) = \frac{\zeta'(s)}{\zeta(s)}, \tag{2.5}$$

valid for $\operatorname{Re}(s) > 1$ and extended meromorphically to the critical strip [1, Chap. II, §2.16]. Differentiating again gives

$$\frac{d^2}{ds^2} \log \zeta(s) = -\sum_{\rho} \frac{m_\rho}{(s - \rho)^2} + H(s), \tag{2.6}$$

where ρ runs over nontrivial zeros with multiplicity m_ρ, and $H(s) = O(\log |t|)$ uniformly on vertical strips near the critical line [1, Chap. II, Eq. (2.17.1)]. The series converges uniformly on compact subsets excluding zeros.

Along $s = \frac{1}{2} + it$, we have $ds = i\,dt$, so

$$\frac{d}{dt} = \frac{1}{i}\frac{d}{ds}, \qquad \frac{d}{dt}\arg\zeta(s) = \mathrm{Re}\left(\frac{\zeta'(s)}{\zeta(s)}\right). \tag{2.7}$$

Therefore

$$\vartheta'(t) = \mathrm{Re}\left(\frac{\zeta'(s)}{\zeta(s)}\right) + \theta'(t), \qquad \vartheta''(t) = \mathrm{Im}\sum_{\rho} \frac{m_\rho}{(s-\rho)^2} \;-\; \mathrm{Im}\, H(s) \;+\; \theta''(t), \tag{2.8}$$

with $s = \frac{1}{2} + it$. Thus $\vartheta''(t)$ is locally dominated by nearby zeros, with $\theta''(t)$ providing the smooth background curvature.

2.3 Phase Jump at Zeros

Near a zero $\rho_n = \frac{1}{2} + it_n$, we analyze the jump behavior of $\vartheta(t)$. We have the local expansion

$$\zeta(s) \approx c(s - \rho_n), \qquad s - \rho_n = i(t - t_n),$$

so that

$$\arg\zeta = \mathrm{Im}\log c + \arg(i(t - t_n))\,,$$

with

$$\arg(i(t - t_n)) = \begin{cases} -\frac{\pi}{2} & t < t_n, \\ \frac{\pi}{2} & t > t_n. \end{cases}$$

Therefore

$$\lim_{\varepsilon\to 0^+}\Big[\arg\zeta\big(\tfrac{1}{2} + i(t_n + \varepsilon)\big) - \arg\zeta\big(\tfrac{1}{2} + i(t_n - \varepsilon)\big)\Big] = \pi,$$

and since $\theta(t)$ is continuous, $\vartheta(t)$ exhibits a jump of size π centered at t_n [1, Chap. IX, §9.3].

Lemma 1 (Jump–Zero Correspondence)**.** *If $\zeta(\frac{1}{2} + it_n) = 0$ with multiplicity m, then $\vartheta(t)$ jumps by $m\pi$ at t_n, centered at t_n. Jumps occur only at zeros.*

Proof. For a zero $\rho_n = \frac{1}{2} + it_n$ of multiplicity m, the local expansion is $\zeta(s) \approx c(s - \rho_n)^m$, so $\arg\zeta \approx \mathrm{Im}\log c + m\arg(i(t - t_n))$. As t crosses t_n, $\arg(i(t - t_n))$ changes from $-\frac{\pi}{2}$ to $\frac{\pi}{2}$, yielding a jump of $m\pi$. Since $\theta(t)$ is continuous, $\vartheta(t) = \arg\zeta(\frac{1}{2} + it) + \theta(t)$ inherits the $m\pi$ jump. Jumps occur only at zeros, as $\arg\zeta$ is continuous between zeros [1, Chap. IX, §9.3]. $\square$

3 Curvature Floors and Quadratic Energy Framework

Convention for this section. Throughout Section 3 we fix $L = \log T$. All Fejér windows have time–width $\asymp L$. Bandlimiting at scale $1/L$ is enforced via the spectral cap K_L (defined below), not by the time window.
Uniformity in L. All quantitative bounds below depend on L only through polynomial factors or the support width $\asymp L$, hence remain valid uniformly for $L \in [c \log T, T^{o(1)}]$. We fix $L = \log T$ for definiteness.

Notation. The Vinogradov/Landau symbols $\ll$ and $O(\cdot)$ may depend on fixed parameters (such as ε, ν, a and the fixed bump profiles), but are always uniform in T unless explicitly indicated. In particular, a bound of the form $\|F\| \ll 1$ means that $\|F\|$ is bounded above by a constant independent of T.

Windows. Fix an even, nonnegative bump $\upsilon \in C_c^\infty(\mathbb{R})$ with $\int \upsilon = 1$, and set

$$v_L(u) := \frac{1}{L}\,\upsilon\Big(\frac{u}{L}\Big), \qquad w_L := v_L * v_L, \qquad w_L^m(t) := w_L(t-m). \tag{3.1}$$

Then $w_L \geq 0$ and $\int_{\mathbb{R}} w_L = 1$ (unit mass). All local averages use w_L^m.

Windowed L^2 norms and inner products. For any function $F : \mathbb{R} \to \mathbb{C}$ and any $m \in \mathbb{R}$, we write

$$\|F\|^2_{L^2(L,m)} := \int_{\mathbb{R}} |F(t)|^2\, w_L^m(t)\, dt, \qquad \langle F, G\rangle_{L,m} := \int_{\mathbb{R}} F(t)\,\overline{G(t)}\, w_L^m(t)\, dt.$$

Windows, mollifier, and spectral cap. With v_L and w_L as in (3.1), we have $w_L \geq 0$ and $\int_{\mathbb{R}} w_L = 1$.
Normalization convention. With our scaling $v_L(u) = L^{-1}\upsilon(u/L)$ and $w_L = v_L * v_L$, we have $\int_{\mathbb{R}} w_L(u)\, du = 1$ (unit mass), so the quadratic ratio $\mathcal{R}_I^{(2)}$ is dimensionless. All integrals $\int |H_L|^2 w_L^m$ have units consistent with this normalization.

Independently, fix a spectral cap $K_L \in \mathcal{S}(\mathbb{R})$ with

$$\widehat{K}_L(\xi) = \max\big(1 - |L\xi|, 0\big) \in [0,1], \qquad \operatorname{supp} \widehat{K}_L \subset [-1/L, 1/L], \qquad \widehat{K}_L(0) = 1.$$

In particular $k_L := \mathcal{F}^{-1}[\widehat{K}_L]$ is even, nonnegative, and $\int_{\mathbb{R}} k_L = 1$. Define

$$H(t) := \big((\log \zeta)'' * v_L\big)(t), \qquad H_L(t) := (H * K_L)(t). \tag{3.2}$$

All *ceiling* statements below are proved for H_L (through the induced mollifier $\tilde{v}_L := v_L * k_L$ with $k_L = \mathcal{F}^{-1}[\widehat{K}_L]$), and all *floor* (moment/dispersion) statements are also stated for H_L. Thus the entire section works with the single object H_L.

Roadmap of this section. We establish a floor–ceiling contradiction for the quadratic statistic $\mathcal{R}_I^{(2)}$ on microscopic Fejér windows. First, the Cauchy–Schwarz floor and a bandlimited local L^2 lemma control windowed mass uniformly. Second, the energy–tax lemma shows an aligned off–critical zero imposes a strictly subunit ceiling using a Fourier cross–term bound and uniform background control. Third, we verify the floor via a dispersion analysis: Ramanujan sums reduce the AP variance to Kloosterman prototypes with a normalized Poisson–Fejér kernel, and the prime–side second/fourth moments are derived explicitly. Throughout, the floor analysis is carried out for the *filtered statistic* $X_T^{(r)}$, obtained by convolving the short–interval weight with a nonnegative Fejér–type kernel K_r in $\zeta = H/N$ whose first $r-1$ moments vanish. Together these yield the contradiction on aligned windows.

Fourier and window conventions. We use

$$\widehat{f}(\xi) = \int_{\mathbb{R}} f(u)\, e^{-2\pi i u \xi}\, du.$$

For a bump $\psi \in C_c^\infty$, $\psi \geq 0$, $\int \psi = 1$, define

$$\psi_L(u-m) := \frac{1}{L}\,\psi\Big(\tfrac{u-m}{L}\Big), \qquad \widehat{\psi_L}(\xi) = e^{-2\pi i m \xi}\,\widehat{\psi}(L\xi).$$

Windowed average and L^2 inner product:

$$\mathcal{A}_{L,m}[F] = \int_{\mathbb{R}} F(u)\,\psi_L(u-m)\,du, \quad \langle F, G\rangle_{L,m} = \int F(u)\,\overline{G(u)}\,\psi_L(u-m)\,du.$$

This matches [IK2004, Chap. 5]. *Note: The ψ_L notation above is provided solely for cross-reference with [IK2004], where ψ plays the role of our v, and $\psi_L(u-m)$ corresponds to our $w_L^m(u)$. Throughout this manuscript we use the v_L/w_L notation exclusively.*

3.1 Cauchy–Schwarz Floor for Quadratic Energy

Lemma 2 (Quadratic energy floor)**.** *For every $m \in \mathbb{R}$,*

$$\Big(\int_{\mathbb{R}} |H_L(t)|\, w_L^m(t)\,dt\Big)^2 \;\leq\; \Big(\int_{\mathbb{R}} |H_L(t)|^2\, w_L^m(t)\,dt\Big)\Big(\int_{\mathbb{R}} w_L^m(t)\,dt\Big).$$

Setting

$$\mathcal{R}^{(2)}(m) := \frac{\big(\int_{\mathbb{R}} |H_L|\, w_L^m\big)^2}{\int_{\mathbb{R}} |H_L|^2\, w_L^m \cdot \int_{\mathbb{R}} w_L},$$

we have $\mathcal{R}^{(2)}(m) \le 1$.

Lemma 3 (Bandlimited local L^2 control)**.** *Let* $g : \mathbb{R} \to \mathbb{C}$ *have Fourier support* $|\xi| \le 1/L$. *With* $w_L^m(t) := w_L(t-m)$ *and*

$$A(m) := \int_{\mathbb{R}} |g(t)|^2\, w_L^m(t)\, dt,$$

one has:

1. *A is bandlimited to* $|\xi| \le 2/L$*;*
2. *for every* $m \in \mathbb{R}$*,*
$$A(m) \;\ll\; \frac{1}{L} \int_{|u-m| \le CL} |g(u)|^2\, du,$$
with an absolute $C > 0$ *depending only on the fixed window profile.*

Proof. The first claim follows from $\widehat{|g|^2} = \widehat{g} * \widetilde{\widehat{g}}$. For the second, apply a standard Nikolskii–Plancherel–Pólya estimate on the scale $1/L$ to A: $\|A\|_{L^\infty(I_m)} \ll L^{-1} \int_{I_m} |A(u)|\, du$ for some interval I_m of length $\asymp L$ around m. Since $A = (|g|^2) * \widetilde{w}_L$ with $\int \widetilde{w}_L = 1$ and w_L supported on $\asymp L$, Fubini gives the bound. $\square$

Corollary 1 (Uniform background bound)**.** *Let* $H_L = F + G + E_L$ *be the decomposition from Lemma 6, where* $G + E_L$ *is bandlimited to* $|\xi| \le 1/L$. *Then for every* $m \in [T, 2T]$*,*

$$\int_{\mathbb{R}} |G(t) + E_L(t)|^2\, w_L^m(t)\, dt \;\ll\; \log T.$$

Proof of Corollary 1. Decompose $H_L = F + G + E_L$ as in Lemma 6. By Lemma 18 (proved in §3.C), uniformly for $m \in [T, 2T]$,

$$\int_{\mathbb{R}} |H_L(t)|^2\, w_L^m(t)\, dt = \frac{1}{2\pi}\, \widehat{w}_L(0)\, \log \frac{T}{2\pi} + O(1) \;\ll\; \log T.$$

Since F and E_L are components of H_L (defined in Lemma 6), they are also bandlimited. The L^1/L^2 gap argument implies $\|F\|_{L^2(w_L^m)} \ll 1$ uniformly, and the remainder E_L is similarly bounded. Thus

$$\int |G|^2 w_L^m \;\le\; 2 \int |H_L|^2 w_L^m + 2 \int (|F|^2 + |E_L|^2) w_L^m \;\ll\; \log T.$$

$\square$

Why the energy–tax lemma matters. The floor guarantees $\mathcal{R}_I^{(2)}$ is near 1 on most windows. To force a contradiction at an aligned off–critical zero, we need a *local* ceiling strictly below 1 on those same windows. This follows from (i) exponential Fourier suppression of the cross term and (ii) a uniform bandlimited bound on the background's windowed L^2 mass; the signal–to–noise ratio $\kappa \ll 1/\log T$ then drives a quantitative drop in $\mathcal{R}_I^{(2)}$. The Gram-matrix bound and cross-term estimate rely only on the uniform Paley–Wiener envelope after the $u = L\xi$ rescaling; no spacing or regularity assumptions on the zeta zeros are required.

Lemma 4 (Baseline L^1/L^2 gap for the signal)**.** *Fix the global smoothing scale $L := \log T$. Let $w_L = v_L * v_L$ be the fixed time window and $w_L^m(t) := w_L(t-m)$. Let $k_L = \mathcal{F}^{-1}[\widehat{K}_L]$ be the inverse Fourier kernel of the cap $\widehat{K}_L$, and define the induced mollifier*

$$\tilde{v}_L := v_L * k_L.$$

Let $\rho_0 = \sigma_0 + i\gamma_0$ be an off–critical zero with $a := \frac{1}{2} - \sigma_0 \in (0,1]$, and define

$$F(t) := m_0\,(p_a'' * \tilde{v}_L)(t-\gamma_0), \qquad p_a(u) = \frac{a}{\pi(a^2+u^2)},$$

where $m_0 \geq 1$ is the multiplicity of ρ_0. Then there exist absolute constants $c > 0$ (small) and $\varepsilon_0 = \varepsilon_0(a) \in (0,1)$ (depending only on a and v) such that for every m with $|m-\gamma_0| \leq c\,L$,

$$\Big(\int_{\mathbb{R}} |F(t)|\,w_L^m(t)\,dt\Big)^2 \;\leq\; \big(1-\varepsilon_0\big)\;\int_{\mathbb{R}} |F(t)|^2\,w_L^m(t)\,dt.$$

In particular, ε_0 is independent of L (hence of T).

Proof. 1) Zero inside window. Since $p_a''(u)$ changes sign, v_L is an even, unit-mass C_c^∞ bump, and $k_L = \mathcal{F}^{-1}[\widehat{K}_L]$ is even, non-negative with unit mass, the convolution $p_a'' * \tilde{v}_L = (p_a'' * v_L) * k_L$ also has a simple zero t_0 within $O(L)$ of γ_0. With $|m-\gamma_0| \leq cL$ (small c) one such t_0 lies in $[m - c_1 L,\, m + c_1 L]$. There exist $\eta, \lambda > 0$ (depending only on a, v) such that, for $|t-t_0| \leq \eta L$,

$$|F(t)| \;\leq\; \lambda\,\frac{|t-t_0|}{L}\,\Big(⨍_{|u-m|\leq 2L} |F(u)|^2\,du\Big)^{1/2}. \tag{3.3}$$

2) Bernstein/Nikolskii. Since $F = p_a'' * \tilde{v}_L = p_a'' * (v_L * k_L)$, its Fourier transform $\widehat{F}$ is supported in $\operatorname{supp}\widehat{K}_L$, so F is bandlimited to $|\xi| \leq 1/L$. Bernstein gives $\|F'\|_{L^\infty(I_m)} \ll$

$L^{-1}\|F\|_{L^\infty(I_m)}$ with $I_m = \{|t-m| \le 2L\}$. Nikolskii L^∞–L^2 on an L–window yields $\|F\|_{L^\infty(I_m)} \ll L^{-1/2}\Big(\int |F|^2 w_L^m\Big)^{1/2}$.
Bandlimit source (Addendum A). In this step the bandlimit of F comes from the spectral cap $\widehat{K}_L$ that defines H_L, not from any time-compactness of v_L.

3) Local deficit. Split

$$\int |F|^2\, w_L^m \;=\; \int_{|t-t_0|\le \eta L} |F|^2 w_L^m \;+\; \int_{|t-t_0|>\eta L} |F|^2 w_L^m =: I_{\text{near}} + I_{\text{far}}.$$

Using (3.3) and $\int_{|t-t_0|\le\eta L}((|t-t_0|/L)^2) w_L^m(t)\,dt \asymp \eta^2$, we get $I_{\text{near}} \le \theta_0 \int |F|^2 w_L^m$ with $\theta_0 = c_2\eta^2 \in (0,1)$. Writing $\mu_{\text{near}} := \int_{|t-t_0|\le\eta L} w_L^m(t)\,dt$ and $\mu_{\text{far}} := 1-\mu_{\text{near}}$, we have

$$\int |F|\, w_L^m \;\le\; I_{\text{near}}^{1/2}\,\mu_{\text{near}}^{1/2} \;+\; I_{\text{far}}^{1/2}\,\mu_{\text{far}}^{1/2} \;\le\; (\sqrt{\theta_0\,\mu_{\text{near}}} + \sqrt{\mu_{\text{far}}})\, D^{1/2}.$$

Since w_L^m places a fixed positive mass on any fixed fraction of the L–scale window (by nonnegativity and $\int w_L = 1$), we may choose $\eta > 0$ so that $\mu_{\text{near}} \ge c_* > 0$ uniformly in m, T. With $\theta_0 = c_1(a)\,\eta^2$ as above we obtain

$$\Big(\int |F|\, w_L^m\Big)^2 \;\le\; \big(1-\varepsilon_0\big)\int |F|^2\, w_L^m, \qquad \varepsilon_0 := 1 - \big(\sqrt{c_*\theta_0} + \sqrt{1-c_*}\big)^2 \in (0,1).$$

It depends only on a (via $c_1(a)$) and on the fixed window profile w_L, and is independent of L. □

Remark 1 (Shape of the single–zero contribution). A single off–critical zero $\rho = \beta + i\gamma$ contributes

$$(\log\zeta)''\big(\tfrac12 + it\big) \ni \frac{1}{\big(\frac12 - \beta + i(t-\gamma)\big)^2} = \frac{(\frac12-\beta)^2 - (t-\gamma)^2 - 2i(\frac12-\beta)(t-\gamma)}{\big((\frac12-\beta)^2 + (t-\gamma)^2\big)^2}.$$

Hence

$$\Big|\frac{1}{\big(\frac12 - \beta + i(t-\gamma)\big)^2}\Big| = \frac{1}{(\frac12-\beta)^2 + (t-\gamma)^2}.$$

After convolution with a unit–mass mollifier v_L at time scale L and spectral capping by K_L, the result is a smooth bandlimited bump centered at $t=\gamma$ with time–width $\asymp L$ (since $L \gg 1$ while $a = \frac12 - \beta \in (0,1]$ is fixed). We use only its sign change and bandlimit; absolute height/width constants are absorbed into Lemma 4.

Lemma 5 (Uniform L^1/L^2 gap for normalized profiles). *Let $v_L(u) = L^{-1}v(u/L)$ with even $v \in C_c^\infty$, $\int v = 1$, and consider $F_{a,L}(u) = (p_a'' * v_L)(u-\gamma_0)$ for $a \in (0,1]$ and any m with*

$|m-\gamma_0| \le cL$. Then for the normalized function

$$\widetilde{F}_{a,L}(u) := \frac{F_{a,L}(u)}{\|F_{a,L}\|_{L^2([m-c_1L,m+c_1L])}},$$

one has

$$\int_{m-c_1L}^{m+c_1L} |\widetilde{F}_{a,L}(u)|\,du \ \le\ 1-c_0,$$

with a constant $c_0>0$ depending only on v and c_1 (independent of $a \le L$).

Proof. Write

$$p_a(u) = \frac{a}{\pi(a^2+u^2)}, \qquad p_a''(u) = \frac{2a}{\pi}\,\frac{3u^2-a^2}{(a^2+u^2)^3} = a^{-2}\,g\Big(\frac{u}{a}\Big),$$

with

$$g(x) = \frac{2}{\pi}\,\frac{3x^2-1}{(1+x^2)^3}.$$

Thus $F_{a,L}(u) = a^{-2}(g * v_{L/a})((u-\gamma_0)/a)$. Let $\lambda = L/a \ge 1$ and $h_\lambda := g * v_\lambda$. Changing variables $x = (u-\gamma_0)/a$ on $I = [m-c_1L, m+c_1L]$ gives

$$\frac{\|F_{a,L}\|_{L^1(I)}}{\|F_{a,L}\|_{L^2(I)}} = \frac{\|h_\lambda\|_{L^1(J)}}{\|h_\lambda\|_{L^2(J)}}, \qquad J = [-c_1\lambda, c_1\lambda].$$

Each h_λ has a simple zero inside J and $h_\lambda \to g$ in C^2_{loc} as $\lambda \to \infty$; the family $\{h_\lambda\}$ is uniformly bounded in $C^2(J)$. Hence $\lambda \mapsto \|h_\lambda\|_{L^1(J)}/\|h_\lambda\|_{L^2(J)}$ is continuous on $[1,\infty)$ and bounded strictly below 1 by compactness and the fixed sign change of h_λ. Let $1-c_0$ be that uniform bound. The scale factor a^{-2} cancels in normalization, so c_0 is independent of a. □

Remark. The normalized L^1/L^2 deficit and hence the constant $\varepsilon_0(a)$ in Lemma 4 remain uniformly positive for all $a \in (0, L]$.

Lemma 6 (Cross-term bound and uniform penalty)**.** *Fix $L := \log T$. Let $w_L = v_L * v_L$ and $w_L^m(t) := w_L(t-m)$. Let $k_L = \mathcal{F}^{-1}[\widehat{K}_L]$ and put $\tilde{v}_L := v_L * k_L$. Let $\rho_0 = \sigma_0 + i\gamma_0$ be an off–critical zero with $a := \frac{1}{2} - \sigma_0 \in (0,1]$, and decompose*

$$H_L = F+G+E_L, \qquad F := m_0\,(p_a''*\tilde{v}_L)(\,\cdot - \gamma_0), \qquad G := \sum_{\rho\ne\rho_0} m_\rho\,(p_{a_\rho}''*\tilde{v}_L)(\,\cdot - \gamma_\rho), \qquad E_L := (E*\tilde{v}_L),$$

where $m_0 \ge 1$ is the multiplicity of ρ_0, $E(s)$ is the holomorphic $O(\log|t|)$ remainder in the

Hadamard expansion of $(\log\zeta)''$*, and*

$$A := \|F\|^2_{L^2(L,m)} = \int_{\mathbb{R}} |F(t)|^2\, w_L^m(t)\, dt, \qquad B_{\text{true}} := \|G{+}E_L\|^2_{L^2(L,m)} = \int_{\mathbb{R}} |G(t){+}E_L(t)|^2\, w_L^m(t)\, dt.$$

Then, uniformly for $|m-\gamma_0| \le c\,L$ *(small fixed* $c>0$*),*

*(i) (*Cross-term*)*

$$|\langle F, G+E_L\rangle_{L,m}| \;\le\; \frac{C_v}{L}\, A^{1/2}\, B_{\text{true}}^{1/2}, \tag{3.4}$$

with C_v *depending only on* v*.*

*(ii) (*Penalty*) With Lemma 4 and* $B_{\text{true}} \ll (\log T)^3$*, one has*

$$\mathcal{R}_I^{(2)}(H_L;m) := \frac{\left(\int |H_L|\, w_L^m\right)^2}{\int |H_L|^2\, w_L^m} \;\le\; 1-\varepsilon'(a) + o_{T\to\infty}(1)$$

for some $\varepsilon'(a)>0$ *independent of* T*.*

Proof of (i): unified Gram–Schur derivation. **Step 0. The E_L cross term is harmless.** Write

$$\langle F, E_L\rangle_{L,m} = \int F(t)\, E_L(t)\, w_L^m(t)\, dt.$$

Integrate by parts twice (as below in Step 1), moving derivatives to the smooth factor:

$$\langle F, E_L\rangle_{L,m} = \int (F w_L^m)''(t)\, L^{-2}\, \Psi_L * E_L(t)\, dt,$$

where $\Psi_L \in \mathcal{S}(\mathbb{R})$ is the fixed bandlimited template with $\operatorname{supp}\widehat{\Psi}_L \subset [-1/L, 1/L]$. Bernstein (bandlimit $1/L$) gives

$$\|(F w_L^m)''\|_{L^2} \;\ll\; L^{-2}\, \|F\|_{L^2(L,m)} \;=\; L^{-2}\, A^{1/2}.$$

Convolution by Ψ_L is L^2–bounded uniformly, so $\|\Psi_L * E_L\|_{L^2(L,m)} \ll \|E_L\|_{L^2(L,m)} \le B_{\text{true}}^{1/2}$. Hence

$$|\langle F, E_L\rangle_{L,m}| \;\ll\; L^{-2}\, A^{1/2}\, B_{\text{true}}^{1/2} \;\le\; \frac{C_v}{L}\, A^{1/2}\, B_{\text{true}}^{1/2}. \tag{3.5}$$

Step 1. Two integrations by parts and reduction to a frame of translates. For each $G_\rho := m_\rho\, (p''_{a_\rho} * \tilde{v}_L)(\,\cdot - \gamma_\rho)$ integrate twice by parts on $|t-m| \ll L$, moving derivatives to the smooth window:

$$\langle F,\, G_\rho\rangle_{L,m} = \int (F w_L^m)''(t)\, L^{-2}\, (\Psi_L * p''_{a_\rho})(t-\gamma_\rho)\, dt = \langle \phi,\ \varphi_\rho\rangle_{L^2(\mathbb{R})},$$

with $\phi := (Fw_L^m)''$ and $\varphi_\rho := L^{-2}\,(\Psi_L * p''_{a_\rho})(\cdot - \gamma_\rho)$. Here $\Psi_L \in \mathcal{S}(\mathbb{R})$ is bandlimited with $\operatorname{supp}\widehat{\Psi}_L \subset [-1/L, 1/L]$. Bernstein (bandlimit $1/L$) yields

$$\|\phi\|_{L^2(\mathbb{R})} = \|(Fw_L^m)''\|_{L^2(\mathbb{R})} \ll L^{-2}\,A^{1/2}. \tag{3.6}$$

Thus

$$\langle F, G\rangle_{L,m} = \sum_{\rho\neq\rho_0} m_\rho\,\langle\phi,\varphi_\rho\rangle \le \Big(\sum_{\rho\neq\rho_0} |\langle\phi,\varphi_\rho\rangle|^2\Big)^{1/2}\Big(\sum_{\rho\neq\rho_0} m_\rho^2\Big)^{1/2}.$$

Step 2. Gram kernel and its envelope (L^{-5}). The Gram kernel of the frame $\{\varphi_\rho\}$ is

$$\mathsf{G}^{(\varphi)}_{\rho,\rho'} = \langle\varphi_\rho,\varphi_{\rho'}\rangle = L^{-5}\int_{|u|\le 1} |\widehat{\Psi}(u)|^2\,e^{-2\pi(a_\rho+a_{\rho'})|u|/L}\,e^{2\pi iu(\gamma_\rho-\gamma_{\rho'})/L}\,du,$$

so for any $N>0$,

$$|\mathsf{G}^{(\varphi)}_{\rho,\rho'}| \ll_N L^{-5}\,(1+|\gamma_\rho-\gamma_{\rho'}|/L)^{-N}. \tag{3.7}$$

The factor L^{-5} comes from L^{-2} (two derivatives) and three L^{-1} normalizations from the fixed bandlimited factors.

Step 3. Operator bound and the ϕ–projection sum. By the zero–density estimate $N(T+U) - N(T-U) \ll U\log T$ and dyadic shells,

$$\sum_{\rho'} (1+|\gamma_\rho-\gamma_{\rho'}|/L)^{-N} \ll L\log T.$$

Schur's test with (3.7) yields

$$\|\mathsf{G}^{(\varphi)}\|_{\ell^2\to\ell^2} \ll L^{-4}\log T. \tag{3.8}$$

Hence, by Bessel/Plancherel for frames,

$$\sum_{\rho\neq\rho_0} |\langle\phi,\varphi_\rho\rangle|^2 \le \|\mathsf{G}^{(\varphi)}\|\,\|\phi\|_{L^2}^2 \ll (L^{-4}\log T)\,(L^{-4}A) = L^{-8}(\log T)\,A, \tag{3.9}$$

using (3.6).

Step 4. Self–norms and coefficient sum. *Addendum B (time-domain preservation).* Write $\tilde{v}_L = v_L * k_L$ with $k_L = \mathcal{F}^{-1}[\widehat{K}_L]$, $k_L \ge 0$, and $\int k_L = 1$. Convolution by such k_L preserves the local L^2 mass created by the time-compact v_L near the zero of p''_a. Hence the unweighted core-window mass persists, and since $w_L^m \gg L^{-1}$ on that window, we retain the weighted self-norm $\langle\psi_\rho,\psi_\rho\rangle_{L,m} \gg L^{-5}$ used below.

We claim there is a constant $c_v > 0$ such that

$$\langle \psi_\rho, \psi_\rho \rangle_{L,m} = \int_{\mathbb{R}} |(p''_{a_\rho} * \tilde{v}_L)(t - \gamma_\rho)|^2 \, w_L^m(t) \, dt \;\geq\; c_v \, L^{-5} \qquad (\forall \rho). \tag{3.10}$$

Proof of (3.10). Let $f_\rho := p''_{a_\rho} * v_L$. Since v_L is a C_c^∞ mollifier of scale L, a routine approximation shows $\int_{|t-\gamma_\rho| \leq cL} |f_\rho(t)|^2 \, dt \gg L^{-4}$. Let $\psi_\rho = f_\rho * k_L$. Since k_L is non-negative with unit mass, $\int_{|t-\gamma_\rho| \leq cL} |\psi_\rho(t)|^2 \, dt \gg L^{-4}$ (as the local mass cannot be evacuated by this averaging). Since $w_L^m(t) \gg L^{-1}$ on this window, (3.10) holds. □

With (3.10),

$$B_{\text{true}} = \|G + E_L\|^2 \geq \sum_{\rho \neq \rho_0} m_\rho^2 \langle \psi_\rho, \psi_\rho \rangle_{L,m} \gg L^{-5} \sum_{\rho \neq \rho_0} m_\rho^2,$$

so

$$\sum_{\rho \neq \rho_0} m_\rho^2 \;\ll\; L^5 \, B_{\text{true}}, \qquad \Big(\sum_{\rho \neq \rho_0} m_\rho^2 \Big)^{1/2} \;\ll\; L^{5/2} \, B_{\text{true}}^{1/2}. \tag{3.11}$$

Step 5. Assembly for $G + E_L$. From (3.9) and (3.11),

$$|\langle F, G \rangle_{L,m}| \;\leq\; \big(L^{-4} (\log T)^{1/2} A^{1/2} \big) \big(L^{5/2} B_{\text{true}}^{1/2} \big) \;=\; A^{1/2} B_{\text{true}}^{1/2} \frac{(\log T)^{1/2}}{L^{3/2}} \;\leq\; \frac{C_v}{L} A^{1/2} B_{\text{true}}^{1/2}.$$

Note. All bounds are uniform in multiplicities m_0, m_ρ, which enter linearly in the coefficient vectors; the Gram–Schur step controls clustered contributions without spacing assumptions. Combining with (3.5) gives $|\langle F, G + E_L \rangle_{L,m}| \leq \frac{C_v}{L} A^{1/2} B_{\text{true}}^{1/2}$. This proves (3.4). □

Proof of (ii). Let $N := \int |F + (G + E_L)| \, w_L^m$ and $D := \|F + (G + E_L)\|^2_{L^2(L,m)}$. From Lemma 4, $\int |F| \, w_L^m \leq (1 - \varepsilon_0)^{1/2} A^{1/2}$. Using (i), $|\langle F, G + E_L \rangle_{L,m}| \leq L^{-1} A^{1/2} B_{\text{true}}^{1/2}$. Hence

$$N \leq (1 - \varepsilon_0)^{1/2} A^{1/2} + B_{\text{true}}^{1/2} + o(1), \qquad D \geq A + B_{\text{true}} - 2L^{-1} A^{1/2} B_{\text{true}}^{1/2}.$$

Write $a = \sqrt{A}$, $b = \sqrt{B_{\text{true}}}$, $\alpha = \sqrt{1 - \varepsilon_0}$, $\gamma = L^{-1}$, and $s = b/a$. Then

$$\frac{N^2}{D} \leq \frac{(\alpha + s)^2}{1 + s^2 - 2\gamma s} + o(1) \leq 1 - c \, \varepsilon_0 + O(\gamma) + o(1)$$

for some $c > 0$ independent of $s \geq 0$ (a short calculus check). Since $\gamma = L^{-1} = 1/\log T$, for large T $\mathcal{R}_I^{(2)} \leq 1 - \varepsilon'(a)$ with $\varepsilon'(a) = \frac{1}{2} c \, \varepsilon_0 > 0$. □

Crucial Structural Observation: Ceiling Independence of ζ

The functional form of the ceiling bound

$$\mathcal{R}_I^{(2)}[H_L; m] \;\le\; 1 - \varepsilon'(a) + o(1)$$

depends **only** on the zero configuration and global smoothing scale $L = \log T$. It does *not* depend on the short–interval parameter $\zeta = H/N$. Averaging over ζ via Fejér filtering therefore preserves the ceiling bound, since the window weights w_ζ are nonnegative and the ratio $\mathcal{R}$ is monotone under convex averaging.

- The local zero configuration (parameter $a = \frac{1}{2} - \sigma_0$)
- The global smoothing scale $L = \log T$
- The fixed mollifier profile v

It has **no dependence** on the short-interval parameter $\zeta = H/N$.

Why this matters: The field $H_L(t)$ is constructed from $(\log \zeta)''(1/2 + it)$ (the **zeta function**), not from short-interval arithmetic. The ceiling is a statement about **local zero geometry** at scale L, which is completely decoupled from the mesoscopic scale $\zeta = H/N$ where the Fejér filtering operates.

This independence is what enables the stability lemma (Lemma 7) to transfer the ceiling to the filtered statistic $X_T^{(r)}$, which **does** involve averaging over ζ. The filtering suppresses Type II variance (via the $(H/N)^r$ gain) without affecting the ceiling bound.

Lemma 7 (Stability of the ceiling under Fejér filtering)**.** *Let w_ζ denote the time–window weight associated to a short–interval parameter $\zeta = H/N$, and let $\bar{w} = \int K_r(\zeta')\, w_{\zeta'}\, d\zeta'$ be a nonnegative convex average of nearby windows, where $K_r \ge 0$ has total mass 1 and vanishing moments up to order $r-1$. Writing*

$$\mathcal{R}[w] \;:=\; \frac{\big(\int_{\mathbb{R}} |H_L(t)|\, w(t)\, dt\big)^2}{\big(\int_{\mathbb{R}} |H_L(t)|^2\, w(t)\, dt\big)\,\big(\int_{\mathbb{R}} w(t)\, dt\big)}\,,$$

assume that for all admissible ζ' one has $\mathcal{R}[w_{\zeta'}] \le 1 - \varepsilon$ with some fixed $\varepsilon > 0$. Then

$$\mathcal{R}[\bar{w}] \;\le\; 1 - \varepsilon \;+\; o_{T\to\infty}(1).$$

Proof. Note that H_L depends only on L and the zero configuration (via a), not on the short-interval parameter $\zeta = H/N$. Thus the ceiling bound $\mathcal{R}[w_{\zeta'}] \le 1 - \varepsilon$ is uniform over the family $\{w_{\zeta'}\}$ for all admissible ζ'.

Set $N(w) := \int |H_L|\, w$, $D_1(w) := \int |H_L|^2\, w$, $D_2(w) := \int w$. For $\bar{w} = \int K_r(\zeta')\, w_{\zeta'}\, d\zeta'$, linearity gives

$$N(\bar{w}) = \int K_r(\zeta')\, N(w_{\zeta'})\, d\zeta', \qquad D_j(\bar{w}) = \int K_r(\zeta')\, D_j(w_{\zeta'})\, d\zeta' \quad (j = 1, 2).$$

By Cauchy–Schwarz with respect to the probability measure $K_r(\zeta')\, d\zeta'$,

$$N(\bar{w})^2 \;\le\; \Big(\int K_r D_1(w_{\zeta'})\, d\zeta'\Big)\Big(\int K_r \frac{N(w_{\zeta'})^2}{D_1(w_{\zeta'})}\, d\zeta'\Big).$$

Divide by $D_1(\bar{w})\, D_2(\bar{w})$ and use $D_2(\bar{w}) = \int K_r D_2(w_{\zeta'})\, d\zeta'$:

$$\mathcal{R}[\bar{w}] \;\le\; \sup_{\zeta'} \mathcal{R}[w_{\zeta'}] \;\le\; 1 - \varepsilon.$$

Any $o(1)$ term comes only from restricting the average to a compact ζ'–support shrinking with T; this vanishes as $T \to \infty$. $\square$

Remark (ζ-independence of ceiling constants). The ceiling bound in Lemma 6 depends only on the global scale $L = \log T$ and the off–critical distance a, not on the short–interval parameter $\zeta = H/N$: the L^1/L^2 gap (Lemma 4) and Gram–Schur cross–term are uniform for all admissible ζ in a fixed compact range. Hence the hypothesis $\mathcal{R}[w_{\zeta'}] \le 1 - \varepsilon$ holds uniformly in ζ', justifying the application of the stability lemma to the Fejér–filtered statistic $X_T^{(r)}$.

Theorem 1 (Refined global moments for $X_T^{(r)}$)**.** *Let $X_T^{(r)}(m) := \mathcal{R}^{(2)}_{I,(r)}(H_L; m)$ for $m \in [T, 2T]$, with $H_L = ((\log\zeta)'' * v_L) * K_L$ and $L = \log T$. Assume N, Q are chosen as in Hypothesis 1, with $Q = T^{1/2-\nu}$, $H = T^{-1+\varepsilon}N$, and $\nu, \varepsilon > 0$ fixed small. Then there exists $\delta > 0$ such that*

$$\mathbb{E}_{[T,2T]}[X_T^{(r)}] = 1 + O((\log T)^{-1-\delta}), \qquad \mathrm{Var}_{[T,2T]}(X_T^{(r)}) = O((\log T)^{-1-\delta}). \tag{3.12}$$

Proof. Note. All polylogarithmic factors $(\log T)^C$ are absorbed into T^η with $\eta > 0$ arbitrarily small; choosing $\nu, \varepsilon > 0$ sufficiently small fixes $\delta > 0$ throughout. By Lemma 18, $\mathbb{E}[X_T^{(r)}]$ equals a diagonal term $1 + o(1)$ plus off–diagonal prime sums weighted by $\Phi_{2,L}$. Lemma 19 yields an analogous expansion for $\mathbb{E}[(X_T^{(r)})^2]$ with weight $\Phi_{4,L}$.

Second–moment off–diagonal (no partition needed). From Lemma 18, after m–averaging one has

$$\mathcal{E}_2(T) := \sum_{u\neq 0} \Phi_{2,L}(u;m) \ \ll_A \ T^{-A} \qquad (\forall A > 0),$$

coming directly from the compact frequency support of Φ_L and the decay of $\widehat{\Psi}(uT)$. This control is independent of the Type I/Type II decomposition.

Fourth–moment off–diagonal (case split by box scale). Decompose the quartic expansion into dyadic boxes $M \sim N$. There are two disjoint regimes.

(a) Small Boxes ($N \leq T^{1/2-\delta}$). For these boxes, the off–diagonal parameter satisfies $|U| \gtrsim 1/N^2$, implying $|UT| \gg T^{2\delta}$. As established in Step 4 of Lemma 19, the m–average yields a factor $\ll (T/N^2)^{-A} \ll T^{-2\delta A}$. Summing over dyadic boxes in this range yields a negligible contribution $O((\log T)^{-1-\delta})$.

(b) Balanced–large boxes $M \asymp N \geq T^{\theta_0}$. Here we do not rely on m–average decay. We invoke the dispersion/Kuznetsov reduction (Lemma 13), the on–shell uniformity of the Poisson–Fejér kernel (Lemma 16), and the moment–vanishing transform gain (Lemma 17) to obtain, uniformly over dyadic $R_2 \leq Q$,

$$\mathcal{J}_{\bullet}(\Phi^*, g; R_2) \ \ll_A (1+\bullet)^{-A} \left(\frac{H}{N}\right)^r,$$

which, combined with the spectral large–sieve bounds (Propositions 4–6), yields $O((\log T)^{-1-\delta})$ for the quartic off–diagonal in these boxes.

Combining (a) and (b) gives $O((\log T)^{-1-\delta})$ for the quartic off–diagonal overall.

Consequently, the off–diagonal contributions to both the mean and the second moment of $X_T^{(r)}$ (indeed T^{-A} for the second moment) are absorbed into $O((\log T)^{-1-\delta})$, while the diagonal pieces contribute $1 + o(1)$. Hence

$$\mathbb{E}[X_T^{(r)}] = 1 + O((\log T)^{-1-\delta}), \qquad \mathbb{E}[(X_T^{(r)})^2] = 1 + O((\log T)^{-1-\delta}),$$

and therefore

$$\mathrm{Var}(X_T^{(r)}) = \mathbb{E}[(X_T^{(r)})^2] - (\mathbb{E}[X_T^{(r)}])^2 = O((\log T)^{-1-\delta}),$$

which proves (3.12). $\square$

Parameter verification. To ensure all estimates in the Type II uniformity and transform–gain lemmas hold uniformly in T, we fix explicit admissible parameters satisfying

$$\nu < \tfrac{1}{3}, \qquad \varepsilon + \theta_0 - 3\nu \leq -\tfrac{1}{2}, \qquad r > \frac{1-2\nu}{1-\varepsilon}.$$

Parameter	Value	Meaning
ε	0.02	Short-interval exponent: $H = T^{-1+\varepsilon}N$
ν	0.2	Spectral cutoff exponent: $Q = T^{1/2-\nu}$
r	2	Fejér filter order (moment-vanishing)
θ_0	0.002	Minimum box size: $N \geq T^{\theta_0}$
L	$\log T$	Time-mollification scale

Table 1: Parameter choices for Type II analysis

Exponent verification (Type II boxes with $M \asymp N \sim T^{\theta}$):

The balanced Type II contribution has exponent

$$\text{Exponent} = 1 - 2\nu - r(1-\varepsilon) + \theta = 1 - 0.4 - 1.96 + \theta = -1.36 + \theta.$$

Box Type	θ Range	Exponent Range	Status
Small boxes	$[0.002, 0.2]$	$[-1.358, -1.16]$	✓ Negative
Mid-range	$[0.2, 0.5]$	$[-1.16, -0.86]$	✓ Negative
Worst case (balanced)	$\theta = 0.5$	$\mathbf{-0.86}$	✓ **Strong saving**

Table 2: Exponent verification across dyadic boxes

Conclusion: All Type II boxes contribute $\ll T^{-0.86}(\log T)^C$, giving strong power saving in variance. The parameter choices above meet all required inequalities with comfortable margins. We emphasize the role separation: m–average decay controls boxes with $N \leq T^{1/2-\delta}$ via $(T/N^2)^{-A}$ (as in Lemma 19), while the $(H/N)^r$–gain neutralizes the spectral Q^2 loss in the balanced–large Type II boxes $M \asymp N \geq T^{\theta_0}$.

With these choices one has

$$\frac{H^{1/2}d^{3/2}}{L^2} \ll 1, \qquad (H/N)^r \ll Q^{-2},$$

for $H = T^{-1+\varepsilon}N$, $N \geq T^{\theta_0}$, $Q = T^{1/2-\nu}$, and $L = \log T$. Hence all implied constants in Lemmas 16–17 are uniform in T, and the bounds

$$|S(\xi)| \ll (H/d)(\log T)^C, \qquad \widehat{\Psi}(UT) \ll (H/N)^r,$$

hold with the stated power savings.

Proposition 1 (Local high–density floor in any unit block)**.** *Let* $X_T^{(r)}(m) := \mathcal{R}^{(2)}_{I,(r)}(H_L; m)$ *for* $m \in [T, 2T]$*. Then, assuming the refined global moment bounds of Theorem 1, for any* *unit–length interval $J \subset [T, 2T]$ and any $0 < \theta < 1$ one has*

$$\frac{1}{|J|}\ \mathrm{meas}\Big\{m \in J:\ X_T^{(r)}(m)\ \geq\ 1 - \theta(\log T)^{-1/2}\Big\}\ \geq\ 1 - o(1).$$

Proof. Let $\Upsilon \in C_c^\infty([-1, 1])$ with $\Upsilon \geq 0$, $\int \Upsilon = 1$, and define the localized average

$$\mathbb{E}_J[f]\ :=\ \frac{1}{|J|}\int_{\mathbb{R}} f(m)\,\Upsilon\Big(\frac{m - m_J}{|J|}\Big)\,dm,$$

where m_J is the midpoint of J. Since $X_T^{(r)}$ is bandlimited in m to width $\ll \log T$, convolution with a fixed Υ preserves moment bounds up to $(1 + o(1))$ factors. Thus by Theorem 1,

$$\mathbb{E}_J[X_T^{(r)}]\ =\ 1 + O((\log T)^{-1-\delta}), \qquad \mathrm{Var}_J(X_T^{(r)})\ =\ O((\log T)^{-1-\delta}).$$

Let $Y(m) := 1 - X_T^{(r)}(m) \geq 0$. Then $\mathbb{E}_J[Y] = O((\log T)^{-1-\delta})$ and $\mathbb{E}_J[Y^2] = \mathrm{Var}_J(X_T^{(r)}) + (\mathbb{E}_J[Y])^2 = O((\log T)^{-1-\delta})$. By Chebyshev's inequality,

$$\frac{1}{|J|}\ \mathrm{meas}\Big\{m \in J:\ X_T^{(r)}(m) < 1 - \theta(\log T)^{-1/2}\Big\}\ =\ \frac{1}{|J|}\ \mathrm{meas}\Big\{Y(m) > \theta(\log T)^{-1/2}\Big\}.$$

$$\frac{1}{|J|}\ \mathrm{meas}\Big\{Y(m) > \theta(\log T)^{-1/2}\Big\}\ \leq\ \frac{\mathbb{E}_J[Y^2]}{\theta^2\,(\log T)^{-1}}\ \ll\ (\log T)^{-\delta}.$$

which tends to 0 as $T \to \infty$. This proves the claim. $\square$

Corollary 2 (Contradiction in aligned block)**.** *Assume an off–critical zero $\rho_0 = \sigma_0 + i\gamma_0$ exists with multiplicity $m \geq 1$ and $a = \frac{1}{2} - \sigma_0 > 0$. Let $\mathcal{I}$ be a block of unit length centered at γ_0. Then for sufficiently large T, the bounds of Theorem 1 and Proposition 1 (for $X_T^{(r)}$) contradict the ceiling bound of Lemma 6.*

Proof. Set $T = \gamma_0$ and take $J = \mathcal{I}$. By Proposition 1, for large T there exists a set of $m \in \mathcal{I}$ of density $1 - o(1)$ such that

$$X_T^{(r)}(m)\ \geq\ 1 - \eta(\log T)^{-1/2}, \qquad 0 < \eta < 1.$$

On the other hand, Lemma 6 shows that for all m aligned with γ_0,

$$X_T^{(r)}(m)\ \leq\ 1 - \varepsilon'(a, m) + o(1),$$

with $\varepsilon'(a, m) \asymp a > 0$ independent of T. For T large, since $(\log T)^{-1/2} < \varepsilon'(a, m)/2$, these bounds are incompatible. Hence the existence of an off–critical zero leads to a contradiction.

$\square$

Synthesis (finitely many zeros). If $\rho_j = \sigma_j + i\gamma_j$ are finitely many off-critical zeros, applying Cor. 2 with $T = \gamma_j$ yields a contradiction in each aligned block. Thus no such zeros exist.

Note on Prime-Side Derivations. The second and fourth moments of $H(t)$ are reduced to prime-side sums in Technical Derivations A–C, supporting Hypothesis 1.

Theorem 2 (The Riemann Hypothesis)**.** *No nontrivial zero of $\zeta(s)$ lies off the critical line* $\mathrm{Re}(s) = 1/2$*.*

Proof. Assume an off–critical zero exists. For any such zero $\rho = \sigma + i\gamma$ with $a = \frac{1}{2} - \sigma > 0$, apply Corollary 2 at $T = \gamma$: the local floor from Theorem 1 and Proposition 1 contradicts the energy–tax ceiling from Lemma 6 on the aligned block. Since this holds for each off–critical zero, none can exist. Hence all nontrivial zeros satisfy $\mathrm{Re}(s) = \frac{1}{2}$. $\square$

3.2 The Main Hypothesis

Hypothesis 1 (Short–Interval BDH with Smooth Weights)**.** *Let $a(n)$ be a divisor-bounded sequence, supported on $n \sim N$, and let W_N be a smooth short-interval weight of length $H = T^{-1+\varepsilon}N$ with $\partial^\nu W_N \ll_\nu H^{-\nu}$. Then there exists $\beta > 0$ such that*

$$\sum_{q \le Q} \sum_{\substack{b \pmod q \\ (b,q)=1}} \left| \sum_{\substack{n \sim N \\ n \equiv b(q)}} a(n)\, W_N\Big(\frac{n-N}{H}\Big) \;-\; \frac{1}{\varphi(q)} \sum_{n \sim N} a(n)\, W_N\Big(\frac{n-N}{H}\Big) \right|^2 \;\ll\; (\log T)^{-\beta}\, HN,$$

uniformly for $Q \le T^{1/2-\varepsilon/4}$.

3.3 Verification of Hypothesis 1 for Type I Sums

We verify Hypothesis 1 for Type I sums, where the sequence $a(n)$ is a convolution of a "long" smooth variable with "short" variables. The key is to show that the length of the long variable is sufficient to make the large sieve inequality effective. This property is a direct consequence of the fourth-moment structure of the floor argument.

Lemma 8 (Product–length constraint from the fourth moment)**.** *Let $H(t) = ((\log\zeta)'' * v_L)(t)(3.2)$ with $L = \log T$, and write H on the critical line by Mellin inversion and the*

Dirichlet–series for $(\log\zeta)''$ *as a short Dirichlet polynomial of effective length* $X = T^{1+o(1)}$*:*

$$H(t) \;=\; \sum_{n\asymp X} \frac{b(n)}{n^{1/2+it}}\, U\Big(\frac{n}{X}\Big) \;+\; O_A(T^{-A}) \qquad (\forall A>0),$$

where $b(n) = \Lambda(n)\log n \ll (\log n)^2$ *and* $U \in \mathcal{S}(\mathbb{R}_{\ge 0})$ *depends only on* v_L *and the fixed* t*–window. Then, in the fourth–moment expansion of*

$$\int_T^{2T} |H(t)|^4\, dt,$$

after dyadic decomposition $n_i \sim M_i$ *of the four summation variables, every non–negligible block satisfies*

$$M_1\, M_2\, M_3\, M_4 \;\ll\; T^{2+o(1)}.$$

Proof. Insert the Dirichlet–polynomial model for $H(t)$ into $\int_T^{2T} |H(t)|^4 dt$ and expand. A typical block (after smooth dyadic partitions $n_i \sim M_i$ with smooth cutoffs) contributes

$$\sum_{n_1\sim M_1} \cdots \sum_{n_4\sim M_4} \frac{b(n_1)b(n_2)b(n_3)b(n_4)}{(n_1n_2n_3n_4)^{1/2}}\, U\Big(\frac{n_1}{X}\Big)\cdots U\Big(\frac{n_4}{X}\Big) \int_T^{2T} e\Big(t\,\Delta(n_\bullet)\Big)\, dt,$$

where $\Delta(n_\bullet) = \frac{1}{2\pi}\log\frac{n_1n_3}{n_2n_4}$. By the standard estimate

$$\int_T^{2T} e\big(t\,\Delta\big)\, dt \;\ll\; \min\Big(T, \frac{1}{|\Delta|}\Big),$$

non–negligible contribution requires $|\Delta(n_\bullet)| \ll 1/T$, i.e.

$$\Big|\log\frac{n_1n_3}{n_2n_4}\Big| \;\ll\; \frac{1}{T} \qquad \Longrightarrow \qquad \Big|\frac{n_1n_3}{n_2n_4} - 1\Big| \;\ll\; \frac{1}{T}.$$

Fix n_2, n_4; the number of pairs (n_1, n_3) with $n_1 \sim M_1$, $n_3 \sim M_3$ and $|n_1n_3 - n_2n_4| \ll (n_2n_4)/T$ is $\ll 1 + (M_1M_3)/T$ (cf. [IK2004, §9.3, Lem. 9.4]). Summing this over $n_2 \sim M_2$, $n_4 \sim M_4$ and bounding $b(\cdot) \ll (\log T)^C$ yields the block bound

$$\ll\; T\,(\log T)^C\, \frac{(M_1M_2M_3M_4)^{1/2}}{T} \Big(1 + \frac{M_1M_3}{T}\Big)^{1/2} \Big(1 + \frac{M_2M_4}{T}\Big)^{1/2}.$$

Thus a block is negligible unless *both* $M_1M_3 \ll T^{1+o(1)}$ and $M_2M_4 \ll T^{1+o(1)}$. Multiplying these two constraints gives the claim:

$$M_1M_2M_3M_4 \;\ll\; T^{2+o(1)}.$$

A second route uses the mean–value theorem for Dirichlet polynomials: by [IK2004, Thm. 9.1],

$$\int_T^{2T} \Big| \sum_{n \sim M} a(n) n^{-it} \Big|^4 dt \;\ll\; (T + M^2)\,(\log T)^C \Big(\sum_{n \sim M} |a(n)|^2 \Big)^2.$$

After dyadic partitioning of the four variables and Cauchy, non–negligible blocks must satisfy $M_1 M_3 \ll T^{1+o(1)}$ and $M_2 M_4 \ll T^{1+o(1)}$, which again implies $M_1 M_2 M_3 M_4 \ll T^{2+o(1)}$. □

Dyadic scale bookkeeping. The global Mellin smoothing with $L = \log T$ produces a single smoothed Dirichlet polynomial for $H(t)$ of effective length $X = T^{1+o(1)}$, which we use only to derive the product–length constraint above. The fourth–moment analysis is then carried out dyadically in boxes $M \sim N$ with $N \le X$. All estimates for log–gaps, m–averaging, and the Type I/II routing are performed on the local scale N of the current box.

Lemma 9 (Type I long side from the product constraint)**.** *Assume a decomposition into four variables with dyadic lengths M_i arises from the fourth–moment expansion above, and suppose a* Type I *block is identified by having three short factors $M_i \le T^{\nu}$ for some fixed $0 < \nu < 1/3$. Then the remaining long side N satisfies*

$$N \;\ge\; T^{1+\nu'} \qquad \text{for some fixed } \;\nu' = 1 - 3\nu \; > 0.$$

Proof. By Lemma 8, non–negligible blocks satisfy

$$N \cdot M_1 M_2 M_3 \;\asymp\; M_1 M_2 M_3 M_4 \;\ll\; T^{2+o(1)}.$$

Under the Type I hypothesis $M_j \le T^{\nu}$ for three indices j, we obtain

$$N \;\gg\; \frac{T^{2+o(1)}}{T^{3\nu}} \;=\; T^{\,2-3\nu+o(1)}.$$

Since $\nu < 1/3$, $2 - 3\nu > 1$. Writing $2 - 3\nu = 1 + \nu'$, we get $N \ge T^{\,1+\nu'}$ for some fixed $\nu' > 0$ (up to the harmless $o(1)$ absorbed by raising ν' slightly). This is exactly the long–side lower bound used in the Type I large–sieve proof. □

We now provide the full proof of the Type I dispersion estimate.

Fejér two–parameter weight. Recall from Section 3 that $v_L(u) = L^{-1}v(u/L)$ and $w_L = v_L * v_L$ with $L = \log T$. We will use the associated two–parameter off–diagonal weight

$$W_L(m,n) \; := \; \int_{\mathbb{R}} v_L(u-m)\, v_L(u-n)\, du \; = \; (v_L * v_L)(m-n) \; = \; w_L(m-n), \qquad (3.13)$$

which satisfies $W_L(m,n) = W_L(n,m) \geq 0$ and $\int_{\mathbb{R}} W_L(m,n)\, dn = 1$ for each fixed m. This is the Fejér–induced coupling used throughout the Type I/II analyses.

Proposition 2 (Two-parameter smoothed short-BDH for Type I sums)**.** *Let $a(n)$ be a Type I sequence supported on $n \sim N$, i.e.*

$$a(n) \; = \; \sum_{m\sim M} \alpha_m \sum_{\substack{r\sim R\\ mr=n}} \beta_r, \qquad \sum_{m\sim M} |\alpha_m|^2 \; \ll \; M(\log T)^B, \quad \sum_{r\sim R} |\beta_r|^2 \; \ll \; R(\log T)^B,$$

with divisor–bounded α_m, β_r and $MR \asymp N$. Let $W_N \in C_c^\infty$ be a short–interval weight of length $H = T^{-1+\varepsilon}N$ with $\partial^\nu W_N \ll_\nu H^{-\nu}$, and let $W_L(m,n)$ be the Fejér–induced two–parameter weight obeying (3.13) *with $L = \log T$. Set $Q = T^{1/2-\nu}$ with small fixed $\nu, \varepsilon > 0$. Assume the* Type I regime

$$R = \frac{N}{M} \; \leq \; T^\nu \qquad \textit{and hence} \qquad M \; \geq \; T^{1+\nu'} \;\; \textit{for some } \nu' > 0,$$

as guaranteed by Lemma 8 and Lemma 9. Then, for any fixed $\beta > 0$,

$$\sum_{q\leq Q} \sum_{\substack{b \pmod q\\ (b,q)=1}} \Bigg| \sum_{\substack{n\sim N\\ n\equiv b\,(q)}} a(n)\, W_L(m,n)\, W_N\Big(\frac{n-N}{H}\Big) - \frac{1}{\varphi(q)} \sum_{n\sim N} a(n)\, W_L(m,n)\, W_N\Big(\frac{n-N}{H}\Big)\Bigg|^2 \; \ll \; (\log T)^{-\beta}\, HN,$$

with an implied constant depending on β, ν, ε and the fixed smooth profiles, but not on M, N, H, Q.

Proof. Write the progression variance in characters (orthogonality):

$$\mathcal{V}_{\mathrm{I}}(M,N;Q) = \sum_{q\leq Q} \frac{1}{\varphi(q)} \sum_{\substack{\chi \pmod q\\ \chi\neq\chi_0}} \Bigg| \sum_{n\sim N} a(n)\, W_L(\cdot,n)\, W_N(n)\, \chi(n) \Bigg|^2.$$

Apply the multiplicative large sieve with smooth weight on n:

$$\sum_{q\le Q}\frac{1}{\varphi(q)}\sum_{\chi\pmod q}\Big|\sum_n c_n\,\chi(n)\Big|^2 \;\ll\; (Q^2+H)\sum_n |c_n|^2,$$

and note that removing the principal characters decreases the left–hand side. With

$$c_n \;:=\; a(n)\,W_L(\cdot,n)\,W_N\Big(\frac{n-N}{H}\Big)\cdot \mathbf{1}_{n\sim N},$$

we obtain

$$\mathcal{V}_{\mathrm{I}}(M,N;Q) \;\ll\; (Q^2+H)\sum_{n\sim N}|c_n|^2. \tag{3.14}$$

Bounding the coefficient energy. The sum to be bounded is $\sum_{n\sim N}|c_n|^2$, where $c_n=a(n)W_L(\cdot,n)W_N(n)$. Since $|W_L|\le 1$ and $|W_N|\le 1$, we have $|c_n|^2\le|a(n)|^2$ for n in the support of W_N. The weight W_N is supported on a short interval of length H. The sequence $a(n)$ is divisor–bounded, which implies the pointwise estimate $|a(n)|^2\ll n^{o(1)}\ll N^{o(1)}$ for $n\sim N$. The sum is therefore over at most H integers, each of size $N^{o(1)}$, giving

$$\sum_{n\sim N}|c_n|^2 \;\ll\; H\cdot N^{o(1)} \;\ll\; H(\log T)^C. \tag{3.15}$$

Conclusion. Insert (3.15) into (3.14):

$$\mathcal{V}_{\mathrm{I}}(M,N;Q) \;\ll\; (Q^2+H)\,H\,(\log T)^C.$$

Normalize by HN:

$$\frac{\mathcal{V}_{\mathrm{I}}(M,N;Q)}{HN} \;\ll\; (\log T)^C\left(\frac{H}{N}+\frac{Q^2}{N}\right).$$

By definition $H/N=T^{-1+\varepsilon}$, and by the Type I length constraint we have $N\ge T^{1+\nu'}$. Since $Q=T^{1/2-\nu}$, we get

$$\frac{Q^2}{N} \;\le\; \frac{T^{1-2\nu}}{T^{1+\nu'}} \;=\; T^{-(2\nu+\nu')}.$$

Thus both H/N and Q^2/N are polynomially small in T. Hence

$$\frac{\mathcal{V}_{\mathrm{I}}(M,N;Q)}{HN} \;\ll\; (\log T)^{-\beta},$$

for any fixed $\beta>0$ (absorbing polylog factors into the saving). This proves the proposition.

$\square$

Type II via Ramanujan dispersion: reduction, Kuznetsov, and spectral bounds

We work on balanced dyadic boxes with $M \asymp N \gg T^{\theta}$ ($\theta > 0$ fixed), where "balanced" means M and N are in the same dyadic range, i.e., $M/2 \leq N \leq 2M$ (as opposed to unbalanced boxes where one variable is much larger than the other).

Type I/Type II partition and threshold. In the Heath–Brown decomposition underlying the fourth–moment expansion, each dyadic box (M, N) satisfies the product–length constraint

$$M_1 M_2 M_3 M_4 \ \ll\ T^{2+o(1)} \qquad \text{(Lemma 8)}.$$

Fix a small constant $\theta_0 > 0$ (for instance $\theta_0 = \nu'/10$, where ν' is from Lemma 9), and route boxes as follows:

- If $M \asymp N \geq T^{\theta_0}$ (i.e. balanced and large), classify the block as *Type II.*
- Otherwise, treat the block as *Type I.*

Justification of full coverage. The product constraint together with Lemma 9 ensures that any block not in the balanced–large regime must contain a long smooth variable: if three of the four dyadic factors in the fourth–moment decomposition satisfy $M_i \leq T^{\nu}$ for some $0 < \nu < 1/3$, then the remaining side obeys

$$N \ \geq\ T^{1+\nu'} \qquad (\nu' = 1 - 3\nu > 0),$$

placing the block within the hypotheses of the Type I large–sieve estimate (Proposition 2). Consequently, an apparently "balanced but small" configuration ($M \asymp N \leq T^{\theta_0}$) cannot occur as an isolated case: such terms arise only as components of a longer decomposition that necessarily includes a long side. Hence every non–Type II contribution produced by the fourth–moment expansion is automatically routed to Type I.

Conclusion. The Type II analysis below applies uniformly for $M \asymp N \geq T^{\theta_0}$. All remaining cases are absorbed by the Type I range through the long–side constraint, so the partition covers all possibilities with no "small–θ" gap. In Theorem 1 and subsequent arguments, all references to Type II implicitly assume this partition.

For concreteness, we fix $\theta_0 = \nu'/10$ throughout.

Why dispersion and Kuznetsov. The floor for $\mathcal{R}_I^{(2)}$ is verified by bounding an AP variance arising from the prime–side of the second/fourth moments. Ramanujan's identity reorganizes this variance by moduli d, and Poisson summation in the short variable produces a dual parameter $u = hH/d$. Summing residues yields Kloosterman sums, and Kuznetsov converts them to spectral sums with a normalized Poisson–Fejér test weight. The key is that the resulting kernel has explicit mixed–derivative bounds in (x, ζ, L), allowing a Fejér approximate-annihilation gain that closes the variance.

Short-interval parameter and local averaging. Let $\zeta := H/N \in (0, \zeta_0]$ be the short–interval parameter. We fix a nonnegative Fejér–type kernel K_r supported on $|\zeta' - \zeta| \ll H/N$, normalized so that $\int K_r = 1$ and with vanishing moments up to order $r-1$. All filtering in ζ below is performed by convolution with K_r.

definition 3 (Moment–vanishing Fejér kernel filter)**.** *Let $K_r : \mathbb{R} \to \mathbb{R}_{\geq 0}$ be a smooth, nonnegative kernel with compact support of diameter $\ll H/N$, normalized so that $\int_{\mathbb{R}} K_r(\zeta)\, d\zeta = 1$, and with vanishing moments*

$$\int_{\mathbb{R}} \zeta^k K_r(\zeta)\, d\zeta = 0 \qquad (1 \leq k \leq r-1).$$

For a function $F(\zeta)$, its filtered version is the convolution

$$F^{(r)}(\zeta) \;:=\; (F * K_r)(\zeta) = \int_{\mathbb{R}} F(\zeta - \zeta')\, K_r(\zeta')\, d\zeta'.$$

Example 1 (Concrete Fejér kernel for $r = 2$)**.** *Let $\delta := H/N$. Define the smooth even bump*

$$K_2(\zeta) := \frac{1}{Z_\delta} \exp\Big(- \frac{1}{1 - (2\zeta/\delta)^2} \Big) \mathbf{1}_{\{|\zeta| < \delta/2\}}, \qquad Z_\delta := \int_{-\delta/2}^{\delta/2} \exp\Big(- \frac{1}{1 - (2u/\delta)^2} \Big)\, du.$$

Then $K_2 \in C_c^\infty(\mathbb{R})$, $K_2 \geq 0$, $\int_{\mathbb{R}} K_2 = 1$, and (being even) $\int_{\mathbb{R}} \zeta\, K_2(\zeta)\, d\zeta = 0$. Thus K_2 satisfies Definition 3 with $r = 2$ and support diameter $\delta = H/N$.

Remark. In this manuscript we fix $r = 2$. Any smooth nonnegative Fejér–type kernel with unit mass and vanishing first moment (e.g. K_2 above) yields the full $(H/N)^2$ gain required to cancel the Q^2 spectral mass; no higher–order moment vanishing is needed.

Lemma 10 (Diagonal–Spectral Identity for the Constant Term)**.** *Let $\mathcal{V}(M, N; Q)$ denote the short–interval variance appearing after Ramanujan dispersion, defined with the main term (the $h = 0$ Poisson mode) already subtracted:*

$$\mathcal{V} = \sum_{q \leq Q} \sum_{b \pmod q}^{*} \Big| \Sigma(m, n; q, b) \;-\; \mathit{MainTerm}_{h=0} \Big|^2.$$

After Poisson summation in the short variable, let $\Phi(y;\zeta)$ be the spectral test weight arising from the $h \neq 0$ frequencies. Then the following identity holds:

$$\boxed{\text{The } \zeta\text{-independent term } \Phi(y;0) \text{ equals the arithmetic diagonal subtracted in the definition of } \mathcal{V}.}$$

Consequently, the off–diagonal *spectral weight entering the Type II analysis is precisely*

$$\Phi_{\text{off}}(y;\zeta) \;:=\; \Phi(y;\zeta) - \Phi(y;0),$$

and satisfies a Taylor expansion beginning at order ζ^1:

$$\Phi_{\text{off}}(y;\zeta) = \zeta\, \partial_\zeta \Phi(y;0) + \frac{\zeta^2}{2} \partial_\zeta^2 \Phi(y;0) + \cdots .$$

Proof. In the Ramanujan–Poisson decomposition of the arithmetic progression sum

$$\Sigma(m,n;q,b) = \sum_{\substack{m\sim M,\, n\sim N \\ mn\equiv b \pmod q}} \alpha_m \beta_n\, W_L(m,n)\, W_N\Big(\frac{n-N}{H}\Big),$$

introduce the Ramanujan identity $c_q(h) = \sum_{d|(q,h)} \mu(q/d)\, d$ and reorganize the variance $\mathcal{V}$ as a weighted sum over frequencies $h \in \mathbb{Z}$. This yields the Poisson expansion

$$\mathcal{V} = \sum_{h\in\mathbb{Z}} \Big(\mathcal{C}(h) \;-\; \delta_{h=0}\, \mathcal{C}(0)\Big), \tag{4.11.1}$$

where $\mathcal{C}(h)$ is the contribution from the h-th Poisson mode and $\delta_{h=0}$ is the Kronecker delta.

By definition of the variance in Hypothesis 1, the term $\mathcal{C}(0)$ is exactly the *arithmetic diagonal* (the mean value over residue classes) and is subtracted before entering any off–diagonal analysis. Thus the effective variance is

$$\mathcal{V}_{\text{off}} = \sum_{\substack{h\in\mathbb{Z} \\ h\neq 0}} \mathcal{C}(h). \tag{4.11.2}$$

Now examine the spectral expansion arising from the $h \neq 0$ modes. For each fixed $d \asymp R_2$ in the Ramanujan reduction, the normalized Poisson–Fejér weight $\mathcal{W}_d(x;\zeta,L)$ depends smoothly on $\zeta = H/N$, and the Kuznetsov test function

$$\Phi(y;\zeta) = y\, \mathcal{W}_d\Big(\big(\tfrac{y}{4\pi}\big)^2; \zeta, L\Big)$$

is its Mellin transform.

Let $\Phi(\cdot;0)$ denote the value at $\zeta=0$. Setting $\zeta=0$ corresponds to collapsing the short–interval weight W_N to its integral, which in the Poisson decomposition kills all modes $h\neq 0$ and preserves exactly the $h=0$ contribution. Therefore,

$$\Phi(y;0) \quad \text{arises solely from} \quad h=0, \tag{4.11.3}$$

and its spectral expansion is the spectral representation of the diagonal term $\mathcal{C}(0)$.

Since $\mathcal{C}(0)$ has already been subtracted in the definition of the variance (cf. (4.11.1)), it follows that the weight that governs the off–diagonal ($h\neq 0$) spectral sums is precisely

$$\Phi_{\text{off}}(y;\zeta)=\Phi(y;\zeta)-\Phi(y;0). \tag{4.11.4}$$

Because $\Phi(\cdot;\zeta)$ is C^r-smooth in ζ uniformly in y (Lemma 16), we may apply Taylor's theorem at $\zeta=0$:

$$\Phi(y;\zeta)=\Phi(y;0)+\zeta\,\partial_\zeta\Phi(y;0)+\frac{\zeta^2}{2}\partial_\zeta^2\Phi(y;0)+\cdots.$$

Subtracting the diagonal component $\Phi(y;0)$ leaves

$$\Phi_{\text{off}}(y;\zeta)=\zeta\,\partial_\zeta\Phi(y;0)+\frac{\zeta^2}{2}\partial_\zeta^2\Phi(y;0)+\cdots. \tag{4.11.5}$$

This shows two things:

- The Taylor series of the off–diagonal spectral weight has no constant term.
- Its smallest-degree term is of order ζ^1.

Finally, subtracting the linear Taylor term (equivalently, replacing $\widehat{\Phi}$ by $\widehat{\Phi}^{(2)}_{\text{off}}$) removes the ζ-linear part in (4.11.5) and leaves an $O(\zeta^2)=O((H/N)^2)$ remainder. (Convolution with K_2 preserves the linear term; the removal is effected by this de-biasing.)

This proves that the constant term $\Phi(\cdot;0)$ contributes only to the removed diagonal, and that the de-biased filter produces the full $(H/N)^2$ gain for the off–diagonal Type II terms. $\square$

Lemma 11 (Moment vanishing and analytic cancellation for the Fejér filter)**.** *Let $K_r:\mathbb{R}\to\mathbb{R}_{\geq 0}$ be a nonnegative Fejér–type kernel with unit mass $\int_{\mathbb{R}}K_r(u)\,du=1$, compact support of diameter $\asymp H/N$, and vanishing moments*

$$\int_{\mathbb{R}}u^kK_r(u)\,du=0 \qquad (1\leq k\leq r-1).$$

Then:

(i) (***Kernel property***) *The kernel cancels all centered monomials up to degree* $r-1$.

(ii) (***Analytic consequence***) *For every* $F \in C^r(\mathbb{R})$,

$$(F * K_r)(\zeta) \;=\; F(\zeta) \;+\; O\Big(\|F^{(r)}\|_\infty \,(H/N)^r\Big).$$

In particular, when $r = 2$, *the degree–*≤ 1 *Taylor polynomial of* F *is preserved and the remainder is* $O(\|F''\|_\infty (H/N)^2)$.

Proof. Expand $F(\zeta - u)$ in a Taylor series about ζ:

$$F(\zeta - u) = \sum_{k=0}^{r-1} \frac{F^{(k)}(\zeta)}{k!}(-u)^k + R_r(\zeta, u),$$

with remainder $|R_r(\zeta, u)| \leq \|F^{(r)}\|_\infty |u|^r / r!$. Convolving against K_r gives

$$(F * K_r)(\zeta) = \sum_{k=0}^{r-1} \frac{F^{(k)}(\zeta)}{k!}(-1)^k \int_{\mathbb{R}} u^k K_r(u)\, du + \int_{\mathbb{R}} R_r(\zeta, u) K_r(u)\, du.$$

By the moment conditions, the integrals vanish for $1 \leq k \leq r-1$. The $k = 0$ term yields $F(\zeta)$. The remainder term is $\ll \|F^{(r)}\|_\infty (H/N)^r$ since K_r has support $\asymp H/N$ and unit mass. This proves both (i) and (ii). □

Application to the dispersion/Kuznetsov step. Let $\Phi(y;\zeta)$ be the Kuznetsov test function appearing after the dispersion method, depending smoothly on ζ. Write its $(r-1)$-st order Taylor expansion at $\zeta = 0$:

$$\Phi(y;\zeta) \;=\; \Phi_{\mathrm{Tay}}(y;\zeta) \;+\; \Phi^*(y;\zeta), \qquad \Phi_{\mathrm{Tay}}(y;\zeta) := \sum_{k=0}^{r-1} \frac{\zeta^k}{k!}\, \partial_\zeta^k \Phi(y;0).$$

Define the *filtered* test function by convolution with K_r:

$$\Phi^{(r)}(y;\zeta) := (\Phi(y;\cdot) * K_r)(\zeta).$$

Because K_r has unit mass and $\int u^k K_r(u)\, du = 0$ for $1 \leq k \leq r-1$, convolution preserves the degree–$< r$ Taylor polynomial:

$$(\Phi(y;\cdot) * K_r)(\zeta) = \Phi_{\mathrm{Tay}}(y;\zeta) + O\big((H/N)^r\big).$$

To force a genuine short–interval gain on the off–diagonal we pass to the de-biased remainder $\Phi^*(y;\zeta) := \Phi(y;\zeta) - \Phi_{\mathrm{Tay}}(y;\zeta)$, whose Mellin transform obeys (3.19) and is $O((H/N)^r)$. The constant (ζ-independent) term belongs to the diagonal by Lemma 10.

Lemma 12 (Off–diagonal sees only the gain–enhanced piece)**.** *Apply the dispersion method and then replace $\Phi(y;\zeta)$ by the de-biased remainder $\Phi^*(y;\zeta) := \Phi(y;\zeta) - \Phi_{\mathrm{Tay}}(y;\zeta)$. Equivalently, at the Mellin level replace $\widehat{\Phi}(s;\zeta)$ by $\widehat{\Phi}^{(r)}_{\mathrm{off}}(s;\zeta) := \widehat{\Phi}(s;\zeta) - P_{r-1}(s;\zeta)$ from Lemma 14. Then, by* (3.19)*, the off–diagonal depends only on this remainder and is $O((H/N)^r)$ uniformly in the spectral parameters; the constant term is diagonal.*

Proof. By Lemma 14, $\widehat{\Phi}(s;\zeta) = P_{r-1}(s;\zeta) + O((H/N)^r(1+|\tau|)^{-A})$. Subtracting P_{r-1} removes the ζ-polynomial of degree $< r$; the surviving transform is $O((H/N)^r)$, and the $k = 0$ term is diagonal by Lemma 10. □

Filtered variance. Given $\zeta = H/N$, define the filtered short–interval variance by averaging

$$\mathcal{V}^{(r)}(M,N;Q) \ := \ \int K_r(\zeta')\,\mathcal{V}(M,N;Q;\zeta-\zeta')\,d\zeta',$$

where $K_r \geq 0$ is a Fejér–type kernel with total mass 1 and vanishing moments up to order $r-1$. This filtering suppresses the Taylor polynomial part to order $O((H/N)^r)$. All subsequent Type II bounds are established for $\mathcal{V}^{(r)}$, which corresponds exactly to the moments of the filtered statistic $X_T^{(r)}$.

Scope of filtering. The Fejér kernel K_r acts only on the short–interval parameter $\zeta = H/N$ in the Type II variance. It does *not* modify the time–windowed observable used in the ceiling argument. Lemma 4 therefore applies to the same Fejér window $w_L^m(t)$ with $L = \log T$, and the stability lemma concerns convex averaging of weights, not a redefinition of F.

Lemma 13 (Ramanujan dispersion to Kloosterman prototype)**.** *Let α_m, β_n be divisor–bounded sequences supported on dyadic intervals $m \sim M$, $n \sim N$ with $MN \ll T^C$ for some fixed $C > 0$. Let $W_L(m,n)$ be the Fejér–induced two–variable weight obeying the bandlimit* (3.13)*, and let $W_N \in C_c^\infty(\mathbb{R})$ be a fixed bump with unit–size support and $\partial_y^j W_N(y) \ll_j 1$, always* *applied as* $W_N\big(\frac{n-N}{H}\big)$ *(or* $W_N\big(\frac{u-x}{H}\big)$ *on the Poisson/Kuznetsov side). Then, for any* $A>0$,

$$\mathcal{V}(M,N;Q) \;:=\; \sum_{q\le Q}\ \sideset{}{^*}\sum_{b \bmod q}\Bigg|\sum_{\substack{m\sim M,\ n\sim N\\ mn\equiv b\ (q)}}\alpha_m\beta_n\,W_L(m,n)\,W_N\Big(\frac{n-N}{H}\Big) \\ -\ \frac{1}{\varphi(q)}\sum_{m\sim M}\sum_{n\sim N}\alpha_m\beta_n\,W_L(m,n)\,W_N\Big(\frac{n-N}{H}\Big)\ \Bigg|^2$$

satisfies

$$\mathcal{V}(M,N;Q)\ \ll\ (\log T)^C\sum_{\substack{R_2\ dyadic\\ R_2\le Q}}\ \sum_{d\asymp R_2}\big|\mathcal{K}(M,N;d)\big|\ +\ O_A\big((\log T)^{-A}MN\big), \tag{3.16}$$

where each $\mathcal{K}(M,N;d)$ *is a Kloosterman–prototype sum of the form*

$$\mathcal{K}(M,N;d)\ :=\ \sum_{m\sim M}\sum_{n\sim N}\alpha_m\,\beta_n\ S(m,n;d)\ \mathcal{W}_d\Big(\tfrac{mn}{d^2};\,\zeta,\,L\Big), \tag{3.17}$$

with $\zeta=H/N$, $S(m,n;d)$ *the classical Kloosterman sum, and test weight*

$$\mathcal{W}_d(x;\zeta,L)\ =\ \int_{\mathbb{R}}W_N\Big(\tfrac{u-x}{H}\Big)\,B_d(u;\zeta,L)\,K_L(u)\,\chi_d(u)\,du, \tag{3.18}$$

where:

- $W_N\in C_c^\infty(\mathbb{R})$ *is a fixed short–interval profile with unit–size support and* $\partial_y^j W_N(y)\ll_j 1$,
- $B_d(\cdot;\zeta,L)\in C^\infty$ *satisfies* $\partial_\zeta^k B_d\ll_k H^{-k}(\log T)^{C_k}$, $\partial_u^\ell B_d\ll_\ell(\log T)^{C_\ell}$,
- $K_L\in\mathcal{S}(\mathbb{R})$ *is a Fejér cap with Fourier support* $|\xi|\le c/L$ *and* $\|K_L^{(\ell)}\|_\infty\ll_\ell L^{-\ell}$,
- $\chi_d\in C_c^\infty(\mathbb{R})$ *localizes* $u\asymp 1$, *uniformly for* $d\asymp R_2$.

uniformly for $d\asymp R_2\le Q$, $x>0$, *and* $\zeta=H/N\in(0,\zeta_0]$.

Proof. 1) Variance expansion with Ramanujan sums. Expand $\mathcal{V}(M,N;Q)$ and insert the identity $c_q(h)=\sum_{d\mid(q,h)}\mu(q/d)\,d$. Swapping the q– and d–sums gives (3.16) up to a factor $(\log T)^C$ from the q–average.

2) Residue decomposition. Fix d and write $n=r+dt$. Insert a smooth cutoff $\omega(t/(H/d))\in C_c^\infty$ to truncate $|t|\ll H/d$. The weight now factors as $\beta_{r+dt}W_L(m,r+dt)W_N(r+dt)\omega(t/(H/d))$.

3) Poisson in the short variable. Apply Poisson to the t–sum:

$$\sum_{t\in\mathbb{Z}} \Xi_{m,r}(t) = \frac{H}{d} \sum_{h\in\mathbb{Z}} \widehat{\Xi}_{m,r}\Big(\tfrac{hH}{d}\Big)\, e\Big(-\tfrac{hr}{d}\Big),$$

where $u := hH/d$. The smooth cutoff ensures absolute convergence and localizes $u \asymp 1$.

4) Summing over r. The sum over r mod d collapses the phases to classical Kloosterman sums $S(m,h;d)$. This produces the prototype structure (3.17) with weight $\mathcal{W}_d$.

5) Structure of the weight. Express $\widehat{W_N}(u)$ by inverse Fourier; the variable x enters as a translation $W_N((u-x)/H)$. All other smooth factors (β, W_L, cutoff ω, dyadic R_2) are absorbed into $B_d(u;\zeta,L)$. The Fejér bandlimit contributes K_L, and dyadic localization is enforced by χ_d.

□

Lemma 14 (Mellin remainder in the short-interval parameter)**.** *Let $\mathcal{W}_d(x;\zeta,L)$ be the weight function from the Type II reduction, whose uniform mixed-derivative bounds are established in Lemma 16. Let $\Phi(y;\zeta,L) = y\,\mathcal{W}_d\big((y/4\pi)^2;\zeta,L\big)$. Fix $\operatorname{Re} s = \sigma'$ and $r \in \mathbb{N}$. Then, uniformly in $\zeta \in (0,\zeta_0]$ and $s = \sigma' + i\tau$,*

$$\widehat{\Phi}(s;\zeta) \;=\; \sum_{m=0}^{r-1} \frac{\zeta^m}{m!}\, \partial_\zeta^m \widehat{\Phi}(s;0) \;+\; O\big((H/N)^r\,(1+|\tau|)^{-A}\big) \qquad (\forall A>0). \tag{3.19}$$

Definition (off–diagonal piece). Let

$$P_{r-1}(s;\zeta) := \sum_{m=0}^{r-1} \frac{\zeta^m}{m!}\, \partial_\zeta^m \widehat{\Phi}(s;0)$$

be the Taylor polynomial of degree $< r$. Here $\partial_\zeta^m \widehat{\Phi}(s;0)$ is the right–limit $\lim_{\zeta\to 0^+} \partial_\zeta^m \widehat{\Phi}(s;\zeta)$, which exists by the uniform bounds in Lemma 16. Define the off–diagonal filtered transform by

$$\widehat{\Phi}^{(r)}_{\text{off}}(s;\zeta) := \widehat{\Phi}(s;\zeta) - P_{r-1}(s;\zeta).$$

Then, by (3.19),

$$\widehat{\Phi}^{(r)}_{\text{off}}(s;\zeta) = O\big((H/N)^r(1+|\tau|)^{-A}\big),$$

and this is the quantity that enters the Type II off–diagonal variance.

Proof. *The uniform mixed-derivative bounds for $\mathcal{W}_d$ established in Lemma 16 justify* *differentiating under the Mellin integral. For any $r \in \mathbb{N}$ and $\theta \in [0,1]$,*

$$\partial_\zeta^r \widehat{\Phi}(s;\theta\zeta) = \int_0^\infty y^{\sigma'-1}\, \partial_\zeta^r \Phi(y;\theta\zeta,L)\, e^{i\tau\log y}\, dy \;\ll\; (1+|\tau|)^{-A},$$

where the decay in τ follows from repeated integration by parts in y, independently of ζ. Taylor's theorem with integral remainder yields

$$\widehat{\Phi}(s;\zeta) = \sum_{m=0}^{r-1} \frac{\zeta^m}{m!}\, \partial_\zeta^m \widehat{\Phi}(s;0) \;+\; \frac{\zeta^r}{(r-1)!} \int_0^1 (1-\theta)^{r-1}\, \partial_\zeta^r \widehat{\Phi}(s;\theta\zeta)\, d\theta.$$

Using the bound on $\partial_\zeta^r \widehat{\Phi}$ gives

$$\widehat{\Phi}(s;\zeta) = \sum_{m<r} \frac{\zeta^m}{m!}\, \partial_\zeta^m \widehat{\Phi}(s;0) \;+\; O\big(\zeta^r (1+|\tau|)^{-A}\big).$$

Since $\zeta = H/N$, this is exactly (3.19).

Lemma 15 (Twofold discrete Abel summation)**.** *Let a_t be supported on $\{1,\dots,H\}$ and set $S(\xi) := \sum_{t=1}^{H} a_t\, e(-\xi t)$ with $e(x) = e^{2\pi i x}$. Define first and second differences $\Delta a_t := a_t - a_{t-1}$ and $\Delta^2 a_t := \Delta(\Delta a_t)$ (with $a_0 = a_{H+1} = 0$).*

Then for every $\xi \in \mathbb{R} \setminus \mathbb{Z}$,

$$S(\xi) \;=\; \frac{1}{\big(e(-\xi)-1\big)^2} \sum_{t=1}^{H} \Delta^2 a_t\, e(-\xi t) \;+\; \mathcal{B}_1(\xi) + \mathcal{B}_2(\xi),$$

where the boundary terms satisfy

$$|\mathcal{B}_1(\xi)| + |\mathcal{B}_2(\xi)| \;\ll\; \frac{1}{|\xi|}\big(|\Delta a_1| + |\Delta a_{H+1}|\big) \;+\; \frac{1}{|\xi|^2}\big(|a_1| + |a_H|\big).$$

Consequently, by Cauchy–Schwarz and $\#\{t\} \asymp H$,

$$|S(\xi)| \;\le\; \frac{1}{(2\pi|\xi|)^2}\, \|\Delta^2 a\|_{\ell^2([1,H])}\, \sqrt{H} \;+\; O\bigg(\frac{\|a\|_{\ell^2}}{|\xi|^2}\bigg).$$

Proof. Let $A(t) := \sum_{u \le t} a_u$ with $A(0) = 0$. One discrete summation by parts gives

$$S(\xi) = \sum_{t=1}^{H} a_t e(-\xi t) = (e(-\xi)-1) \sum_{t=1}^{H-1} A(t)\, e(-\xi t) + a_H e(-\xi H).$$

Apply summation by parts once more to the A–sum, introducing $B(t) := \sum_{u \le t} A(u)$ (so that $\Delta B(t) = A(t)$ and $\Delta^2 B(t) = a_t$):

$$\sum_{t=1}^{H-1} A(t)\, e(-\xi t) = (e(-\xi) - 1) \sum_{t=1}^{H-2} B(t)\, e(-\xi t) + A(H-1)e(-\xi(H-1)).$$

Combining, we obtain

$$S(\xi) = (e(-\xi) - 1)^2 \sum_{t=1}^{H-2} B(t)\, e(-\xi t) + \mathcal{B}_1(\xi) + \mathcal{B}_2(\xi),$$

where the boundary terms $\mathcal{B}_1, \mathcal{B}_2$ are as in the statement. Since $e(-\xi) - 1 = -2\pi i\xi\, \omega(\xi)$ with $|\omega(\xi)| \asymp 1$ for $|\xi| \le 1/2$,

$$S(\xi) = (2\pi i\xi)^2 \omega(\xi)^2 \sum_{t=1}^{H-2} B(t)\, e(-\xi t) + \mathcal{B}_1(\xi) + \mathcal{B}_2(\xi).$$

Finally, using $\Delta^2 B(t) = a_t$ and reversing the previous steps yields

$$\sum_{t=1}^{H-2} B(t)\, e(-\xi t) = \frac{1}{\big(e(-\xi) - 1\big)^2} \sum_{t=1}^{H} \Delta^2 a_t\, e(-\xi t),$$

which proves the main identity and the boundary bounds. The ℓ^2 consequence follows by Cauchy–Schwarz with $\#\{t\} \asymp H$. $\square$

Lemma 16 (Uniformity across dyadic moduli)**.** *Let R_2 be dyadic with $R_2 \le Q$, and fix a dyadic block of moduli $d \asymp R_2$. For the normalized Poisson–Fejér weight*

$$\mathcal{W}_d(x; \zeta, L) \;=\; \int_{\mathbb{R}} W_N\Big(\frac{u-x}{H}\Big)\, B_d(u; \zeta, L)\, K_L(u)\, \chi_d(u)\, du,$$

arising in the Type II reduction, the mixed derivatives satisfy, for all $j, k, \ell \ge 0$,

$$\sup_{d \asymp R_2} \sup_{x>0} \big|\partial_x^j\, \partial_\zeta^k\, \partial_L^\ell\, \mathcal{W}_d(x; \zeta, L)\big| \;\ll_{j,k,\ell}\; H^{-j}\, H^{-k}\, L^{-\ell}\, \frac{H^2}{R_2}\, (\log T)^{C_{j,k,\ell}}, \tag{3.20}$$

uniformly over all d in the dyadic shell $d \in [R_2/2, 2R_2]$, $x > 0$, and $\zeta = H/N \in (0, \zeta_0]$.

Proof. **(A) Dependence on ζ.** The short parameter $\zeta = H/N$ enters only through the factor $W_N((u-x)/H)$. Here N is regarded as fixed when differentiating in ζ, so $H = \zeta N$ and each ∂_ζ incurs a factor of H^{-1} by the chain rule through $W_N((u-x)/H)$. This explains the factor H^{-k} in (3.20). *(In applications we later specialize to $\zeta = T^{-1+\varepsilon}$; the differentiation is carried out before this specialization.)*

(B) Reduction to a bound for B_d. Differentiating under the u–integral gives

$$\partial_x^j \partial_\zeta^k \partial_L^\ell \mathcal{W}_d = \int_{\mathbb{R}} \Big(\partial_x^j W_N\big(\tfrac{u-x}{H}\big)\Big)\, B_d(u;\zeta,L)\, \Big(\partial_L^\ell K_L(u)\Big)\, \chi_d(u)\, du.$$

Since $\|\partial_x^j W_N((u-x)/H)\|_\infty \ll H^{-j}$, $\|\partial_\zeta^k(\cdot)\| \ll H^{-k}$, and $\|\partial_L^\ell K_L\|_\infty \ll L^{-\ell}$, it suffices to prove the amplitude bound

$$\sup_{d \asymp R_2} \sup_{u \asymp 1} |B_d(u;\zeta,L)| \;\ll\; \frac{H^2}{R_2}\,(\log T)^C, \tag{3.21}$$

for then inserting the derivative costs into the compact u–integral immediately yields (3.20).

(C) Structure of B_d and its Fourier side. From the Type II setup,

$$B_d(u;\zeta,L) = \frac{H}{d} \sum_{r \,(\bmod d)} e\Big(-\frac{hr}{d}\Big)\widehat{\Xi}_{m,r}\Big(\frac{hH}{d}\Big), \qquad u = \frac{hH}{d},$$

where

$$\Xi_{m,r}(t) = \beta_{r+dt}\,\mathsf{S}_m(r+dt), \qquad \mathsf{S}_m(n) = W_L(m,n)\,W_N(n)\,\omega\Big(\frac{t}{H/d}\Big),$$

and $t = (n-r)/d$ is supported on $|t| \ll H/d$. Divisor–boundedness gives $\sum_t |\beta_{r+dt}|^2 \ll (H/d)(\log T)^C$.

(D) Fourier–Plancherel estimate for discrete differences. Let $a_t := \beta_{r+dt}\,\mathsf{S}_m(r+dt)$ and $\widehat{a}(\eta) = \sum_t a_t e(-\eta t)$. For $k=2$,

$$\|\Delta^2 a\|_{\ell_t^2} = \big\|\,(e^{-2\pi i\eta}-1)^2\widehat{a}(\eta)\,\big\|_{L_\eta^2} \;\ll\; \sup_{|\eta| \ll d/H + d/L} |e^{-2\pi i\eta}-1|^2\,\|\widehat{a}\|_{L_\eta^2}.$$

By Young and Plancherel, $\|\widehat{a}\|_{L^2} \le \|\widehat{\beta}\|_{L^2}\|\widehat{\mathsf{S}}\|_{L^1} = \|\beta\|_{\ell^2}\|\widehat{\mathsf{S}}\|_{L^1}$. For the smooth bump S_m, standard Paley–Wiener/Nikolskii bounds give $\|\widehat{\mathsf{S}}\|_{L^1} \ll 1$ and $\operatorname{supp}\widehat{\mathsf{S}} \subset \{|\eta| \ll d/H + d/L\}$. Hence

$$|e^{-2\pi i\eta}-1|^2 \ll (d/H + d/L)^2 \ll (d/H)^2 + (d/L)^2,$$

and with $\|\beta\|_{\ell^2} \ll (H/d)^{1/2}(\log T)^C$, we obtain

$$\|\Delta^2 a\|_{\ell_t^2} \;\ll\; \Big(\frac{d^2}{H^2} + \frac{d^2}{L^2}\Big)\Big(\frac{H}{d}\Big)^{1/2}(\log T)^C. \tag{3.22}$$

(E) Twofold Abel summation and explicit power bookkeeping. For any $\xi \in \mathbb{R}\backslash\mathbb{Z}$, Lemma 15 and Cauchy–Schwarz give

$$|S(\xi)| = \Big|\sum_t a_t e(-\xi t)\Big| \ \le\ \frac{1}{(2\pi|\xi|)^2}\,\|\Delta^2 a\|_{\ell^1_t} + O\Big(\frac{\|a\|_{\ell^2}}{|\xi|^2}\Big) \ \ll\ \frac{1}{|\xi|^2}\,\|\Delta^2 a\|_{\ell^2_t}\,(H/d)^{1/2},$$

since $\|\Delta^2 a\|_{\ell^1_t} \le (\#\text{support})^{1/2}\,\|\Delta^2 a\|_{\ell^2_t}$ and $\#\{t\} \asymp H/d$. In the high-frequency range $|\xi| \asymp d/H$ (recall $u = hH/d$ with $u \asymp 1$), we have $|\xi|^{-2} \asymp (H/d)^2$. Thus, inserting (3.22),

$$\begin{aligned}|S(\xi)| &\ll (H/d)^2\ \Big[\Big(\frac{d^2}{H^2}+\frac{d^2}{L^2}\Big)\Big(\frac{H}{d}\Big)^{1/2}(\log T)^C\Big]\ \Big(\frac{H}{d}\Big)^{1/2} \\ &= \Big((H/d)^2\frac{d^2}{H^2}+(H/d)^2\frac{d^2}{L^2}\Big)\ \frac{H}{d}\,(\log T)^C \\ &= \Big(1+\frac{H^2}{L^2}\Big)\ \frac{H}{d}\,(\log T)^C\ \ll\ \frac{H}{d}\,(\log T)^C.\end{aligned}$$

Therefore the discrete Fourier sum is bounded by $|S(\xi)| \ll (H/d)(\log T)^C$. Finally,

$$B_d(u;\zeta,L) = \frac{H}{d}\sum_{r \bmod d} e(-hr/d)\,S(\xi), \qquad \xi = \tfrac{ud}{H}.$$

The geometric sum over r has modulus $\le d$, so

$$|B_d(u;\zeta,L)|\ \ll\ \frac{H}{d}\cdot d\cdot|S(\xi)|\ \ll\ \frac{H}{d}\cdot d\cdot\Big(\frac{H}{d}(\log T)^C\Big)\ =\ \frac{H^2}{d}\,(\log T)^C, \tag{3.23}$$

which is exactly the amplitude bound (3.21) for all d in the dyadic shell $d \in [R_2/2, 2R_2]$.

(F) Conclusion and parameter bookkeeping. Substituting (3.23) into the u–integral for $\mathcal{W}_d$ and re-inserting the derivative costs from (B) gives (3.20). Moreover, because $\chi_d(u)$ localizes $u \asymp 1$, we evaluate $S(\xi)$ on–shell[1] at $|\xi| = |ud/H| \asymp d/H$. the d/L Fourier lobe would contribute only for $|\xi| \asymp d/L$ (equivalently $u \asymp H/L \ll 1$), which lies outside the $u \asymp 1$ support of χ_d. Thus the d/L lobe does not contribute at the sampled frequency. This yields $|S(\xi)| \ll (H/d)(\log T)^C$ and hence $|B_d(u)| \ll (H^2/d)(\log T)^C$, as claimed.

□

[1]The terminology "on-shell" refers to the natural frequency scale $\xi \sim d/H$ where the Poisson kernel is concentrated; "off-shell" refers to frequencies outside this band. This language is borrowed from dispersion-relation analysis in physics.

Kuznetsov skeleton with a short–interval transform gain

For each dyadic $R_2 \le Q$, aggregate the Kloosterman–prototype sums produced by Lemma 13 at moduli $d \asymp R_2$ into

$$\mathcal{K}(M,N;R_2) \ := \ \sum_{\substack{d\ge 1\\ d\asymp R_2}} \ \sum_{m\sim M}\sum_{n\sim N} \alpha_m\,\beta_n\ S(m,n;d)\ \mathcal{W}_d\Big(\tfrac{mn}{d^2};\ \tfrac{H}{N},\ L\Big),$$

where $\mathcal{W}_d$ is smooth and satisfies the uniform mixed–derivative bounds of Lemma 16. Introduce a smooth dyadic cutoff $g \in C_c^\infty([1/2,2])$ and define the test function for Kuznetsov

$$\Phi(y) \ := \ y\,\mathcal{W}\Big(\Big(\tfrac{y}{4\pi}\Big)^2;\ \tfrac{H}{N},\ L\Big) \ \in C_c^\infty((0,\infty)), \tag{3.24}$$

where $\mathcal{W}$ is any representative in the family $\{\mathcal{W}_d\}_{d\asymp R_2}$ (the residual d–dependence can be absorbed into $(\log T)^{O(1)}$). Then, writing c for d,

$$\mathcal{K}(M,N;R_2) = \sum_{m\sim M}\sum_{n\sim N}\ \sum_{c\ge 1} \frac{S(m,n;c)}{c}\ g\Big(\frac{c}{R_2}\Big)\ \Phi\Big(\frac{4\pi\sqrt{mn}}{c}\Big)\ +\ O_A\big((\log T)^{-A}\big) \tag{3.25}$$

(for any fixed $A > 0$), where the error comes from the negligible tails in the Poisson/partition steps of Lemma 13.

Proposition 3 (Kuznetsov trace formula with dyadic level)**.** *Let $g \in C_c^\infty([1/2,2])$ and $\Phi \in C_c^\infty((0,\infty))$. For positive integers m,n one has*

$$\sum_{c\ge 1} \frac{S(m,n;c)}{c}\, g\Big(\frac{c}{R_2}\Big)\, \Phi\Big(\frac{4\pi\sqrt{mn}}{c}\Big) = \mathcal{H}_{m,n}[\Phi,g;R_2]\ +\ \mathcal{M}_{m,n}[\Phi,g;R_2]\ +\ \mathcal{E}_{m,n}[\Phi,g;R_2], \tag{3.26}$$

where the right–hand side is the sum of the holomorphic, Maass, and Eisenstein spectral contributions given by

$$\mathcal{H}_{m,n}[\Phi,g;R_2] = \sum_{\substack{k\ge 2\\ k\ \text{even}}}\ \sum_{f\in\mathcal{B}_k} i^k\,\mathcal{J}_k(\Phi,g;R_2)\,\rho_f(m)\,\overline{\rho_f(n)}, \tag{3.27}$$

$$\mathcal{M}_{m,n}[\Phi,g;R_2] = \sum_{f\in\mathcal{B}}\ \frac{1}{\cosh(\pi t_f)}\,\mathcal{J}^{\pm}_{t_f}(\Phi,g;R_2)\,\rho_f(m)\,\overline{\rho_f(n)}, \tag{3.28}$$

$$\mathcal{E}_{m,n}[\Phi,g;R_2] = \frac{1}{4\pi}\int_{\mathbb{R}}\ \frac{1}{\cosh(\pi t)}\,\mathcal{J}^{\pm}_{t}(\Phi,g;R_2)\,\rho_t(m)\,\overline{\rho_t(n)}\,dt, \tag{3.29}$$

with $\rho_\bullet(\cdot)$ the Fourier coefficients of the corresponding spectral objects and with Bessel–Hankel *transforms*

$$\mathcal{J}_k(\Phi, g; R_2) = \int_0^\infty \Phi(y)\, J_{k-1}(y)\, \frac{dy}{y}, \qquad \mathcal{J}_t^{\pm}(\Phi, g; R_2) = \int_0^\infty \Phi(y) \Big(J_{\pm 2it}(y) - J_{\mp 2it}(y) \Big) \frac{dy}{y}, \tag{3.30}$$

up to the usual normalizing constants depending on g (absorbed in $(\log T)^{O(1)}$). Moreover, for every $A > 0$,

$$\mathcal{J}_k(\Phi, g; R_2) \ll_A (1+k)^{-A}, \qquad \mathcal{J}_t^{\pm}(\Phi, g; R_2) \ll_A (1+|t|)^{-A}. \tag{3.31}$$

Proof. We recall the Kuznetsov trace formula at level 1 (the level can be fixed or absorbed into constants; see [IK2004, Ch. 16]). Let $W : (0,\infty) \times (0,\infty) \to \mathbb{C}$ be a smooth test kernel. The formula asserts that for positive integers m, n,

$$\sum_{c \ge 1} \frac{S(m,n;c)}{c} W\Big(\frac{4\pi\sqrt{mn}}{c}, c\Big) = \mathcal{H}_{m,n}[W] + \mathcal{M}_{m,n}[W] + \mathcal{E}_{m,n}[W], \tag{3.32}$$

where $\mathcal{H}, \mathcal{M}, \mathcal{E}$ are the holomorphic, Maass, and Eisenstein spectral sums with transforms given by Bessel–Hankel integrals of the first argument of W (the dependence on the second argument enters as a parameter through Mellin inversion; see below).

We choose the separable test

$$W\Big(\frac{4\pi\sqrt{mn}}{c}, c\Big) := g\Big(\frac{c}{R_2}\Big)\, \Phi\Big(\frac{4\pi\sqrt{mn}}{c}\Big),$$

where $g \in C_c^\infty([1/2, 2])$ is compactly supported and $\Phi \in C_c^\infty((0,\infty))$; this matches the left–hand side of (3.26). To bring this into the standard framework of (3.32), one notes that the dependence on c through $g(c/R_2)$ can be inserted by Mellin inversion:

$$g\Big(\frac{c}{R_2}\Big) = \frac{1}{2\pi i} \int_{(\sigma)} \widehat{g}(s) \Big(\frac{c}{R_2}\Big)^{-s} ds, \qquad \widehat{g}(s) = \int_0^\infty g(u) u^{s-1}\, du,$$

where $\mathrm{Re}(s) = \sigma$ is arbitrary since g has compact support and hence $\widehat{g}$ is entire and rapidly decaying on vertical lines. Inserting this into (3.32) and interchanging sum and integral (justified by absolute convergence from the rapid decay of $\widehat{g}$ and the compact support of Φ), we obtain

$$\sum_{c \ge 1} \frac{S(m,n;c)}{c} g\Big(\frac{c}{R_2}\Big) \Phi\Big(\frac{4\pi\sqrt{mn}}{c}\Big) = \frac{1}{2\pi i} \int_{(\sigma)} \widehat{g}(s)\, R_2^s \sum_{c \ge 1} \frac{S(m,n;c)}{c^{1+s}}\, \Phi\Big(\frac{4\pi\sqrt{mn}}{c}\Big)\, ds.$$

Inserting (3.32) with $W(y, c) = c^{-(1+s)}\Phi(y)$ yields spectral terms whose Bessel transforms depend on s; averaging in s with weight $\widehat{g}(s)R_2^s$ defines

$$\mathcal{J}_\bullet(\Phi, g; R_2) := \frac{1}{2\pi i}\int_{(\sigma)} \widehat{g}(s)\, R_2^s\, \mathcal{J}_\bullet(\Phi_s)\, ds.$$

By this definition, all subsequent occurrences of $\mathcal{J}_\bullet(\Phi, g; R_2)$ refer to these s–averaged transforms, so the s–dependence has been absorbed into the weights; the bounds (3.31) follow from the rapid decay of $\widehat{g}$ and the compact support of Φ.

Applying (3.32) to the inner c–sum with kernel $c^{-(1+s)}\Phi(4\pi\sqrt{mn}/c)$ yields

$$\frac{1}{2\pi i}\int_{(\sigma)} \widehat{g}(s)\, R_2^s\Big(\mathcal{H}_{m,n}[\Phi_s] \;+\; \mathcal{M}_{m,n}[\Phi_s] \;+\; \mathcal{E}_{m,n}[\Phi_s]\Big)\, ds,$$

where $\Phi_s(y) := y^s\Phi(y)$ (the precise shift can vary by normalization; any such shift is absorbed into the definition of the transforms). Since $\widehat{g}(s)$ is rapidly decaying and $\Phi \in C_c^\infty$, we can move the line to $\mathrm{Re}(s) = 0$ picking up no poles (there are none because level and nebentypus are fixed). Evaluating the s–integral formally gives (3.26) with transforms as in (3.30) and overall normalizing constants depending only on g and absorbed into $(\log T)^{O(1)}$.

Finally, the classical decay bounds (3.31) follow by repeated integration by parts in (3.30): since $\Phi \in C_c^\infty((0,\infty))$, for every $A > 0$ one has $\int_0^\infty \Phi(y)J_\nu(y)\, dy/y \ll_A (1+|\nu|)^{-A}$ uniformly in $\nu \in \{k-1, \pm 2it\}$. This is standard; see, e.g., [IK2004, Lem. 16.2]. □

Lemma 17 (Short–interval transform gain)**.** ***Uniform Taylor–Bessel interchange.*** *Before proving the main estimate we note that, by Lemma 16, for all integers $j, k, \ell \geq 0$,*

$$\sup_{\zeta,\, x>0} x^j\left|\partial_x^j\partial_\zeta^k\partial_L^\ell\Phi(x;\zeta, L)\right| \ll H^{-j}H^{-k}L^{-\ell}\,\Xi(x),$$

where Ξ is integrable against every Bessel kernel: $\int_0^\infty \Xi(y)|J_\nu(y)|\,\frac{dy}{y} \ll 1$ uniformly in ν. Hence the Taylor expansion $\Phi(y;\zeta) = \sum_{m<r}\frac{\zeta^m}{m!}\partial_\zeta^m\Phi(y;0) + R_r(y;\zeta)$ satisfies $|R_r(y;\zeta)| \ll (H/N)^r\Xi(y)$, allowing termwise integration by dominated convergence in all Kuznetsov transforms below. Convolution in ζ with K_r preserves the degree–$<r$ polynomial part; subtracting $\Phi_{\mathrm{Tay}}(y;\zeta)$ removes it and leaves an $O((H/N)^r)$ remainder.

Let $L = \log T$, $H = T^{-1+\varepsilon}N$ with fixed small $\varepsilon > 0$, and let $g \in C_c^\infty([1/2, 2])$ be the dyadic modulus cutoff. The following bounds hold uniformly for all $d \asymp R_2 \leq Q$. There exists a filtered Kuznetsov test function $\Phi^* \in C_c^\infty((0,\infty))$, supported where Φ in (3.24) is supported and with the same derivative bounds up to $(\log T)^{O(1)}$, such that for any fixed $A > 0$ and uniformly for dyadic $R_2 \leq Q$ one has

$$\mathcal{J}_k(\Phi^*, g; R_2) \;\ll_A\; (1+k)^{-A}\left(\frac{H}{N}\right)^r, \qquad \mathcal{J}_t^\pm(\Phi^*, g; R_2) \;\ll_A\; (1+|t|)^{-A}\left(\frac{H}{N}\right)^r, \tag{3.33}$$

for any chosen integer $r \geq 1$. Moreover, for all $a, b \in \mathbb{N}$,

$$\partial_{R_2}^a \, \partial_L^b \, \mathcal{J}_\bullet(\Phi^*, g; R_2) \;\ll_{a,b,A}\; R_2^{-a} \, L^{-b} \, (\log T)^{C_{a,b,A}} \, (1+\bullet)^{-A} \left(\frac{H}{N}\right)^r, \qquad \bullet \in \{k, t\}. \quad (3.34)$$

Proof. Write

$$\Phi(y;\zeta) = \Phi_{\mathrm{Tay}}(y;\zeta) + R_r(y;\zeta), \qquad \Phi_{\mathrm{Tay}}(y;\zeta) := \sum_{m<r} \frac{\zeta^m}{m!} \partial_\zeta^m \Phi(y;0).$$

Define the filtered, de-biased test function

$$\Phi^*(y;\zeta) := (\Phi(y;\cdot) * K_r)(\zeta) - \Phi_{\mathrm{Tay}}(y;\zeta) = (R_r(\,\cdot\,;\cdot) * K_r)(y,\zeta).$$

By Lemma 11, $|\Phi^*(y;\zeta)| \ll (H/N)^r \, \Xi(y)$, where Ξ satisfies

$$\int_0^\infty \Xi(y) |J_\nu(y)| \, \frac{dy}{y} \ll 1$$

uniformly in ν. Consequently,

$$\left|\mathcal{J}_k(\Phi^*, g; R_2)\right| = \left| \int_0^\infty \Phi^*(y;\zeta) \, J_{k-1}(y) \, \frac{dy}{y} \right| \ll (H/N)^r \int_0^\infty \Xi(y) \, |J_{k-1}(y)| \, \frac{dy}{y} \ll (H/N)^r,$$

and the same argument gives $\mathcal{J}_t^{\pm}(\Phi^*, g; R_2) \ll (H/N)^r$. The derivative bounds (3.34) follow by differentiating under the integral sign and using Lemma 16 together with the same domination by Ξ.

□

Corollary 3 (Type II variance bound with full gain)**.** *In the Type II range, the entire off–diagonal contribution to the variance is controlled with the $(H/N)^r$ gain by combining Lemmas 12–17 together with the spectral large–sieve bounds (Propositions 4–6). Consequently, the short–interval dispersion estimate stated in Hypothesis 1 holds with the indicated exponents.*

Remark 2 (Optimizing r). Since $H/N = T^{-1+\varepsilon}$, choosing r so that $(H/N)^r \ll Q^{-2}$ (e.g. $r > \frac{2(1/2-\nu)}{1-\varepsilon}$ when $Q = T^{1/2-\nu}$) ensures that the $(H/N)^r$ saving exactly cancels the Q^2 loss from the spectral large–sieve bounds in Propositions 4–6. Any fixed integer r satisfying this inequality suffices; in this section we take $r = 2$ in all proofs, and the optimization remark is only contextual.

Spectral large–sieve bounds: formal statements and proofs

We retain the notation of Proposition 3 and Lemma 17. In particular,

$$\mathcal{K}(M,N;R_2)=\sum_{m\sim M}\sum_{n\sim N}\sum_{c\geq 1}\frac{S(m,n;c)}{c}\,g\Big(\frac{c}{R_2}\Big)\,\Phi\Big(\frac{4\pi\sqrt{mn}}{c}\Big),$$

with $g\in C_c^\infty([1/2,2])$ and $\Phi\in C_c^\infty((0,\infty))$ built from $\mathcal{W}$ as in (3.24), and the transforms $\mathcal{J}_\bullet(\Phi,g;R_2)$ defined in (3.30). The short–interval transform gain is recorded in (3.33).

Proposition 4 (Spectral large–sieve bound: holomorphic channel)**.** *Let* $\mathcal{H}_{m,n}[\Phi,g;R_2]$ *be as in* (3.27)*. Then for any* $A>0$*,*

$$\Big|\sum_{m\sim M}\sum_{n\sim N}\alpha_m\beta_n\,\mathcal{H}_{m,n}[\Phi,g;R_2]\Big|\ \ll_A\ (R_2^2+M)^{\frac12}(R_2^2+N)^{\frac12}\,(\log T)^{C_A}\left(\frac{H}{N}\right)^r\|\alpha\|_2\,\|\beta\|_2,$$

uniformly for dyadic $R_2\leq Q$*. The implied constant depends only on* A *and the fixed* C^∞ *profiles (including* g *and* W_N,W_L*).*

Proof. By (3.27) and the triangle inequality,

$$\sum_{m,n}\alpha_m\beta_n\,\mathcal{H}_{m,n}=\sum_{\substack{k\geq 2\\ k\text{ even}}}\sum_{f\in\mathcal{B}_k}i^k\,\mathcal{J}_k(\Phi,g;R_2)\Big(\sum_{m\sim M}\alpha_m\rho_f(m)\Big)\overline{\Big(\sum_{n\sim N}\beta_n\rho_f(n)\Big)}.$$

Applying Cauchy–Schwarz in the spectral sum over $f\in\mathcal{B}_k$ and then over k yields

$$\Big|\sum_{m,n}\alpha_m\beta_n\,\mathcal{H}_{m,n}\Big|\ \leq\ \Big(\sum_k\ |\mathcal{J}_k|^2\sum_{f\in\mathcal{B}_k}\Big|\sum_{m\sim M}\alpha_m\rho_f(m)\Big|^2\Big)^{1/2}\Big(\sum_k\ \sum_{f\in\mathcal{B}_k}\Big|\sum_{n\sim N}\beta_n\rho_f(n)\Big|^2\Big)^{1/2}.$$

By the spectral large–sieve inequality for holomorphic cusp forms at fixed level (Iwaniec-Kowalski [IK2004, Thm. 16.5, p. 387]), for any $T\geq 1$,

$$\sum_{\substack{k\text{ even}\\ k\leq T}}\sum_{f\in\mathcal{B}_k}\Big|\sum_{m\sim M}\alpha_m\rho_f(m)\Big|^2\ \ll\ (M+T^2)\,(\log T)^C\,\|\alpha\|_2^2,$$

and similarly for the n–sum with β. In our application, the dyadic modulus cutoff $g(c/R_2)$ localizes the geometric side at $c\asymp R_2$; hence the spectral parameter effectively ranges up to $T\asymp R_2$ (the transforms outside that range decay rapidly by (3.31)). Using this with $T\asymp R_2$ and the bound $|\mathcal{J}_k|\ll_A(1+k)^{-A}\left(\frac{H}{N}\right)^r$ from (3.33) (the $\left(\frac{H}{N}\right)^r$ factor is uniform in k and

R_2), we get

$$\sum_k |\mathcal{J}_k|^2 \sum_{f\in\mathcal{B}_k} \Big|\sum_{m\sim M} \alpha_m \rho_f(m)\Big|^2 \ \ll\ \Big(\frac{H}{N}\Big)^{2r} (M+R_2^2)(\log T)^C \, \|\alpha\|_2^2,$$

and likewise

$$\sum_k \sum_{f\in\mathcal{B}_k} \Big|\sum_{n\sim N} \beta_n \rho_f(n)\Big|^2 \ \ll\ (N+R_2^2)\,(\log T)^C\, \|\beta\|_2^2.$$

Taking square roots yields the claimed bound. □

Proposition 5 (Spectral large–sieve bound: Maass channel)**.** *Let $\mathcal{M}_{m,n}[\Phi,g;R_2]$ be as in (3.28). Then for any $A>0$,*

$$\Big|\sum_{m\sim M}\sum_{n\sim N} \alpha_m\beta_n\, \mathcal{M}_{m,n}[\Phi,g;R_2]\Big| \ \ll_A\ (R_2^2+M)^{\frac12}(R_2^2+N)^{\frac12}\,(\log T)^{C_A}\Big(\frac{H}{N}\Big)^r \|\alpha\|_2\,\|\beta\|_2,$$

uniformly for dyadic $R_2 \le Q$.

Proof. Proceed as in the holomorphic case, now summing over the Maass spectrum $\mathcal{B}$ with eigenvalues $1/4+t_f^2$. Cauchy–Schwarz gives

$$\Big|\sum_{m,n} \alpha_m\beta_n\, \mathcal{M}_{m,n}\Big| \ \le\ \Big(\sum_{f\in\mathcal{B}} \frac{|\mathcal{J}_{t_f}^{\pm}|^2}{\cosh(\pi t_f)} \Big|\sum_m \alpha_m\rho_f(m)\Big|^2\Big)^{1/2}\Big(\sum_{f\in\mathcal{B}} \frac{1}{\cosh(\pi t_f)} \Big|\sum_n \beta_n\rho_f(n)\Big|^2\Big)^{1/2}.$$

By (3.33), $|\mathcal{J}_t^{\pm}| \ll_A (1+|t|)^{-A}\Big(\frac{H}{N}\Big)^r$. Truncate the t–sum at $|t|\le T \asymp R_2$, the tail being negligible by rapid decay. Then apply the Maass spectral large–sieve (Iwaniec-Kowalski [IK2004, Thm. 16.5, p. 387]): for $|t_f|\le T$,

$$\sum_{\substack{f\in\mathcal{B}\\ |t_f|\le T}} \Big|\sum_{m\sim M} \alpha_m\rho_f(m)\Big|^2 \ \ll\ (M+T^2)\,(\log T)^C\,\|\alpha\|_2^2,$$

and similarly for β. The claimed bound follows. □

Proposition 6 (Spectral large–sieve bound: Eisenstein channel)**.** *Let $\mathcal{E}_{m,n}[\Phi,g;R_2]$ be as in (3.29). Then for any $A>0$,*

$$\Big|\sum_{m\sim M}\sum_{n\sim N} \alpha_m\beta_n\, \mathcal{E}_{m,n}[\Phi,g;R_2]\Big| \ \ll_A\ (R_2^2+M)^{\frac12}(R_2^2+N)^{\frac12}\,(\log T)^{C_A}\Big(\frac{H}{N}\Big)^r \|\alpha\|_2\,\|\beta\|_2,$$

uniformly for dyadic $R_2 \le Q$.

Proof. Identical in spirit: Cauchy–Schwarz in $t \in \mathbb{R}$ with weight $1/\cosh(\pi t)$ and $\mathcal{J}_t^{\pm}$, truncate at $|t| \le T \asymp R_2$ using (3.33), and apply the continuous spectral large–sieve (Iwaniec-Kowalski [IK2004, Thm. 16.5, p. 387], continuous spectrum case):

$$\int_{|t|\le T} \Big| \sum_{m\sim M} \alpha_m \rho_t(m) \Big|^2 dt \ \ll\ (M+T^2)\,(\log T)^C\,\|\alpha\|_2^2,$$

and likewise for β. Combine as above. □

Corollary 4 (Fixed–modulus Kloosterman–prototype bound)**.** *Let $\mathcal{K}(M,N;R_2)$ be as in (3.25). Then for any $A>0$,*

$$|\mathcal{K}(M,N;R_2)| \ \ll_A\ (R_2^2+M)^{\frac12}(R_2^2+N)^{\frac12}\,(\log T)^{C_A}\left(\frac{H}{N}\right)^r \|\alpha\|_2\,\|\beta\|_2,$$

uniformly for dyadic $R_2 \le Q$.

Proof. Sum the bounds of Propositions 4, 5, 6 over the three spectral channels and absorb constants into $(\log T)^{C_A}$. □

Parameters at a glance. Recall $H/N = T^{-1+\varepsilon}$ and $Q = T^{1/2-\nu}$. Choose an integer $r \ge 1$ so that

$$\left(\frac{H}{N}\right)^r \ \le\ Q^{-2} \ =\ T^{-1+2\nu}.$$

For example, any $r > \frac{1-2\nu}{1-\varepsilon}$ suffices. With this choice, the $(H/N)^r$ factor from Lemma 17 neutralizes the Q^2 loss in the spectral large sieve. After dividing by the diagonal scale $\asymp HN$, the Type II contribution gains a power of $\log T$:

$$\mathcal{V}_{\mathrm{II}}(M,N) \ \ll\ (\log T)^{-\beta}\,HN.$$

Outcome. The Type II variance on a single balanced box obeys (3.16) with a *short–interval gain* $\left(\frac{H}{N}\right)^r$. This bound feeds directly into the final optimization: with $H = T^{-1+\varepsilon}N$ and $Q = T^{1/2-\nu}$, the $\left(\frac{H}{N}\right)^r$ factor compensates for the Q^2–terms so that, after dividing by the diagonal scale $\sim HN$, a log–power saving survives (for fixed small $\nu>0$), uniformly over all Type II boxes.

Lemma 18 (Prime-side second moment identity, refined)**.** *Let $H_L = ((\log\zeta)'' * v_L) * K_L$ with $L = \log T$, $w_L = v_L * v_L$, and $m \in [T, 2T]$. Then*

$$E_I(m) := \int_{\mathbb{R}} |H_L(t)|^2\, w_L^m(t)\,dt = \mathcal{M}_2(T;m) + \mathcal{Z}_2(T;m),$$

with explicit diagonal main term

$$\mathcal{M}_2(T;m) = \frac{1}{2\pi}\,\widehat{w}_L(0)\,\log\frac{T}{2\pi} + O(1),$$

and off–diagonal term

$$\mathcal{Z}_2(T;m) = \sum_p \sum_{k\geq 1} \frac{\log p}{p^{k/2}}\,\Phi_{2,L}(k\log p;m),$$

where $\Phi_{2,L}(u;m)$ is smooth, supported on $|u| \leq c/L$, and after m–averaging

$$\mathbb{E}_T^{(m)}[\Phi_{2,L}(u;m)] = \widehat{\Psi}(uT)\,B_L(u), \qquad \mathcal{E}_2(T) := \sum_p \sum_{k\geq 1} \frac{\log p}{p^{k/2}}\,\Phi_{2,L}(k\log p;m) \;\ll_A T^{-A}.$$

Proof. 1) Kernel. Define

$$\mathcal{K}_L(\eta,\xi) := \widehat{K}_L(\eta)\,\overline{\widehat{K}_L(\eta-\xi)}\,\widehat{K}_L(\xi),$$

supported on $|\eta|, |\eta-\xi|, |\xi| \leq 1/L$. Then

$$E_I(m) = \frac{1}{(2\pi)^2}\int_{|\xi|\leq 1/L}\int_{\mathbb{R}} \widehat{H}_L(\eta)\,\overline{\widehat{H}_L(\eta-\xi)}\,\mathcal{K}_L(\eta,\xi)\,e^{i\xi m}\,d\eta\,d\xi.$$

2) Splitting. Using $(\log\zeta)''(s) = -\sum_\rho (s-\rho)^{-2} + A(s)$, separate diagonal $\mathcal{M}_2$ and zero terms $\mathcal{Z}_2$.

3) Contour integral and decay. Define

$$\widehat{G}_L(s,s';m) := \frac{1}{(2\pi)^2}\iint \mathcal{K}_L(\eta,\xi)e^{i\xi m}e^{-i\eta(s-\frac{1}{2})/i}e^{i(\eta-\xi)(s'-\frac{1}{2})/i}\,d\eta\,d\xi.$$

Because $\mathcal{K}_L \in C_c^\infty$, repeated integration by parts shows $|\partial_s^a\partial_{s'}^b\widehat{G}_L(s,s';m)| \ll_{a,b,N} (1+|\operatorname{Im} s|+|\operatorname{Im} s'|)^{-N}$, allowing contour shifts. Moving $\operatorname{Re} s, \operatorname{Re} s'$ from $1/2+\epsilon$ to $1+\epsilon$ crosses only the pole at $s=1$.

4) Residue at $s=1$. Since $\zeta'/\zeta(s) \sim -1/(s-1)$, the double residue at $(1,1)$ yields

$$\mathcal{M}_2(T;m) = \frac{1}{2\pi}\,\widehat{w}_L(0)\,\log\frac{T}{2\pi} + O(1),$$

as $\widehat{w}_L(0) = \int w_L = 1$.

5) Prime–side form. On $\operatorname{Re} s > 1$, $\zeta'/\zeta(s) = -\sum_{n\geq 1}\Lambda(n)n^{-s}$. Insert, exchange sums/integrals, and invert Mellin transforms:

$$\mathcal{Z}_2(T;m) = \sum_p \sum_{k\geq 1} \frac{\log p}{p^{k/2}}\, \Phi_{2,L}(k\log p;m),$$

with

$$\Phi_{2,L}(u;m) = \frac{1}{(2\pi)^2}\int_{|\xi|\leq 1/L}\left(\int_{\mathbb{R}} e^{-i\eta u}\widehat{K}_L(\eta)\overline{\widehat{K}_L(\eta-\xi)}\,d\eta\right)\widehat{K}_L(\xi)\,e^{i\xi m}\,d\xi,$$

smooth and supported on $|u|\leq c/L$.

6) Averaging in m. Let $\Psi\in C_c^\infty([1,2])$ with $\int\Psi=1$ and define

$$\mathbb{E}_T^{(m)}[F] = \frac{1}{T}\int_{\mathbb{R}} F(m)\Psi(m/T)\,dm.$$

Then

$$\mathbb{E}_T^{(m)}[\Phi_{2,L}(u;m)] = \widehat{\Psi}(uT)\,B_L(u), \quad |B_L(u)|\ll 1,\ |u|\leq c/L.$$

For $u\neq 0$, $|\widehat{\Psi}(uT)|\ll_A (1+|u|T)^{-A}$, so

$$\mathcal{E}_2(T) := \sum_{u\neq 0}\Phi_{2,L}(u;m) \ \ll_A T^{-A},$$

a polynomial decay stronger than any log–power saving, since $|u|\leq c/L = O(\log T)$. This completes the proof. $\square$

Remark (Bilinear off–diagonals and the partition). The bilinear off–diagonal sums arising from the second moment are already controlled by the compact frequency support of Φ_L together with the m–average, yielding $\mathcal{E}_2(T)\ll T^{-A}$ for all $A>0$. Thus the Type I/II decomposition is *not* required for the second moment. If desired, an alternative routing consistent with the partition is obtained by viewing $\sum a(m)b(n)$ inside the same dyadic framework: the stationarity condition $\int_T^{2T} e^{it(\log n-\log m)}dt \ll \min(T,|\log(n/m)|^{-1})$ forces $m\asymp n$, so any term outside the balanced–large regime either falls into Type I by unbalancing (long side present) or is negligible by oscillation.

C. Fourth moment: prime-side formulation and m–average

Lemma 19 (Prime-side fourth moment identity, refined)**.** *Let $H_L = ((\log\zeta)''*v_L)*K_L$ with $L=\log T$, and $w_L = v_L * v_L$. Fix $m\in[T,2T]$. Then*

$$\int_{\mathbb{R}} |H_L(t)|^4\, w_L^m(t)\,dt \ = \ \mathcal{M}_4(T;m) \ + \ \mathcal{E}_4(T;m),$$

where the diagonal main term satisfies

$$\mathbb{E}_T^{(m)}[\mathcal{M}_4(T;m)] \;=\; \mathcal{M}_2(T)^2\,(1+o(1)),$$

and the off–diagonal term admits a prime–side expansion supported on $|U| \le c/L$ *with* $U = \log(n_1 n_3/n_2 n_4)$. *After* m*–smoothing one has, for every* $A > 0$,

$$\mathbb{E}_T^{(m)}[\Phi_{4,L}(U;m)] \;\ll_A\; (1+|UT|)^{-A}.$$

Consequently, for dyadic boxes with $N \le T^{1/2-\delta}$,

$$\mathbb{E}_T^{(m)}[\mathcal{E}_4(T;m)\big|_N] \;\ll_A\; T^{-A}.$$

Proof. We prove the stated fourth–moment identity and bounds for the spectrally-capped field H_L, with $w_L = v_L * v_L$, $w_L^m(t) = w_L(t-m)$, $L = \log T$, and $m \in [T, 2T]$.

1) Fourfold Plancherel and bandlimit. Let $\widehat{f}(\xi) = \int_{\mathbb{R}} f(t)e^{-2\pi i t\xi}\,dt$. With the spectral cap $\widehat{K}_L$ supported in $|\xi| \le 1/L$, write

$$\int_{\mathbb{R}} |H_L(t)|^4\, w_L^m(t)\,dt = \underset{|\eta_j|\le 1/L}{\int\!\cdots\!\int} \widehat{H}_L(\eta_1)\,\overline{\widehat{H}_L(\eta_2)}\,\widehat{H}_L(\eta_3)\,\overline{\widehat{H}_L(\eta_4)}\,\mathcal{K}_L^{(4)}(\eta_\bullet)\,e^{2\pi i(\eta_1-\eta_2+\eta_3-\eta_4)m}\,d\eta_1\,d\eta_2\,d\eta_3\,d\eta_4,$$

where the smooth kernel

$$\mathcal{K}_L^{(4)}(\eta_\bullet) := \widehat{K}_L(\eta_1)\overline{\widehat{K}_L(\eta_2)}\,\widehat{K}_L(\eta_3)\overline{\widehat{K}_L(\eta_4)}\,\widehat{w}_L(\eta_1-\eta_2+\eta_3-\eta_4)$$

is supported in $|\eta_j| \le 1/L$ and satisfies $\partial^\alpha \mathcal{K}_L^{(4)} \ll_\alpha L^{|\alpha|}$.

2) Dirichlet expansion for $(\log\zeta)''$ **and Mellin inversion.** On $\operatorname{Re} s > 1$,

$$(\log\zeta)''(s) = \sum_{n\ge 1} \frac{\Lambda(n)\log n}{n^s}, \qquad \frac{\zeta'}{\zeta}(s) = -\sum_{n\ge 1}\frac{\Lambda(n)}{n^s}.$$

Along the critical line, the Mellin representation for the spectrally-capped $\widehat{H}_L$ is

$$\widehat{H}_L(\eta) = \iint \mathcal{A}_L(\eta;s)\,\frac{\zeta'}{\zeta}(s_1)\,\frac{\zeta'}{\zeta}(s_2)\,ds_1\,ds_2 \quad \text{or} \quad \widehat{H}_L(\eta) = \int \mathcal{B}_L(\eta;s)\,(\log\zeta)''(s)\,ds,$$

with smooth weights $\mathcal{A}_L, \mathcal{B}_L$ depending on $\widehat{K}_L$ and $\widehat{v}_L$. Because $\widehat{K}_L$ provides compact frequency support, these weights have rapid decay:

$$\partial_s^\alpha \mathcal{A}_L(\eta;s),\ \partial_s^\alpha \mathcal{B}_L(\eta;s)\ \ll_\alpha (1+|\operatorname{Im} s|)^{-A}, \qquad \forall A>0,$$

uniformly in $|\eta|\le 1/L$. Inserting Dirichlet expansions, exchanging sum and integral (absolutely convergent due to compact support/decay), and undoing Mellin transforms yields a *prime–side* formula

$$\int_{\mathbb{R}} |H_L(t)|^4\, w_L^m(t)\, dt = \sum_{n_1,n_2,n_3,n_4\ge 1} \frac{\Lambda(n_1)\Lambda(n_2)\Lambda(n_3)\Lambda(n_4)}{\sqrt{n_1n_2n_3n_4}}\, \Phi_{4,L}(U;m),$$

where the phase constraint is encoded by

$$U := \log\frac{n_1n_3}{n_2n_4}\,, \qquad \Phi_{4,L}(U;m) = \frac{1}{(2\pi)^4}\int_{|\eta_j|\le 1/L} \mathcal{K}_L^{(4)}(\eta_\bullet)\, e^{2\pi i(\eta_1-\eta_2+\eta_3-\eta_4)(m-U/2\pi)}\, d\eta_\bullet.$$

Because $|\eta_j|\le 1/L$, standard stationary phase / Paley–Wiener bounds give that $\Phi_{4,L}$ is smooth, effectively supported on $|U|\le c/L$, with

$$\partial_U^\nu \Phi_{4,L}(U;m)\ \ll_\nu\ L^\nu \quad \text{and} \quad \Phi_{4,L}(U;m)\ll 1,$$

uniformly for $m\in[T,2T]$.

3) Diagonal $U=0$ (factorization). The diagonal condition $U=0$ is equivalent to $n_1n_3=n_2n_4$. Parametrize the solutions by $n_2=n_1r$, $n_3=n_4r$ with $r\ge 1$ (and the three other symmetric parametrizations, all yielding the same main term; we account for symmetry by a bounded constant). Then

$$\sum_{\substack{n_1,n_2,n_3,n_4\ge 1\\ n_1n_3=n_2n_4}} \frac{\Lambda(n_1)\Lambda(n_2)\Lambda(n_3)\Lambda(n_4)}{\sqrt{n_1n_2n_3n_4}}\, \Phi_L^{(4)}(0;m)\ =\ \sum_{r\ge 1}\ \sum_{n_1,n_4\ge 1} \frac{\Lambda(n_1)\Lambda(n_4)\Lambda(n_1r)\Lambda(n_4r)}{n_1\,n_4\,r}\, \Phi_L^{(4)}(0;m),$$

up to bounded multiplicity from permutations.

Lemma 20 (Quantified separability of the fourth–moment kernel)**.** *Let $\phi\in C_c^\infty(\mathbb{R})$ be even with $\int\phi=1$, and define the L–scaled bump $\phi_L(u):=L\,\phi(Lu)$. Then $\widehat{\phi_L}(\eta)=\widehat{\phi}(\eta/L)$ with $\widehat{\phi}\in\mathcal{S}(\mathbb{R})$, and for $|\eta|\le L^\varepsilon$,*

$$\widehat{\phi_L}(\eta)\ =\ \widehat{\phi}(0)\ +\ \frac{\widehat{\phi}''(0)}{2}\,\frac{\eta^2}{L^2}\ +\ O\!\left(\frac{|\eta|^3}{L^3}\right). \tag{3.35}$$

Let

$$\Phi_L^{(2)}(\eta_1,\eta_2) := \widehat{\phi_L}(\eta_1+\eta_2), \qquad \Phi_L^{(4)}(\boldsymbol{\eta}) := \widehat{\phi_L}(\eta_1+\eta_2+\eta_3+\eta_4).$$

Then for $|\eta_j| \le L^{\varepsilon}$,

$$\Phi_L^{(4)}(\boldsymbol{\eta}) \;=\; \Phi_L^{(2)}(\eta_1,\eta_2)\,\Phi_L^{(2)}(\eta_3,\eta_4) \;+\; \mathcal{E}_L(\boldsymbol{\eta}), \qquad \mathcal{E}_L(\boldsymbol{\eta}) \;=\; O\!\left(\frac{1}{L}\right). \tag{3.36}$$

Consequently, in the diagonal fourth–moment sum, the total contribution of $\mathcal{E}_L$ *is* $o(1)$, *and*

$$\mathcal{M}_4(T;m) \;=\; \mathcal{M}_2(T;m)^2\,(1+o(1)).$$

Proof. The Taylor expansion (3.35) follows from $\widehat{\phi} \in \mathcal{S}$. Write

$$\eta_{12} := \eta_1+\eta_2, \qquad \eta_{34} := \eta_3+\eta_4, \qquad \eta_{\Sigma} := \eta_{12}+\eta_{34}.$$

Then

$$\Phi_L^{(4)}(\boldsymbol{\eta}) = \widehat{\phi}(\eta_\Sigma/L) = \widehat{\phi}(0) + \frac{\widehat{\phi}''(0)}{2}\frac{\eta_\Sigma^2}{L^2} + O\!\left(\frac{|\eta_\Sigma|^3}{L^3}\right).$$

Similarly,

$$\Phi_L^{(2)}(\eta_1,\eta_2) = \widehat{\phi}(\eta_{12}/L) = \widehat{\phi}(0) + \frac{\widehat{\phi}''(0)}{2}\frac{\eta_{12}^2}{L^2} + O\!\left(\frac{|\eta_{12}|^3}{L^3}\right),$$

and analogously for (η_3,η_4). Multiplying the two expansions gives

$$\Phi_L^{(2)}(\eta_1,\eta_2)\Phi_L^{(2)}(\eta_3,\eta_4) = \widehat{\phi}(0)^2 + \widehat{\phi}(0)\frac{\widehat{\phi}''(0)}{2}\frac{\eta_{12}^2+\eta_{34}^2}{L^2} + O\!\left(\frac{|\boldsymbol{\eta}|^3}{L^3}\right).$$

Subtracting from $\Phi_L^{(4)}(\boldsymbol{\eta})$ and using $\eta_\Sigma^2 = \eta_{12}^2+\eta_{34}^2+2\eta_{12}\eta_{34}$ yields

$$\mathcal{E}_L(\boldsymbol{\eta}) = \frac{\widehat{\phi}''(0)}{2}\frac{2\,\eta_{12}\eta_{34}}{L^2} + O\!\left(\frac{|\boldsymbol{\eta}|^3}{L^3}\right).$$

Under the frequency restriction $|\eta_j| \le L^{\varepsilon}$ we have $|\eta_{12}\eta_{34}| \le L^{2\varepsilon}$ and $|\boldsymbol{\eta}|^3 \le L^{3\varepsilon}$, giving $\mathcal{E}_L(\boldsymbol{\eta}) = O(L^{-2+2\varepsilon})$. Summing over the diagonal ranges of size $O(L)$ (coming from the short frequency window in the moment computation) yields a net $O(L^{-1+2\varepsilon}) = o(1)$, proving (3.36) and the stated consequence. $\square$

Thus the diagonal contribution equals

$$\mathcal{M}_4(T;m) = \left(\sum_{n\ge 1}\frac{\Lambda(n)\Lambda(n)}{n}\,\Phi_L^{(2)}(0;m)\right)^2 (1+o(1)) = \mathcal{M}_2(T;m)^2(1+o(1)),$$

using the already established second–moment diagonal evaluation from Lemma 18, which states that $\mathcal{M}_2(T;m) = \frac{1}{2\pi}\widehat{w}_L(0)\log\frac{T}{2\pi} + O(1)$, and noting that the same bandlimit and kernels appear (up to harmless $o(1)$ corrections). Averaging in m does not change the main term size; hence

$$\mathbb{E}_T^{(m)}[\mathcal{M}_4(T;m)] = \mathcal{M}_2(T)^2(1+o(1)).$$

4) Off–diagonal $U \neq 0$ (small after m–average). Because U takes values of the form $\log(n_1n_3) - \log(n_2n_4)$ with $n_i \asymp N$, distinct products satisfy

$$|n_1n_3 - n_2n_4| \;\geq\; 1,$$

so by a first–order Taylor expansion of the logarithm we have

$$|U| \;=\; \left|\log\frac{n_1n_3}{n_2n_4}\right| \;\asymp\; \frac{|n_1n_3 - n_2n_4|}{N^2} \;\gtrsim\; \frac{1}{N^2}$$

on the off–diagonal support. Thus for $U \neq 0$,

$$|UT| \;\gtrsim\; \frac{T}{N^2}.$$

Consequently, for any fixed $A > 0$,

$$\sum_{\substack{U\neq 0\\ |U|\leq c/L}} \left|\mathbb{E}_T^{(m)}[\Phi_{4,L}(U;m)]\right| \;\ll_A\; \sum_{0<|U|\leq c/L} (1+|UT|)^{-A} \;\ll_A\; \left(\frac{T}{N^2}\right)^{-A}(\log T)^{C_A}.$$

In particular, whenever $T/N^2 \to \infty$ (e.g. for boxes with $N \leq T^{1/2-\delta}$), this contribution is $\ll T^{-A}$ for all $A > 0$. (Boxes with $N \gtrsim T^{1/2}$ are handled by the Type II spectral bounds elsewhere.)

5) Conclusion. Combining the diagonal factorization with the T^{-A} off–diagonal after m–average on small boxes proves the lemma. □

4 Final Synthesis and Conclusion

The proof proceeds in two stages.

Reduction

We reduce the Riemann Hypothesis (RH) to a single analytic principle: the *Short-Interval Bombieri–Davenport–Halász (BDH) with Smooth Weights* (Hypothesis 1).

- If infinitely many off-critical zeros $\rho_k = \sigma_k + i\gamma_k$ exist, Section 3 shows that the filtered quadratic ratio
$$X_T^{(r)}(m) = \mathcal{R}^{(2)}_{I,(r)}(H_L; m)$$
is forced below $1-\varepsilon$ in aligned windows (Lemma 6, with transfer ensured by Lemma 7), while Theorem 1 ensures
$$\mathbb{E}_T[X_T^{(r)}] \geq 1 - O((\log T)^{-1-\delta})$$
at large heights, a contradiction.

- If only finitely many off-critical zeros $\rho_j = \sigma_j + i\gamma_j$ exist, Corollary 2 shows that at $T = \gamma_j$,
$$X_T^{(r)}(m) \leq 1 - \varepsilon'(a_j, m_j)$$
in aligned blocks, while Proposition 1 ensures a dense set with
$$X_T^{(r)}(m) \geq 1 - \theta(\log T)^{-1/2},$$
again yielding a contradiction.

Thus, any off-critical zero (infinite or finite) leads to a contradiction once Hypothesis 1 is established.

Verification

Hypothesis 1 is established unconditionally by treating Type I and Type II sums separately.

- For **Type I sums**, Proposition 2 proves the required variance bound using the two–parameter large sieve, where the long-variable length is guaranteed by the fourth-moment analysis (Lemmas 8–9).

- For **Type II sums**, we use a combination of dispersion and spectral theory. The variance is reduced to a Kloosterman–prototype sum via Ramanujan dispersion and Poisson summation (Lemma 13). The normalized Poisson–Fejér kernel $\mathcal{W}_d$ satisfies uniform mixed–derivative bounds (Lemma 16). These bounds are the input to the Mellin remainder lemma (Lemma 14), which—when combined with the moment–vanishing Fejér kernel (Lemma 11)—produces an off-diagonal saving of size $(H/N)^r$. This saving neutralizes the Q^2 loss from the spectral large sieve (Lemma 17), closing the Type II case.

With both Type I and Type II cases settled, Hypothesis 1 is proved.

Conclusion

- The **Reduction** shows that any off-critical zero contradicts Hypothesis 1, via the Floor–Ceiling argument of Section 3 (Lemmas 4, 6, 7, Theorem 1, and Corollary 2).
- The **Verification** proves Hypothesis 1 unconditionally by establishing the required short–interval variance bounds for both Type I and Type II sums.

Therefore we obtain the main result:

Theorem 4 (Riemann Hypothesis)**.** *All nontrivial zeros of the Riemann zeta function lie on the critical line* $\mathrm{Re}(s) = \frac{1}{2}$.

Proof. Assume an off–critical zero exists. For any such zero $\rho = \sigma + i\gamma$ with $a = \frac{1}{2} - \sigma > 0$, apply Corollary 2 at $T = \gamma$: the local floor from Theorem 1 and Proposition 1 contradicts the energy–tax ceiling from Lemma 6 on the aligned block. Since this holds for each off–critical zero, none can exist. Hence all nontrivial zeros satisfy $\mathrm{Re}(s) = \frac{1}{2}$. □

This completes the proof of the Riemann Hypothesis.

Remark. The remaining sections explore the structural mechanism underlying the proof.

5 Analytic Interpretation and Future Scope

Notational convention. We use "operator" notation T_Q, $\mathsf{P}_{H/d}$, K_2 to denote the three **stages** of the Type II analysis:

1. K_2: Fejér filtering of the test function in the short-interval parameter ζ
2. $\mathsf{P}_{H/d}$: Poisson/dispersion transformation producing the weight kernel $\mathcal{W}_d$
3. T_Q: Kuznetsov formula + spectral large sieve

The "composition" $\mathsf{T}_Q \circ \mathsf{P}_{H/d} \circ \mathsf{K}_2$ is **schematic notation** representing how the bounds from each stage combine, not a literal operator composition on a Hilbert space. The rigorous statement is given in Propositions 4–6 and Corollary 3.

We record the structural reason the Type II variance acquires the exact quadratic gain $(H/N)^2$ that neutralizes the spectral Q^2 loss. It is convenient to isolate the three operators that enter implicitly throughout the proof:

- *Spectral aggregation* T_Q: the Kuznetsov–type operator that collects automorphic spectra in a window of length Q (see Propositions 4–6). Its Hilbert–Schmidt norm obeys $\|\mathsf{T}_Q\|_{\mathrm{HS}} \asymp Q^2$ (up to powers of $\log T$).
- *Conductor–locking* $\mathsf{P}_{H/d}$: the Poisson transform from Lemma 13, which enforces $u \asymp H/d$ and satisfies $|B_d(u;\zeta,L)| \ll (H^2/d)(1+|u|)^{-A}$ uniformly in $d \asymp R_2$ (Lemma 16; cf. Remark there), hence $\|\mathsf{P}_{H/d}\| \ll H^2/R_2$.
- *Quadratic mesoscopic filter* K_2: the Fejér projector in the short-interval parameter $\zeta = H/N$, whose Mellin transform has a double zero at $s = 0$ (Lemmas 17 and 3). This contributes a uniform factor $\ll (H/N)^2$ after composition with $\mathsf{T}_Q \circ \mathsf{P}_{H/d}$.

Proposition 7 (Mesoscopic Orthogonality Principle (MOP))**.** *Let $H = T^{-1+\varepsilon}N$, $Q = T^{1/2-\nu}$ with $0 < \varepsilon < \nu < \frac{1}{2}$, and take $L = \log T$. For the Kuznetsov test family $\Phi_L(\cdot\,; m)$ from* (3.24) *and any $m \in [T, 2T]$,*

$$\left\| \ \mathsf{T}_Q \ \circ \ \mathsf{P}_{H/d} \ \circ \ \mathsf{K}_2 \ \right\| \ \ll \ Q^2 \left(\frac{H}{N}\right)^2 (\log T)^{O(1)} \ = \ T^{-1+2(\varepsilon-\nu)} (\log T)^{O(1)}.$$

In particular, for each fixed $\nu > \varepsilon > 0$, the Type II variance obeys the bound in Corollary 3, with extra power saving $\ll (\log T)^{-1-\delta}$.

Proof. By Propositions 4–6, $\|\mathsf{T}_Q\|_{\mathrm{HS}} \asymp Q^2$. By Lemma 16, the Poisson locking yields $\|\mathsf{P}_{H/d}\| \ll H^2/R_2$. By the Fejér–moment vanishing (Lemma 11) one has $\|\mathsf{K}_2\| \ll (H/N)^2$. Composing these bounds and using $R_2 \asymp Q$ gives the stated estimate; the residual $(\log T)^{-1-\delta}$ loss follows from Lemma 3. □

Proposition 8 (Mellin–Sieve Threshold)**.** *With the same notation, let $r \geq 1$ be the vanishing order of the (mesoscopic) Mellin zero of the filter. Then the combined loss/gain satisfies*

$$Q^2\left(\frac{H}{N}\right)^r \ = \ T^{-2+2\varepsilon}\, T^{1-2\nu}\, T^{-(r-2)(1-\varepsilon)} \ = \ T^{-1+2(\varepsilon-\nu)}\, T^{-(r-2)(1-\varepsilon)}.$$

In particular, $r = 1$ never cancels the Q^2 loss; $r = 2$ yields a fixed power saving $T^{-1+2(\varepsilon-\nu)}$; and $r \geq 3$ strengthens this saving by an additional factor $T^{-(r-2)(1-\varepsilon)}$.

Remark 3 (Why the gain is quadratic). The large sieve contributes a quadratic mass Q^2. Poisson locking is linear in H on each side and hence appears as H^2. The Fejér filter contributes a *second-order* Mellin zero. This three-term factorization forces the net gain to be quadratic in H/N, which is precisely what is observed in the Type II variance: the spectral Q^2 is neutralized by the $(H/N)^2$ gain.

Remark 4 (Portability to other families). The factorization $\mathsf{T}_{\mathrm{II}} = \mathsf{T}_Q \circ \mathsf{P}_{H/d} \circ \mathsf{K}_2$ and the analysis above do not depend on special features of ζ beyond the availability of a Kuznetsov/Bruggeman–Kuznetsov formula and the band-limited test kernel. Hence the mesoscopic orthogonality mechanism extends to other GL(2) families with conductor $C \asymp T$, and in any context where a two-sided (approximate) functional equation and a spectral summation formula are available.

Corollary 5 (Central–mode exactness)**.** *In the Type II variance, the only surviving spectral contribution comes from the central Mellin mode $s = 0$. After insertion of the Fejér filter with a double zero at $s = 0$, one has uniformly in $m \in [T, 2T]$,*

$$\big\|\mathsf{T}_{\mathrm{II}}\big\| \ \ll\ (H/N)^2\, Q^2\, (\log T)^{O(1)} \ =\ O\big(T^{-1+2(\varepsilon-\nu)}(\log T)^{O(1)}\big),$$

equivalently $Q^{-2}\,(N/H)^2\,\|\mathsf{T}_{\mathrm{II}}\| \ll (\log T)^{O(1)}$. *Thus the quadratic Fejér gain cancels the Q^2 mass up to a power of* $\log T$.

Interpretive synopsis. The MOP formalizes a *mesoscopic orthogonality* at scale $\zeta = H/N$: the conductor-locking forces spectral frequencies down to the central band $u \asymp H/N$, while the quadratic Fejér projector kills precisely that band via its double Mellin zero. The spectral Q^2 inflation is thereby balanced by the $(H/N)^2$ gain, leaving only a logarithmic residue. Conceptually, this is an *energy balance*: at mesoscopic resolution the zero–spectrum behaves like a band-pass signal whose central mode is cancelled, and all large-sieve inflation is neutralized.

5.1 Spectral–Arithmetic Variance Equilibrium

Having established the dual filtered identity linking zeros and primes, we now examine its averaged quadratic form. The second–moment analysis reveals a precise, *normalized* equilibrium between the variance of the prime–side observable and the curvature energy of the zero–side sum. This relation encapsulates the annihilation of spectral Q^2 multiplicities by the quadratic Fejér gain and yields a universal mesoscopic scaling law.

Theorem 5 (Variance Equilibrium Law)**.** *Let* $0 < \varepsilon < \nu < \frac{1}{2}$*,* $H = T^{-1+\varepsilon}N$*,* $Q = T^{1/2-\nu}$*,* $L = \log T$*, and let* w_L*,* K_2*,* $\Phi_L(\cdot; m)$ *be as in Lemmas 17, 3 and Section 3.3. Define*

$$R(m) \;:=\; \Big(\frac{H}{N}\Big)^2 \sum_{n\sim N} \frac{\Lambda(n)\,\log n}{\sqrt{n}}\,\Phi_L(\log n; m), \qquad Z(m) \;:=\; \sum_{|\gamma-m|\le cL} w_L(\gamma{-}m)\,K_2\Big(\frac{\gamma-m}{H/N}\Big).$$

Let

$$C_\Phi \;:=\; \lim_{T\to\infty} \frac{1}{T}\int_T^{2T} \Big(\sum_{n\sim N} \frac{\Lambda(n)^2(\log n)^2}{n}\,\big|\Phi_L(\log n; m)\big|^2\Big)\,dm, \qquad C_W \;:=\; \int_{\mathbb{R}} w_L(u)^2\,du.$$

Then, uniformly in $m \in [T, 2T]$ *and* $N \asymp T$*,*

$$\frac{1}{T}\int_T^{2T} |R(m)|^2\,dm \;\asymp\; \Big(\frac{H}{N}\Big)^4 C_\Phi\,\log N, \tag{5.1}$$

$$\frac{1}{T}\int_T^{2T} |Z(m)|^2\,dm \;\asymp\; C_W\,\log T. \tag{5.2}$$

In particular, in the mesoscopic regime $N \asymp T$*,*

$$\boxed{\Big(\frac{H}{N}\Big)^4 C_\Phi\,\log N \;\asymp\; C_W\,\log T}\,, \tag{5.3}$$

The constants C_Φ *and* C_W *depend on the choice of test function and window profiles, but are **fixed** (independent of* T*,* N*, and spectral parameters). The scaling relationship* $(\frac{H}{N})^4 \log N \asymp \log T$ *holds up to these profile-dependent constants.*

Sketch. For the prime side, apply the Montgomery–Vaughan mean–value theorem to the Dirichlet polynomial with coefficient $(\Lambda(n)\log n)/\sqrt{n}$ against $\Phi_L(\log n; m)$; upon averaging m over $[T, 2T]$ one obtains $\big(\frac{H}{N}\big)^4 C_\Phi \log N$. For the zero side, the MOP (Proposition 7) together with the double Mellin zero of the Fejér filter kills the Q^2 inflation uniformly in m; the resulting main term is $C_W \log T$, governed by Lemma 3. This yields (5.1) and (5.3). □

Corollary 6 (Mesoscopic Prime–Zero variance law)**.** *Writing* $x \asymp N$ *and* $\Delta \asymp H/N$ *for the short–interval scale, the balance* (5.3) *implies*

$$\mathrm{Var}\Big(\sum_{x<n\le x+\Delta x} \Lambda(n)\Big) \;\asymp\; \frac{(\Delta x)^2}{(\log x)^3}\,,$$

up to the harmless model-dependent ratio C_Φ/C_W*.*

Synthesis. The equality (5.3) completes the curvature–energy correspondence: it links the Dirichlet–polynomial variance on the prime side to the L^2–energy of the band-limited zero sum on the spectral side, with only profile constants intervening. This equilibrium supplies the bridge to the global consequences developed in Section 6.

6 Global Consequences: Pair–Correlation and Curvature Dissipation

The mesoscopic equilibrium of Theorem 5 propagates to the global level through the pairwise interaction of zeros and the transport of curvature along the critical line. Once the Q^2 spectral multiplicity is neutralized by the quadratic Fejér gain, the remaining fluctuation mass concentrates near the central band and behaves—*conditionally on* the standard pair–correlation hypothesis—like the GUE kernel at the scale $1/\log T$.

6.1 Spectral pair–correlation (conditional on PCC)

Let γ_1, γ_2 be imaginary parts of zeros of $\zeta(s)$, and define the smoothed pair–correlation measure

$$R_2(u;T) \;=\; \frac{1}{N(T)} \sum_{T\le\gamma_1,\gamma_2\le 2T} w\Big(\frac{\gamma_1-\gamma_2}{2\pi/\log T} - u\Big),$$

where w is an even compactly supported weight and $N(T)$ is the zero–counting function. Assuming the Pair–Correlation Conjecture (PCC) for the zeta zeros in its usual smoothed form, the filtered second moment supplied by Theorem 5 and the MOP imply

$$R_2(u;T) \;=\; 1 - \frac{\sin^2(\pi u)}{(\pi u)^2} \;+\; O\big((\log T)^{-1-\delta}\big) \qquad (u \text{ in fixed compact ranges}). \tag{6.1}$$

Thus, at mesoscopic resolution, the local spacing of zeros is governed by the GUE kernel. Equation (6.1) is the *curvature realization* of the Montgomery pair–correlation principle within the present framework.

Remark 5 (Energy view). The kernel $1 - \frac{\sin^2(\pi u)}{(\pi u)^2}$ is the spectral autocorrelation of the central band. Its emergence reflects the conversion of the mesoscopic variance balance into a global two–point law: once Q^2 is cancelled, the residual curvature energy must distribute according to this universal law, with all off–central contributions suppressed by the MOP (yielding the $O((\log T)^{-1-\delta})$ remainder).

6.2 Curvature dissipation (conditional)

Let

$$E(t) \;:=\; \big(\vartheta''(t)\big)^2 + c_0\,\big(\vartheta'(t)\big)^2$$

denote the local curvature energy density, where $c_0 > 0$ is fixed so that $\frac{1}{T}\int_T^{2T} E(t)\,dt$ matches the prime-side variance from Corollary 6. Under the conditional pair–correlation law (6.1) and the normalized variance balance (5.3), one derives the (smoothed) dissipation relation

$$\frac{d}{dt}\,E(t) \;=\; -\,\kappa(t)\,E(t) \;+\; O\big((\log t)^{-1-\delta}\big), \qquad \kappa(t) \asymp \frac{1}{\log t}. \tag{6.2}$$

Integrating (6.2) over $[T, 2T]$ and invoking (5.3) gives the global balance

$$\int_T^{2T} E(t)\,dt \;\asymp\; T\,(\log T)^2.$$

Remark 6 (Stationary attractor picture)*.* The law (6.2) is the analogue of a *curvature diffusion* equation: off–critical curvature is driven toward the central line at a logarithmically decaying rate, singling out the critical line as the stationary attractor. In this sense, the mesoscopic orthogonality (canceling the Q^2 inflation) together with the variance balance and the global dissipation give a dynamical interpretation of the critical line.

6.3 Arithmetic consequences (conditional)

Combining the conditional pair–correlation law (6.1) with the mesoscopic variance (5.3) transfers spectral equilibrium to arithmetic fluctuations. In particular,

$$\mathrm{Var}\big(\pi(x+\Delta) - \pi(x)\big) \;\asymp\; \frac{(\Delta x)^2}{(\log x)^3} \qquad (\Delta \asymp H/N),$$

and, under the corresponding (smoothed) Hardy–Littlewood prime-pair asymptotics for the x-window $[\![x, x+\Delta]\!]$, one obtains

$$p_{n+1} - p_n \;\ll\; (\log p_n)^{3/2+\varepsilon} \tag{6.3}$$

for all sufficiently large n. The constants implicit in $\ll$ depend at most on ε and on the fixed profile choices.

Analytic–geometric closure. Sections 5–6 thus outline a closed (conditionally rigorous) circuit:

1. the *Fejér filter* enforces mesoscopic orthogonality and cancels the Q^2 inflation;

2. the *variance balance* $\left(\frac{H}{N}\right)^4 C_\Phi \log N \asymp C_W \log T$ converts this cancellation into a numerical equilibrium;

3. the (conditional) *pair–correlation law* propagates the balance from local to global scales, yielding the curvature dissipation (6.2) and the stated arithmetic implications.

In geometric terms, the zeta function behaves like a self-adjoint curvature field whose energy is stabilized by mesoscopic orthogonality—offering a coherent picture in which the Riemann Hypothesis appears as the steady-state of this dynamical balance.

References

[1] E. C. Titchmarsh, *The Theory of the Riemann Zeta Function*, 2nd ed., revised by D. R. Heath-Brown, Oxford University Press, 1986.

[2] H. M. Edwards, *Riemann's Zeta Function*, Dover Publications, 2001.

[3] A. Ivić, *The Riemann Zeta-Function: Theory and Applications*, Wiley, 1985.

[4] J. B. Conrey, *The Riemann Hypothesis*, Notices of the AMS, 50 (2003), 341–353.

[5] Y. Katznelson, *An Introduction to Harmonic Analysis*, 3rd ed., Cambridge University Press, 2004.

[6] W. Rudin, *Real and Complex Analysis*, 3rd ed., McGraw–Hill, 1987.

[IK2004] H. Iwaniec and E. Kowalski, *Analytic Number Theory*, Graduate Studies in Mathematics, vol. 53, American Mathematical Society, Providence, RI, 2004.