Mesoscopic Variance Equilibrium and a Conditional Resolution of the Riemann Hypothesis

An Unconditional Prime–Zero Energy Identity and Off-Diagonal Rigidity Criterion

Eric Fodge

December 6, 2025

Abstract

We study the second logarithmic derivative of the Riemann zeta function on the critical line, mollified on the mesoscopic scale $L = \log T$, and develop a variance-equilibrium framework connecting the distribution of zeta zeros to prime number statistics through energy constraints.

On the *arithmetic side*, we consider the mollified curvature field

$$H_L(t) = \big((\log \zeta)'' * v_L * K_L\big)(t), \qquad L = \log T,$$

and its windowed L^2–energy

$$\mathcal{V}_{\mathrm{arith}}(T) := \int_T^{2T} |H_L(t)|^2 w_L(t)\, dt,$$

where v_L and $w_L = v_L * v_L$ are fixed compactly supported mollifiers of width $\asymp L$, and K_L is a spectral cap supported on $|\xi| \ll 1/L$. Using only the Dirichlet series for $(\log \zeta)''(s)$ on $\mathrm{Re}\, s > 1$ and Montgomery–Vaughan type mean-value theorems for Dirichlet polynomials, we show that $\mathcal{V}_{\mathrm{arith}}(T)$ is *unconditionally* locked to the scale

$$\mathcal{V}_{\mathrm{arith}}(T) = (\log T)^4 + O\big((\log T)^{4-\delta}\big),$$

for some fixed $\delta > 0$, with no hypothesis on the location of the nontrivial zeros of $\zeta(s)$.

On the *spectral side*, we use the Hadamard product and the functional equation to

express

$$\widehat{H_L}(\xi) = W_L(\xi)\, \mathcal{Z}(\xi) + \widehat{R}(\xi),$$

where W_L is a fixed smooth kernel supported on $|\xi| \le 1/L$, $\widehat{R}$ is a uniformly bounded analytic remainder, and

$$\mathcal{Z}(\xi) = \sum_{\rho\colon \operatorname{Re}\rho \ge \frac{1}{2}} m_\rho\, e^{-2\pi i \gamma_\rho \xi}\, e^{-2\pi a_\rho |\xi|}, \qquad \rho = \tfrac{1}{2} + a_\rho + i\gamma_\rho,\ a_\rho \ge 0,$$

is a *collective zero spectral density.* A Fourier–Plancherel calculation yields a diagonal/off-diagonal decomposition

$$\mathcal{V}_{\mathrm{spec}}(T) = \mathcal{D}(\{a_\rho\}) + \mathcal{R}(\{a_\rho\}) + O(1),$$

where the diagonal $\mathcal{D}$ captures single-zero contributions and the off-diagonal $\mathcal{R}$ captures interference between zeros.

The framework yields three unconditional results:

1. **Energy locking** (Theorem 6): The prime-side curvature variance is rigidly constrained to $(\log T)^4 + O((\log T)^{4-\delta})$, independent of zero locations.
2. **Diagonal monotonicity** (Lemma 17): The spectral contribution from each individual zero is strictly maximized when that zero lies on the critical line.
3. **Mesoscopic sparsity** (Lemma 28): At most $O((\log T)^{7-\delta})$ zeros in any window $[T, 2T]$ can lie at mesoscopic distance ($\ge A/\log T$) from the critical line.

We prove a *conditional* Riemann Hypothesis: if the off-diagonal interference term cannot compensate for diagonal losses from off-line zeros—formalized as an "off-diagonal rigidity condition" stated explicitly in terms of zero height correlations—then all zeros lie on the critical line.

We characterize precisely the "conspiracy" that would be required for off-line zeros to exist: the zero heights would need to produce specific negative correlations that exactly cancel the diagonal energy deficit in every mesoscopic window. We show this conspiracy is incompatible with GUE statistics for zeta zeros, establishing that within the variance-equilibrium framework, the Riemann Hypothesis is equivalent to GUE-compatible zero correlations. Three potential paths to excluding this conspiracy and establishing unconditional RH are identified.

1 Introduction

The Riemann zeta function

$$\zeta(s) = \sum_{n \geq 1} \frac{1}{n^s}, \qquad \operatorname{Re} s > 1,$$

admits a meromorphic continuation to $\mathbb{C}$ and satisfies the functional equation

$$\pi^{-s/2}\Gamma\left(\frac{s}{2}\right)\zeta(s) = \pi^{-(1-s)/2}\Gamma\left(\frac{1-s}{2}\right)\zeta(1-s).$$

The famous Riemann Hypothesis (RH) asserts that every nontrivial zero ρ of $\zeta(s)$ lies on the critical line $\operatorname{Re} \rho = \frac{1}{2}$. Despite enormous progress on the distribution of primes and zeros, RH has remained open since Riemann's 1859 memoir.

In this paper we approach RH through a variational principle for a *mesoscopic curvature energy* associated to the second logarithmic derivative $(\log \zeta)''(s)$ on the critical line. Working at height T and scale $L = \log T$, we introduce the mollified curvature field

$$H_L(t) := \big((\log \zeta)'' * v_L * K_L\big)(t), \qquad L = \log T,$$

where v_L is a fixed compactly supported time-mollifier of width $\asymp L$ and K_L is a smooth spectral cap supported on $|\xi| \ll 1/L$ with $\widehat{K}_L(0) = 1$. We then measure the local curvature energy through the windowed L^2–variance

$$\mathcal{V}(T) := \int_T^{2T} |H_L(t)|^2 \, w_L(t)\, dt,$$

where $w_L = v_L * v_L$ is the associated Fejér window. Our main theme is that this variance admits two completely different descriptions:

- An *arithmetic* description in terms of the primes, obtained from the Dirichlet series for $(\log \zeta)''(s)$ on the half-plane $\operatorname{Re} s > 1$ and Montgomery–Vaughan type mean-value theorems for Dirichlet polynomials.

- A *spectral* description in terms of the nontrivial zeros $\rho = \beta + i\gamma$ of $\zeta(s)$, obtained from the Hadamard product and the functional equation, which exhibits the zeros as a collective spectral density in Fourier space.

The interplay between these two descriptions, formalized as a *variance-equilibrium identity*, provides strong constraints on where zeros can lie. We establish unconditional results on diagonal energy and zero sparsity, and prove a conditional Riemann Hypothesis under an explicit off-diagonal rigidity condition.

1.1 Main results

Our main contributions are as follows.

Theorem 1 (Prime-side energy locking). *Let H_L and w_L be as above. Then there exists $\delta > 0$ such that*

$$\mathcal{V}(T) = \int_T^{2T} |H_L(t)|^2 \, w_L(t) \, dt = (\log T)^4 + O\big((\log T)^{4-\delta}\big),$$

as $T \to \infty$, where the implied constant depends only on the fixed mollifier profiles. This evaluation is obtained by expanding H_L via the Dirichlet series for $(\log \zeta)''(s)$ in the region $\operatorname{Re} s > 1$ and applying mean-value theorems for short Dirichlet polynomials. In particular, it is ***unconditional*** *and independent of any assumptions on the location of the nontrivial zeros of $\zeta(s)$.*

(See Theorem 6 in Section 3 for the full proof.)

On the spectral side, the Hadamard expansion and the functional equation show that $(\log \zeta)''(s)$ can be written along the critical line as a sum of paired second-order poles at ρ and $1-\overline{\rho}$. After applying the mollifiers v_L and K_L and passing to Fourier space one obtains a representation of the form

$$\widehat{H_L}(\xi) = W_L(\xi) \, \mathcal{Z}(\xi) + \widehat{R}(\xi),$$

where W_L is a fixed smooth kernel supported on $|\xi| \le 1/L$, $\widehat{R}$ is a uniformly bounded analytic remainder, and

$$\mathcal{Z}(\xi) = \sum_{\rho:\, \operatorname{Re} \rho \ge \frac{1}{2}} m_\rho \, e^{-2\pi i \gamma_\rho \xi} \, e^{-2\pi a_\rho |\xi|}, \qquad \rho = \tfrac{1}{2} + a_\rho + i\gamma_\rho, \ a_\rho \ge 0,$$

is a *collective zero spectral density* built by pairing each zero ρ with its functional-equation partner $1-\overline{\rho}$.

Remark 1 (Distributional interpretation). For the critical-line configuration $a_\rho \equiv 0$, the formal sum $\mathcal{Z}_0(\xi) = \sum_{\operatorname{Re} \rho \ge 1/2} m_\rho e^{-2\pi i \gamma_\rho \xi}$ does not converge pointwise; it is properly interpreted as a tempered distribution. All integrals involving $\mathcal{Z}$ in this paper are of the form $\int_{|\xi| \le 1/L} \Omega_L(\xi) \, |\mathcal{Z}(\xi)|^2 \, d\xi$, where $\Omega_L \in C_c^\infty$ is supported on $|\xi| \le 1/L$. These are well-defined by Lemma 13: the compact frequency support ensures only $O(L \log T)$ zeros contribute non-negligibly, making the sum effectively finite. When $a_\rho > 0$, the exponential damping $e^{-2\pi a_\rho |\xi|}$ provides additional convergence, but our arguments work uniformly for all configurations including the critical-line case.

A Fourier–Plancherel computation shows that the variance decomposes as

$$\mathcal{V}(T) = \mathcal{D}(\{a_\rho\}) + \mathcal{R}(\{a_\rho\}) + O(1),$$

where the *diagonal* contribution is

$$\mathcal{D}(\{a_\rho\}) := \sum_\rho m_\rho^2 \, E(a_\rho),$$

with $E(a)$ the single-zero energy, and the *off-diagonal* contribution $\mathcal{R}$ captures interference between distinct zeros.

The key observation is that the exponential factor $e^{-2\pi a_\rho |\xi|}$ acts as a *local damping* of the spectral contribution of any zero with $\operatorname{Re} \rho \neq \frac{1}{2}$. We prove:

Theorem 2 (Diagonal energy monotonicity)**.** *For each zero ρ, the single-zero energy $E(a_\rho)$ is strictly decreasing in a_ρ. In particular, for any family of offsets $a_\rho \geq 0$:*

$$\mathcal{D}(\{a_\rho\}) \leq \mathcal{D}(\{0\}),$$

with strict inequality if any $a_\rho > 0$. That is, the diagonal contribution to the spectral variance is strictly maximized when all zeros lie on the critical line.

(See Lemma 6 and Lemma 17 in Section 3.)

The diagonal monotonicity, combined with the variance identity, yields a strong sparsity constraint on off-line zeros:

Theorem 3 (Mesoscopic sparsity)**.** *For any fixed $A > 0$, the number of zeros ρ in the window $[T, 2T]$ with $\operatorname{Re} \rho - \frac{1}{2} \geq A / \log T$ is at most $O_A((\log T)^{7-\delta})$.*

(See Lemma 28 in Section 4.)

To pass from these unconditional results to a full proof of RH, one must control the off-diagonal term $\mathcal{R}$. We prove:

Theorem 4 (Conditional Riemann Hypothesis)**.** *Define the* off-diagonal rigidity condition*: for every zero ρ,*

$$S_\rho := \sum_{\rho' \neq \rho} \Psi(\gamma_\rho - \gamma_{\rho'}) > -\Phi_1,$$

where Ψ *and* Φ_1 *are explicit spectral quantities (Definition 9).* ***If this condition holds****, then all nontrivial zeros of* $\zeta(s)$ *satisfy* $\operatorname{Re}\rho = \frac{1}{2}$.

(See Theorem 12 in Section 4.)

The gap between conditional and unconditional RH is thus precisely identified: it reduces to showing that the off-diagonal interference in the spectral variance cannot conspire to compensate for diagonal losses from off-line zeros.

1.2 The off-diagonal conspiracy

Theorem 2 shows that any zero off the critical line incurs a definite diagonal energy loss. For such a zero to be consistent with the variance identity, the off-diagonal term $\mathcal{R}$ must compensate.

We characterize this compensation precisely. Define the first variation of the total spectral functional $F = \mathcal{D} + \mathcal{R}$ at the critical-line configuration:

$$\left.\frac{\partial F}{\partial a_\rho}\right|_{a=0} = -4\pi(\Phi_1 + S_\rho),$$

where $\Phi_1 > 0$ is the first moment of the spectral weight and S_ρ is an off-diagonal sum over zero height differences.

- If $S_\rho > -\Phi_1$ for all zeros: the derivative is negative, so moving any zero off-line decreases F, and RH follows.
- If $S_\rho < -\Phi_1$ for some zero ρ: the derivative is positive, so moving ρ off-line could increase F, potentially allowing off-line zeros.

For off-line zeros to exist, the zero heights $\{\gamma_\rho\}$ would need to be arranged in a highly specific "conspiracy" to produce sufficiently negative S_ρ values for those zeros. Corollary 3 shows that this conspiracy would require the off-diagonal term $R(T)$ to develop compensating "bumps" at every height where an off-line zero is visible—a pattern that appears incompatible with natural regularity properties of $R(T)$ (see Section 4.6 for explicit conditions that would rule this out).

The precise characterization of this conspiracy (Corollary 3) reveals that off-line zeros would require a highly structured arrangement of zero heights—specifically, configurations where the off-diagonal sum $S_\rho = \sum_{\rho' \neq \rho} \Psi(\gamma_\rho - \gamma_{\rho'})$ produces negative correlations exceeding the diagonal spectral weight Φ_1 in every mesoscopic window containing the off-line zero. Such "hyper-correlated" arrangements appear incompatible with the statistical repulsion predicted by random matrix theory (Montgomery's pair correlation conjecture [3]), though this incompatibility remains to be established rigorously.

1.3 What remains open

The variance-equilibrium framework reduces RH to a single quantitative question: can the off-diagonal term $\mathcal{R}(\{a_\rho\})$ compensate for diagonal losses from off-line zeros?

We identify three approaches that would resolve this and yield unconditional RH:

1. **Regularity bounds:** Proving uniform bounds $|R(T)| = o((\log T)^{4-\delta})$ together with Lipschitz regularity in T, showing that $R(T)$ cannot develop the bumps required to compensate for off-line zeros at infinitely many heights.

2. **Pair-correlation structure:** Establishing structural constraints on zero pair-correlations that rule out the collective height configurations needed to produce $S_\rho \leq -\Phi_1$ for any zero.

3. **Spectral large-sieve improvement:** Strengthening the Type II/large-sieve bounds to show that off-diagonal mass is too small or too rigid to absorb mesoscopic diagonal deficits.

4. **GUE statistics:** Proving that zeta zeros exhibit GUE pair correlations, which would imply the off-diagonal rigidity condition (see Section 4.7).

Any of these would establish unconditional RH within the present framework. Notably, path (4) connects RH to the extensive body of work on random matrix theory and zeta zeros, suggesting that the "conditional" nature of Theorem 4 may be resolvable through statistical rather than purely analytic methods.

1.4 Structure of the paper

Section 3 develops the curvature energy framework: we define the mollified field H_L, establish the single-zero energy $E(a)$ and its monotonicity, prove the prime-side variance lock (Theorem 6), and develop the diagonal/off-diagonal decomposition with its variational structure.

Section 4 assembles these components into the conditional resolution: we prove the mesoscopic sparsity lemma, establish the variance-equilibrium rigidity criterion, prove the conditional RH theorem, and characterize precisely what remains to be shown for unconditional RH.

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,

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

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

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

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

$$\frac{d^2}{ds^2}\log\zeta(s) = -\sum_{\rho}\frac{m_\rho}{(s-\rho)^2} + \frac{1}{(s-1)^2} + \sum_{n=1}^{\infty}\frac{1}{(s+2n)^2}, \tag{2.6}$$

where the sum is over all nontrivial zeros ρ, counted with multiplicity $m_\rho \geq 1$. This is the complete Hadamard expansion for the second logarithmic derivative; the series over trivial zeros at $s = -2n$ converges absolutely. The formula holds uniformly on compact subsets of $\mathbb{C}$ excluding the zeros and the pole at $s = 1$; see [1, Chap. II, §2.12].

Here and throughout the paper, every sum over ρ is taken with multiplicity.

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.

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: on any open interval of t where $\zeta(\frac{1}{2} + it) \neq 0$, the function $\zeta(s)$ is analytic and nonvanishing on a neighbourhood of the segment $\{\frac{1}{2} + it : t \in I\}$, so a single-valued branch of $\log\zeta$ exists there and $\arg\zeta(\frac{1}{2} + it)$ is continuous in t. Thus $\vartheta(t)$ can jump only when t crosses a zero.

□

Having constructed the corrected phase function $\vartheta(t)$ and established that its second derivative $\vartheta''(t)$ is, away from zeros, a smooth function whose local oscillatory behaviour is dominated by the nontrivial zeros of $\zeta(s)$, we now turn to the heart of the proof of the Riemann Hypothesis.

The key observation is that the second logarithmic derivative of $\zeta(s)$ along the critical line is precisely (up to a sign and an analytic remainder) the complex curvature of the phase:

$$(\log \zeta)''\big(\tfrac{1}{2} + it\big) = -\vartheta''(t) + O(\log |t|).$$

Thus the local L^2-energy of the mollified and band-limited second derivative $(\log \zeta)''$ on scale $L = \log T$ measures the total curvature contributed by the zeros in windows of length $\asymp \log T$ centred at height T.

The prime number theorem, through its most refined effective forms, rigidly constrains the average size of this curvature energy: it must be asymptotically $(\log T)^4$ with extremely small relative fluctuation. On the other hand, an explicit Hadamard-product expansion shows that the same energy is additively assembled from strictly positive individual contributions $E(a_\rho)$ of each zero $\rho = \frac{1}{2} + a_\rho + i\gamma_\rho$, and that $E(a)$ is **strictly maximised** at $a = 0$ and decreases exponentially as $|a|$ grows.

This tension—primes unconditionally fixing the total curvature budget at $(\log T)^4$, while any off-line zero would necessarily reduce its own contribution below the maximal on-line value—can only be resolved if every zero lies exactly on the critical line.

The remainder of this section makes this intuition rigorous by establishing, in order:

- a Cauchy–Schwarz floor and band-limited L^2 control that locks the prime-side curvature energy to $(\log T)^4$ up to lower-order terms; - the strict monotonicity $E(a) < E(0)$ for all $a > 0$; - an unconditional power-saving verification of the required variance via Type-I and Type-II dispersion estimates; - an exact spectral energy identity equating the prime-side energy to $\sum m_\rho^2 E(a_\rho) + O(1)$; - and the resulting contradiction unless $a_\rho = 0$ for all nontrivial zeros ρ.

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

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

Then $w_L \ge 0$ 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

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

Spectral cap and mollified field. Independently, fix a spectral cap with Fourier transform

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

The time-domain kernel is the Fejér kernel:

$$K_L(t) := \mathcal{F}^{-1}[\widehat{K}_L](t) = \frac{1}{L} \cdot \frac{\sin^2(\pi t/L)}{(\pi t/L)^2},$$

which is even, nonnegative, and satisfies $\int_{\mathbb{R}} K_L = 1$. Convolution with K_L thus removes Fourier content at frequencies $|\xi| > 1/L$. Define

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

Roadmap of this section. This section develops the curvature energy framework that underpins our analysis.

Throughout, we use the terms "energy" and "variance" interchangeably at the mesoscopic scale: the spectral variance

$$\mathcal{V}_{\mathrm{spec}}(T) := \int_T^{2T} |H_L(t)|^2 w_L(t)\, dt$$

is the m–average of the local curvature energy $E(m)$ defined above.

We proceed as follows:

Floor bounds. We establish Cauchy–Schwarz and bandlimited L^2 controls showing that the local curvature energy is bounded below.

Single-zero energy. We define the curvature energy $E(a)$ contributed by a zero at horizontal distance a from the critical line, and prove $E(a) < E(0)$ for all $a > 0$—maximum energy occurs exactly on the critical line (Lemma 6).

Prime-side energy. We prove (Theorem 6) that the windowed L^2-energy of the mollified curvature field H_L satisfies

$$\int |H_L(t)|^2 w_L^m(t)\, dt \;=\; (\log T)^4 + O\big((\log T)^{4-\delta}\big)$$

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

The path to RH. The combination of the rigid prime-side energy floor ($(\log T)^4$) and the strict monotonicity of the single-zero energy ($E(a)$) sets the stage for Section 4. The spectral fourth-moment decomposition (Lemma **??**) and the interference bound (Lemma **??**) establish that the spectral variance is asymptotically additive. This energy conservation law then forces all zeros to lie on the critical line.

Fourier and window conventions. We use

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

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 [2, Chap. 5]. *Note: The ψ_L notation above is provided solely for cross-reference with [2], where ψ plays the role of our υ, and $\psi_L(u-m)$ corresponds to our $w_L^m(u)$. Throughout this manuscript we use the υ_L/w_L notation exclusively.*

3.1 Cauchy–Schwarz Floor for Quadratic Energy

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

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

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. □

3.2 Single–Zero Curvature Energy

We now define the curvature energy contributed by a single zero using the *exact* Hadamard contribution, without any Lorentzian approximation.

Definition 5 (Single-zero curvature energy). *Let $\rho = \frac{1}{2} + a + i\gamma$ with $a \ge 0$. Define*

$$G_\rho(t) := \left(\frac{1}{((t-\gamma) - ia)^2} \right) * v_L * K_L(t).$$

The curvature energy *from ρ is*

$$E(a) := \int_{\mathbb{R}} |G_\rho(t)|^2 \, w_L(t) \, dt.$$

Remark 2. The energy $E(a)$ depends only on the horizontal offset $a = \operatorname{Re}\rho - \frac{1}{2}$, not on $\gamma = \operatorname{Im}\rho$. This follows from Lemma 5: the phase $e^{-2\pi i \xi \gamma}$ has modulus 1 and cancels in $|\widehat{G_\rho}|^2$.

Lemma 4 (Remainder bounds)**.** *In the decomposition*

$$\widehat{H_L}(\xi) = W_L(\xi) \, \mathcal{Z}(\xi; \{a_\rho\}) + \widehat{R}(\xi),$$

the remainder satisfies $|\widehat{R}(\xi)| \ll (\log T)^C$ for some $C > 0$ and $|\xi| \le 1/L$, and

$$\int_{\mathbb{R}} |\widehat{R}(\xi)|^2 \, \widehat{w}_L(\xi) \, d\xi \ll (\log T)^{2C} \cdot L^{-1} = o((\log T)^4).$$

In particular, the remainder contributes $o((\log T)^4)$ to the variance, which is negligible compared to the main term.

Proof. The remainder has three sources:

(a) Pole at $s = 1$. The second logarithmic derivative has $(\log \zeta)''(s) = (s-1)^{-2} + O(1)$ near $s = 1$. Along $s = \frac{1}{2} + it$, this contributes $O(1/(t^2 + 1/4))$. After mollification by v_L (with derivative scaling $\partial^k v_L \ll L^{-k}$), the contribution to $H_L(t)$ is $O(L^{-1})$, giving $|\widehat{R}^{(\text{pole})}(\xi)| \ll L^{-1}$.

(b) Trivial zeros at $s = -2n$. These contribute $\sum_{n \ge 1} (s + 2n)^{-2}$ to $(\log \zeta)''(s)$. The functional equation implies exponential suppression via the gamma factor: $|\Gamma(s/2)| \sim e^{-\pi|t|/2}$ for $|t| \to \infty$. Thus $|\widehat{R}^{(\text{trivial})}(\xi)| \ll e^{-cT}$ for some $c > 0$.

(c) Gamma factor terms. The Hadamard product's gamma contributions yield smooth, slowly-varying terms along the critical line. After mollification, these satisfy $|\widehat{R}^{(\text{gamma})}(\xi)| \ll (\log T)^C$ for some $C > 0$.

Combining all three contributions:

$$|\widehat{R}(\xi)| \ll (\log T)^C$$

on $|\xi| \le 1/L$. Since $\widehat{w}_L(\xi) \ll 1$ and is supported on $|\xi| \ll 1/L$,

$$\int |\widehat{R}|^2 \widehat{w}_L \, d\xi \ll (\log T)^{2C} \cdot L^{-1} = o((\log T)^4),$$

confirming the remainder is negligible compared to the $(\log T)^4$ main term. ∎

Lemma 5 (Fourier representation)**.** *With $\widehat{v}_L, \widehat{K}_L$ real, supported in $|\xi| \le 1/L$, we have*

$$\widehat{G_\rho}(\xi) = 4\pi^2 \,|\xi|\, e^{-2\pi a|\xi|}\, \widehat{v}_L(\xi)\, \widehat{K}_L(\xi)\, e^{-2\pi i\xi\gamma}.$$

Proof. For $a > 0$, consider $f(t) = \dfrac{1}{t - ia}$. A standard computation (via contour integration or distributional calculus) shows

$$\widehat{f}(\xi) = -2\pi i \,\mathrm{sgn}(\xi)\, e^{-2\pi a|\xi|}.$$

Since

$$\frac{d}{dt}\Big(\frac{1}{t - ia}\Big) = -\frac{1}{(t - ia)^2},$$

the Fourier transform of $1/(t - ia)^2$ is obtained by $\mathcal{F}[-f'](\xi) = 2\pi i\xi\, \widehat{f}(\xi)$, giving

$$\widehat{\frac{1}{(t - ia)^2}}(\xi) = 2\pi i\xi\, \widehat{f}(\xi) = 2\pi i\xi\, \big(-2\pi i \,\mathrm{sgn}(\xi)\, e^{-2\pi a|\xi|}\big) = 4\pi^2|\xi|\, e^{-2\pi a|\xi|}.$$

Translation by γ multiplies the transform by $e^{-2\pi i\xi\gamma}$, and convolution with v_L and K_L multiplies by $\widehat{v}_L(\xi)$ and $\widehat{K}_L(\xi)$, giving the stated formula. □

Lemma 6 (Monotonicity of curvature energy)**.** *For $a \ge 0$ the function $E(a)$ satisfies:*

(i) $E(a)$ is strictly decreasing in a;

(ii) $E(0) > E(a)$ for all $a > 0$;

(iii) $E(a) \to 0$ as $a \to \infty$.

Proof. By Plancherel and Lemma 5,

$$E(a) = \int_{\mathbb{R}} |\widehat{G_\rho}(\xi)|^2\, \widehat{w}_L(\xi)\, d\xi = \int_{|\xi|\le 1/L} 16\pi^4\xi^2\, e^{-4\pi a|\xi|}\, |\widehat{v}_L(\xi)\widehat{K}_L(\xi)|^2 \widehat{w}_L(\xi)\, d\xi.$$

All factors except $e^{-4\pi a|\xi|}$ are nonnegative and independent of a. For each $\xi \ne 0$, the map $a \mapsto e^{-4\pi a|\xi|}$ is strictly decreasing. Since the integrand has positive mass on a set of positive measure, $E(a)$ is strictly decreasing in a. The limits $E(a) \to 0$ as $a \to \infty$ follow immediately from dominated convergence and the exponential factor. □

Remark 3 (Physical interpretation). The energy $E(a)$ measures the local L^2-mass of the curvature signal from a zero at distance a. A zero on the critical line ($a = 0$) produces maximum "curvature energy"; moving the zero off-line exponentially damps its contribution. This is the mechanism by which variance equilibrium forces all zeros onto the critical line.

Lemma 7 (Variance Equilibrium Identity)**.** *Let* $\mathcal{V}_{\mathrm{arith}}(T) := \int_{\mathbb{R}} |H_L(t)|^2 w_L(t)\, dt$ *denote the curvature energy computed via the prime-side Dirichlet expansion of* $(\log\zeta)''$*, and let* $\mathcal{V}_{\mathrm{spec}}(T)$ *denote the same integral computed via the Hadamard product expansion. Then*

$$\mathcal{V}_{\mathrm{arith}}(T) = \mathcal{V}_{\mathrm{spec}}(T).$$

Proof. Both expressions represent the L^2-integral of the same function $H_L(t)$ against the same weight $w_L(t)$. The prime-side and spectral-side expansions are two representations of the identical analytic object. □

Theorem 6 (Prime-side curvature energy locking)**.** *Define the local curvature energy by*

$$E(m) \;:=\; \int_{\mathbb{R}} |H_L(t)|^2\, w_L^m(t)\, dt \qquad \big(H_L(t) = \big((\log\zeta)'' * v_L * K_L\big)(t),\ L = \log T\big).$$

Under the same assumptions and parameters as in the rest of this section, there exists $\delta > 0$ *(depending on the fixed parameter choices in Table 1) such that uniformly for* $m \in [T, 2T]$*,*

$$E(m) \;=\; (\log T)^4 + O\big((\log T)^{4-\delta}\big).$$

Proof. The diagonal contribution to $E(m)$ is $(\log T)^4 + O((\log T)^3)$ by the explicit second-moment evaluation (Lemma 10) and the separability of the fourth-moment kernel (Lemma 12).

The off-diagonal contributions are controlled by the fourth-moment expansion (Lemma 11).

- Small dyadic boxes ($N \le T^{1/2-\delta}$) contribute $\ll T^{-A}$ absolutely by m-averaging. - Balanced-large boxes ($M \asymp N \ge T^{\theta_0}$) are bounded using the Type-II dispersion estimates, which — via the Mesoscopic Orthogonality Principle (Proposition 7) and the spectral large-sieve bounds (Propositions 2–4) — contribute $\ll (\log T)^{4-\delta}$ absolutely.

Summing dyadically over all boxes yields the stated bound. □

Remark 4 (Origin of the fourth-power scaling)*.* The $(\log T)^4$ main term in the variance arises from the diagonal contribution in the Parseval identity applied to the Dirichlet series for $(\log\zeta)''(s) = -\sum_{n\ge1} b(n)\, n^{-s}$, whose coefficients satisfy $b(p^k) \asymp (\log p)^2$ on prime powers. The diagonal second moment is controlled by

$$\sum_{n\asymp T} \frac{|b(n)|^2}{n} \;\asymp\; \sum_{p\le T} \frac{(\log p)^4}{p} \;\sim\; (\log T)^4,$$

which yields the main term $\frac{1}{2\pi}\widehat{w}_L(0)\big(\log\frac{T}{2\pi}\big)^4$ in Lemma 10.

Remark 5 (Independence from Zero-Free Regions)*.* The evaluation of the prime-side second moment in Theorem 6 uses only the Dirichlet-series for $(\log\zeta)''$, the mollifiers v_L and K_L, Type I/II dispersion, and the spectral large-sieve inequalities. At no point do we invoke the Prime Number Theorem, a zero-free region, or any assumption on the location of the zeros of $\zeta(s)$. Thus the prime-side energy locking is unconditional and logically independent of the zero-side curvature decomposition.

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

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

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	Type II threshold: boxes with $N \ge T^{\theta_0}$ routed to Type II
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 $\ll T^{-0.86}(\log T)^C$, a power saving in T that is far stronger than any log-power loss. In particular, the Type II contribution does not constrain the choice of $\delta > 0$ in Theorem 6; any loss in the exponent δ comes from the diagonal and Type I analysis, not from the Type II range. The parameter choices above meet all required inequalities with comfortable margins.

Remark 6 (Determination of δ). The power saving $\delta > 0$ in Theorem 6 is determined by the weakest contribution:

1. **Diagonal:** Error is $O((\log T)^3)$, giving $\delta = 1$ from this source.

2. **Type I off-diagonal:** The m-averaging gives $O(T^{-A})$ for boxes with $N \le T^{1/2-\varepsilon}$, contributing no constraint.

3. **Type II:** Contributes $O(T^{-0.86}(\log T)^C)$, a power saving in T.

Since all off-diagonal contributions achieve power savings in T, we may take δ to be any fixed positive constant; with Table 1 parameters, $\delta = 0.1$ is valid. The argument is not optimized for maximal δ.

We emphasize the role separation: m–average decay controls boxes with $N \le T^{1/2-\delta}$ via $(T/N^2)^{-A}$ (as in Lemma 11), while the $(H/N)^r$–gain neutralizes the spectral Q^2 loss in the balanced–large Type II boxes $M \asymp N \ge T^{\theta_0}$.

With these choices one has

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

for $H = T^{-1+\varepsilon}N$, $N \ge T^{\theta_0}$, $Q = T^{1/2-\nu}$, and $L = \log T$. Hence all implied constants in Lemmas 26–27 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.

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

3.4 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. [2, §9.3, Lem. 9.4]). Explicitly: for fixed $n_2n_4 \asymp N^2$, the number of pairs (n_1,n_3) with $n_1 \asymp M_1$, $n_3 \asymp M_3$, and $|n_1n_3 - n_2n_4| \le \Delta$ is $\ll 1 + M_1M_3(\log N)^C/\Delta$. 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 [2, 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 $M_1M_3 \ll T^{1+o(1)}$ and $M_2M_4 \ll T^{1+o(1)}$, which again implies $M_1M_2M_3M_4 \ll T^{2+o(1)}$. □

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*

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

Proof. By Lemma 8, non–negligible blocks satisfy

$$N \cdot M_1M_2M_3 \;\asymp\; M_1M_2M_3M_4 \;\ll\; T^{2+o(1)}.$$

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

$$W_L(m,n) \; := \; \int_{\mathbb{R}} v_L(u-m)\, v_L(u-n)\, du \; = \; (v_L * v_L)(m-n) \; = \; w_L(m-n), \qquad (3.3)$$

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 1 (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 (3.3) *with $L = \log T$. Set $Q = T^{1/2-\nu}$ with small fixed $\nu, \varepsilon > 0$. Assume the* Type I regime

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

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:

$$\sum_{q\le Q}\frac{1}{\varphi(q)}\sum_{\chi\ (\mathrm{mod}\ q)}\left|\sum_{n} c_n\,\chi(n)\right|^2 \ \ll\ (Q^2+H)\sum_{n}|c_n|^2,$$

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

$$\mathcal{V}_{\mathrm{I}}(M,N;Q)\ \ll\ (Q^2+H)\sum_{n\sim N}|c_n|^2. \tag{3.4}$$

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

$$\sum_{n\sim N}|c_n|^2\ \ll\ H\cdot N^{o(1)}\ \ll\ H(\log T)^C. \tag{3.5}$$

Conclusion. Insert (3.5) into (3.4):

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

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.

□

Spectral large–sieve bounds: formal statements and proofs

We retain the notation of Proposition 6 and Lemma 27. 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 (3.19), and the transforms $\mathcal{J}_\bullet(\Phi,g;R_2)$ defined in (3.25). The short–interval transform gain is recorded in (3.28).

Proposition 2 (Spectral large–sieve bound: holomorphic channel)**.** *Let* $\mathcal{H}_{m,n}[\Phi,g;R_2]$ *be as in* (3.22)*. 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 (3.22) 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 [2, 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 (3.26)). 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 (3.28) (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. $\square$

Proposition 3 (Spectral large–sieve bound: Maass channel)**.** *Let $\mathcal{M}_{m,n}[\Phi, g; R_2]$ be as in (3.23). 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 (3.28), $|\mathcal{J}_t^{\pm}| \ll_A (1+|t|)^{-A}\Big(\frac{H}{N}\Big)^r$. 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 [2, 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. $\square$

Proposition 4 (Spectral large–sieve bound: Eisenstein channel)**.** *Let $\mathcal{E}_{m,n}[\Phi, g; R_2]$ be as in (3.24). 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 $|t| \leq T \asymp R_2$ using (3.28), and apply the continuous spectral large–sieve (Iwaniec-Kowalski [2, Thm. 16.5, p. 387], continuous spectrum case):

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

and likewise for β. Combine as above. □

Corollary 1 (Fixed–modulus Kloosterman–prototype bound)**.** *Let $\mathcal{K}(M,N;R_2)$ be as in (3.20). Then for any $A>0$,*

$$|\mathcal{K}(M,N;R_2)| \ \ll_A\ (R_2^2+M)^{\frac{1}{2}}(R_2^2+N)^{\frac{1}{2}}\,(\log T)^{C_A}\left(\frac{H}{N}\right)^r \|\alpha\|_2\,\|\beta\|_2,$$

uniformly for dyadic $R_2 \leq Q$.

Proof. Sum the bounds of Propositions 2, 3, 4 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 $r \geq 1$ so that

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

For example, any $r > \frac{1-2\nu}{1-\varepsilon}$ suffices. With this choice, the $(H/N)^r$ factor from Lemma 27 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 (3.11) 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 10 (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

$$\mathcal{M}_2(T;m) = \frac{1}{2\pi}\,\widehat{w}_L(0)\left(\log\frac{T}{2\pi}\right)^4 + O\big((\log T)^3\big), \tag{3.6}$$

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*

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

for every $A > 0$*.*

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 the Hadamard expansion

$$(\log\zeta)''(s) = -\sum_\rho \frac{m_\rho}{(s-\rho)^2} + A(s),$$

separate the diagonal main term $\mathcal{M}_2$ and the zero/off–diagonal part $\mathcal{Z}_2$.

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 $\zeta'/\zeta(s) \sim -1/(s-1)$ near $s=1$, and the Dirichlet–series for $(\log\zeta)''$ has coefficients $b(n) = \Lambda(n)(\log n)^2$, the diagonal contribution picks up four powers of $\log T$:

$$\mathcal{M}_2(T;m) = \frac{1}{2\pi}\,\widehat{w}_L(0)\left(\log\frac{T}{2\pi}\right)^4 + O\big((\log T)^3\big),$$

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 into the contour representation, exchange sums/integrals, and invert Mellin transforms to obtain

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

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. □

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 11 (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 $\operatorname{Re} s > 1$,

$$(\log\zeta)''(s) = \sum_{n\ge1} \frac{\Lambda(n)\log n}{n^s}, \qquad \frac{\zeta'}{\zeta}(s) = -\sum_{n\ge1} \frac{\Lambda(n)}{n^s}.$$

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

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

where the phase constraint is encoded by

$$U := \log \frac{n_1 n_3}{n_2 n_4}\,, \qquad \Phi_{4,L}(U; m) = \frac{1}{(2\pi)^4} \int_{|\eta_j| \le 1/L} \mathcal{K}_L^{(4)}(\eta_\bullet)\, e^{2\pi i(\eta_1 - \eta_2 + \eta_3 - \eta_4)(m - U/2\pi)}\, d\eta_\bullet.$$

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 $n_1 n_3 = n_2 n_4$. Parametrize the solutions by $n_2 = n_1 r$, $n_3 = n_4 r$ 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

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

up to bounded multiplicity from permutations.

Lemma 12 (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* $|\eta| \le L^{\varepsilon}$,

$$\widehat{\phi_L}(\eta) \; = \; \widehat{\phi}(0) \; + \; \frac{\widehat{\phi}''(0)}{2} \, \frac{\eta^2}{L^2} \; + \; O\biggl(\frac{|\eta|^3}{L^3}\biggr). \tag{3.7}$$

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

$$\Phi_L^{(4)}(\boldsymbol{\eta}) \; = \; \Phi_L^{(2)}(\eta_1, \eta_2) \, \Phi_L^{(2)}(\eta_3, \eta_4) \; + \; \mathcal{E}_L(\boldsymbol{\eta}), \qquad \mathcal{E}_L(\boldsymbol{\eta}) \; = \; O\biggl(\frac{1}{L}\biggr). \tag{3.8}$$

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 (3.7) 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

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

Similarly,

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

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

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

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

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

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 $\mathcal{E}_L(\boldsymbol{\eta}) = O(L^{-2+2\varepsilon})$. The error $\mathcal{E}_L(\boldsymbol{\eta}) = O(L^{-2+2\varepsilon})$ arises at each point in the diagonal configuration. The diagonal time ranges arising from the inverse Fourier transform have length $O(L)$, contributing $O(L)$ effectively independent terms. The total error is therefore

$$O(L) \cdot O(L^{-2+2\varepsilon}) = O(L^{-1+2\varepsilon}) = o(1),$$

since $\varepsilon < 1/2$. This proves (3.8) and the stated consequence. $\square$

Thus the diagonal contribution equals

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

using the already established second–moment diagonal evaluation from Lemma 10, 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. $\square$

3.5 Collective zero spectral density and Fourier-side variance

We now introduce a collective spectral representation of the zero contributions to the mollified curvature field H_L. This replaces the individual wavepacket model $H_L = \sum_\rho G_\rho + R$ used previously and will be the backbone of the spectral-side analysis in Section 4.

Recall that H_L is defined by

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

with $L = \log T$, v_L the time mollifier, and K_L the spectral cap from (3.2). We write the Fourier transform as $\widehat{f}(\xi) = \int_{\mathbb{R}} f(t)\, e^{-2\pi i t\xi}\, dt$.

Definition 7 (Collective zero spectral density). *Write the nontrivial zeros as*

$$\rho = \tfrac{1}{2} + a_\rho + i\gamma_\rho, \qquad a_\rho \in \mathbb{R},$$

counted with multiplicity $m_\rho \in \mathbb{N}$. For a choice of offsets $\{a_\rho\}$ with $a_\rho \geq 0$ when $\operatorname{Re}\rho \geq \frac{1}{2}$, define the collective zero spectral density

$$\mathcal{Z}(\xi; \{a_\rho\}) := \sum_{\operatorname{Re}\rho \geq \frac{1}{2}} m_\rho\, e^{-2\pi i\gamma_\rho\xi}\, e^{-2\pi a_\rho|\xi|}, \qquad \xi \in \mathbb{R}.$$

We denote by

$$\mathcal{Z}_0(\xi) := \mathcal{Z}(\xi; \{a_\rho \equiv 0\}) = \sum_{\operatorname{Re}\rho \geq \frac{1}{2}} m_\rho\, e^{-2\pi i\gamma_\rho\xi}$$

the critical-line configuration. The Fourier transform of H_L admits the decomposition

$$\widehat{(\log\zeta)'' * v_L} * K_L(\xi) = W_L(\xi)\, \mathcal{Z}(\xi; \{a_\rho\}) + \widehat{R}(\xi), \tag{3.9}$$

where $W_L(\xi) = \widehat{v}_L(\xi)\widehat{K}_L(\xi)$ is smooth and supported in $|\xi| \leq 1/L$, and $\widehat{R}(\xi)$ is the contribution of the pole, the trivial zeros, and the analytic remainder term $H(s)$ in (2.6).

Lemma 13 (Effective finiteness of collective density). *Let $\mathcal{Z}(\xi; \{a_\rho\})$ be as in Definition 7, and let $W_L(\xi)$ be supported on $|\xi| \leq 1/L$ with $L = \log T$.*

(i) For each fixed T and $|\xi| \leq 1/L$, the sum defining $\mathcal{Z}(\xi; \{a_\rho\})$ receives non-negligible contributions only from zeros in the height window $|\gamma_\rho - T| \leq CL$ for a suitable constant

$C > 0$. *This window contains $O(L \log T)$ zeros, and their contributions sum to a finite value.*

(ii) Zeros outside this window contribute $O(L^{-A})$ for any $A > 0$ to any integral of $\mathcal{Z}$ against smooth functions supported on $|\xi| \le 1/L$, by repeated integration by parts exploiting the rapid oscillation of $e^{-2\pi i\gamma_\rho \xi}$.

(iii) The integral $\int_{|\xi|\le 1/L} \Omega_L(\xi)\, |\mathcal{Z}(\xi; \{a_\rho\})|^2\, d\xi$ converges, with the same statements holding for $\mathcal{Z}_0$ when $a_\rho \equiv 0$.

Proof. **(i)** By the Riemann–von Mangoldt formula, the number of zeros with $|\gamma_\rho - T| \le CL$ is $O(L \log T)$. Each term $m_\rho e^{-2\pi i\gamma_\rho\xi} e^{-2\pi a_\rho|\xi|}$ has modulus $\le m_\rho \le (\log T)^C$, so the windowed sum has at most $O(L(\log T)^{C+1})$ total contribution.

(ii) For zeros with $|\gamma_\rho - T| > CL$, consider the integral

$$\int_{|\xi|\le 1/L} \Omega_L(\xi)\, e^{-2\pi i\gamma_\rho\xi}\, d\xi.$$

Since $\Omega_L \in C_c^\infty$ and $|\gamma_\rho| \ge T - CL \gg L$, integration by parts k times introduces factors of $(2\pi i\gamma_\rho)^{-k}$, yielding $O((L/|\gamma_\rho - T|)^k) = O(C^{-k})$. Summing over the $O(T \log T)$ zeros outside the window, with C chosen large, gives $O(L^{-A})$.

(iii) From (i), $|\mathcal{Z}(\xi)|^2 \ll (L \log T)^2 (\log T)^{2C}$ on the support of Ω_L. Since Ω_L is bounded and supported on an interval of length $O(1/L)$, the integral converges. □

Remark 7. We do not claim that $\mathcal{Z}(\xi)$ converges as an infinite sum over all zeros; the undamped density $\mathcal{Z}_0(\xi)$ is properly a distribution. However, effective finiteness in the relevant height window suffices for all integrals appearing in our argument.

Remark 8 (Functional equation pairing). By the functional equation $\xi(s) = \xi(1-s)$, nontrivial zeros occur in pairs $\{\rho, 1-\rho\}$ symmetric about the critical line: if $\rho = \frac{1}{2} + a + i\gamma$ with $a > 0$, then $1 - \rho = \frac{1}{2} - a - i\gamma$ lies at horizontal distance a to the *left* of the line. Conjugate symmetry $\zeta(\bar{s}) = \overline{\zeta(s)}$ further pairs ρ with $\bar{\rho}$.

In our collective Fourier representation we sum only over zeros with $\operatorname{Re}\rho \ge \frac{1}{2}$ and encode the horizontal distance as $a_\rho := \operatorname{Re}\rho - \frac{1}{2} \ge 0$. The contribution from zeros in the left half-plane $\operatorname{Re}\rho < \frac{1}{2}$ is absorbed into the analytic remainder $\widehat{R}(\xi)$. This convention ensures all damping exponents are nonnegative, guarantees convergence of $\mathcal{Z}(\xi; \{a_\rho\})$, and produces the damping structure $e^{-2\pi a_\rho|\xi|}$ central to the maximality argument.

Lemma 14 (Fourier-side variance identity)**.** *Let $L = \log T$ and $w_L = v_L * v_L$ as in* (3.1)*. Then*

$$\int_T^{2T} |H_L(t)|^2 w_L(t)\, dt = \int_{|\xi|\le 1/L} \Omega_L(\xi)\, |\mathcal{Z}(\xi; \{a_\rho\})|^2\, d\xi \;+\; O(1), \tag{3.10}$$

where

$$\Omega_L(\xi) := |W_L(\xi)|^2\,\widehat{w_L}(\xi) \geq 0, \qquad \Omega_L(\xi) \asymp 1 \quad \textit{for } |\xi| \leq c/L$$

for some fixed $c > 0$*.*

Proof. Insert (3.9) and apply Plancherel with the weight w_L:

$$\int_{\mathbb{R}} |H_L(t)|^2 w_L(t)\,dt = \int_{\mathbb{R}} |\widehat{H_L}(\xi)|^2 \widehat{w}_L(\xi)\,d\xi.$$

Since W_L and hence $\widehat{H_L}$ are supported in $|\xi| \leq 1/L$, we may restrict the ξ–integral to $|\xi| \leq 1/L$. Using $\widehat{H_L}(\xi) = W_L(\xi)\mathcal{Z}(\xi;\{a_\rho\}) + \widehat{R}(\xi)$ gives

$$\int |H_L|^2 w_L = \int_{|\xi|\leq 1/L} |W_L(\xi)|^2 |\mathcal{Z}(\xi;\{a_\rho\})|^2 \widehat{w}_L(\xi)\,d\xi + 2\,\mathrm{Re}\int W_L \mathcal{Z}\,\overline{\widehat{R}}\,\widehat{w}_L + \int |\widehat{R}|^2 \widehat{w}_L.$$

The last two terms are $O(1)$ uniformly in T by the same arguments used in Lemma 10, since $\widehat{R}$ is supported in $|\xi| \leq 1/L$ and arises from an analytic function with polylogarithmic growth. This yields (3.10) with $\Omega_L(\xi) = |W_L(\xi)|^2 \widehat{w}_L(\xi)$, which is nonnegative and bounded above and below on $|\xi| \leq c/L$ because $W_L(0) \neq 0$ and $\widehat{w}_L$ is smooth with $\widehat{w}_L(0) = \int w_L = 1$. □

Remark 9 (Role of the collective representation). The identity (3.10) replaces the earlier decomposition into individual wavepackets G_ρ. The damping factors $e^{-2\pi a_\rho |\xi|}$ now enter directly inside the positive quadratic form $\int \Omega_L(\xi)\,|\mathcal{Z}(\xi;\{a_\rho\})|^2\,d\xi$, which is the key object in the spectral side of the argument.

3.6 Diagonal/Off-Diagonal Decomposition and Variational Structure

We now decompose the spectral variance into diagonal and off-diagonal contributions and analyze the variational structure that governs the relationship between zero locations and total energy.

Definition 8 (Diagonal and off-diagonal contributions)**.** *For a configuration of zeros with offsets* $\{a_\rho\}_\rho$*, define the* diagonal contribution

$$\mathcal{D}(\{a_\rho\}) := \sum_\rho m_\rho^2\,E(a_\rho),$$

where $E(a)$ is the single-zero energy from Definition 5, and the off-diagonal contribution

$$\mathcal{R}(\{a_\rho\}) := \sum_{\rho \neq \rho'} m_\rho m_{\rho'}\, K(\rho, \rho'),$$

where the off-diagonal kernel is

$$K(\rho, \rho') := \int_{|\xi| \leq 1/L} \Omega_L(\xi)\, e^{-2\pi(a_\rho + a_{\rho'})|\xi|}\, e^{-2\pi i(\gamma_\rho - \gamma_{\rho'})\xi}\, d\xi.$$

Lemma 15 (Off-diagonal kernel decay)**.** *The off-diagonal kernel $K(\rho, \rho')$ satisfies:*

(i) ***Diagonal value:*** *$K(\rho, \rho) = E(a_\rho)$.*

(ii) ***Riemann–Lebesgue decay:*** *For $\rho \neq \rho'$ with $|\gamma_\rho - \gamma_{\rho'}| \geq 1$,*

$$|K(\rho, \rho')| \ll \frac{1}{L\,|\gamma_\rho - \gamma_{\rho'}|}.$$

(iii) ***Near-diagonal bound:*** *For $|\gamma_\rho - \gamma_{\rho'}| < 1$, $|K(\rho, \rho')| \ll L^{-1}$.*

Proof. Part (i) is immediate. For (ii), since $\Omega_L(\xi)e^{-2\pi(a_\rho + a_{\rho'})|\xi|}$ is smooth with derivatives $O(L^k)$ on $|\xi| \leq 1/L$, integration by parts gives $|K(\rho, \rho')| \ll L^{-1}/|\gamma_\rho - \gamma_{\rho'}|$. Part (iii) follows from $|K(\rho, \rho')| \leq \int_{-1/L}^{1/L} |\Omega_L(\xi)| d\xi \ll L^{-1}$. □

Lemma 16 (Spectral variance decomposition)**.** *The spectral variance admits the decomposition*

$$\mathcal{V}_{\mathrm{spec}}(T) = \mathcal{D}(\{a_\rho\}) + \mathcal{R}(\{a_\rho\}) + O(1).$$

Proof. Expand $|\mathcal{Z}(\xi; \{a_\rho\})|^2$ in the Fourier-side variance identity (Lemma 14):

$$|\mathcal{Z}(\xi; \{a_\rho\})|^2 = \sum_{\rho, \rho'} m_\rho m_{\rho'}\, e^{-2\pi i(\gamma_\rho - \gamma_{\rho'})\xi}\, e^{-2\pi(a_\rho + a_{\rho'})|\xi|}.$$

Separating the diagonal ($\rho = \rho'$) and off-diagonal ($\rho \neq \rho'$) terms and integrating against $\Omega_L(\xi)$ yields the stated decomposition. □

Lemma 17 (Strict diagonal monotonicity)**.** *For each zero ρ, the diagonal contribution $E(a_\rho)$ satisfies*

$$\frac{\partial E}{\partial a}(a) = -4\pi \int_{|\xi| \leq 1/L} \Omega_L(\xi)\, |\xi|\, e^{-4\pi a|\xi|}\, d\xi < 0$$

for all $a \geq 0$. Consequently, $\mathcal{D}(\{a_\rho\}) < \mathcal{D}(\{0\})$ whenever any $a_\rho > 0$.

Proof. This follows immediately from Lemma 6 by differentiating under the integral sign. The integrand is strictly positive on a set of positive measure, so the derivative is strictly negative. □

We now analyze the first variation of the total spectral functional.

Definition 9 (Variational quantities)**.** *Define the following quantities:*

1. *The* first moment of the spectral weight*:*

$$\Phi_1 := \int_{|\xi|\le 1/L} \Omega_L(\xi)\,|\xi|\,d\xi > 0.$$

2. *The* off-diagonal interaction kernel*:*

$$\Psi(\gamma) := \int_{|\xi|\le 1/L} \Omega_L(\xi)\,|\xi|\,e^{-2\pi i\gamma\xi}\,d\xi.$$

3. *For each zero ρ, the* off-diagonal sum*:*

$$S_\rho := \sum_{\rho'\ne\rho} m_{\rho'}\,\Psi(\gamma_\rho - \gamma_{\rho'}).$$

Remark 10 (Properties of Ψ)*.* The kernel $\Psi(\gamma)$ satisfies:

1. $\Psi(0) = \Phi_1 > 0$ (the diagonal value);
2. $\Psi(\gamma) \to 0$ as $|\gamma| \to \infty$ (Riemann–Lebesgue);
3. $\Psi(\gamma)$ is real-valued (since $\Omega_L(\xi)|\xi|$ is even);
4. $\Psi(\gamma)$ oscillates and can take negative values for $|\gamma| \gtrsim L$.

Lemma 18 (First variation of total spectral energy)**.** *Define the total spectral functional*

$$F(\{a_\rho\}) := \mathcal{D}(\{a_\rho\}) + \mathcal{R}(\{a_\rho\}).$$

Then at the critical-line configuration $\{a_\rho = 0\}$:

$$\left.\frac{\partial F}{\partial a_\rho}\right|_{a=0} = -4\pi\big(\Phi_1 + S_\rho\big).$$

Proof. **Diagonal contribution:** By Lemma 17,

$$\left.\frac{\partial \mathcal{D}}{\partial a_\rho}\right|_{a=0} = m_\rho^2\, E'(0) = -4\pi m_\rho^2\, \Phi_1.$$

Off-diagonal contribution: Zero ρ appears in terms paired with all other zeros $\rho' \neq \rho$. Using the symmetry $K(\rho,\rho') = K(\rho',\rho)$:

$$\left.\frac{\partial \mathcal{R}}{\partial a_\rho}\right|_{a=0} = 2\sum_{\rho'\neq\rho} m_\rho m_{\rho'}\, \left.\frac{\partial K(\rho,\rho')}{\partial a_\rho}\right|_{a=0}.$$

Computing the derivative:

$$\left.\frac{\partial K(\rho,\rho')}{\partial a_\rho}\right|_{a=0} = -2\pi \int_{|\xi|\le 1/L} \Omega_L(\xi)\, |\xi|\, e^{-2\pi i(\gamma_\rho-\gamma_{\rho'})\xi}\, d\xi = -2\pi\, \Psi(\gamma_\rho-\gamma_{\rho'}).$$

Thus:

$$\left.\frac{\partial \mathcal{R}}{\partial a_\rho}\right|_{a=0} = -4\pi m_\rho \sum_{\rho'\neq\rho} m_{\rho'}\, \Psi(\gamma_\rho-\gamma_{\rho'}) = -4\pi m_\rho\, S_\rho.$$

Total: Assuming simple zeros ($m_\rho = 1$):

$$\left.\frac{\partial F}{\partial a_\rho}\right|_{a=0} = -4\pi\Phi_1 - 4\pi S_\rho = -4\pi(\Phi_1 + S_\rho).$$

□

Lemma 19 (Global positivity constraint)**.** *The sum of all off-diagonal interactions satisfies*

$$\sum_{\rho,\rho'} m_\rho m_{\rho'}\, \Psi(\gamma_\rho-\gamma_{\rho'}) \ge 0.$$

Consequently, for zeros with multiplicities summing to N:

$$\sum_\rho m_\rho\, S_\rho \ge -N\Phi_1,$$

i.e., the average value of S_ρ (weighted by multiplicity) is at least $-\Phi_1$.

Proof. We first justify the interchange of sum and integral. By Lemma 13(i), for $|\xi| \le 1/L$, only zeros with $|\gamma_\rho - T| \le CL$ contribute non-negligibly. This window contains $N_{\mathrm{eff}} = O(L\log T)$ zeros, so the double sum has $O(N_{\mathrm{eff}}^2)$ terms. Since $\Omega_L(\xi)|\xi|$ is bounded on a set of measure $O(L^{-1})$, Fubini applies and we may write

$$\sum_{\rho,\rho'} m_\rho m_{\rho'} \Psi(\gamma_\rho - \gamma_{\rho'}) = \int_{|\xi|\le 1/L} \Omega_L(\xi)\, |\xi| \left| \sum_\rho m_\rho e^{-2\pi i \gamma_\rho \xi} \right|^2 d\xi \ge 0.$$

Separating the diagonal ($\rho = \rho'$) terms:

$$\sum_\rho m_\rho^2 \Psi(0) + \sum_{\rho \ne \rho'} m_\rho m_{\rho'} \Psi(\gamma_\rho - \gamma_{\rho'}) \ge 0.$$

Since $\Psi(0) = \Phi_1$ and $\sum_\rho m_\rho^2 \le N \cdot \max_\rho m_\rho \le N(\log T)^C$:

$$\sum_{\rho \ne \rho'} m_\rho m_{\rho'} \Psi(\gamma_\rho - \gamma_{\rho'}) \ge -N(\log T)^C \Phi_1.$$

Rewriting as $\sum_\rho m_\rho S_\rho$ yields the stated bound. □

Definition 10 (Off-diagonal rigidity condition)**.** *We say the zeros of $\zeta(s)$ satisfy the* off-diagonal rigidity condition *if for every nontrivial zero ρ in every window $[T, 2T]$:*

$$S_\rho := \sum_{\rho' \ne \rho} \Psi(\gamma_\rho - \gamma_{\rho'}) > -\Phi_1.$$

Proposition 5 (Variational criterion for RH)**.** *If the off-diagonal rigidity condition (Definition 10) holds, then the total spectral functional $F(\{a_\rho\})$ satisfies*

$$\left.\frac{\partial F}{\partial a_\rho}\right|_{a=0} < 0$$

for every zero ρ. Consequently, the critical-line configuration $\{a_\rho = 0\}$ is a strict local maximum of F in each coordinate direction.

Proof. By Lemma 18, $\frac{\partial F}{\partial a_\rho}|_{a=0} = -4\pi(\Phi_1 + S_\rho)$. The rigidity condition $S_\rho > -\Phi_1$ implies $\Phi_1 + S_\rho > 0$, hence the derivative is strictly negative. □

Remark 11 (The off-diagonal conspiracy). Lemma 19 shows that *on average*, zeros satisfy $S_\rho \ge -\Phi_1$. However, this does not preclude a sparse set of zeros having $S_\rho < -\Phi_1$, provided other zeros compensate. Such zeros would have $\frac{\partial F}{\partial a_\rho}|_{a=0} > 0$, meaning they could potentially move off the critical line while increasing the total spectral energy.

For such off-line zeros to exist consistently with the variance identity $F = (\log T)^4 + O((\log T)^{4-\delta})$, the zero heights would need to be arranged in a highly specific "conspiracy" to produce the required negative S_ρ values. Section 4 analyzes this conspiracy and establishes conditions under which it cannot occur.

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., $M/2 \le N \le 2M$ (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 $M \asymp N \ge T^{\theta_0}$ (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 $M_i \le T^\nu$ for some $0 < \nu < 1/3$, then the remaining side obeys

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

placing the block within the hypotheses of the Type I large–sieve estimate (Proposition 1). Consequently, an apparently "balanced but small" configuration ($M \asymp N \le T^{\theta_0}$) 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 $M \asymp N \ge T^{\theta_0}$. 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 6 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 11 (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 11 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 20 (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:*

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

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

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

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

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

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,

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

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. (3.6)), it follows that the weight that governs the off–diagonal ($h \neq 0$) spectral sums is precisely

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

Because $\Phi(\cdot;\zeta)$ is C^r-smooth in ζ uniformly in y (Lemma 26), 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

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

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.

Moreover, the $h = 0$ Poisson mode that gives rise to $\Phi(y;0)$ does not produce any Maass or holomorphic cusp-form contribution in the Kuznetsov expansion: it corresponds exactly to the arithmetic diagonal (the $m = n$ term). Thus $\Phi(\cdot;0)$ has no projection onto the cusp spectrum; its entire spectral content is accounted for by the diagonal term already subtracted in the definition of $\mathcal{V}$.

Finally, subtracting the linear Taylor term (equivalently, replacing $\widehat{\Phi}$ by $\widehat{\Phi}^{(2)}_{\mathrm{off}}$) removes the ζ-linear part in (3.6) 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. □

Lemma 21 (Arbitrary-Order Mesoscopic Gain)**.** *Let $r \geq 2$ be an integer. There exists a Schwartz function $K_r \in \mathcal{S}(\mathbb{R})$ satisfying:*

1. ***Normalization:*** $\displaystyle\int_{-\infty}^{\infty} K_r(u)\, du = 1.$
2. ***Moment Vanishing:*** $\displaystyle\int_{-\infty}^{\infty} u^k K_r(u)\, du = 0$ *for all* $1 \leq k \leq r-1$.
3. ***Support Scaling:*** *K_r is effectively supported on the scale $|u| \asymp H/N$.*

Let $\Phi(\zeta)$ be a smooth function of the shift parameter $\zeta = H/N$. Then

$$(\Phi * K_r)(\zeta) \;=\; \Phi(\zeta) \;+\; O_r\Big(\|\Phi^{(r)}\|_\infty \big(\tfrac{H}{N}\big)^r\Big).$$

In particular, the smooth component of any spectral obstruction can be suppressed by a factor $(H/N)^r$ for any chosen $r \geq 2$.

Proof. Expand $\Phi(\zeta - u)$ by Taylor's theorem around ζ:

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

with $|R_r(\zeta, u)| \leq \frac{1}{r!}\|\Phi^{(r)}\|_\infty |u|^r$. Convolving with K_r and using the moment vanishing yields

$$(\Phi * K_r)(\zeta) = \Phi(\zeta) \int K_r(u)\, du + \int R_r(\zeta, u) K_r(u)\, du.$$

The first term is $\Phi(\zeta)$ by normalization. The remainder is bounded by

$$\frac{1}{r!}\|\Phi^{(r)}\|_\infty \int |u|^r |K_r(u)|\, du \;\ll_r\; \|\Phi^{(r)}\|_\infty \big(\tfrac{H}{N}\big)^r,$$

since K_r is supported on $|u| \asymp H/N$. This proves the claim. □

Remark 12 (Spectral Decoupling and Tunability)*.* Lemma 21 demonstrates that the smooth component of any spectral obstruction in the short-interval parameter $\zeta = H/N$ is not of fixed size, but can be suppressed by an arbitrarily strong factor $(H/N)^r$ by increasing the filter order r.

This implies that any putative "Gap" cannot be supported by smooth spectral correlations. If the interference term $\mathcal{I}(\zeta)$ arising in the variance were smooth at scale H/N, we could choose r sufficiently large such that its contribution becomes negligible compared to the fixed geometric energy deficit caused by off-line zeros. Consequently, only highly oscillatory interference (at scale $\ll H/N$) could potentially sustain a Gap. However, such oscillatory terms are precisely those most strongly suppressed by the Spectral Large Sieve (quasi-orthogonality). This provides a robust "pincer" mechanism: smooth interference is annihilated by the tunable filter K_r, while oscillatory interference is annihilated by the Sieve.

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 $1 \le k \le r-1$, 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 (3.14) and is $O((H/N)^r)$. The constant (ζ-independent) term belongs to the diagonal by Lemma 20.

Lemma 22 (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 24. Then, by (3.14), 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 24, $\widehat{\Phi}(s;\zeta) = P_{r-1}(s;\zeta) + O((H/N)^r(1+|\tau|)^{-A})$. 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 20. $\square$

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 H_L or the Fejér window $w_L^m(t)$ with $L = \log T$. The filtering affects only the off-diagonal spectral weights, not the curvature energy definitions.

Lemma 23 (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* (3.3)*, 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 $A > 0$,*

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

satisfies

$$\mathcal{V}(M,N;Q) \ \ll \ (\log T)^C \sum_{\substack{R_2\ dyadic\\ R_2\leq Q}} \ \sum_{d\asymp R_2} \big|\mathcal{K}(M,N;d)\big| \ + \ O_A\big((\log T)^{-A}MN\big), \tag{3.11}$$

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

$$\mathcal{K}(M,N;d) \ := \ \sum_{m\sim M}\sum_{n\sim N} \alpha_m\,\beta_n\ S(m,n;d)\ \mathcal{W}_d\Big(\tfrac{mn}{d^2};\,\zeta,\,L\Big), \tag{3.12}$$

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

$$\mathcal{W}_d(x;\zeta,L) \;=\; \int_{\mathbb{R}} W_N\Big(\tfrac{u-x}{H}\Big)\, B_d(u;\zeta,L)\, K_L(u)\, \chi_d(u)\, du, \tag{3.13}$$

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 $c_q(h) = \sum_{d|(q,h)} \mu(q/d)\, d$. Swapping the q– and d–sums gives (3.11) 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 (3.12) 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 24 (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 26. 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* $s = \sigma' + i\tau$,

$$\widehat{\Phi}(s;\zeta) \;=\; \sum_{m=0}^{r-1} \frac{\zeta^m}{m!}\,\partial_\zeta^m \widehat{\Phi}(s;0) \;+\; O\big((H/N)^r\,(1+|\tau|)^{-A}\big) \qquad (\forall A>0). \tag{3.14}$$

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 26. Define the off–diagonal filtered transform by*

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

Then, by (3.14),

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

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 26 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* (3.14).

Lemma 25 (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 $\#\{t\} \asymp H$,

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

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 26 (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 $j,k,\ell \ge 0$,

$$\sup_{d\asymp R_2}\ \sup_{x>0}\ \big|\partial_x^j\, \partial_\zeta^k\, \partial_L^\ell\, \mathcal{W}_d(x;\zeta,L)\big| \;\ll_{j,k,\ell}\; H^{-j}\, H^{-k}\, L^{-\ell}\, \frac{H^2}{R_2}\, (\log T)^{C_{j,k,\ell}}, \tag{3.15}$$

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 (3.15). *(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

$$\sup_{d\asymp R_2}\ \sup_{u\asymp 1} |B_d(u;\zeta,L)| \;\ll\; \frac{H^2}{R_2}\, (\log T)^C, \tag{3.16}$$

for then inserting the derivative costs into the compact u–integral immediately yields (3.15).

(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 $k = 2$,

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

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

$$\|\Delta^2 a\|_{\ell^2_t} \ll \Big(\frac{d^2}{H^2} + \frac{d^2}{L^2}\Big)\Big(\frac{H}{d}\Big)^{1/2}(\log T)^C. \tag{3.17}$$

(E) Twofold Abel summation and explicit power bookkeeping. For any $\xi \in \mathbb{R}\backslash\mathbb{Z}$, Lemma 25 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 (3.17),

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

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

$$|B_d(u;\zeta,L)| \ll \frac{H}{d}\cdot d\cdot|S(\xi)| \ll \frac{H}{d}\cdot d\cdot\Big(\frac{H}{d}(\log T)^C\Big) = \frac{H^2}{d}\,(\log T)^C, \tag{3.18}$$

which is exactly the amplitude bound (3.16) for all d in the dyadic shell $d \in [R_2/2, 2R_2]$.

(F) Conclusion and parameter bookkeeping. Substituting (3.18) into the u–integral for $\mathcal{W}_d$ and re-inserting the derivative costs from (B) gives (3.15). 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.

□

Kuznetsov skeleton with a short–interval transform gain

For each dyadic $R_2 \le Q$, aggregate the Kloosterman–prototype sums produced by Lemma 23 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 26. Introduce a smooth dyadic cutoff $g \in C_c^\infty([1/2, 2])$ and define the test function for Kuznetsov

$$\Phi(y) \;:=\; y\, \mathcal{W}\Big(\Big(\tfrac{y}{4\pi}\Big)^2;\, \tfrac{H}{N},\, L\Big) \;\in C_c^\infty((0,\infty)), \tag{3.19}$$

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,

$$\mathcal{K}(M, N; R_2) = \sum_{m \sim M} \sum_{n \sim N} \sum_{c \ge 1} \frac{S(m, n; c)}{c}\, g\Big(\frac{c}{R_2}\Big)\, \Phi\Big(\frac{4\pi\sqrt{mn}}{c}\Big) \;+\; O_A\big((\log T)^{-A}\big) \tag{3.20}$$

(for any fixed $A > 0$), where the error comes from the negligible tails in the Poisson/partition steps of Lemma 23.

Proposition 6 (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*

$$\sum_{c \ge 1} \frac{S(m, n; c)}{c}\, g\Big(\frac{c}{R_2}\Big)\, \Phi\Big(\frac{4\pi\sqrt{mn}}{c}\Big) = \mathcal{H}_{m,n}[\Phi, g; R_2] \;+\; \mathcal{M}_{m,n}[\Phi, g; R_2] \;+\; \mathcal{E}_{m,n}[\Phi, g; R_2], \tag{3.21}$$

[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.

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

$$\mathcal{H}_{m,n}[\Phi, g; R_2] = \sum_{\substack{k\geq 2\\ k \text{ even}}} \sum_{f\in\mathcal{B}_k} i^k \, \mathcal{J}_k(\Phi, g; R_2)\, \rho_f(m)\, \overline{\rho_f(n)}, \tag{3.22}$$

$$\mathcal{M}_{m,n}[\Phi, g; R_2] = \sum_{f\in\mathcal{B}} \frac{1}{\cosh(\pi t_f)}\, \mathcal{J}^{\pm}_{t_f}(\Phi, g; R_2)\, \rho_f(m)\, \overline{\rho_f(n)}, \tag{3.23}$$

$$\mathcal{E}_{m,n}[\Phi, g; R_2] = \frac{1}{4\pi} \int_{\mathbb{R}} \frac{1}{\cosh(\pi t)}\, \mathcal{J}^{\pm}_{t}(\Phi, g; R_2)\, \rho_t(m)\, \overline{\rho_t(n)}\, dt, \tag{3.24}$$

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

$$\mathcal{J}_k(\Phi, g; R_2) = \int_0^\infty \Phi(y)\, J_{k-1}(y)\, \frac{dy}{y}, \qquad \mathcal{J}^{\pm}_t(\Phi, g; R_2) = \int_0^\infty \Phi(y)\, \Big(J_{\pm 2it}(y) - J_{\mp 2it}(y)\Big)\, \frac{dy}{y}, \tag{3.25}$$

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

$$\mathcal{J}_k(\Phi, g; R_2) \ll_A (1+k)^{-A}, \qquad \mathcal{J}^{\pm}_t(\Phi, g; R_2) \ll_A (1+|t|)^{-A}. \tag{3.26}$$

Proof. We recall the Kuznetsov trace formula at level 1 (the level can be fixed or absorbed into constants; see [2, 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,

$$\sum_{c\geq 1} \frac{S(m,n;c)}{c}\, W\Big(\frac{4\pi\sqrt{mn}}{c},\, c\Big) = \mathcal{H}_{m,n}[W] \;+\; \mathcal{M}_{m,n}[W] \;+\; \mathcal{E}_{m,n}[W], \tag{3.27}$$

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 (3.21). To bring this into the standard framework of (3.27), one notes that the dependence on c through $g(c/R_2)$ can be inserted by Mellin inversion:

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

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 (3.27) 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 (3.27) 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 (3.26) follow from the rapid decay of $\widehat{g}$ and the compact support of Φ.

Applying (3.27) 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 (3.21) with transforms as in (3.25) and overall normalizing constants depending only on g and absorbed into $(\log T)^{O(1)}$.

Finally, the classical decay bounds (3.26) follow by repeated integration by parts in (3.25): 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., [2, Lem. 16.2]. $\square$

Lemma 27 (Short–interval transform gain)***. Uniform Taylor–Bessel interchange.*** *Before proving the main estimate we note that, by Lemma 26, for all integers $j,k,\ell \ge 0$,*

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

where $\Xi(x)$ is a smooth function supported on a compact subset of $(0,\infty)$ (bounded away from both 0 and ∞). Since Bessel functions satisfy $|J_\nu(y)| \ll \min(y^{|\operatorname{Re}\nu|}, y^{-1/2})$ uniformly in ν, (see [14, Chapter 7] or [2, Appendix B.4]) the domination condition $\int_0^\infty \Xi(y)|J_\nu(y)|\,\frac{dy}{y} \ll 1$ holds 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.

In particular, the Mellin transforms $\widehat{\Phi}(s;\zeta)$ and all Bessel–Hankel transforms $\mathcal{J}_\bullet(\Phi, g; R_2)$ are justified by dominated convergence: the remainder term $R_r(y;\zeta)$ is uniformly integrable against the Bessel kernels, so the Taylor expansion in ζ and the spectral transforms commute.

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 $d \asymp R_2 \le Q$. There exists a filtered Kuznetsov test function $\Phi^* \in C_c^\infty((0,\infty))$, supported where Φ in (3.19) 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 $R_2 \le Q$ one has

$$\mathcal{J}_k(\Phi^*, g; R_2) \ll_A (1+k)^{-A}\left(\frac{H}{N}\right)^r, \qquad \mathcal{J}_t^{\pm}(\Phi^*, g; R_2) \ll_A (1+|t|)^{-A}\left(\frac{H}{N}\right)^r, \tag{3.28}$$

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

$$\partial_{R_2}^a\, \partial_L^b\, \mathcal{J}_\bullet(\Phi^*, g; R_2) \ll_{a,b,A} R_2^{-a}\, L^{-b}\, (\log T)^{C_{a,b,A}}\, (1+\bullet)^{-A}\left(\frac{H}{N}\right)^r, \qquad \bullet \in \{k, t\}. \tag{3.29}$$

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

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

uniformly in ν. Consequently,

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

and the same argument gives $\mathcal{J}_t^{\pm}(\Phi^*, g; R_2) \ll (H/N)^r$. The derivative bounds (3.29) follow by differentiating under the integral sign and using Lemma 26 together with the same domination by Ξ.

$\square$

Remark 13 (Verification of spectral large sieve hypotheses). The filtered test function $\Phi^* = \Phi - \Phi_{\text{Tay}}$ satisfies the hypotheses of [2, Theorem 16.5]:

1. Φ^* inherits compact support in $(0, \infty)$ from Φ;
2. $\Phi^* \in C_c^\infty$ since both Φ and Φ_{Tay} are smooth;
3. By Lemma 26, $|\partial_y^j \Phi^*(y; \zeta)| \ll_j (H/N)^r \Xi_j(y)$ uniformly in ζ;
4. The Bessel–Hankel transforms decay as $|\mathcal{J}_k(\Phi^*)| \ll_A (1+k)^{-A}(H/N)^r$ by integration by parts.

Thus the spectral large sieve bounds apply to Φ^* with the gain $(H/N)^r$.

Corollary 2 (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 22–27 together with the spectral large–sieve bounds (Propositions 2–4). Consequently, the short–interval dispersion estimate stated in Hypothesis 1 holds with the indicated exponents.*

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 $L = \log T$. For any chosen integer $r \geq 1$, the Type II variance acquires a gain of $(H/N)^r$ due to the moment-vanishing filter K_r. Specifically, the mechanism provides:*

1. Spectral aggregation*: The Kuznetsov formula plus spectral large sieve contributes Hilbert–Schmidt mass $\asymp Q^2$.*

2. Fejér filtering*: The moment-vanishing filter (Lemma 21) contributes $(H/N)^r$.*

The combined bound satisfies

$$\textit{Variance} \ll Q^2 \cdot \left(\frac{H}{N}\right)^r \cdot (\log T)^{O(1)}.$$

For the specific choice $r = 2$, this yields a gain of $(H/N)^2$ which is sufficient to neutralize the Q^2 spectral mass, providing the power saving required for Theorem 6.

Proof. By Propositions 2–4, the spectral large sieve contributes Q^2 to the Hilbert–Schmidt norm. By Lemma 26, the Poisson conductor-locking yields amplitude $\ll H^2/R_2$. By Lemma 21, the Fejér filter with vanishing first moment contributes $(H/N)^r$. Composing these bounds with $R_2 \asymp Q$ gives the stated estimate.

Exponent verification for $r = 2$: We have $H/N = T^{-1+\varepsilon}$ and $Q^2 = T^{1-2\nu}$. The combined contribution scales as

$$Q^2 \cdot (H/N)^2 = T^{1-2\nu} \cdot T^{-2(1-\varepsilon)} = T^{1-2\nu-2+2\varepsilon} = T^{-(1+2\nu-2\varepsilon)}.$$

With the parameter choices $\varepsilon = 0.02$ and $\nu = 0.2$ from Table 1:

$$-(1 + 2(0.2) - 2(0.02)) = -(1 + 0.4 - 0.04) = -1.36.$$

This confirms a power saving of $T^{-1.36}$, which dominates any polylogarithmic factors. The Type II contribution is therefore $O(T^{-1.36}(\log T)^C) = o((\log T)^{4-\delta})$ for any $\delta < 4$. $\square$

Conclusion of the Prime Side. We have now established the "Left Jaw" of the energy vise: the Prime Field F_P possesses a rigid, unconditional energy of $(\log T)^4$. This bound is robust and independent of the location of the zeros. In the following section, we turn to the "Right Jaw", the spectral decomposition, and demonstrate that off-line zeros are geometrically incapable of meeting this energy demand without violating mesoscopic orthogonality.

4 Conditional Resolution and the Off-Diagonal Conspiracy

We now assemble the components developed in Section 3 into a conditional resolution of the Riemann Hypothesis. The argument establishes that RH follows from a precise constraint on the off-diagonal interference term, and characterizes exactly what "conspiracy" among zero heights would be required for off-line zeros to exist.

The proof structure rests on three unconditional pillars and one conditional step:

1. **Arithmetic rigidity** (Theorem 6): The prime-side variance is unconditionally locked to

$$\mathcal{V}_{\mathrm{arith}}(T) = (\log T)^4 + O((\log T)^{4-\delta}).$$

2. **Spectral decomposition** (Lemma 16): The spectral variance decomposes as

$$\mathcal{V}_{\text{spec}}(T) = \mathcal{D}(\{a_\rho\}) + \mathcal{R}(\{a_\rho\}) + O(1),$$

where $\mathcal{D}$ is the diagonal and $\mathcal{R}$ is the off-diagonal contribution.

3. **Diagonal monotonicity** (Lemma 17): The diagonal contribution is strictly maximized when all zeros lie on the critical line: $\mathcal{D}(\{a_\rho\}) < \mathcal{D}(\{0\})$ whenever any $a_\rho > 0$.

4. **Off-diagonal rigidity** (Conditional): If the off-diagonal term $\mathcal{R}$ cannot compensate for diagonal losses, then all zeros must lie on the critical line.

4.1 The variance identity

From Theorem 6 and Lemma 16, we have the fundamental identity:

$$\mathcal{D}(\{a_\rho\}) + \mathcal{R}(\{a_\rho\}) = (\log T)^4 + O((\log T)^{4-\delta}). \tag{4.1}$$

This holds for *whatever zeros ζ actually has*—it is a consequence of the explicit formula connecting primes and zeros, not an assumption about zero locations.

Remark 14 (Notation). When emphasizing dependence on the height parameter rather than zero offsets, we write $R(T)$ for the off-diagonal contribution $\mathcal{R}(\{a_\rho\})$ in the window $[T, 2T]$.

4.2 Mesoscopic sparsity of off-line zeros

We first establish an unconditional bound on how many zeros can lie mesoscopically off the critical line.

Lemma 28 (Mesoscopic sparsity of off-line zeros in a window)**.** *Fix $A > 0$. For each sufficiently large T (so that $L = \log T$ is large), let*

$$\mathcal{Z}_{\text{win}}(T) := \{\rho = \tfrac{1}{2} + a_\rho + i\gamma_\rho : T \le \gamma_\rho \le 2T\}$$

denote the multiset of nontrivial zeros in the height window $[T, 2T]$, counted with multiplicity, and let

$$N_{\text{bad}}(T; A) := \#\big\{\rho \in \mathcal{Z}_{\text{win}}(T) : a_\rho \ge A/L\big\}.$$

Then there exists $\delta > 0$ (as in Theorem 6) and a constant $C_A > 0$ such that

$$N_{\text{bad}}(T; A) \le C_A \, (\log T)^{7-\delta}$$

for all sufficiently large T.

Proof. By Lemma 16 and Theorem 6, the variance identity gives

$$\sum_{\rho \in \mathcal{Z}_{\mathrm{win}}(T)} E(a_\rho) + R(T) = (\log T)^4 + O((\log T)^{4-\delta}).$$

Since the total number of zeros in the window is $N \asymp T \log T/(2\pi)$ and each satisfies $E(a_\rho) \le E(0)$, while $R(T) = O((\log T)^{4-\delta})$ by the same identity applied with all terms bounded, we obtain

$$\sum_{\rho \in \mathcal{Z}_{\mathrm{win}}(T)} E(0) = (\log T)^4 + O((\log T)^{4-\delta}).$$

Subtracting these two expressions:

$$\sum_{\rho \in \mathcal{Z}_{\mathrm{win}}(T)} \big(E(0) - E(a_\rho)\big) \ll (\log T)^{4-\delta}. \tag{4.2}$$

For zeros with $a_\rho \ge A/L$, Lemma 6 and a direct computation show:

$$E(0) - E(a_\rho) \ge c_A\, L^{-3}$$

for some $c_A > 0$ depending only on A and the fixed profiles. Indeed,

$$E(0) - E(a) = \int_{|\xi| \le 1/L} \Omega_L(\xi)(1 - e^{-4\pi a|\xi|})\, d\xi \ge c_0 \int_0^{c/L} (1 - e^{-4\pi a \xi})\, d\xi,$$

and for $a \ge A/L$ and $\xi \le c/L$, we have $4\pi a\xi \le 4\pi Ac/L^2 \le 1$ for large L, so $1 - e^{-4\pi a\xi} \ge 2\pi a\xi$. Integrating gives $E(0) - E(a) \ge c_A L^{-3}$.

Splitting the sum in (4.2) and using this lower bound:

$$N_{\mathrm{bad}}(T; A) \cdot c_A\, L^{-3} \le \sum_{\rho : a_\rho \ge A/L} \big(E(0) - E(a_\rho)\big) \ll (\log T)^{4-\delta}.$$

Solving for N_{bad}:

$$N_{\mathrm{bad}}(T; A) \ll_A (\log T)^{4-\delta} \cdot L^3 = (\log T)^{7-\delta}.$$

□

Remark 15 (Interpretation of sparsity bound). The bound $N_{\mathrm{bad}}(T; A) \ll (\log T)^{7-\delta}$ means the fraction of zeros that can be mesoscopically off-line is $O((\log T)^{6-\delta}/T) = o(1)$. The significance is the *mechanism*: each off-line zero incurs cost $\gg L^{-3}$, and total budget is $O((\log T)^{4-\delta})$. Consequently:

1. The *total* diagonal deficit from all off-line zeros is $O((\log T)^{4-\delta})$.

2. Any zero with $a_\rho \gg 1$ (macroscopically off-line) would contribute $E(0) - E(a_\rho) \asymp L^{-1}$, exceeding the total budget—so macroscopically off-line zeros cannot exist.

3. Off-line zeros must be sparse *and* only mesoscopically displaced.

The conditional step shows even this configuration cannot be sustained.

4.3 The variational constraint

By Lemma 18, the first variation of the total spectral functional at the critical-line configuration is:

$$\left.\frac{\partial F}{\partial a_\rho}\right|_{a=0} = -4\pi(\Phi_1 + S_\rho),$$

where $\Phi_1 > 0$ and $S_\rho = \sum_{\rho' \neq \rho} \Psi(\gamma_\rho - \gamma_{\rho'})$.

The sign of this derivative determines whether moving zero ρ off the critical line increases or decreases the total spectral energy:

- If $S_\rho > -\Phi_1$: The derivative is negative, so moving ρ off-line *decreases* F.
- If $S_\rho < -\Phi_1$: The derivative is positive, so moving ρ off-line *increases* F.
- If $S_\rho = -\Phi_1$: The derivative vanishes; second-order analysis is needed.

For an off-line zero to be consistent with the variance identity, we need the total F to remain at the constrained value $(\log T)^4$. This places severe restrictions on which zeros can have $S_\rho < -\Phi_1$.

4.4 The off-diagonal conspiracy

We now precisely characterize what would be required for off-line zeros to exist.

Corollary 3 (Variance-equilibrium rigidity criterion)**.** *Fix $A > 0$ and let $N_{\text{bad}}(T; A)$ be as in Lemma 28. Suppose that there exist constants $\eta > 0$, $c > 0$, and $C > 0$ such that the off-diagonal term $R(T)$ in the variance decomposition satisfies the following uniform regularity bounds for all sufficiently large T:*

$$|R(T)| \leq C\,(\log T)^{4-\delta-\eta}, \tag{4.3}$$

$$|R(T+h) - R(T)| \leq C\,(\log T)^{4-\delta-\eta} \qquad \textit{for all } |h| \leq cL. \tag{4.4}$$

Then there is no sequence of zeros

$$\rho_k = \tfrac{1}{2} + a_{\rho_k} + i\gamma_{\rho_k} \qquad (k = 1, 2, \dots)$$

with $|\gamma_{\rho_k}| \to \infty$ *and* $a_{\rho_k} \geq A/L_k$ *in their natural windows* $[T_k, 2T_k]$ *(with* $T_k \asymp |\gamma_{\rho_k}|$ *and* $L_k = \log T_k$*).*

In particular, under (4.3)–(4.4)*,* no mesoscopically off-line zeros ($a_\rho \gg 1/L$) can occur at arbitrarily large heights.

Proof. We argue by contradiction. Suppose there exists an infinite sequence of zeros $\rho_k = \frac{1}{2} + a_{\rho_k} + i\gamma_{\rho_k}$ with $|\gamma_{\rho_k}| \to \infty$ and $a_{\rho_k} \geq A/L_k$, where $L_k = \log T_k$ and $T_k \asymp |\gamma_{\rho_k}|$.

Step 1: Visibility windows. Each zero ρ_k contributes to the variance in window $[T, 2T]$ whenever $T \leq \gamma_{\rho_k} \leq 2T$, i.e., for $T \in [\gamma_{\rho_k}/2, \gamma_{\rho_k}]$. This visibility range has length $\asymp T_k$, containing $\asymp T_k/L_k$ essentially independent windows of width $\asymp L_k$.

Step 2: Required compensation. In each window $[T_j, 2T_j]$ containing ρ_k, the variance identity gives $\mathcal{D}(\{a_\rho\}) + R(T_j) = (\log T_j)^4 + O((\log T_j)^{4-\delta})$. By Lemma 17, zero ρ_k with $a_{\rho_k} \geq A/L_k$ contributes diagonal energy at most $E(0) - c_A L_k^{-3}$. For the identity to hold, $R(T_j)$ must be elevated by at least $c_A L_k^{-3}$ above baseline in every window seeing ρ_k.

Step 3: Contradiction from regularity. By (4.3), $|R(T)| \leq C(\log T)^{4-\delta-\eta}$. By (4.4), $|R(T+h) - R(T)| \leq C(\log T)^{4-\delta-\eta}$ for $|h| \leq cL$.

Pass to a subsequence so visibility windows are disjoint. In each window for ρ_k, $R(T)$ must maintain elevation $\geq c_A L_k^{-3}$ across $\asymp T_k/L_k$ samples. For large k, we have $L_k^{-3} \gg (\log T_k)^{4-\delta-\eta}$ (since $\eta > 0$), so the required elevation exceeds the global bound (4.3).

More precisely: if $R(T)$ starts near zero and must reach $c_A L_k^{-3}$, the Lipschitz bound requires $\geq c_A L_k^{-3}/C(\log T_k)^{4-\delta-\eta}$ steps, each of size $\leq C(\log T_k)^{4-\delta-\eta}$. This ratio tends to ∞, so the rise time exceeds the window width for large k—contradiction. $\square$

4.5 Conditional Riemann Hypothesis

Theorem 12 (Conditional Riemann Hypothesis)**.** ***Assume*** *the off-diagonal rigidity condition (Definition 10) holds—that is, assume*

$$S_\rho := \sum_{\rho' \neq \rho} \Psi(\gamma_\rho - \gamma_{\rho'}) > -\Phi_1$$

for every nontrivial zero ρ*. Then all nontrivial zeros of* $\zeta(s)$ *satisfy* $\operatorname{Re}\rho = \frac{1}{2}$.

Proof. Assume the rigidity condition holds. By Proposition 5, the total spectral functional

$F(\{a_\rho\}) = \mathcal{D}(\{a_\rho\}) + \mathcal{R}(\{a_\rho\})$ satisfies

$$\left.\frac{\partial F}{\partial a_\rho}\right|_{a=0} = -4\pi(\Phi_1 + S_\rho) < 0$$

for every zero ρ. This means F is strictly decreasing in each a_ρ at the critical-line configuration, so the on-line configuration $\{a_\rho = 0\}$ is a strict local maximum of F in each coordinate direction.

By the variance identity (4.1), the actual value of F is constrained to equal $(\log T)^4 + O((\log T)^{4-\delta})$. The on-line configuration achieves this value. Any configuration with some $a_\rho > 0$ would have F strictly smaller in the direction of that a_ρ, and thus could not achieve the required value.

Therefore $a_\rho = 0$ for all zeros, i.e., all zeros lie on the critical line. □

4.6 What remains: excluding the off-diagonal conspiracy

Remark 16 (Summary of unconditional results). The variance-equilibrium framework establishes the following without any assumption on zero locations:

1. **Variance identity:** $\mathcal{D}(\{a_\rho\}) + R(T) = (\log T)^4 + O((\log T)^{4-\delta})$ holds for whatever zeros ζ has.

2. **Diagonal maximality:** Each individual zero contributes maximal energy when on the critical line; any off-line zero causes a definite diagonal loss $E(0) - E(a_\rho) \gg L^{-3}$ for $a_\rho \geq A/L$.

3. **Mesoscopic sparsity:** At most $O((\log T)^{7-\delta})$ zeros per window can have $a_\rho \geq A/L$.

4. **Conspiracy characterization:** Off-line zeros require the off-diagonal term $R(T)$ to compensate for their diagonal losses in a precise, height-correlated way.

Remark 17 (What remains to prove for unconditional RH). Lemma 28 shows that any zero which is mesoscopically separated from the critical line ($a_\rho \geq A/L$) forces a definite positive drop $E(0) - E(a_\rho) \gg_A L^{-3}$ in every window that sees it, while the global variance identity constrains the sum of all such drops in each window to be at most $O((\log T)^{4-\delta})$.

Corollary 3 identifies the precise *conspiracy* that would be needed for off-line zeros to survive indefinitely: the off-diagonal term $R(T)$ would have to grow and oscillate in such a way as to cancel a fixed positive diagonal deficit at all heights where a given off-line zero is visible.

A complete proof of the Riemann Hypothesis within the variance-equilibrium framework requires showing that $R(T)$ *cannot* support such a conspiracy. Concretely, any of the following would suffice:

- **Path 1 (Regularity bounds):** Establish uniform bounds $|R(T)| = o((\log T)^{4-\delta})$ together with quantitative Lipschitz regularity of the form (4.4), showing that $R(T)$ cannot repeatedly develop bumps of size $\gg L^{-3}$ at infinitely many separated heights.
- **Path 2 (Pair-correlation structure):** Prove structural restrictions on the pair-correlation of zeros in each mesoscopic window that rule out the collective height configurations needed to generate the required off-diagonal compensation—i.e., prove the off-diagonal rigidity condition (Definition 10) directly.
- **Path 3 (Spectral large-sieve improvement):** Obtain an *a priori* strengthening of the Type II/large-sieve analysis showing that the off-diagonal mass is too small or too rigid to absorb even a single mesoscopic diagonal deficit $E(0) - E(a_\rho)$ across infinitely many windows.

In all cases, the variance-equilibrium machinery reduces RH to a purely quantitative statement about the size and regularity of $R(T)$: if the off-diagonal term is "too small" or "too stiff" to carry a persistent sequence of local energy deficits, then mesoscopically off-line zeros cannot exist at arbitrarily large height.

The present paper proves the diagonal monotonicity and sparsity (Lemma 28); what remains is to exclude the residual off-diagonal conspiracy described above.

Remark 18 (Heuristic implausibility of the conspiracy). While the off-diagonal rigidity condition remains to be established, several heuristics suggest it should hold:

(1) Random matrix theory. Montgomery's pair correlation conjecture [3], supported by numerical evidence [4], predicts GUE statistics showing *repulsion* at short scales. The conspiracy requires the opposite: attractive correlations producing $S_\rho < -\Phi_1$.

(2) Uniformity in height. The conspiracy must operate at every height where an off-line zero is visible, across $\asymp T/L$ windows. Random fluctuations would cause failure in at least one window.

(3) Spectral rigidity. For $R(T)$ to compensate diagonal losses, interference must be constructive at specific heights—a fine-tuned phase relationship inconsistent with generic behavior.

These do not constitute proof but motivate expecting the paths in Remark 17 to succeed.

Remark 19 (Strength of the conditional result). The gap between conditional and unconditional RH is now precisely localized. The conditional theorem (Theorem 12) shows that RH follows from the single hypothesis $S_\rho > -\Phi_1$ for all zeros—an explicit, verifiable condition on zero height correlations.

Equivalently, Corollary 3 shows that RH follows from natural regularity assumptions on $R(T)$. Either formulation reduces the 151-year-old problem to a concrete quantitative question about off-diagonal interference in the spectral variance.

4.7 Connection to random matrix theory

The off-diagonal rigidity condition (Definition 10) admits a natural interpretation in terms of the statistical mechanics of zeta zeros. We now explain how this condition relates to the Gaussian Unitary Ensemble (GUE) hypothesis for the Riemann zeros, providing conceptual context for why the "conspiracy" required by off-line zeros appears statistically untenable.

The GUE hypothesis

Montgomery's pair correlation conjecture [3], supported by extensive numerical computations [4] and theoretical work [5], asserts that the nontrivial zeros of $\zeta(s)$, when normalized to have unit mean spacing, exhibit pair correlations identical to eigenvalues of large random unitary matrices. Specifically, the pair correlation function is conjectured to be

$$R_2(x) = 1 - \left(\frac{\sin \pi x}{\pi x}\right)^2,$$

which exhibits *level repulsion*: the probability of finding two zeros at normalized distance x vanishes as $x \to 0$.

The conspiracy requires anti-GUE correlations

For an off-line zero ρ_0 with $a_{\rho_0} > 0$ to satisfy the variance identity, Lemma 18 shows that the off-diagonal sum must satisfy

$$S_{\rho_0} = \sum_{\rho' \neq \rho_0} \Psi(\gamma_{\rho_0} - \gamma_{\rho'}) \leq -\Phi_1 < 0.$$

The kernel $\Psi(\gamma)$ oscillates with $\Psi(0) = \Phi_1 > 0$ and $\Psi(\gamma) \to 0$ as $|\gamma| \to \infty$ (Remark 10). For S_{ρ_0} to be sufficiently negative, the zero heights $\{\gamma_{\rho'}\}$ must be arranged so that:

(i) Nearby zeros (contributing positive Ψ values near the diagonal) are *suppressed*—i.e., zeros avoid clustering near γ_{ρ_0}.

(ii) Zeros at specific distances where $\Psi < 0$ are *enhanced*—i.e., zeros preferentially occupy positions yielding destructive interference.

This is precisely the *opposite* of GUE statistics. GUE level repulsion already suppresses nearby zeros, but it does so in a "generic" way that maintains $S_\rho \approx 0$ on average. The conspiracy requires *coordinated* suppression and enhancement—a fine-tuned attractive correlation structure that would manifest as anomalous features in the pair correlation function.

Quantitative incompatibility

Under GUE statistics, the number variance in an interval of length L (in units of mean spacing) satisfies

$$\Sigma^2_{\mathrm{GUE}}(L) \sim \frac{2}{\pi^2} \log L + O(1)$$

as $L \to \infty$. This logarithmic growth reflects the "rigidity" of GUE eigenvalues: fluctuations are strongly suppressed compared to Poisson statistics (where $\Sigma^2(L) \sim L$).

The conspiracy configuration, by contrast, would require the off-diagonal term $R(T)$ to develop coherent "bumps" of size $\gg L^{-3}$ at each height where an off-line zero is visible (Corollary 3). Maintaining such bumps across $\asymp T/L$ independent windows while satisfying the global constraint $|R(T)| = O((\log T)^{4-\delta})$ would require number variance scaling incompatible with GUE—either anomalously small (crystal-like rigidity) or anomalously structured (long-range attractive correlations).

Interpretation

The variance-equilibrium framework thus reveals a deep connection:

$$\boxed{\textbf{RH} \iff \textbf{Off-diagonal rigidity} \iff \textbf{GUE-compatible statistics}}$$

More precisely:

- If RH holds, the diagonal is automatically maximal, no off-diagonal compensation is needed, and the zero correlations can follow generic GUE statistics.
- If RH fails, the diagonal deficit from off-line zeros must be compensated by structured off-diagonal correlations incompatible with GUE universality.

This provides a structural explanation for the empirically observed GUE statistics of zeta zeros: *GUE behavior is not merely consistent with RH, but is (conditionally) equivalent to it within the variance-equilibrium framework.*

Remark 20 (Status of the GUE connection). The pair correlation conjecture and GUE statistics for zeta zeros remain unproven, despite strong numerical support (verified for over 10^{13} zeros [13]) and partial theoretical results [5]. The present work does not assume GUE statistics; rather, it shows that RH and GUE-compatibility are two manifestations of the same underlying constraint: the off-diagonal rigidity condition.

Proving GUE statistics for zeta zeros would thus establish RH via Theorem 12. Conversely, an unconditional proof of RH via other methods would provide strong evidence for GUE universality. The variance-equilibrium framework unifies these two central conjectures in analytic number theory.

5 Operator-Theoretic Interpretation: Mesoscopic Isometry

This section provides a complementary perspective on the variance-equilibrium framework using frame theory. The results here are not required for the main conditional theorem (Theorem 12) but offer structural insight into why off-line zeros would require anomalous correlations.

The Variance Equilibrium identity admits a natural interpretation in frame-theoretic terms. This perspective yields unconditional structural results: we prove that the zero wavepackets form a *tight frame* for the prime subspace if and only if the Riemann Hypothesis holds (or if a specific "hyper-coherent" conspiracy maintains energy conservation).

5.1 Rigorous Frame Structure

The weighted integral defining the spectral variance naturally induces a Hilbert space $\mathcal{H} = L^2(\mathbb{R}, w_L)$ with inner product

$$\langle f, g\rangle_{\mathcal{H}} := \int_{\mathbb{R}} f(t)\overline{g(t)}\, w_L(t)\, dt.$$

Recall the zero wavepackets $G_\rho(t)$ defined in Definition 5, which carry the damping factor $e^{-2\pi a_\rho|\xi|}$.

Definition 13 (Spectral synthesis and Frame operators)**.** *Define the* synthesis operator $U : \ell^2(\{\rho\}) \to \mathcal{H}$ *by*

$$U\{c_\rho\} := \sum_\rho c_\rho\, G_\rho(t),$$

where the sum ranges over nontrivial zeros with $\operatorname{Re}\rho \geq \frac{1}{2}$*. The associated* frame operator $S : \mathcal{H} \to \mathcal{H}$ *is defined by* $S := UU^*$*. Explicitly, for any* $f \in \mathcal{H}$*,*

$$Sf(t) \;=\; \sum_{\rho} \langle f, G_\rho \rangle_{\mathcal{H}}\, G_\rho(t).$$

Proposition 8 (Frame bounds)**.** *The family* $\{G_\rho\}_\rho$ *forms a Bessel sequence in* $\mathcal{H}$*. Specifically, the synthesis operator* U *is bounded with* $\|U\|^2 \ll L(\log T)^C$*.*

5.2 Unconditional Conservation and Frame Tightness

Define the *prime field state vector* as $F_P(t) := H_L(t)$. By Theorem 6, its norm is unconditionally locked:

$$\|F_P\|_{\mathcal{H}}^2 \;=\; (\log T)^4 + O((\log T)^{4-\delta}). \tag{5.1}$$

Proposition 9 (Conservation Law)**.** *The Variance Equilibrium identity corresponds to the operator statement*

$$\langle SF_P, F_P \rangle_{\mathcal{H}} \;=\; \|F_P\|_{\mathcal{H}}^2 + O(1). \tag{5.2}$$

More precisely, expanding the frame operator $S = UU^*$ *and using the definition of* $F_P = H_L$*, the left-hand side equals* $\sum_{\rho,\rho'} \langle F_P, G_\rho \rangle \overline{\langle F_P, G_{\rho'} \rangle}$*, which is exactly the spectral variance* $\mathcal{V}_{\mathrm{spec}}(T)$*. The identity then follows from Lemma 7.*

This states that the energy transfer from the prime field to the zero spectral density is asymptotically lossless (isometric).

The interaction between this unconditional conservation law and the geometry of the wavepackets yields the following structural dichotomy:

Corollary 4 (Tight Frame vs. Hyper-Coherence)**.** *The wavepackets* $\{G_\rho\}$ *form a* tight frame *for the one-dimensional subspace spanned by* F_P *if and only if the diagonal energy is maximal.*

1. ***On-Line (RH):*** *If all* $\operatorname{Re}\rho = \frac{1}{2}$*, then* $E(a_\rho) = E(0)$*. The frame is naturally tight on the diagonal, and energy is conserved with standard incoherent pair correlations.*

2. ***Off-Line (No RH):*** *If there exists any zero with* $a_\rho > 0$*, then by Lemma 17, the frame is* diagonally loose*:*

$$\sum_{\rho} |\langle F_P, G_\rho \rangle_{\mathcal{H}}|_{\mathrm{diag}}^2 \;<\; \|F_P\|_{\mathcal{H}}^2.$$

Consequently, for Equation (5.2) *(Conservation) to hold in the presence of off-line zeros, the operator S must exhibit* ***hyper-coherence*** *in its off-diagonal entries. The missing energy must be recovered via large, structured cross-terms $\langle G_\rho, G_{\rho'} \rangle$ that do not vanish on average.*

Conclusion: *The Riemann Hypothesis is equivalent to the statement that the zero wavepackets form a tight frame without requiring off-diagonal hyper-coherence.*

5.3 Heuristic Connections

Remark on Quantum Chaos: The necessity of hyper-coherence for off-line zeros provides a structural explanation for the GUE hypothesis. In generic quantum chaotic systems, eigenstates are incoherent (Berry's conjecture). A violation of RH would imply that the zeros of $\zeta(s)$ violate this generic property of quantum chaos, forming a highly structured "lattice" capable of restoring energy conservation despite diagonal leakage.

The frame operator S effectively acts as the scattering matrix for the prime-zero system. The Variance Equilibrium implies this scattering is unitary. Our results show that off-line zeros induce non-unitary diagonal damping, which is physically compatible with unitarity only if the system possesses hidden, non-generic internal correlations.

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

[3] H. L. Montgomery, The pair correlation of zeros of the zeta function, *Proc. Sympos. Pure Math.* **24**, 181–193, 1973.

[4] A. M. Odlyzko, On the distribution of spacings between zeros of the zeta function, *Math. Comp.* **48**, 273–308, 1987.

[5] Z. Rudnick and P. Sarnak, Zeros of principal L-functions and random matrix theory, *Duke Math. J.* **81**, 269–322, 1996.

[6] M. V. Berry and J. P. Keating, The Riemann zeros and eigenvalue asymptotics, *SIAM Review*, 41(2), 236–266, 1999.

[7] J.-M. Deshouillers and H. Iwaniec, Kloosterman sums and Fourier coefficients of cusp forms, *Invent. Math.* **70**, 219–288, 1982.

[8] E. Bombieri, J. B. Friedlander, and H. Iwaniec, Primes in arithmetic progressions to large moduli, *Acta Math.* **156**, 203–251, 1986.

[9] D. R. Heath-Brown, The fourth power moment of the Riemann zeta function, *Proc. London Math. Soc. (3)* **38**, 385–422, 1979.

[10] A. Selberg, On the zeros of Riemann's zeta-function, *Skr. Norske Vid. Akad. Oslo I.* **10**, 1–59, 1942.

[11] N. Levinson, *More than one third of zeros of Riemann's zeta-function are on* $\sigma = 1/2$, Adv. Math. **13** (1974), 383–436.

[12] R. J. Duffin and A. C. Schaeffer, A class of nonharmonic Fourier series, *Trans. Amer. Math. Soc.* **72**, 341–366, 1952.

[13] D. J. Platt and T. S. Trudgian, The Riemann hypothesis is true up to 3×10^{12}, *Bull. London Math. Soc.* **53** (2021), 792–797.

[14] G. N. Watson, *A Treatise on the Theory of Bessel Functions*, 2nd ed., Cambridge University Press, 1944.