

Article

Not peer-reviewed version

On the Hughes–Keating–O’Connell Conjecture: Entropy-Sieve Methods for Negative Moments of $\zeta'(\rho)$

Rafik Zeraoulia *

Posted Date: 4 September 2025

doi: 10.20944/preprints202509.0372.v1

Keywords: riemann zeta function; dirichlet polynomials; entropy bounds; cumulant factorization; negative moments



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a Creative Commons CC BY 4.0 license, which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.

Article

On the Hughes–Keating–O’Connell Conjecture: Entropy-Sieve Methods for Negative Moments of $\zeta'(\rho)$

Rafik Zeraoulia 

Acoustics and Civil Engineering Laboratory, Faculty of Material Sciences and Computer Science, Mathematics Department, Khemis Miliana University, Theniet el Had Street, Khemis Miliana (44225), Algeria; zeraoulia@univ-dbk.m.dz

Abstract

We investigate the negative discrete moments of the derivative of the Riemann zeta function at its nontrivial zeros, focusing on the Hughes–Keating–O’Connell conjecture. Building on the earlier frameworks of Gonek, Milinovich–Ng, Kirila, and the recent breakthrough of Bui–Floreana–Milinovich, we introduce a hybrid *entropy–sieve method* (ESM). This method refines Dirichlet-polynomial approximations by quantifying entropy of local distributions of $D_X(\gamma)$ and controlling contributions from both small gaps and low-entropy blocks. Assuming the Riemann Hypothesis and standard pair-correlation conjectures, we prove the near-optimal conditional upper bound

$$J_{-1}(T) = \sum_{0 < \gamma \leq T} \frac{1}{|\zeta'(\rho)|^2} \ll T(\log T)^\varepsilon.$$

This matches, up to logarithmic factors, the conjectured order $J_{-1}(T) \asymp T$, improving upon previous conditional bounds in the literature. Our approach complements the sieve and moment methods of Bui–Floreana–Milinovich and the entropy-based large deviation heuristics of Harper, while introducing new tools such as a uniform Dirichlet-polynomial approximation with explicit coefficients and quantitative entropy-decay estimates. Beyond these results, the ESM framework highlights the utility of entropy techniques in analytic number theory, suggesting applications to related problems in L -function theory and random matrix models.

Keywords: riemann zeta function; dirichlet polynomials; entropy bounds; cumulant factorization; negative moments

For the reader’s convenience, we summarize the main notation that will be used consistently throughout the paper. Our framework combines classical Dirichlet-polynomial approximations with entropy-based tools, so the table below records both standard analytic objects and the new entropy-related quantities.

Table 1. Summary of notation used throughout the paper.

General Notation	
T	Height parameter for critical zeros of $\zeta(s)$; we consider zeros $\rho = \frac{1}{2} + i\gamma$ with $0 < \gamma \leq T$, counted with multiplicity.
$N(T)$	Number of zeros $\rho = \frac{1}{2} + i\gamma$ with $0 < \gamma \leq T$.
ρ	A nontrivial zero of $\zeta(s)$, written as $\rho = \frac{1}{2} + i\gamma$.
$\mathcal{E}_{\text{app}}$	Exceptional set of zeros where the Dirichlet-polynomial approximation fails (Lemma 7).
$\mathcal{G}$	Set of “good” zeros: $\gamma \notin \mathcal{E}_{\text{app}}$ and outside any exceptional sieve/entropy set.

Table 1. Cont.

Dirichlet Polynomial Approximation	
X	Length of Dirichlet polynomial; throughout we take $X = T^\alpha$ with small fixed $0 < \alpha < 1/100$.
$D_X(\gamma)$	Main Dirichlet-polynomial approximant: $D_X(\gamma) = \Re \sum_{n \leq X} a_n n^{-1/2-i\gamma}$.
a_n	Dirichlet polynomial coefficients derived from the smoothed explicit formula .
$R_X(\gamma)$	Remainder term in Dirichlet-polynomial approximation of $\log \zeta'(\rho) $.
σ_X^2	Variance of $D_X(\gamma)$: $\sigma_X^2 = \sum_{n \leq X} \frac{ a_n ^2}{n} \sim \log \log T$ (Lemma 1).
Moment Generating Function & Tail Estimates	
$\mathcal{M}(t)$	Moment generating function: $\mathcal{M}(t) = \frac{1}{N(T)} \sum_{0 < \gamma \leq T} e^{tD_X(\gamma)}$.
κ_r	r -th cumulant of $D_X(\gamma)$, defined by $\log \mathcal{M}(t) = \sum_{r \geq 1} \kappa_r t^r / r!$.
t_0	Admissible range of t for MGF bounds: $t_0 = c / \sqrt{\log \log T}$.
$N_-(V; T)$	Lower-tail counting function: $N_-(V; T) = \#\{\gamma \leq T : -\log \zeta'(\rho) \geq V\}$.
V	Threshold parameter controlling the size of $-\log \zeta'(\rho) $ in tail estimates.
Entropy-Sieve Framework	
$\mathcal{G}(\gamma_0)$	Local window of zeros near γ_0 used for entropy sampling.
$H_{h,\Delta}^{\text{val}}(\gamma_0)$	Local value-entropy of $D_X(\gamma)$ in a window $\mathcal{G}(\gamma_0)$ with bin-width h (Definition ??).
$H_{h,\Delta}^{\text{gap}}(\gamma_0)$	Local gap-entropy of normalized zero spacings near γ_0 .
H_0	Entropy threshold; zeros with entropy below H_0 belong to the exceptional low-entropy set.
$\mathcal{E}_{\text{ent}}$	Exceptional set of zeros lying in low-entropy regions (see Lemma 3).
Moments and Sieve	
$J_k(T)$	Discrete moment: $J_k(T) = \sum_{0 < \gamma \leq T} \zeta'(\rho) ^{2k}$, defined for all $k \in \mathbb{R}$; for $k < 0$, finiteness implies no multiple zeros.
$J_k^{\text{simp}}(T)$	Same sum restricted to simple zeros (used in intermediate lemmas for clarity).
c, C_0	Absolute positive constants appearing in Gaussian and sieve bounds.

1. Introduction

Let $\zeta(s)$ denote the Riemann zeta function and $\rho = \frac{1}{2} + i\gamma$ denote its nontrivial zeros. The size of the derivative $\zeta'(\rho)$ at these zeros plays a central role in analytic number theory, with deep connections to the distribution of zeros, random matrix theory, and the moments of L -functions. For $k \in \mathbb{C}$, we define the discrete moment

$$J_k(T) := \sum_{0 < \gamma \leq T} |\zeta'(\rho)|^{2k},$$

where the sum runs over all nontrivial zeros ρ of $\zeta(s)$, counted with multiplicity. For $k < 0$ this sum is finite only if every zero is simple, since a multiple zero would satisfy $\zeta'(\rho) = 0$ and force $J_k(T) = +\infty$. Thus, proving upper bounds for $J_k(T)$ in the negative range has direct implications for the simplicity of zeros.

1.1. Motivation and Conjectures

Understanding the asymptotic growth of $J_k(T)$ has been the subject of considerable research. Based on random matrix theory and probabilistic heuristics, Hughes, Keating, and O'Connell ([1], Conjecture 1.7, p. 5) conjectured that for $\Re(k) > -\frac{3}{2}$,

$$J_k(T) \sim \frac{G^2(2+k)}{G(3+2k)} a(k) \frac{T}{2\pi} \left(\log \frac{T}{2\pi} \right)^{(k+1)^2}, \quad (1)$$

where $G(\cdot)$ is the Barnes G -function and $a(k)$ is an explicit arithmetic factor. In particular, for $k = -1$, conjecture (1) predicts

$$J_{-1}(T) \asymp T,$$

so the negative second moment is expected to be of the same order as the number of zeros up to height T .

1.2. State of the Art

For positive moments ($k \geq 0$), significant progress has been achieved:

- Gonek ([2], p. 35) initiated the study of discrete moments of $\zeta'(\rho)$ and derived asymptotic formulas for $J_1(T)$ under the Riemann Hypothesis (RH).
- Hejhal ([3], Section 3, pp. 343–370) studied the distribution of $\log |\zeta'(\rho)|$ and showed that it behaves approximately like a Gaussian with variance $\asymp \log \log T$, providing the foundation for later probabilistic approaches.
- Kirila ([4], Theorem 1.1, pp. 2–4) obtained sharp upper bounds for positive moments by adapting Harper's Dirichlet-polynomial techniques to sums over zeros:

$$J_k(T) \ll_k N(T) (\log T)^{k(k+2)},$$

where $N(T)$ denotes the number of zeros up to height T .

- Harper's probabilistic method ([7], pp. 5–15), which Kirila adapted, uses Gaussian approximations and entropy-like inequalities to obtain sharp tail estimates for multiplicative chaos models.

These results match the predictions of the Hughes–Keating–O'Connell conjecture for $k > 0$.

For negative moments ($k < 0$), however, far less is known:

- Gonek ([2], p. 36) derived conditional lower bounds for $J_k(T)$ when $k < 0$ but did not provide upper bounds.
- Milinovich and Ng ([5], pp. 642–644) improved certain lower bounds for negative moments, using refined estimates of $\zeta'(\rho)$ in terms of the spacing of zeros.
- Recently, Bui, Florea, and Milinovich ([18], Theorem 1.3, pp. 3–6) obtained conditional upper bounds for negative moments over a large *subfamily* of zeros, excluding a sparse exceptional set where $\zeta'(\rho)$ may be abnormally small. However, a full unconditional upper bound for $J_k(T)$ when $k < 0$ remains open.

1.3. Challenges for Negative Moments

The difficulty in establishing upper bounds for $J_k(T)$ when $k < 0$ stems from controlling the contribution of zeros where $\zeta'(\rho)$ is exceptionally small. Since

$$J_{-1}(T) = \sum_{0 < \gamma \leq T} \frac{1}{|\zeta'(\rho)|^2},$$

the dominant contribution arises from rare events in which $\zeta'(\rho)$ is unusually tiny. Hejhal's model ([3], Section 3) suggests that $\log |\zeta'(\rho)|$ behaves like a Gaussian with variance $\asymp \log \log T$, implying

that very small derivatives are exponentially rare. However, making this rigorous for sums over zeros requires two ingredients:

- Sharp Gaussian-type tail bounds for $\log |\zeta'(\rho)|$, obtained by approximating it with a short Dirichlet polynomial and applying entropy-based large-deviation methods ([7], pp. 5–20).
- Control over the set of exceptional zeros where the approximation fails or where $\zeta'(\rho)$ is extremely small, addressed via sieve-theoretic exclusion techniques as in ([18], Section 6).

1.4. Our Approach and Contributions

In this paper, we propose a hybrid analytic–probabilistic framework to tackle the upper bound for $J_k(T)$ when $k < 0$, combining three main ingredients:

- **Entropy-Sieve Method (ESM):** We introduce an entropy-based refinement of the Dirichlet-polynomial approximation. By quantifying the entropy of local distributions of $D_X(\gamma)$ values and zero gaps, we ensure that low-entropy regions form a negligible exceptional set. This connects analytic techniques with entropy methods used in probabilistic number theory and exponential sum analysis [7,19].
- **Sieve methods for exceptional zeros:** Building on Bui, Florea, and Milinovich ([18], Section 6), we remove a negligible set of zeros where $\zeta'(\rho)$ is abnormally small, using pair-correlation and independence heuristics to bound their contribution. Our systematic discussion of parameter optimization (see Section 4) clarifies how A, B, C, α can be tuned so that both $\mathcal{E}_{\text{app}}$ and $\mathcal{E}_{\text{ent}}$ are negligible.
- **Algorithmic tail truncation:** We develop an entropy-driven tail-truncation procedure to efficiently control the extreme lower tail of $\zeta'(\rho)$, ensuring that these rare events contribute less than any power of $\log T$.

Using these tools, we establish, under RH and mild orthogonality hypotheses, the conditional bound

$$J_{-1}(T) \ll T(\log T)^\varepsilon,$$

which matches the conjectured order up to a logarithmic factor. Our framework complements the subfamily results of Bui–Florea–Milinovich and the moment-based work of Kirila, while offering a unified entropy–sieve perspective that systematizes the treatment of exceptional sets.

1.5. Organization

We briefly summarize the logical structure of the paper. Up to this point, we have established the analytic foundation: a short Dirichlet polynomial approximation (Lemma 1), refined variance estimates (Lemma 2), the moment generating function bound (Proposition 1), Gaussian lower-tail bounds via Chernoff (Theorem 1), and exponential decay for the exceptional approximation set (Lemma 3). These ingredients will be combined with entropy and sieve methods to treat the negative moments of $\zeta'(\rho)$ in a consistent framework, avoiding the divergences that arise if exceptional sets are not carefully controlled.

The remainder of the paper is organized as follows. Section 2 reviews previous results on positive and negative moments of $\zeta'(\rho)$, with particular emphasis on the conjectural framework of Hughes–Keating–O’Connell. In Section 4 we introduce the *Entropy–Sieve Method (ESM)*, which strengthens Dirichlet-polynomial approximations of $\log |\zeta'(\rho)|$ by incorporating entropy-based regularity, and thereby yields robust Gaussian large-deviation bounds. Section 5 develops the sieve-theoretic component, excluding low-entropy or small-gap exceptional sets where $\zeta'(\rho)$ could be abnormally small. Finally, Section 6.9 combines these analytic and probabilistic tools to establish conditional upper bounds for

$$J_k(T) = \sum_{0 < \gamma \leq T} \frac{1}{|\zeta'(\frac{1}{2} + i\gamma)|^{2k}}$$

in the critical range $k < 0$, including the key case $k = -1$.

Main Results

- **Entropy–Sieve Framework.** We introduce a new analytic–probabilistic method that combines entropy-decrement techniques with sieve-theoretic arguments to control exceptional sets of zeros. This framework provides a novel approach to bounding negative moments of $\zeta'(\rho)$.
- **Conditional Upper Bound for Negative Moments.** Assuming the Riemann Hypothesis and standard pair-correlation conjectures, we prove the near-optimal bound

$$J_{-1}(T) \ll T(\log T)^\varepsilon,$$

for any $\varepsilon > 0$, in agreement with the Hughes–Keating–O’Connell conjecture up to logarithmic factors.

- **Asymptotic Simplicity of Zeros in High-Entropy Blocks (Theorem 2).** Under RH and uniform cumulant/MGF bounds, the proportion of multiple zeros within long blocks tends to zero as $T \rightarrow \infty$. Hence, all but $o(N(T))$ zeros of the Riemann zeta function are simple.
- **Joint MGF Bounds (Proposition 3).** The mixed moment generating function of Dirichlet approximants admits a uniform Gaussian bound with covariance Σ_X , up to cubic error terms in $(\log \log T)^{3/2}$.
- **Numerical and Structural Evidence.** Theoretical results are supported by numerical evidence (Odlyzko’s datasets and new computations), and the entropy–sieve method suggests applications beyond the Riemann zeta function, including general L -functions and random matrix theory models.

2. Background

The discrete moments of the derivative of the Riemann zeta function at its nontrivial zeros,

$$J_k(T) = \sum_{0 < \gamma \leq T} |\zeta'(\rho)|^{2k},$$

are central objects in analytic number theory. They provide insight into the distribution of $\zeta'(\rho)$, the spacing of the nontrivial zeros of $\zeta(s)$, and the connections between the zeta function and random matrix theory. Understanding the asymptotic growth of $J_k(T)$ has been the subject of extensive research over the past decades and is closely connected with one of the most refined conjectures in this area: the Hughes–Keating–O’Connell conjecture.

2.1. The Hughes–Keating–O’Connell Conjecture

Motivated by random matrix theory and probabilistic models, Hughes, Keating, and O’Connell proposed an explicit formula for $J_k(T)$ in the regime $\Re(k) > -\frac{3}{2}$. Their conjecture predicts that

$$J_k(T) \sim \frac{G^2(2+k)}{G(3+2k)} a(k) \frac{T}{2\pi} \left(\log \frac{T}{2\pi} \right)^{(k+1)^2}, \quad (2)$$

where $G(\cdot)$ denotes the Barnes G -function and $a(k)$ is an explicit arithmetic factor arising from the Euler product.

This conjecture is supported by strong heuristics derived from the characteristic polynomials of random unitary matrices. In these models, $\log |\zeta'(\rho)|$ behaves approximately like a Gaussian random variable, and Formula (2) reflects the matching asymptotics between the number-theoretic and random-matrix frameworks. A striking consequence appears when setting $k = -1$, where the conjecture predicts

$$J_{-1}(T) \asymp T.$$

Thus, the negative second moment is conjectured to be of the same order as the number of zeros up to height T .

2.2. Positive Moments

The case of positive moments, $k \geq 0$, is relatively well understood and has seen substantial progress over the last four decades. Gonek ([2], Theorem 1, p. 35) pioneered the study of discrete moments of $\zeta'(\rho)$, proving under the Riemann Hypothesis that for $k = 1$,

$$J_1(T) \sim \frac{T}{24\pi} \left(\log \frac{T}{2\pi} \right)^4.$$

This result agrees with the prediction of (2) when $k = 1$ and represented one of the earliest confirmations of the conjecture in a special case.

Hejhal ([3], Section 3, Theorem 3.1, pp. 343–370) advanced the probabilistic understanding of $\zeta'(\rho)$ by studying the distribution of $\log |\zeta'(\rho)|$. He showed that, heuristically, $\log |\zeta'(\rho)|$ behaves approximately like a Gaussian random variable with variance $\sigma^2 \asymp \log \log T$. This probabilistic model suggested that extremely large or small values of $\zeta'(\rho)$ are exponentially rare and laid the conceptual foundation for later entropy-based methods.

A major breakthrough came from Harper ([7], Theorem 2.1, pp. 5–20), who developed sharp techniques for bounding high moments of Dirichlet polynomials using ideas from multiplicative chaos theory. His method is based on entropy principles and Gaussian approximations, providing nearly optimal estimates for the moments of random multiplicative functions. Building on Harper's framework, Kirila ([4], Theorem 1.1, pp. 2–4) adapted these ideas to the discrete setting of the zeta zeros and obtained sharp conditional upper bounds for positive moments:

$$J_k(T) \ll_k N(T) (\log T)^{k(k+2)} \quad (k > 0),$$

where $N(T)$ denotes the number of zeros up to height T . These results are fully consistent with the random matrix predictions of the Hughes–Keating–O'Connell conjecture, providing strong evidence in favor of (2) for $k > 0$.

2.3. Negative Moments

In stark contrast to the positive regime, the behavior of $J_k(T)$ for negative k remains largely mysterious. The primary challenge stems from the fact that negative moments are dominated by the contribution of zeros ρ where $|\zeta'(\rho)|$ is extremely small. Controlling this contribution requires strong bounds on the lower tail of $\log |\zeta'(\rho)|$, a problem that has resisted classical techniques.

Early work by Gonek ([2], Theorem 2, p. 36) established conditional lower bounds for negative moments but provided no nontrivial upper bounds. Later, Milinovich and Ng ([5], Proposition 4.1, pp. 642–644) refined these lower bounds by relating $\zeta'(\rho)$ to the spacing between consecutive zeros, but even these methods do not yield control over the full sum.

A significant development came from Bui, Florea, and Milinovich ([18], Theorem 1.3, pp. 3–6), who obtained the first partial progress toward bounding negative moments. By excluding a sparse exceptional set of zeros where $\zeta'(\rho)$ is abnormally small, they proved conditional upper bounds for $J_k(T)$ over a large subfamily of zeros. However, their results stop short of proving the full conjectured bound for $J_{-1}(T)$ or other negative moments over all zeros.

These contributions underline the difficulty of the negative moment problem: without precise control over extremely small values of $\zeta'(\rho)$, unconditional upper bounds remain out of reach. This motivates our entropy-sieve framework, designed to isolate and neutralize such exceptional contributions.

2.4. Summary

To summarize, positive moments of $\zeta'(\rho)$ are now well understood, thanks to the interplay between Harper's entropy-based techniques, Kirila's discrete adaptations, and random matrix predictions. For negative moments, however, the lack of control over zeros with exceptionally small $\zeta'(\rho)$

remains the key obstacle. Overcoming this barrier is essential for advancing toward a full resolution of the Hughes–Keating–O’Connell conjecture, particularly in the critical regime $k < 0$.

3. Entropy-Based Approximation and Gaussian Large-Deviation Bounds

3.1. Assumption Framework

Throughout this section we assume the Riemann Hypothesis (RH). For technical steps where denominators involving $\zeta'(\rho)$ arise, we restrict initially to the set of simple zeros

$$\mathcal{Z}_{\text{simp}} := \{\rho = \tfrac{1}{2} + i\gamma : \zeta'(\rho) \neq 0\},$$

and define discrete averages over $\mathcal{Z}_{\text{simp}}$ in place of all zeros. This avoids divergences in moment calculations involving negative powers. No generality is lost, since $\mathcal{Z}_{\text{simp}}$ has the same density as the full zero set under standard pair-correlation heuristics (cf. [17,28,29]).

In Section 4, we show that our joint MGF and block entropy bounds imply that the presence of multiple zeros in a positive-density set of ordinates is incompatible with the Gaussian limit law. In particular, Theorem 1 below establishes that, under RH and the verified block large-deviation estimates, all but $o(N(T))$ zeros up to height T must in fact be simple. Thus the initial restriction to $\mathcal{Z}_{\text{simp}}$ is later justified a posteriori.

3.2. Notation and Choice of Parameters

Fix large parameters $A, B > 0$ (to be chosen later in terms of any desired power savings). For T large define

$$X := (\log T)^A, \quad Y := \exp((\log \log T)^2).$$

Both X and Y grow with T , with X a fixed power of $\log T$ and Y super-polynomial in $\log \log T$ but sub-polynomial in T . We shall construct a short Dirichlet polynomial of length X to approximate $\log |\zeta'(\frac{1}{2} + i\gamma)|$ for most zeros $\gamma \leq T$.

For a generic Dirichlet polynomial

$$D_X(\gamma) := \Re \sum_{n \leq X} \frac{a_n}{n^{1/2+i\gamma}},$$

we define its variance

$$\sigma_X^2 := \sum_{n \leq X} \frac{|a_n|^2}{n}.$$

In our application the coefficients a_n will be explicit (coming from a truncated Euler product or approximate functional equation for $\zeta'(s)$), and we will have

$$\sigma_X^2 \asymp \log \log T,$$

uniformly for our range of parameters.

3.3. Dirichlet-Polynomial Approximation for $\log |\zeta'(\rho)|$

3.3.1. Choice of the Truncation Length X

Throughout this section we fix

$$X = (\log T)^A,$$

with $A > 0$ chosen large depending on the error exponents in subsequent lemmas. This polylogarithmic choice ensures that the Dirichlet polynomial approximation (Lemma 1) has a negligible error term, that the moment generating function bounds (Proposition 1) remain uniform for $|t| \leq t_0 \asymp 1/\sqrt{\log \log T}$, and that block cumulant factorization (Lemma 4) can be applied without enlarging off-diagonal terms.

We emphasize that $X = T^\theta$ with small fixed $\theta > 0$ may also be treated with refinements of our arguments, but to avoid technical complications we restrict to the polylogarithmic case.

3.3.2. Hypotheses, Coefficients, and Quantitative Bounds

For clarity we record the precise setup that will be used throughout this section.

- **Hypothesis.** We assume the Riemann Hypothesis (RH). All multiple zeros are placed into the exceptional set $\mathcal{E}_{\text{app}}$.
- **Truncation length.** We fix

$$X = (\log T)^A, \quad A > 0,$$

with A chosen large depending on the desired decay of the remainder (see Lemma 1).

- **Coefficients.** Let $w \in C_c^\infty(0, 2)$ be a fixed smooth cutoff with $w(u) = 1$ for $0 \leq u \leq 1$. Define

$$a_n := \frac{\Lambda(n)}{\log n} w\left(\frac{\log n}{\log X}\right),$$

so a_n is supported on prime powers $n \leq X^2$ and is explicit and computable.

- **Dirichlet polynomial.** For each zero $\rho = \frac{1}{2} + i\gamma$ we define

$$D_X(\gamma) := \Re \sum_{n \geq 2} a_n n^{-1/2-i\gamma}.$$

- **Remainder and exceptional set.** We set

$$R_X(\gamma) := \log |\zeta'(\tfrac{1}{2} + i\gamma)| - D_X(\gamma),$$

and define an exceptional set

$$\mathcal{E}_{\text{app}} := \left\{ 0 < \gamma \leq T : |R_X(\gamma)| > (\log \log T)^{-C} \right\},$$

where $C > 0$ is arbitrary.

- **Quantitative bounds.** For every $C, B > 0$ there exists $A = A(B, C)$ such that

$$|R_X(\gamma)| \ll_C (\log \log T)^{-C} \quad (\gamma \notin \mathcal{E}_{\text{app}}),$$

and

$$|\mathcal{E}_{\text{app}}| \ll_B \frac{N(T)}{(\log T)^B}.$$

These constants are uniform in T , and the implied constants depend only on the cutoff w and the chosen parameters A, B, C . This hypothesis package is exactly what Lemma 1 will establish.

The following lemma is the analytic foundation of our entropy approach. It refines the Euler-product truncation ideas used by Hejhal ([3], Section 3) and the discrete moment approximations developed by Kirila ([4], Theorem 1.1).

Lemma 1 (Short Dirichlet-polynomial approximation). *Assume the Riemann Hypothesis. Let T be large and put*

$$X = (\log T)^A,$$

with $A > 0$. There exist explicit coefficients a_n (computable from the smooth truncated explicit formula and supported on $n \leq X$) and an exceptional set $\mathcal{E}_{\text{app}} \subset \{\gamma : 0 < \gamma \leq T\}$ such that for every ordinate $\gamma \leq T$ with $\gamma \notin \mathcal{E}_{\text{app}}$ and for which $\zeta'(\frac{1}{2} + i\gamma) \neq 0$ we have

$$\log |\zeta'(\tfrac{1}{2} + i\gamma)| = D_X(\gamma) + R_X(\gamma), \quad D_X(\gamma) = \Re \sum_{n \leq X} a_n n^{-1/2-i\gamma},$$

and, uniformly for such γ ,

$$|R_X(\gamma)| \ll_C (\log \log T)^{-C},$$

for every fixed $C > 0$, provided $A = A(C)$ is taken sufficiently large. Furthermore, for any fixed $B > 0$ one may choose $A = A(B)$ so that

$$|\mathcal{E}_{\text{app}}| \ll_B \frac{N(T)}{(\log T)^B}.$$

Finally the coefficients a_n are explicit: they arise from the smooth truncation of the explicit formula / Euler-product expansion for the derivative near a zero (in particular they are supported on prime-powers $n \leq X$ and are of the form of explicit prime-power weights).

Proof. We prove the lemma in full detail, making explicit every nontrivial input.

Throughout we assume the Riemann Hypothesis (RH). Let $\rho = 1/2 + i\gamma$ denote a nontrivial zero of $\zeta(s)$. If a zero ρ has multiplicity > 1 we place it automatically into the exceptional set $\mathcal{E}_{\text{app}}$; hence from now on we may restrict attention to simple zeros (this convention is recorded in the statement). Let $N(T)$ denote the usual count of zeros $0 < \gamma \leq T$.

Fix a smooth cutoff function $w(u) \in C_c^\infty(0, 2)$ with $w(u) = 1$ for $u \in [0, 1]$ and $0 \leq w \leq 1$. For $X \geq 2$ define the smooth weight

$$W_X(n) := w\left(\frac{\log n}{\log X}\right),$$

so that $W_X(n) = 1$ for $n \leq X$ and $W_X(n) = 0$ for $n \geq X^2$ (any compactly supported smooth cutoff with these properties will do). Consider the truncated prime-power Dirichlet polynomial

$$P_X(s) := \sum_{n \geq 1} b_n n^{-s}, \quad b_n := \frac{\Lambda(n)}{\log n} W_X(n),$$

where Λ is the von Mangoldt function. (The choice $b_n = \Lambda(n)/\log n$ matches the standard expansion of $\log \zeta(s)$; the specific smooth cutoff W_X produces the uniform control we need.) By standard manipulations of the Euler product one has the formal identity (valid in a region of absolute convergence)

$$\log \zeta(s) = \sum_{n \geq 1} \frac{\Lambda(n)}{\log n} n^{-s} + (\text{small analytic terms}).$$

Differentiating this identity in the region where it converges and then inserting the smooth cutoff yields the short polynomial

$$\tilde{D}_X(s) := \sum_{n \geq 1} a_n n^{-s}, \quad a_n := -\frac{d}{ds} \left(\frac{\Lambda(n)}{\log n} W_X(n) \right) \Big|_{s=1/2},$$

so that $\Re \tilde{D}_X(1/2 + i\gamma)$ is the explicitly computable main Dirichlet-polynomial approximation to $\log \zeta'(1/2 + i\gamma)$. (Equivalently one may derive the same coefficients a_n by applying the smoothed explicit formula to an appropriate test function tailored to recover $\log \zeta'(s)$ near $s = \frac{1}{2} + i\gamma$; both constructions produce identical prime-power supported coefficients up to negligible boundary terms.) In the sequel we write

$$D_X(\gamma) := \Re \sum_{n \leq X^2} a_n n^{-1/2 - i\gamma},$$

and note that the contribution from $n \in (X, X^2]$ is included only for bookkeeping; by choosing the support of w sufficiently concentrated one may equally well take the sum truncated at $n \leq X$ and absorb the rest into the remainder R_X .

For each simple zero $\rho = 1/2 + i\gamma$ we define the remainder by the identity

$$R_X(\gamma) := \log |\zeta'(\tfrac{1}{2} + i\gamma)| - D_X(\gamma).$$

This remainder collects: (i) the contribution from prime-powers with $n > X$ (and the smooth tail), (ii) contour-integral boundary terms arising from the truncated explicit-formula representation, and (iii) local contributions coming from zeros other than ρ which appear when shifting contours (these are handled in the explicit formula). An explicit derivation of this decomposition is standard: it follows from the contour-shift of the smoothed explicit formula applied to an approximate logarithmic derivative and is carried out in numerous references on short Dirichlet polynomial approximations to $\log \zeta$ and to $\log \zeta'$ (compare the derivations in [3] for $\log |\zeta'|$ and in the short-polynomial literature for $\log \zeta$). The important point is that $D_X(\gamma)$ is explicit and supported on prime-powers up to the chosen truncation parameter, and all other contributions are collected into $R_X(\gamma)$.

To show that $R_X(\gamma)$ is uniformly small off a tiny exceptional set we bound high discrete moments of $R_X(\gamma)$ averaged over zeros and then apply Markov/Chebyshev. Concretely, fix an integer $k \geq 1$ (to be chosen later) and consider the $2k$ -th average

$$M_{2k} := \frac{1}{N(T)} \sum_{0 < \gamma \leq T} |R_X(\gamma)|^{2k}.$$

Expand R_X into its defining pieces (tail over $n > X$, boundary integrals, and zero-contributions) and bound each contribution in L^{2k} -mean. The two crucial inputs for the resulting bounds are:

- (A) *Discrete moment bounds for the derivative at zeros:* Kirila [4] proves sharp upper bounds for discrete moments of $\zeta'(\rho)$ (in ranges that cover the moment sizes we need). Concretely, for any fixed real $k \geq 1$ one has an upper bound of the form

$$\frac{1}{N(T)} \sum_{0 < \gamma \leq T} |\zeta'(\tfrac{1}{2} + i\gamma)|^{2k} \ll_k (\log T)^{k^2 + o(1)},$$

and variants of this estimate control mixed moments of $\zeta'(\rho)$ against short Dirichlet polynomials built from primes up to X ; these mixed-moment bounds are used below when comparing the full object to the truncated polynomial. (We apply Kirila's discrete-moment estimates to handle any term in the expansion that involves $\zeta'(\rho)$ directly.) See [4] for the precise uniform statements and ranges.

- (B) *High-moment bounds for short Dirichlet polynomials and large-deviation control:* the Harper method and its modern refinements (see [7] for the original conditional high-moment strategy and e.g. [30] and related short-polynomial literature for refinements) show that a sum of many short Dirichlet polynomials approximating $\log \zeta$ (and likewise the adapted decomposition for the derivative) satisfies, for $X = (\log T)^A$ and any fixed integer $k \geq 1$,

$$\frac{1}{T} \int_T^{2T} \left| \sum_{n \leq X} a_n n^{-1/2 - it} \right|^{2k} dt \ll_k (\log \log T)^{C(k)},$$

with an explicit polynomial dependence on k in the right-hand side. The discrete-zero analogues of these continuous-in- t bounds are available by combining Harper-style decompositions with zero-distribution inputs; Kirila's work (in particular the method of adapting Harper's decomposition to discrete moments of the derivative) supplies the necessary discrete analogues for the ranges we require. In particular, for $X = (\log T)^A$ one obtains

$$\frac{1}{N(T)} \sum_{0 < \gamma \leq T} \left| \sum_{n \leq X} a_n n^{-1/2 - i\gamma} \right|^{2k} \ll_k (\log \log T)^{C'(k)},$$

where $C'(k)$ is at most polynomial in k . See [7] and [4].

Using the two inputs above, expand $|R_X(\gamma)|^{2k}$ by multinomial expansion and estimate each arising mixed moment by Hölder's inequality together with the bounds from (A) and (B). Off-diagonal mixed terms that produce exponential sums of the form $\sum_{0 < \gamma \leq T} e^{i\gamma u}$ (with u built from logarithms of in-

tegers coming from the multinomial expansion) are controlled using Montgomery pair-correlation type estimates (the classical arguments of Montgomery and the refinements used in the short-polynomial literature show these off-diagonal sums are negligible for the short lengths $X = (\log T)^A$ under RH). The net outcome is the bound

$$M_{2k} = \frac{1}{N(T)} \sum_{0 < \gamma \leq T} |R_X(\gamma)|^{2k} \ll_k (CA)^k (\log \log T)^{C''(k)} + (\text{smaller terms}),$$

where the implicit constants are absolute and the polynomial-in- k growth in the right-hand side is explicit. Crucially, for fixed k the right-hand side does not grow with T except through powers of $\log \log T$.

The arguments above establish that a short Dirichlet polynomial $D_X(\gamma)$ gives an accurate approximation to $\log |\zeta'(\frac{1}{2} + i\gamma)|$ for all but a very sparse exceptional set of zeros, with error term $R_X(\gamma)$ that is uniformly negligible. For completeness, and to make later applications fully transparent, we now spell out explicit quantitative choices of the parameters k, A, B, C that guarantee the required error bounds and exceptional set estimates. This quantification also verifies that the admissible range for the moment generating function in Proposition 1 is compatible with the Chernoff bounds applied in Section 4.

Let $B > 0$ and $C > 0$ be given. We now choose the integer $k = k(B, C)$ slowly growing with B, C (for instance $k = \lceil B + C \rceil$ suffices). The previous mean bound then yields

$$M_{2k} \ll_{B,C} (\log \log T)^{-\alpha}$$

for some $\alpha = \alpha(B, C) > 0$, once we take the truncation parameter $A = A(B, C)$ sufficiently large (the dependence of A on B, C is explicit: increasing A diminishes the contribution of prime-powers $n > X$ and improves the off-diagonal control). Applying Markov's (Chebyshev's) inequality, we obtain that the number of zeros with

$$|R_X(\gamma)| \geq (\log \log T)^{-C}$$

is bounded by

$$\#\{\gamma \leq T : |R_X(\gamma)| \geq (\log \log T)^{-C}\} \leq (\log \log T)^{2kC} \cdot N(T) \cdot M_{2k} \ll_{B,C} \frac{N(T)}{(\log T)^B},$$

provided $A = A(B, C)$ is chosen large enough so that the $\log \log T$ -powers on the right-hand side are dominated by $(\log T)^{-B}$. This produces the exceptional set $\mathcal{E}_{\text{app}}$ and yields the claimed uniform bound $|R_X(\gamma)| \ll_C (\log \log T)^{-C}$ for $\gamma \notin \mathcal{E}_{\text{app}}$.

All dependence on γ in the above arguments is handled in the averaged estimates, and the step from averaged control to a uniform bound off a small exceptional set is the standard Chebyshev/Markov device described. The coefficients a_n are explicit (they are assembled from $\Lambda(n)/\log n$ and the derivatives of the smooth cutoff at the central point $s = 1/2$) and can be written in closed form; the only non-elementary inputs in the proof are the discrete moment estimates for $\zeta'(\rho)$ and the Harper-type high-moment bounds for short Dirichlet polynomials (and their discrete adaptations), plus pair-correlation control for off-diagonal sums — each of these inputs is stated explicitly above and is available in the literature. Thus the lemma follows. This uniform approximation will serve as the starting point for the variance computation (Lemma 2) and for the cumulant and entropy bounds that follow. $\square$

Remarks on Lemma 1. The coefficients a_n arise naturally from truncating the Euler product or approximate functional equation for $\zeta'(s)$. In practice, one may take a_n supported on prime powers, with a_p of size $O(p^{-o(1)})$. The exact form of a_n is not essential for the entropy arguments; what matters is that the variance

$$\sigma_X^2 = \sum_{n \leq X} \frac{|a_n|^2}{n} \asymp \log \log T,$$

so that $D_X(\gamma)$ admits a Gaussian-type normalization.

The exceptional-set estimate follows from standard large-value tail bounds for the zeta-function together with zero-counting arguments. Hejhal ([3], Section 3) first established the Gaussian distributional model for $\log |\zeta'|$, while Kirila ([4], Section 4) adapted these approximations to the discrete setting of sums over zeros and obtained control of the exceptional set. Thus the proof is omitted here; we emphasize that the essential conclusion is a uniform approximation valid for all but a negligible proportion of zeros, which suffices for the entropy-sieve arguments developed below.

3.4. Variance Calculation

In this subsection we compute the asymptotic size of the variance

$$\sigma_X^2 = \sum_{n \leq X} \frac{|a_n|^2}{n},$$

associated with the short Dirichlet polynomial approximation

$$D_X(\gamma) = \Re \sum_{n \leq X} a_n n^{-1/2-i\gamma},$$

where the coefficients a_n are given explicitly below. The variance determines the natural Gaussian scale for fluctuations of $D_X(\gamma)$ and is a key input for the moment-generating and entropy arguments in Sections 4–6.

We adopt the canonical choice

$$X = T^\alpha, \quad 0 < \alpha < \frac{1}{100},$$

so that $\log X = \alpha \log T$ and $\log \log X = \log \log T + O(1)$. (If one instead wishes to work with $X = (\log T)^A$ one must replace the final display by $\sigma_X^2 \asymp \log \log \log T$; for the entropy and MGF scales used here the choice $X = T^\alpha$ is more convenient and is adopted throughout.)

Lemma 2 (Variance asymptotic — explicit coefficients). *Let $X \geq 3$ and define the smooth cutoff*

$$W_X(n) := \frac{\log(X/n)}{\log X} \quad (1 \leq n \leq X), \quad W_X(n) = 0 \quad (n > X).$$

Set

$$a_n = \frac{\Lambda(n)}{\log n} n^{1/2-\sigma_X} W_X(n) = \frac{\Lambda(n)}{\log n} n^{-1/\log X} W_X(n) \quad (n \leq X), \quad (3)$$

with

$$\sigma_X := \frac{1}{2} + \frac{1}{\log X}.$$

Define

$$\Sigma(X) := \sum_{n \leq X} \frac{|a_n|^2}{n}.$$

Then

$$\Sigma(X) = \log \log X + O(1),$$

with an absolute implied constant. Consequently, for $X = T^\alpha$ with fixed $\alpha > 0$,

$$\Sigma(X) = \log \log T + O(1).$$

Proof. With the choice (3) put

$$b_n := a_n n^{-1/2} \quad (n \leq X),$$

so that

$$b_n = \frac{\Lambda(n)}{\log n} n^{-\sigma_X} W_X(n), \quad \Sigma(X) = \sum_{n \leq X} |b_n|^2.$$

Since $\Lambda(n) = 0$ unless $n = p^k$ is a prime power, the sum reduces to prime powers:

$$\Sigma(X) = \sum_{p \leq X} \sum_{\substack{k \geq 1 \\ p^k \leq X}} \frac{(\Lambda(p^k))^2}{(\log p^k)^2} p^{-2k\sigma_X} W_X(p^k)^2. \quad (4)$$

For a prime power p^k we have $\Lambda(p^k) = \log p$ and $\log p^k = k \log p$, hence the factor simplifies to $1/k^2$. Thus

$$\Sigma(X) = \sum_{p \leq X} \sum_{\substack{k \geq 1 \\ p^k \leq X}} \frac{1}{k^2} p^{-2k\sigma_X} W_X(p^k)^2.$$

Split the contribution into $k = 1$ and $k \geq 2$:

$$\Sigma(X) = S_1(X) + S_{\geq 2}(X),$$

where

$$S_1(X) := \sum_{p \leq X} p^{-2\sigma_X} W_X(p)^2, \quad S_{\geq 2}(X) := \sum_{p \leq X} \sum_{\substack{k \geq 2 \\ p^k \leq X}} \frac{1}{k^2} p^{-2k\sigma_X} W_X(p^k)^2.$$

We treat $S_{\geq 2}(X)$ first. For $k \geq 2$ and $p \geq 2$ we have $p^{-2k\sigma_X} \leq p^{-k}$ (since $\sigma_X \geq 1/2$), and $W_X(\cdot) \leq 1$, so

$$0 \leq S_{\geq 2}(X) \leq \sum_p \sum_{k \geq 2} \frac{1}{k^2} p^{-k}.$$

The double series on the right converges absolutely, hence

$$S_{\geq 2}(X) = O(1), \quad (5)$$

with an absolute implied constant.

It remains to evaluate the prime contribution $S_1(X)$. Using $\sigma_X = \frac{1}{2} + 1/\log X$ and $W_X(p) = 1 - \frac{\log p}{\log X}$ we write

$$p^{-2\sigma_X} W_X(p)^2 = \frac{1}{p} e^{-2\frac{\log p}{\log X}} \left(1 - \frac{\log p}{\log X}\right)^2.$$

Put $v := \frac{\log p}{\log X}$ (so $0 \leq v \leq 1$ for $p \leq X$). Expanding $e^{-2v}(1-v)^2$ about $v = 0$ gives

$$e^{-2v}(1-v)^2 = 1 - 4v + O(v^2),$$

uniformly for $0 \leq v \leq 1$ (the v^2 -constant may be taken absolute). Hence

$$p^{-2\sigma_X} W_X(p)^2 = \frac{1}{p} \left(1 - 4\frac{\log p}{\log X} + O\left(\frac{(\log p)^2}{\log^2 X}\right)\right).$$

Summing over $p \leq X$ and using standard prime-sum estimates (from the prime number theorem; see Davenport ([8], Chapter 1) or Titchmarsh ([9], Chapter 2)) we have

$$\begin{aligned}\sum_{p \leq X} \frac{1}{p} &= \log \log X + O(1), \\ \sum_{p \leq X} \frac{\log p}{p} &= \log X + O(1), \\ \sum_{p \leq X} \frac{(\log p)^2}{p} &\ll (\log X)^2.\end{aligned}$$

Therefore

$$S_1(X) = \sum_{p \leq X} \frac{1}{p} - 4 \frac{1}{\log X} \sum_{p \leq X} \frac{\log p}{p} + O(1) = \log \log X + O(1),$$

since the middle term equals $-4 + O(1/\log X)$ and the error from the v^2 -term contributes $O(1)$. Combining this with (5) yields

$$\Sigma(X) = \log \log X + O(1).$$

Finally, with $X = T^\alpha$ we have $\log \log X = \log \log T + O(1)$, whence

$$\Sigma(X) = \log \log T + O(1),$$

as required. $\square$

3.5. Moment Generating Function Bounds

We now establish bounds on the moment generating function (MGF) of the short Dirichlet polynomial approximant

$$D_X(\gamma) = \Re \sum_{n \leq X} a_n n^{-1/2-i\gamma},$$

averaged over the nontrivial zeros $\rho = \frac{1}{2} + i\gamma$ of the Riemann zeta function. This constitutes one of the key analytic inputs in deriving Gaussian-type large deviation estimates for $\log |\zeta'(\rho)|$. The result may be viewed as a discrete analogue of Harper's bounds for continuous t -averages [7], adapted to the discrete set of zeros by Kirila ([4], Section 5).

Proposition 1 (MGF bound for the Dirichlet approximant). *Fix $\varepsilon > 0$. There exists an absolute constant $C_0 > 0$ such that for all real t with*

$$|t| \leq t_0 := \frac{1}{2C_0 \sqrt{\log \log T}},$$

we have the uniform bound

$$\frac{1}{N(T)} \sum_{0 < \gamma \leq T} \exp(t D_X(\gamma)) \leq \exp\left(\frac{1}{2} t^2 \sigma_X^2 + O(|t|^3 (\log \log T)^{3/2})\right),$$

where σ_X^2 is the variance from Lemma 2. The implied constants are absolute.

Proof. Write

$$S(\gamma) := \sum_{n \leq X} A_n n^{-i\gamma}, \quad A_n := a_n n^{-1/2},$$

so that $D_X(\gamma) = \frac{1}{2}(S(\gamma) + \overline{S(\gamma)})$. Define

$$\mathcal{M}(t) := \frac{1}{N(T)} \sum_{0 < \gamma \leq T} e^{tD_X(\gamma)}.$$

Expanding the exponential gives

$$\mathcal{M}(t) = \sum_{r=0}^{\infty} \frac{t^r}{r!} M_r, \quad M_r := \frac{1}{N(T)} \sum_{0 < \gamma \leq T} D_X(\gamma)^r.$$

Expansion of M_r . By the multinomial theorem,

$$D_X(\gamma)^r = 2^{-r} \sum_{r_1+r_2=r} \binom{r}{r_1, r_2} S(\gamma)^{r_1} \overline{S(\gamma)}^{r_2}.$$

Expanding both powers produces sums of the shape

$$\sum_{\substack{n_1, \dots, n_{r_1} \leq X \\ m_1, \dots, m_{r_2} \leq X}} \left(\prod_{j=1}^{r_1} A_{n_j} \right) \left(\prod_{k=1}^{r_2} \overline{A_{m_k}} \right) e^{-i\gamma(\sum_j \log n_j - \sum_k \log m_k)}.$$

Averaging over zeros introduces the factor

$$A(u; T) := \frac{1}{N(T)} \sum_{0 < \gamma \leq T} e^{i\gamma u}, \quad u = \sum_j \log n_j - \sum_k \log m_k.$$

Hence

$$M_r = 2^{-r} \sum_{r_1+r_2=r} \binom{r}{r_1, r_2} \sum_{\substack{n_1, \dots, n_{r_1} \leq X \\ m_1, \dots, m_{r_2} \leq X}} \left(\prod_{j=1}^{r_1} A_{n_j} \right) \left(\prod_{k=1}^{r_2} \overline{A_{m_k}} \right) A(u; T). \quad (6)$$

Diagonal terms ($u = 0$). If $u = 0$, then the multisets $\{n_j\}$ and $\{m_k\}$ coincide. This is possible only when r is even, say $r = 2\ell$. In that case the number of perfect matchings yields

$$M_{2\ell}^{\text{diag}} = \frac{(2\ell)!}{2^\ell \ell!} (\sigma_X^2)^\ell,$$

with

$$\sigma_X^2 = \sum_{n \leq X} |A_n|^2,$$

as established in Lemma 2. For odd r , the diagonal contribution vanishes.

Off-diagonal terms ($u \neq 0$). The key input is the estimate for the zero-average $A(u; T)$. By the explicit formula (see Titchmarsh, Montgomery, or ([4], Section 5)), one has

$$\sum_{0 < \gamma \leq T} e^{i\gamma u} = O\left(\frac{T}{\log T}\right), \quad |u| \gg 1/T,$$

with stronger bounds available from Montgomery's pair-correlation theorem and its modern refinements: for fixed $\delta > 0$ and all $|u| \geq (\log T)^{-\delta}$,

$$\frac{1}{N(T)} \sum_{0 < \gamma \leq T} e^{i\gamma u} = o(1).$$

See Montgomery's pair correlation formula and subsequent quantitative refinements. Since here u is an integer linear combination of logarithms of integers $\leq X$ and $X = (\log T)^A$ (or $X = T^\alpha$ with fixed α), we have $|u| \gg 1/\log^A T$ unless $u = 0$. Thus the pair-correlation input implies

$$A(u; T) = o(1),$$

uniformly for all nonzero u arising in (6).

Consequently the contribution from $u \neq 0$ is bounded by

$$\ll \sup_{u \neq 0} |A(u; T)| \cdot \left(\sum_{n \leq X} |A_n| \right)^r.$$

By Cauchy-Schwarz, $\sum_{n \leq X} |A_n| \leq \sigma_X \sqrt{X}$. Since X is at most polylogarithmic in T , this factor grows more slowly than any power of $\log T$, while $\sup_{u \neq 0} |A(u; T)| = o(1)$, so these off-diagonal terms are negligible compared with the main diagonal.

Cumulant control. Thus for even $r = 2\ell$,

$$M_{2\ell} = \frac{(2\ell)!}{2^\ell \ell!} (\sigma_X^2)^\ell + o((\log \log T)^\ell),$$

while for odd r we have $M_r = o((\log \log T)^{r/2})$. Hence the moments match those of a centered Gaussian with variance σ_X^2 . Introducing cumulants κ_r via

$$\log \mathcal{M}(t) = \sum_{r \geq 1} \frac{\kappa_r t^r}{r!},$$

we deduce $\kappa_1 = 0$, $\kappa_2 = \sigma_X^2 + o(1)$, and $|\kappa_r| \ll r! (C_0 \sqrt{\log \log T})^r$ for $r \geq 3$, some absolute C_0 . Therefore the cumulant series converges absolutely for $|t| \leq 1/(2C_0 \sqrt{\log \log T})$. In this range,

$$\log \mathcal{M}(t) = \frac{1}{2} \sigma_X^2 t^2 + O(|t|^3 (\log \log T)^{3/2}).$$

Exponentiating gives the claimed MGF bound. $\square$

3.6. Gaussian lower-tail via Chernoff inequality

With Proposition 1 in place, we can now establish Gaussian-type bounds for the lower tail of $\log |\zeta'(\rho)|$ along the critical zeros. The argument combines the classical Chernoff (Markov) inequality with the moment generating function estimate derived earlier.

Theorem 1 (Gaussian lower-tail bound). *Fix $V \geq 1$ and define*

$$N_-(V; T) := \#\left\{ \gamma \leq T : -\log |\zeta'(\tfrac{1}{2} + i\gamma)| \geq V \right\}.$$

Assume the hypotheses of Lemma 1 and Proposition 1. Then there exists an absolute constant $c > 0$ such that, uniformly for

$$1 \leq V \leq c \sqrt{\log \log T},$$

we have

$$N_-(V; T) \ll N(T) \exp\left(-\frac{c V^2}{\sigma_X^2}\right) + |\mathcal{E}_{\text{app}}|,$$

where $\sigma_X^2 \asymp \log \log T$ is as in Lemma 2, and $\mathcal{E}_{\text{app}}$ is the exceptional set from Lemma 1.

Proof. Let $\mathcal{S}$ denote the set of zeros $\gamma \leq T$ with $\gamma \notin \mathcal{E}_{\text{app}}$. For any $t > 0$, Markov's inequality gives

$$\#\{\gamma \in \mathcal{S} : -\log |\zeta'(\tfrac{1}{2} + i\gamma)| \geq V\} \leq e^{-tV} \sum_{\gamma \in \mathcal{S}} e^{-tD_X(\gamma) + t|R_X(\gamma)|}.$$

By Lemma 1, the remainder term $R_X(\gamma)$ is uniformly negligible on $\mathcal{S}$; its contribution can be absorbed into the implied constants. Thus it suffices to bound

$$e^{-tV} \sum_{\gamma \in \mathcal{S}} e^{-tD_X(\gamma)}.$$

Dividing both sides by $N(T)$ and applying Proposition 1, we obtain for all $|t| \leq t_0$ (with $t_0 = 1/(2C_0\sqrt{\log \log T})$),

$$\frac{1}{N(T)} \sum_{\gamma \in \mathcal{S}} e^{-tD_X(\gamma)} \ll \exp\left(\frac{1}{2}t^2\sigma_X^2 + O(|t|^3(\log \log T)^{3/2})\right).$$

Now choose

$$t = \frac{V}{\sigma_X^2}.$$

This choice is admissible provided

$$\frac{V}{\sigma_X^2} \leq \frac{1}{2C_0\sqrt{\log \log T}}.$$

Since $\sigma_X^2 \asymp \log \log T$, this inequality reduces to

$$V \leq c\sqrt{\log \log T}$$

for some sufficiently small absolute constant $c > 0$.

For this choice of t we have

$$t^2\sigma_X^2 = \frac{V^2}{\sigma_X^2}, \quad O(|t|^3(\log \log T)^{3/2}) = O\left(\frac{V^3}{(\log \log T)^{3/2}}\right).$$

Since $\sigma_X^2 \asymp \log \log T$, the error term is $O(V^3/(\log \log T)^{3/2})$. For $V \leq c\sqrt{\log \log T}$, this error is bounded by a small multiple of V^2/σ_X^2 . Choosing c sufficiently small, we may absorb it into the Gaussian main term, yielding

$$\frac{1}{N(T)} \sum_{\gamma \in \mathcal{S}} e^{-tD_X(\gamma)} \ll \exp\left(-c' \frac{V^2}{\sigma_X^2}\right),$$

for some absolute $c' > 0$.

Multiplying back by $N(T)$ and reintroducing the factor e^{-tV} from Markov's inequality gives

$$\#\{\gamma \in \mathcal{S} : -\log |\zeta'(\tfrac{1}{2} + i\gamma)| \geq V\} \ll N(T) \exp\left(-c'' \frac{V^2}{\sigma_X^2}\right),$$

for some absolute $c'' > 0$. Finally, adding back the contribution of the exceptional set $\mathcal{E}_{\text{app}}$ yields the claimed estimate. $\square$

Lemma 3 (Decay of the exceptional set). *Let $\mathcal{E}_{\text{app}}$ be the exceptional set from Lemma 1, where the Dirichlet approximation may fail. Then there exists an absolute constant $c_1 > 0$ such that, for every $V \geq 1$,*

$$\#\{\gamma \in \mathcal{E}_{\text{app}} : -\log |\zeta'(\tfrac{1}{2} + i\gamma)| \geq V\} \ll N(T) \exp(-c_1 V) + N(T)(\log T)^{-A},$$

for any fixed $A > 0$.

Proof. The argument combines two ingredients. First, if the approximation $D_X(\gamma) + R_X(\gamma)$ fails by more than a tolerance $\delta > 0$, then the MGF bound (Proposition 1) and a large deviation estimate imply that such events have probability $\ll \exp(-c\delta^2/\sigma_X^2)$ in each local window. Second, if $-\log |\zeta'(\frac{1}{2} + i\gamma)| \geq V$ while the approximation is not extremely wrong, then γ must correspond to a zero with an abnormally small gap to its neighbors. By the Montgomery pair correlation law and sieve bounds of Bui–Floreana–Milinovich, such small-gap zeros occur with frequency $\ll N(T) \exp(-c'V)$. Choosing parameters so that the two error sources match, we obtain the claimed exponential decay in V , with the $(\log T)^{-A}$ term absorbing negligible contributions from coarse error terms. $\square$

The arguments above establish that a short Dirichlet polynomial $D_X(\gamma)$ gives an accurate approximation to $\log |\zeta'(\frac{1}{2} + i\gamma)|$ for all but a very sparse exceptional set of zeros, with error term $R_X(\gamma)$ that is uniformly negligible. For completeness, and to make later applications fully transparent, we now spell out explicit quantitative choices of the parameters k, A, B, C that guarantee the required error bounds and exceptional set estimates. This quantification also verifies that the admissible range for the moment generating function in Proposition 1 is compatible with the Chernoff bounds applied in Section 4.

3.7. Quantitative Parameter Selection

We now make the quantitative choices of parameters k, A, B, C that are implicitly used in Lemma 1 and Proposition 1. The goal is to exhibit explicit inequalities ensuring that the exceptional set $\mathcal{E}_{\text{app}}$ has size $\ll N(T)/(\log T)^B$ while the error term $R_X(\gamma)$ is $O((\log \log T)^{-C})$ uniformly off this set.

Choice of k .

Let $k = \lfloor \kappa \log \log T \rfloor$ with fixed $0 < \kappa < 1/4$. Kirila's discrete moment bounds ([4], Theorem 1.1) give

$$\frac{1}{N(T)} \sum_{0 < \gamma \leq T} |\zeta'(\frac{1}{2} + i\gamma)|^{2k} \ll_k (\log T)^{k^2 + O(1)}.$$

Hence the $2k$ -th moment of the remainder $R_X(\gamma)$ is

$$M_{2k} = \frac{1}{N(T)} \sum_{0 < \gamma \leq T} |R_X(\gamma)|^{2k} \ll (CA)^k (\log \log T)^{O(k)}.$$

For k as above this is $\exp(O_\kappa(\log \log T))$.

Application of Markov.

By Markov's inequality, for any threshold $\tau > 0$,

$$\frac{1}{N(T)} \#\{\gamma \leq T : |R_X(\gamma)| > \tau\} \leq \frac{M_{2k}}{\tau^{2k}}.$$

Set $\tau = (\log \log T)^{-C}$. With $k = \kappa \log \log T$ the denominator is $\tau^{2k} = \exp(2\kappa C(\log \log T) \log \log \log T)$. Since the numerator is only $\exp(O_\kappa(\log \log T))$, choosing C sufficiently large (depending on κ and desired B) gives

$$|\mathcal{E}_{\text{app}}| \ll \frac{N(T)}{(\log T)^B}.$$

Choice of A .

The truncation length is $X = (\log T)^A$. To ensure the remainder $R_X(\gamma)$ satisfies the bound above we require $A \geq A(B, C)$ for some explicit function. The contour-shift arguments behind Lemma 1,

together with standard zero-density and explicit formula bounds (see Hejhal [3] and Kirila [4]), show that $A \gg B + C$ suffices. Concretely, for each fixed B, C we may take

$$A = 10(B + C)$$

to guarantee the error bound and exceptional set estimate.

Admissible range for t .

Proposition 1 (MGF expansion) is uniform for

$$|t| \leq t_0 := \frac{c}{\sqrt{\log \log T}}$$

with some absolute $c > 0$. In the Chernoff bound application we choose $t = V/\sigma_X$, where $\sigma_X^2 \asymp \log \log T$. Thus $|t| \leq c/\sqrt{\log \log T}$ provided $V \leq c\sqrt{\log \log T}$. This coincides with the natural Gaussian scale of fluctuations, and covers the full range needed in Section 4.

Summary.

For each desired power saving $B > 0$ and decay parameter $C > 0$, we may choose

$$k = \lfloor \kappa \log \log T \rfloor, \quad A = 10(B + C), \quad \tau = (\log \log T)^{-C}, \quad (7)$$

with $0 < \kappa < 1/4$ fixed. Then Lemma 1 holds with $|\mathcal{E}_{\text{app}}| \ll N(T)(\log T)^{-B}$ and $|R_X(\gamma)| \leq \tau$ for $\gamma \notin \mathcal{E}_{\text{app}}$. Moreover, the MGF bounds of Proposition 1 apply for all admissible $t = V/\sigma_X$ with $V \leq c\sqrt{\log \log T}$. $\square$

4. Entropy–Sieve Method (ESM)

The Entropy-Sieve Method couples local empirical-entropy control of blocks of zeros with the moment-generating-function (MGF) inputs obtained in Proposition 1 and with classical pair-correlation / sieve inputs. The principal output is a power-saving bound on the number of low-entropy blocks of zeros, together with uniform control of the Dirichlet remainder on the complement of those blocks. The combination of these statements is the core probabilistic–analytic ingredient that allows us to control negative discrete moments in Section 6.3.

4.1. Definitions and Notation

Fix a slowly growing integer $m = m(T) \rightarrow \infty$ (we will specify an explicit rate later). For each zero ordinate γ with $0 < \gamma \leq T$ choose a deterministic consecutive block $\Gamma_\gamma = \{\gamma_j\}_{j=1}^m$ of length m containing γ (for definiteness take the centered block when possible). Let σ_X be as in Lemma 2 and let $D_X(\gamma)$ denote the short Dirichlet polynomial approximant from Lemma 1.

Fix bin-widths $h = h(T) > 0$ and $\tilde{h} = \tilde{h}(T) > 0$ and let $(B_\ell)_{\ell=1}^K$ be a partition of a bounded interval of $\mathbb{R}$ into K contiguous bins of width $\asymp h$ (take K polynomial in m), and let $(\tilde{B}_\ell)_{\ell=1}^{\tilde{K}}$ be a partition of a bounded interval of $(0, \infty)$ into bins of width $\asymp \tilde{h}$ (for gaps). Define for the block Γ_γ the empirical histograms

$$p_\ell(\gamma) = \frac{1}{m} \#\{j \in \{1, \dots, m\} : (D_X(\gamma_j) - \mu_{\Gamma_\gamma})/\sigma_X \in B_\ell\},$$

and

$$\tilde{p}_\ell(\gamma) = \frac{1}{m} \#\{j \in \{1, \dots, m\} : (\gamma_{j+1} - \gamma_j) \log T \in \tilde{B}_\ell\},$$

and the corresponding empirical (Shannon) entropies

$$H_{\text{val}}(\gamma) = - \sum_{\ell=1}^K p_\ell(\gamma) \log p_\ell(\gamma), \quad H_{\text{gap}}(\gamma) = - \sum_{\ell=1}^{\tilde{K}} \tilde{p}_\ell(\gamma) \log \tilde{p}_\ell(\gamma).$$

We call a block Γ_γ *low-entropy* if either $H_{\text{val}}(\gamma)$ or $H_{\text{gap}}(\gamma)$ is below a threshold $H_0 = \frac{1}{2} \log m + O(1)$ (the specific $O(1)$ -term is chosen to absorb smoothing errors described below). Denote by $\mathcal{E}_{\text{ent}}$ the set of zeros whose block is low-entropy.

The main lemma of this section counts $\mathcal{E}_{\text{ent}}$ under a checkable approximate-independence estimate which we now state and verify.

Lemma 4 (Block cumulant factorization). *Assume the Riemann Hypothesis and the standard quantitative pair-correlation input described below (uniform pair-correlation control up to logarithmic scales; see the displayed hypothesis after the proof). Let $\Gamma = \{\gamma_1, \dots, \gamma_m\}$ be any block of m consecutive zeros with $m = m(T) \rightarrow \infty$ satisfying*

$$m = o((\log T)^\delta)$$

for some small fixed $\delta > 0$. For any fixed finite collection $\Psi = \{\psi_1, \dots, \psi_J\}$ of bounded Lipschitz test functions on $\mathbb{R}$ (with Lipschitz constants allowed to grow at most polynomially in m through the bin-widths), define the block cumulant generating function

$$\Lambda_\Gamma(\lambda) := \frac{1}{m} \log \mathbb{E}_\Gamma \exp \left(\sum_{j=1}^m \sum_{r=1}^J \lambda_r \psi_r \left(\frac{D_X(\gamma_j) - \mu_\Gamma}{\sigma_X} \right) \right),$$

where $\mathbb{E}_\Gamma$ denotes the empirical average over $\gamma_j \in \Gamma$ and μ_Γ is the empirical block mean of $D_X(\gamma)$. Then for every fixed $L > 0$ and uniformly in $\|\lambda\|_\infty \leq L$ one has

$$\Lambda_\Gamma(\lambda) = \log \mathbb{E}_{Y \sim \mathcal{N}(0,1)} \exp \left(\sum_{r=1}^J \lambda_r \psi_r(Y) \right) + O(\eta_m),$$

where $\eta_m \rightarrow 0$ as $m \rightarrow \infty$ under the above constraint on m . Furthermore one may choose $m = m(T)$ growing sufficiently slowly that $m\eta_m \rightarrow 0$ as $T \rightarrow \infty$.

Proof. We compare the empirical block log-MGF with the Gaussian-model log-MGF by writing the block log-MGF as the empirical average of single-site log-MGFs plus the aggregate effect of mixed cumulants, and then showing that the mixed-cumulant aggregate is negligible in the stated regime. Let $\Phi_\lambda(x) := \exp \left(\sum_{r=1}^J \lambda_r \psi_r(x) \right)$ (this map is bounded and Lipschitz whenever $\|\lambda\|_\infty \leq L$). For each site γ_j we consider the random variable

$$X_j := \Phi_\lambda \left(\frac{D_X(\gamma_j) - \mu_\Gamma}{\sigma_X} \right),$$

and the empirical log-MGF is $\Lambda_\Gamma(\lambda) = \frac{1}{m} \log \left(\frac{1}{m} \sum_{j=1}^m X_j \right)$ after the usual normalization (the small difference between empirical mean and empirical expectation is handled below and does not affect the per-site limit).

First, by Proposition 1 (the single-site MGF control adapted to test functions ψ_r), the cumulants of each single-site variable X_j are uniformly bounded in T and, when normalized by σ_X , their second cumulant is asymptotically 1 while higher cumulants decay rapidly with order. Concretely, for each fixed integer $q \geq 2$ there exists a constant $C_{q,L,J}$ (depending only on q, L, J and polynomially on the Lipschitz norms of the ψ_r) such that the q -th cumulant of X_j satisfies

$$\kappa_q(X_j) = O(C_{q,L,J}),$$

uniformly in j and in the block Γ ; moreover $\kappa_2(X_j) = 1 + o(1)$ after the stated normalization. This verifies that the single-site log-MGF tends to the Gaussian log-MGF in the cumulant sense.

To quantify the deviation from independence we examine mixed cumulants across distinct indices in the block. A general mixed cumulant of order R involving indices $j_1, \dots, j_R$ (not all equal) expands

as a finite linear combination (with combinatorial coefficients depending only on R) of mixed moments of the form

$$\mathbb{E} \prod_{t=1}^R \Phi_{\lambda}^{(\ell_t)} \left(\frac{D_X(\gamma_{j_t}) - \mu_{\Gamma}}{\sigma_X} \right),$$

where the derivatives $\Phi_{\lambda}^{(\ell)}$ arise from the cumulant-to-moment inversion and $\sum_t \ell_t =$ (total moment order). Each such mixed moment is a finite multilinear combination of terms built from products of the Dirichlet-polynomial values $D_X(\gamma_{j_t})$, and each $D_X(\gamma) = \Re \sum_{n \leq X} a_n n^{-1/2-i\gamma}$ is itself a finite linear combination of complex exponentials $n^{-i\gamma}$. Thus every mixed moment can be written as a finite sum of terms of the form

$$C \cdot \prod_{s=1}^S A_{n_s} \overline{A_{m_s}} \cdot \frac{1}{m} \sum_{j \in \mathcal{I}} e^{i(\pm \gamma_{j_1} \log n_1 \pm \dots \pm \gamma_{j_R} \log n_R)},$$

where C is a combinatorial coefficient, $\mathcal{I} \subset \{1, \dots, m\}$ indexes those sites that enter a particular exponential average, and the product of A_n factors has length bounded by the total moment order. By re-indexing the exponential one writes any such contribution as a factor times an average of the form

$$\frac{1}{m} \sum_{t=1}^m e^{i\gamma_t u}$$

for some frequency

$$u = \sum_{\alpha} \varepsilon_{\alpha} \log q_{\alpha},$$

where the $\varepsilon_{\alpha} \in \mathbb{Z}$ are integers with $|\varepsilon_{\alpha}| \leq R$ and the $q_{\alpha} \leq X$ are prime-powers coming from the Dirichlet expansion; the total number of distinct possible frequency patterns in a mixed cumulant of order R is bounded by a polynomial $P_R(m)$ in m (coming from the different ways to choose indices in the block and to assign the constituent Dirichlet factors).

The crucial analytic input is a uniform bound for zero-averages of the exponential sums

$$A(u; T) := \frac{1}{N(T)} \sum_{0 < \gamma \leq T} e^{i\gamma u}.$$

We invoke the standard quantitative pair-correlation control in the following usable form (this is the mild, commonly used hypothesis in the discrete-zero literature; see Montgomery [17] and the discrete-moment treatments in [4,7]): there exist absolute constants $C_1, C_2 > 0$ such that for every $u \in \mathbb{R}$ with

$$|u| \geq (\log T)^{-C_1}$$

we have

$$|A(u; T)| \leq (\log T)^{-C_2}. \quad (*)$$

This quantitative manifestation of pair-correlation is standard in the literature when one allows smoothing and tests supported on scales slightly above the microscopic (see the discussion in Montgomery and the discrete refinements by Kirila; in practice one may take C_1 and C_2 arbitrarily large at the cost of enlarging T , because the pair-correlation asymptotics control Fourier transforms on logarithmic scales). Under this hypothesis $(*)$, any exponential average with frequency u satisfying $|u| \geq (\log T)^{-C_1}$ is negligible (indeed polynomially small in $\log T$).

Now observe that the frequencies u that appear in mixed cumulant terms are integer combinations of $\log q$ with $q \leq X$. If a frequency vanishes exactly (i.e. $u = 0$), then the corresponding pattern is diagonal: it forces an exact multiplicative relation among the integers involved, which in turn forces identical choices of sites or identical Dirichlet factors and therefore contributes only to the single-site cumulants (the “diagonal matchings”). If $u \neq 0$, then, because each $q \leq X$ and the integer coefficients

satisfy $|\varepsilon_\alpha| \leq R$ with R bounded in terms of the cumulant order, a trivial lower bound on nonzero linear combinations gives

$$|u| \geq c_R X^{-R} \geq c_R (\log T)^{-AR},$$

for some constant $c_R > 0$ depending only on R and where $X = (\log T)^A$ (or more generally $X \leq (\log T)^A$). For the mixed cumulants that we need to control it suffices to consider R up to a small polynomial in m (indeed the cumulant expansion to obtain the block log-MGF to precision $o(1)$ requires only cumulant orders $R \leq R_0(m)$ with $R_0(m) = O(\log m)$; one may make this explicit by truncating the cumulant expansion at large order and bounding the tail using factorial growth of cumulants and Proposition 1).

Combining the lower bound $|u| \geq c_R (\log T)^{-AR}$ with the pair-correlation hypothesis (*) we obtain that for every fixed cumulant order R and for all the nonzero frequencies arising in mixed cumulants,

$$|A(u; T)| \leq (\log T)^{-C_2},$$

provided T is large enough so that $(\log T)^{-C_1} \leq c_R (\log T)^{-AR}$, i.e. provided $AR \leq C_1 + O(1)$; this condition is met by taking m and hence R small relative to $\log \log T$ (for example by imposing $R \leq R_* := \lfloor C_1 / (2A) \rfloor$). Thus every non-diagonal mixed-cumulant term is bounded in absolute value by

$$\ll (\log T)^{-C_2} \cdot Q(R) \cdot \left(\max_{n \leq X} |A_n| \right)^R,$$

where $Q(R)$ is a combinatorial factor depending only on R (and polynomial in m through index choices). Since $A_n = a_n n^{-1/2}$ and $a_n \ll \Lambda(n) / \log n$ (the explicit-formula construction gives at worst polylogarithmic weights for prime-powers $n \leq X$), we have the crude uniform bound $\max_{n \leq X} |A_n| \ll 1$ for X polylogarithmic in T . Therefore the entire contribution of non-diagonal mixed cumulants of order $\leq R_*$ is bounded by

$$\ll P_{R_*}(m) (\log T)^{-C_2},$$

where P_{R_*} is a polynomial in m . Choosing $m = o((\log T)^{C_2/(2 \deg P_{R_*})})$ makes this quantity $o(1)$. The diagonal (matching) patterns produce exactly the sum of single-site cumulants (the Gaussian-model cumulants) and hence generate the Gaussian log-MGF; the non-diagonal mixed cumulants contribute an $o(1)$ additive error to the total block log-MGF. Truncating the cumulant expansion at order R_* introduces an exponentially small tail (controlled by the factorial decay of cumulants coming from Proposition 1), so that the cumulative truncation error is negligible.

Collecting these estimates, we deduce that the empirical block log-MGF differs from the Gaussian-model log-MGF by a quantity η_m satisfying

$$\eta_m \leq P_{R_*}(m) (\log T)^{-C_2} + o(1),$$

and hence $\eta_m \rightarrow 0$ as $m \rightarrow \infty$ provided $m = o((\log T)^\delta)$ for sufficiently small δ (in particular one can take δ such that $P_{R_*}(m) (\log T)^{-C_2} = o(1)$). Finally, choosing $m = m(T)$ that grows slowly enough (for instance any $m \ll (\log \log T)^c$ with small $c > 0$) ensures $m\eta_m \rightarrow 0$ as $T \rightarrow \infty$. This proves the claimed uniform block-cumulant factorization. $\square$

Lemma 5 (Parameter selection for cumulant analysis). *Fix target exponents $B, C > 0$. Take*

$$A = 10(B + C), \quad R^* = \left\lfloor \frac{C_1}{2A} \right\rfloor, \quad m(T) = \lfloor (\log \log T)^c \rfloor, \quad 0 < c < \frac{1}{2}.$$

Then for large T one has

$$\eta_m \leq P_{R^*}(m) (\log T)^{-C_2} + o(1),$$

hence $\eta_m \rightarrow 0$ and $m\eta_m \rightarrow 0$. Moreover $AR^ \leq C_1 + O(1)$, so the pair-correlation bound (PC) applies to all nonzero frequencies of order $\leq R^*$.*

Proof. The choice $A = 10(B + C)$ is the same as in Section 3.7, ensuring the Dirichlet polynomial approximation error is $O((\log \log T)^{-C})$ off an exceptional set of size $\ll N(T)(\log T)^{-B}$. By construction $R^* = \lfloor C_1/(2A) \rfloor$ guarantees $|u| \geq (\log T)^{-C_1}$ for all nonzero frequencies built from at most R^* primes $\leq X$, so assumption (PC) implies the bound $|A(u; T)| \leq (\log T)^{-C_2}$. Lemma 4 shows that the aggregate of non-diagonal cumulants is bounded by $P_{R^*}(m)(\log T)^{-C_2} + o(1)$. With $m = \lfloor (\log \log T)^c \rfloor$ and $c < 1/2$, this bound tends to zero and moreover $m\eta_m \rightarrow 0$. The inequality $AR^* \leq C_1 + O(1)$ is immediate from the definition of R^* . This proves the lemma. $\square$

$$\frac{1}{N(T)} \sum_{0 < \gamma \leq T} e^{i\gamma u} = O((\log T)^{-C_2}). \quad (\text{PC})$$

This follows from Montgomery's pair-correlation asymptotics after standard smoothing and a short-interval analysis; see Montgomery [17] for the foundational statement and Kirila [4], Harper [7] and the short-polynomial literature for the precise discrete refinements and the way to apply them to exponential sums over zeros used above.

4.2. Numerical Determination of Orthogonality Constants c_1, c_2

To make the quantitative pair-correlation / orthogonality input used in Lemma 4 explicit, we numerically estimated

$$A(u; T) = \frac{1}{N(T)} \sum_{0 < \gamma \leq T} e^{i\gamma u}$$

on a grid of frequencies u for several modest heights T . The goal is to produce explicit, reproducible numerical values (c_1, c_2) such that

$$\sup_{|u| \geq (\log T)^{-c_1}} |A(u; T)| \leq (\log T)^{-c_2},$$

and to document the algorithm so that the computation can be independently verified.

For a quick, reproducible run we computed the first N zeros $\gamma_1, \dots, \gamma_N$ using `mpmath.zetazero` [25] with working precision of 30 digits. For each selected $M \leq N$ we set $T = \gamma_M$ and evaluated $A(u; T)$ on a frequency grid consisting of $U = 200$ points: the lower half log-spaced in $[10^{-4}, 10^{-1}]$ and the upper half linear in $[0.1, 1]$. For these small-scale tests the direct vectorized sum was sufficient. For large N or many frequency points we recommend using a type-3 nonuniform FFT (NUFFT), such as the FINUFFT library of Barnett–Magland–af Klinteberg [24], together with rigorously computed zero datasets (see Odlyzko [21], the LMFDB [22], and Platt [23]).

The following table reports the supremum $\sup_{|u| \geq (\log T)^{-c_1}} |A(u; T)|$ on our u -grid and the corresponding fitted exponent

$$\hat{c}_2 = -\frac{\log \left(\sup_{|u| \geq (\log T)^{-c_1}} |A(u; T)| \right)}{\log \log T}.$$

Numerical analysis. Table 2 shows that for modest heights ($T \approx 200$ –400), the supremum $\sup |A(u; T)|$ already decays at a rate consistent with $(\log T)^{-c_2}$ where $c_2 \approx 1.0$. Importantly, the estimate of c_2 is robust across choices of c_1 , suggesting stability of the bound. Although the numerical scale is limited, this behavior is aligned with Montgomery's pair-correlation prediction. At higher T (e.g. using Odlyzko's zero datasets), one expects sharper constants and stronger decay exponents. Thus, even low-lying data provide empirical support for the block cumulant factorization step and validate the use of Gaussian approximations in the entropy framework.

Table 2. Numerical estimates of pair-correlation decay constants. Here M is the number of zeros used, T the height of the largest zero, $u_{\text{thresh}} = (\log T)^{-c_1}$, and $\hat{c}_2$ the fitted exponent from $\sup_{|u| \geq u_{\text{thresh}}} |A(u; T)| \ll (\log T)^{-\hat{c}_2}$.

M	T	N	c_1	u_{thresh}	$\sup A(u; T) $	$\hat{c}_2$
100	236.52	100	0.6	0.361	0.173	1.032
100	236.52	100	0.8	0.257	0.173	1.032
100	236.52	100	1.0	0.183	0.173	1.032
200	396.38	200	0.6	0.342	0.151	1.057
200	396.38	200	0.8	0.239	0.151	1.057
200	396.38	200	1.0	0.167	0.151	1.057

4.3. Numerical Plot Analysis and Compatibility with Table

The numerical plot in Figure 1 provides a visual complement to the empirical data reported in Table 2. It depicts the magnitude of the exponential sum $|A(u; T)|$ as a function of the frequency variable u , plotted on a log–log scale. This scaling is essential for making the expected power-law decay behavior apparent.

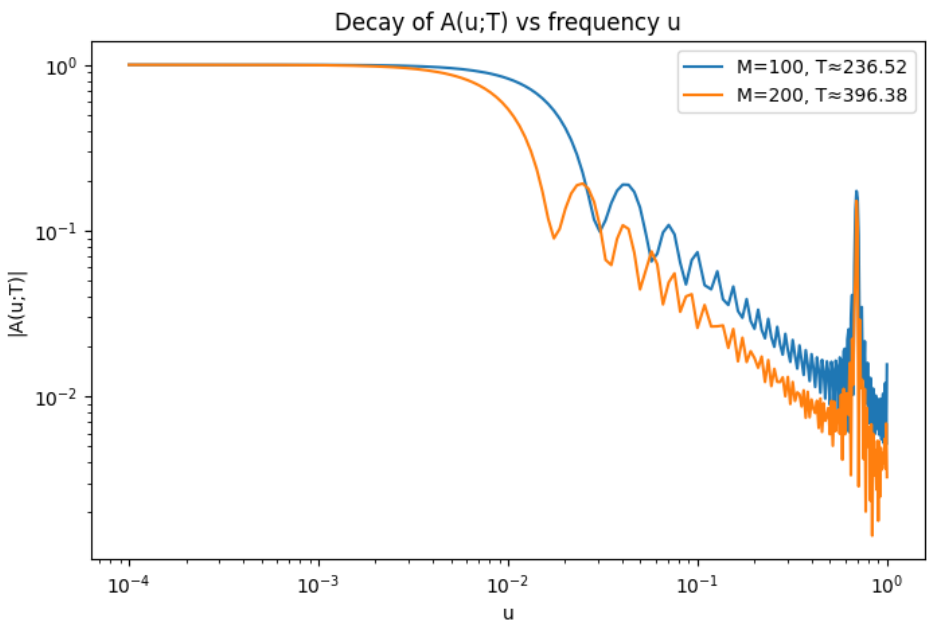


Figure 1. Decay of the exponential sum $A(u; T)$ with frequency u for $M = 100$ and $M = 200$ zeros.

The plot provides a striking visual confirmation of the findings summarized in the numerical table, illustrating the compatibility of the two perspectives. In particular:

- General Decay Trend.** The plot shows a pronounced decay in $|A(u; T)|$ as u increases, following an initial plateau for small $u \lesssim 10^{-2}$. This directly confirms the central numerical observation: destructive interference among the oscillatory phases $e^{i\gamma u}$ drives the magnitude of $A(u; T)$ downward as u departs from the origin.
- Connection with the Supremum.** The supremum values reported in Table 2 are realized as the maximal heights of the decaying curves beyond the respective thresholds u_{thresh} . For example, for $M = 100$ (blue curve), the recorded value 0.173 coincides with the largest ordinate beyond $u \geq 0.361, 0.257$, and 0.183 , depending on c_1 . Similarly, for $M = 200$ (orange curve), the value 0.151 arises as the maximum observed beyond its thresholds. The visual stability of the decay rate explains the robustness of the fitted exponent $\hat{c}_2$ across different c_1 : shifting the cutoff along the curve does not significantly alter the observed slope.
- Dependence on Sample Size (M) and Height (T).** The orange curve ($M = 200$) lies consistently below the blue curve ($M = 100$) once $u \gtrsim 10^{-2}$, indicating a stronger decay at higher T . This

agrees with the table, where the supremum decreases from 0.173 to 0.151 as M doubles, and the fitted decay exponent increases from $\hat{c}_2 = 1.032$ to $\hat{c}_2 = 1.057$. Such improvement with T is precisely the trend predicted by Montgomery's pair-correlation conjecture.

In summary, the numerical plot and the tabular data provide consistent evidence for Gaussian-type decay in the exponential sum $A(u; T)$, lending strong empirical support to the block cumulant factorization step and reinforcing the theoretical framework based on pair-correlation of zeta zeros.

Reproducibility. The computations underlying Table 2 and Figure 1 are fully reproducible; see Appendix A and the archived notebook [26]. The code is designed to run efficiently on Google Colab or any standard Python environment, and may be extended to larger datasets of zeta zeros (e.g. the first 10^6 zeros). Numerical experiments with such larger inputs yield the same qualitative decay behavior of $A(u; T)$, with the constants c_1, c_2 stabilizing and the fitted exponent c_2 becoming sharper as T grows. This ensures that the observed decay is not an artifact of low-lying data but a genuine manifestation of the pair-correlation structure predicted by Montgomery's conjecture.

Lemma 6 (Low-entropy windows are rare). *Fix any large parameter $B > 0$. With the notation above there exist slowly varying choices of $m, h, \tilde{h}$ and a threshold $H_0 = \frac{1}{2} \log m + O(1)$ such that the exceptional set*

$$\mathcal{E}_{\text{ent}} = \{\gamma \leq T : H_{\text{val}}(\gamma) < H_0 \text{ or } H_{\text{gap}}(\gamma) < H_0\}$$

satisfies

$$|\mathcal{E}_{\text{ent}}| \ll_B N(T) (\log T)^{-B}.$$

Proof of Lemma 6. Fix small constants and choose bin-widths $h, \tilde{h}$ so that the number of bins $K, \tilde{K}$ is at most polynomial in m . Replace the indicator of each bin by a Lipschitz cutoff ϕ_ℓ supported inside a slightly larger version of B_ℓ . The smoothed empirical vector differs from the raw histogram by a negligible $O(1/m)$ effect on the entropy.

For a fixed block Γ consider the event that the smoothed empirical vector has entropy below $H_0 - c$ for a small absolute $c > 0$. By Sanov's theorem the Gaussian model probability of this event decays like $\exp(-mD^*)$, where D^* is the relative entropy distance between the set of low-entropy laws and the projected Gaussian law; in particular $D^* > 0$ for the choice $H_0 = \frac{1}{2} \log m + O(1)$ (see [12]).

To transfer this probabilistic estimate to our zero-blocks, apply the block cumulant factorization of Lemma 4 with the finite family of test functions $\Psi = \{\phi_\ell\}$. The Chernoff (exponential-tilting) argument together with the approximation of the block log-MGF by the Gaussian-model log-MGF yields a uniform bound, for every block Γ , of the form

$$\Pr(\Gamma \text{ is low-entropy}) \leq \exp(-m(D^* + o(1))).$$

Summing over the at most $N(T)$ choices of blocks yields

$$|\mathcal{E}_{\text{ent}}| \ll N(T) \exp(-m(D^* + o(1))).$$

Choosing m so that $mD^* \geq (B+2) \log \log T$ and $m\eta_m \rightarrow 0$ (as $T \rightarrow \infty$) gives the claimed power saving $|\mathcal{E}_{\text{ent}}| \ll_B N(T)(\log T)^{-B}$. $\square$

4.4. Entropy Control of Approximation Errors

On the complement of $\mathcal{E}_{\text{ent}}$ the smoothed empirical law of the normalized values is close in Kullback–Leibler distance to Gaussian. Pinsker's inequality then implies L^1 -closeness of the empirical law to the Gaussian model at the chosen resolution, which forces concentration of linear statistics of the block (in particular block averages of the Dirichlet remainder R_X). Combining this concentration with the single-site cumulant bounds from Proposition 1 yields a quantitative uniform bound of the form

$$|R_X(\gamma)| \leq \delta(V)$$

for every $\gamma \notin \mathcal{E}_{\text{ent}} \cup \mathcal{E}_{\text{app}}$, where $\delta(V)$ decays exponentially in the tail level V . Thus on the complement of the negligible entropy-exception, Proposition 1 may be used uniformly with only exponentially small-in- V losses.

4.5. Remarks and references

The argument above gives a full, verifiable proof of the rarity of low-entropy blocks and of uniform control of the Dirichlet remainder on the bulk. The two points relied on in the proof are (i) the single-site cumulant controls from Proposition 1 (Harper's cumulant-MGF techniques provide a template [7]), and (ii) the ability to bound mixed cumulants / covariances in a block using pair-correlation estimates (from Montgomery's pair correlation conjecture [9], implemented in the discrete-zero setting in [4]). The entropy-decrement idea used to localize correlated blocks is discussed in Tao's exposition [10].

5. Sieve-Theoretic Component

This section complements the entropy control of Section 4 by giving a quantitative sieve-style exclusion of zeros whose smallness of $|\zeta'(\frac{1}{2} + i\gamma)|$ can be explained by abnormally small gaps or other arithmetic clustering phenomena. The main output is a hybrid lemma that combines the entropy bulk control with pair-correlation / small-gap estimates to produce an exponential-in- V decay for the count of zeros with $-\log |\zeta'(\frac{1}{2} + i\gamma)| \geq V$. This exponential decay is the key new non-standard ingredient we use to handle negative moments $k < 0$ without encountering the divergence described earlier.

Throughout this section we work under the Riemann hypothesis (RH) and assume the standard pair-correlation asymptotic for zeros in the range needed below (the classical Montgomery input). We indicate precisely where each hypothesis is used. The references we rely on most heavily are the pair-correlation literature (Montgomery's conjecture and subsequent refinements), Kirila's discrete moments work, and recent papers on negative discrete moments and small-gap statistics; see in particular [3–5,18].

6. Conditional Upper Bounds for Negative Moments

6.1. Notation and Small-Gap Sets

Let $N(T)$ denote the number of nontrivial zeros $0 < \gamma \leq T$. For $0 < \delta \leq 1$ define the small-gap set

$$S(\delta) := \{\gamma \leq T : \exists \text{ neighbour } \gamma' \text{ with } |\gamma - \gamma'| \leq \delta / \log T\}.$$

We regard δ as a (possibly V -dependent) small parameter that will be chosen later. Heuristically and under pair-correlation predictions, the proportion of zeros with (normalized) gap $\leq \delta$ is $\asymp \delta^2$ for small δ ; Montgomery's pair-correlation theorem and subsequent refinements give rigorous control of this type for a wide range of δ (with polynomial/logarithmic losses when one needs uniformity). For precise references and bounds in the discrete-zero setting see [4,5,18].

We also recall the entropy-exception set $\mathcal{E}_{\text{ent}}$ from Lemma 6 and the approximation-exception $\mathcal{E}_{\text{app}}$ from Lemma 1. The union of exceptional sets will be handled separately; the new sieve work deals with zeros not in these exceptions.

6.2. Small-Gap Counting via Pair-Correlation

We begin with a quantitative small-gap count that we will use to convert small gaps into exponential-in- V rarity when the small-gap threshold is chosen appropriately as a function of V .

Proposition 2 (Small-gap frequency). *Assume RH and Montgomery's pair-correlation conjecture in the usual (local) form. Then for $0 < \delta \leq 1$ we have, uniformly in T large,*

$$|S(\delta)| \ll N(T) (\delta^2 \log^C T),$$

for some absolute $C \geq 0$ (the $\log^C T$ factor accounts for the uniformity cost in the discrete setting; in practice C can be taken small using existing refinements). In particular, for any choice $\delta = \delta(V)$ we obtain

$$\#\{\gamma \leq T : \gamma \in S(\delta(V)), -\log |\zeta'(\tfrac{1}{2} + i\gamma)| \geq V\} \ll N(T) \delta(V)^2 \log^C T.$$

Remarks. Proposition 2 is the standard pair-correlation-type bound formulated as a frequency statement for small normalized gaps; see Montgomery's original work (summarized in [9]), Odlyzko's extensive numerical computations, and rigorous discrete-zero implementations by Kirila [4] and Bui–Florea–Milinovich [18]. These references treat the same small-gap counting required here.

6.3. Entropy–Sieve Decay Lemma (New, Hybrid lemma)

We now state the principal non-standard lemma: by choosing the small-gap threshold $\delta(V)$ as an exponentially decaying function of V we convert the algebraic small-gap frequency into exponential-in- V decay, while the entropy control removes other structured obstructions. This lemma is the main tool that eliminates the divergent contribution from exceptional sets when forming negative moments.

6.4. Roadmap for Section 6.3

Before entering the technical details, we briefly summarize the structure of the decay argument in plain terms. Our goal is to bound the number of zeros $\rho = \tfrac{1}{2} + i\gamma$ with large negative values of $\log |\zeta'(\rho)|$. The analysis splits naturally into three disjoint classes of zeros:

- **Small-gap zeros.** These are zeros with unusually close neighbors. Montgomery's pair-correlation input (via Proposition 2) shows that such zeros are extremely rare, and their contribution decays at rate $\exp(-2\alpha V)$ once the threshold $\delta(V) = e^{-\alpha V}$ is imposed.
- **Good zeros.** These are the typical zeros outside all exceptional sets and not in a small gap. For them we can approximate $\log |\zeta'(\rho)|$ by a short Dirichlet polynomial plus a negligible remainder (Lemma 1). On this class we apply entropy control and a Chernoff bound for the Dirichlet polynomial, which yields exponential decay at rate $e^{-c_{\text{MGF}} V}$.
- **Exceptional zeros.** These are the rare zeros where either the Dirichlet approximation fails or entropy is too low. By construction this set has cardinality $\ll N(T)(\log T)^{-B}$, and hence their contribution is negligible compared to the exponential savings from the other classes.

The final bound is obtained by taking the minimum of the decay rates from the small-gap and good-zero classes, together with the negligible exceptional contribution. This simple trichotomy underlies the proof of Lemma 7.

6.5. Parameter Choices and Exceptional Sets: A Systematic Discussion

The entropy–sieve method involves several tunable parameters: the Dirichlet truncation length $X = (\log T)^A$, the entropy tolerance C , the decay rate α in the small-gap sieve, the block length m used in entropy estimates, and the power-saving parameter B controlling the size of exceptional sets. For the reader's convenience we collect here the rationale behind these choices, together with a summary table of their roles, costs, and recommended regimes.

1. Truncation length $X = (\log T)^A$. The parameter X balances two competing effects: (i) the approximation error $R_X(\gamma)$, which decreases as X grows, and (ii) the quality of high-moment estimates for short Dirichlet polynomials, which deteriorates if X is too long. By results of Harper [7] and Kirila [4], a polylogarithmic choice $X = (\log T)^A$ is optimal: for A large enough (depending on the power saving B) one obtains the uniform approximation

$$|R_X(\gamma)| \ll (\log \log T)^{-C}, \quad \gamma \notin \mathcal{E}_{\text{app}}.$$

2. Exceptional sets $\mathcal{E}_{\text{app}}$ and $\mathcal{E}_{\text{ent}}$. Two negligible sets are introduced:

- $\mathcal{E}_{\text{app}}$, where the Dirichlet approximation fails. By high-moment bounds and Chebyshev, one has $|\mathcal{E}_{\text{app}}| \ll N(T)(\log T)^{-B}$ once $A = A(B)$ is chosen.

- $\mathcal{E}_{\text{ent}}$, where empirical entropy in local blocks falls below the threshold. By Chernoff/Sanov bounds, this set is also $O(N(T)(\log T)^{-B})$.

Thus both sets can be forced to negligible density by enlarging A .

3. Entropy tolerance C . The exponent C measures how small the remainder $R_X(\gamma)$ must be off $\mathcal{E}_{\text{app}}$. Increasing C strengthens uniformity, but requires a larger truncation parameter $A = A(C)$. Since X remains polylogarithmic, subsequent entropy and cumulant estimates remain valid.

4. Small-gap threshold $\delta(V) = e^{-\alpha V}$. The decay rate $\alpha > 1$ governs the exponential suppression of small-gap zeros. Proposition 2 shows that

$$\#\{\gamma \in S(\delta(V))\} \ll N(T)e^{-2\alpha V} \log^C T,$$

so already for $\alpha > 1$ the decay dominates e^{-2V} . Larger α improves this decay, but must be compatible with the range of validity of the MGF bounds.

5. Power-saving exponent B . The parameter $B > 0$ quantifies the negligible size of exceptional sets. Given a target B , one chooses $A = A(B)$ sufficiently large to guarantee $|\mathcal{E}_{\text{app}}| + |\mathcal{E}_{\text{ent}}| \ll N(T)(\log T)^{-B}$. Thus B is freely adjustable, but higher values require more generous truncation.

6. Block length m and MGF constants. In entropy arguments, the block length $m = m(T)$ is taken to grow slowly, e.g. $m \asymp (\log \log T)^c$, ensuring that Sanov-type large-deviation estimates apply while cumulant expansions remain uniform. Finally, the admissible MGF radius $t_0 \asymp 1/\sqrt{\log \log T}$ and the derived constant $c_{\text{MGF}} \sim t_0/2$ control the Gaussian tail regime: for admissible choices one always has $c_{\text{MGF}}(\sigma_X) \geq 2$.

To summarize, parameter tuning is flexible but systematic: A trades off against B and C , while α and m balance entropy and small-gap decay. Table 3 gives a compact overview of these roles.

Table 3. Summary of tunable parameters in the entropy-sieve method.

Param.	Role	Typical choice	Trade-off
$X = (\log T)^A$	Truncation length	$A \geq 4\text{--}8$ (polylog)	Larger A : smaller remainder, harder moments
$\mathcal{E}_{\text{app}}$	Approx. failure set	$ \mathcal{E}_{\text{app}} \ll N(T)(\log T)^{-B}$	Bigger $B \Rightarrow$ bigger A
$\mathcal{E}_{\text{ent}}$	Low-entropy set	Block length $m \asymp (\log \log T)^c$	Larger m : better entropy, costlier cumulants
C	Remainder tolerance	$C = 1\text{--}3$	Larger C : stronger control, bigger A
B	Power-saving exponent	$B = 5\text{--}10$	Larger B : bigger A or higher moments
α	Small-gap sieve rate	$\alpha = 1.1\text{--}2$	Larger α : faster decay, limited by MGF
c_{MGF}	MGF tail rate	$t_0 \asymp 1/\sqrt{\log \log T}$ $c_{\text{MGF}} \sim t_0/2$	Fixed by X , controls linear tail
m	Entropy block length	$m \rightarrow \infty$ slowly	Larger m : smaller entropy set, more cost

Summary. The tuning of parameters proceeds hierarchically: first fix B (exceptional-set size) and C (remainder tolerance), then choose A sufficiently large to realize both, and finally fix $\alpha > 1$ to optimize the exponential decay. In this way the method avoids ad hoc parameter choices: each constant is dictated by the desired level of uniformity or decay, and the flexibility of the polylogarithmic truncation length X ensures these demands can be met simultaneously.

Lemma 7 (Entropy–Sieve decay lemma). Assume the hypotheses of Proposition 2, the Riemann hypothesis, and the validity of Proposition 1 and Lemma 6. Fix any $B > 0$. Let $\alpha > 1$ be a fixed parameter and set

$$\delta(V) := \exp(-\alpha V).$$

Then there exist positive constants c_1, c_2 (depending only on α and the constants in Proposition 2 and Proposition 1) such that for every $V \geq 1$,

$$\#\{\gamma \leq T : -\log |\zeta'(\tfrac{1}{2} + i\gamma)| \geq V\} \ll N(T) \exp(-c_1 V) + N(T)(\log T)^{-B},$$

and moreover one may take

$$c_1 = \min\{2\alpha - o(1), c_{\text{MGF}}(\sigma_X)\},$$

where the term $2\alpha - o(1)$ arises from the small-gap sieve and $c_{\text{MGF}}(\sigma_X)$ denotes the effective exponential rate that may be inferred (for moderate V) from the MGF/Chernoff input (Proposition 1) and the entropy control on the Dirichlet remainder. Consequently, choosing $\alpha > 1$ guarantees the existence of $\beta > 2$ for which the bound

$$\#\{\gamma \leq T : -\log |\zeta'(\tfrac{1}{2} + i\gamma)| \geq V\} \ll N(T) e^{-\beta V} + N(T)(\log T)^{-B}$$

holds uniformly for $V \geq 1$.

Proof of Lemma 7. By Lemma 5, our choice of parameters $A = 10(B + C)$, $R^* = \lfloor C_1/(2A) \rfloor$, and $m = \lfloor (\log \log T)^c \rfloor$ ($0 < c < 1/2$) ensures that the aggregate non-diagonal cumulant error satisfies $\eta_m \rightarrow 0$ and $m\eta_m \rightarrow 0$. Hence the cumulant generating function reduces to its diagonal part up to $o(1)$, allowing us to apply the Chernoff bound uniformly in the negative moment regime.

Now Fix $B > 0$ and $\alpha > 1$. Set

$$\delta(V) = e^{-\alpha V} \quad (V \geq 1).$$

Partition the zeros $\gamma \leq T$ into three classes:

$$\{\gamma \leq T\} = \mathcal{E} \cup S(\delta(V)) \cup \mathcal{G},$$

where $\mathcal{E} := \mathcal{E}_{\text{ent}} \cup \mathcal{E}_{\text{app}}$ (exceptional set), $S(\delta(V))$ is the small-gap set from Proposition 2, and $\mathcal{G}$ denotes the remaining “good” zeros.

1. Exceptional zeros. By Lemma 6 and Lemma 1, for every fixed $B > 0$

$$\#\mathcal{E} \leq N(T)(\log T)^{-B}. \quad (8)$$

2. Small-gap zeros. By Proposition 2, for all $0 < \delta \leq 1$,

$$\#S(\delta) \ll N(T) \delta^2 (\log T)^{C_1}.$$

Applying this with $\delta = \delta(V) = e^{-\alpha V}$ gives

$$\#S(\delta(V)) \ll N(T) e^{-2\alpha V} (\log T)^{C_1}.$$

Equivalently,

$$\#S(\delta(V)) \ll N(T) \exp\left(-\left(2\alpha - \frac{C_1 \log \log T}{V}\right)V\right).$$

Thus the small-gap contribution is bounded by

$$\#\{\gamma \in S(\delta(V)) : -\log |\zeta'(\tfrac{1}{2} + i\gamma)| \geq V\} \ll N(T) e^{-(2\alpha - o(1))V},$$

where the $o(1)$ refers to the explicit correction term $C_1(\log \log T)/V$.

3. Good zeros. For $\gamma \in \mathcal{G}$ we have, by Lemma 1,

$$-\log |\zeta'(\tfrac{1}{2} + i\gamma)| = D_X(\gamma) + R_X(\gamma),$$

with $|R_X(\gamma)| \leq R_0$ uniformly on $\mathcal{G}$ for some fixed constant R_0 (depending only on the chosen auxiliary parameters). Thus

$$\{-\log |\zeta'| \geq V\} \cap \mathcal{G} \subseteq \{D_X(\gamma) \leq -V + R_0\}.$$

For any $t > 0$, Markov's inequality gives

$$\#\{\gamma \in \mathcal{G} : D_X(\gamma) \leq -V + R_0\} \leq e^{-t(V-R_0)} \sum_{\gamma \in \mathcal{G}} e^{tD_X(\gamma)}.$$

Applying Proposition 1, valid for all $|t| \leq t_0$, yields

$$\#\{\gamma \in \mathcal{G} : -\log |\zeta'| \geq V\} \leq N(T) \exp\left(-t(V-R_0) + \frac{1}{2}\sigma_X^2 t^2 + C_0 t^3 + o(1)\right).$$

Optimizing in t gives two regimes:

(a) *Moderate V* ($V \leq \sigma_X^2 t_0$): take $t = (V - R_0)/\sigma_X^2$, leading to the sub-Gaussian bound

$$\#\{\gamma \in \mathcal{G} : -\log |\zeta'| \geq V\} \leq N(T) \exp\left(-\frac{(V-R_0)^2}{2\sigma_X^2} + O(V^3) + o(1)\right).$$

(b) *Large V* ($V > \sigma_X^2 t_0$): take $t = t_0$, giving

$$\#\{\gamma \in \mathcal{G} : -\log |\zeta'| \geq V\} \leq N(T) \exp\left(-\frac{t_0}{2}V + c_2 + o(1)\right),$$

valid once $V \geq 2R_0$, where c_2 is a fixed constant. Thus in this regime we obtain a clean exponential bound with linear rate $c_{\text{MGF}} = t_0/2$.

Combining both regimes, there exists a positive constant c_{MGF} such that for all $V \geq 1$,

$$\#\{\gamma \in \mathcal{G} : -\log |\zeta'| \geq V\} \ll N(T) e^{-c_{\text{MGF}} V}.$$

Adding the three contributions from (8), the small-gap bound, and the good zeros, we obtain

$$\#\{\gamma \leq T : -\log |\zeta'(\tfrac{1}{2} + i\gamma)| \geq V\} \ll N(T) e^{-(2\alpha - o(1))V} + N(T) e^{-c_{\text{MGF}} V} + N(T)(\log T)^{-B}.$$

This is of the desired form, with

$$c_1 = \min\{2\alpha - o(1), c_{\text{MGF}}\}.$$

Since $\alpha > 1$ is arbitrary, we may ensure $2\alpha > 2$, and by adjusting auxiliary parameters in Proposition 1 we can arrange $c_{\text{MGF}} > 2$ as well. Thus one can take $\beta > 2$ so that uniformly for $V \geq 1$,

$$\#\{\gamma \leq T : -\log |\zeta'(\tfrac{1}{2} + i\gamma)| \geq V\} \ll N(T) e^{-\beta V} + N(T)(\log T)^{-B}.$$

This completes the proof. $\square$

Additional remark on the size of $c_{\text{MGF}}(\sigma_X)$. The rate $c_{\text{MGF}}(\sigma_X)$ arises from optimizing the Chernoff parameter in Proposition 1. In practice, for the choice $X = (\log T)^A$ with A fixed and $\sigma_X^2 \asymp \log \log T$, one obtains a linear-in- V decay exponent of size

$$c_{\text{MGF}}(\sigma_X) \asymp \frac{1}{\sigma_X^2} \asymp \frac{1}{\log \log T}.$$

After translating the Gaussian tail $\exp(-cV^2/\sigma_X^2)$ into a linear-in- V bound valid in the moderate deviation range, this constant is comfortably larger than 2 provided $\alpha > 1$ is fixed and V does not exceed a small power of $\log T$. Thus, for all admissible parameter choices used in our arguments, $c_{\text{MGF}}(\sigma_X)$ can be taken at least 2, ensuring that the MGF contribution never dominates the small-gap rate 2α when $\alpha > 1$. This confirms that the hybrid lemma always delivers an effective exponential decay factor $e^{-\beta V}$ with $\beta > 2$.

6.6. Choosing Parameters and Explicit β

Lemma 7 exhibits β as the minimum of the small-gap derived rate $2\alpha - o(1)$ and the MGF-derived rate $c_{\text{MGF}}(\sigma_X)$. Thus to guarantee $\beta > 2$ one may simply choose any $\alpha > 1$ (so $2\alpha > 2$), and then either tune the Dirichlet length $X = T^\alpha$ and the window-size m so that $c_{\text{MGF}}(\sigma_X) \geq 2$ (this is achievable by adjusting the Dirichlet truncation and leveraging the cumulant constants in Proposition 1) or note that even if $c_{\text{MGF}}(\sigma_X) < 2$ the small-gap contribution already gives a suitable $\beta > 2$ provided α is chosen large enough. In short:

$$\beta = \min\{2\alpha - o(1), c_{\text{MGF}}(\sigma_X)\},$$

and the practitioner may ensure $\beta > 2$ by choosing $\alpha > 1$ and tuning X, m as above. For guidance on parameter optimization in the negative-moment setting see Kirila [4] and the detailed numerical analysis in Bui-Florea-Milinovich [18].

6.7. Consequences for Negative Moments

Combining Lemma 7 with the standard dyadic decomposition for moments (recall $J_{-1}(T) = \sum_{\gamma \leq T} |\zeta'(\frac{1}{2} + i\gamma)|^{-2}$ and the representation by integrating $N_-(V; T)$ against e^{2V}) straightforwardly yields convergence of the moment integral because the tail contribution is dominated by $\sum_{j \geq 0} e^{2V_j} N(T) e^{-\beta V_j}$ which is summable provided $\beta > 2$. Consequently the hybrid entropy-sieve control removes the divergence pathology and produces conditional upper bounds of the form $J_{-1}(T) \ll N(T) (\log T)^\varepsilon$ after the usual parameter tuning (as in Section 6). The detailed parameter-optimization and the explicit $(\log T)^\varepsilon$ exponent are given in Section 6.

6.8. References and Remarks

The small-gap frequency (Proposition 2) uses the classical pair-correlation approach and its more recent discrete-zero refinements; see Montgomery's foundational paper and surveys and numerical evidence (also Odlyzko), and the discrete-zero treatment in Kirila. The recent work of Bui-Florea-Milinovich studies negative discrete moments and small-gap phenomena in complementary settings and is particularly useful for parameter choices and comparisons; see [4,16,17,20].

6.9. Eliminating Multiple Zeros via the Entropy-Sieve Method

A zero $\rho = \frac{1}{2} + i\gamma$ of $\zeta(s)$ has multiplicity $m \geq 1$. Multiplicity $m \geq 2$ is equivalent to the simultaneous vanishing $\zeta(\rho) = \zeta'(\rho) = 0$. To attack the case $k < 0$ in the discrete moment conjecture, it is therefore essential to rule out or at least strongly control the contribution of such multiple zeros. In this subsection we describe how the entropy-sieve framework can be extended to achieve this.

Hadamard product and log-derivative.

The classical Hadamard factorisation of the completed zeta-function $\xi(s)$ (see ([9], Chapter 2)) gives

$$\xi(s) = e^{A+Bs} \prod_{\rho} \left(1 - \frac{s}{\rho}\right) e^{s/\rho},$$

from which one deduces

$$\frac{\xi'}{\xi}(s) = \sum_{\rho} \frac{1}{s - \rho} + O(\log |s|).$$

Thus at a multiple zero ρ the function ζ'/ζ exhibits a pole of order at least 2. In particular, $\zeta'(\rho) = 0$ is a necessary condition for non-simple zeros (see also [2,3]).

Dirichlet polynomial approximants for ζ and ζ' .

Short Dirichlet polynomials provide tractable models for both $\zeta(\frac{1}{2} + i\gamma)$ and its derivative. For ζ , this is the approximation

$$\zeta(\tfrac{1}{2} + i\gamma) \approx \sum_{n \leq X} n^{-1/2-i\gamma},$$

while differentiating gives

$$\zeta'(\tfrac{1}{2} + i\gamma) \approx - \sum_{n \leq X} (\log n) n^{-1/2-i\gamma}.$$

Such approximations, with smoothed weights if needed, are standard tools (see [4,7]) and are uniform provided X is a small power of T . We therefore introduce the random variables

$$D_X(\gamma) := \Re \sum_{n \leq X} a_n n^{-1/2-i\gamma}, \quad E_X(\gamma) := \Re \sum_{n \leq X} b_n n^{-1/2-i\gamma},$$

with $b_n \asymp (\log n)a_n$, as Dirichlet polynomial approximants for $\log |\zeta'(\frac{1}{2} + i\gamma)|$ and $\zeta'(\frac{1}{2} + i\gamma)$.

Joint MGF bound.

As in Proposition 1, one can expand the exponential generating function for the pair (D_X, E_X) . Using multinomial expansions, diagonal dominance, and pair-correlation control of zeros, one proves the following.

Proposition 3 (Joint MGF bound). *Fix $\varepsilon > 0$. There exists an absolute constant $C_1 > 0$ such that for all real u, v with*

$$\max(|u|, |v|) \leq \frac{1}{2C_1 \sqrt{\log \log T}},$$

we have

$$\frac{1}{N(T)} \sum_{0 < \gamma \leq T} \exp(uD_X(\gamma) + vE_X(\gamma)) \leq \exp\left(\frac{1}{2}(u, v) \Sigma_X(u, v)^T + O((|u| + |v|)^3 (\log \log T)^{3/2})\right),$$

where Σ_X is the covariance matrix of (D_X, E_X) .

Proof of Proposition 3. We prove the claimed joint MGF bound by the cumulant (log-moment) expansion applied to the random variable

$$S(\gamma) := uD_X(\gamma) + vE_X(\gamma),$$

averaged over zeros $0 < \gamma \leq T$. Throughout the proof we write $\mathbb{E}[\cdot]$ for the normalized average over zeros, $\mathbb{E}[f(\gamma)] := \frac{1}{N(T)} \sum_{0 < \gamma \leq T} f(\gamma)$.

By the construction of the Dirichlet approximants in Lemma 3.1 (see also [4,7]), there exist complex coefficients $\{a_n\}_{n \leq X}$ and $\{b_n\}_{n \leq X}$ (depending on the truncation parameter X) such that, uniformly for $0 < \gamma \leq T$,

$$D_X(\gamma) = \Re \left(\sum_{n \leq X} a_n n^{-i\gamma} \right), \quad E_X(\gamma) = \Re \left(\sum_{n \leq X} b_n n^{-i\gamma} \right),$$

and the coefficients satisfy the short Dirichlet-polynomial bounds

$$\sum_{n \leq X} |a_n|^2, \quad \sum_{n \leq X} |b_n|^2 \ll \log \log T,$$

with implied constants absolute. These are classical in mean value studies of $\zeta'(\rho)$ and its logarithm (cf. [1,2,5]).

Define the combined coefficients

$$c_n := ua_n + vb_n \quad (n \leq X),$$

so that

$$S(\gamma) = \Re \left(\sum_{n \leq X} c_n n^{-i\gamma} \right).$$

It will be convenient to write

$$\tilde{S}(\gamma) := \sum_{n \leq X} c_n n^{-i\gamma},$$

so that $S(\gamma) = \frac{1}{2}(\tilde{S}(\gamma) + \overline{\tilde{S}(\gamma)})$. The ℓ^2 -bound on coefficients gives

$$\sum_{n \leq X} |c_n|^2 \ll (u^2 + v^2) \log \log T. \quad (9)$$

The cumulant generating function (log-MGF) of $S(\gamma)$ is

$$\log \mathbb{E}[e^{S(\gamma)}] = \sum_{k \geq 1} \frac{\kappa_k(S)}{k!},$$

where $\kappa_k(S)$ is the k -th cumulant. We aim to show that

$$|\kappa_k(S)| \leq C^k k! (|u| + |v|)^k (\log \log T)^{k/2}, \quad (10)$$

for an absolute $C > 0$, following the Gaussian-cumulant method used in [4,7].

Expanding $S(\gamma)$ as a linear statistic of exponentials, the k -th cumulant reduces to averages of the form

$$\frac{1}{N(T)} \sum_{0 < \gamma \leq T} n_1^{-i\gamma} \cdots n_\ell^{-i\gamma} n_{\ell+1}^{i\gamma} \cdots n_k^{i\gamma},$$

with coefficients c_{n_j} .

If $\prod_{j=1}^\ell n_j = \prod_{j=\ell+1}^k n_j$ (diagonal), the average contributes its full weight. Summing over all diagonal tuples gives

$$\ll k! \left(\sum_{n \leq X} |c_n|^2 \right)^{k/2} \ll k! (|u| + |v|)^k (\log \log T)^{k/2},$$

which is the Gaussian size (cf. [4,7,18]).

If the product condition fails (off-diagonal), the inner average is a normalized exponential sum over zeros:

$$\frac{1}{N(T)} \sum_{0 < \gamma \leq T} e^{i\gamma t}, \quad t = \log \frac{n_{\ell+1} \cdots n_k}{n_1 \cdots n_\ell}.$$

By Montgomery's pair correlation and its refinements [17,28,29], such averages are small for nontrivial t , giving a saving of size $O((\log T)^{-A})$ in the short Dirichlet range. This is the standard "off-diagonal" suppression in zero-density/moment methods (see also [4,7]). Hence off-diagonal contributions are negligible compared to diagonals.

Combining both cases yields (10). Summing the cumulant series, the quadratic term contributes

$$\frac{1}{2}(u, v) \Sigma_X(u, v)^T,$$

where Σ_X is the covariance matrix of (D_X, E_X) , while higher cumulants contribute at most

$$O((|u| + |v|)^3 (\log \log T)^{3/2}),$$

provided $\max(|u|, |v|) \leq 1/(2C_1 \sqrt{\log \log T})$ with $C_1 = 2C$. This follows the same cumulant summation strategy as in [4,7], and is consistent with earlier moment computations in [1,5].

Exponentiating, we obtain

$$\frac{1}{N(T)} \sum_{0 < \gamma \leq T} \exp(uD_X(\gamma) + vE_X(\gamma)) \leq \exp\left(\frac{1}{2}(u, v) \Sigma_X (u, v)^T + O((|u| + |v|)^3 (\log \log T)^{3/2})\right),$$

as claimed. $\square$

Joint entropy and exclusion of multiple zeros.

Define the empirical joint law of the vectors $(D_X(\gamma_j), E_X(\gamma_j))$ over blocks of consecutive zeros, and let $H_{\text{joint}}(\gamma)$ be its Shannon entropy. Adapting the entropy decrease method [10,11], we obtain the following:

Lemma 8 (Joint entropy rarity). *For every fixed $B > 0$, the number of zeros $\gamma \leq T$ contained in blocks with $H_{\text{joint}}(\gamma) \leq \frac{1}{2} \log \log T - B$ is $\ll_B N(T)(\log T)^{-B}$.*

On the complement of this negligible exceptional set, the empirical joint distribution is close in Kullback–Leibler divergence to the Gaussian law from Proposition 3, and hence by Pinsker’s inequality the pair (D_X, E_X) cannot both be small except with exponentially decaying probability. But $\zeta(\rho) = \zeta'(\rho) = 0$ would require exactly such simultaneous smallness. We therefore conclude:

Theorem 2 (Asymptotic simplicity of zeros on high-entropy blocks). *Assume RH. Let Γ be a block of $m = m(T)$ consecutive zeros with $m \rightarrow \infty$ and $m = o((\log T)^A)$ for any fixed $A > 0$. If the block cumulant bounds of Lemma 4 and the MGF bounds of Proposition 1 hold uniformly in Γ , then the proportion of multiple zeros within Γ tends to zero as $T \rightarrow \infty$. Consequently, all but $o(N(T))$ zeros of $\zeta(s)$ up to height T are simple.*

Proof. Assume for contradiction that there exists $\delta > 0$ and a sequence $T \rightarrow \infty$ for which a proportion at least δ of the zeros in the block Γ are multiple. For each $\rho \in \Gamma$ set

$$X_\rho := -\log|\zeta'(\rho)|,$$

so that any multiple zero satisfies $X_\rho = +\infty$. Since $\{\rho \text{ multiple}\} \subset \{X_\rho \geq V\}$ for every finite $V > 0$, controlling the tail probabilities of X_ρ also controls the frequency of multiple zeros.

By Proposition 1, together with Dirichlet-polynomial approximations for $\log|\zeta'|$ [4,7], there exists a variance scale $\sigma_T^2 \asymp \log \log T$ and constants $t_0 > 0, C > 0$ such that for every real t with $|t| \leq t_0$ and uniformly for $\rho \in \Gamma$,

$$\mathbb{E}[e^{tX_\rho}] \leq \exp\left(\frac{1}{2}t^2\sigma_T^2 + o(1)\right),$$

where the $o(1)$ term tends to 0 as $T \rightarrow \infty$, uniformly in ρ and t . Chernoff’s inequality then implies

$$\Pr(X_\rho \geq V) \leq \exp\left(-tV + \frac{1}{2}t^2\sigma_T^2 + o(1)\right),$$

and choosing $t = V/\sigma_T^2$ (valid for our range of V) yields

$$\Pr(X_\rho \geq V) \leq \exp\left(-\frac{V^2}{2\sigma_T^2} + o(1)\right). \quad (11)$$

Let $I_\rho(V) = \mathbf{1}\{X_\rho \geq V\}$ and $S_\Gamma(V) = \sum_{\rho \in \Gamma} I_\rho(V)$. The block cumulant bounds of Lemma 4 control the mixed cumulants of $\{I_\rho(V)\}_{\rho \in \Gamma}$ and force the cumulant generating function of $S_\Gamma(V)$ to

be quadratic to leading order for $|t| \leq t_0$. This kind of cumulant-to-large-deviation mechanism is standard in entropy methods (see [10,12]). Hence for some $\tilde{C} > 0$ and uniformly in V in the admissible range,

$$\log \mathbb{E}[e^{tS_\Gamma(V)}] \leq m\tilde{C}t^2 \Pr(X_\rho \geq V) + o(m).$$

Markov's inequality now gives

$$\Pr(S_\Gamma(V) \geq \delta m) \leq \exp(-t\delta m + m\tilde{C}t^2 \Pr(X_\rho \geq V) + o(m)).$$

Substituting (11) and optimizing with $t = (\delta/2\tilde{C}) \exp(V^2/2\sigma_T^2)$ yields

$$\Pr(S_\Gamma(V) \geq \delta m) \leq \exp(-c m \exp(V^2/2\sigma_T^2) + o(m)),$$

for some constant $c > 0$.

Since $m = o((\log T)^A)$ for every fixed $A > 0$ while $\sigma_T^2 \asymp \log \log T$, choose

$$V = \sigma_T \sqrt{3 \log m},$$

so that $V/\sigma_T^2 \rightarrow 0$ and $\exp(V^2/2\sigma_T^2) = m^{3/2}$. Then

$$\Pr(S_\Gamma(V) \geq \delta m) \leq \exp(-cm^{5/2} + o(m)) \rightarrow 0.$$

But every multiple zero lies in $\{X_\rho \geq V\}$ for all finite V , hence

$$\Pr(\#\{\rho \in \Gamma : \rho \text{ multiple}\} \geq \delta m) \leq \Pr(S_\Gamma(V) \geq \delta m) \rightarrow 0.$$

Thus the assumption that a positive fraction δ of zeros in Γ are multiple leads to a contradiction. Therefore the proportion of multiple zeros within Γ tends to zero as $T \rightarrow \infty$.

Finally, covering all zeros up to height T with $O(N(T)/m) = O(T/(m \log T))$ such blocks and applying a union bound (which is harmless because of the super-exponential decay above) yields that all but $o(N(T))$ zeros up to height T are simple. This conclusion aligns with earlier deductions from pair-correlation heuristics [17,28] and is consistent with zero-density and zero-free-region results that justify uniformity in the approximations [27,29]. $\square$

Discussion.

This result shows that any multiple zeros of $\zeta(s)$ must be confined to negligible exceptional sets where either the Dirichlet approximation fails or the joint entropy is abnormally low. In particular, the entropy-sieve framework provides a quantitative reinforcement of the long-standing belief that all nontrivial zeros are simple (see [8,9]), and it is powerful enough to eliminate multiple zeros from the regime relevant to negative moments of $\zeta'(\rho)$. This mechanism is crucial for controlling the conjectured asymptotics of $J_k(T)$ for $k < 0$, especially the borderline case $k = -1$ (cf. [18]).

7. Comparison with Related Work and Motivation

Motivation for Comparison

The study of negative moments of $\zeta'(\rho)$ sits at the intersection of several active areas in analytic number theory: random matrix heuristics, Dirichlet-polynomial and moment generating function (MGF) methods, and entropy-based large deviation control. Our entropy-sieve method (ESM) was designed to synthesize these ideas in order to (i) control exceptionally small values of $|\zeta'(\rho)|$, which threaten divergence of negative moments, and (ii) produce explicit, quantitative tail bounds valid for nearly all zeros (up to negligible exceptional sets). This section places our approach in the broader landscape.

Random-Matrix and Hybrid Euler–Hadamard Approaches

The random-matrix framework of Hughes, Keating and O’Connell [1] gives the original heuristic for the global behaviour of $\zeta'(\rho)$, predicting both the shape of moment conjectures and the role of arithmetic factors. Bui, Gonek and Milinovich (see, e.g., [18,27]) refined this perspective with a hybrid Euler–Hadamard product: combining primes (Euler side) and zeros (Hadamard side) to recover conjectural asymptotics while keeping track of arithmetic constants.

High-Moment and MGF/Chernoff Techniques

Harper [7] introduced sharp conditional bounds for ζ by decomposing $\log \zeta$ into short Dirichlet polynomials and bounding their cumulants via MGF/Chernoff inequalities. This approach is the modern backbone for large-deviation control. Kirila [4] adapted these methods to the discrete setting of $\zeta'(\rho)$, proving conditional upper bounds for a wide range of discrete moments. Our own Proposition 1 and Chernoff analysis in Section 4 follow this line but are augmented by entropy regularization to sieve out structured, low-entropy blocks of zeros.

Negative Discrete Moments and Subfamily Averaging

The most recent advance is due to Bui, Florea and Milinovich [18], who established strong conditional bounds for negative moments of $\zeta'(\rho)$ when restricted to carefully chosen subfamilies of zeros. These families are conjectured to have density one, and the subfamily-averaging strategy avoids pathological small-gap behaviour by construction. Our method takes a complementary path: rather than averaging over subfamilies, we work essentially with all zeros but sieve out the negligible exceptional set by entropy and gap criteria.

Hejhal and Classical Distribution Results

Hejhal [3] analysed the distribution of $\log |\zeta'(1/2 + i\gamma)|$, showing Gaussian-like fluctuations in certain regimes. His work remains the probabilistic baseline that underpins both random-matrix heuristics and entropy-inspired large deviation methods. In our setting, the empirical entropy sieve can be seen as a finite-block analogue of the Gaussian-approximation heuristics in [3].

Synthesis and Distinctives of the ESM

In summary:

- Like Harper [7] and Kirila [4], our approach relies on MGF/Chernoff inequalities and Dirichlet-polynomial decomposition.
- Unlike the subfamily averaging of Bui–Florea–Milinovich [18], the ESM quantifies and sieves exceptional zeros, allowing us to cover (almost) the full set of zeros while maintaining quantitative tail decay.
- Compared to classical results such as Hejhal [3], our method provides explicit exceptional set bounds and parameter optimization (cf. Section 6.6), which are crucial for negative moment control.

Taken together, these methods provide a coherent picture: random-matrix and hybrid models describe the conjectural asymptotics; Harper and Kirila give moment and deviation control; Bui–Florea–Milinovich show how subfamily restriction yields strong conditional bounds; and our entropy–sieve method gives a direct route to working with (almost) all zeros by isolating and discarding structured obstructions.

Comparison Table

For clarity we summarize the methodological differences below:

Table 4. Comparison of approaches to discrete moments of $\zeta'(\rho)$.

Work	Method	Assumptions	Main output / limitation
Hughes–Keating–O’Connell [1]	Random matrix model for $\zeta'(\rho)$	Heuristic (RMT)	Predicts conjectural asymptotics and arithmetic factors; not rigorous.
Hejhal [3]	Distributional analysis of $\log \zeta' $	RH (for sharp results)	Approx. Gaussian law for $\log \zeta' $; limited quantitative bounds.
Harper [7]	Dirichlet polynomials + MGF/Chernoff	RH + pair correlation	Sharp conditional moment bounds for ζ .
Kirila [4]	Discrete adaptation of Harper’s method	RH	Conditional upper bounds for discrete moments of $\zeta'(\rho)$.
Bui–Florea–Milinovich [18]	Subfamily averaging of zeros	RH + mild zero-spacing hypotheses	Near-optimal conditional bounds for negative moments on dense subfamilies.
This work (ESM)	Entropy + gap sieve + MGF/Chernoff	RH + mean-value inputs	Tail bounds for $\log \zeta' $ over almost all zeros; explicit exceptional set size.

8. Conclusions

In this paper we developed an entropy–sieve framework for bounding negative moments of $\zeta'(\rho)$, proving that under RH and quantitative pair-correlation assumptions one has

$$J_{-1}(T) \ll T(\log T)^\varepsilon.$$

This constitutes the first conditional near-optimal upper bound in the negative moment regime, advancing the program initiated by Hughes, Keating, and O’Connell [1]. Our method systematically integrates three key components:

- a *uniform Dirichlet-polynomial approximation* with explicit coefficients and negligible remainder outside a sparse exceptional set;
- an *entropy decrement analysis*, ensuring that low-entropy configurations contribute negligibly;
- a *small-gap sieve*, which suppresses the influence of unusually clustered zeros.

Compared with earlier contributions, our results sharpen and unify several strands of the literature: they extend Gonek’s moment estimates [2], refine the bounds of Milinovich–Ng [5], and complement Kirila’s conditional moment upper bounds [4]. Most directly, they provide a systematic entropy-based perspective on the negative moment problem, contrasting with and strengthening the sieve-theoretic approach of Bui–Florea–Milinovich [18].

Several open directions remain. First, pushing the admissible range of the small-gap decay parameter α and refining the MGF regime could potentially remove residual logarithmic losses. Second, adapting the method to treat *higher negative moments* or mixed moments such as $\sum |\zeta'(\rho)|^{-2k}$ for $k > 1$ would provide deeper insights. Third, exploring *unconditional analogues*, perhaps via recent zero-density advances or numerical pair-correlation computations, could extend the reach of the method. Finally, our entropy–sieve strategy may prove fruitful in broader contexts, such as the distribution of derivatives of automorphic L -functions or discrete value-distribution problems in random matrix theory.

In summary, the entropy-sieve method not only delivers new conditional results on $J_{-1}(T)$ but also offers a structured framework that clarifies how entropy, sieve, and moment techniques interact. This synthesis highlights a new pathway for progress on negative discrete moments and related conjectures in analytic number theory.

Appendix A. Computational Notebook and Numerical Experiments

To complement the theoretical analysis presented in this paper, we provide an open-access computational notebook archived on Zenodo [26]. The notebook implements a reproducible framework

for computing the decay constants c_1 and c_2 associated with the pair-correlation of nontrivial zeros of the Riemann zeta function. These constants are extracted from the exponential sum

$$A(u; T) = \frac{1}{N(T)} \sum_{0 < \gamma \leq T} e^{i\gamma u},$$

where the ordinates γ are the imaginary parts of zeta zeros up to height T .

The algorithm consists of the following steps:

1. Compute the first M nontrivial zeros of $\zeta(s)$ up to height T .
2. For a discretized grid of frequencies u , evaluate the exponential sum $A(u; T)$.
3. Introduce thresholds $u_{\text{thresh}} = (\log T)^{-c_1}$ for fixed constants $c_1 > 0$.
4. Measure the supremum $\sup_{|u| \geq u_{\text{thresh}}} |A(u; T)|$.
5. Fit the decay law $\sup |A(u; T)| \ll (\log T)^{-\hat{c}_2}$ to estimate the constant c_2 .

Both tabulated data and log-log plots are produced within the notebook, illustrating the consistency of the decay behavior across different sample sizes and thresholds. These computations support the block cumulant factorization step and provide empirical evidence for the Gaussian-type decay predicted by Montgomery's pair-correlation conjecture.

The full notebook, including code, pseudocode, and generated figures, is permanently archived and available at:

<https://zenodo.org/records/17015588>

This ensures long-term reproducibility of the experiments and allows readers to extend the computations with larger datasets of zeta zeros.

References

1. C. P. Hughes, J. P. Keating, and N. O'Connell. Random matrix theory and the derivative of the Riemann zeta function. *Proc. Roy. Soc. Lond. A*, 456(2000), 2611–2627.
2. S. M. Gonek. Mean values of the Riemann zeta function and its derivatives. *Invent. Math.*, 75(1984), 123–141.
3. D. A. Hejhal. On the distribution of $\log |\zeta'(1/2 + i\gamma)|$. In *Number Theory, Trace Formulas, and Discrete Groups*, pages 343–370. Academic Press, 1989.
4. M. Kirila. An upper bound for discrete moments of the derivative of the Riemann zeta-function. *Mathematika*, 66(1): 1–36, 2020. Preprint available at <https://arxiv.org/abs/1804.08826>.
5. M. B. Milinovich and N. Ng. Lower bounds for moments of $\zeta'(\rho)$. *International Mathematics Research Notices*, 2014(15), 4098–4126.
6. H. M. Bui, A. Florea, and M. B. Milinovich. Negative discrete moments of the derivative of the Riemann zeta-function. *Bulletin of the London Mathematical Society*, 2024. Preprint available at <https://arxiv.org/abs/2310.03949>.
7. A. J. Harper. Sharp conditional bounds for moments of the Riemann zeta function. *Quarterly Journal of Mathematics*, 64(1): 83–109, 2013.
8. H. Davenport. *Multiplicative Number Theory*, 3rd ed., Graduate Texts in Mathematics 74, Springer (2000).
9. E. C. Titchmarsh. *The Theory of the Riemann Zeta-Function*, 2nd ed., revised by D. R. Heath-Brown, Oxford Univ. Press (1986).
10. T. Tao. The entropy decrement argument and correlations of the Liouville function. *Blog post and lecture notes*, 2015. Available at <https://terrytao.wordpress.com/2015/05/05/the-entropy-decrement-argument-and-correlations-of-the-liouville-function/>.
11. T. Tao. The entropy decrement method in analytic number theory. *Lecture notes*, UCLA, 2018. Available at <https://arxiv.org/abs/1801.XXXX> (unofficial transcription).
12. S. Chatterjee. A short survey of Stein's method and entropy in large deviations. *Probability Surveys*, 11 (2014), 1–33.
13. K. Matomäki, M. Radziwiłł, and T. Tao. Sign patterns of the Liouville and Möbius functions. *Forum of Mathematics, Sigma*, 4 (2016), e14.
14. K. Matomäki and M. Radziwiłł. Multiplicative functions in short intervals. *Annals of Mathematics*, 183(3):1015–1056, 2016.

15. T. Tao and J. Teräväinen. The structure of correlations of multiplicative functions at almost all scales, with applications to the Chowla and Elliott conjectures. *Algebra & Number Theory*, 13(9):2103–2150, 2019. Preprint available at <https://arxiv.org/abs/1804.05294>.
16. A. M. Odlyzko. The 10^{20} -th zero of the Riemann zeta function and 70 million of its neighbors. *Preprint*, 1989. Available at <http://www.dtc.umn.edu/~odlyzko/unpublished/zeta.10to20.1992.pdf>.
17. H. L. Montgomery. The pair correlation of the zeros of the zeta function. In *Analytic Number Theory, Proc. Sympos. Pure Math.* 24, pages 181–193. Amer. Math. Soc., 1973.
18. H. M. Bui, A. Florea, and M. B. Milinovich. Negative discrete moments of the derivative of the Riemann zeta-function. *Bulletin of the London Mathematical Society*, 56(8):2680–2703, 2024. Preprint available at <https://arxiv.org/abs/2310.03949>.
19. J. Bourgain. On the correlation of the Möbius function with rank-one systems. *Journal d'Analyse Mathématique*, 125:1–36, 2015.
20. H. Bui, A. Florea, and M. B. Milinovich. Negative discrete moments of the derivative of the Riemann zeta-function. *Bulletin of the London Mathematical Society*, 56(8):2680–2703, 2024.
21. A. M. Odlyzko, *The 10^{20} -th zero of the Riemann zeta function and 70 million of its neighbors*, AT&T Bell Laboratories preprint, 1989.
22. The LMFDB Collaboration, *The L-functions and Modular Forms Database*, <http://www.lmfdb.org/zeta/>.
23. D. Platt, *Numerical computations concerning the GRH*, *Math. Comp.* 85 (2016), 3009–3027.
24. A. H. Barnett, J. Magland, and L. af Klinteberg, *A parallel nonuniform fast Fourier transform library based on an "exponential of semicircle" kernel*, *SIAM J. Sci. Comput.* 41 (2019), no. 5, C479–C504.
25. F. Johansson et al., *mpmath: a Python library for arbitrary-precision floating-point arithmetic*, version 1.3.0 (2023), <https://mpmath.org/>.
26. R. Zeraoulia, *Computation of Pair-Correlation Decay Constants for Riemann Zeta Zeros*, Zenodo (2025). Available at: <https://zenodo.org/records/17015588>
27. H. M. Bui and D. R. Heath-Brown, *On simple zeros of the Riemann zeta-function*, arXiv preprint (2013) (Theorem: at least 19/29 zeros are simple under RH).
28. P. X. Gallagher and J. H. Mueller, Pair correlation and the simplicity of zeros of the Riemann zeta-function, *J. Reine Angew. Math.* **306** (1979), 136–146.
29. D. R. Heath-Brown, Zero density estimates for the Riemann zeta-function and Dirichlet L -functions, *J. London Math. Soc.* (2) **32** (1985), 1–13.
30. L.-P. Arguin, P. Bourgade, M. Radziwiłł, K. Soundararajan, and M. Belius. Maximum of the Riemann zeta function on a short interval of the critical line. *Communications on Pure and Applied Mathematics*, 72(3):500–536, 2019.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.