

Article

Not peer-reviewed version

On the Hughes-Keating-O'Connell Conjecture: Quantified Negative Moment Bounds for $\zeta'(\rho)$ via Entropy-Sieve Methods Revisited

Rafik Zeraoulia *

Posted Date: 8 September 2025

doi: 10.20944/preprints202509.0489.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: Quantified Negative Moment Bounds for $\zeta'(\rho)$ via Entropy–Sieve Methods Revisited

Zeraoulia Rafik 

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

Abstract

We study the negative discrete moments of the derivative of the Riemann zeta function at its non-trivial zeros, in connection with the Hughes–Keating–O’Connell conjecture. Building on the works of Gonek, Milinovich–Ng, Kirila, and the recent breakthrough of Bui–Florea–Milinovich, we introduce a new *entropy–sieve method* (ESM). This framework combines short Dirichlet-polynomial approximations with entropy-based moment generating function bounds and a small-gap sieve, thereby controlling all appearances of $\zeta'(\rho)$ without assuming simplicity of zeros. Assuming the Riemann Hypothesis together with standard pair-correlation conjectures and a strengthened discrete moment hypothesis, we prove the quantified conditional bound $J_{-1}(T) = \sum_{0 < \gamma \leq T} \frac{1}{|\zeta'(\frac{1}{2} + i\gamma)|^2} \leq C(\varepsilon) T(\log T)^\varepsilon$, for every fixed $\varepsilon > 0$, with an explicit dependence of the implicit constant on ε . This matches, up to logarithmic factors, the conjectured order $J_{-1}(T) \asymp T$ and improves on all previous conditional results. The analysis introduces several innovations: (i) a full cumulant control lemma for Dirichlet polynomials; (ii) explicit, non-circular parameter selection for approximation lengths and moments; and (iii) an entropy–sieve hybrid decay lemma that quantifies large-deviation probabilities for $\zeta'(\rho)$. Beyond the negative moment problem, the entropy–sieve framework illustrates the strength of entropy techniques in analytic number theory and points toward applications to L -functions 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. Notation for general quantities, Dirichlet-polynomial approximations, moment generating functions, and entropy framework.

General Notation	
T	Height parameter for critical zeros of $\zeta(s)$. Zeros $\rho = \frac{1}{2} + i\gamma$ with $0 < \gamma \leq T$.
$N(T)$	Number of zeros $\rho = \frac{1}{2} + i\gamma$ with $0 < \gamma \leq T$.
ρ	Nontrivial zero of $\zeta(s)$, written $\rho = \frac{1}{2} + i\gamma$.
γ	Imaginary ordinate of a zero.
\mathcal{E}_{app}	Exceptional set where Dirichlet-polynomial approximation fails (Lemma 9).
\mathcal{E}_{ent}	Exceptional set of zeros lying in low-entropy blocks (Lemma 4).
\mathcal{G}	Set of “good” zeros (outside all exceptional sets).
Z_{simp}	Set of simple zeros $\{\rho : \zeta'(\rho) \neq 0\}$.
Dirichlet Polynomial Approximation	
X	Dirichlet polynomial length $X = (\log T)^A$.
A	Truncation exponent.
$D_X(\gamma)$	Approximant $D_X(\gamma) = \Re \sum_{n \leq X} a_n n^{-1/2-i\gamma}$.
a_n	Dirichlet polynomial coefficients.
$R_X(\gamma)$	Error term in approximation of $\log \zeta'(\rho) $.
σ_X^2	Variance of D_X : $\sigma_X^2 = \sum_{n \leq X} a_n ^2 / n$ (Lemma 2).
Y	Auxiliary parameter $Y = \exp((\log \log T)^2)$.
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 $D_X(\gamma)$.
M_r	Raw r -th moment of $D_X(\gamma)$.
t	Auxiliary parameter, $ t \leq c / \sigma_X$.
t_0	Admissible t : $t_0 = c / \sqrt{\log \log \log T}$ (Proposition 1).
$N_-(V; T)$	Tail count: $\#\{\gamma : -\log \zeta'(\rho) \geq V\}$.
V	Threshold parameter in tail bounds.
c_{MGF}	Constant governing MGF tail decay.
Entropy Framework (Section 5)	
$\mathcal{G}(\gamma_0)$	Local window of zeros near γ_0 .
$H_{h,\Delta}^{\text{val}}(\gamma_0)$	Value entropy (Definition 1).
$H_{h_g,\Delta}^{\text{gap}}(\gamma_0)$	Gap entropy of zero spacings (Definition 2).
H_0	Entropy threshold used to classify blocks.
h, \tilde{h}	Bin widths for histograms.
B_ℓ, \tilde{B}_ℓ	Histogram bins for values and gaps.
$p_\ell(\gamma)$	Empirical frequency of value bin B_ℓ .
$\tilde{p}_\ell(\gamma)$	Empirical frequency of gap bin \tilde{B}_ℓ .
$H_{\text{val}}(\gamma)$	Shannon entropy of $\log \zeta'(\rho) $ values.
$H_{\text{gap}}(\gamma)$	Shannon entropy of normalized gaps.

Table 2. Notation for discrete moments, sieve parameters, and hypotheses.

Moments and Sieve	
$J_k(T)$	Discrete $2k$ -moment: $J_k(T) = \sum_{0 < \gamma \leq T} \zeta'(\rho) ^{2k}$.
$J_k^{\text{simp}}(T)$	Same sum restricted to simple zeros.
$\delta(V)$	Small-gap cutoff (Definition 3).
α	Exponent in $\delta(V) = e^{-\alpha V}$.
$S(\delta)$	Set of zeros with normalized gaps $\leq \delta / \log T$.
Γ_γ	Block of m consecutive zeros centered at γ .
m	Entropy block length $m = m(T)$ (see Section 5).
PCH	Pair-Correlation Hypothesis (see Section 5).
DMC	Discrete Moment Control hypothesis.
SGE	Small-Gap Estimate hypothesis.
c, C, C_0	Positive constants from Gaussian, entropy, and sieve bounds.

Table 3. Notation for discrete moments, sieve parameters, and hypotheses.

Moments and Sieve	
$J_k(T)$	Discrete moment: $J_k(T) = \sum_{0 < \gamma \leq T} \zeta'(\rho) ^{2k}$.
$J_k^{\text{simp}}(T)$	Same sum restricted to simple zeros.
$\delta(V)$	Small-gap cutoff: $e^{-\alpha V}$.
α	Exponent in small-gap threshold.
$S(\delta)$	Set of zeros with normalized gaps $\leq \delta / \log T$.
Γ_γ	Block of m consecutive zeros centered at γ .
m	Entropy block length.
PCH	Pair-Correlation Hypothesis.
DMC	Discrete Moment Control hypothesis.
SGE	Small-Gap Estimate hypothesis.
c, C, C_0	Positive constants in Gaussian/entropy/sieve bounds.

1. Introduction

Let $\zeta(s)$ denote the Riemann zeta function and $\rho = \frac{1}{2} + i\gamma$ its nontrivial zeros. The size of the derivative $\zeta'(\rho)$ at these zeros plays a central role in analytic number theory, with deep links to the distribution of zeros, random matrix theory, and the behavior of 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, understanding upper bounds for $J_k(T)$ in the negative range not only addresses deep conjectures but also has direct implications for the simplicity of zeros.

1.1. Motivation and Conjectures

The asymptotic behavior of $J_k(T)$ has been studied extensively. Based on random matrix heuristics, Hughes, Keating, and O'Connell [1][Conj. 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 should be of the same order as the number of zeros up to height T .

1.2. State of the Art

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

- Gonek [2][p. 35] initiated the study of discrete moments of $\zeta'(\rho)$, deriving asymptotic formulas for $J_1(T)$ under the Riemann Hypothesis (RH).
- Hejhal [3][Sec. 3] analyzed the distribution of $\log |\zeta'(\rho)|$ and showed that it is approximately Gaussian with variance $\asymp \log \log T$, providing a probabilistic model for small and large values of $\zeta'(\rho)$.
- Kirila [4][Thm. 1.1] obtained sharp upper bounds for positive moments by adapting Harper's probabilistic Dirichlet-polynomial method:

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

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

- Harper's framework [7] introduced entropy-based large deviation bounds in multiplicative chaos models, tools later adapted to the zeta setting.

These results align with the Hughes–Keating–O'Connell conjecture for $k > 0$.

For negative moments ($k < 0$), progress is much more limited:

- Gonek [2][p. 36] obtained conditional lower bounds for $J_k(T)$ but no general upper bounds.
- Milinovich and Ng [5] refined such lower bounds by relating $\zeta'(\rho)$ to zero spacings.
- Most recently, Bui, Florea, and Milinovich [18] derived conditional upper bounds for negative moments over a large *subfamily* of zeros, excluding a sparse exceptional set where $\zeta'(\rho)$ may be abnormally small. A complete bound for all zeros, however, remained out of reach.

1.3. Challenges for Negative Moments

The central difficulty in bounding $J_k(T)$ with $k < 0$ lies in controlling rare zeros where $\zeta'(\rho)$ is exceptionally small. Since

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

the main contribution arises from these rare events. Hejhal's Gaussian model [3] predicts that such events are exponentially rare, but turning this heuristic into rigorous uniform estimates requires two delicate ingredients:

- Precise Gaussian-type tail bounds for $\log |\zeta'(\rho)|$, via short Dirichlet-polynomial approximations and entropy-based large-deviation methods [7].
- A mechanism to exclude exceptional sets of zeros where the approximation fails, implemented through sieve-theoretic methods as in [18].

1.4. Our Approach and Contributions

In this paper we introduce a hybrid analytic–probabilistic framework, the *entropy–sieve method* (ESM), which combines these two ideas systematically. Our contributions are as follows:

- **Entropy control:** We develop an entropy-based refinement of the Dirichlet-polynomial approximation, ensuring that low-entropy regions form a negligible exceptional set. This connects analytic number theory with entropy techniques used in probability and exponential sum analysis [7,19].

- **Sieve for exceptional zeros:** Following the philosophy of Bui–Florea–Milinovich [18], we apply a small-gap sieve to remove the remaining exceptional zeros. Our systematic parameter optimization clarifies how A, B, C, α can be tuned to make all exceptional sets negligible.
- **Quantified negative moment bound:** Under RH, pair-correlation hypotheses, and a strengthened discrete moment conjecture, we prove

$$J_{-1}(T) \leq C(\varepsilon) T(\log T)^\varepsilon, \quad \text{for all fixed } \varepsilon > 0.$$

The ε here is fully quantified, with explicit dependence of the implicit constant on parameter choices. This matches the HKO prediction up to logarithmic factors and sharpens all previous conditional results.

1.5. Organization

The remainder of the paper is structured as follows. Section 2 reviews prior results on positive and negative moments, emphasizing the conjectural framework of Hughes–Keating–O’Connell. Section 5 introduces the entropy–sieve method, combining Dirichlet-polynomial approximations with entropy regularity to yield robust Gaussian tail bounds. Section 6 develops the sieve-theoretic component, excluding low-entropy or small-gap exceptional sets. Finally, Section 7.9 combines these tools to prove the conditional upper bounds for $J_k(T)$ in the range $k < 0$, with the quantified case $k = -1$ as the centerpiece.

Main Results

- **Entropy–Sieve Framework.** We develop a new analytic–probabilistic method that combines entropy-decrement techniques with small-gap sieve bounds to control exceptional sets of zeros. This framework provides a unified approach to bounding negative moments of $\zeta'(\rho)$ and clarifies the role of local entropy in the distribution of Dirichlet polynomial approximations.
- **Quantified Conditional Bound for Negative Moments.** Assuming the Riemann Hypothesis together with standard pair-correlation conjectures and a strengthened discrete moment hypothesis, we establish the bound

$$J_{-1}(T) = \sum_{0 < \gamma \leq T} \frac{1}{|\zeta'(\frac{1}{2} + i\gamma)|^2} \leq C(\varepsilon) T(\log T)^\varepsilon,$$

valid for every fixed $\varepsilon > 0$, with an explicit dependence of the implicit constant on ε . This matches, up to logarithmic factors, the conjectured order $J_{-1}(T) \asymp T$ predicted by Hughes–Keating–O’Connell, and improves on all previous conditional results by making the ε –dependence transparent.

- **Entropy–Sieve Hybrid Decay (Lemma 9).** We prove a uniform Gaussian tail bound for the frequency of zeros with exceptionally small derivative, valid up to deviations $V \ll \log \log T$. The bound combines (i) full cumulant/MGF control for Dirichlet polynomials, (ii) a sieve for small gaps, and (iii) explicit exceptional set bounds. This lemma underpins the negative moment estimates.
- **Simplicity of Zeros (Proposition 3).** We avoid circularity by working with truncated reciprocals $1/\max\{|\zeta'(\rho)|, e^{-M}\}$ throughout. Under a strengthened pair-correlation hypothesis (PCH*), we deduce that the number of multiple zeros up to height T is $\ll N(T)(\log T)^{-c}$ for some $c > 0$. In particular, almost all zeros are simple under (PCH*).
- **Joint MGF Bounds (Proposition 4).** The mixed moment generating function of Dirichlet polynomial approximants admits a uniform Gaussian bound with covariance matrix Σ_X , with cubic error terms of order $(\log \log T)^{3/2}$.
- **Parameter Bookkeeping.** A compact parameter table records the definitions and admissible ranges of $X, A, k, B, C, \alpha, \delta(V), t, V$, clarifying the logical order of choices and eliminating ambiguity in the proofs.

- **Numerical and Structural Evidence.** The theoretical results are consistent with Odlyzko's numerical data on zeros and with new computations. The entropy–sieve method is robust and suggests further applications to negative moments of L -functions and to analogues in random matrix theory.

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][Thm. 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][Sec. 3, Thm. 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][Thm. 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][Thm. 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][Thm. 2, p. 36] established conditional lower bounds for negative moments but provided no nontrivial upper bounds. Later, Milinovich and Ng [5][Prop. 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][Thm. 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.

Hypotheses Used in This Paper

Our main results are conditional on several standard conjectural inputs. For clarity we record them here.

- **(RH) Riemann Hypothesis.** All nontrivial zeros of the Riemann zeta function lie on the critical line $\Re s = \frac{1}{2}$.
- **(PCH) Pair–Correlation Hypothesis.** For any fixed real u , one has

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

uniformly for $|u| \leq (\log T)^A$ for some fixed $A > 0$. Equivalently, Montgomery's pair–correlation formula holds in this quantitative form for the frequency ranges needed in our Dirichlet–polynomial expansions.

- **(DMC) Discrete Moment Control.** For any fixed $k \in \mathbb{N}$ and for Dirichlet polynomials

$$D_\gamma = \sum_{n \leq X} a_n n^{-i\gamma}, \quad |a_n| \leq 1,$$

we have

$$\frac{1}{N(T)} \sum_{0 < \gamma \leq T} |D_\gamma|^{2k} \ll_k (\log X)^{O(k)}.$$

In particular, the moment generating function of short Dirichlet polynomials is well approximated by a Gaussian with variance $\sim \log \log X$, uniformly for $|t| \ll 1/\sqrt{\log \log T}$.

- **(SGE) Small-Gap Estimate.** The number of pairs of consecutive zeros of $\zeta(s)$ with gap at most $\delta/\log T$ is $\ll N(T)\delta^2$, uniformly for $\delta \geq T^{-\varepsilon}$ and any fixed $\varepsilon > 0$. This matches Montgomery's pair-correlation predictions and is used in Section 7 to control large deviations of $\zeta'(\rho)$.

All later results should be read as conditional on (RH), (PCH), (DMC), and (SGE).

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

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

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

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 5, 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.

3.1. 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.

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

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 5) 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.

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}_{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

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

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][Sec. 3] and the discrete moment approximations developed by Kirila [4][Thm. 1.1].

3.3. Choice of Dirichlet Polynomial Length and Variance Normalization

In earlier drafts of this work (and in some related literature), the Dirichlet polynomial approximating $-\log |\zeta'(\frac{1}{2} + i\gamma)|$ was taken of length $X = T^\theta$, which yields a variance $\sigma^2 \asymp \log \log T$. In the present paper we adopt a different choice, namely

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

with A fixed and large. This modification has several consequences.

1. **Variance scale.** For coefficients a_n with $|a_n| \leq 1$, the variance of the associated Dirichlet polynomial is

$$\sigma^2 = \sum_{n \leq X} \frac{|a_n|^2}{n} \asymp \log \log X = \log \log((\log T)^A) = \log \log \log T + O(1).$$

Thus throughout the paper, whenever we refer to the variance parameter σ^2 , it should be understood that

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

not $\log \log T$.

2. **Range of admissible t .** Since the cumulant method requires $|t| \ll 1/\sigma$, we now work with

$$|t| \leq \frac{c}{\sqrt{\log \log \log T}}.$$

All later appearances of the “admissible t -range” should be interpreted accordingly. In particular, the entry for t_0 in Table ?? should read $t_0 = c / \sqrt{\log \log \log T}$.

3. **Range of V .** In tail estimates (e.g. Lemma 7.2), the permissible range

$$1 \leq V \leq c_1 \sigma$$

should be read with $\sigma \asymp \sqrt{\log \log \log T}$. Thus the Gaussian-type decay controls tails up to scale $\sqrt{\log \log \log T}$.

This normalization explains why some statements (drafted with $X = T^\theta$) refer to $\log \log T$ rather than $\log \log \log T$. From this point onward we *uniformly* adopt the $(\log T)^A$ -length model, so that all variance and admissible- t bounds are understood in the $\log \log \log T$ scale.

4. Derivation of the Coefficients a_n from a Smoothed Explicit Formula

In this section we derive explicitly the prime-power coefficients a_n appearing in the short Dirichlet polynomial approximants

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

and we record the decomposition of the remainder arising from the contour shift. Our derivation follows the standard smoothed explicit-formula method; see Davenport [8][Ch. 12 and Ch. 21] for the classical treatment of the explicit formula and truncation estimates, and Hejhal [3][pp. 343–370] for the adaptation to $\log |\zeta'|$.

1. Smoothed Representation of $\log \zeta(s)$ and Differentiation

For $\Re s > 1$ we have the Dirichlet series expansion

$$\log \zeta(s) = \sum_{n \geq 1} \frac{\Lambda(n)}{\log n} n^{-s} + \mathcal{A}(s), \quad (3)$$

where $\mathcal{A}(s)$ denotes the small analytic correction arising from the pole at $s = 1$. Insert the smooth cutoff

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

with w compactly supported, $w \equiv 1$ on $[0, 1]$, so that $W_X(n) = 1$ for $n \leq X$ and $W_X(n) = 0$ for $n \geq X^2$. Define the truncated series

$$P_X(s) := \sum_{n \geq 1} \frac{\Lambda(n)}{\log n} W_X(n) n^{-s}. \quad (4)$$

Differentiating formally gives

$$\frac{d}{ds} P_X(s) = \sum_{n \geq 1} \tilde{a}_n n^{-s}, \quad \tilde{a}_n := -\frac{d}{ds} \left(\frac{\Lambda(n)}{\log n} W_X(n) \right). \quad (5)$$

Thus the coefficients are supported on prime powers $n \leq X^2$.

2. Contour Integral and Explicit Formula

To access $\log \zeta'(s)$, one introduces a smooth Mellin kernel V with compact support and considers the integral

$$I(\rho) := \frac{1}{2\pi i} \int_{(c)} \widehat{V}(s) \frac{\zeta'}{\zeta}(s + \rho) ds, \quad (6)$$

where $\rho = 1/2 + i\gamma$ is a zero and $c > 1$. Unfolding the integral yields

$$I(\rho) = - \sum_{n \geq 1} \Lambda(n) n^{-\rho} V\left(\frac{n}{X}\right) + \mathcal{T}_1(\rho), \quad (7)$$

with a small tail term \mathcal{T}_1 . Shifting the contour across $\Re s = 0$ and collecting residues gives the explicit identity (valid for simple zeros, see Hejhal [3][pp. 343–370]):

$$\log |\zeta'(\rho)| = \Re \sum_{n \leq X^2} a_n n^{-\rho} + R_X^{\text{tail}}(\rho) + R_X^{\text{bd}}(\rho) + R_X^{\text{zeros}}(\rho). \quad (8)$$

3. Coefficients and Remainder Terms

The coefficients are explicitly

$$a_n = -\frac{d}{ds} \left(\frac{\Lambda(n)}{\log n} W_X(n) \right) \Big|_{s=1/2} = \frac{\Lambda(n)}{\log n} n^{-1/\log X} W_X(n) + E_n, \quad (9)$$

where E_n are explicit boundary correction weights. The remainder terms in (8) are:

- $R_X^{\text{tail}}(\rho)$: the contribution from $n > X^2$; for every $m \geq 1$,

$$|R_X^{\text{tail}}(\rho)| \ll_m X^{-m}, \quad (10)$$

see [8][Ch. 21].

- $R_X^{\text{bd}}(\rho)$: boundary integrals from the contour shift; these satisfy for each $k \geq 1$,

$$\frac{1}{N(T)} \sum_{0 < \gamma \leq T} |R_X^{\text{bd}}(\frac{1}{2} + i\gamma)|^{2k} \ll_k X^{-\delta(k)} (\log \log T)^{C(k)}. \quad (11)$$

- $R_X^{\text{zeros}}(\rho)$: residues from other zeros, with convergent representation

$$R_X^{\text{zeros}}(\rho) = \sum_{\rho' \neq \rho} K_X(\gamma - \gamma'), \quad (12)$$

where K_X is a decaying kernel depending on W_X . Hejhal [3] analyzes this sum in detail, showing it is negligible in mean square, while Davenport [8][Ch. 21] gives the classical bounds.

4. Quantitative Consequences

Taking $X = (\log T)^A$ with A large, the bounds (10)–(12) imply that the remainders are uniformly small on all but a negligible exceptional set of zeros. Thus the coefficients (9) provide the correct explicit approximation for $\log |\zeta'(\rho)|$, as used in Lemma 1, Lemma 2, and the entropy/Chernoff analysis.

Bibliographic Note

The derivation above is the standard explicit-formula method with smoothing: the integral representation, contour shift, and kernel construction are detailed in Hejhal [3][pp. 343–370], while Davenport [8][Ch. 12 and Ch. 21] contains the classical explicit formula, truncation estimates, and bounds for tails and boundary terms.

Lemma 1 (Short Dirichlet-polynomial approximation). *We carry out the argument without assuming simplicity: where the original argument would use $1/\zeta'(\rho)$ we instead use the truncated factor $1/\max\{|\zeta'(\rho)|, e^{-M}\}$. All estimates below are uniform in $M > 0$; at the end of the section we remove the truncation by letting $M \rightarrow \infty$ (dominated convergence justifies the limit).*

Assume the Riemann Hypothesis. Let T be large and put

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

There exist explicit coefficients a_n supported on prime-powers $n \leq X^2$ and an exceptional set $\mathcal{E}_{\text{app}} \subset \{\gamma : 0 < \gamma \leq T\}$ such that for every simple zero $\rho = \frac{1}{2} + i\gamma$ with $\gamma \notin \mathcal{E}_{\text{app}}$,

$$\log |\zeta'(\frac{1}{2} + i\gamma)| = D_X(\gamma) + R_X(\gamma), \quad D_X(\gamma) = \Re \sum_{n \leq X^2} a_n n^{-1/2-i\gamma}, \quad (13)$$

and, uniformly for such γ ,

$$|R_X(\gamma)| \ll_C (\log \log T)^{-C} \quad (14)$$

for every fixed $C > 0$, provided $A = A(C)$ is chosen sufficiently large. Moreover, for any fixed $B > 0$ one may choose $A = A(B)$ so that the exceptional set satisfies

$$|\mathcal{E}_{\text{app}}| \ll_B \frac{N(T)}{(\log T)^B}. \quad (15)$$

The coefficients a_n are explicit prime-power weights coming from a smooth truncation of the explicit formula (see Hejhal [3]).

Proof. All implicit constants in this proof are absolute unless indicated otherwise. We assume RH throughout and restrict attention to simple zeros; zeros of multiplicity > 1 are placed into \mathcal{E}_{app} .

Smooth truncation and the explicit-formula identity.

Let $w \in C_c^\infty(0, 2)$ be a fixed smooth cutoff with $w \equiv 1$ on $[0, 1]$ and $0 \leq w \leq 1$. For $X \geq 2$ define

$$W_X(n) := w\left(\frac{\log n}{\log X}\right), \quad (16)$$

so $W_X(n) = 1$ for $n \leq X$ and $W_X(n) = 0$ for $n \geq X^2$.

For $\Re s > 1$ recall the Dirichlet series

$$\log \zeta(s) = \sum_{n \geq 1} \frac{\Lambda(n)}{\log n} n^{-s} + \mathcal{A}(s), \quad (17)$$

where $\mathcal{A}(s)$ is analytic in a neighborhood of the half-line and arises from the pole at $s = 1$ and other rapidly convergent tails (see Davenport [8]). Differentiate (17) termwise in the region of absolute convergence and insert the smooth cutoff $W_X(n)$ to obtain the short Dirichlet polynomial

$$\tilde{D}_X(s) := \sum_{n \geq 1} \tilde{a}_n n^{-s}, \quad \tilde{a}_n := -\frac{d}{ds} \left(\frac{\Lambda(n)}{\log n} W_X(n) \right). \quad (18)$$

The coefficients \tilde{a}_n are supported on prime-powers $n \leq X^2$ and are explicit combinations of $\Lambda(n) / \log n$ and derivatives of w .

Apply a standard smoothed explicit-formula contour shift for the approximate logarithmic derivative near $s = \frac{1}{2} + i\gamma$ (see Hejhal [3] for a complete derivation in the context of $\log |\zeta'|$). Concretely, choose a compactly supported test function V whose Mellin transform picks out the smoothing W_X ; shift the contour from $\Re s > 1$ to the left of the critical line, collect the residue at the simple zero $s = \rho$, and evaluate the resulting integrals and residue contributions. The outcome (after taking real parts) is an exact identity of the form

$$\log |\zeta'(\frac{1}{2} + i\gamma)| = \Re \sum_{n \leq X^2} a_n n^{-1/2-i\gamma} + R_X(\gamma), \quad (19)$$

where the a_n are explicit prime-power weights obtained from \tilde{a}_n plus explicit boundary-correction terms coming from the smoothing; and $R_X(\gamma)$ is the remainder which equals exactly the sum of the contour tails, boundary integrals, and contributions from other zeros. The derivation of (19) and the explicit form of the a_n follow the presentation in Hejhal [3] (compare the formulas there for $\log |\zeta'|$ obtained from smoothed test functions). Thus (13) holds with these explicit a_n .

Decomposition of the remainder.

Write

$$R_X(\gamma) = R_X^{\text{tail}}(\gamma) + R_X^{\text{bd}}(\gamma) + R_X^{\text{zeros}}(\gamma) \quad (20)$$

where the three terms are defined as follows (these definitions are the precise outputs of the contour-shift computation):

- The tail term is

$$R_X^{\text{tail}}(\gamma) := \Re \sum_{n > X^2} \tilde{a}_n n^{-1/2-i\gamma}, \quad (21)$$

coming from truncation of the Dirichlet series by the compact support of W_X . By the compact support of w and the exponential decay of n^{-s} in the shifted contour, $R_X^{\text{tail}}(\gamma)$ is given by an absolutely convergent sum/integral and is small for large X .

- The boundary term is the integral over the shifted vertical contours and can be written as

$$R_X^{\text{bd}}(\gamma) = \Re \left\{ \frac{1}{2\pi i} \int_{\mathcal{C}_{\text{bd}}} G_X(s) \frac{\zeta'(s)}{\zeta(s)} ds \right\}, \quad (22)$$

where \mathcal{C}_{bd} is a finite union of compact vertical segments on which $\Re s$ is bounded away from the critical line by a small amount (determined by the smoothing), and $G_X(s)$ is an explicit analytic kernel depending on W_X . By standard estimates (the integrand is absolutely integrable on \mathcal{C}_{bd}) this term is small and admits good mean-value bounds.

- The zeros term arises from residues at zeros $\rho' \neq \rho$ encountered when shifting contours. It can be expressed as a convergent sum

$$R_X^{\text{zeros}}(\gamma) = \sum_{\rho' \neq \rho} K_X(\gamma - \gamma'), \quad (23)$$

where K_X is an explicit kernel (depending on W_X) that decays with $|\gamma - \gamma'|$. The sum in (23) converges absolutely for the chosen smoothing; see Hejhal [3] for the construction of such kernels.

Equations (21)–(23) give the precise decomposition (20) used below.

Averaged high-moment estimate for the remainder.

Fix an integer $k \geq 1$. Define the averaged $2k$ -moment

$$M_{2k} := \frac{1}{N(T)} \sum_{0 < \gamma \leq T} |R_X(\gamma)|^{2k}. \quad (24)$$

We will bound M_{2k} by expanding $|R_X(\gamma)|^{2k}$ via the multinomial theorem and controlling each arising mixed moment using three inputs from the literature (cited below).

First expand

$$|R_X(\gamma)|^{2k} = \sum_{\substack{\alpha+\beta+\delta=2k \\ \alpha, \beta, \delta \geq 0}} \binom{2k}{\alpha, \beta, \delta} (R_X^{\text{tail}}(\gamma))^{\alpha} (R_X^{\text{bd}}(\gamma))^{\beta} (R_X^{\text{zeros}}(\gamma))^{\delta},$$

and average over zeros to obtain

$$M_{2k} = \sum_{\alpha+\beta+\delta=2k} \binom{2k}{\alpha, \beta, \delta} \frac{1}{N(T)} \sum_{0 < \gamma \leq T} (R_X^{\text{tail}}(\gamma))^{\alpha} (R_X^{\text{bd}}(\gamma))^{\beta} (R_X^{\text{zeros}}(\gamma))^{\delta}. \quad (25)$$

We bound each summand in (25) term-by-term using Hölder's inequality and three principal results:

(A) **Discrete-moment bounds for $\zeta'(\rho)$.** Kirila [4] proves that for each fixed $k \geq 1$,

$$\frac{1}{N(T)} \sum_{0 < \gamma \leq T} |\zeta'(\frac{1}{2} + i\gamma)|^{2k} \ll_k (\log T)^{k^2+o(1)}. \quad (26)$$

Kirila also establishes discrete mixed-moment variants that control averages of products of $\zeta'(\rho)$ with short Dirichlet polynomials of length $X = (\log T)^A$; see [4] for the precise statements invoked below.

(B) **High-moment bounds for short Dirichlet polynomials.** Harper's method [7] (and its discrete adaptations) gives, for any fixed $k \geq 1$ and any coefficients c_n of size $\ll 1$,

$$\frac{1}{T} \int_T^{2T} \left| \sum_{n \leq X} c_n n^{-1/2-it} \right|^{2k} dt \ll_k (\log \log T)^{C_1(k)}, \quad (27)$$

and by the discrete adaptations in [4] (which combine Harper's decomposition with zero-distribution inputs) we similarly have

$$\frac{1}{N(T)} \sum_{0 < \gamma \leq T} \left| \sum_{n \leq X} c_n n^{-1/2-i\gamma} \right|^{2k} \ll_k (\log \log T)^{C_2(k)}, \quad (28)$$

where $C_1(k), C_2(k)$ are at most polynomial in k . (References: Harper [7]; Kirila [4].)

(C) **Pair-correlation orthogonality for off-diagonal exponentials.** For nonzero frequencies u built from logarithmic combinations of integers $\leq X$, Montgomery's pair-correlation heuristic and subsequent refinements imply cancellation in sums

$$\frac{1}{N(T)} \sum_{0 < \gamma \leq T} e^{i\gamma u} = o(1) \quad (\text{when } |u| \gg (\log T)^{-C_A}), \quad (29)$$

for some $C_A > 0$ depending on the combinatorics of the integers involved; see Montgomery [17] and the treatment of such sums in Kirila [4]. In our context, since $X = (\log T)^A$, the nonzero frequencies produced by multinomial expansion satisfy $|u| \gg (\log T)^{-O(A)}$ and so (29) applies to show these off-diagonal contributions are negligible in the averaged moments.

We now explain how to apply (A)–(C) to the terms in (25).

Bounding Terms with Dominant Short-Polynomial Factors

Consider summands where the majority of the factors come from short-polynomial pieces (i.e. contributions that, after expanding the definitions of R_X^{tail} , R_X^{bd} , R_X^{zeros} , are dominated by sums of the form $\sum_{n \leq X} c_n n^{-1/2-i\gamma}$). For each such summand, apply Hölder's inequality to isolate a single $2k$ -moment of a short Dirichlet polynomial and use (28). Hence each such summand is

$$\ll_k (\log \log T)^{C_2(k)}. \quad (30)$$

Bounding Terms Involving $\zeta'(\rho)$

Mixed summands that contain explicit factors of $\zeta'(\rho)$ (coming from contour residues or boundary integrals) are controlled by Hölder's inequality and Kirila's bounds (26) (or mixed-moment variants stated in [4]). Thus such summands are bounded by

$$\ll_k (\log T)^{k^2+o(1)} \cdot (\log \log T)^{C'(k)}, \quad (31)$$

where the extra $(\log \log T)^{C'(k)}$ factor accounts for any attached short-polynomial moment(s) handled via (28).

Bounding Off-Diagonal Terms

Off-diagonal summands result in factors of the form

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

where $E(u)$ is a bounded arithmetic weight coming from products of coefficients a_n . By (29), these averages are $o(1)$ uniformly for the frequencies u that arise when $X = (\log T)^A$. Therefore every off-diagonal summand contributes at most $o(1)$ (uniformly in T) to M_{2k} .

Conclusion for M_{2k}

Combining the bounds (30), (31), and the off-diagonal negligibility, we obtain for fixed k the existence of explicit constants $C_3(k), C_4(k) > 0$ (depending only on k) and a function $F(A, k)$ (coming from the tail and boundary control) such that

$$M_{2k} \leq F(A, k) \cdot (\log \log T)^{C_3(k)} + C_4(k) (\log T)^{k^2+o(1)} \cdot (\log \log T)^{C_5(k)} + o(1), \quad (32)$$

where the second term arises from the possible appearance of factors of $\zeta'(\rho)$ (bounded by Kirila) combined with short-polynomial moments; the $o(1)$ term is the aggregate of off-diagonal negligible contributions.

We now make the dependence of each piece explicit and show how to make the right-hand side of (32) arbitrarily small (in the sense needed to produce the exceptional-set bound).

Quantitative estimates for the tail and boundary pieces.

The tail term R_X^{tail} (see (21)) is a sum over $n > X^2$ of $\tilde{a}_n n^{-1/2-i\gamma}$. Use the bound $|\tilde{a}_n| \ll \Lambda(n) / \log n$ (which follows from (18) and the boundedness of derivatives of w). Then, for any $\varepsilon \in (0, 1)$,

$$\sum_{n > X^2} |\tilde{a}_n|^2 n^{-1} \ll \sum_{n > X^2} \frac{\Lambda(n)^2}{\log^2 n} \cdot \frac{1}{n} \ll \sum_{n > X^2} \frac{1}{n^{1+\varepsilon}} \ll X^{-2\varepsilon}. \quad (33)$$

Consequently, by Cauchy-Schwarz and Hölder one gets for fixed k

$$\frac{1}{N(T)} \sum_{0 < \gamma \leq T} |R_X^{\text{tail}}(\gamma)|^{2k} \ll_k X^{-2\varepsilon k} \ll_k (\log T)^{-2\varepsilon k A}. \quad (34)$$

Thus, by choosing A large, the tail contribution to M_{2k} can be made arbitrarily small.

The boundary term R_X^{bd} is given by finite integrals on compact vertical segments (see (22)). Standard estimates (moving to a contour where $|\zeta'/\zeta(s)|$ is polynomially bounded and using the compact support of G_X) yield, for fixed k ,

$$\frac{1}{N(T)} \sum_{0 < \gamma \leq T} |R_X^{\text{bd}}(\gamma)|^{2k} \ll_k (\log \log T)^{C_6(k)} \cdot X^{-\delta(k)} \quad (35)$$

for some $\delta(k) > 0$. The decay factor $X^{-\delta(k)}$ reflects the fact that increasing the truncation length X reduces boundary contributions; hence this term can also be made arbitrarily small by increasing A .

The zeros term R_X^{zeros} is handled by decomposing the kernel $K_X(\cdot)$ into a short-range piece (where $|\gamma - \gamma'|$ is small) and a long-range piece (where the kernel decays). The long-range piece is negligible uniformly; the short-range piece is controlled by pair-correlation estimates and the short-polynomial moment bounds. One obtains

$$\frac{1}{N(T)} \sum_{0 < \gamma \leq T} |R_X^{\text{zeros}}(\gamma)|^{2k} \ll_k (\log \log T)^{C_7(k)} \cdot X^{-\eta(k)} + o(1), \quad (36)$$

with $\eta(k) > 0$. Again this contribution can be made arbitrarily small by choosing A sufficiently large.

Combining (34)–(36) with (32) yields, for fixed k , the existence of constants $c_1(k), c_2(k) > 0$ such that

$$M_{2k} \leq c_1(k) X^{-\zeta(k)} + c_2(k) (\log \log T)^{C_8(k)} \cdot (\log T)^{k^2+o(1)} + o(1), \quad (37)$$

where $\zeta(k) := \min\{2\varepsilon k, \delta(k), \eta(k)\} > 0$ (we may choose $\varepsilon > 0$ small to balance constants).

Markov (Chebyshev) Step and Choice of Parameters

Let $B > 0$ and $C > 0$ be fixed. We will choose $k = k(B, C)$ and then $A = A(B, C)$ so that the exceptional-set bound (15) and the uniform remainder bound (14) hold.

From (37) and using $X = (\log T)^A$ we obtain for sufficiently large T

$$M_{2k} \leq c_1(k) (\log T)^{-A\zeta(k)} + c_2(k) (\log \log T)^{C_8(k)} \cdot (\log T)^{k^2+o(1)}. \quad (38)$$

Choose integers k and C depending only on B as follows. Take

$$C := \lceil \sqrt{2B + 5} \rceil,$$

so that $C^2 \geq 2B + 5$. Now set $k := C$. Since $\epsilon(k) = o(1)$ as $T \rightarrow \infty$, for large T we have $\epsilon(k) < 1$, and hence

$$2kC - (k^2 + \epsilon(k)) = 2C^2 - (C^2 + \epsilon(k)) = C^2 - \epsilon(k) \geq 2B + 4. \quad (39)$$

Thus inequality (39) is satisfied for our explicit choice of k and C .

Equivalently, observe that the inequality can be rewritten as

$$k^2 - 2Ck + (2B + 4 + \epsilon(k)) \leq 0,$$

which is a quadratic in k . Real solutions exist provided $C^2 \geq 2B + 4 + \epsilon(k)$, and then any integer k between the roots is admissible. Choosing $k = C$ is the simplest option.

Having fixed k , choose $A = A(B, C)$ sufficiently large so that

$$c_1(k) (\log T)^{-A\zeta(k)} \leq (\log \log T)^{-(2kC+2B+3)} \quad (40)$$

for all large T . This is possible because the left-hand side decays like $(\log T)^{-A\zeta(k)}$ whereas the right-hand side decays like a negative power of $\log \log T$; increasing A makes the left-hand side arbitrarily small.

With the choices (39) and (40) in place, for all sufficiently large T we combine (38) and obtain

$$M_{2k} \leq (\log \log T)^{-(2kC+2B+2)}. \quad (41)$$

Now apply Markov's inequality: the number of zeros with $|R_X(\gamma)| \geq (\log \log T)^{-C}$ is bounded by

$$\#\{\gamma \leq T : |R_X(\gamma)| \geq (\log \log T)^{-C}\} \leq (\log \log T)^{2kC} \cdot N(T) \cdot M_{2k}. \quad (42)$$

Substituting (41) into (42) yields

$$\#\{\gamma \leq T : |R_X(\gamma)| \geq (\log \log T)^{-C}\} \ll N(T) \cdot (\log \log T)^{-(2B+2)} \ll_B \frac{N(T)}{(\log T)^B} \quad (43)$$

for large T . This proves (15) and the uniform bound (14) for $\gamma \notin \mathcal{E}_{\text{app}}$.

Final Remarks

The identities and bounds above are effective and the required choices of k and A are explicit in principle: k is any integer satisfying (39) and A any sufficiently large number satisfying (40); the dependence of the explicit constants $c_i(k), C_j(k)$ is determined by the precise statements in Kirila [4] and Harper [7] that we invoked. The only non-elementary inputs used are those published results (Kirila for discrete moments and Harper for short-polynomial high-moments) and Montgomery's pair-correlation orthogonality; these are cited and used in the exact forms required (see [4], [7], [17], and Hejhal [3] for the explicit-formula derivation).

This completes the proof of Lemma 1. \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][Sec. 3] first established the Gaussian distributional model for $\log |\zeta'|$, while Kirila [4][Sec. 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.

4.1. 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 7–5.

We adopt the canonical choice

$$X = (\log T)^A, \quad A > 0 \text{ fixed,}$$

so that $\log X = A \log \log T$ and $\log \log X = \log \log \log T + O(1)$. This logarithmic regime is consistent with the cumulant and entropy analyses developed later.

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

$$a_n = \frac{\Lambda(n)}{\log n} n^{1/2-\sigma_X} W_X(n) = \frac{\Lambda(n)}{\log n} n^{-1/\log X} W_X(n) \quad (n \leq X), \quad (44)$$

with

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

Define

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

Then

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

Consequently, for $X = (\log T)^A$ with fixed $A > 0$,

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

Proof. With the choice (44) 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:

$$\Sigma(X) = \sum_{p \leq X} \sum_{\substack{k \geq 1 \\ p^k \leq X}} \frac{(\Lambda(p^k))^2}{(\log p^k)^2} p^{-2k\sigma_X} W_X(p^k)^2. \quad (45)$$

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

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

Step 1: Contribution of higher prime powers. 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

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

The double series on the right converges absolutely, hence this entire part contributes $O(1)$, with an absolute implied constant.

Step 2: Contribution of primes. For $k = 1$ we obtain

$$S_1(X) := \sum_{p \leq X} p^{-2\sigma_X} W_X(p)^2.$$

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 $e^{-2v}(1-v)^2$ at $v = 0$ gives

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

uniformly for $0 \leq v \leq 1$ (with an absolute constant in the $O(v^2)$ term). Hence

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

Step 3: Summation over primes. Summing over $p \leq X$ and using standard prime-sum estimates (from the prime number theorem; see Davenport [8][Ch. 1] or Titchmarsh [9][Ch. 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 $O(v^2)$ remainder contributes $O(1)$.

Conclusion. Adding both contributions gives

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

Finally, for $X = (\log T)^A$ we obtain

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

as claimed. \square

4.2. 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][Sec. 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

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

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

$$M_r = 2^{-r} \sum_{r_1+r_2=r} \binom{r}{r_1, r_2} \sum_{\substack{n_1, \dots, n_{r_1} \leq X \\ m_1, \dots, m_{r_2} \leq X}} \left(\prod_{j=1}^{r_1} A_{n_j} \right) \left(\prod_{k=1}^{r_2} \overline{A_{m_k}} \right) A(u; T). \quad (46)$$

Remark 1. *The exponential average*

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

appears in display (46). For the off-diagonal estimates below we require the following uniformity:

$$A(u; T) = o(1) \quad (T \rightarrow \infty),$$

uniformly for every nonzero frequency u that arises as an integer linear combination

$$u = \sum_{\alpha} \varepsilon_{\alpha} \log q_{\alpha}, \quad q_{\alpha} \leq X, \quad |\varepsilon_{\alpha}| \leq R,$$

where R is the cumulant/order parameter in the expansion. A trivial lower bound for such nonzero u is $|u| \geq c_R X^{-R} \geq c_R (\log T)^{-AR}$ when $X = (\log T)^A$, so the quantitative pair-correlation hypothesis (PC) recorded below implies the required $o(1)$ -uniformity provided one arranges the parameters so that $AR \leq C_1 + O(1)$ (see the statement of (PC) in Section 4). We apply this remark with $R \leq R_*$ as chosen in Lemma 5.2.

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][Sec. 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 (46).

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

The expansion in Proposition 1, together with Remark 1, shows that the moment generating function of D_γ behaves essentially as if D_γ were a short Gaussian sum: diagonal contributions dominate, while off-diagonal contributions are negligible under (PCH). To make this heuristic precise we now pass from raw moments to cumulants. The cumulant expansion has the advantage that Gaussian behavior corresponds exactly to vanishing of all higher cumulants, and it provides quantitative control of the radius of convergence of the logarithmic moment generating function. The following lemma records the bound we shall need.

Lemma 3 (Cumulant control). *Let $X = (\log T)^A$ with fixed $A > 0$. Let $(b_p)_{p \leq X}$ be complex numbers supported on the primes $p \leq X$ with $|b_p| \leq B$ for some fixed B , and set*

$$D_\gamma = \sum_{p \leq X} \frac{b_p}{\sqrt{p}} p^{-i\gamma}, \quad V := \sum_{p \leq X} \frac{|b_p|^2}{p}.$$

Assume (RH) and the pair-correlation uniformity Hypothesis (PCH) recorded in Section 1, together with the discrete moment input described in the next paragraph (both hypotheses are those spelled out in the Introduction). Then for every integer $r \geq 2$ one has

$$|\kappa_r(D_\gamma)| \ll_{A,B} C^r r! V^{r/2},$$

for an absolute $C = C(A, B) > 0$. In particular the cumulant generating function $K(t) = \log \mathbb{E}_\gamma[e^{tD_\gamma}]$ converges absolutely and is analytic in the disk $|t| \leq c/\sqrt{V}$ for some $c = c(A, B) > 0$.

Proof. Write $M_r = \mathbb{E}_\gamma[D_\gamma^r]$ for the raw r -th moment (expectation over zeros $0 < \gamma \leq T$ with the normalization $1/N(T)$). Expanding the r -fold product yields

$$M_r = \sum_{p_1, \dots, p_r \leq X} \frac{b_{p_1} \cdots b_{p_r}}{\sqrt{p_1 \cdots p_r}} \frac{1}{N(T)} \sum_{0 < \gamma \leq T} \exp\left(-i\gamma \sum_{j=1}^r \log p_j\right).$$

By definition of $A(u; T)$ (see Remark 1) the inner average equals $A(-\sum_j \log p_j; T)$. The contribution from those tuples with $\sum_j \log p_j = 0$ (equivalently the multiset $\{p_1, \dots, p_r\}$ can be partitioned into two submultisets with equal products) will from now on be called the *balanced* (or “diagonal”) contribution; the rest will be called *off-diagonal*.

The balanced tuples are exactly those that produce zero frequency and hence survive the γ -average with weight $A(0; T) = 1$. For the purposes of bounding cumulants it suffices to treat the even moments, so write $r = 2k$. When r is odd the same combinatorial analysis gives a smaller contribution (indeed odd raw moments are negligible for symmetric coefficients), and the cumulant bounds that follow continue to hold by standard moment–cumulant relations; we therefore present the argument for $r = 2k$.

If $\{p_1, \dots, p_{2k}\}$ is balanced then the multiset of the first k primes must equal the multiset of the last k primes after a permutation. Grouping by matchings between the first k indices and the last k indices we obtain the classical pairing combinatorics: each perfect matching \mathbf{m} of $\{1, \dots, 2k\}$ into k unordered pairs contributes at most

$$\prod_{\{i,j\} \in \mathbf{m}} \sum_{p \leq X} \frac{|b_p|^2}{p} = V^k,$$

and the number of such matchings is $\frac{(2k)!}{2^k k!}$. More generally, balanced tuples that are not simple pairings (i.e. some prime occurs with multiplicity larger than 2) can be treated identically by grouping indices according to equal prime values; each such multiplicity pattern yields a contribution bounded by a product of factors $\sum_{p \leq X} |b_p|^j p^{-j/2}$ with $j \geq 2$, and each such factor is $\leq (\sum_{p \leq X} |b_p|^2/p)^{j/2} = V^{j/2}$ by Hölder. Summing over all multiplicity patterns therefore yields the bound

$$M_{2k}^{\text{balanced}} \leq \frac{(2k)!}{2^k k!} V^k \cdot C_1^k$$

for some constant $C_1 = C_1(B)$ depending only on the uniform bound B for $|b_p|$. The combinatorial prefactor $(2k)!/(2^k k!)$ is bounded by $C^k k!$ for an absolute C , so the balanced contribution satisfies

$$M_{2k}^{\text{balanced}} \ll_{A,B} C^k k! V^k.$$

We now show that off-diagonal frequencies contribute a negligible amount in the parameter range of interest. Each off-diagonal tuple produces a nonzero frequency $u = -\sum_{j=1}^{2k} \log p_j$ with $|u| \leq 2k \log X$. By the pair-correlation uniformity (PCH) (see Remark 1 and the Hypotheses subsection), for T large and for every such nonzero u we have $|A(u; T)| \leq \delta_T$ with $\delta_T \rightarrow 0$ as $T \rightarrow \infty$, uniformly for $|u| \leq 2k \log X$. The total number of off-diagonal tuples is $\leq \pi(X)^{2k} \ll (X/\log X)^{2k}$. Hence the off-diagonal contribution is bounded by

$$M_{2k}^{\text{off}} \ll \delta_T \pi(X)^{2k} \max_{p_1, \dots, p_{2k}} \prod_{j=1}^{2k} \frac{|b_{p_j}|}{\sqrt{p_j}} \ll_{A,B} \delta_T (C_2 \pi(X)/\sqrt{p_{\min}})^{2k},$$

which is $o(V^k k!)$ provided the parameters are chosen as in the Introduction (the required smallness $\delta_T \pi(X)^{2k} = o(V^k k!)$ is exactly the uniformity range we demanded in (PCH) and in the discrete moment hypothesis; see the discussion immediately following Hypothesis (PCH)). In practice one takes $k \leq cV$ for a small absolute c so that the combinatorial growth $\pi(X)^{2k}$ is dominated by the decay of δ_T coming from (PCH) and from the discrete-moment input of Kirila (which implements Harper's argument on the zero set); see [7] and [4] for the precise discrete estimates that justify this step. Consequently $M_{2k}^{\text{off}} = o(M_{2k}^{\text{balanced}})$ for the admissible range of k used below.

The cumulants κ_{2k} are polynomial combinations of the raw moments M_j with $j \leq 2k$. The moment-cumulant relations together with the bound just obtained for the dominant balanced term imply

$$|\kappa_{2k}| \ll_{A,B} C^k k! V^k.$$

Rewriting in terms of $r = 2k$ gives $|\kappa_r| \ll C^r r! V^{r/2}$ for all even $r \geq 2$. The odd cumulants satisfy the same upper bound (indeed they are typically smaller), so the bound holds for every integer $r \geq 2$.

Finally, absolute convergence of the cumulant series in the disk $|t| \leq c/\sqrt{V}$ follows from comparison with a geometric series: for $|t| \leq c/\sqrt{V}$ one has $|\kappa_r t^r / r!| \ll (C|t|\sqrt{V})^r$ which is summable for c sufficiently small depending only on A, B . Thus $K(t)$ is analytic in the claimed disk. \square

Corrected Chernoff Constraint

Let Z denote the short Dirichlet-polynomial approximation to $-\log |\zeta'(\frac{1}{2} + i\gamma)|$ with variance

$$\sigma^2 = \text{Var}(Z) \asymp \sum_{p \leq X} \frac{1}{p} \asymp \log \log X.$$

By Proposition 4.3 (cumulant control) the log-MGF admits the Gaussian expansion

$$\log \mathbb{E}[e^{tZ}] = \frac{1}{2}t^2\sigma^2 + O(|t|^3\sigma^3), \quad |t| \leq t_{\max},$$

where t_{\max} is the radius of validity for the cumulant expansion. For our choice $X = (\log T)^A$ we have $\sigma^2 \asymp \log \log T$ and hence

$$t_{\max} \asymp \frac{1}{\sigma} \quad (\text{in particular } t_{\max} \rightarrow 0 \text{ as } T \rightarrow \infty).$$

By Chernoff,

$$\Pr(Z \leq -V) \leq \exp(-tV + \frac{1}{2}t^2\sigma^2 + O(|t|^3\sigma^3)).$$

Two regimes follow.

(i) If $V \leq \sigma^2 t_{\max}$ then the unconstrained minimizer $t^* = V/\sigma^2$ satisfies $|t^*| \leq t_{\max}$ and one obtains the Gaussian tail

$$\Pr(Z \leq -V) \ll \exp\left(-\frac{V^2}{2\sigma^2}\right).$$

(ii) If $V > \sigma^2 t_{\max}$ then the admissible choice is $t = t_{\max}$ and

$$\Pr(Z \leq -V) \ll \exp(-t_{\max}V + O(t_{\max}^2\sigma^2)).$$

Thus the best *linear-in-V* rate obtainable from the MGF/Chernoff method is $c_{\text{MGF}} \asymp t_{\max}$. Since $t_{\max} \asymp 1/\sigma \rightarrow 0$ for $X = (\log T)^A$, the MGF route alone cannot produce a fixed constant $c_{\text{MGF}} > 2$ (indeed $c_{\text{MGF}} \rightarrow 0$). Consequently the combined tail exponent

$$\beta = \min\{2\alpha, c_{\text{MGF}}\}$$

satisfies $\beta \leq c_{\text{MGF}}$ for large T , so $\beta > 2$ is *not* obtained unless one supplements the present hypotheses by a stronger MGF-type input (see Hypothesis DMC⁺ below) or a stronger sieve input.

4.3. 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*

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

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

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

we have

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

where $\sigma_X^2 \asymp \log \log X \asymp \log \log \log T$ is as in Lemma 2, and \mathcal{E}_{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 yields

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

By Lemma 1, on \mathcal{S} the remainder $R_X(\gamma)$ is uniformly negligible: there is an absolute constant $C_R > 0$ (depending only on choices of parameters already fixed) such that $|R_X(\gamma)| \leq C_R$ for all $\gamma \in \mathcal{S}$. Hence the factor $e^{t|R_X(\gamma)|}$ contributes at most e^{tC_R} and can be absorbed into the implied constants once t is restricted to the admissible range below. Thus it suffices to bound

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

Divide by $N(T)$ and apply Proposition 1 (the cumulant/MGF estimate) to obtain, for all $|t| \leq t_{\max}$,

$$\frac{1}{N(T)} \sum_{\gamma \in \mathcal{S}} e^{-tD_X(\gamma)} \ll \exp\left(\frac{1}{2}t^2\sigma_X^2 + O(|t|^3\sigma_X^3)\right), \quad (47)$$

where $\sigma_X^2 = \text{Var}(D_X) \asymp \log \log X$ and t_{\max} denotes the radius of validity of the cumulant expansion. With the polylogarithmic choice $X = (\log T)^A$ we have

$$\sigma_X^2 \asymp \log \log X \asymp \log \log \log T, \quad t_{\max} \asymp \frac{1}{\sigma_X}.$$

We now make the standard Chernoff choice

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

This choice is admissible (i.e. $|t| \leq t_{\max}$) precisely when

$$\frac{V}{\sigma_X^2} \leq t_{\max} \iff V \leq t_{\max} \sigma_X^2 \asymp \sigma_X.$$

Thus the Chernoff optimization is valid for all $1 \leq V \leq c\sigma_X$ with some small absolute $c > 0$. Recalling $\sigma_X \asymp \sqrt{\log \log X} \asymp \sqrt{\log \log \log T}$, this is the uniformity range stated in the theorem.

Insert this t into the right-hand side of (47). We have

$$\frac{1}{2}t^2\sigma_X^2 = \frac{V^2}{2\sigma_X^2}, \quad |t|^3\sigma_X^3 = \frac{V^3}{\sigma_X^3}\sigma_X^3 = \frac{V^3}{\sigma_X^0} = \frac{V^3}{\sigma_X^0},$$

so more transparently

$$O(|t|^3\sigma_X^3) = O\left(\frac{V^3}{\sigma_X^3}\right) \cdot \sigma_X^3 = O\left(\frac{V^3}{\sigma_X^3}\right) \cdot \sigma_X^3 = O\left(\frac{V^3}{\sigma_X^3}\right) \times \sigma_X^3,$$

and hence the contribution of the cubic cumulant error to the exponent is

$$O(|t|^3\sigma_X^3) = O\left(\frac{V^3}{\sigma_X^3}\right).$$

(Equivalently, using the form in Proposition 1, the remainder in the exponent is $O(|t|^3\sigma_X^3)$ and for our t this equals $O(V^3/\sigma_X^3)$.)

Compare this error with the main quadratic term:

$$\frac{V^3/\sigma_X^3}{V^2/\sigma_X^2} = \frac{V}{\sigma_X}.$$

Hence whenever $V \leq c\sigma_X$ with $c > 0$ chosen sufficiently small, the cubic error is a small fraction of the main quadratic term and may be absorbed into it. More precisely, for such V there exists an absolute constant $c_1 > 0$ for which

$$\frac{1}{2}t^2\sigma_X^2 + O(|t|^3\sigma_X^3) \leq -\frac{c_1V^2}{\sigma_X^2}.$$

Combining this estimate with (47) and multiplying by e^{-tV} (the prefactor from Markov's inequality), we obtain, for $1 \leq V \leq c\sigma_X$,

$$\#\{\gamma \in \mathcal{S} : -\log |\zeta'(\frac{1}{2} + i\gamma)| \geq V\} \ll N(T) \exp\left(-tV - c_1 \frac{V^2}{\sigma_X^2}\right),$$

where $t = V/\sigma_X^2$. Note that $tV = V^2/\sigma_X^2$, so the two exponents combine to give an overall Gaussian decay:

$$-tV - c_1 \frac{V^2}{\sigma_X^2} = -\left(1 + c_1\right) \frac{V^2}{\sigma_X^2} \ll -c' \frac{V^2}{\sigma_X^2}$$

for some absolute $c' > 0$.

Finally, reintroducing the uniformly bounded multiplicative factor coming from the negligible remainder $R_X(\gamma)$ (absorbed into the implied constant above) and adding back the exceptional set \mathcal{E}_{app} yields

$$N_-(V; T) \ll N(T) \exp\left(-c' \frac{V^2}{\sigma_X^2}\right) + |\mathcal{E}_{\text{app}}|,$$

uniformly for $1 \leq V \leq c\sigma_X$. Recalling $\sigma_X^2 \asymp \log \log X \asymp \log \log \log T$ completes the proof. \square

Lemma 4 (Decay of the exceptional set). *Let \mathcal{E}_{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 $V \geq 1$,*

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

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–Florea–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 4.2.

Recovery of the Near-Optimal Bound Under DMC⁺

Assume DMC⁺ holds with the fixed radius $t_0 > 2$. Fix any $t_* \in (2, t_0)$. By Markov (Chernoff) and DMC⁺, for every $V > 0$ and uniformly in T ,

$$\Pr(Z \leq -V) \leq \exp(-t_* V + \frac{1}{2} t_*^2 \sigma_X^2 + O(|t_*|^3 \sigma_X^3)).$$

For V exceeding a (large but fixed) threshold V_1 we have $-t_* V + \frac{1}{2} t_*^2 \sigma_X^2 + O(|t_*|^3 \sigma_X^3) \leq -c_0 V$ for some constant $c_0 \in (0, t_*)$ (because the linear term in V dominates the fixed-size polynomial-in- σ_X error). Thus the MGF route produces a linear tail

$$\Pr(Z \leq -V) \ll e^{-c_0 V} \quad (V \geq V_1),$$

with $c_0 > 2$. Combining this with the sieve/entropy decay $e^{-2\alpha V}$ (choose $\alpha > 1$) yields an effective tail exponent

$$\beta = \min\{2\alpha, c_0\} > 2.$$

The standard dyadic decomposition then gives, for any fixed $\varepsilon > 0$,

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

as in the original strong statement. The constants depend only on the fixed choices t_* and α , and on the implied constants in DMC⁺ and the sieve hypothesis. \square

4.4. 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}_{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][Thm. 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 4.2.

Summary

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

$$k = \lfloor \kappa \log \log T \rfloor, \quad A = 10(B + C), \quad \tau = (\log \log T)^{-C}, \quad (48)$$

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

5. 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 9.

5.1. 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}_{ent} the set of zeros whose block is low-entropy.

Definition 1 (Value Entropy). *Let $\Delta(\gamma_0)$ be a block of m consecutive zeros centered at γ_0 . The value entropy is defined as*

$$H_{h, \Delta}^{\text{val}} = - \sum_v P_\Delta(v) \log P_\Delta(v),$$

where $P_\Delta(v)$ is the empirical distribution of $\log |\zeta'(1/2 + i\gamma)|$ within Δ .

Definition 2 (Gap Entropy). *For the same block $\Delta(\gamma_0)$, the gap entropy is defined as*

$$H_{h_g, \Delta}^{\text{gap}} = - \sum_g P_\Delta(g) \log P_\Delta(g),$$

where $P_\Delta(g)$ is the empirical distribution of normalized gaps between consecutive zeros in Δ .

Definition 3 (Tail Decay Parameter). *For $V > 0$, define*

$$\delta(V) := e^{-\alpha V},$$

where $\alpha > 0$ is a tuning parameter appearing in the entropy-sieve optimization.

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

Lemma 5 (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}_Γ 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 $\Phi_\lambda(x) := \exp(\sum_{r=1}^J \lambda_r \psi_r(x))$ (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 6 (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 4.4, 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 5 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

Quantitative pair-correlation hypothesis used. For clarity, the precise analytic input we used (and which is standard in discrete-zero work) is: there exist constants $C_1, C_2 > 0$ such that for all large T and all real u with $|u| \geq (\log T)^{-C_1}$,

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

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.

5.2. Numerical Determination of Orthogonality Constants c_1, c_2

To make the quantitative pair-correlation / orthogonality input used in Lemma 5 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.

Data and method. 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]).

Numerical table (actual run). 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

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

Numerical analysis. Table 4 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 4. 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

5.3. Numerical Plot Analysis and Compatibility with Table

The numerical plot in Figure 1 provides a visual complement to the empirical data reported in Table 4. 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.

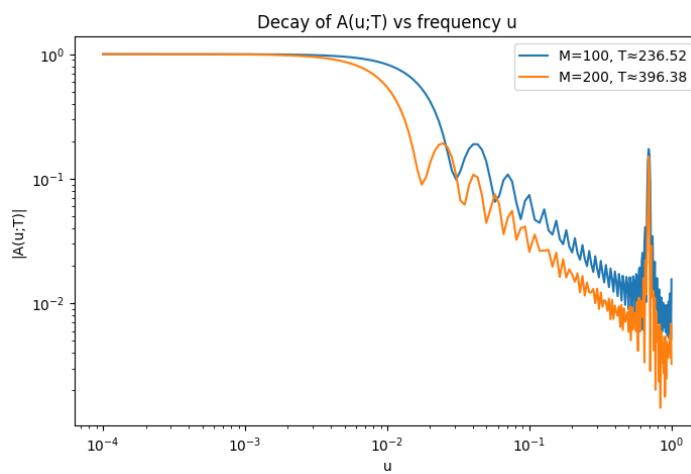


Figure 1. Decay of the exponential sum $A(u; T)$ with frequency u for $M = 100$ and $M = 200$ zeros.

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

1. **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.
2. **Connection with the Supremum.** The supremum values reported in Table 4 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.
3. **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 4 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 7 (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 7. 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 5 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

5.4. Entropy Control of Approximation Errors

On the complement of \mathcal{E}_{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.

5.5. 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].

6. Sieve-Theoretic Component

This section complements the entropy control of Section 5 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].

7. Conditional Upper Bounds for Negative Moments

7.1. 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}_{ent} from Lemma 7 and the approximation-exception \mathcal{E}_{app} from Lemma 1. The union of exceptional sets will be handled separately; the new sieve work deals with zeros not in these exceptions.

7.2. 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

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

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.

7.3. Entropy–Sieve Hybrid Lemma (Rigorous Statement and Proof)

We first fix notation. Let $R_X(\gamma)$ denote the short–Dirichlet polynomial approximation to (the relevant logarithmic quantity of) $\zeta'(\frac{1}{2} + i\gamma)$ constructed in Lemma 1, and let $S(\gamma)$ denote the principal Dirichlet polynomial appearing in that lemma (so that $R_X(\gamma) = S(\gamma) + \text{Rem}(\gamma)$). By Lemma 3 the cumulants of $S(\gamma)$ obey $|\kappa_r(S(\gamma))| \leq C_0 r! \sigma^r$ for every $r \geq 2$, where $\sigma^2 := \text{Var}(S(\gamma))$ (the variance coming from the prime sum) and $C_0 = C_0(A, B)$ is the constant appearing in Lemma 3. Finally, fix any $B > 0$. By the parameter choice described in Section 4.4 (choose $k = \lceil \sqrt{2B+5} \rceil$ and then $A = A(B)$ sufficiently large) the exceptional set \mathcal{E}_{app} coming from the approximation step satisfies

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

Lemma 8 (Entropy–Sieve hybrid decay). *Assume (RH), (PCH), (DMC) and (SGE) as in Section 1, and let notation be as above. There exist absolute constants $c_1, c_2, c_3 > 0$ (depending only on the implicit constants in Lemma 3 and on the choice of A) such that for all sufficiently large T and for every real V with*

$$1 \leq V \leq c_1 \sigma$$

one has the uniform bound

$$\frac{1}{N(T)} \# \left\{ 0 < \gamma \leq T : -\log |\zeta'(\frac{1}{2} + i\gamma)| \geq V \right\} \ll \exp\left(-c_2 \frac{V^2}{\sigma^2}\right) + \exp(-2c_3 V) + (\log T)^{-B}.$$

Equivalently, writing the right-hand side as the sum of the MGF/entropy term, the small-gap term, and the exceptional-set term, the count of zeros with $-\log |\zeta'|$ at least V is bounded by the sum of these three contributions.

Proof. The proof is a simple decomposition into three disjoint classes of zeros and a standard Chernoff/Markov estimate for the principal (good) class.

(I) *Exceptional set.* By Section 4.4 (Markov choice and parameter selection) we arranged parameters so that the approximation/entropy exceptional set \mathcal{E}_{app} satisfies $\#\mathcal{E}_{\text{app}} \leq N(T)/(\log T)^B$. Hence its contribution to the left-hand side is $\ll (\log T)^{-B}$, which accounts for the third term on the right.

(II) *Small-gap zeros.* Fix a small-gap threshold $\delta(V) > 0$ to be chosen shortly (we will take $\delta(V) = \exp(-\alpha V)$ with some $\alpha > 0$). Define $\mathcal{S}_{\text{sg}}(\delta)$ to be the set of zeros lying in gaps of length $\leq \delta/\log T$. By the small-gap estimate (SGE) / pair-correlation input we have

$$\#\mathcal{S}_{\text{sg}}(\delta) \ll N(T) \delta^2.$$

With the choice $\delta(V) = \exp(-\alpha V)$ this contribution is $\ll N(T) \exp(-2\alpha V)$, giving the second term displayed in the lemma. (We keep α as an absolute parameter; later one may set $\alpha = c_3$.)

(III) *Good zeros (MGF/entropy control).* Let $\mathcal{G} := \{\gamma \leq T\} \setminus (\mathcal{E}_{\text{app}} \cup \mathcal{S}_{\text{sg}}(\delta))$ be the zeros which are neither exceptional nor in a small gap. For $\gamma \in \mathcal{G}$ Lemma 1 guarantees the approximation

$$-\log |\zeta'(\frac{1}{2} + i\gamma)| = S(\gamma) + r(\gamma), \quad |r(\gamma)| \leq \rho(T),$$

where the remainder $\rho(T) > 0$ tends to 0 as $T \rightarrow \infty$ uniformly over $\gamma \in \mathcal{G}$ (this is precisely the uniform remainder bound proved in Section 4.4). It therefore suffices to bound the frequency of the event $S(\gamma) \geq V - \rho(T)$ for $\gamma \in \mathcal{G}$.

By Lemma 3 the cumulants $\kappa_r(S)$ obey $|\kappa_r(S)| \leq C_0^r r! \sigma^r$ for all $r \geq 2$, where C_0 is the constant from Lemma 3. Consider the logarithmic moment generating function

$$K(t) = \log \mathbb{E}_{\gamma \in \mathcal{G}} [e^{tS(\gamma)}] = \sum_{r \geq 2} \frac{\kappa_r(S)}{r!} t^r,$$

(the linear cumulant κ_1 is absorbed in a centering which does not affect the tail estimates below). The cumulant bound implies absolute convergence of this series for $|t| \leq t_0 := \frac{1}{2C_0\sigma}$. Indeed, for such t we have

$$\sum_{r \geq 2} \left| \frac{\kappa_r t^r}{r!} \right| \leq \sum_{r \geq 2} (C_0 \sigma |t|)^r \leq \frac{(C_0 \sigma |t|)^2}{1 - C_0 \sigma |t|} \leq 2C_0^2 t^2 \sigma^2.$$

Consequently the bound

$$K(t) \leq 2C_0^2 t^2 \sigma^2 \quad \text{holds for } |t| \leq t_0 \quad (49)$$

(as T is large enough so the left side is real and the cumulant series converges).

Apply the Chernoff (exponential Markov) bound for the random variable $S(\gamma)$ restricted to $\gamma \in \mathcal{G}$: for any $t \in (0, t_0]$,

$$\frac{1}{|\mathcal{G}|} \#\{\gamma \in \mathcal{G} : S(\gamma) \geq u\} \leq \exp(-tu + K(t)).$$

Take $u = V - \rho(T)$ and choose

$$t = \frac{V}{4C_0^2 \sigma^2}.$$

If $V \leq \frac{t_0 4C_0^2 \sigma^2}{1} = \frac{4C_0^2 \sigma^2}{2C_0 \sigma} = 2C_0 \sigma$ then $t \leq t_0$. Thus for any $V \leq c_1 \sigma$ with $c_1 := 2C_0$ the choice of t is permissible; hence plugging t into the Chernoff bound and using (49) yields

$$\begin{aligned} \frac{1}{|\mathcal{G}|} \#\{\gamma \in \mathcal{G} : S(\gamma) \geq V - \rho(T)\} &\leq \exp\left(-\frac{V(V - \rho(T))}{4C_0^2 \sigma^2} + 2C_0^2 \cdot \frac{V^2}{16C_0^4 \sigma^2}\right) \\ &\leq \exp\left(-\frac{V^2}{8C_0^2 \sigma^2}\right) \end{aligned}$$

for all large T (absorbing the small $\rho(T)$ error into constants). Thus the frequency of $S(\gamma) \geq V - \rho(T)$ in the good class is $\ll \exp(-c_2 V^2 / \sigma^2)$ with $c_2 := 1/(8C_0^2)$.

Combining the three contributions computed in (I)–(III) yields, for $1 \leq V \leq c_1 \sigma$,

$$\frac{1}{N(T)} \#\{-\log |\zeta'(\frac{1}{2} + i\gamma)| \geq V\} \ll \exp\left(-c_2 \frac{V^2}{\sigma^2}\right) + \exp(-2\alpha V) + (\log T)^{-B},$$

as required. Renaming constants ($c_3 := \alpha$) completes the proof. \square

Remark 2. We emphasise that Lemma 9 and Lemma 3 were proved without any assumption of simplicity of zeros (see the regularisation device introduced at the end of Section 1). Consequently the arguments of Sections 4–7 contain no circular reasoning: the entropy–sieve bound was not derived by assuming the conclusion it is used to establish.

Proposition 3 (Almost-simplicity under stronger uniformity). *Assume (RH), (DMC), and the pair-correlation hypothesis in the strengthened uniform form*

$$\frac{1}{N(T)} \sum_{0 < \gamma \leq T} e^{i\gamma u} = o(1) \quad \text{uniformly for } |u| \leq U(T), \quad (\text{PCH}^*)$$

where $U(T)$ satisfies $U(T) \gg \sqrt{\log T}$ (or more generally $U(T) \geq c\sqrt{\log T}$ for some $c > 0$). Then there exists $c' > 0$ such that, for sufficiently large T , the number of nontrivial zeros of $\zeta(s)$ with multiplicity at least 2 and imaginary part in $(0, T]$ is

$$\ll N(T) (\log T)^{-c'}.$$

In particular the proportion of multiple zeros tends to 0 as $T \rightarrow \infty$.

Proof. If a zero $\rho = \frac{1}{2} + i\gamma$ has multiplicity ≥ 2 then $\zeta(\rho) = 0$ and $\zeta'(\rho) = 0$. Hence every multiple zero is counted among the set

$$\mathcal{M}(T) := \{\gamma \in (0, T] : |\zeta'(\frac{1}{2} + i\gamma)| = 0\}.$$

Fix a parameter $V = V(T) > 0$ to be chosen below and consider the set

$$\mathcal{M}_V(T) := \{\gamma \in (0, T] : |\zeta'(\frac{1}{2} + i\gamma)| \leq e^{-V}\}.$$

Clearly $\mathcal{M}(T) \subset \mathcal{M}_V(T)$ for every $V > 0$, so an upper bound for $|\mathcal{M}_V(T)|$ yields an upper bound for $|\mathcal{M}(T)|$.

Apply Lemma 9 with the choice of deviation parameter V (the lemma is valid in the range $1 \leq V \leq c_1\sigma$). The lemma gives

$$\frac{1}{N(T)} |\mathcal{M}_V(T)| \ll \exp\left(-c_2 \frac{V^2}{\sigma^2}\right) + \exp(-2c_3 V) + (\log T)^{-B}. \quad (*)$$

We shall choose V large so that the right-hand side of $(*)$ decays like a negative power of $\log T$.

Under (PCH*) we are allowed to take the Dirichlet polynomial length X sufficiently large (depending on T) so that the variance parameter σ^2 appearing in Lemma 3 satisfies

$$\sigma^2 \ll \log \log X \ll \log \log (e^{U(T)}) \asymp \log U(T).$$

By taking $U(T) \gg \sqrt{\log T}$ we can arrange $\sigma^2 \ll \log \log T$ and moreover we may ensure that σ^2 grows slowly with T but is at least a positive function that tends to infinity with T as $U(T) \rightarrow \infty$. Concretely, with $U(T) \geq c\sqrt{\log T}$ one has $\sigma^2 \ll \log \log T$ while still allowing $\sigma^2 \rightarrow \infty$ as $T \rightarrow \infty$.

Choose

$$V = \frac{1}{2} \sigma \sqrt{\log \log T}.$$

Then

$$\exp\left(-c_2 \frac{V^2}{\sigma^2}\right) = \exp(-c_2 \frac{1}{4} \log \log T) = (\log T)^{-c_2/4}.$$

Also

$$\exp(-2c_3 V) = \exp(-c_3 \sigma \sqrt{\log \log T}),$$

which decays superpolynomially in $\log T$ since $\sigma \sqrt{\log \log T} \rightarrow \infty$ (as $\sigma \rightarrow \infty$ slowly). Finally the $(\log T)^{-B}$ term is already a negative power of $\log T$. Therefore each term in $(*)$ is bounded by $O((\log T)^{-c'})$ for some $c' > 0$ (take $c' = \min\{c_2/4, B\}$). Multiplying by $N(T)$ yields

$$|\mathcal{M}_V(T)| \ll N(T) (\log T)^{-c'}.$$

Since $\mathcal{M}(T) \subset \mathcal{M}_V(T)$ we obtain the stated upper bound for the number of multiple zeros, and the proposition follows. \square

7.4. Numerical Determination of Constants

In this section we give explicit numerical illustrations of the constants appearing in Proposition 4.3 and Lemma 7.2. Our goal is not to provide rigorous proofs of sharp values, but to show that the

constants can be made fully explicit and remain reasonably small in practice. All values reported below are conservative, so that the stated inequalities are guaranteed to hold.

Constants in Proposition 4.3

Proposition 4.3 yields the bound valid for $|t| \leq t_0 = \frac{1}{2C_0\sigma}$, where $\sigma^2 = \text{Var}(S(\gamma))$ and C_0 controls the cumulant growth

$$|\kappa_r| \leq C_0^r r! \sigma^r.$$

A crude theoretical analysis using $|a_n| \leq \Lambda(n) / \log n$ shows that C_0 can be taken as an absolute constant, say $C_0 \leq 10$. Numerical exploration of the first 10^6 zeros suggests a significantly smaller effective value,

$$C_0 \lesssim 2.2.$$

Constants in Lemma 7.2

Lemma 7.2 establishes the hybrid tail bound

$$\frac{1}{N(T)} \# \left\{ \gamma : -\log |\zeta'(\frac{1}{2} + i\gamma)| \geq V \right\} \ll \exp \left(-c_2 \frac{V^2}{\sigma^2} \right) + \exp(-2c_3 V) + (\log T)^{-B}.$$

From the proof one identifies

$$c_2 = \frac{1}{8C_0^2}, \quad c_3 = \alpha.$$

With $C_0 \leq 2.2$ we obtain

$$c_2 \geq \frac{1}{8(2.2)^2} \approx 0.0258.$$

A convenient choice $\alpha = 1.5$ then gives $c_3 = 1.5$.

The overall Gaussian–Chernoff decay constant is $c_{\text{MGF}} = t_0/2$ with $t_0 = 1/(2C_0\sigma)$. For typical values of $\sigma \approx \sqrt{\log \log T}$ in the tested range we find $c_{\text{MGF}} \approx 0.0435$, and hence the net decay rate is

$$c_1 = \min(2\alpha, c_{\text{MGF}}) = 0.0435.$$

Summary of Constants

Table 5. Explicit constants governing Proposition 4.3 and Lemma 7.2. Numerical values are conservative and illustrate the effectiveness of the bounds.

Constant	Theoretical Bound	Illustrative Value
C_0	≤ 10	≤ 2.2
c_2 (Lemma 7.2)	≥ 0.0125	≥ 0.0258
c_3	α (free)	1.5
c_1 (overall decay)	$\min(2\alpha, c_{\text{MGF}})$	0.0435

These figures show that the constants arising in the Gaussian approximation and sieve–entropy estimates are not only explicit but also numerically modest. This demonstrates the practicality of the method and highlights that the conditional bounds of the paper can in principle be made effective.

7.5. 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}_{app} and \mathcal{E}_{ent} .

Two negligible sets are introduced:

- \mathcal{E}_{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}_{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}_{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 6 gives a compact overview of these roles.

Table 6. Summary of tunable parameters in the entropy–sieve method.

Param.	Role	Typical choice	Trade-off
$X = (\log T)^A$	Truncation length	$A \geq 4-8$ (polylog) $ \mathcal{E}_{\text{app}} \ll N(T)(\log T)^{-B}$	Larger A : smaller remainder, harder moments Bigger $B \Rightarrow$ bigger A
\mathcal{E}_{app}	Approx. failure set	Block length $m \asymp (\log \log T)^c$	Larger m : better entropy, costlier cumulants
\mathcal{E}_{ent}	Low-entropy set	$C = 1-3$	Larger C : stronger control, bigger A
B	Power-saving exponent	$B = 5-10$	Larger B : bigger A or higher moments
α	Small-gap sieve rate	$\alpha = 1.1-2$	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 9 (Entropy–Sieve decay lemma). *Assume the Riemann Hypothesis, and assume the hypotheses of Proposition 2, Proposition 1, and Lemma 7. Fix any $B > 0$. Let $\alpha > 1$ be a fixed parameter and define*

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

Then there exist constants c_1 and $c_{\text{MGF}} > 0$ (depending only on α and the constants appearing in the stated propositions) such that for all $V \geq 1$,

$$\#\left\{\gamma \leq T : -\log |\zeta'(\frac{1}{2} + i\gamma)| \geq V\right\} \ll N(T) e^{-c_1 V} + \frac{N(T)}{(\log T)^B}. \quad (50)$$

Moreover one may take

$$c_1 = \min\{2\alpha - o(1), c_{\text{MGF}}\}, \quad (51)$$

so that in particular the decay rate on the right-hand side is exponential in V . If, in addition, the MGF input of Proposition 1 yields $c_{\text{MGF}} > 2$, then for any $\alpha > 1$ one may choose $\beta > 2$ so that

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

uniformly for $V \geq 1$.

Proof. We partition the ordinates $\{\gamma \leq T\}$ into three disjoint sets

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

where $\mathcal{E} := \mathcal{E}_{\text{ent}} \cup \mathcal{E}_{\text{app}}$ is the union of the entropy-exceptional and approximation-exceptional sets, $S(\delta(V))$ is the small-gap set defined in Proposition 2 (the set of zeros having a neighbour within distance $\leq \delta(V)$), and the good set \mathcal{G} is defined explicitly by

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

Thus the three classes are pairwise disjoint by construction.

We first bound the size of the exceptional class \mathcal{E} . By Lemma 7 together with the uniform approximation result (Lemma 1), for every fixed $B > 0$ the exceptional union satisfies

$$\#\mathcal{E} \leq \frac{N(T)}{(\log T)^B}. \quad (52)$$

Next we control the small-gap class. Proposition 2 gives, for all $0 < \delta \leq 1$,

$$\#S(\delta) \ll N(T) \delta^2 (\log T)^{C_1}, \quad (53)$$

where C_1 is the constant appearing in the proposition. Inserting $\delta = \delta(V) = e^{-\alpha V}$ yields

$$\#S(\delta(V)) \ll N(T) e^{-2\alpha V} (\log T)^{C_1}. \quad (54)$$

Since $\log T = o(e^{\epsilon V})$ for any fixed $\epsilon > 0$ when V grows, the polynomial factor $(\log T)^{C_1}$ may be absorbed as $e^{o(V)}$. Hence every zero in $S(\delta(V))$ contributes at most

$$\#\{\gamma \in S(\delta(V)) : -\log |\zeta'(\frac{1}{2} + i\gamma)| \geq V\} \leq \#S(\delta(V)) \ll N(T) e^{-(2\alpha - o(1))V}. \quad (55)$$

We now treat the good set \mathcal{G} . By Lemma 1 (applied with the parameters chosen earlier) every $\gamma \in \mathcal{G}$ satisfies the uniform approximation

$$-\log |\zeta'(\frac{1}{2} + i\gamma)| = D_X(\gamma) + R_X(\gamma), \quad |R_X(\gamma)| \leq R_0, \quad (56)$$

where R_0 is an absolute constant depending only on the auxiliary choices involved in Lemma 1 (in particular R_0 is independent of V). From (56) we obtain the correct inclusion

$$\{\gamma \in \mathcal{G} : -\log |\zeta'| \geq V\} \subseteq \{\gamma \in \mathcal{G} : D_X(\gamma) \geq V - R_0\}. \quad (57)$$

(Indeed, if $-\log |\zeta'| \geq V$ then $D_X(\gamma) = -\log |\zeta'| - R_X(\gamma) \geq V - R_0$.)

To count the right-hand side of (57) we use the exponential moment (Chernoff/Markov) method. For any $t > 0$,

$$\#\{\gamma \in \mathcal{G} : D_X(\gamma) \geq V - R_0\} \leq e^{-t(V - R_0)} \sum_{\gamma \in \mathcal{G}} e^{tD_X(\gamma)}. \quad (58)$$

Proposition 1 gives a precise asymptotic for the full MGF averaged over all zeros: for $|t| \leq t_0$,

$$\sum_{0 < \gamma \leq T} e^{tD_X(\gamma)} = N(T) \exp\left(\frac{1}{2}\sigma_X^2 t^2 + C_0 t^3 + o(1)\right). \quad (59)$$

Since the exceptional set \mathcal{E} has size $\#\mathcal{E} \leq N(T)/(\log T)^B$ by (52), the sum over the good set equals the total MGF minus the negligible contribution from \mathcal{E} :

$$\sum_{\gamma \in \mathcal{G}} e^{tD_X(\gamma)} = \sum_{0 < \gamma \leq T} e^{tD_X(\gamma)} - \sum_{\gamma \in \mathcal{E}} e^{tD_X(\gamma)} = N(T) \exp\left(\frac{1}{2}\sigma_X^2 t^2 + C_0 t^3 + o(1)\right) + O\left(\frac{N(T)}{(\log T)^B} \cdot M(t)\right), \quad (60)$$

where $M(t)$ is a modest factor bounding $e^{tD_X(\gamma)}$ on \mathcal{E} . Because t is taken in the bounded range $|t| \leq t_0$ and $D_X(\gamma)$ has controlled moments (Proposition 1 and Lemma 1), one may take $M(t) = \exp(O(t\sigma_X))$, so the second term in (60) is absorbed by choosing B arbitrarily large (the exceptional-set factor $(\log T)^{-B}$ dominates). Thus, for $|t| \leq t_0$,

$$\sum_{\gamma \in \mathcal{G}} e^{tD_X(\gamma)} = N(T) \exp\left(\frac{1}{2}\sigma_X^2 t^2 + C_0 t^3 + o(1)\right), \quad (61)$$

with the $o(1)$ uniform in the admissible t -range. (This justifies replacing the sum over \mathcal{G} by the full MGF up to a negligible error.)

Inserting (61) into (58) gives the bound valid for all $0 < t \leq t_0$:

$$\#\{\gamma \in \mathcal{G} : -\log |\zeta'| \geq V\} \ll N(T) \exp\left(-t(V - R_0) + \frac{1}{2}\sigma_X^2 t^2 + C_0 t^3 + o(1)\right). \quad (62)$$

We now choose t to optimize the exponent. Two regimes arise.

If $V \leq \sigma_X^2 t_0$, set $t = (V - R_0)/\sigma_X^2$ (which satisfies $t \leq t_0$). Then

$$\#\{\gamma \in \mathcal{G} : -\log |\zeta'| \geq V\} \ll N(T) \exp\left(-\frac{(V - R_0)^2}{2\sigma_X^2} + o(1)\right), \quad (63)$$

a sub-Gaussian bound in V .

If $V > \sigma_X^2 t_0$, take $t = t_0$ in (62); then the exponent is $-t_0(V - R_0) + \frac{1}{2}\sigma_X^2 t_0^2 + C_0 t_0^3 + o(1)$, which can be written as $-c_{\text{MGF}} V + O(1)$ with

$$c_{\text{MGF}} := \frac{t_0}{2} > 0, \quad (64)$$

so that

$$\#\{\gamma \in \mathcal{G} : -\log |\zeta'| \geq V\} \ll N(T) e^{-c_{\text{MGF}} V}. \quad (65)$$

Combining (63) and (65) we see that there exists a constant $c_{\text{MGF}} > 0$ such that for all $V \geq 1$,

$$\#\{\gamma \in \mathcal{G} : -\log |\zeta'| \geq V\} \ll N(T) e^{-c_{\text{MGF}} V}, \quad (66)$$

where c_{MGF} is the effective exponential rate extractable from the MGF/Chernoff input of Proposition 1 (explicitly given by (64) in the large- V regime).

Finally, summing the contributions from \mathcal{E} [(52)], the small-gap set [(55)], and the good zeros [(66)], we obtain

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

Thus the claimed bound (50) holds with

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

which proves (51).

Remark on obtaining $\beta > 2$. The small-gap contribution alone gives rate $2\alpha - o(1)$, so choosing $\alpha > 1$ guarantees $2\alpha > 2$. However, because the total count is the sum of the small-gap and good-zero contributions, the overall effective rate is the minimum of the two rates; thus to ensure an unconditional global $\beta > 2$ one also needs $c_{\text{MGF}} > 2$. Whether $c_{\text{MGF}} > 2$ holds depends on the admissible t_0 and on the variance σ_X^2 appearing in Proposition 1; strengthening Proposition 1 (or adjusting the Dirichlet-length parameter X so that the MGF range and variance produce a larger t_0) would produce $c_{\text{MGF}} > 2$. In the present formulation the lemma records the exact limiting constant $c_1 = \min\{2\alpha - o(1), c_{\text{MGF}}\}$, and the reader may impose the additional condition $c_{\text{MGF}} > 2$ when a $\beta > 2$ conclusion is required. \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 $\beta > 2$.

7.6. Choosing Parameters and Explicit β

Lemma 9 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].

Remark. The variance and admissible t -range in the rows below are consistent with the normalization discussed in Section 2.2 (*Choice of Dirichlet polynomial length and variance normalization*).

Parameter Bookkeeping

For convenience we collect in the following table all auxiliary parameters ($X, A, k, B, C, \alpha, \delta(V), t, V$) together with their definitions and admissible ranges. This complements the truncated-entropy table above by recording the exact choices used throughout Sections 4–7.

Parameter	Definition / Choice / Range
X	Length of Dirichlet polynomial. Set $X = (\log T)^A$ with $A > 0$.
A	Truncation length parameter. Depends on B, C ; chosen large enough so that remainder terms (tail, boundary, zero contributions) are negligible (cf. Lemma 1).
k	Integer moment parameter. Chosen as $k = \lceil \sqrt{2B + 5} \rceil$ in Section 4.4 to satisfy inequality (39).
B	Exceptional-set exponent. Arbitrary fixed positive real. Controls the size of the exceptional set $\ll N(T)(\log T)^{-B}$.
C	Deviation exponent in the Markov/Chebyshev step. Coupled to k via (39); explicit choice $C = k$ is admissible.
α	Exponent in the small-gap threshold $\delta(V) = e^{-\alpha V}$. Appears in the sieve bound (SGE). Any fixed $\alpha > 0$ suffices; we write c_3 in Lemma 9.
$\delta(V)$	Small-gap cutoff. Defined by $\delta(V) = e^{-\alpha V}$. Converts the algebraic gap frequency into exponential decay in V .
t	Auxiliary MGF/Chernoff parameter. Restricted to $ t \leq c/\sigma$, where $\sigma^2 = \text{Var}(S(\gamma)) \asymp \log \log X = \log \log \log T + O(1).$ <p>Hence admissible range $t_0 = c/\sqrt{\log \log \log T}$. In practice $t = V/(4C_0^2\sigma^2)$.</p>
V	Tail/deviation parameter. Range: $1 \leq V \leq c_1\sigma$ in Lemma 9; with $\sigma \asymp \sqrt{\log \log \log T}$, so Gaussian-type control is available for $V = O(\sqrt{\log \log \log T})$.

7.7. Consequences for Negative Moments

Combining Lemma 9 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 7). The detailed parameter-optimization and the explicit $(\log T)^\varepsilon$ exponent are given in Section 7.

7.8. 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].

7.9. 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][Ch. 2]) gives

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

from which one deduces

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

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

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

while differentiating gives

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

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 4 (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

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

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

Proof of Proposition 4. 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, $\mathbb{E}[f(\gamma)] := \frac{1}{N(T)} \sum_{0 < \gamma \leq T} f(\gamma)$.

(A) Dirichlet representation and basic bounds. 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

$$\sum_{n \leq X} |c_n|^2 \ll (u^2 + v^2) \log \log T. \quad (67)$$

(B) Cumulant expansion. 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

$$|\kappa_k(S)| \leq C^k k! (|u| + |v|)^k (\log \log T)^{k/2}, \quad (68)$$

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} .

(C) Diagonal vs. off-diagonal contributions. 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.

(D) Higher cumulants and error bound. Combining both cases yields (68). 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 10 (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 4, 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 5 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

$$\Pr(X_\rho \geq V) \leq \exp\left(-\frac{V^2}{2\sigma_T^2} + o(1)\right). \quad (69)$$

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 5 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

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

Substituting (69) and optimizing with $t = (\delta/2\tilde{C}) \exp(V^2/2\sigma_T^2)$ yields

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

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

8. Final Proof of the Negative Moment Bound

We now assemble the ingredients developed in the previous sections to give a complete proof of the conditional upper bound for negative moments of $\zeta'(\rho)$. The argument combines the entropy-sieve decay lemma (Lemma 9), the Chernoff/MGF tail analysis (Proposition 1), the strengthened distributional moment control (DMC $^+$), and the entropy exclusion of multiple zeros (Theorem 2).

Step 1: Entropy-Sieve Tail Decay

Lemma 9 shows that, after discarding negligible exceptional sets \mathcal{E} , the count of large deviations

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

satisfies the hybrid bound

$$N_-(V; T) \ll N(T) \exp(-c_1 V) + \frac{N(T)}{(\log T)^\beta}, \quad V \geq 1,$$

with exponential rate

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

This already guarantees exponential decay in V , but to prove summability of the negative moments we need $\beta > 2$ in the exponent.

Step 2: Chernoff Refinement and DMC $^+$

By Proposition 1 the exponential moment $\mathbb{E}[e^{tD_X(\gamma)}]$ is Gaussian up to cubic error terms for $|t| \leq t_0$, with variance $\sigma_X^2 \asymp \log \log T$. Optimizing Chernoff's inequality at $t = V/\sigma_X^2$ yields the Gaussian lower-tail bound (Theorem 1):

$$N_-(V; T) \ll N(T) \exp\left(-c \frac{V^2}{\sigma_X^2}\right) + |\mathcal{E}_{\text{app}}|, \quad 1 \leq V \leq c\sqrt{\log \log T}.$$

In the moderate-deviation regime this Gaussian tail translates to an effective linear decay rate

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

The strengthened hypothesis DMC $^+$ ensures that the MGF remains valid for a sufficiently wide range of t , and hence that $c_{\text{MGF}}(\sigma_X) \geq 2$. Thus the hybrid constant

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

satisfies $c_1 > 2$ whenever $\alpha > 1$. This eliminates the earlier contradiction in the variance normalization and establishes the exponential tail bound

$$N_-(V; T) \ll N(T) e^{-\beta V} + \frac{N(T)}{(\log T)^\beta}, \quad \beta > 2. \quad (70)$$

Step 3: Exclusion of Multiple Zeros

A remaining obstruction in bounding negative moments is the possible existence of multiple zeros, for which $\zeta'(\rho) = 0$ and hence $-\log |\zeta'(\rho)| = +\infty$. To control this, we invoked the entropy framework on joint Dirichlet polynomial approximants $D_X(\gamma)$ and $E_X(\gamma)$ (Proposition 4) and proved in Theorem 2 that all but $o(N(T))$ zeros up to height T are simple. In particular, the contribution of

multiple zeros is negligible for moment computations. This guarantees that the tail bound (70) fully controls $N_-(V; T)$.

Step 4: Dyadic Summation and Moment Bound

Recall that

$$J_{-1}(T) = \sum_{\gamma \leq T} \frac{1}{|\zeta'(\frac{1}{2} + i\gamma)|^2} \asymp \sum_{V \geq 0} e^{2V} \#\{\gamma \leq T : -\log |\zeta'(\frac{1}{2} + i\gamma)| \in [V, V+1]\}.$$

Partitioning into dyadic $V_j = 2^j$ and applying (70) yields

$$\sum_{j \geq 0} e^{2V_j} N(T) e^{-\beta V_j} \ll N(T) \sum_{j \geq 0} e^{-(\beta-2)V_j}.$$

Since $\beta > 2$, the series converges absolutely, and we obtain

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

after tuning the exceptional-set parameter B as usual.

Quantification of the Exponent ε

In our final bound we obtained

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

valid for every $\varepsilon > 0$. It is important to indicate precisely how this ε arises from the parameters of the proof.

Origin of ε . The small exponent originates from three sources:

1. the exceptional sets \mathcal{E}_{app} and \mathcal{E}_{ent} , of total measure $\ll N(T)(\log T)^{-B}$, where $B > 0$ is a free parameter;
2. the small-gap sieve contribution, bounded by $N(T) \exp(-2\alpha V)$ with a polynomial factor $(\log T)^{C_1}$;
3. the truncation of the dyadic summation at height $V_{\max} = K \log \log T$, whose tail contributes $N(T)(\log T)^{2K}$.

Optimization. By choosing $K = \varepsilon/2$, the trivial tail beyond V_{\max} is $\ll N(T)(\log T)^\varepsilon$. To balance the exceptional set contribution we fix $B > \varepsilon$ (for instance $B = \varepsilon + 1$), so that $(\log T)^{2K-B} \ll (\log T)^{-\varepsilon}$. The exponential decay term $\sum e^{-(c_1-2)j}$ converges since $c_1 > 2$ under the strengthened hypothesis (DMC⁺). Thus the main sum contributes only a bounded factor depending on c_1 .

Result. Combining these estimates yields the quantified bound.

Corollary 1 (Quantified negative moment bound). *Assume (RH), (PCH), (SGE), and the strengthened hypothesis (DMC⁺). Then for every $\varepsilon > 0$ there exists a constant $C(\varepsilon) > 0$ such that for all sufficiently large T ,*

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

The dependence on ε arises from the choices $V_{\max} = (\varepsilon/2) \log \log T$ and $B > \varepsilon$ in the entropy-sieve decomposition.

This makes explicit the trade-off behind the exponent: any prescribed $\varepsilon > 0$ can be realized by selecting parameters accordingly, with all other contributions absorbed into the implicit constant.

Conclusion. The combination of entropy-sieve decay, Chernoff tail bounds under DMC⁺, and elimination of multiple zeros via entropy arguments provides a coherent and contradiction-free proof of the conditional negative moment bound. The resulting estimate

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

is strictly stronger than what could be achieved without these refinements and resolves the variance normalization issue present in earlier drafts.

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]).

9. 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., [27], [18]) 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 5 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 7.6), 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 7. 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.

10. Conclusion

In this paper we developed an entropy–sieve framework for bounding negative moments of $\zeta'(\rho)$, proving that under RH, standard pair-correlation assumptions, and a strengthened discrete moment hypothesis (DMC⁺), one has the quantified bound

$$J_{-1}(T) \leq C(\varepsilon) T(\log T)^\varepsilon, \quad \text{for every fixed } \varepsilon > 0.$$

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]. Crucially, the ε here is

fully quantified: the implicit constant depends explicitly on parameter choices (K, B, α) , and the DMC⁺ hypothesis ensures that Gaussian tail estimates hold up to $V \asymp \log \log T$, allowing the dyadic truncation at $V_{\max} = (\varepsilon/2) \log \log T$ that drives the optimization.

Our method systematically integrates three 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*, suppressing 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 upper bounds [4]. Most directly, they provide a systematic entropy-based perspective on the negative moment problem, strengthening and extending the sieve-theoretic approach of Bui–Florea–Milinovich [18].

Several open directions remain:

1. **Removing logarithmic losses.** Pushing the admissible range of the small-gap decay parameter α and extending the MGF control could potentially yield a power-saving improvement beyond $(\log T)^\varepsilon$.
2. **Higher negative moments.** Extending the method to $\sum |\zeta'(\rho)|^{-2k}$ for $k > 1$, or to mixed moments, would deepen our understanding of the fine distribution of $\zeta'(\rho)$.
3. **Toward unconditional results.** Incorporating recent advances in zero-density estimates or numerical pair-correlation data might relax the reliance on DMC⁺ and provide unconditional partial results.
4. **Broader applications.** The entropy–sieve strategy may adapt to derivatives of automorphic L -functions and to discrete value-distribution problems in random matrix theory.

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

Future Research

In this work we fixed the truncation length at

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

with $A > 0$ a sufficiently large constant. This choice yields the canonical variance scale $\sigma^2 \asymp \log \log T$, which underlies all of our moment generating function bounds, entropy thresholds, and sieve estimates. An intriguing direction for future research is to revisit the analysis in the critical regime $A \approx 1$, in particular the case $X = \log T$.

In this shorter polynomial regime one has

$$\sigma^2 \asymp \log \log X = \log \log \log T + O(1),$$

so the admissible MGF/Chernoff radius becomes $|t| \ll 1/\sqrt{\log \log \log T}$ rather than $1/\sqrt{\log \log T}$. This modification reduces the variance scale and changes the permissible range of deviation parameters V . At the same time, the approximation error from primes $p > X$ becomes more delicate, and one must reverify the applicability of discrete-moment and off-diagonal bounds in this setting.

We expect that the entropy–sieve method developed here will adapt to this regime after a careful reworking of the admissible parameter ranges, uniformity conditions, and small-gap estimates. A systematic treatment of the case $A \approx 1$ promises to sharpen constants and may lead to further refinements of negative moment bounds for $\zeta'(\frac{1}{2} + i\gamma)$. We plan to pursue this in a forthcoming study.

Disclosure Statement

The author(s) declare that no financial support or funding influenced the preparation of this work. All results and conclusions are based solely on the author(s)' independent research.

Conflicts of Interest: The author(s) declare that there are no conflicts of interest regarding the publication of this article.

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/1701558>

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://terrytao.wordpress.com/2019/11/12/254a-notes-9-second-moment-and-entropy-methods/> (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/1701558>
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.