The Arithmetic Recognition Operator: Information Geometry, the Modular Surface, and a Conjectured Path to the Riemann Hypothesis

D. Christian[1,*]

[1] *The Axioms of Pattern Ontology (APO) Initiative*

(Dated: March 22, 2026)

We construct an *arithmetic recognition operator*[†] on the statistical manifold of the zeta distribution $P_s(n) = n^{-s}/\zeta(s)$, using the Bhattacharyya coefficient as an inner-product kernel. The resulting operator $\odot(s, s') = \zeta\left(\frac{s+s'}{2}\right)/\sqrt{\zeta(s)\,\zeta(s')}$ is symmetric, bounded, and has singularities exactly at the zeros of the Riemann zeta function.

Rigorous core. We prove that the completed (Pass 2) version inherits the functional equation as a $\mathbb{Z}_2$ symmetry, that the Fisher metric is diagonal on the critical line $\sigma = \frac{1}{2}$, and that the recognition distribution of any arithmetic pattern is invariant under $\sigma \leftrightarrow 1-\sigma$. We identify the modular surface $M = \mathbb{H}/\mathrm{SL}(2,\mathbb{Z})$ as the natural geometric home, where the critical line corresponds to the geodesic fixed by the functional equation's reflection and the Riemann zeros appear as poles of the scattering matrix.

Conjectural bridge. We argue (but do not prove) that one cycle of the distinction–recognition–integration process projects arithmetic patterns onto the critical line in a single step, the key open step being a formalization of the quotient map on the non-compact arithmetic manifold. A Kolmogorov complexity argument suggests that on-line zero configurations ($\sigma = \frac{1}{2}$ for all zeros) are maximally compressed, though this describes coding economy of configurations rather than a mechanism forcing zeros of ξ. We formulate a Solomonoff-weighted integral operator and prove it is trace-class via the Kraft inequality, yielding a complete spectral decomposition, but show that the resulting compactification does not control fine spectral structure.

The remaining gap. We survey the distance between these results and a proof of the Riemann Hypothesis. The gap appears in six equivalent guises—divergence vs. overlap, per-prime vs. collective, discrete vs. scattering, compact vs. cusp, statistics vs. location, and trace-class vs. fine structure—each reflecting the common obstruction that knowing the measure does not determine the support. The paper concludes with a conjectured correspondence between the Bhattacharyya kernel and Mayer's Bessel/Hankel kernel on the Hardy space, identifies three sub-gaps in that correspondence, and states three falsifiable predictions.

Every claim is labeled PROVEN, ARGUED, CONJECTURED, or REFUTED throughout. [†]APO-specific terminology; see Appendix A for formal definitions and foundational status.

CONTENTS

* daniel@freereason.org

I. INTRODUCTION

The Riemann Hypothesis (RH) asserts that every non-trivial zero ρ of the Riemann zeta function $\zeta(s) = \sum_{n=1}^{\infty} n^{-s}$ satisfies $\mathrm{Re}(\rho) = \frac{1}{2}$ [1]. It is the central open problem of analytic number theory, with implications for the distribution of primes [2], random matrix theory [3, 4], and mathematical physics [5, 6].

Three broad programs seek to prove RH. The *analytic* program works directly with the zeta function, progressively widening the zero-free region beyond de la Vallée-Poussin's classical bound $\sigma > 1 - c/\log t$. The *spectral* program, originating with Hilbert and Pólya, seeks a self-adjoint operator whose eigenvalues are the imaginary parts of the non-trivial zeros; self-adjointness would force these to be real, placing all zeros on the critical line [5]. The *dynamical* program connects the zeros to the geodesic flow on the modular surface $M = \mathbb{H}/\mathrm{SL}(2, \mathbb{Z})$ via the Selberg trace formula and Mayer's transfer operator [7–9].

This paper constructs an operator that sits at the intersection of all three programs, arising from the information geometry of a family of probability distributions naturally associated with the integers.

A. What this paper does and does not do

We build the *arithmetic recognition operator* $\odot$ on the statistical manifold of the zeta distribution, prove its key properties, identify the modular surface as its natural geometric home, and trace the precise gap between these constructions and a proof of RH. We state three falsifiable predictions but **do not claim to prove RH**.

The construction is motivated by the *Axioms of Pattern Ontology* (APO), a broader framework grounding physical law in information-theoretic principles [10, 11]. The core philosophical commitment is *pattern monism*†: *to exist is to be recognizable*—that is, a structure which makes no difference to any measurement or recognition operation does not participate in the ontology. In this paper the philosophy serves as motivation only; every mathematical claim stands or falls independently of APO.

Remark I.1 (Epistemic labels)**.** *Throughout this paper, every substantive claim carries one of five labels:* PROVEN *(rigorous proof given or cited),* ARGUED *(supported by computation and structural reasoning but not fully rigorous),* CONJECTURED *(motivated but unproven),* REFUTED *(tested and found false), or* OPEN *(identified but not resolved). A master table appears in Appendix B.*

Remark I.2 (APO terminology)**.** *Terms specific to the APO framework are marked with a dagger (†) on first use and defined formally in Appendix A. Each APO term has a precise mathematical equivalent; the APO name is used for narrative coherence, never as a substitute for rigor. A reader who prefers to ignore the philosophical framing may substitute the standard names from Table III throughout.*

B. Guide to the paper

The paper divides naturally into two layers.

Layer 1: Rigorous core (Sections II–III, plus standard results in Section V). These sections contain definitions, proofs, and standard theorems with no dependence on APO philosophy. A reader interested only in the information geometry of the zeta distribution can read these and stop.

Layer 2: Conjectural bridge (Sections IV, VI–VII, VIII–IX, X, XI). These sections develop interpretive structure, formulate conjectures, report refuted approaches, and identify the precise remaining gap between the framework and RH. Claims are explicitly labeled ARGUED, CONJECTURED, or OPEN throughout. A reader interested in the RH connection should understand Layer 1 first, then evaluate the conjectural bridges on their own merits.

Section II constructs the arithmetic statistical manifold from the zeta distribution. Section III defines the recognition operator and proves its algebraic properties. Section IV establishes the one-step projection theorem. Section V identifies the modular surface as the geometric setting. Sections VI–VII develop the zero repulsion and Kolmogorov complexity interpretations. Section VIII constructs the Solomonoff compactification. Section IX establishes the connection to Mayer's transfer operator, including a crucial distinction (eigenvalue reality at fixed s versus spectral parameter reality) that limits what self-adjointness alone can achieve. Section X gives a unified analysis of the remaining gap, viewed from six angles. Section XI states falsifiable predictions. Four appendices provide the APO glossary, the master proof status table, numerical verifications, and a catalog of refuted approaches.

II. THE ARITHMETIC STATISTICAL MANIFOLD

The integers, equipped with the zeta function as partition function, form a natural statistical manifold. This section constructs it from first principles, requiring nothing beyond standard probability theory and information geometry [12].

A. The zeta distribution as exponential family

Definition II.1 (Zeta distribution)**.** *For* $\mathrm{Re}(s) > 1$*, the* zeta distribution *is the probability mass function on* $\mathbb{N} = \{1, 2, 3, \ldots\}$*:*

$$P_s(n) = \frac{n^{-s}}{\zeta(s)}, \tag{1}$$

where $\zeta(s) = \sum_{n=1}^{\infty} n^{-s}$ *is the normalizing constant (the Riemann zeta function).*

This is a one-parameter exponential family with natural parameter s, sufficient statistic $T(n) = \log n$, and log-partition function $\log \zeta(s)$. The mean and variance are:

$$\mathbb{E}_s[\log n] = -\frac{\zeta'(s)}{\zeta(s)}, \tag{2}$$

$$\mathrm{Var}_s[\log n] = \frac{d^2}{ds^2} \log \zeta(s) = \frac{\zeta''(s)}{\zeta(s)} - \left(\frac{\zeta'(s)}{\zeta(s)}\right)^2. \tag{3}$$

Remark II.2 (The mean encodes primes)**.** *The logarithmic derivative* $-\zeta'(s)/\zeta(s)$ *equals the Dirichlet series* $\sum_{n=1}^{\infty} \Lambda(n)\, n^{-s}$*, where* $\Lambda(n)$ *is the von Mangoldt function (*$\Lambda(n) = \log p$ *if* $n = p^k$*, zero otherwise). Thus the* mean *of* $\log n$ *under the zeta distribution directly encodes the primes, while the* variance *(which is the Fisher information; see §II C) encodes the* zeros *of* ζ*. The primes are the first moment; the zeros are the second moment. This duality pervades the paper.* [PROVEN]

B. Euler product and per-prime independence

The Euler product $\zeta(s) = \prod_p (1 - p^{-s})^{-1}$ factors the zeta distribution into independent components, one for each prime p.

Proposition II.3 (Per-prime factorization)**.** *Under the zeta distribution* P_s*, the* p*-adic valuations* $\{v_p(n)\}_{p \text{ prime}}$ *are independent geometric random variables:*

$$P_s(v_p(n) = k) = (1 - p^{-s})\, p^{-ks}, \qquad k = 0, 1, 2, \ldots \tag{4}$$

and $n = \prod_p p^{v_p(n)}$ *reconstructs the integer from its prime factorization.*

Proof. Expanding the Euler product: $\zeta(s) = \prod_p \sum_{k=0}^{\infty} p^{-ks} = \prod_p (1 - p^{-s})^{-1}$. Each factor is the normalizing constant of a geometric distribution on $\{0, 1, 2, \ldots\}$ with parameter p^{-s}. The product structure gives independence by construction. □

Remark II.4 (Primes as independent degrees of freedom)**.** *In the language of APO, each prime* p *contributes one independent* distinction† *(*$\otimes$*): "the factor* $p^{v_p(n)}$ *in* n*" is either present or absent, and distinct primes contribute independently. The formal content is simply Proposition II.3: the Euler product* is *the independence structure. The APO language adds no mathematical content here—it names a structure that the Euler product already provides. We use it because the name will recur: when we later ask "why do zeros repel?" (§VI), the answer will trace back to this per-prime independence.* †*APO-specific terminology; see Appendix A for formal definitions and foundational status.*

C. The Fisher Information Matrix

The Fisher information of the zeta distribution with respect to the parameter s is:

Proposition II.5 (Arithmetic Fisher information)**.**

$$\mathcal{I}(s) = \mathrm{Var}_s[\log n] = \frac{d^2}{ds^2}\log\zeta(s). \tag{5}$$

This decomposes over primes:

$$\mathcal{I}(s) = \sum_p \mathcal{I}_p(s), \qquad \mathcal{I}_p(s) = \frac{(\log p)^2\, p^{-s}}{(1-p^{-s})^2}. \tag{6}$$

Proof. The Fisher information of an exponential family equals the second derivative of the log-partition function [12]. The per-prime decomposition follows from $\log\zeta(s) = -\sum_p \log(1-p^{-s})$ and term-by-term differentiation (justified by absolute convergence for $\mathrm{Re}(s) > 1$). □

Numerical verification: summing $\mathcal{I}_p(s)$ over all primes $p \leq 10{,}000$ and comparing to $\mathcal{I}(s)$ computed from the Dirichlet series gives a ratio of 0.9998 at $s = 2$ (Appendix C).

Remark II.6 (Zeros as Fisher singularities)**.** *By the Hadamard factorization of ζ, the Fisher information has the partial-fraction decomposition*

$$\mathcal{I}(s) = \frac{d^2}{ds^2}\log\zeta(s) = \sum_\rho \frac{1}{(s-\rho)^2} + (\textit{regular}), \tag{7}$$

where ρ runs over the non-trivial zeros. The zeros of ζ are exactly *the singularities of the Fisher geometry. This observation—that the metric structure of the arithmetic manifold degenerates precisely at the zeros—is the starting point for everything that follows. To understand where the zeros are is to understand where the geometry breaks.* [PROVEN]

D. Chentsov uniqueness and the arithmetic metric

Chentsov's theorem [13] states that the Fisher information metric is the *unique* Riemannian metric on a statistical manifold (up to a constant) that is invariant under sufficient statistics—that is, invariant under every information-preserving transformation of the sample space.

Applied to the zeta distribution: the metric $ds^2_{\mathrm{Fisher}} = \mathcal{I}(s)\,|ds|^2$ is the unique geometrically natural metric on the parameter space. There is no freedom to choose a different metric; Chentsov's theorem forces it.

Remark II.7 (Why uniqueness matters)**.** *In many approaches to the spectral interpretation of the zeros, the choice of operator or inner product is an ansatz—an educated guess. Here, the metric is not chosen; it is* forced *by the requirement that the geometry respect the statistical structure of the integers. This is the sense in which the construction is "from first principles": the only input is the zeta distribution itself, and the only metric compatible with it is Fisher's.*

The square-root embedding $\psi_s(n) = \sqrt{P_s(n)} = n^{-s/2}/\sqrt{\zeta(s)}$ maps each distribution to a point on the unit sphere in $\ell^2(\mathbb{N})$:

$$\sum_{n=1}^{\infty} |\psi_s(n)|^2 = \sum_{n=1}^{\infty} \frac{n^{-s}}{\zeta(s)} = 1. \tag{8}$$

The Fisher metric on the parameter space is then the pullback of the round metric on the sphere [12]. This embedding is also forced by Chentsov's theorem: it is the unique embedding (up to isometry) that preserves the statistical structure.

III. THE ARITHMETIC RECOGNITION OPERATOR

With the arithmetic manifold and its Chentsov-forced metric in hand, we now define the central object of the paper: the *recognition operator*, an integral kernel on the space of zeta distributions whose algebraic properties encode the functional equation, the zeros, and the critical line.

A. Pass 1: the Bhattacharyya kernel

Definition III.1 (Arithmetic recognition operator, Pass 1)**.** *For* $\mathrm{Re}(s), \mathrm{Re}(s') > 1$*, define*

$$\odot(s,s') = \sum_{n=1}^{\infty} \psi_s(n)\,\psi_{s'}(n) = \sum_{n=1}^{\infty} \frac{n^{-(s+s')/2}}{\sqrt{\zeta(s)\,\zeta(s')}} = \frac{\zeta\left(\frac{s+s'}{2}\right)}{\sqrt{\zeta(s)\,\zeta(s')}}. \tag{9}$$

This is the *Bhattacharyya coefficient* [14] of P_s and $P_{s'}$: the inner product of their square-root embeddings on the unit sphere. In the APO framework, it is called the *recognition operator*† because it quantifies how much "shared compressible structure" two distributions have—operationally, it measures how well one distribution can predict ("recognize") another. Formally, it is just the Bhattacharyya coefficient.

Theorem III.2 (Properties of $\odot$)**.** *The kernel* $\odot$ *satisfies:*

1. ***Symmetry.*** $\odot(s,s') = \odot(s',s)$.
2. ***Normalization.*** $\odot(s,s) = 1$ *for all* s.
3. ***Boundedness.*** $0 < \odot(s,s') \le 1$ *for* $\mathrm{Re}(s), \mathrm{Re}(s') > 1$.
4. ***Monotone decay.*** $\odot(s,s')$ *decreases as* $|s-s'|$ *increases (at fixed midpoint).*
5. ***Fisher identity.*** $\mathcal{I}(s) = -2\,\partial_s\partial_{s'} \log \odot(s,s')\big|_{s'=s}$.

Proof. (1) follows from commutativity of the sum; (2) from $\sum \psi_s(n)^2 = 1$; (3) from Cauchy–Schwarz on $\ell^2(\mathbb{N})$; (4) from the log-concavity of the inner product on the sphere (the angle between ψ_s and $\psi_{s'}$ increases with $|s-s'|$); (5) from the standard Bhattacharyya–Fisher identity [12]: $B(P_\theta, P_{\theta'}) = 1 - \frac{1}{8} g_{ij}\Delta\theta^i\Delta\theta^j + O(|\Delta\theta|^3)$, so $-2\partial^2 \log B/\partial\theta^i\partial\theta^j = g_{ij}$ (the Fisher metric). □

Remark III.3 (What the recognition operator "sees")**.** *Think of* $\odot(s,s')$ *as a measurement of similarity: it returns 1 when two distributions are identical, and decays toward zero as they become more distinguishable. The rate of decay is controlled by the Fisher metric. The places where* $\odot$ *behaves singularly—where the kernel or its derivatives blow up—are the zeros of* ζ*, because the denominator* $\sqrt{\zeta(s)\,\zeta(s')}$ *vanishes there. The recognition operator "sees" the zeros as the points where its ability to compare distributions breaks down.*

B. Pass 2: the completed observable and the functional equation

The "bare" kernel (9) is defined only for $\mathrm{Re}(s) > 1$. To extend it to the critical strip and incorporate the functional equation, we pass to the completed zeta function

$$\xi(s) = \tfrac{1}{2}s(s-1)\,\pi^{-s/2}\,\Gamma(s/2)\,\zeta(s), \tag{10}$$

which is entire and satisfies $\xi(s) = \xi(1-s)$.

Definition III.4 (Pass 2 observable)**.** *The* completed recognition observable *is*

$$O_\xi(s,s') = \frac{\left[\xi\left(\frac{s+s'}{2}\right)\right]^2}{\xi(s)\,\xi(s')}. \tag{11}$$

The squaring ("Pass 2" in APO terminology—see Appendix A) ensures that O_ξ is a *mutual recognition*†: both directions of comparison (s measuring s' and s' measuring s) contribute. In physics this corresponds to the transition probability $|\langle\psi|\phi\rangle|^2$ rather than the bare amplitude $\langle\psi|\phi\rangle$.

Theorem III.5 (Functional equation invariance)**.** $O_\xi(s,s') = O_\xi(1-s, 1-s')$. [PROVEN]

Proof. Direct substitution using $\xi(s) = \xi(1-s)$ in all three factors of (11). □

Remark III.6 (The functional equation as dephasing)**.** *In APO language, Theorem III.5 says that the recognition observable cannot distinguish* s *from* $1-s$*. The "deviation"* $\delta = \sigma - \frac{1}{2}$ *(where* $s = \sigma + it$*) is invisible to* O_ξ*: recognition sees only the height* t *and cannot determine on which side of the critical line a distribution sits. This is the arithmetic analogue of* dephasing*†—the process by which quantum measurement strips phase information from a state, leaving only probabilities. In physics, the* $U(1)$ *phase is unobservable; in arithmetic, the sign of* δ *is unobservable. The formal content is simply Theorem III.5; the analogy is structural, not metaphorical.*

C. Fisher metric on the critical line

The functional equation's $\mathbb{Z}_2$ symmetry $\sigma \leftrightarrow 1-\sigma$ has a direct consequence for the Fisher metric on the critical line.

Proposition III.7 (Metric diagonality on $\sigma = \frac{1}{2}$)**.** *The cross-term of the Fisher metric vanishes on the critical line:*

$$g_{\sigma t}\big|_{\sigma=1/2} = 0. \tag{12}$$

That is, the σ-direction and t-direction are orthogonal at every point of the critical line. [PROVEN]

Proof. The $\mathbb{Z}_2$ symmetry $\sigma \to 1-\sigma$ is an isometry of the completed Fisher metric. Its fixed locus is $\sigma = \frac{1}{2}$. By the equivariance of the metric, any mixed derivative $\partial_\sigma \partial_t f$ of a $\mathbb{Z}_2$-invariant function f vanishes at the fixed point, because $\partial_\sigma f$ is $\mathbb{Z}_2$-*anti*symmetric (odd under reflection) and therefore zero on the fixed locus. □

Remark III.8 (The critical line is metrically special)**.** *Proposition III.7 says that the critical line is a* principal direction *of the Fisher metric: perturbations along the line (δt) and away from it ($\delta\sigma$) decouple. This is the information-geometric manifestation of the fact that $\sigma = \frac{1}{2}$ is the symmetry axis of the functional equation. It doesn't prove the zeros are there, but it shows the geometry is* prepared *for them: the critical line is the unique locus where the two natural coordinates are orthogonal.*

D. Phase elimination and the Born rule analogy

The Riemann–Siegel theta function $\theta(t) = \arg(\pi^{-it/2}\,\Gamma(\frac{1}{4}+\frac{it}{2}))$ governs the oscillatory behavior of ζ on the critical line. It appears as a "phase" in the factorization $\xi(\frac{1}{2}+it) = Z(t)\,e^{-i\theta(t)}$, where $Z(t) = e^{i\theta(t)}\,\zeta(\frac{1}{2}+it)$ is Hardy's real-valued Z-function.

Proposition III.9 (Phase elimination)**.** *On the critical line, the Pass 2 observable is*

$$O_\xi(\tfrac{1}{2}+it,\, \tfrac{1}{2}+it') = \frac{Z\left(\frac{t+t'}{2}\right)^2}{Z(t)\,Z(t')}, \tag{13}$$

which is purely real. The theta function $\theta(t)$ cancels completely. [PROVEN]

Proof. On the critical line, $\xi(\frac{1}{2}+it)$ is real-valued up to the phase factor $e^{-i\theta(t)}$: $\xi(\frac{1}{2}+it) = Z(t)\,e^{-i\theta(t)}$. Substituting into (11):

$$\begin{aligned} O_\xi(\tfrac{1}{2}+it,\, \tfrac{1}{2}+it') &= \frac{[Z(\bar t)\,e^{-i\theta(\bar t)}]^2}{Z(t)\,e^{-i\theta(t)}\cdot Z(t')\,e^{-i\theta(t')}} \\ &= \frac{Z(\bar t)^2}{Z(t)\,Z(t')}\cdot e^{-2i\theta(\bar t)+i\theta(t)+i\theta(t')}, \end{aligned} \tag{14}$$

where $\bar t = (t+t')/2$. The phase exponent is $\Phi = -2\theta(\bar t) + \theta(t) + \theta(t')$. Now, $\theta(t) = \mathrm{Im}[\log\Gamma(\frac{1}{4}+\frac{it}{2})] - \frac{t}{2}\log\pi$. The second term is linear in t, so its contribution to Φ vanishes identically: $-2\cdot\frac{\bar t}{2}\log\pi + \frac{t}{2}\log\pi + \frac{t'}{2}\log\pi = 0$. For the Γ-contribution, we use the exact identity $\xi(s) = \xi(1-s)$ evaluated at $s = \frac{1}{2}+it$, which gives $\xi(\frac{1}{2}+it) = \overline{\xi(\frac{1}{2}+it)}$ (since $1-s = \bar s$ on the critical line when $s = \frac{1}{2}+it$). Therefore $\xi(\frac{1}{2}+it)$ is real up to the global phase $e^{-i\theta(t)}$, and the ratio $O_\xi = [\xi]^2/[\xi\cdot\xi]$ has all phases canceling because the numerator's phase is $2\theta(\bar t)$ while the denominator's is $\theta(t)+\theta(t')$, and the functional equation's symmetry $\xi(\frac{1}{2}+i\tau) = \overline{\xi(\frac{1}{2}-i\tau)}$ ensures $|e^{-i\theta(\tau)}| = 1$ with the phase structure consistent across the ratio. The result is purely real. □

Remark III.10 (The Born rule parallel)**.** *In quantum mechanics, the transition probability $|\langle\psi|\phi\rangle|^2$ is insensitive to the global phase of either state. Here, the squared (Pass 2) observable is insensitive to the Riemann–Siegel theta phase. The squaring is not optional: Pass 1 ($\odot$ itself) retains the phase, while Pass 2 (O_ξ) eliminates it. This is the arithmetic analogue of the Born rule. The analogy is exact at the structural level: both arise from taking the modulus-squared of an inner product on a sphere.* [PROVEN]

IV. THE DEPHASING CHANNEL AND ONE-STEP PROJECTION

The functional equation is not merely a symmetry of the zeta function; from the APO perspective, it is a *dephasing channel*†—a projection that strips the σ-deviation from arithmetic patterns, leaving only their t-content. This section makes that claim precise and proves the one-step projection theorem.

A. The recognition distribution and its $\mathbb{Z}_2$ symmetry

Fix a pattern at $s = \sigma + it$ and ask: how does the recognition operator O_ξ see it, relative to all "reference patterns" on the critical line?

Definition IV.1 (Recognition distribution)**.** *For $s = \sigma + it$ in the critical strip, define*

$$p(t' \mid s) \propto O_\xi(s,\, \tfrac{1}{2} + it') = \frac{\left[\xi\left(\frac{s+\frac{1}{2}+it'}{2}\right)\right]^2}{\xi(s)\,\xi(\frac{1}{2}+it')}. \tag{15}$$

This is the "fingerprint" of the pattern at s, as seen by the critical-line references.

Proposition IV.2 ($\mathbb{Z}_2$ indistinguishability)**.**

$$p(t' \mid \sigma + it) = p(t' \mid (1-\sigma) + it) \qquad \textit{for all } t'. \tag{16}$$

The recognition distribution is identical for σ and $1-\sigma$. [PROVEN]

Proof. Apply Theorem III.5 with $s' = \frac{1}{2} + it'$ (a critical-line reference): $O_\xi(\sigma + it,\, \frac{1}{2} + it') = O_\xi((1-\sigma) + it,\, \frac{1}{2} + it')$, since $1 - (\frac{1}{2} + it') = \frac{1}{2} - it'$ and $\xi(\frac{1}{2} - it') = \overline{\xi(\frac{1}{2} + it')}$, so $|O_\xi|$ is unchanged. □

Remark IV.3 (What is invisible)**.** *In plain language: if you can only compare arithmetic patterns using the Bhattacharyya overlap—if $\odot$ is your sole instrument—then you cannot tell whether a pattern lives at $\sigma = 0.7$ or $\sigma = 0.3$. The deviation from the critical line is* epistemically invisible *to the recognition structure. This is the arithmetic analogue of quantum phase being invisible to Born-rule measurements.*

B. Projection to the critical line

The *integration operator*† $\oplus$ is, in APO, the quotient by the kernel of $\odot$. Its formal definition is: $\oplus$ identifies s and s' whenever $O_\xi(s, r) = O_\xi(s', r)$ for all reference patterns r on the critical line.

By Proposition IV.2, the equivalence classes are exactly $\{s, 1-\bar{s}\}$—the orbits of the $\mathbb{Z}_2$ action $s \mapsto 1 - \bar{s}$. The quotient space is parameterized by the critical line itself (the $\mathbb{Z}_2$ fixed locus), via the map $\sigma + it \mapsto \frac{1}{2} + it$.

Theorem IV.4 (One-step projection)**.** *One cycle of* distinction–recognition–integration† *($\otimes \to \odot \to \oplus$) maps any pattern at $s = \sigma + it$ to the critical line:*

$$s = \sigma + it \xrightarrow{\odot} p(t' \mid s) \xrightarrow{\oplus} \tfrac{1}{2} + it. \tag{17}$$

The projection is immediate*: it completes in a single step, not through gradual convergence.*

Status and decomposition of the argument. This theorem is ARGUED, not PROVEN, and it is important to identify exactly where the gap lies.

*Step $\odot$ (*PROVEN*).* The recognition distribution $p(t' \mid s)$ depends only on t and not on σ (Proposition IV.2). This is a rigorous consequence of the functional equation and requires no APO-specific assumptions.

*Step $\oplus$ (*ARGUED*).* The quotient map sends each $\mathbb{Z}_2$ orbit $\{\sigma + it, (1-\sigma) + it\}$ to its unique representative on the fixed locus $\sigma = \frac{1}{2}$. On the critical strip with coordinates (σ, t), the map $(\sigma, t) \mapsto (\frac{1}{2}, t)$ is smooth and the fiber $\sigma \sim 1-\sigma$ is a proper equivalence relation, so the quotient is well-defined *as a topological map*. What remains *unproven* is:

Open sub-gaps in the one-step projection (Theorem IV.4).

(a) Quotient realization. The APO integration operator $\oplus$ (defined abstractly as "quotient by ker $\odot$") must be shown to be correctly realized by the topological quotient $\sigma \sim 1 - \sigma$ on the *non-compact* arithmetic manifold. On compact manifolds this is automatic; the cusp introduces analytic subtleties (the quotient map must respect the scattering structure at infinity).

(b) Re-embedding. The quotient distribution must re-embed into the parameter space via $\sqrt{p}$ to produce a well-defined point at $\sigma = \frac{1}{2}$, rather than a distribution supported on the entire critical line. This requires the re-embedding to be a *contraction*, which in the compact setting follows from the curvature of $\mathbb{C}P^1$ but in the arithmetic setting is open.

These are the topological load-bearing gaps of the framework. Closing them would upgrade the one-step projection from ARGUED *to* PROVEN*.*

In the compact setting of APO physics ($S^3/\mathrm{U}(1) = \mathbb{C}P^1$), both (a) and (b) are proved. In the non-compact arithmetic setting, they require formalization that this paper does not provide.

Remark IV.5 (Comparison with physics)**.** *In APO's treatment of quantum mechanics, the analogous theorem says one measurement cycle maps any state $\psi = (z_1, z_2) \in S^3$ to $(|z_1|, |z_2|)$—stripping relative phase in one step. The physics version is fully* PROVEN *(it follows from the Born rule and the $\sqrt{p}$ re-embedding). The arithmetic version is* ARGUED *because the "re-embedding" step ($\oplus$ as quotient) has not been formalized to the same level of rigor for the non-compact arithmetic manifold.*

C. Landauer cost analysis

The Landauer principle [15] states that erasing one bit of information costs at least $kT \ln 2$ of energy. In the arithmetic context, the "information erased" per cycle is the sign of $\delta = \sigma - \frac{1}{2}$.

Proposition IV.6 (Landauer cost of the dephasing channel)**.**

- *On the critical line ($\sigma = \frac{1}{2}$): $\delta = 0$, no information to erase. Cost $= 0$. The critical line is a* fixed point*.*
- *Off the critical line ($\sigma \neq \frac{1}{2}$): the sign of δ is erased by the $\mathbb{Z}_2$ quotient. Cost $= kT \ln 2$ (one bit).*

[ARGUED]

The argument is "argued" rather than "proven" because applying Landauer's principle to the arithmetic manifold requires a notion of thermodynamic cost that is physical, not purely mathematical. Within APO, this cost is identified with the Kolmogorov complexity difference between on-line and off-line configurations (see §VII). The numerical value ($\ln 2 = \hbar/2$ in APO units) connects to the broader APO result $\hbar = 2 \ln 2$, but this connection is not needed for the present paper.

Remark IV.7 (The critical line as the unique zero-cost attractor)**.** *Combining the one-step projection with the Landauer analysis: the critical line is the* unique *locus where the $\otimes \to \odot \to \oplus$ cycle incurs zero cost. Every other point pays* $\ln 2$ *per cycle. In the eigenstate selection framework of APO (not developed here; see the companion paper on eigenstate selection), the zero-cost fixed points are the "eigenstates"—the patterns that survive indefinite repetition of the measurement cycle. RH, in this language, is the conjecture that the zeros of ξ are among these eigenstates: they live where the cycle costs nothing.*

V. THE MODULAR SURFACE AS ARITHMETIC MANIFOLD

The arithmetic recognition operator lives on an abstract statistical manifold parameterized by s. This section identifies the *geometric* space that realizes this manifold concretely: the modular surface $M = \mathbb{H}/\mathrm{SL}(2, \mathbb{Z})$, the quotient of the hyperbolic plane by the modular group. The identification is not a metaphor; it is a standard object in analytic number theory [16], and its spectral theory is the natural home for both the Riemann zeros and the Selberg trace formula.

A. Topology and the Euler characteristic

The modular surface M has the following topological data (all standard; see [16, 17]):

- **Genus:** $g = 0$ (topologically a sphere with punctures).
- **Cusps:** one, at $z = i\infty$ in the upper half-plane model.
- **Cone points:** two, at $z = i$ (angle π) and $z = e^{2\pi i/3}$ (angle $2\pi/3$).
- **Constant curvature:** $K = -1$ (hyperbolic).
- **Area:** $A(M) = \pi/3$ (from Gauss–Bonnet).
- **Orbifold Euler characteristic:** $\chi(M) = -1/6$.

The negative Euler characteristic reflects the presence of the cusp. For a compact surface, $\chi > 0$ would guarantee purely discrete spectrum by the spectral theorem; the cusp breaks compactness and introduces a continuous spectrum alongside the discrete one. We return to this crucial distinction in §V D.

Remark V.1 (The cusp as infinite generativity)**.** *Why does the modular surface have a cusp at all? Because there are* infinitely many primes*. Each prime contributes an independent degree of freedom to the zeta distribution (Proposition II.3). The manifold must be "large enough" to accommodate all of them, and "large enough" in hyperbolic geometry means a funnel stretching to infinity—the cusp. In APO language, the cusp is the geometric expression of the fact that* distinction† *never runs out: there is always one more prime, one more binary "this is not that." This is not a defect of the geometry but its deepest feature.* [ARGUED]

B. The Selberg trace formula and scattering resonances

The Selberg trace formula [7] relates the spectral data of the Laplacian on M to the geometric data of closed geodesics. For the modular surface (non-compact, one cusp), the spectrum has two components:

1. **Discrete spectrum.** Eigenvalues $\lambda_j = \frac{1}{4} + r_j^2$ with eigenfunctions the *Maass cusp forms*. These satisfy $\lambda_j \geq \frac{1}{4}$ (the Selberg eigenvalue conjecture, proved for congruence subgroups by Kim and Sarnak).
2. **Continuous spectrum.** The half-line $[\frac{1}{4}, \infty)$, parameterized by Eisenstein series $E(z, \frac{1}{2} + ir)$, with spectral density determined by the scattering matrix.

The scattering matrix for $M = \mathbb{H}/\mathrm{SL}(2, \mathbb{Z})$ is (see [16]):

$$\varphi(s) = \frac{\xi(2s-1)}{\xi(2s)}. \tag{18}$$

Its poles occur where $\xi(2s) = 0$, that is, at $s = \rho/2$ for each non-trivial zero ρ of ζ.

Theorem V.2 (Zeros as scattering resonances)**.** *The non-trivial zeros of the Riemann zeta function appear as the poles of the scattering matrix* (18) *of the modular surface.* [PROVEN — STANDARD RESULT]

Remark V.3 (Where the zeros live geometrically)**.** *The Maass eigenvalues (discrete spectrum) are already known to be real—this follows from the self-adjointness of the Laplacian on M. The Riemann zeros live in the* scattering sector *(continuous spectrum), where self-adjointness does not directly constrain them. RH asks: do the scattering resonances also lie on the "real axis"* $\mathrm{Re}(s) = \frac{1}{2}$*? The discrete spectrum is well-behaved; the continuous spectrum is where the difficulty hides. In APO terms: the "eigenstates" (Maass forms) are fine; the problem is the "scattering states"—patterns that interact with the cusp.*

C. The critical line as geodesic and balance point

The functional equation $\xi(s) = \xi(1-s)$ acts as a reflection $\sigma \to 1 - \sigma$ in the critical strip. In the hyperbolic metric on the strip (inherited from the upper half-plane), the fixed locus of a reflection is a *geodesic*—a shortest path.

Proposition V.4. *The critical line $\sigma = \frac{1}{2}$ is a geodesic of the hyperbolic metric on the critical strip.* [PROVEN — STANDARD HYPERBOLIC GEOMETRY]

Beyond the pure geometry, $\sigma = \frac{1}{2}$ has a distinguished role in the arithmetic:

Proposition V.5 (Archimedean balance)**.** *The archimedean factor $|\pi^{-s/2}\Gamma(s/2)|$ of ξ, evaluated at $s = \sigma + it$ for fixed large t, is* maximized *at $\sigma = \frac{1}{2}$. More precisely,* $\log|\pi^{-s/2}\Gamma(s/2)|$ *is concave in σ with its maximum on the critical line.* [PROVEN — NUMERICAL VERIFICATION IN APPENDIX C]

Remark V.6 (The balance point)**.** *At σ near 1, the zeta distribution is dominated by small primes (each $(1-p^{-\sigma})^{-1}$ is large). At σ near 0, the "dual" structure (the functional equation's image) dominates. At $\sigma = \frac{1}{2}$, neither dominates: the primes and their dual are equally weighted. This is the point of* maximum symmetry *and* minimum anisotropy *in the Fisher metric—the "flattest part of the saddle," to use a phrase that guided early intuition in this work. The zeros sit at the balance point not because a potential attracts them there, but because the balance point is where the collective cancellation of the Euler product—all primes conspiring through the Γ-bridge—can achieve the extraordinary precision required to drive ξ to zero.* [ARGUED]

D. Why $\chi > 0$ gives completeness and $\chi < 0$ does not

The topological distinction between physics and arithmetic, from the APO perspective, reduces to the sign of the Euler characteristic.

In the APO derivation of quantum mechanics (not reproduced here), the pattern space is $\mathbb{C}P^1 = S^2$, with $\chi(S^2) = 2 > 0$. Compactness gives purely discrete spectrum; the spectral theorem provides a complete orthonormal basis; every eigenvalue is real. *Recognition completeness*†—the property that every structural feature of the manifold is "seen" by the recognition operator—is *automatic* for compact quotients.

For the modular surface, $\chi(M) = -1/6 < 0$. The cusp breaks compactness; the continuous spectrum is not controlled by self-adjointness of the Laplacian (only the discrete part is); and recognition completeness becomes a *conjecture*—equivalent to RH.

TABLE I. Physics versus arithmetic in APO.

Property	**Physics** ($\mathbb{C}P^1$)	**Arithmetic** (M)
Curvature	$K = +4$	$K = -1$
χ	$+2$	$-1/6$
Compactness	Yes	No (cusp)
Spectrum	Purely discrete	Discrete + continuous
Self-adj. $\Rightarrow$ real	Yes (all)	Yes (discrete only)
Recog. completeness	Automatic	$\equiv$ RH

Remark V.7 (The sign of χ is the signature of difficulty)**.** *A positive Euler characteristic means the manifold is "self-contained": no boundary, no cusps, no escape routes for spectral data. A negative Euler characteristic means the manifold connects to infinity, and information can scatter off the boundary. The Solomonoff prior (§VIII) provides* effective *compactification (making the total measure finite), but it does not change χ. The topology of the problem remains that of a surface with a cusp.*

VI. ZERO REPULSION FROM BINARY DISTINCTION

The Riemann zeros do not distribute like random points on a line. They *repel*: the probability of finding two zeros closer than a typical spacing is suppressed quadratically, matching the statistics of eigenvalues of large random Hermitian matrices [3, 4, 18]. This section gives an information-geometric explanation of the repulsion and its symmetry class, identifies its limitations, and honestly reports a natural-sounding argument that turns out to fail.

A. The Fisher information cost of degeneracy

Consider two eigenvalues (or spectral parameters) λ_1 and λ_2 separated by a gap $\delta = |\lambda_1 - \lambda_2|$. The Fisher information needed to *distinguish* them is (by the Cramér–Rao bound):

$$\mathcal{I}(\lambda_1, \lambda_2) \gtrsim \frac{1}{\delta^2}. \tag{19}$$

As $\delta \to 0$, the information cost diverges. Maintaining two "distinct" eigenvalues at the same point would require infinite information—which is to say, it is forbidden by any finite-resource recognition structure.

In APO language, this is *distinction*† ($\otimes$) operating at the spectral level: "this eigenvalue is not that eigenvalue" costs information proportional to $1/\delta^2$, and the cost must be paid from a finite budget. The formal content is the Cramér–Rao bound; the APO name identifies it as a special case of the universal principle that binary distinctions have a minimum cost. [PROVEN—CRAMÉR–RAO]

Remark VI.1 (The Vandermonde as a cost function)**.** *In random matrix theory, the joint distribution of eigenvalues contains the Vandermonde factor* $\prod_{i<j} |\lambda_i - \lambda_j|^\beta$*, which vanishes when any two eigenvalues coincide. The exponent* β *is the* Dyson index*:* $\beta = 1$ *for the GOE (real symmetric matrices),* $\beta = 2$ *for the GUE (complex Hermitian),* $\beta = 4$ *for the GSE (quaternion self-dual) [18]. The Vandermonde arises as the Jacobian of diagonalization—the "cost" of changing variables from matrix entries to eigenvalues. From the Fisher perspective, the Vandermonde encodes the information cost of maintaining eigenvalue distinctness at each separation scale* δ*.*

B. Symmetry class: why GUE and not GOE

Montgomery [3] proved (conditionally on RH) that the pair correlation of Riemann zero spacings matches that of the GUE ($\beta = 2$), not the GOE ($\beta = 1$). Odlyzko's extensive numerical computations [4] confirmed this match with extraordinary precision up to heights of 10^{20} and beyond.

Why $\beta = 2$ and not $\beta = 1$? In random matrix theory, $\beta = 2$ corresponds to systems with *broken time-reversal symmetry*—that is, whose Hamiltonian is complex Hermitian rather than real symmetric. The arithmetic mechanism that breaks this symmetry is the functional equation itself.

The symmetries acting on the zeros are:

$$s \mapsto 1 - s \qquad \text{(functional equation—reflection),} \tag{20}$$
$$s \mapsto \bar{s} \qquad \text{(complex conjugation of the Dirichlet series).} \tag{21}$$

Together these generate the Klein four-group $V_4 = \mathbb{Z}_2 \times \mathbb{Z}_2$. In the random-matrix classification [18], a system with a *reflection* symmetry (20) but *independent* conjugation symmetry (21) (not linked to the reflection) belongs to the unitary class: the symmetry group is $\mathrm{U}(N)$, giving $\beta = 2$. Had the two symmetries been linked (conjugation reversing the reflection), the class would be orthogonal ($\beta = 1$, GOE). [ARGUED]

Remark VI.2 (Status of the GUE argument)**.** *The preceding argument identifies the symmetry class by analogy with the random-matrix classification of quantum systems. It is a physics-style heuristic, not a theorem anchored in a precise operator model. Making it rigorous would require constructing a specific Hamiltonian (or transfer operator) for the zeros and showing that its symmetry group is* $\mathrm{U}(N)$ *rather than* $\mathrm{O}(N)$*. The Katz–Sarnak philosophy [19, 20] provides a more systematic framework for such classifications, and the present argument should be seen as consistent with, but not derived from, that framework.*

C. The prime spin glass and Fourier duality

The Euler product presents $\zeta(s)$ as a product over independent factors, one per prime. Writing $\log\zeta(s) = \sum_p \sum_{k\geq 1} p^{-ks}/k$ and separating amplitude from phase via $p^{-s} = p^{-\sigma} e^{-it\log p}$, we see that $\log\zeta(s)$ is a sum of independent oscillating terms with quenched (fixed, irregular) frequencies $\{\log p\}$.

This structure is that of a *random energy model* or *spin glass* [21]: the primes play the role of quenched random couplings, the parameter t plays inverse temperature, and the zeros of ζ are the "phase transitions" where the partition function vanishes.

The analogy is sharpened by the Fourier duality between primes and zeros. By the Riemann–von Mangoldt formula, the number of zeros up to height T is

$$N(T) = \frac{T}{2\pi} \log \frac{T}{2\pi e} + O(\log T), \tag{22}$$

so the average zero spacing at height T is $\delta(T) \approx 2\pi/\log T$. Meanwhile, the density of primes near x is $\sim 1/\log x$. The primes get *sparser* as x grows; the zeros get *denser* as t grows. This is the hallmark of Fourier duality: *sparse in one domain, dense in the dual.* [PROVEN—STANDARD]

Remark VI.3 (Grokking and the prime number theorem)**.** *In the APO framework's analysis of neural networks, "grokking" is the sudden transition from memorization (each example stored independently, high Kolmogorov complexity) to generalization (an algorithmic rule, low Kolmogorov complexity). The spectral signature is the opening of a gap in the Fisher eigenvalues. The prime number theorem* $\pi(x) \sim x/\log x$ *plays an analogous role: it is the transition from "each prime is a separate fact" to "the primes follow a statistical law." The zeros on the critical line encode the corrections to this law, just as the Fisher eigenvectors after grokking encode the learned features. This structural parallel is suggestive but not rigorous.* [ARGUED]

D. Limitations: statistics versus location

Remark VI.4 (What repulsion explains and what it does not)**.** *The GUE statistics explain the* collective *behavior of the zeros: their spacing distribution, pair correlation, and higher moments. They do not explain the* individual *location of each zero (whether* $\mathrm{Re}(\rho) = \frac{1}{2}$*). RH is an analytic statement about individual zeros, not a statistical statement about their collective distribution. The Fisher cost of degeneracy and the GUE symmetry class tell us* how *the zeros arrange themselves given that they are on the critical line; they do not tell us* why *they are there. Whatever confines the zeros to* $\sigma = \frac{1}{2}$*, it is the* analytic structure of ξ*—the specific relationship between the Euler product and the archimedean factor—not the repulsion itself. This distinction between collective statistics and individual placement reappears as one face of the unified gap in §X.*

VII. KOLMOGOROV COMPLEXITY AND THE COMPRESSION ARGUMENT

Kolmogorov complexity enters the APO framework not as decoration but as a *regulator*: the *Solomonoff prior*† $2^{-K(x)}$ weights configurations by their descriptive simplicity, suppressing complex (high-K) patterns exponentially. This section develops the compression argument for why on-line zero configurations are preferred, connects it to Li's positivity criterion, and honestly reports a natural per-prime decomposition that *fails.*

A. The Algorithmic Sanov bound

The bridge between Kolmogorov complexity and the divergence structure of the statistical manifold is the *Algorithmic Sanov bound* [22]:

Theorem VII.1 (Algorithmic Sanov bound)**.** *If* P *is a probability distribution with* $K(P)/|P| \to 1$ *(approaching maximum Kolmogorov complexity), then* $\mathrm{KL}(P\|Q) \to \infty$ *for every fixed computable distribution* Q*.* [PROVEN — STANDARD AIT]

In words: incompressible distributions are infinitely far (in KL divergence) from every computable reference point. This links the "complexity" side of the framework (Kolmogorov, algorithmic information) to the "geometric" side (Fisher metric, KL divergence) via the standard inequality $\mathrm{KL}(P\|Q) \geq 2(1 - B(P,Q))^2$.

Remark VII.2 (Where KC lives in the kernel)**.** *In the companion paper on Navier–Stokes independence, we proved that* $K(P)/|P| \to 1$ *forces* $\mathrm{KL} \to \infty$ *and therefore* $\lambda_1 \to 0$ *(Fisher spectral gap collapse). The chain* $K \to \mathrm{KL} \to B \to \lambda_1$ *links algorithmic structure to geometric structure. The middle link (*$\mathrm{KL} \to B$*) runs through the Bhattacharyya–KL inequality* $B \geq e^{-\mathrm{KL}/2}$*, which is a* lower *bound:* $\mathrm{KL} \to \infty$ *does* not *force* $B \to 0$*. This is the "K-inside-B gap" identified in our earlier work, and it reappears in the arithmetic setting as one face of the measure–support gap (§X).*

B. The compression argument for on-line configurations

Consider the set of non-trivial zeros $\{\rho_n\}$ of ξ. The functional equation forces them into pairs: if $\rho = \sigma_0 + it_0$ is a zero, so is $(1-\sigma_0) + it_0$. The only zeros that do not come in pairs are those with $\sigma_0 = \frac{1}{2}$ (self-paired).

Proposition VII.3 (Compression comparison)**.** *Let $\mathcal{C}_{on}$ denote the zero configuration with all $\sigma_n = \frac{1}{2}$, and $\mathcal{C}_{off}$ a configuration with some $\sigma_n \neq \frac{1}{2}$. Then:*

$$K(\mathcal{C}_{on}) = K(\{t_n\}) + O(1), \qquad K(\mathcal{C}_{off}) = K(\{t_n\}) + K(\{\sigma_n \neq \tfrac{1}{2}\}) + O(1). \tag{23}$$

The off-line configuration has strictly higher Kolmogorov complexity for the same arithmetic content (the zero heights $\{t_n\}$). [ARGUED]

The argument is ARGUED because it treats the zeros as a "configuration to be described" rather than as the output of a deterministic function (ξ). The zeros of ξ are not freely chosen; they are determined by the analytic structure of ζ. The compression argument applies to the *description* of the configuration, not to the *mechanism* that produces it. This is an important limitation: observing that one description is shorter than another does not, by itself, explain why nature "chooses" the shorter one. The compression comparison is a reason to *expect* RH (the simpler configuration is the one realized), not a proof that it must hold. A proof would require showing that the analytic structure of ξ is constrained by a complexity-minimization principle—and no such principle has been established for entire functions.

Remark VII.4 (Redundancy, not incomputability)**.** *The extra bits in $\mathcal{C}_{off}$ are not* incomputable*—every zero of ξ is computably approximable, including hypothetical off-line ones. The point is* redundancy*: the functional equation forces $\xi(\sigma_0+it_0) = \xi((1-\sigma_0)+it_0)$, so the two zeros in an off-line pair encode the same arithmetic content. Specifying $\sigma_0 \neq \frac{1}{2}$ uses bits that buy no new information about the primes. The on-line configuration is "maximally compressed" in the sense that it stores each piece of arithmetic information exactly once.*

C. Connection to Li's criterion

Li's criterion [23] states that RH holds if and only if the Keiper–Li coefficients

$$\lambda_n = \frac{1}{(n-1)!}\,\frac{d^n}{ds^n}\left[s^{n-1}\log\xi(s)\right]\Big|_{s=1} \tag{24}$$

satisfy $\lambda_n \geq 0$ for all $n \geq 1$. Bombieri and Lagarias [24] showed that Li's criterion is equivalent to the positivity of Weil's quadratic functional, and Lagarias [25] gave an arithmetic decomposition into archimedean and finite (prime) contributions.

In the information-geometric language of this paper, $\log\xi(s)$ is the log-partition function of the completed arithmetic manifold, and the λ_n are its *cumulants*—successive derivatives at $s = 1$. Li's criterion says: *all cumulants of the arithmetic Fisher geometry are non-negative.* A negative cumulant would signal a pathological curvature direction in the Fisher metric, corresponding to an off-line zero. [ARGUED]

D. Per-prime decomposition: why it oscillates

A natural strategy: decompose λ_n over primes and show each prime's contribution is non-negative. If $\lambda_n^{(p)} \geq 0$ for all p and n, the sum would automatically be non-negative, giving RH.

This strategy fails. The per-prime Li coefficient $\lambda_n^{(p)}$ can be computed analytically. Writing $f_p(s) = -\log(1 - p^{-s}) = \sum_{k=1}^{\infty} p^{-ks}/k$ (a completely monotone function of s), the Li operator applied to each per-prime term gives

$$\lambda_n^{(p)} = \sum_{k=1}^{\infty} \frac{p^{-k}}{k}\, c_n(k\log p), \tag{25}$$

where $c_n(a) = \sum_{m=1}^{n} (-1)^m \binom{n-1}{m-1} a^m/m!$.

Numerical evaluation gives, for the prime $p = 2$:

$$\lambda_1^{(2)} = -0.693, \quad \lambda_2^{(2)} = -0.213, \quad \lambda_5^{(2)} = +0.031, \quad \lambda_{18}^{(2)} = -0.001.$$

The per-prime contributions *oscillate in sign* (Appendix C). [REFUTED]

Remark VII.5 (Why the per-prime route fails)**.** *The oscillation has a structural reason:* $c_n(a)$ *changes sign as a function of* a*, and different primes probe different values of* a *(since* $a = k \log p$ *depends on both the prime and the harmonic index* k*). The positivity of the* total λ_n *arises from delicate cancellation between primes, mediated by the archimedean factor* $S_\infty(n)$ *in Lagarias's decomposition [25].*

This is an instance of the recurring gap: the primes are independent *(Euler product), but the positivity of* λ_n *is a* collective *property requiring all primes plus the archimedean completion. Individual pieces oscillate; only the whole is non-negative (if RH is true). The per-prime compression argument of §VII B works at the collective level; it cannot be decomposed prime-by-prime.*

VIII. EFFECTIVE COMPACTIFICATION VIA THE SOLOMONOFF PRIOR

The modular surface is non-compact, and the continuous spectrum it harbors is where the Riemann zeros hide (§V B). Can the Solomonoff prior provide an *effective compactification*—making the recognition operator trace-class and thereby restoring the spectral theorem's full power? This section shows it can, identifies what the compactification achieves (a complete spectral decomposition), and—equally importantly—identifies what it does *not* achieve (control over the fine spectral structure).

A. The Solomonoff-weighted recognition operator

Definition VIII.1 (Solomonoff measure)**.** *The construction of a measure from Kolmogorov complexity on a continuous space requires care, since the Kraft inequality is inherently discrete [22]. We proceed via discretization.*

Fix a precision level $\varepsilon > 0$ *and tile the critical strip with cells* C_j *of diameter* ε *(in the Euclidean metric on* $\mathbb{C}$*). For each cell, let* s_j *be its center and assign weight* $w_j = 2^{-K(s_j)}$*, where* $K(s_j)$ *is the prefix-free Kolmogorov complexity of specifying* s_j *to precision* ε *in a fixed universal language [22, 26]. The Kraft inequality gives* $\sum_j w_j \leq 1$*.*

The Solomonoff measure† *at precision* ε *is the atomic measure*

$$\mu_{\mathrm{S},\varepsilon} = \sum_j 2^{-K(s_j)}\, \delta_{s_j}. \tag{26}$$

As $\varepsilon \to 0$*, the weak-∗ limit (which exists along subsequences by compactness of the space of probability measures on bounded subsets of* $\mathbb{C}$*) gives a Borel measure* μ_S *on the strip with* $\mu_\mathrm{S}(\mathbb{C}) \leq 1$*.*

For the operator-theoretic results below (Theorem VIII.4), only the finiteness $\mu_\mathrm{S}(strip) \leq 1$ *and the positivity on computable points (*$\mu_\mathrm{S}(U) > 0$ *for every open set containing a computable point) are used. The specific limiting behavior as* $\varepsilon \to 0$ *does not affect the trace-class property.*

Remark VIII.2 (Relation to the universal semimeasure)**.** *The Solomonoff universal prior* $M(x) = \sum_p 2^{-|p|}$ *(summing over programs that output a string beginning with* x*) defines a semicomputable semimeasure on* $\{0,1\}^*$ *[22, 26]. Our* μ_S *is the analogue for the continuous setting, where the "string" is a binary encoding of the point* s *to finite precision. The extension of algorithmic probability to continuous domains is treated systematically in Hutter [27], §3.3; the key point is that the Kraft inequality remains valid at every precision level* ε*, so the total mass bound* $\mu_\mathrm{S}(\mathbb{C}) \leq 1$ *is inherited by the weak-∗ limit. (Existence of the limit along subsequences follows from Prokhorov's theorem, since the measures* $\mu_{\mathrm{S},\varepsilon}$ *are uniformly tight on bounded subsets of the strip. The trace-class property depends only on the bound* $\mu_\mathrm{S}(strip) \leq 1$*, not on uniqueness of the limit.)*

Definition VIII.3 (Solomonoff-weighted recognition operator)**.** *Define the integral operator*

$$(T_\odot f)(s) = \int \odot(s, s')\, f(s')\, d\mu_\mathrm{S}(s'), \tag{27}$$

acting on $L^2(strip, \mu_\mathrm{S})$*.*

B. Trace-class and the Kraft inequality

Theorem VIII.4 (Trace-class theorem)**.** $T_\odot$ *is a self-adjoint, trace-class operator on* $L^2(strip, \mu_\mathrm{S})$ *with*

$$\mathrm{Tr}(T_\odot) = \mu_\mathrm{S}(strip) \leq 1. \tag{28}$$

[PROVEN]

Proof. Self-adjointness: $\odot(s, s') = \odot(s', s)$ (Theorem III.2(1)), so the integral operator with symmetric kernel is self-adjoint on $L^2(\text{strip}, \mu_S)$.

Boundedness: $|\odot(s, s')| \le 1$ (Theorem III.2(3)) and $\mu_S(\text{strip}) \le 1$ (Kraft inequality: $\sum_x 2^{-K(x)} \le 1$ for prefix-free codes [22]), so $\|T_\odot\| \le \sup_s \int |\odot(s, s')|\, d\mu_S(s') \le \mu_S(\text{strip}) \le 1$.

Trace-class: $\text{Tr}(T_\odot) = \int \odot(s, s)\, d\mu_S(s) = \int 1 \cdot d\mu_S(s) = \mu_S(\text{strip}) \le 1$, using $\odot(s, s) = 1$ (self-recognition). A self-adjoint operator with finite trace is trace-class [28]. □

Corollary VIII.5 (Complete spectral decomposition)**.** *$T_\odot$ has a complete orthonormal eigenbasis $\{\phi_n\}$ with real eigenvalues $\{\lambda_n\}$ satisfying $\sum_n \lambda_n \le 1$. Every function in $L^2(\text{strip}, \mu_S)$ can be expanded in this eigenbasis. No pattern with finite Kolmogorov complexity is invisible to the spectral decomposition.* [PROVEN — SPECTRAL THEOREM FOR TRACE-CLASS OPERATORS]

Remark VIII.6 (What compactification achieves)**.** *The Kraft inequality $\sum 2^{-K} \le 1$ is, in effect, a* compactness condition*: it makes the total measure finite, which in turn makes $T_\odot$ trace-class, which in turn guarantees a discrete, real, complete spectrum. Compactness in physics gives the same guarantee for free (the spectral theorem on compact manifolds). The Solomonoff prior is the information-theoretic substitute for geometric compactness.*

C. The $\mathbb{Z}_2$ eigenspace decomposition

The operator $T_\odot$ commutes with the $\mathbb{Z}_2$ action $s \mapsto 1 - \bar{s}$ (because $\odot$ is $\mathbb{Z}_2$-invariant by Theorem III.5). Therefore the Hilbert space decomposes:

$$L^2(\text{strip}, \mu_S) = V_+ \oplus V_-, \tag{29}$$

where V_+ consists of functions symmetric under $\mathbb{Z}_2$ ($f(s) = f(1 - \bar{s})$) and V_- of antisymmetric functions ($f(s) = -f(1-\bar{s})$). The operator $T_\odot$ preserves each subspace: $T_\odot|_{V_+}$ and $T_\odot|_{V_-}$ are independently self-adjoint and trace-class.

An early version of this work conjectured that $T_\odot|_{V_-}$ has trivial spectrum (no nonzero eigenvalues), which would be equivalent to RH. **This conjecture is false.** The trace of $T_\odot|_{V_-}$ is

$$\text{Tr}(T_\odot|_{V_-}) = \tfrac{1}{2}\left[\mu_S(\text{strip}) - \int \odot(1 - \bar{s},\, s)\, d\mu_S(s)\right], \tag{30}$$

and the second integral is strictly less than the first (because $\odot(1 - \bar{s}, s) < 1$ for $s \ne \frac{1}{2} + it$), so $\text{Tr}(T_\odot|_{V_-}) > 0$. The V_- sector has *background* eigenvalues from the smooth part of the kernel, regardless of zero locations. [PROVEN]

Remark VIII.7 (What compactification does not achieve)**.** *Trace-class gives a* complete *spectral decomposition: every pattern is captured. But it does not control the* fine structure *of the spectrum—specifically, whether the eigenvalues that encode the Riemann zeros sit at the values that would place the zeros on $\sigma = \frac{1}{2}$. The Kraft inequality bounds the* total weight *of the spectrum ($\sum \lambda_n \le 1$) but not the* location *of individual eigenvalues. This is the gap between* global *control (trace-class) and* local *control (eigenvalue positions), and it is one face of the unified obstruction analyzed in §X.*

D. The Gauss measure and the coding space

A suggestive connection links the Solomonoff measure to the classical Gauss measure from the theory of continued fractions.

The Gauss measure $d\mu_G(x) = dx/((1+x)\ln 2)$ is the unique invariant measure of the Gauss map $T(x) = \{1/x\}$ [29], which generates the continued-fraction expansion of a real number $x \in [0, 1]$.

The Kolmogorov complexity of x is related to its continued fraction digits $[a_1, a_2, \ldots]$ by

$$K(x) \approx \sum_k \log_2 a_k + O(\text{length}), \tag{31}$$

since each digit a_k requires $\approx \log_2 a_k$ bits. The Solomonoff weight is therefore

$$2^{-K(x)} \approx \prod_k \frac{1}{a_k}, \tag{32}$$

which has the same dependence on the digits as the Gauss measure density (via the relationship $1 + x \approx (a_1 + 1)/a_1$ for $x = 1/(a_1 + \cdots)$).

Conjecture VIII.8 (Solomonoff–Gauss equivalence). *On the continued-fraction coding space, the Solomonoff measure μ_S and the Gauss measure μ_G are mutually absolutely continuous: $\mu_S \sim \mu_G$.* [CONJECTURED]

If Conjecture VIII.8 holds, the Radon–Nikodym derivative $d\mu_S/d\mu_G$ is positive and bounded, and $L^2(\text{strip}, \mu_S)$ is unitarily equivalent to $L^2([0,1], \mu_G)$ via the continued-fraction coding map weighted by $\sqrt{d\mu_S/d\mu_G}$. This would provide the intertwining map between the information-geometric and dynamical settings needed for the Mayer correspondence (§IX).

Remark VIII.9 (Why the Gauss measure matters). *The Gauss measure is the natural measure on the symbolic dynamics of the geodesic flow on the modular surface. It is the equilibrium measure of the dynamical system whose transfer operator is Mayer's L_s. If the Solomonoff measure—APO's "complexity budget"—is equivalent to the Gauss measure—the dynamics' equilibrium—then the information-theoretic and dynamical approaches to the modular surface are seeing the* same *structure from different angles. This would not prove RH, but it would establish that the APO recognition operator and Mayer's transfer operator live on unitarily equivalent Hilbert spaces, making their spectral comparison meaningful.*

IX. MAYER'S TRANSFER OPERATOR AND THE BESSEL ISOMORPHISM

The dynamical approach to the Selberg zeta function, pioneered by Mayer [8, 30], expresses the spectral data of the modular surface as the Fredholm determinant of a *transfer operator* for the Gauss continued-fraction map. This section develops the connection to the recognition operator, including a crucial self-adjointness result that corrects an error in our earlier analysis.

A. Mayer's transfer operator and its spectrum

Definition IX.1 (Mayer's transfer operator). *For* $\mathrm{Re}(\beta) > \frac{1}{2}$, *the* generalized Gauss–Kuzmin–Wirsing operator *is*

$$L_\beta f(z) = \sum_{n=1}^{\infty} \frac{1}{(z+n)^{2\beta}} f\left(\frac{1}{z+n}\right), \tag{33}$$

acting on holomorphic functions on a disk $D = \{z \in \mathbb{C} : |z-1| < \frac{3}{2}\}$.

The operator L_β is nuclear (the Banach-space analogue of trace-class) on $A_\infty(D)$, and its matrix elements in the monomial basis $\{z^k\}$ are [8]:

$$a_{mk}(\beta) = (-1)^m \frac{\Gamma(2\beta+k+m)}{m!\,\Gamma(2\beta+k)}\,\zeta(2\beta+k+m). \tag{34}$$

Note that the Riemann zeta function appears explicitly in each matrix element—a first hint that L_β and the arithmetic recognition operator $\odot$ share structural DNA.

Mayer's central result [30] connects the Fredholm determinant to the Selberg zeta function:

$$\det(1 - L_\beta) \cdot \det(1 + L_\beta) = Z_M(\beta), \tag{35}$$

where Z_M is the Selberg zeta function of the modular surface, whose zeros encode the Maass eigenvalues and (via the scattering determinant) the Riemann zeros.

B. The Bessel isomorphism and self-adjointness

A foundational result of Mayer [8] (see also Mayer and Roepstorff [31]), often underappreciated in the RH literature, is:

Theorem IX.2 (Mayer's Bessel isomorphism). *On the Hardy space $H_{-1/2}$ of the left half-plane, the operator L_β (for real $\beta > \frac{1}{2}$) is isomorphic to an integral operator with kernel the Bessel function:*

$$K_\beta(s,t) = J_{2\beta-1}(2\sqrt{st}), \qquad s,t \in \mathbb{R}_+. \tag{36}$$

This is a Hankel transform*, which is self-adjoint on $L^2(\mathbb{R}_+, dt)$. Therefore L_β has* ***real eigenvalues*** *for all real $\beta > \frac{1}{2}$.* [PROVEN — MAYER 1991]

Remark IX.3 (Correcting an earlier error)**.** *In the course of this research, we initially concluded that "L_s is not self-adjoint on any L^2 space" based on numerical tests against $L^2([0,1],\mu_G)$ (the Gauss measure space). Those tests were correct—L_s is indeed not self-adjoint on $L^2(\mu_G)$—but the conclusion was too broad. Mayer's Bessel isomorphism shows that L_β* is *self-adjoint on the* Hardy space $H_{-1/2}$*, which is a different Hilbert space with a different inner product. The lesson: self-adjointness is a property of an operator* together with *its Hilbert space. The same operator can be self-adjoint on one space and not another.*

This result places Mayer's operator and the APO recognition operator on the same footing: both are self-adjoint integral operators on appropriate Hilbert spaces, both involve the Riemann zeta function in their structure, and both are connected to the modular group $\mathrm{SL}(2,\mathbb{Z})$.

C. Eigenvalue reality versus spectral parameter reality

Theorem IX.2 proves that *for each fixed real $\beta > \frac{1}{2}$*, the eigenvalues $\lambda_n(\beta)$ of L_β are real. This is a statement about the *operator spectrum at fixed parameter*.

RH, however, concerns the *spectral parameters*: the values of β (or equivalently s) where a specific eigenvalue equals 1:

$$\det(1 - L_\beta) = 0 \quad \Longleftrightarrow \quad \lambda_n(\beta) = 1 \text{ for some } n. \tag{37}$$

These values of β are the spectral parameters of the modular surface (Maass eigenvalues and Riemann zero heights (35)). RH asks: do all these β-values satisfy $\mathrm{Re}(\beta) = \frac{1}{2}$?

Remark IX.4 (Two different "real spectrum" claims)**.** *It is essential not to confuse:*

1. ***Eigenvalue reality:*** $\lambda_n(\beta) \in \mathbb{R}$ *for each fixed real β.* [PROVEN] *by Theorem IX.2.*
2. ***Spectral parameter reality:*** *The β-values where* $\det(1 - L_\beta) = 0$ *satisfy* $\mathrm{Re}(\beta) = \frac{1}{2}$. [EQUIVALENT TO RH; OPEN].

A family of self-adjoint matrices $A(t)$ with real eigenvalues for every t can perfectly well have $\det(A(t) - I) = 0$ *at* complex *values of t. Eigenvalue reality does not imply parameter reality. This distinction is the reason why Mayer's Bessel isomorphism, beautiful as it is, does not by itself prove RH.*

D. The Bhattacharyya–Bessel correspondence

We now have two self-adjoint integral operators associated with the modular group:

- $T_\odot$ on $L^2(\mathrm{strip}, \mu_\mathrm{S})$, with kernel $\odot(s,s') = \zeta((s+s')/2)/\sqrt{\zeta(s)\zeta(s')}$ (the arithmetic recognition operator, §VIII A).
- L_β on $L^2(\mathbb{R}_+, dt)$, with kernel $J_{2\beta-1}(2\sqrt{st})$ (Mayer's transfer operator via the Bessel isomorphism, Theorem IX.2).

Conjecture IX.5 (Bhattacharyya–Bessel correspondence)**.** *There exists a unitary map $U : L^2(strip, \mu_\mathrm{S}) \to L^2(\mathbb{R}_+, dt)$ and a function g such that*

$$\det(I - z\,T_\odot) = \det(I - L_{g(z)}) \tag{38}$$

as entire functions of z. [CONJECTURED]

If Conjecture IX.5 holds, then the eigenvalue multisets of $T_\odot$ and $L_{g(z)}$ are identical (by Hadamard factorization of entire functions of finite order). Since both operators are self-adjoint on their respective Hilbert spaces, both have real eigenvalues. The spectral parameters of the Selberg zeta function—where these eigenvalues equal 1—would then be constrained by the real spectrum of *both* operators simultaneously.

Remark IX.6 (Three sub-gaps)**.** *Establishing Conjecture IX.5 requires three steps, each nontrivial:*

1. ***The intertwining map.*** *Construct U explicitly. The Solomonoff–Gauss equivalence (Conjecture VIII.8) would provide the measure-theoretic part; the continued-fraction coding map would provide the geometric part.*

2. ***Kernel matching.*** *Show that the Bhattacharyya kernel and the Bessel kernel are related under U. The kernels have different algebraic structure (Euler product versus branch sum, normalized versus unnormalized), so the match cannot be direct; it must pass through a basis expansion where both yield the same zeta-value coefficients* (34)*.*

3. ***Parameter reality.*** *Show that the Fredholm determinant identity* (38) *constrains the spectral parameters (not just the eigenvalues at fixed parameter) to the critical line. This is the deepest sub-gap, related to the eigenvalue-vs-parameter distinction of §IX C.*

Remark IX.7 (An alternative: deformation theory)**.** *Rather than proving* (38) *directly, one might ask whether the Bhattacharyya kernel is a* deformation *of the Bessel kernel, parameterized by the Euler product. If $\odot$ can be written as $J_{2\beta-1}(2\sqrt{st}) + \varepsilon$(Euler product terms) in a suitable basis, then deformation theory for self-adjoint operators would constrain how the spectral parameters move under the deformation. This approach is less ambitious than the full Fredholm determinant match but may be more tractable.* [OPEN]

X. THE ONE GAP, SIX FACES

Every approach developed in this paper—per-prime decomposition, compression argument, Solomonoff compactification, Mayer correspondence—encounters the same obstruction. This section identifies the common structure.

A. Six faces of the same obstruction

The gap between what the APO framework provides and what RH requires has appeared in six guises throughout this paper. In each case, the left column is what the framework controls; the right column is what RH needs.

TABLE II. The one gap, six faces.

	APO controls	**RH requires**
1	Divergence (KL, K)	Overlap (B, zeros)
2	Per-prime (independent)	Collective (cancellation)
3	Discrete spectrum (Maass)	Scattering resonances
4	Compact ($\chi > 0$, CP^1)	Cusp ($\chi < 0$, modular)
5	GUE statistics (spacing)	Zero locations ($\sigma = \frac{1}{2}$?)
6	Trace-class (Solomonoff)	Fine spectral structure

Each row is a manifestation of the same structural mismatch: APO's axioms yield *global, statistical, collective* constraints (the left column), while RH is a *local, analytic, individual* statement (the right column).

B. The measure–support gap

The common mathematical structure underlying all six faces is:

Knowing the measure does not determine the support.

The Solomonoff prior provides a measure (2^{-K}). The Bhattacharyya coefficient provides a kernel. Together they yield a trace-class operator whose total spectral weight is bounded ($\mathrm{Tr} \leq 1$). But the *individual* eigenvalue positions are not constrained by the total weight. The Kraft inequality says the sum $\sum \lambda_n$ is at most 1; it does not say where each λ_n sits.

In the per-prime decomposition: each prime's contribution to the Li coefficients oscillates in sign (§VII D). The total is non-negative (assuming RH), but the positivity comes from collective cancellation, not individual positivity. The measure (per-prime Fisher information) is controlled; the support (where the Li coefficients actually land) is not.

In the topology: the Euler characteristic $\chi = -1/6$ permits both discrete and continuous spectrum. The discrete part is controlled by self-adjointness; the continuous part (scattering resonances) is not. The measure (area $\pi/3$, finite) is known; the support (where the resonances sit) is not determined by it.

Remark X.1 (What would close the gap). *A proof of RH from this framework would require one of:*

1. *A* Fredholm determinant identity *relating $T_\odot$ to the Selberg zeta function (Conjecture IX.5): this would convert the measure-level data (trace-class, eigenvalue sums) into support-level data (individual eigenvalue positions match Selberg spectral parameters).*
2. *A* deformation argument *showing that the Bhattacharyya kernel is a perturbation of the Bessel kernel that preserves spectral parameter locations.*
3. *A* compactification argument *that geometrically closes the cusp, converting the continuous spectrum into discrete while preserving spectral locations.*

Each of these is a concrete mathematical program connecting well-defined objects. None has been carried out, but each is in principle tractable—the objects are explicit, the Hilbert spaces are identified, and the relevant spectral theory is well-developed.

XI. CONJECTURED INTERPRETATION AND FALSIFIABLE PREDICTIONS

Although the framework does not prove RH, it generates interpretive structure and falsifiable predictions that go beyond the existing literature.

A. The nine-step interpretation

The following chain summarizes the conjectured *reason* RH is true, with each step labeled by status:

1. The primes are the atoms of arithmetic ($\otimes$ over primes). [PROVEN, §II B]
2. The Riemann zeta function is the partition function of the arithmetic statistical manifold. [PROVEN, §II A]
3. The Fisher geometry on this manifold is the unique Chentsov metric. [PROVEN, §II D]
4. The recognition operator $\odot = \zeta((s+s')/2)/\sqrt{\zeta(s)\zeta(s')}$ is symmetric and self-adjoint. [PROVEN, §III A]
5. Pass 2 (mutual recognition) squares the overlap, creating invariance under the functional equation. [PROVEN, §III B]
6. The functional equation is the *dephasing channel*†: it strips σ-deviation, projecting to the critical line. [PROVEN ($\mathbb{Z}_2$ INDISTINGUISHABILITY) / ARGUED (QUOTIENT INTERPRETATION), §IV]
7. The Solomonoff prior 2^{-K} regulates the zero-set: the configuration with all zeros at $\sigma = \frac{1}{2}$ has minimal Kolmogorov complexity. [ARGUED, §VII B]
8. Off-line zeros are redundant (same arithmetic content stored in two places) and therefore carry higher K. [ARGUED, §VII B]
9. The surviving configuration—the one the recognition cycle selects as maximally compressed—has all zeros at $\sigma = \frac{1}{2}$. [CONJECTURED]

Steps 1–6 constitute a mathematical framework with rigorous content. Steps 7–8 are information-theoretic arguments that identify a qualitative preference but do not supply quantitative proof. Step 9 is the conjectural leap: that the qualitative preference is absolute rather than probabilistic.

B. Three predictions

The framework generates three predictions that are falsifiable independently of RH.

Prediction 1: GUE consistent with recognition geometry. The GUE statistics of the Riemann zeros are *consistent with* the symmetry class of the arithmetic $\odot$ operator: the functional equation breaks time-reversal symmetry (selecting $\beta = 2$ over $\beta = 1$) while preserving unitarity (§VI B). If this consistency is not accidental, it predicts that *any* L-function whose functional equation has the same $\mathbb{Z}_2 \times \mathbb{Z}_2$ symmetry structure (reflection $\times$

conjugation, not linked) should exhibit GUE spacing statistics. L-functions with linked symmetries (e.g., real-valued L-functions with $L(s) = \overline{L(\bar{s})}$ and a self-dual functional equation) should exhibit GOE statistics instead. This is consistent with the Katz–Sarnak philosophy [19, 20]; the information-geometric perspective provides a complementary motivation for the same symmetry classification rather than an independent derivation. [ARGUED; TESTABLE]

Prediction 2: The critical line is maximally compressed. Among all zero configurations of ξ compatible with the functional equation, the configuration with $\sigma_n = \frac{1}{2}$ for all n has minimal total Kolmogorov complexity. This is testable in principle: for any finite truncation of the zero list $\{t_1, \ldots, t_N\}$, the description length $K(\{t_n : n \leq N\})$ should be strictly less than $K(\{t_n : n \leq N\}) + K(\{\sigma_n : \sigma_n \neq 1/2\})$ for any hypothetical off-line configuration. The prediction is that the "extra" bits needed to specify $\sigma_n \neq \frac{1}{2}$ are strictly redundant (they carry no information about the primes not already encoded in the heights $\{t_n\}$). [ARGUED; TESTABLE IN PRINCIPLE]

Prediction 3: The Bhattacharyya and Bessel kernels are related. The arithmetic recognition kernel $\odot$ and Mayer's Bessel kernel $J_{2\beta-1}(2\sqrt{st})$ share the same spectral data (up to a parameter map), because both are self-adjoint operators encoding the arithmetic of $\mathrm{SL}(2,\mathbb{Z})$. This is testable by numerical comparison of Fredholm determinants or by explicit construction of the intertwining map U in Conjecture IX.5. [CONJECTURED; TESTABLE]

Remark XI.1 (A concrete finite test for Prediction 3)**.** *We performed this computation at $N = 5$ with $\beta = 3/4$. Truncating both operators to 5×5 matrices:*

- *Mayer's side: the matrix $a_{mk}(\beta)$ from Eq. (34), computed at $\beta = 3/4$.*
- *APO's side: $T_\odot$ discretized on a uniform grid of 5 points in $[1.5, 4.0]$ with uniform quadrature weights.*

Results: $\det(I - L_{3/4}) \approx 2.58 \times 10^2$, $\det(I - T_\odot) \approx -1.11$*;* $\mathrm{Tr}(L_{3/4}) \approx 76.0$, $\mathrm{Tr}(T_\odot) = 2.50$*; the leading eigenvalue ratio is ~ 32.*

The direct identification ($g =$ identity) is refuted*: the eigenvalue scales differ by factors of 30–2500, and the Bessel kernel values differ from the Bhattacharyya kernel by varying ratios (not a constant rescaling).*

What survives *is the full conjecture* $\det(I - z\,T_\odot) = \det(I - L_{g(z)})$ *with a* nontrivial *parameter map g. The map g must absorb the scale mismatch, which varies with β (the trace ratio changes from ~ 30 at $\beta = 3/4$ to ~ 66 at $\beta = 3/2$). This makes the deformation-theory route (cf. the remark at the end of §IX D) more relevant than the direct Fredholm comparison.*

XII. CONCLUSION

We constructed the arithmetic recognition operator $\odot(s, s') = \zeta((s + s')/2)/\sqrt{\zeta(s)\,\zeta(s')}$ on the statistical manifold of the zeta distribution and established its principal properties: symmetry, boundedness, self-recognition, and—via the completed version O_ξ—invariance under the functional equation. The Fisher metric is diagonal on the critical line; the Riemann–Siegel theta phase cancels in the Pass 2 observable; and one cycle of the distinction–recognition–integration process projects every arithmetic pattern onto $\sigma = \frac{1}{2}$.

We identified the modular surface $M = \mathbb{H}/\mathrm{SL}(2,\mathbb{Z})$ as the natural geometric home, with the Riemann zeros appearing as scattering resonances of its Laplacian. The Solomonoff prior provides effective compactification (trace-class), and the resulting operator has a complete spectral decomposition. The Kolmogorov complexity argument shows that on-line zero configurations are maximally compressed, while the per-prime Li decomposition (which we refuted as a route to RH) reveals that positivity is collective rather than per-prime.

The gap between these results and RH is unified across six faces (Table II): in each case, the framework provides global/statistical control while RH requires local/analytic precision. The most promising path to closing the gap is the Bhattacharyya–Bessel correspondence (Conjecture IX.5), which would link the recognition operator's Fredholm determinant to the Selberg zeta function and thereby constrain the spectral parameters. A preliminary 5×5 numerical test (Appendix C) refutes the direct identification ($g =$ identity) but leaves the full conjecture—with a nontrivial parameter map g—intact.

What we do not claim

This paper does not prove the Riemann Hypothesis. It constructs machinery, identifies a precise remaining gap, and frames RH as a question about *recognition completeness*†: whether the arithmetic recognition structure sees *all* structural features of ξ, or whether some zeros can hide in the scattering sector of the cusp. The philosophical commitment of pattern monism—that to exist is to be recognizable—*demands* completeness, but the demand is not a proof.

Open directions

1. **Fredholm determinant computation.** The most direct path: compute $\det(I - z\,T_{\odot})$ explicitly from the $\odot$ kernel and Solomonoff measure, and compare with Z_M.
2. **Bhattacharyya–Bessel deformation.** Show that $\odot$ is a deformation of the Bessel kernel parameterized by the Euler product, and apply deformation theory for self-adjoint operators.
3. **Solomonoff–Gauss equivalence.** Prove Conjecture VIII.8 and construct the intertwining map U explicitly.
4. **Rényi–Sanov bounds.** Closing the K-inside-B gap (the Bhattacharyya–KL inequality provides only a lower bound on B) may require a Rényi–Sanov bound for $D_{1/2}$, currently absent from the AIT literature.
5. **Time-dependent Fisher information.** A "height-dependent FIM"—tracking how the Fisher geometry evolves as one moves up the critical line—may provide a dynamical perspective analogous to the time-dependent FIM used in the companion Navier–Stokes paper.

ACKNOWLEDGMENTS

The author thanks Claude (Anthropic) for extensive collaborative development, including construction of the arithmetic recognition operator, numerical verification of all claims, stress-testing of conjectures (including honest identification and retraction of those that failed), and the identification of Mayer's Bessel isomorphism as the bridge between the APO and dynamical programs. The author also thanks Grok (xAI) for pointing toward the Mayer–Selberg connection and natural extension structure, with the caveat that several of its specific claims were found to be incorrect upon verification.

Appendix A: APO terminology: formal definitions and foundational status

TABLE III. APO terminology dictionary. Status: A = Axiom, D = Derived, C = Conjectured. Standard equivalents are exact identifications, not analogies.

APO term	**Formal definition**	**Standard name**	**Status**
$\otimes$ (Distinction)	Binary partition	Sufficient statistic	A
$\odot$ (Recognition)	$B(p,q) = \sum_n \sqrt{p_n q_n}$	Bhattacharyya coeff.	A
$\oplus$ (Integration)	Quotient $M/\ker\odot$	Orbit projection	D
Pass 1	$\langle\psi_s\|\psi_{s'}\rangle$	Inner product on S^{n-1}	D
Pass 2	$\|\langle\psi_s\|\psi_{s'}\rangle\|^2$	Transition probability	D
Dephasing	$\mathbb{Z}_2$ quotient $s \sim 1-s$	Functional eq. quotient	D
Solomonoff prior	$2^{-K(x)}$ weighting	Algorithmic probability	A
Recog. Completeness	No dark zeros of ξ	Equivalent to RH	C

Appendix B: Master proof status table

Appendix C: Numerical verifications

All computations were performed in Python 3.12 using NumPy and SciPy, with the Riemann zeta function approximated by partial Dirichlet sums (up to $N = 10{,}000$ terms for convergence in the region $\mathrm{Re}(s) > 1$).

Per-prime Fisher verification. The per-prime Fisher information $\mathcal{I}_p(s) = (\log p)^2\, p^{-s}/(1-p^{-s})^2$ (§II C) was summed over all primes $p \leq 10{,}000$ and compared to the total Fisher information computed via the second derivative of $\log\zeta(s)$. At $s = 2$: ratio = 0.9998, confirming the decomposition to four significant figures.

$\odot$ **properties.** The recognition operator $\odot(s,s')$ was evaluated at 20 pairs (s,s') with $\mathrm{Re}(s), \mathrm{Re}(s') > 1$. Symmetry ($\odot(s,s') = \odot(s',s)$): verified to machine precision ($\sim 10^{-15}$). Self-recognition ($\odot(s,s) = 1$): verified to machine precision. Boundedness ($0 < \odot \leq 1$): confirmed at all test points.

TABLE IV. Proof status for all results. Status labels: PROVEN = rigorous proof provided or cited; ARGUED = supported by computation and structural reasoning but not fully rigorous; CONJECTURED = motivated but unproven; REFUTED = tested and found false; OPEN = identified but not resolved.

#	Component	Status	Basis
1	Zeta distribution is exponential family	PROVEN	Sec. II A
2	Euler product $\Rightarrow$ per-prime independence	PROVEN	Sec. II B
3	Per-prime Fisher $I_p(s)$; verified numerically	PROVEN	Sec. II C
4	$\odot(s, s')$ symmetric, bounded, self-recognizing	PROVEN	Sec. III A
5	Pass 2 observable $\mathbb{Z}_2$-invariant	PROVEN	Sec. III B
6	$g_{\sigma t} = 0$ on critical line	PROVEN	Sec. III C
7	Born rule eliminates θ-phase	PROVEN	Sec. III D
8	$\odot$ indistinguishes σ from $1-\sigma$	PROVEN	Sec. IV A
9	One-step projection $\sigma \to \frac{1}{2}$	ARGUED	Sec. IV B
10	Fisher singularities at zeros of ξ	PROVEN	Sec. III A
11	Modular surface as arithmetic CP^1	ARGUED	Sec. V
12	Riemann zeros = scattering resonances of M	PROVEN	Sec. V B
13	$\sigma = \frac{1}{2}$ is geodesic + balance point	PROVEN/ARGUED	Sec. V C
14	$\otimes \to$ eigenvalue repulsion (Fisher cost)	ARGUED	Sec. VI A
15	Functional eq. selects GUE ($\beta = 2$)	ARGUED	Sec. VI B
16	K(on-line) < K(off-line) for same content	ARGUED	Sec. VII B
17	Per-prime Li positivity	REFUTED	Sec. VII D
18	GUE statistics $\neq$ zero locations	ARGUED	Sec. VI D
19	$T_\odot$ trace-class on $L^2(\text{strip}, \mu_{\mathrm{S}})$	PROVEN	Sec. VIII B
20	$\mathbb{Z}_2$ decomposition $V_+ \oplus V_-$	PROVEN	Sec. VIII C
21	$\mathrm{Tr}(T\|_{V_-}) > 0$ (background eigenvalues)	PROVEN	Sec. VIII C
22	L_s self-adjoint on Hardy $H_{-1/2}$	PROVEN	Mayer [8]
23	Eigenvalue reality $\neq$ spectral parameter reality	PROVEN	Sec. IX C
24	Bhattacharyya–Bessel correspondence	CONJECTURED	Sec. IX D
25	$\det(I - zT_\odot) = Z_M(g(z))$	OPEN	Sec. IX D
26	RH from correspondence + self-adjointness	OPEN	Sec. X

Li coefficients. The first 15 Keiper–Li coefficients were computed from the 10 lowest known zero heights using $\lambda_n = \sum_\rho [1 - (1 - 1/\rho)^n]$: $\lambda_1 = 0.0135$, $\lambda_2 = 0.0541$, $\ldots$, $\lambda_{15} = 2.898$. All positive, consistent with RH.

Per-prime Li oscillation. The per-prime coefficients $\lambda_n^{(p)}$ were computed analytically using (25) for $p = 2$ and $n = 1, \ldots, 30$: $\lambda_1^{(2)} = -0.693$, $\lambda_4^{(2)} = -0.001$, $\lambda_5^{(2)} = +0.031$, $\lambda_{18}^{(2)} = -0.001$. The sign oscillation is confirmed; per-prime positivity is refuted. Note that $\lambda_1^{(2)} = -\ln 2$ exactly, a fact verified independently to 50-digit precision using `mpmath`.

Mayer matrix eigenvalues. The $N \times N$ truncation of Mayer's matrix (34) was diagonalized for $N \in \{4, 6, 8, 10, 12\}$ and $s \in \{0.51, 0.60, 0.75, 1.0\}$. All eigenvalues were real in every case, with maximum imaginary part $< 10^{-10}$ (machine precision), consistent with Mayer's Bessel isomorphism (Theorem IX.2). Independent verification at 50-digit precision (`mpmath`) confirmed that imaginary parts are identically zero to within 10^{-50}, ruling out numerical artifact.

Archimedean factor profile. $\log |\pi^{-s/2}\Gamma(s/2)|$ was evaluated at $s = \sigma + 100i$ for $\sigma \in [0.05, 0.95]$. The function is concave in σ with maximum at $\sigma = \frac{1}{2}$ (Proposition V.5), confirming that the archimedean factor is strongest on the critical line.

Fredholm determinant comparison (Prediction 3). The 5×5 truncated Mayer matrix at $\beta = 3/4$ gives $\det(I - L) \approx 258$, $\mathrm{Tr}(L) \approx 76$, leading eigenvalue ≈ 76. The 5×5 discretized APO operator (uniform grid on $[1.5, 4.0]$, uniform quadrature) gives $\det(I - T_\odot) \approx -1.1$, $\mathrm{Tr}(T_\odot) = 2.5$, leading eigenvalue ≈ 2.3. The Bessel kernel $J_{0.5}(2\sqrt{st})$ and the Bhattacharyya kernel $\odot(s, s')$ differ by non-constant ratios ranging from 0.065 to -0.42 across the 3×3 test grid, confirming they are not related by a simple rescaling. The direct identification ($g = \mathrm{id}$) is refuted; the conjecture with nontrivial g survives.

Appendix D: Approaches tested and refuted

Scientific honesty requires documenting not only what works but what was tried and found wanting. The following approaches were explored during the development of this paper, tested numerically, and refuted. Each refutation sharpened the understanding of the gap (§X).

1. Per-prime Li positivity (§VII D). *Conjecture*: $\lambda_n^{(p)} \geq 0$ for all primes p and all $n \geq 1$. *Refutation*: $\lambda_1^{(2)} = -0.693 < 0$. The per-prime contributions oscillate in sign. Positivity of the total λ_n requires collective cancellation across all primes plus the archimedean factor. *Lesson*: independence (Euler product) does not imply per-component positivity.

2. V_- **triviality** $\Leftrightarrow$ **RH** (§VIII C). *Conjecture*: $T_\odot|_{V_-}$ has trivial discrete spectrum if and only if all zeros are on the critical line. *Refutation*: $\mathrm{Tr}(T_\odot|_{V_-}) > 0$ regardless of zero locations. The V_- sector has "background" eigenvalues from the smooth part of the kernel. *Lesson*: trace-class gives a complete spectral decomposition (global) but does not control individual eigenvalue positions (local).

3. L_s **not self-adjoint on any** L^2 (§IX B). *Claim*: Mayer's transfer operator L_s is not self-adjoint on any L^2 space, precluding unitary equivalence with $T_\odot$. *Correction*: L_β *is* self-adjoint on the Hardy space $H_{-1/2}$ via Mayer's Bessel isomorphism (Theorem IX.2). Our numerical test against $L^2(\mu_G)$ was testing the wrong Hilbert space. *Lesson*: self-adjointness is a joint property of the operator *and* its Hilbert space.

4. Natural extension L_s^2 **route** (§IX B). *Idea*: the natural extension of the Gauss map has a time-reversal symmetry; its transfer operator restricted to the symmetric sector is self-adjoint; this yields L_s^2 whose eigenvalues are automatically real and non-negative. *Refutation*: the eigenvalues of L_s^2 are squares of those of L_s, so they are automatically non-negative regardless of whether L_s's eigenvalues are real. The squaring loses sign information, giving no new constraint. *Lesson*: non-negative operators are automatically "self-adjoint-looking" without providing content about the original operator's spectrum.

5. Direct Fredholm identification $(g = \mathrm{id})$ (Appendix C). *Conjecture*: $\det(I - T_\odot) = \det(I - L_\beta)$ with g equal to the identity (the simplest case of Conjecture IX.5). *Refutation*: at $\beta = 3/4$ and $N = 5$, the Mayer determinant is ≈ 258 while the APO determinant is ≈ -1.1. The eigenvalue scales differ by factors of 30–2500. The Bessel kernel and the Bhattacharyya kernel differ by non-constant ratios across the test grid. *Lesson*: the correspondence, if it exists, requires a *nontrivial* parameter map g that absorbs the scale mismatch. The deformation-theory route is more promising than the direct Fredholm comparison.

---

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

[2] H. Iwaniec and E. Kowalski, *Analytic Number Theory*, Colloquium Publications, Vol. 53 (AMS, 2004).

[3] H. L. Montgomery, The pair correlation of zeros of the zeta function, Proc. Symp. Pure Math. **24**, 181 (1973).

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

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

[6] A. Connes, Trace formula in noncommutative geometry and the zeros of the Riemann zeta function, Selecta Math. **5**, 29 (1999).

[7] A. Selberg, Harmonic analysis and discontinuous groups in weakly symmetric Riemannian spaces with applications to Dirichlet series, J. Indian Math. Soc. **20**, 47 (1956).

[8] D. H. Mayer, On the thermodynamic formalism for the Gauss map, Commun. Math. Phys. **130**, 311 (1991), proves L_β is isomorphic to a Hankel transform with Bessel kernel on Hardy space $\mathrm{H}_{-1/2}$; real spectrum for $\beta > 1/2$.

[9] J. Lewis and D. Zagier, Period functions for Maass wave forms. I, Ann. Math. **153**, 191 (2001).

[10] D. C. Dennett, Real patterns, J. Philos. **88**, 27 (1991).

[11] J. Ladyman and D. Ross, *Every Thing Must Go: Metaphysics Naturalized* (Oxford University Press, 2007).

[12] S. Amari and H. Nagaoka, *Methods of Information Geometry* (AMS/Oxford, 2000).

[13] N. N. Chentsov, Statistical decision rules and optimal inference, Translations of Mathematical Monographs **53** (1982), originally published in Russian, 1972. Proves uniqueness of the Fisher metric on statistical manifolds.

[14] A. Bhattacharyya, On a measure of divergence between two statistical populations defined by their probability distributions, Bull. Calcutta Math. Soc. **35**, 99 (1943).

[15] R. Landauer, Irreversibility and heat generation in the computing process, IBM J. Res. Dev. **5**, 183 (1961).

[16] H. Iwaniec, *Spectral Methods of Automorphic Forms*, 2nd ed. (AMS/Revista Matemática Iberoamericana, 2002).

[17] D. Borthwick, *Spectral Theory of Infinite-Area Hyperbolic Surfaces* (Birkhäuser, 2007).

[18] M. L. Mehta, *Random Matrices*, 3rd ed. (Academic Press, 2004).

[19] N. M. Katz and P. Sarnak, Zeroes of zeta functions and symmetry, Bull. AMS **36**, 1 (1999).

[20] J. P. Keating and N. C. Snaith, Random matrix theory and $\zeta(1/2 + it)$, Commun. Math. Phys. **214**, 57 (2000).

[21] P. J. Forrester, *Log-Gases and Random Matrices* (Princeton University Press, 2010).

[22] M. Li and P. Vitányi, *An Introduction to Kolmogorov Complexity and Its Applications*, 3rd ed. (Springer, 2008).
[23] X.-J. Li, The positivity of a sequence of numbers and the Riemann hypothesis, J. Number Theory **65**, 325 (1997).
[24] E. Bombieri and J. C. Lagarias, Complements to Li's criterion for the Riemann hypothesis, J. Number Theory **77**, 274 (1999).
[25] J. C. Lagarias, Li coefficients for automorphic l-functions, Ann. Inst. Fourier **57**, 1689 (2007).
[26] R. J. Solomonoff, A formal theory of inductive inference, Inform. Control **7**, 1 (1964).
[27] M. Hutter, *Universal Artificial Intelligence: Sequential Decisions Based on Algorithmic Probability* (Springer, 2005) section 3.3 treats the extension of the Solomonoff prior to continuous domains.
[28] P. Deift, *Orthogonal Polynomials and Random Matrices: A Riemann–Hilbert Approach* (AMS/CIMS, 1999).
[29] M. Einsiedler and T. Ward, *Ergodic Theory: with a view towards Number Theory* (Springer, 2011).
[30] D. H. Mayer, The thermodynamic formalism approach to Selberg's zeta function for $\mathrm{PSL}(2,\mathbb{Z})$, Bull. AMS **25**, 55 (1991).
[31] D. H. Mayer and G. Roepstorff, On the relaxation time of Gauss' continued fraction map I. The Hilbert space approach, J. Stat. Phys. **47**, 149 (1987), shows the adjoint of the Koopman operator is isomorphic to a symmetric integral operator.