# Variational Persistence under Information Causality
# Viability selection for nonlocal correlations in concatenated random-access codes

Alberto Alejandro Duarte[1, *] and Carlos Paul Avalos Soto[1]

[1] *Paradox Systems R&D, La Paz, Baja California Sur, México*

(Dated: February 28, 2026)

The Tsirelson bound $\rho \leq 2\sqrt{2}$ is the maximal violation of the CHSH inequality compatible with quantum theory. We show that it also emerges as a *variational optimum* of persistence for agents that rely on nonlocal correlations to solve a stochastic family of random-access code (RAC) tasks of unbounded difficulty. Using Information Causality (IC) as a hard viability constraint, we derive a closed-form fitness functional $\bar{W}(\rho) = G(\rho)\, S_\infty(\rho)$, where $G(\rho)$ is the per-task information benefit and $S_\infty(\rho)$ is the long-horizon survival probability. For all task distributions with unbounded support, $S_\infty(\rho)$ reduces to a sharp phase transition at $\rho = 2\sqrt{2}$; consequently the unique persistent optimum is $\rho^* = 2\sqrt{2}$. The benefit function $G(\rho) = 1 - h_2\big(\frac{4-\rho}{8}\big)$ is derived entirely from Shannon entropy and the RAC protocol, with zero free parameters, and Shannon entropy is shown to be the unique measure compatible with the concatenation structure. We further obtain an exact *Benefit–Metric Correspondence*: $G''(\rho) = \mathcal{I}_F(\rho)/\ln 2$, where $\mathcal{I}_F(\rho)$ is the classical Fisher information of the effective binary symmetric channel induced by $\rho$. This identity establishes a direct link between viability selection and information geometry, and provides a geometric reinterpretation of the $(2\eta^2)^n$ scaling identified by Carmi and Moskovich: IC is equivalent to stability of the Fisher–Rao metric pullback under protocol concatenation, with the Tsirelson bound as the critical eigenvalue $\lambda = 2\eta^2 = 1$.

## I. INTRODUCTION

The CHSH inequality [1] bounds the strength of bipartite correlations: classical theories satisfy $\rho \leq 2$, while quantum mechanics permits $\rho \leq 2\sqrt{2}$ [2]. The no-signaling principle alone allows $\rho \leq 4$, realized by Popescu–Rohrlich (PR) boxes [3]. A central question in the foundations of quantum mechanics is: *why does nature stop at $2\sqrt{2}$?*

Several information-theoretic principles have been proposed to explain this boundary. Van Dam [4] and Brassard et al. [5] showed that superquantum correlations trivialize communication complexity (CC), but their bound fell short of $2\sqrt{2}$, leaving a gap at $\rho \lesssim 3.27$. Pawłowski et al. [6] introduced Information Causality (IC) and demonstrated that IC, applied to concatenated random access code (RAC) protocols, recovers the Tsirelson bound exactly. Carmi and Moskovich [7] independently showed that the accumulated Fisher information in the van Dam protocol scales as $(2c^2)^n$ and derived the Tsirelson bound from a "Statistical No-Signaling" condition. Numerous other device-independent principles recover the CHSH Tsirelson bound, including no advantage for nonlocal computation [8], macroscopic locality [9], the exclusivity principle [10], and relativistic independence [11].

Here we show that the IC convergence at $2\sqrt{2}$ acquires a persistence interpretation within the *Variational Principle of Persistence* (VPP): the quantum boundary emerges not merely as a prohibition wall, but as the unique global optimum of a persistence functional $\bar{W}(\rho) = G(\rho) \cdot S_\infty(\rho)$. The benefit function $G(\rho)$ is derived entirely from Shannon's binary entropy [12] and the RAC protocol structure—with zero free parameters.

Our specific contributions beyond the existing literature are:

1. The derivation of $G(\rho) = 1 - h_2\big(\frac{4-\rho}{8}\big)$ from first principles, revealing that $G$ is strictly convex ($G'' > 0$), not concave as in standard diminishing-returns frameworks.

2. The construction of a complete functional $\bar{W}(\rho)$ with zero free parameters, where $P(n)$ is shown to naturally have unbounded support from algorithmic, maximum-entropy, and thermodynamic arguments.

3. The identification of a *generalized Systemic Reduction Paradox* (SRP/PRS): an interior optimum arising from accelerating benefit against a hard viability wall, rather than the classical "concave benefit vs. convex cost" mechanism.

4. A proof that Shannon entropy is the *unique* entropy measure compatible with concatenated RAC structure, grounded in its exclusive satisfaction of the strong chain rule—reinforced by the independent finding that Rényi-based alternatives yield incorrect bounds [13].

5. Extension to $N$-partite Mermin inequalities [14] within the bipartite-decomposable regime, conditional on structural assumptions stated explicitly, with acknowledgment of the Gallego–Wolf–Acín impossibility result [15].

6. Integration of communication complexity [4, 5] as a weaker but independent viability channel that is strictly subsumed by the IC wall.

7. An exact *Benefit–Metric Correspondence* ($G'' = \mathcal{I}_F/\ln 2$) linking the curvature of the VPP benefit

* albertoduarte@paradoxsystems.xyz

function to the classical Fisher information of the effective BSC induced by $\rho$, and a *metric criticality theorem*: IC is equivalent to stability of the Fisher–Rao metric pullback under protocol concatenation, with Tsirelson as the critical eigenvalue $\lambda = 2\eta^2 = 1$.

The paper is organized as follows. Section II establishes information-theoretic primitives and the necessity of Shannon entropy. Section III introduces the RAC protocol and derives the benefit function. Section IV formalizes IC and the viability horizon. Section V constructs the VPP functional and proves the main theorem. Sections VI–VIII address closure of $P(n)$, communication complexity, and multipartite extensions. Section IX summarizes computational validation. Section X establishes the information-geometric structure, proving the Benefit–Metric Correspondence and the metric criticality theorem. Section XI discusses implications and limitations.

## II. INFORMATION-THEORETIC PRIMITIVES

*a. Notation.* Throughout this paper, $\rho$ denotes the CHSH parameter (Bell value), $\rho \in [2, 4]$, not a density matrix. No density matrices appear in the analysis. We define $\eta := \rho/4 \in [1/2, 1]$ as the box visibility; for comparison, Carmi and Moskovich [7] use $c$ for the same quantity.

**Definition 1** (Binary Shannon entropy). *For $e \in (0, 1)$:*

$$h_2(e) = -e \log_2 e - (1 - e) \log_2(1 - e). \tag{1}$$

*The function $h_2$ satisfies $h_2(e) \geq 0$, $h_2(1/2) = 1$, $h_2(e) = h_2(1 - e)$, and is strictly concave: $h_2''(e) = -1/[e(1 - e) \ln 2] < 0$.*

**Definition 2** (Binary symmetric channel (BSC)). *A BSC with crossover probability $e$ transmits bit $U \in \{0, 1\}$ producing output $\hat{G}$ with $\Pr(\hat{G} \neq U) = e$.*

**Proposition 3** (BSC capacity). *For a BSC with uniform input and error $e$:*

$$I(U : \hat{G}) = 1 - h_2(e). \tag{2}$$

*Proof.* $H(U) = 1$ (uniform input), $H(U|\hat{G}) = h_2(e)$ (channel noise). By definition, $I(U : \hat{G}) = H(U) - H(U|\hat{G}) = 1 - h_2(e)$. □

**Lemma 4** (Convexity of $1 - h_2$). *The function $f(e) = 1 - h_2(e)$ is strictly convex on $(0, 1)$:*

$$f''(e) = \frac{1}{e(1 - e) \ln 2} > 0. \tag{3}$$

**Remark 5** (Necessity of Shannon entropy). *The IC proof of the Tsirelson bound relies on three properties of mutual information: (i) the* data processing inequality, *(ii) the* strong chain rule $I(X_1, X_2 : Y) = I(X_1 : Y) + I(X_2 : Y|X_1)$*, and (iii) the constraint that local operations cannot create correlations. Among all entropy functionals in the Rényi family $H_\alpha$ (which includes Shannon as the $\alpha \to 1$ limit),* only *Shannon entropy satisfies the strong chain rule. More broadly, the Khinchin–Faddeev theorem [20, 21] characterizes Shannon entropy as the unique functional satisfying continuity, maximality at the uniform distribution, and the chain rule—a result that extends well beyond the Rényi family.*

*Oughton and Timpson [13] independently confirmed this structural dependence: replacing Shannon with Rényi entropy in the IC framework yields incorrect bounds, either failing to reach $2\sqrt{2}$ or admitting superquantum correlations. Their result demonstrates that the Shannon measure carries physical content—the chain rule encodes the compositional structure of information processing—rather than being an arbitrary choice among alternatives. In the VPP framework, this means $G(\rho) = 1 - h_2(\frac{4-\rho}{8})$ is the* unique *benefit function compatible with concatenated information protocols.*

## III. RAC PROTOCOL AND BENEFIT FUNCTION

**Definition 6** (Isotropic no-signaling box). *An isotropic box with visibility $\eta \in [0, 1]$ takes inputs $(x, y) \in \{0, 1\}^2$ and produces outputs $(a, b)$ satisfying*

$$\Pr(a \oplus b = x \cdot y) = \frac{1 + \eta}{2}, \tag{4}$$

*with uniform marginals (no-signaling). The CHSH value is $\rho = 4\eta$.*

Key values: $\eta = 1/2$ ($\rho = 2$, classical); $\eta = 1/\sqrt{2}$ ($\rho = 2\sqrt{2}$, Tsirelson); $\eta = 1$ ($\rho = 4$, PR-box).

**Definition 7** ($(2,1)$-RAC protocol). *Alice holds $\mathbf{u} = (u_0, u_1) \in \{0, 1\}^2$; Bob wants $u_y$ for uniform $y$. Protocol: Alice computes $x = u_0 \oplus u_1$, inputs $x$ to the box receiving $a$, and sends $\mathrm{msg} = a \oplus u_0$ (one classical bit). Bob inputs $y$ to the box receiving $b$, and computes $\hat{g} = \mathrm{msg} \oplus b$.*

**Proposition 8** (RAC success probability).

$$p_{\text{correct}} = \Pr(\hat{g} = u_y) = \frac{1 + \eta}{2}. \tag{5}$$

*Proof.* We have $\hat{g} = u_0 \oplus (a \oplus b)$.

*Case* $y = 0$: We need $\hat{g} = u_0$, i.e., $a \oplus b = 0$. Since $x \cdot y = 0$, the box gives $\Pr(a \oplus b = 0) = (1 + \eta)/2$.

*Case* $y = 1$: We need $a \oplus b = u_0 \oplus u_1 = x$. Since $x \cdot y = x$, the box gives $\Pr(a \oplus b = x) = (1 + \eta)/2$. □

**Corollary 9.** *The RAC implements an effective BSC with error $e(\eta) = (1 - \eta)/2$.*

**Theorem 10** (Benefit function—derived). *For CHSH value $\rho \in [2, 4]$ with $\eta = \rho/4$, the informational benefit per RAC query is:*

$$\boxed{G(\rho) = 1 - h_2\left(\frac{4 - \rho}{8}\right).} \tag{6}$$

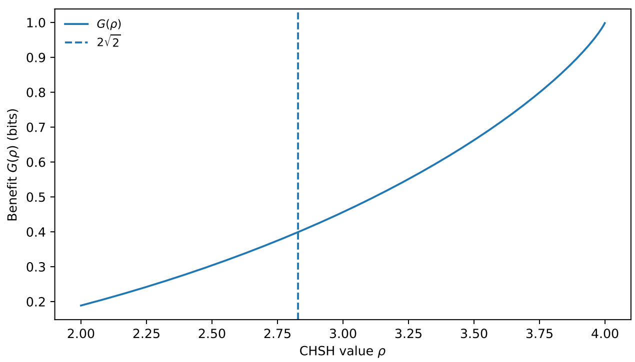


FIG. 1. (a) Benefit function $G(\rho)$ derived from Shannon entropy and the RAC protocol. (b) Analytic derivatives: $G' > 0$ (monotonically increasing) and $G'' > 0$ (strictly convex) throughout $(2,4)$. Dashed line: Tsirelson point $\rho = 2\sqrt{2}$.

*Proof.* By Corollary 9, $e = (1-\eta)/2 = (4-\rho)/8$. By Proposition 3, $G(\rho) = I(U_k : \hat{G}_k) = 1 - h_2(e)$. □

**Proposition 11** (Properties of $G$). *On* $[2,4]$*:*
*(a) $G$ is strictly increasing:*

$$G'(\rho) = \frac{1}{8\ln 2} \ln \frac{4+\rho}{4-\rho} > 0. \tag{7}$$

*(b) $G$ is strictly convex:*

$$G''(\rho) = \frac{1}{\ln 2\,(16-\rho^2)} > 0. \tag{8}$$

*Proof.* Let $e(\rho) = (4-\rho)/8$ so $e' = -1/8$.
(a) $G' = -h_2'(e)\cdot e' = h_2'(e)/8 = \frac{1}{8\ln 2}\ln\frac{1-e}{e} = \frac{1}{8\ln 2}\ln\frac{4+\rho}{4-\rho}$. For $\rho \in (2,4)$: $4+\rho > 4-\rho > 0$, so $G' > 0$.
(b) $G'' = \frac{d}{d\rho}\left[\frac{1}{8\ln 2}\ln\frac{4+\rho}{4-\rho}\right] = \frac{1}{8\ln 2}\left(\frac{1}{4+\rho} + \frac{1}{4-\rho}\right) = \frac{1}{\ln 2\,(16-\rho^2)} > 0$. □

**Remark 12** (Significance of convexity). *That $G'' > 0$ is crucial: benefits accelerate with increasing correlation. There are no diminishing returns. The optimal $\rho^*$ therefore cannot arise from the classical SRP mechanism of concave benefit against convex cost; it must come from a hard viability constraint.*

Figure 1 shows $G(\rho)$ and its derivatives.

## IV. INFORMATION CAUSALITY AND VIABILITY HORIZON

### A. Concatenated RAC protocol

Following Pawłowski et al. [6], the $(2^n, 1)$-RAC protocol is built by $n$-level recursive concatenation of the base RAC.

**Proposition 13** (Effective bias after concatenation). *After $n$ levels of concatenation, the effective bias is $\eta_{\text{eff}}(n) = \eta^n$.*

*Proof.* By induction. Base: $\eta_{\text{eff}}(1) = \eta$. Step: if level $n-1$ gives bias $\eta^{n-1}$, one additional RAC layer with box bias $\eta$ produces $\eta \cdot \eta^{n-1} = \eta^n$ (multiplicativity of bias in concatenated channels; see Ref. [6], Supplementary Lemma 1). □

The effective error at level $n$ is $e(n) = (1-\eta^n)/2$, and the success probability is $p_{\text{correct}}(n) = (1+\eta^n)/2$.

### B. Information Causality functional

**Definition 14.** *The total accessible information at concatenation level $n$ is*

$$I_{\text{total}}(n,\eta) = 2^n\left[1 - h_2\left(\frac{1-\eta^n}{2}\right)\right], \tag{9}$$

*using the symmetry of the protocol (all $2^n$ bits see the same effective BSC).*

**Axiom 15** (Information Causality [6]). *If Alice sends $m$ classical bits to Bob, then $I_{\text{total}} \le m$. In our protocol, $m = 1$:*

$$I_{\text{total}}(n,\eta) \le 1 \quad \forall n \ge 1. \tag{10}$$

### C. Asymptotic analysis

**Lemma 16.** *For $\delta \to 0^+$: $h_2(1/2-\delta) = 1 - 2\delta^2/\ln 2 - 4\delta^4/(3\ln 2) + O(\delta^6)$.*

**Proposition 17** (Asymptotic $I_{\text{total}}$). *For large $n$ and fixed $\eta \in (1/\sqrt{2}, 1)$ (i.e., excluding the PR-box endpoint $\eta = 1$ where $\eta^n \not\to 0$):*

$$I_{\text{total}}(n,\eta) \sim \frac{(2\eta^2)^n}{2\ln 2}. \tag{11}$$

*This diverges if $2\eta^2 > 1$ (i.e., $\rho > 2\sqrt{2}$) and converges to zero if $2\eta^2 < 1$. At the critical point $\eta = \eta_T = 1/\sqrt{2}$: $I_{\text{total}} \to 1/(2\ln 2) \approx 0.721 < 1$.*

*Proof.* Assume $\eta < 1$ so that $\eta^n \to 0$ and the expansion in Lemma 16 applies. Set $\delta_n = \eta^n/2$. Then $e(n) = 1/2 - \delta_n$ and by Lemma 16: $1 - h_2(e(n)) \approx 2\delta_n^2/\ln 2 = \eta^{2n}/(2\ln 2)$. Therefore $I_{\text{total}} \approx 2^n \cdot \eta^{2n}/(2\ln 2) = (2\eta^2)^n/(2\ln 2)$. □

### D. A uniform quadratic bound (global)

The asymptotic result identifies the critical scaling but does not upper-bound $I_{\text{total}}$ uniformly for all finite $n$. The following exact bound closes this gap.

**Lemma 18** (Global quadratic upper bound)**.** *For $z \in [0,1]$:*

$$1 - h_2\left(\frac{1-z}{2}\right) \le z^2. \tag{12}$$

*Proof.* Define $g(z) = 1 - h_2\left(\frac{1-z}{2}\right)$ for $z \in [0,1]$. Differentiating gives

$$g'(z) = \frac{1}{2\ln 2} \ln \frac{1+z}{1-z} = \frac{1}{\ln 2} \operatorname{artanh}(z). \tag{13}$$

Using the power series $\operatorname{artanh}(z) = \sum_{k\ge 0} \frac{z^{2k+1}}{2k+1}$ for $|z| < 1$ and integrating termwise,

$$g(z) = \frac{1}{\ln 2} \sum_{k\ge 0} \frac{z^{2k+2}}{(2k+1)(2k+2)} = \frac{z^2}{\ln 2} \sum_{k\ge 0} \frac{z^{2k}}{(2k+1)(2k+2)}. \tag{14}$$

For $z \in [0,1]$ the series is bounded by its value at $z = 1$, and

$$\sum_{k\ge 0} \frac{1}{(2k+1)(2k+2)} = \sum_{k\ge 0} \left(\frac{1}{2k+1} - \frac{1}{2k+2}\right) = \ln 2. \tag{15}$$

Therefore $g(z) \le z^2$. $\square$

**Corollary 19** (Global bound on $I_{\text{total}}$)**.** *For all $n \ge 1$ and $\eta \in [0,1]$:*

$$I_{\text{total}}(n,\eta) \le (2\eta^2)^n. \tag{16}$$

*Proof.* Apply Lemma 18 with $z = \eta^n$ and multiply by $2^n$:

$$I_{\text{total}}(n,\eta) = 2^n g(\eta^n) \le 2^n (\eta^n)^2 = (2\eta^2)^n.$$

$\square$

### E. Tsirelson dichotomy

**Theorem 20** (Tsirelson dichotomy via IC)**.** *Let $\eta_T = 1/\sqrt{2}$ ($\rho_T = 2\sqrt{2}$).*
*(a) For $\eta \le \eta_T$: $I_{\text{total}}(n,\eta) \le 1$ for all $n \ge 1$.*
*(b) For $\eta > \eta_T$: there exists finite $n_0(\eta)$ such that $I_{\text{total}}(n_0,\eta) > 1$.*
*(c) $\lim_{\eta\to\eta_T^+} n_0(\eta) = \infty$.*

*Proof.* (a) By Corollary 19: $I_{\text{total}}(n,\eta) \le (2\eta^2)^n$. If $\eta \le \eta_T$, then $2\eta^2 \le 1$ and $(2\eta^2)^n \le 1$ for all $n$.

(b) For $\eta > \eta_T$, Proposition 17 gives $I_{\text{total}}(n,\eta) \sim (2\eta^2)^n/(2\ln 2) \to \infty$, hence $I_{\text{total}}$ exceeds 1 for some finite $n_0(\eta)$.

(c) As $\eta \to \eta_T^+$: $\ln(2\eta^2) \to 0^+$, and the asymptotic estimate $n_0(\eta) \approx \ln(2\ln 2)/\ln(2\eta^2)$ diverges. $\square$

**Definition 21** (Viability horizon)**.**

$$n_{\max}(\rho) = \sup\{n \in \mathbb{N} : I_{\text{total}}(k,\rho/4) \le 1,\ \forall\, k \le n\}, \tag{17}$$

*with the convention $\sup\emptyset = 0$ (i.e., $n_{\max} = 0$ when $I_{\text{total}}(1,\rho/4) > 1$). By Theorem 20: $n_{\max} = \infty$ iff $\rho \le 2\sqrt{2}$.*

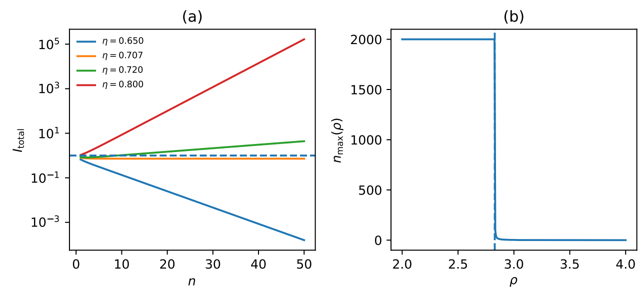


FIG. 2. (a) $I_{\text{total}}(n,\eta)$ for four values of $\eta$. Below Tsirelson ($\eta \le 1/\sqrt{2}$), $I_{\text{total}}$ remains below the IC bound; above, it diverges exponentially. (b) Viability horizon $n_{\max}(\rho)$: exact computation vs. analytic approximation $0.116/\varepsilon$.

**Proposition 22.** *For $\varepsilon = \eta - 1/\sqrt{2} \ll 1$:*

$$n_{\max} \approx \frac{\ln(2\ln 2)}{2\sqrt{2}\,\varepsilon} \approx \frac{0.116}{\varepsilon}. \tag{18}$$

Figure 2 shows $I_{\text{total}}(n,\eta)$ for several values and the $n_{\max}(\rho)$ curve.

## V. VPP FUNCTIONAL AND MAIN THEOREM

### A. Persistence framework

The Variational Principle of Persistence (VPP) formalizes the intuition: a system that extracts greater benefit from its environment persists longer, but only if it remains viable across the full spectrum of challenges it encounters. The optimal operating point maximizes benefit subject to viability.

**Definition 23** (Task environment)**.** *A task environment is a sequence of tasks $\{T_i\}_{i=1}^{\infty}$ where each $T_i$ has complexity $n_i$ (concatenation level required), drawn i.i.d. from distribution $P(n)$ on $\mathbb{N}$.*

**Definition 24** (Viability)**.** *A system with correlation $\rho$ is* viable *in task $T_i$ iff $n_i \le n_{\max}(\rho)$.*

**Definition 25** (Survival)**.** *The survival probability after $T$ tasks is $S(T|\rho) = [F(n_{\max}(\rho))]^T$, where $F(n) = \mathbb{P}(n_i \le n)$ is the CDF of $P$.*

**Definition 26** (VPP functional)**.** *The expected informational benefit weighted by long-horizon survival:*

$$\bar{W}(\rho) = G(\rho)\cdot S_\infty(\rho), \quad S_\infty(\rho) = \lim_{T\to\infty} [F(n_{\max}(\rho))]^T. \tag{19}$$

*For any $P$ with support not entirely contained in $\{1,\dots,n_{\max}\}$: $F(n_{\max}) < 1$, hence $S_\infty = 0$. For $n_{\max} = \infty$: $F(\infty) = 1$, hence $S_\infty = 1$.*

If $P$ has unbounded support:

$$\bar{W}(\rho) = \begin{cases} G(\rho) & \text{if } \rho \le 2\sqrt{2}, \\ 0 & \text{if } \rho > 2\sqrt{2}. \end{cases} \tag{20}$$

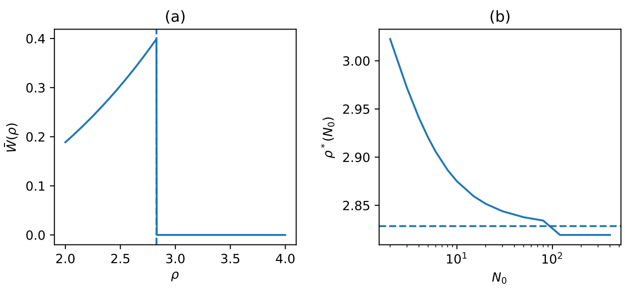


FIG. 3. (a) Main result: $\bar{W}(\rho) = G(\rho) \cdot S_\infty(\rho)$ with a viability wall at $\rho = 2\sqrt{2}$. The star marks the unique global optimum. (b) Convergence: $\rho^*(N_0) \to 2\sqrt{2}$ for bounded-support task distributions as $N_0 \to \infty$.

### B. Main result

**Theorem 27** (VPP–Tsirelson)**.** *Let $P(n)$ be any task distribution with unbounded support on $\mathbb{N}$. Then:*

$$\boxed{\rho^* = \arg\max_{\rho \in [2,4]} \bar{W}(\rho) = 2\sqrt{2}.} \tag{21}$$

*Proof.* Partition $[2, 4]$ into Region I ($\rho \leq 2\sqrt{2}$) and Region II ($\rho > 2\sqrt{2}$).

In Region I: $n_{\max}(\rho) = \infty$ (Theorem 20a), so $S_\infty = 1$ and $\bar{W} = G(\rho)$.

In Region II: $n_{\max}(\rho) < \infty$ (Theorem 20b). Since $P$ has unbounded support, $F(n_{\max}) < 1$, so $S_\infty = 0$ and $\bar{W} = 0$.

In Region I, $\bar{W}(\rho) = G(\rho)$ is strictly increasing (Proposition 11a). Therefore the maximum on $[2, 2\sqrt{2}]$ is attained at $\rho = 2\sqrt{2}$.

Finally, $\bar{W}(2\sqrt{2}) > 0 = \bar{W}(\rho)$ for all $\rho > 2\sqrt{2}$, so $\rho^* = 2\sqrt{2}$. □

**Corollary 28** (Universality)**.** *The result is independent of: the specific $P(n)$ (if unbounded support), the time horizon $T$ (as $T \to \infty$), and any relative channel weights.*

**Corollary 29** (Bounded convergence)**.** *If $P$ has support on $\{1, \dots, N_0\}$, then $\rho^*(N_0) \to 2\sqrt{2}$ as $N_0 \to \infty$.*

**Corollary 30** (Generalized SRP/PRS)**.** *The theorem instantiates a generalized form of the Systemic Reduction Paradox. In the* standard *SRP, a concave benefit $G$ competes against a convex cost $C$, producing a smooth interior optimum where $G' = \lambda C'$. Here, a convex benefit meets a hard viability wall, producing a corner optimum at the constraint boundary.*

**Remark 31** (Robustness: not an artifact of the "hard death" rule)**.** *The step-function form of $S_\infty$ might suggest the result is an artifact of a "one failure kills" assumption. It is not. Under soft survival models where failure probability is exponentially penalized over a long horizon, the optimum converges to the same boundary because the key driver is the sharp transition in $n_{\max}(\rho)$ at $\rho = 2\sqrt{2}$.*

### C. Lagrangian equivalence

The problem is equivalent to:

$$\rho^* = \arg\max_{\rho \in [2,4]} G(\rho) \quad \text{s.t.} \quad \Phi(\rho) \leq 1, \tag{22}$$

where $\Phi(\rho) = \sup_n I_{\text{total}}(n, \rho/4)$. The feasible set is $[2, 2\sqrt{2}]$. KKT conditions give $G'(\rho^*) = \mu^* \Phi'(\rho^*)$ with $\mu^* > 0$ (active constraint), confirming the corner-solution structure.

## VI. TASK DISTRIBUTIONS: CLOSURE ARGUMENTS

By Corollary 28, the result holds for any $P(n)$ with unbounded support. We now argue that this is a natural condition—not an exotic assumption—under standard physical and computational principles. If it is relaxed (e.g., by imposing a finite complexity ceiling), the optimum converges to $2\sqrt{2}$ as the ceiling grows (Corollary 29).

### A. Algorithmic argument

Under the universal prior [16, 17], the probability of a task of complexity $n$ satisfies $P(n) \geq 2^{-K(n)+O(1)}$, where $K(n)$ is the Kolmogorov complexity. Since $K(n) \approx \log_2 n + O(1)$ for most $n$, we get $P(n) = \Omega(1/n)$—a harmonic distribution with unbounded support. No finite cutoff is consistent with Turing completeness.

### B. Maximum entropy argument

Given only $\mathbb{E}[n] = \mu < \infty$, the maximum entropy distribution on $\mathbb{N}$ subject to this constraint is geometric: $P(n) = (1 - e^{-\lambda})e^{-\lambda(n-1)}$ [18]. For any finite $\mu$, this distribution has unbounded support.

### C. Thermodynamic argument

Mapping task complexity to energy $E(n) = \alpha n$, Boltzmann statistics at temperature $T > 0$ give $P(n) \propto g(n)e^{-\beta\alpha n}$, which has unbounded support for any $T > 0$ and sub-exponential degeneracy $g(n)$. Bounded support requires $T = 0$, which is unreachable by the Third Law of thermodynamics.

**Remark 32** (Unbounded support as a natural condition)**.** *In any setting with (i) Turing-complete computation, (ii) positive temperature ($T > 0$), and (iii) no arbitrary complexity cutoff, the arguments above yield distributions with unbounded support. With finite resources the effective support is bounded, but Corollary 29 guarantees $\rho^*(N_0) \to 2\sqrt{2}$ as the resource ceiling $N_0$ grows.*

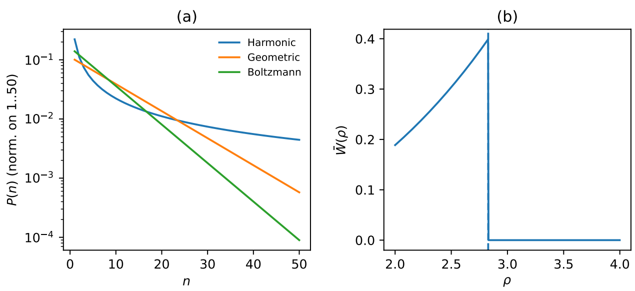


FIG. 4. (a) Representative task distributions derived from algorithmic, MaxEnt, and thermodynamic arguments (all with unbounded support). (b) The VPP optimum $\rho^* = 2\sqrt{2}$ is identical for any unbounded $P(n)$.

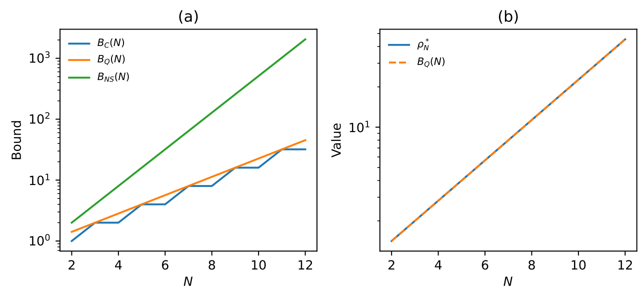


FIG. 5. (a) Classical, quantum, and no-signaling bounds for the $N$-party Mermin inequality. (b) Conditional VPP optimum $\rho^*_N$ matches the quantum bound $B_Q(N)$ for the tested range (illustrative).

## VII. COMMUNICATION COMPLEXITY AS INDEPENDENT CONSTRAINT

Van Dam [4] and Brassard et al. [5] showed that superquantum correlations trivialize communication complexity: with sufficiently strong correlations, the communication cost of computing any Boolean function collapses to $O(1)$ for large inputs. The tightest bound obtained from CC arguments is $\rho \lesssim 3.27$ [5], which is strictly weaker than the IC wall at $2\sqrt{2}$.

Define the CC viability set:

$$\mathcal{K}_{\mathrm{CC}} = [2, \rho_{\mathrm{CC}}], \qquad \rho_{\mathrm{CC}} \approx 3.27, \tag{23}$$

and the IC viability set:

$$\mathcal{K}_{\mathrm{IC}} = [2, 2\sqrt{2}]. \tag{24}$$

Since $2\sqrt{2} < \rho_{\mathrm{CC}}$, the combined viability set is $\mathcal{K} = \mathcal{K}_{\mathrm{IC}}$. IC strictly subsumes CC; the interval $(2\sqrt{2}, \rho_{\mathrm{CC}})$ is a "phantom zone" preserved by CC but forbidden by IC.

## VIII. MULTIPARTITE EXTENSION (SKETCH)

For $N$ parties, the Mermin–Klyshko inequality [14] has bounds: $B_C(N) = 2^{\lfloor (N-1)/2 \rfloor}$ (classical), $B_Q(N) = 2^{(N-1)/2}$ (quantum/Tsirelson), $B_{NS}(N) = 2^{N-1}$ (no-signaling). The VPP argument extends to the Mermin class under two structural assumptions: (i) the $N$-party RAC benefit is monotone increasing, and (ii) a bipartition-wise IC-type constraint creates a wall at $B_Q(N)$, as holds for bipartite-decomposable inequalities [15].

**Proposition 33** (Multipartite VPP—conditional)**.** *If (i)–(ii) hold for the $N$-party Mermin inequality, then for any task distribution with unbounded support:*

$$\rho^*_N = B_Q(N) = 2^{(N-1)/2}. \tag{25}$$

**Remark 34** (Limitations)**.** *Gallego, Wolf, and Acín [15] proved that no set of bipartite information principles can fully characterize the multipartite quantum correlation set. Proposition 33 is limited to the bipartite-decomposable class (Mermin). For genuinely multipartite inequalities, a natively multipartite viability principle would be required.*

TABLE I. Computational validation summary.

| Proposition | Method | Result |
|---|---|---|
| 3: BSC capacity | Monte Carlo $10^6$ | error $< 0.001$ |
| 4: Convexity | $f''$ evaluation | $f'' > 0$ everywhere |
| 8: RAC $p_{\mathrm{correct}}$ | Monte Carlo $5 \times 10^5$ | error $< 0.001$ |
| 10: $G(\rho)$ | analytic vs. numeric | exact match |
| 11: $G', G''$ | analytic vs. numeric | $< 10^{-6}$ |
| 17: Asymptotics | $n = 5$–$50$ | rel. error $< 0.01$ |
| 18: Global bound | $z \in [0, 1]$ grid | satisfied |
| 20: Dichotomy | sweep $\rho \in [2, 4]$ | verified |
| 22: $n_{\max}$ approx. | $\varepsilon \in [10^{-3}, 0.05]$ | rel. error $< 5\%$ |
| 27: Main theorem | multiple $P(n)$ | $\Delta < 0.002$ |
| 29: Convergence | $N_0 \in [2, 500]$ | $\rho^* \to 2\sqrt{2}$ |

## IX. COMPUTATIONAL VALIDATION

All propositions and theorems are verified numerically using a stable implementation that avoids catastrophic cancellation in $1 - h_2(e)$ for $e \to 1/2$ via series expansion. Table I summarizes key tests.

## X. INFORMATION GEOMETRY OF THE VIABILITY WALL

### A. Classical Fisher information of the BSC family

The RAC protocol maps the CHSH parameter $\rho$ to an effective BSC with error $e(\rho) = (4 - \rho)/8$. Under uniform

TABLE II. VPP functional values (illustrative).

| $\rho$ | $\eta$ | $G(\rho)$ | $n_{\max}$ | $S_\infty$ | $\bar{W}$ |
|---|---|---|---|---|---|
| 2.000 | 0.500 | 0.189 | $\infty$ | 1 | 0.189 |
| 2.600 | 0.650 | 0.307 | $\infty$ | 1 | 0.307 |
| 2.800 | 0.700 | 0.387 | $\infty$ | 1 | 0.387 |
| $2\sqrt{2}$ | $1/\sqrt{2}$ | 0.399 | $\infty$ | 1 | 0.399 |
| 2.830 | 0.708 | 0.400 | 293 | 0 | 0 |
| 2.900 | 0.725 | 0.429 | 6 | 0 | 0 |
| 3.000 | 0.750 | 0.456 | 2 | 0 | 0 |
| 4.000 | 1.000 | 1.000 | 0 | 0 | 0 |

input $U$, the BSC output $\hat{G}$ has the marginal distribution

$$p_\rho(\hat{g}) = \begin{cases} 1 - e(\rho) = \frac{4+\rho}{8}, & \hat{g} = u \ \text{(correct)}, \\ e(\rho) = \frac{4-\rho}{8}, & \hat{g} \neq u \ \text{(error)}. \end{cases} \tag{26}$$

This defines a one-parameter Bernoulli family indexed by $\rho$.

**Definition 35** (Fisher information of the BSC family)**.** *The Fisher information of the marginal output family $\{p_\rho\}$ for the parameter $\rho$ is*

$$\mathcal{I}_F(\rho) = \mathbb{E}\left[\left(\frac{\partial}{\partial\rho}\log p_\rho(\hat{G})\right)^2\right]. \tag{27}$$

**Theorem 36** (Fisher information of the RAC channel)**.**

$$\mathcal{I}_F(\rho) = \frac{1}{16-\rho^2}. \tag{28}$$

*Proof.* The error probability $e(\rho) = (4-\rho)/8$ has derivative $e'(\rho) = -1/8$. Reparameterizing the Bernoulli family by $\rho$ gives:

$$\mathcal{I}_F(\rho) = \frac{[e'(\rho)]^2}{e(\rho)[1-e(\rho)]} = \frac{1/64}{\frac{4-\rho}{8}\cdot\frac{4+\rho}{8}} = \frac{1}{16-\rho^2}. \tag{29}$$

□

### B. Benefit–Metric Correspondence

**Theorem 37** (Benefit–Metric Correspondence)**.** *Let $G(\rho) = 1 - h_2\left(\frac{4-\rho}{8}\right)$ and $\mathcal{I}_F(\rho)$ the Fisher information above. Then for all $\rho \in (2,4)$:*

$$\boxed{G''(\rho) = \frac{\mathcal{I}_F(\rho)}{\ln 2}.} \tag{30}$$

*Proof.* From Proposition 11(b): $G''(\rho) = 1/[\ln 2\,(16-\rho^2)]$. From Theorem 36: $\mathcal{I}_F(\rho) = 1/(16-\rho^2)$. The ratio is $1/\ln 2$. □

### C. Metric accumulation under concatenation

At concatenation depth $n$, each leaf experiences bias $\delta_n = \eta^n/2$. Near zero bias, the mutual information per leaf is quadratic in $\delta_n$, matching the local Fisher metric norm; summing over $2^n$ leaves yields $(2\eta^2)^n$ scaling. This yields a geometric stability condition: the metric pullback is stable iff its spectral radius $\lambda(\eta) = 2\eta^2 \leq 1$.

**Theorem 38** (Tsirelson as metric criticality)**.** *Let $\lambda(\eta) = 2\eta^2$ be the one-step growth factor of the accumulated Fisher–Rao metric norm over the concatenation tree. Then IC holds for all $n$ iff $\lambda(\eta) \leq 1$, i.e., $\rho \leq 2\sqrt{2}$. The value $\rho = 2\sqrt{2}$ is the unique marginally stable point $\lambda = 1$.*

*Proof.* By Lemma 18, $I_{\text{total}}(n,\eta) = 2^n g(\eta^n) \leq 2^n(\eta^n)^2 = (2\eta^2)^n = \lambda^n$. Hence if $\lambda \leq 1$, IC holds for all $n$. If $\lambda > 1$, Proposition 17 implies $I_{\text{total}}$ diverges and exceeds 1 at finite depth. □

## XI. DISCUSSION

### A. What is the fundamental contribution?

The key contribution is conceptual and structural: IC becomes a *viability wall* inside a variational persistence functional, turning “forbidden beyond $2\sqrt{2}$” into “persistence-optimal at $2\sqrt{2}$” for agents facing unbounded task difficulty. Mathematically, this is driven by three facts: (i) the benefit $G(\rho)$ is strictly increasing and strictly convex (no diminishing returns), (ii) concatenation produces the eigenvalue $\lambda(\eta) = 2\eta^2$ controlling stability of information accumulation, and (iii) unbounded task support forces $S_\infty(\rho)$ into a step-like viability filter.

The generalized SRP/PRS is the interpretive payoff: the optimum is not created by a cost curve but by a hard constraint. The Benefit–Metric Correspondence adds the geometric payoff: the same Fisher–Rao metric determines both benefit curvature and the concatenation stability.

### B. Limitations

IC has not been proven to characterize the full quantum set. For CHSH it reproduces the quantum boundary, but it does not (currently) single out $Q$ vs. $\tilde{Q}$ beyond CHSH scenarios [19]. The multipartite extension is conditional (Remark 34). Finally, our geometric correspondence uses *classical* Fisher information of the effective BSC; a direct link to the quantum Fisher information or quantum geometric tensor of the underlying state remains open.

### C. Open questions

Can VPP yield novel constraints beyond those already captured by device-independent principles? Can one design experiments where a measurable resource "cost" of approaching the viability wall is quantified? Can a natively multipartite viability principle be formulated to evade the Gallego–Wolf–Acín limitation?

## XII. CONCLUSION

We derived the Tsirelson point $\rho^* = 2\sqrt{2}$ as the unique optimum of the VPP functional $\bar{W}(\rho) = G(\rho)\, S_\infty(\rho)$, with $G$ fixed by Shannon/RAC and viability fixed by IC. The generalized SRP/PRS mechanism is "accelerating benefit + hard viability wall". The Benefit–Metric Correspondence $G'' = \mathcal{I}_F / \ln 2$ and the metric criticality theorem provide a unified information-geometric interpretation: the same Fisher–Rao metric governs both benefit curvature and the viability boundary via the eigenvalue $\lambda = 2\eta^2$.


## ACKNOWLEDGMENTS

This work was developed within the Variational Principle of Persistence program at Paradox Systems R&D.


### Appendix A: Reproducibility: exact numerical procedure

To guarantee reproducibility, we provide: (i) a Colab notebook that generates Figures 1–5, plus CSV outputs for key curves, (ii) a stable implementation of $1 - h_2(e)$ near $e = 1/2$ using a series expansion, and (iii) a one-command recipe to regenerate all PDFs and compile the manuscript.

### Appendix B: Numerical stability

For $e \to 1/2$, the direct formula for $1 - h_2(e)$ suffers catastrophic cancellation in floating-point arithmetic. We use the Taylor expansion $1 - h_2(1/2 - \delta) = \frac{2\delta^2}{\ln 2}\left(1 + \frac{2}{3}\delta^2 + \frac{16}{15}\delta^4 + \cdots\right)$ for $|\delta| < 10^{-2}$, switching to the direct formula otherwise.


[1] J. F. Clauser, M. A. Horne, A. Shimony, and R. A. Holt, "Proposed experiment to test local hidden-variable theories," *Phys. Rev. Lett.* **23**, 880 (1969).

[2] B. S. Cirel'son, "Quantum generalizations of Bell's inequality," *Lett. Math. Phys.* **4**, 93 (1980).

[3] S. Popescu and D. Rohrlich, "Quantum nonlocality as an axiom," *Found. Phys.* **24**, 379 (1994).

[4] W. van Dam, "Implausible consequences of superstrong nonlocality," *Nat. Comput.* **12**, 9 (2013); preprint arXiv:quant-ph/0501159 (2005).

[5] G. Brassard, H. Buhrman, N. Linden, A. A. Méthot, A. Tapp, and F. Unger, "Limit on nonlocality in any world in which communication complexity is not trivial," *Phys. Rev. Lett.* **96**, 250401 (2006).

[6] M. Pawłowski, T. Paterek, D. Kaszlikowski, V. Scarani, A. Winter, and M. Żukowski, "Information causality as a physical principle," *Nature* **461**, 1101 (2009).

[7] A. Carmi and D. Moskovich, "Tsirelson's bound prohibits communication through a disconnected channel," *Entropy* **20**, 151 (2018).

[8] N. Linden, S. Popescu, A. J. Short, and A. Winter, "Quantum nonlocality and beyond: Limits from nonlocal computation," *Phys. Rev. Lett.* **99**, 180502 (2007).

[9] M. Navascués and H. Wunderlich, "A glance beyond the quantum model," *Proc. R. Soc. A* **466**, 881 (2010).

[10] A. Cabello, "Simple explanation of the quantum violation of a fundamental inequality," *Phys. Rev. Lett.* **110**, 060402 (2013).

[11] A. Carmi and E. Cohen, "Relativistic independence bounds nonlocality," *Sci. Adv.* **5**, eaav8370 (2019).

[12] C. E. Shannon, "A mathematical theory of communication," *Bell Syst. Tech. J.* **27**, 379 (1948).

[13] W. Oughton and C. G. Timpson, "Bounding quantum correlations: The role of the Shannon information in the information causality principle," *Entropy* **26**, 562 (2024).

[14] N. D. Mermin, "Extreme quantum entanglement in a superposition of macroscopically distinct states," *Phys. Rev. Lett.* **65**, 1838 (1990).

[15] R. Gallego, L. E. Würflinger, A. Acín, and M. Navascués, "Quantum correlations require multipartite information principles," *Phys. Rev. Lett.* **107**, 210403 (2011).

[16] R. J. Solomonoff, "A formal theory of inductive inference. Part I," *Inf. Control* **7**, 1 (1964).

[17] L. A. Levin, "Laws of information conservation (nongrowth) and aspects of the foundation of probability theory," *Probl. Peredachi Inf.* **10**, 30 (1974).

[18] E. T. Jaynes, "Information theory and statistical mechanics," *Phys. Rev.* **106**, 620 (1957).

[19] M. Navascués, Y. Guryanova, M. J. Hoban, and A. Acín, "Almost quantum correlations," *Nat. Commun.* **6**, 6288 (2015).

[20] A. I. Khinchin, *Mathematical Foundations of Information Theory* (Dover, New York, 1957).

[21] D. K. Faddeev, "On the concept of entropy of a finite probabilistic scheme," *Uspekhi Mat. Nauk* **11**(1), 227 (1956).