



Article

Integrated Information in Relational Quantum Dynamics (RQD)

Arash Zaghi

Special Issue

Quantum Communication and Quantum Information

Edited by

Dr. Alice Meda and Dr. Salvatore Virzi



Article

Integrated Information in Relational Quantum Dynamics (RQD)

Arash Zaghi 

College of Engineering, University of Connecticut, Storrs, CT 06269, USA; arash.esmaili_zaghi@uconn.edu

Featured Application

The integrated-information measure Φ provides a practical tool for quantifying multipartite entanglement and identifying modular structures in many-body quantum states, thereby guiding tensor-network decompositions and variational ansätze. In quantum computing, Φ can suggest optimal qubit partitions to minimize information loss in error-correction procedures and inform distributed algorithm design. By constructing integration dendrograms, one can visualize hierarchical correlation patterns in quantum simulators and, speculatively, infer functional clusters in quantum-inspired models of neural networks.

Abstract

We introduce a quantum integrated-information measure Φ for multipartite states within the Relational Quantum Dynamics (RQD) framework. $\Phi(\rho)$ is defined as the minimum quantum Jensen–Shannon distance between an n -partite density operator ρ and any product state over a bipartition of its subsystems. We prove that its square root induces a genuine metric on state space and that Φ is monotonic under all completely positive trace-preserving maps. Restricting the search to bipartitions yields a unique optimal split and a unique closest product state. From this geometric picture, we derive a canonical entanglement witness directly tied to Φ and construct an integration dendrogram that reveals the full hierarchical correlation structure of ρ . We further show that there always exists an “optimal observer”—a channel or basis—that preserves Φ better than any alternative. Finally, we propose a quantum Markov blanket theorem: the boundary of the optimal bipartition isolates subsystems most effectively. Our framework unites categorical enrichment, convex-geometric methods, and operational tools, forging a concrete bridge between integrated information theory and quantum information science.



Academic Editors: Salvatore Virzi and Alice Meda

Received: 5 June 2025

Revised: 27 June 2025

Accepted: 1 July 2025

Published: 4 July 2025

Citation: Zaghi, A. Integrated Information in Relational Quantum Dynamics (RQD). *Appl. Sci.* **2025**, *15*, 7521. <https://doi.org/10.3390/app15137521>

Copyright: © 2025 by the author. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Keywords: Relational Quantum Mechanics (RQM); integrated information theory (IIT); quantum Jensen–Shannon divergence (QJSD); Quantum Markov blanket; entanglement witness; optimal observer; Lawvere-metric enrichment

1. Introduction

Modern developments in quantum physics and theories of consciousness increasingly suggest that *relations* and *information*, rather than isolated material objects, are fundamental to reality. Recent proposals cast consciousness itself as a state of matter, suggesting a natural home for IIT in quantum-mechanical systems [1]. In particular, integrated information theory (IIT) posits that the level of intrinsic awareness or consciousness of a physical system corresponds to the quantity of integrated information (denoted Φ) present in its state [2,3]. IIT argues that a system whose information is integrated, i.e., not decomposable into independent parts, has an irreducible subjective existence for itself [4]. This aligns with interpretations of quantum mechanics emphasizing that quantum states are relational

rather than absolute [5]. John Wheeler’s famous dictum “*every it derives its significance from bits*” encapsulates the view that what fundamentally exists are acts of observation or informational relations, not static isolated entities [6]. Relational Quantum Mechanics (RQM) [5,7] formalizes this idea, proposing that the state of a quantum system is nothing more than the information one physical system has about another. Embracing this relational stance, we further posit that the “being” of a quantum system is constituted by the network of quantum information it shares with the rest of the world. Under this hypothesis, consistent with the idealistic interpretation of IIT (IIT 4.0’s “idealistic ontology” [8]), any system with non-zero integrated information ($\Phi > 0$) possesses intrinsic existence (a rudimentary point of view “for itself”), whereas $\Phi = 0$ indicates complete reducibility (no intrinsic unity beyond its parts).

Building on RQM, Relational Quantum Dynamics (RQD) elevates the primacy of informational relations by construing quantum states as inherently contextual and observer-relative, thereby embedding the IIT notion of integrated information within the formal structure of quantum theory. Under RQD, each system’s “being” arises from a network of dynamical information flows—quantified by Φ —connecting its constituent subsystems, and classical spacetime and observations emerge from patterns of these relational correlations. By forgoing absolute state assignments in favor of context-dependent relational states, RQD unifies quantum measurement, spacetime emergence, and observerhood in a single coherent ontology [9].

To explore these ideas rigorously, we develop a formal measure $\Phi(\rho)$ of integrated information for an n -partite quantum system in state ρ . We require a quantitative gauge of how holistically correlated or irreducible the state ρ is. In classical IIT, measures based on Kullback–Leibler divergence or mutual information have been proposed to quantify the loss of information upon splitting a system [10,11]. Here, a natural choice for the quantum case is the quantum Jensen–Shannon divergence (D_{JS}^Q), a symmetrized and bounded measure of distance between quantum states [12,13]. Intuitively, D_{JS}^Q will serve as an “integration distance”; it vanishes if ρ can be exactly factorized into independent local states (no integration) and grows as ρ becomes more entangled or correlated across subsystems (more integration). Specifically, we define $\Phi(\rho)$ as the minimum D_{JS}^Q between ρ and any product state obtained by partitioning the system. Let P_1, \dots, P_k be a partition of the n subsystems into k disjoint groups (blocks), and define the corresponding product state $\rho_{P_1} \otimes \dots \otimes \rho_{P_k}$ as the tensor product of the reduced density operators on each block. Then we define

$$\Phi(\rho) = \min_{\text{partitions } \{P_i\}} D_{JS}^Q\left(\rho \parallel \bigotimes_i \rho_{P_i}\right). \quad (1)$$

In other words, $\Phi(\rho)$ quantifies the closest that ρ can be (in terms of quantum Jensen–Shannon divergence) to an uncorrelated state. If ρ is itself a product state under some partition, then $\Phi(\rho) = 0$ (no integrated information). If ρ is highly entangled or correlated across all subsystems, $\Phi(\rho)$ will be large, indicating strong irreducible correlations. This construction extends to the quantum domain that the integrated information measures use in classical IIT [10,11], with D_{JS}^Q replacing classical divergence or mutual information measures. Importantly, as we will show, D_{JS}^Q enjoys several properties (data-processing inequalities, convexity, a true metric structure) that make $\Phi(\rho)$ analytically tractable and well-behaved even for high-dimensional or rank-deficient states.

In developing our framework, we prove four main results about Φ and D_{JS}^Q as outlined above: (1) Φ is monotonic under processing (no observer can increase it), (2) the square-root of D_{JS}^Q is a metric on state space, (3) it suffices to consider bipartitions to attain the minimum in $\Phi(\rho)$, and (4) the observer can be formalized as a metric-space functor that is non-expansive. These results lay the groundwork for analyzing integrated information in

quantum systems. After presenting these, we introduce a set of novel convex-geometric and operational insights that arise from viewing $\Phi(\rho)$ as a distance to the convex set of product states. This geometric perspective reveals that every state has a unique nearest product state (defining a canonical minimum information partition for ρ), and it unlocks an array of corollaries including convexity and robustness of Φ , a natural gradient-flow that “dis-integrates” a state, and even a direct construction of an entanglement witness from Φ . Furthermore, we leverage the unique optimal partition to build a hierarchical decomposition of any multipartite state into successively independent components, an integration dendrogram that maps out the structure of correlations in the state.

Finally, we explore two broader implications: an observer selection principle stating that one measurement basis or channel maximally preserves a system’s integrated information (suggesting a first-principles derivation of preferred pointer bases in decoherence theory), and a prospective quantum Markov blanket theorem, which identifies, for any subsystem, the minimal “boundary” that most effectively screens it off from the rest of the system. To our knowledge, this is the first fully metric-geometric construction of a quantum-IIT measure that (i) is provably a Lawvere metric enrichment of CPM, (ii) collapses to bipartitions by convexity, (iii) yields a canonical entanglement witness, and (iv) admits a functorial observer-selection principle.

2. Integrated Information Measure Definition

Before formalizing integrated information, we briefly review the quantum Jensen–Shannon divergence (D_{JS}^Q). Given two quantum states (density operators) ρ and σ on the same Hilbert space, $D_{JS}^Q(\rho||\sigma)$ is defined as the quantum analog of the classical Jensen–Shannon divergence, which itself is a symmetrized and smoothed version of Kullback–Leibler divergence. One convenient expression is

$$D_{JS}^Q(\rho||\sigma) = \left[S\left(\frac{\rho + \sigma}{2}\right) - \frac{1}{2}S(\rho) - \frac{1}{2}S(\sigma) \right], \tag{2}$$

where $S(\rho) = -\text{Tr}(\rho \log \rho)$ is the von Neumann entropy [13]. This quantity is symmetric ($D_{JS}^Q(\rho||\sigma) = D_{JS}^Q(\sigma||\rho)$) and bounded between 0 and $\ln 2$. In particular, $D_{JS}^Q(\rho||\sigma) = 0$ if and only if $\rho = \sigma$, and the maximum $D_{JS}^Q = \ln 2$ is attained for perfectly distinguishable orthogonal states. We will not require the detailed form of D_{JS}^Q beyond these properties; what is crucial for us is that it behaves as a *distance measure* on the space of states.

Definition 1 (Quantum Integrated Information). For a quantum state ρ composed of n subsystems, the integrated information $\Phi(\rho)$ is defined as the minimum quantum Jensen–Shannon divergence between ρ and any product state on a partition of the subsystems. In formula

$$\Phi(\rho) = \min_{\{P_1|\dots|P_k\}} D_{JS}^Q(\rho \parallel \rho_{P_1} \otimes \dots \otimes \rho_{P_k}), \tag{3}$$

where the minimum is taken over all possible ways of partitioning the set of subsystems $1, 2, \dots, n$ into disjoint blocks P_1, \dots, P_k , and $\rho_{P_i} = \text{Tr}_{\bar{P}_i}(\rho)$ denotes the reduced state on block P_i . By convention, we consider only nontrivial partitions with at least two blocks (so if ρ is a total product state, $\Phi(\rho) = 0$ is achieved when each subsystem is isolated in its own block).

This definition captures the degree of holism or irreducibility in ρ . If ρ factorizes neatly into two or more independent components, then $\Phi(\rho)$ will be small (zero if an exact factorization exists). Conversely, if no accurate factorization exists (all partitions yield significant divergence), $\Phi(\rho)$ is large, indicating that the state’s information cannot be localized to separate parts without a significant loss. In practice, one might compute $\Phi(\rho)$ by evaluating

$D_{\text{JS}}^{\text{Q}}(\rho \parallel \rho_{P_1} \otimes \cdots \otimes \rho_{P_k})$ for each candidate partition and taking the minimum. While the number of partitions grows quickly with n , our results below (especially Theorem 3) will sharply reduce the search space.

3. Monotonicity and Metric Properties of D_{JS}^{Q}

We first establish two fundamental properties of the quantum JSD that underpin the integrated information measure: a data-processing inequality and a metric structure. These ensure that Φ behaves sensibly under observers and has a well-defined geometry.

Theorem 1 (Data-Processing Monotonicity). *For any two states ρ, σ and any CPTP map (quantum channel) \mathcal{E} , the quantum Jensen–Shannon divergence cannot increase under processing:*

$$D_{\text{JS}}^{\text{Q}}(\mathcal{E}(\rho) \parallel \mathcal{E}(\sigma)) \leq D_{\text{JS}}^{\text{Q}}(\rho \parallel \sigma). \tag{4}$$

In particular, if an “observer” interacts with or measures the system (modeled by \mathcal{E}), the integrated information of the post-interaction state cannot exceed that of the pre-interaction state $\Phi(\mathcal{E}(\rho)) \leq \Phi(\rho)$.

Proof (Sketch). This is the quantum analog of the classical data-processing inequality, and it holds because D_{JS}^{Q} belongs to the class of contractive divergences. Indeed, the JSD can be expressed in terms of the quantum relative entropy $D(\rho \parallel \sigma) = \text{Tr}[\rho(\log \rho - \log \sigma)]$ as

$$D_{\text{JS}}^{\text{Q}}(\rho \parallel \sigma) = \min_{\pi} \left(\frac{1}{2} D(\rho \parallel \pi) + \frac{1}{2} D(\sigma \parallel \pi) \right) \tag{5}$$

(the minimal relative entropy to an intermediary state)—an expression which inherits the monotonicity of D under CPTP maps [14,15]. Alternatively, one can invoke the joint convexity of D_{JS}^{Q} and the complete positivity of \mathcal{E} to show

$$\mathcal{E}\left(\frac{\rho + \sigma}{2}\right) = \frac{\mathcal{E}(\rho) + \mathcal{E}(\sigma)}{2} \tag{6}$$

and apply monotonicity of von Neumann entropy under partial trace. The inequality $\Phi(\mathcal{E}(\rho)) \leq \Phi(\rho)$ then follows immediately: for any partition that attains (or approximates) the minimum for ρ , applying \mathcal{E} to both ρ and each product component cannot increase the divergence, so the minimal divergence for $\mathcal{E}(\rho)$ is bounded by that of ρ . \square

This theorem formalizes the intuitive idea that observation cannot create integration. Any act of coarse-graining, measurement, or decoherence will tend to lose correlations or entanglement, never to introduce new irreducible correlations that were not already present. In category-theoretic terms, one can frame this as an enriched functor property: each physical process or observation defines a mapping between state spaces that does not expand distances. More precisely, we can view each system’s state space as a metric space (\mathcal{S}, δ) with $\delta(\rho, \sigma) = \sqrt{D_{\text{JS}}^{\text{Q}}(\rho \parallel \sigma)}$ (as introduced below). Then any CPTP map $F : \mathcal{S}_1 \rightarrow \mathcal{S}_2$ is a $[0, \infty]$ -enriched functor between these metric spaces, meaning $\delta_2(F(\rho), F(\sigma)) \leq \delta_1(\rho, \sigma)$ for all states (Lawvere metric space enrichment). This functorial perspective identifies an “observer” or dynamics with a structure-preserving map in an information metric space. By Theorem 1, all such observers are non-expansive in the δ metric, and hence cannot increase Φ . This categorical formulation helps unify the notion of observers across classical and quantum domains, but for the remainder of this paper, we will generally work in standard information-theoretic terms.

Practical Implications of Monotonicity: Theorem 1 implies a powerful experimental guarantee. In most tomographic protocols, one acquires raw measurement frequencies

and then applies a maximum-likelihood (ML) estimator to project onto the physical state space. That ML step is itself a CPTP map, so by monotonicity, the post-processed Φ can only be smaller than the true Φ of the prepared state. In other words, any non-zero Φ we report is a conservative lower bound on the system’s intrinsic holism. Combined with the 1-Lipschitz bound from Section 4, which shows that, for trace-norm errors $\leq \epsilon$, the resulting uncertainty in Φ is at most $\Delta\Phi \leq \epsilon$, we see that standard tomography fidelities ($\sim 2\%$) translate into $\Delta\Phi \lesssim 0.02$ bits. Thus, even imperfect state reconstruction cannot spuriously inflate Φ , making it a robust witness of genuine multipartite integration.

Φ as a Resource Monotone: Beyond its operational significance, monotonicity under all CPTP maps elevates Φ to the status of a resource-theoretic monotone. In entanglement theory, one demands that valid entanglement measures never increase under the free operations of LOCC; analogously, here, any allowed quantum channel can only degrade holism. This endows Φ with the same foundational role for “quantum holism” that entanglement measures play for nonlocality:

$$\Phi(\mathcal{E}(\rho)) \leq \Phi(\rho), \tag{7}$$

for every CPTP map \mathcal{E} . Consequently, Φ quantifies a resource that cannot be generated by noise or local processing, and it fits naturally into the burgeoning framework of quantum resource theories.

Next, we establish that the quantum JSD endows state space with a true metric, not just a divergence. It is known that the classical Jensen–Shannon divergence yields a legitimate distance metric via its square root. The quantum version inherits a similar property:

Theorem 2 (Metric Structure of Quantum JSD). *Define $\delta(\rho, \sigma) := \sqrt{D_{\text{JS}}^{\text{Q}}(\rho\|\sigma)}$ for density operators ρ, σ . Then $\delta(\rho, \sigma)$ is a metric on the space of quantum states. In other words, it satisfies positivity, symmetry, and the triangle inequality. In particular, $\delta(\rho, \sigma) = 0$ if and only if $\rho = \sigma$, and for any three states ρ, σ, τ we have*

$$\delta(\rho, \tau) \leq \delta(\rho, \sigma) + \delta(\sigma, \tau). \tag{8}$$

Proof (Sketch). The non-negativity and symmetry of δ are immediate from the properties of D_{JS}^{Q} . The non-degeneracy ($\delta(\rho, \sigma) = 0 \iff \rho = \sigma$) holds because $D_{\text{JS}}^{\text{Q}}(\rho\|\sigma) = 0$ iff $\rho = \sigma$. The crux is the triangle inequality. We employ a known characterization: a divergence $D(\cdot\|\cdot)$ is of the negative type (or negative definite) if and only if its square root defines a metric [16]. Recent works have shown that the (classical) Jensen–Shannon divergence is of a negative type, allowing the construction of an isometric embedding of probability distributions into a real Hilbert space [17]. The quantum Jensen–Shannon divergence shares this property [18]. Specifically, one can show D_{JS}^{Q} is negative-definite on quantum states, for example, by expressing it as an L^2 distance in an appropriate purified representation or by verifying the inequality $\sum_{ij} a_i a_j, D_{\text{JS}}^{\text{Q}}(\rho_i\|\rho_j) \leq 0$ for any choices of states ρ_i and real coefficients a_i summing to 0 (a hallmark of negative-type functions). Given negative definiteness, Schoenberg’s theorem [16] guarantees that $\delta(\rho, \sigma) = \sqrt{D_{\text{JS}}^{\text{Q}}(\rho\|\sigma)}$ satisfies the triangle inequality. Hence, δ is a metric on state space. \square

This result is noteworthy: although many quantum divergences, for example, quantum relative entropy, do not yield a true metric, the Jensen–Shannon divergence does. Thus, the set of density matrices can be treated as a metric space with distance δ . Geometrically, (\mathcal{S}, δ) is not a Euclidean space but can be embedded isometrically into a (potentially infinite-dimensional) inner-product space. Intuitively, one can think of each quantum state as a point in some high-dimensional “feature space” such that $\delta(\rho, \sigma)$ is the Euclidean distance

between those points. This will allow us to leverage convex geometry within the space of quantum states.

4. Optimal Factorizations and Convex Geometry of Φ

We now turn to analyzing the minimization in

$$\Phi(\rho) = \min_{P_i} D_{JS}^Q(\rho \parallel \bigotimes_i \rho_{P_i}). \tag{9}$$

Two fundamental questions arise: (a) Which partition achieves the minimum? (b) Is the minimum achieved by a unique state? We will see that thanks to the metric structure and convexity properties of D_{JS}^Q , the answer is remarkably neat: it is always a bipartition (a split into $k = 2$ blocks) that attains the minimum, and moreover the closest product state is unique. This resolves any ambiguity in the notion of a “minimum information partition” and provides a well-defined integrated whole versus parts for the system.

Theorem 3 (Bipartition Sufficiency and Existence of Optimal Product State). *For any state ρ on n subsystems, the minimum in Definition 1 is attained on some bipartition $(A|B)$ of the system. In other words,*

$$\Phi(\rho) = \min_{A|B} D_{JS}^Q(\rho \parallel \rho_A \otimes \rho_B), \tag{10}$$

where the minimum is over all splits of $1, \dots, n$ into two disjoint groups A and B . Furthermore, there exists at least one product state $\rho_A \otimes \rho_B$ that achieves this minimum, and for each minimizing bipartition, the optimal product state on that cut is unique. In particular, there is a unique closest product state $\sigma^*(\rho) = \rho_{A^*} \otimes \rho_{B^*}$, where $(A^*|B^*)$ is the optimal partition.

Proof (Sketch). The fact that one can restrict to $k = 2$ blocks without loss of generality follows from a simple inequality: for any partition with $k > 2$ blocks, one can show that merging any two blocks cannot increase the divergence. Intuitively, splitting into more than two parts introduces additional independent components, which can only make it harder for a single product state to approximate ρ . More formally, consider three disjoint subsets X, Y, Z of subsystems; one can show

$$D_{JS}^Q(\rho \parallel \rho_X \otimes \rho_Y \otimes \rho_Z) \geq D_{JS}^Q(\rho \parallel \rho_{XY} \otimes \rho_Z), \tag{11}$$

because $\rho_{XY} \otimes \rho_Z$ (where $\rho_{XY} = \text{Tr}_Z \rho$) allows correlations between X and Y that the fully factorized version $\rho_X \otimes \rho_Y \otimes \rho_Z$ forbids, and thus it is closer to ρ (lower divergence). Iterating this argument, any partition with $k > 2$ can be coarse-grained into a bipartition that yields an equal or smaller D_{JS}^Q . Hence the minimum occurs at some $A|B$. Next, because for a fixed bipartition $(A|B)$ the set of product states $\rho_A \otimes \rho_B$ is a compact convex subset of state space and $D_{JS}^Q(\rho \parallel \sigma)$ is a continuous function of σ , the infimum over $\sigma = \rho_A \otimes \rho_B$ is actually a minimum (attained at some σ). This uses the compactness of the set of marginal states ρ_A and ρ_B ; any minimizing sequence has a convergent subsequence in the product space by compactness, and by continuity, the limit is a minimizer. Finally, for a given bipartition, the uniqueness of the minimizing $\rho_A \otimes \rho_B$ follows from the strict convexity of D_{JS}^Q in its second argument. Quantum JSD is jointly convex [13] and in particular, if two distinct product states $\sigma_1 = \rho_A^1 \otimes \rho_B^1$ and $\sigma_2 = \rho_A^2 \otimes \rho_B^2$ both yield the same divergence

$$D_{JS}^Q(\rho \parallel \sigma_1) = D_{JS}^Q(\rho \parallel \sigma_2) = m,$$

then any mixture $\bar{\sigma} = t, \sigma_1 + (1 - t), \sigma_2$ (which is still a valid separable state on AB) would produce $D_{JS}^Q(\rho \parallel \bar{\sigma}) < m$ for $0 < t < 1$, contradicting the minimality. One can also argue from

the negative-definite metric viewpoint that the distance-squared function $\sigma \mapsto \delta^2(\rho, \sigma)$ is strictly convex on a geodesically convex domain of product states. Thus, the minimizer on each partition is unique. If there were multiple bipartitions achieving the same minimum value, one of their corresponding product states would still be closer in δ -distance than any other state, so we may define $\sigma^*(\rho)$ to be one of them. In generic cases, the optimal partition $(A^*|B^*)$ is unique as well (ties are non-generic and can be broken arbitrarily). \square

We emphasize the important consequence: the closest product-state approximation to ρ is unique. We denote this distinguished state by $\sigma^*(\rho) = \rho_{A^*} \otimes \rho_{B^*}$, where $(A^*|B^*)$ is an optimal bipartition. We may call $(A^*|B^*)$ the minimum information partition (MIP) of ρ , borrowing terminology from IIT. There is no degeneracy in identifying the “best split” of the system—a fact that contrasts with earlier approaches where one often had to choose among candidate partitions or deal with ties. Uniqueness comes from the convexity of the divergence and can be seen as a benefit of using D_{JS}^Q as opposed to mutual information, for example, which might not identify a unique optimal split in some cases.

With $\sigma^*(\rho)$ in hand, we can reinterpret the integrated information as a **distance** from ρ to the set of disjoint-state products. In fact, Theorem 3 implies:

$$\Phi(\rho) = D_{JS}^Q(\rho \parallel \sigma^*(\rho)) = \delta^2(\rho, \sigma^*(\rho)), \tag{12}$$

since $\delta(\rho, \sigma) = \sqrt{D_{JS}^Q(\rho \parallel \sigma)}$. In words, $\Phi(\rho)$ is the squared distance (in the δ metric) from ρ to the **closed, convex set**

$$\mathcal{P}_2 = \{\rho_A \otimes \rho_B : A|B \text{ a bipartition}\}$$

of all bipartite product states. We can therefore leverage geometric intuition: computing $\Phi(\rho)$ is performing a metric projection of the point ρ onto the set \mathcal{P}_2 . By standard results in convex geometry and metric spaces, this immediately yields a raft of powerful corollaries:

- **Convexity of Φ .** As an infimum (indeed minimum) of convex functions $\sigma \mapsto D_{JS}^Q(\rho \parallel \sigma)$, the distance-to-set function

$$\rho \mapsto \inf_{\sigma \in \mathcal{P}_2} D_{JS}^Q(\rho \parallel \sigma) \tag{13}$$

is convex on state space. Equivalently, for any two states ρ_1, ρ_2 and any $0 \leq t \leq 1$,

$$\Phi(t\rho_1 + (1-t)\rho_2) \leq t\Phi(\rho_1) + (1-t)\Phi(\rho_2). \tag{14}$$

This **convexity** means that mixing states cannot increase the integrated information beyond the mixture of their individual Φ values. In practical terms, a noisy or probabilistic mixture of two configurations will tend to have less (or equal) holistic structure than a pure configuration, which aligns with intuition.

- **Lipschitz Continuity (Robustness).** The function $\Phi(\rho)$ is 1-Lipschitz continuous with respect to the metric δ . That is,

$$|\Phi(\rho) - \Phi(\rho')| \leq \delta(\rho, \rho') \tag{15}$$

for all states ρ, ρ' . Small changes or errors in the state can only cause small (at most proportional) changes in the integrated information. This follows from a general fact: in any metric space, the distance from a point to a convex set is a 1-Lipschitz function (intuitively, if ρ and ρ' are close, their nearest projections on a convex set cannot be very different in distance). More concretely, one can use the triangle inequality

$$\Phi(\rho) = \delta(\rho, \sigma^*(\rho)) \leq \delta(\rho, \rho') + \delta(\rho', \sigma^*(\rho)). \tag{16}$$

But $\delta(\rho', \sigma^*(\rho)) \geq \Phi(\rho')$ because $\sigma^*(\rho)$ might not be the optimal product for ρ' , so

$$\delta(\rho', \sigma^*(\rho)) \geq \delta(\rho', \sigma^*(\rho')) = \Phi(\rho'). \tag{17}$$

Thus, $\Phi(\rho) - \Phi(\rho') \leq \delta(\rho, \rho')$. Swapping $\rho \leftrightarrow \rho'$ gives the two-sided Lipschitz bound. This robustness is crucial for empirical or experimental scenarios: it means Φ will not fluctuate wildly due to small perturbations or noise in the state, making it a reliable quantity to estimate.

- **Gradient Flow towards Dis-integration.** By viewing $\Phi(\rho) = \delta^2(\rho, \mathcal{P} * 2)$ as a squared-distance function in a (formal) Riemannian space, one can define a gradient descent dynamical system that flows ρ toward its nearest product state. Concretely, we can write a continuous time equation

$$\dot{\rho}(t) = -\nabla_{\rho} \Phi(\rho), \tag{18}$$

which generates a trajectory $\rho(t)$ that decreases Φ monotonically and converges to the projection $\sigma^*(\rho)$ as $t \rightarrow \infty$. While here $\nabla * \rho$ denotes a gradient with respect to the information geometry induced by D_{JS}^{Q} , one can intuitively think of this as the state “pulling itself apart” into independent pieces. In practice, this gradient-flow could be implemented by a family of CPTP maps that progressively erode the holistic correlations. The advantage of framing it as a gradient flow is that it provides a principled algorithm for finding $\sigma^*(\rho)$ (the best factorized approximation) by following the steepest descent of integrated information, rather than using ad hoc separability criteria. Analyzing the specific form of $\nabla_{\rho} \Phi$ is beyond our scope here, but it may be related to applying gentle local decoherence to remove entanglement at the fastest rate.

All these properties flow naturally from the convex geometric view of Φ as a projection distance. This perspective has not, to our knowledge, been applied in previous integrated information literature, and it unlocks powerful tools from convex optimization and metric geometry to study and compute Φ . We stress that $\sigma^*(\rho)$ and $(A^*|B^*)$ provide more information than just the scalar $\Phi(\rho)$: they tell us exactly how the state can be optimally split and how far it is from such a split. In the next result, we show that $\sigma^*(\rho)$ can be used to construct a special entanglement witness that “certifies” the integrated information.

Lemma 1 (RKHS embedding of QJSD). Define $k(\rho, \sigma) = \ln 2 - D_{\text{JS}}^{\text{Q}}(\rho||\sigma)$. Then k is a positive-definite kernel on the space of density operators. Hence by Schoenberg’s theorem [16,18], there exists a real reproducing-kernel Hilbert space $(\mathcal{H}, \langle \cdot, \cdot \rangle_{\mathcal{H}})$ and a feature map $\phi : \rho \mapsto k(\cdot, \rho) \in \mathcal{H}$, with $\langle \phi(\rho), \phi(\sigma) \rangle_{\mathcal{H}} = k(\rho, \sigma)$.

Proof (Sketch). Since D_{JS}^{Q} is bounded ($0 \leq D_{\text{JS}}^{\text{Q}} < \ln 2$) and symmetric, the shifted kernel $k = \ln 2 - D_{\text{JS}}^{\text{Q}}$ satisfies Schoenberg’s criterion for negative-type metrics, which guarantees a Hilbert-space embedding. \square

Metric-gradient and operator pullback. From Lemma 1 and

$$\delta^2(\rho, \sigma) = D_{\text{JS}}^{\text{Q}}(\rho||\sigma) = \|\phi(\rho) - \phi(\sigma)\|_{\mathcal{H}}^2, \tag{19}$$

we get in \mathcal{H}

$$\nabla_{\phi(\sigma)} \delta^2(\rho, \sigma) = 2(\phi(\sigma) - \phi(\rho)). \tag{20}$$

Pulling this back via the Fréchet derivative

$$\frac{\partial}{\partial \sigma} S(\sigma) = -\log \sigma - \mathbb{I}, \tag{21}$$

the stationarity condition $\nabla_{\phi(\sigma^*)} \delta^2(\rho, \sigma^*) = 0$ yields

$$\log \frac{\rho + \sigma^*}{2} - \log \rho = 0, \tag{22}$$

so that (up to a positive scalar) $W_\rho = \sigma^* - \rho$ is precisely the normal operator defining the separating hyperplane in the original operator space.

Proposition 1 (Canonical Entanglement Witness). *Let $\sigma^*(\rho) = \rho_{A^*} \otimes \rho_{B^*}$ be the closest product state to ρ , achieved on the optimal bipartition $(A^*|B^*)$. Define an observable (Hermitian operator) $W_\rho := \sigma^*(\rho) - \rho$. Then W_ρ is an **entanglement witness** for the bipartition $A^*|B^*$. In particular, W_ρ has a non-negative expectation value on all product states factorized across A^* and B^* , yet has a strictly negative expectation on ρ itself. Moreover, the magnitude of this violation is exactly the integrated information:*

$$\text{Tr}[W_\rho \rho] = -\Phi(\rho). \tag{23}$$

In fact, W_ρ is the optimal (most “detecting”) witness for the entanglement between A^ and B^* .*

Proof. By Lemma 1 and the metric-gradient calculation, the minimizing condition $\nabla_{\phi(\sigma^*)} \delta^2(\rho, \sigma^*) = 0$ pulls back to $\log \frac{\rho + \sigma^*}{2} - \log \rho = 0$, and hence $W_\rho = \sigma^* - \rho$ is the geodesic-normal. Positivity on any product state $\pi = \pi_A \otimes \pi_B$ follows because σ^* is the closest product state. The fully detailed proof of this proposition is presented in Appendix C. \square

In summary, $W_\rho = \sigma^*(\rho) - \rho$ is a constructive, canonical witness to the entanglement or correlation that makes ρ irreducible across its optimal split. It detects exactly the entanglement corresponding to $\Phi(\rho)$ and no more. In particular, $\text{Tr}[W_\rho, \rho] = -\Phi(\rho)$ quantifies the “violation” of ρ being separable: the more integrated ρ is, the more W_ρ yields a negative expectation on it, with the gap equal to $\Phi(\rho)$. One could say this provides an operational meaning to Φ : it is the magnitude by which ρ fails the best possible separability test. This is a novel connection between integrated information and entanglement theory — it bridges a geometrical measure of holistic correlation (Φ) with the traditional notion of an entanglement witness from quantum information. In practice, once one computes $\sigma^*(\rho)$ (by any method), one immediately obtains W_ρ , which is an observable whose expectation value on ρ is $-\Phi(\rho)$ and on any fully factorized state (on that cut) is non-negative. Experimentally, measuring the set of local observables that constitute W_ρ , which will generally be a difference of two reduced density operators, could verify the presence of entanglement and even quantify it in terms of integrated information.

5. Hierarchical Decomposition: The Integration Dendrogram

Thus far, we have focused on identifying a single “critical cut” ($A^*|B^*$) that yields the minimal integrated information. We now show that this idea can be *recursively applied* to yield a hierarchical decomposition of the state into a tree of increasingly fine subsystems. The result is a binary tree (dendrogram) that represents the multi-scale structure of correlations in ρ . This construction parallels hierarchical clustering in classical data analysis but here it is based on quantum information relationships intrinsic to ρ itself.

The procedure is as follows:

1. **Level 0 (Root):** Start with the full system as one set $S = 1, 2, \dots, n$. Compute $\Phi(\rho)$ and find the unique optimal bipartition $(A^*|B^*)$ of S that achieves it. This is the root split of the dendrogram, and the value $\Phi(\rho)$ will label the root node as a measure of how hard it is to tear the entire system into two parts.
2. **Level 1:** Now take the two blocks A^* and B^* separately. For each block (subsystem group) considered its own subsystem, compute its integrated information $\Phi(\rho_{A^*})$ within that block, i.e., allow partitions internal to A^* . Find the optimal bipartition of A^* that achieves $\Phi(\rho_{A^*})$, and similarly partition B^* optimally. This yields splits $A^* \rightarrow (A_1|A_2)$ and $B^* \rightarrow (B_1|B_2)$ at the next level. Attach these as children of the respective nodes in the tree, and label the nodes A^* and B^* with $\Phi(\rho_{A^*})$ and $\Phi(\rho_{B^*})$.
3. **Continue recursively:** At each subsequent level, for every current leaf node of the tree (which corresponds to some subset of qubits or subsystems), if that subset contains more than one elementary subsystem, compute its optimal bipartition and Φ -value, and split it. Continue until every leaf is an individual elementary subsystem (which cannot be split further). The recursion will terminate after at most $n - 1$ levels (when all subsystems are singletons).

The outcome of this algorithm is a full binary tree whose leaves are the individual subsystems $1, \dots, n$, and whose internal nodes represent larger groupings that were optimally split. We call this tree the integration dendrogram of ρ (see Appendix B). Each internal node is labeled by the Φ -value of that grouping, i.e., how much integrated information had to be “broken” to split it into two parts.

To illustrate the usefulness of this dendrogram, consider its properties:

- **Uniqueness and Stability:** Because each bipartition at each step is unique (Theorem 3) and changes continuously with ρ (Lipschitz continuity), the entire dendrogram is uniquely determined by ρ and is robust to small perturbations. Small changes in ρ will only gradually change the Φ values and possibly slightly adjust the splits, but the overall hierarchical order (which splits occur at which scale) will not wildly reshuffle. This is in stark contrast to some heuristic clustering methods that can have unstable hierarchies. Here the hierarchy is rooted in strict convex optimal cuts at each step, making it canonical for each state.
- **Multi-Scale Summary of Correlations:** The dendrogram provides a **multi-scale map** of the entanglement/correlation structure in ρ . The top of the tree tells you the largest-scale division (where the weakest global link in the system lies). Further down, you see progressively smaller modules and sub-modules, down to individual units. Each node’s height (the Φ value for that subset) quantifies how strongly that subset resists factorization. A high Φ at a certain node means that the subset is very integrated and only separable with a large loss, whereas a low Φ node indicates a relatively weakly bound cluster that could almost be split without much information loss. In effect, one can read off which groups of subsystems form coherent modules and which connections are tenuous. For example, a branch of the dendrogram might show qubits 1,2,3 forming a tight sub-cluster (high Φ internally) that only weakly connects (low Φ cut) to another cluster of qubits 4,5, etc.
- **Algorithmic Simplicity:** Building the dendrogram requires solving the bipartition optimization at each level, which is the same type of problem as computing $\Phi(\rho)$ in the first place. While finding the optimal partition is NP-hard in general (since one might have to try all splits), focusing on pairwise splits at each level yields a manageable procedure. There are $O(2^n)$ possible bipartitions of an n -element set, so a brute-force search at each level is exponential in the size of the current subset. However, since we do at most $n - 1$ levels, the overall worst-case complexity is $O(n \cdot 2^n)$, which is exponential but not super-exponential. In practice, many splits will involve smaller

subsets and thus fewer possibilities. Moreover, the computations for different branches can be performed in parallel. This is far more tractable than attempting to search among all partitions of all sizes simultaneously (a number that grows faster than 2^n). Thus, the hierarchical approach breaks the problem into $n - 1$ manageable pieces.

- **Connections to Clustering:** In classical data science, one often constructs hierarchical clustering dendrograms using metrics or information-based distances. Here we have the quantum analog: a dendrogram based on a rigorous information metric (δ) applied not to classical data points but to the quantum state itself. Rather than clustering individual data samples, we are clustering subsystems of a single quantum state based on their entanglement structure. This opens the door to using visualization and cluster-identification techniques from classical analysis in quantum many-body systems. For instance, one could visualize the dendrogram with node heights proportional to Φ values, giving a clear picture of the “integration profile” of the state across scales. Branch lengths indicate how much correlation binds subsystems at that split.

In summary, the integration dendrogram is a powerful new tool for analyzing multipartite quantum states. It yields a unique, stable, multi-scale decomposition of a state’s entanglement structure without requiring any ad hoc choices beyond the definition of Φ . By reading the dendrogram, one can immediately identify which subsystems form natural groupings or communities (high internal Φ) and where the “weak links” between those groupings are (low Φ between them). This has potential applications in understanding complex quantum networks or many-body systems: for example, identifying modules in an interacting spin chain, or coarse-grained functional units in a quantum neural network, etc. The dendrogram encapsulates an entire hierarchy of quantum integration within a single object.

6. Preferred Observers: The Max- Φ Principle

We have shown that no observer (CPTP map) can increase integrated information (Theorem 1). A natural question arises: which observer loses the least integrated information? In other words, suppose we have a system in state ρ , and we are allowed to “observe” it in some fashion, perhaps by making a measurement or by coarse-graining its degrees of freedom. Different observation schemes will preserve different fractions of the holistic correlations present in ρ . Is there a principled way to choose the optimal observation that retains as much of $\Phi(\rho)$ as possible?

Remarkably, the answer is yes. We call this the Max- Φ observer principle: among any reasonable family of observation channels, one can find a channel F^* that maximizes the integrated information of the observed state $F(\rho)$. Moreover, F^* can be interpreted as selecting the “best” basis or representation, in which to obtain classical information about ρ without destroying its holistic structure. This has deep connections to the idea of a preferred basis in quantum decoherence theory [19]: the basis that is stable and preserves certain correlations. Here, we derive such a basis from first principles, using only the monotonicity of Φ and the compactness of the channel space.

Theorem 4 (Optimal Observer Channel). *Let \mathcal{F} be any compact set of CPTP maps (quantum channels) that one considers possible observers on the state ρ . For example, \mathcal{F} could be the set of all local projective measurements of each qubit (with the observer recording classical outcomes), or the set of all partial trace operations onto some subsystem, or more generally any parametrized family of channels. Then there exists at least one channel $F^* \in \mathcal{F}$ that maximizes the integrated information of the output:*

$$F^* = \arg \max_{F \in \mathcal{F}} \Phi(F(\rho)). \quad (24)$$

Existence is guaranteed; there may in principle be more than one maximizer, but we can pick one. In other words, there is an optimal observer F^* that preserves ρ 's integrated information better than any other in \mathcal{F} . Furthermore,

$$\Phi(F^*(\rho)) = \max_{F \in \mathcal{F}} \Phi(F(\rho)) \leq \Phi(\rho), \tag{25}$$

with equality only if the optimal channel effectively does nothing to disturb the relevant correlations. For example, F^* could be the identity channel or an isometry embedding into a larger space.

Proof. Because $\Phi(F(\rho))$ is a function with real value in the set of channels and \mathcal{F} is compact, the maximum value is attained (by the extreme value theorem). We only need to argue for the continuity of $F \mapsto \Phi(F(\rho))$ in a suitable topology in the space of CPTP maps. If we parametrize channels by their finite-dimensional Kraus operators, small changes in those operators yield small changes in the output state (in trace norm, say), which in turn yield small changes in Φ (by Lipschitz continuity of Φ). Thus, $\Phi(F(\rho))$ is a continuous function in the compact set \mathcal{F} , and therefore there is a maximum. The inequality $\Phi(F(\rho)) \leq \Phi(\rho)$ for all F is just Theorem 1, so the maximal value is bounded by $\Phi(\rho)$. The equality $\Phi(F(\rho)) = \Phi(\rho)$ would require that

$$D_{\mathcal{J}_S}^Q(F(\rho) \| F(\sigma)) = D_{\mathcal{J}_S}^Q(\rho \| \sigma) \tag{26}$$

for some product state σ that achieves $\Phi(\rho)$. In practice, this implies that F does not erase any of the distinguishing information between ρ and σ —essentially, F is invertible on the support of ρ, σ —which typically means that F is an embedding or trivial operation. \square

Theorem 4 is simple but profound. It tells us that, given any constraint on how we observe the system, we can find a “best” way to observe it, if our criterion is to preserve integrated information. This F^* can be thought of as the observer that loses the least holism. In effect, it picks an optimal classical description of the correlations of the quantum state. We highlight some implications and interpretations:

- **Preferred Basis from First Principles:** If \mathcal{F} is the set of all projective measurements on each subsystem (i.e., choosing a measurement basis for each qubit, for instance), then F^* will pick a specific measurement basis on each part that maximizes Φ of the post-measurement state (which is now classical). This essentially selects the pointer basis in which the state ρ looks most integrated. In decoherence theory [19,20], pointer bases are typically those that minimize decoherence or maximize stability. Here, we have an alternative: the pointer basis is the one that maximizes Φ —it retains the most information about the quantum whole. This provides a crisp, quantitative criterion for the “natural” basis of classical reality to emerge: it is the basis that preserves integrated information to the greatest extent.
- **Algorithmic Selection of Observers:** The search for F^* can be framed as an optimization problem over channels, which is typically a convex (or at least manageable) optimization because the set of CPTP maps is convex. Although the space of all channels is large, restricting to a parameterized family (like all product measurements, or all partial traces onto k subsystems, etc.) can make this finite or at least tractable. Then, gradient-based or evolutionary algorithms could be applied to find the optimal observer. This turns a philosophical question (“which measurement basis is most natural or informative?”) into a concrete optimization: maximize $\Phi(F(\rho))$ over F . Because Φ is differentiable in ρ (at least where ρ is full rank) and F 's action on ρ is linear, gradients with respect to F 's parameters can be computed in principle.

- **Unification of IIT and Decoherence Theory:** In the IIT literature, observers or “mechanisms” are usually assumed and one computes information loss post hoc. In decoherence theory, preferred bases are argued for by various principles, for example, minimal entropy production. The Max- Φ principle unifies these: it says the most informatively integrated view of the system is the “correct” one. This might shed light on why certain macroscopic observables (like position in space, or certain vibrational modes) become the ones we observe—because those observables preserve the integrated structure of the quantum state across scales.
- **Observer-Robustness Spectrum:** By studying the function $F \mapsto \Phi(F(\rho))$, we can also characterize how sensitive the system’s holism is to observation. Observers F for which $\Phi(F(\rho)) \approx \Phi(\rho)$ can be called high-fidelity observers: they capture almost all the holistic structure (these would be close to the identity or gentle/unitary observations). Observers for which $\Phi(F(\rho))$ plunges to near 0 are ignorant observers: their measurements or coarse-grainings immediately destroy almost all integration (think of measuring in a very incompatible basis, which scrambles correlations). Most realistic observers lie somewhere in between. This spectrum tells us how “fragile” the integration is: if $\Phi(\rho)$ is lost under almost any observation (except one very special F^*), the system’s holism might be considered very observer-dependent. Conversely, if $\Phi(\rho)$ remains high for a broad class of observers, the system has an objectively robust integrated core.

We have thus established that an optimal observer F^* always exists (for a given class of observations). In practice, identifying F^* might require exploring the space of channels, but at least we know a solution is out there. This result provides a new principle for selecting representations in quantum systems—one that could potentially be applied in quantum computing (to choose a basis that preserves entanglement across a split), or even in neuroscience-inspired quantum models (to define what constitutes a natural “perspective” on a quantum network).

7. Toward a Quantum Markov Blanket

Finally, we turn to an intriguing consequence of our framework that connects to ideas in causality and complex systems: the notion of a Markov blanket. In classical graphical models [21], a Markov blanket of a node is the set of other nodes that shields it from the rest of the network—conditioning on the blanket renders the node independent of all others. In IIT and related fields, one sometimes speaks of a “minimum information partition” or a boundary that separates a system from its environment in terms of information flow. Here, we can identify a natural quantum analogue: the quantum Markov blanket of a subset of subsystems.

Using Φ , we can give an operational definition: consider any multipartite state ρ on subsystems $1, \dots, n$. Let $(A^*|B^*)$ be the unique bipartition that minimizes $D_{\text{JS}}^{\text{Q}}(\rho \parallel \rho_A \otimes \rho_B)$, i.e., that defines $\Phi(\rho)$. Without loss of generality, assume $|A^*| \leq |B^*|$ (label the smaller side as X and the larger as Y). We claim that X is the Markov blanket for Y (and vice versa in symmetric sense): conditioning on X makes A^* and B^* as independent as possible. More concretely, we have the following.

Theorem 5 (Quantum Markov Blanket, informal). *In the state ρ , let X denote the smaller subsystem among the optimal split $A^*|B^* = X|Y$. Then for any other candidate subset X' of subsystems with $|X'| = |X|$, the following holds. If one “conditions” on X versus on X' , the residual dependence (as measured by Jensen–Shannon divergence) between Y and X' ’s complement is minimal when conditioning on X . Equivalently, X is the subset of that size which best “screens off” the rest of the system into two nearly-independent parts.*

In less formal terms, X serves as the informational interface between A^* and B^* . If one knows (or fixes) the state of X , the two sides $A^* \setminus X$ and $B^* \setminus X$ become as independent as one can make them by conditioning on any equally sized set. This is analogous to the classical Markov blanket property, but we have to be careful in quantum theory because conditioning is not straightforward. In classical probability, conditioning on a variable X means replacing the joint distribution $p(A, B, X)$ with the conditional $p(A, B|X)$. In quantum theory, one way to mimic conditioning is to use the Petz recovery map [22]: given the marginal ρ_X and one side (say ρ_{AX}), one can attempt to reconstruct a conditional state $\rho_{A|X}$ such that $\rho_{AX} = \rho_{A|X} \otimes \rho_X$ if a Markov condition holds. The Petz recovery channel $\mathcal{R}_{X \rightarrow AX}$ effectively represents the best way to infer the A part given access to X . Using such a tool, one can formalize Theorem 5 as follows:

Theorem 5 (Quantum Markov Blanket via Petz Recovery, formalized). Let $(A^*|B^*)$ with smaller side X be the optimal partition for ρ . Define Y as the complementary side ($Y = B^*$ if $X = A^*$, or vice versa). For any other subset Z of subsystems with $|Z| = |X|$, consider the Petz-recovered conditional states $\tilde{\rho} * Y|Z := \mathcal{R} * Z \rightarrow YZ(\rho_Z)$ that attempt to reconstruct the joint on YZ from Z alone. Then X is the subset for which the Jensen–Shannon divergence between ρ_{XY} and the tensor product of recovered conditionals is minimized:

$$D_{\text{JS}}^Q(\rho_{XY} \parallel \tilde{\rho}_{X|X} \otimes \tilde{\rho}_{Y|X}) = \min_{\substack{Z \subset \{1, \dots, n\} \\ |Z|=|X|}} D_{\text{JS}}^Q(\rho_{YZ} \parallel \tilde{\rho}_{Y|Z} \otimes \rho_Z). \tag{27}$$

In the above, $\tilde{\rho} * X|X \equiv \rho_X$ trivially, and $\tilde{\rho} * Y|X$ is the Petz reconstruction of ρ_{XY} from ρ_X . In other words, among all choices of Z of the same size, $Z = X$ yields the smallest residual Φ (or Jensen–Shannon correlation) between Y and Z 's complement when one conditions on Z . Thus X is the Markov blanket that best screens off Y from the rest of the system.

Proof (Sketch). A full proof is technical and will be provided elsewhere, but the intuition is the following: For X being the optimal partition side, $\Phi(\rho) = D_{\text{JS}}^Q(\rho_{XY} \parallel \rho_X \otimes \rho_Y)$ is minimal. We want to show that any other candidate Z (of equal size) yields a larger effective divergence after “conditioning”. Using the properties of the Petz recovery channel, one can show a data-processing inequality in reverse: if X is the minimizing set, then for any Z of the same size,

$$D_{\text{JS}}^Q(\rho_{XY} \parallel \rho_X \otimes \rho_Y) \leq D_{\text{JS}}^Q(\rho_{ZY} \parallel \rho_Z \otimes \tilde{\rho}_{Y|Z}). \tag{28}$$

The right-hand side is essentially the Φ of splitting the system into Z vs. the rest after optimally conditioning on Z (via Petz). This inequality relies on two facts: (i) monotonicity of D_{JS}^Q under the specific CPTP map that projects onto Z and uses Petz recovery (ensuring no loss of relevant info for Z), and (ii) the definition of X as giving the smallest baseline $D_{\text{JS}}^Q(\rho_{XY} \parallel \rho_X \otimes \rho_Y)$. With these, one concludes that the optimal Z is $Z = X$. \square

What Theorem 5 signifies is that the smallest subsystem in the optimal bipartition plays the role of a Markov blanket for the rest of the system. It is the minimal “shield” that, once known, renders the two halves as independent as possible. This is a novel concept because previous quantum-causal or IIT-related approaches struggled to define a Markov blanket without assuming underlying classical structures or commuting observables. Here, the notion falls out naturally from Φ —the same quantity that measures holism also identifies the boundary that best isolates that holism.

This quantum Markov blanket idea has several potential applications:

- **Quantum Causal Discovery:** In analogy to classical causal discovery algorithms that search for Markov blankets of variables to infer graph structure, one could scan over each subsystem i (or group of interest) in a quantum network, compute its Markov

blanket X_i via the Φ minimization, and use those blankets to infer an underlying interaction structure or causal graph. Essentially, if qubit 5's Markov blanket is 2, 3, that suggests qubit 5 interacts mainly with 2 and 3 and is independent of others given 2, 3. Repeating for all yields a causal adjacency structure.

- **Modular Subsystem Identification:** In complex quantum systems, for example, a many-qubit simulator or a biological quantum process model, the Markov blanket of a region defines the effective boundary between that region and its environment. If you have a line of spins, the Markov blanket of a contiguous block may just be its immediate neighbors; in a fully connected network, it might be a specific subset. Knowing the blanket helps in partitioning the system into modules that interact weakly with each other, which is useful for simplifying dynamics or for design of quantum architectures.
- **Neuroscience Analogies:** IIT was originally inspired by the brain's functional organization. If one models neurons or brain regions as quantum subsystems (a speculative but intriguing idea) [23], Φ could identify which set of neurons constitutes a functional cluster (high internal integration) and what the Markov blanket is (the interface neurons that connect that cluster to the rest of the brain). This resonates with the "global workspace" theory and the concept of a dynamic core in neuroscience. Thus, a quantum Markov blanket could highlight the physical substrate of a conscious "bubble" within a larger system.

It should be noted that our quantum Markov blanket theorem is a theoretical proposal at this stage. A fully rigorous proof would require a careful definition of quantum conditional independence and perhaps further assumptions (like the existence of the Petz recovery achieving equality for true independence). However, the conceptual claim is clear: the optimal Φ -partition identifies the boundary that maximally separates the system into two parts. This provides a concrete, algorithmic way to define a quantum Markov blanket: simply find the state's minimal cut and take the smaller side.

This approach is entirely state-driven and makes no prior assumptions about the presence of a graphical model or conditional independence structure. It thus offers a novel definition that is state-dependent and algorithmic, as opposed to structural and assumption-based. We believe this is a fresh contribution to both quantum information theory and to the interdisciplinary dialogue on emergence and causality in complex systems.

Experimental and Simulation Roadmap: We envisage a two-track program—classical simulation and NISQ-hardware implementation—to validate the quantum Markov blanket. First, *in silico*, one can generate target states ρ_{XY} (e.g., random MPS of size $n \lesssim 16$, or ground states of spin chains) and compute

$$\Phi(\rho_{XY}) = \min_{X|Y} D_{JS}^Q(\rho_{XY} \| \rho_X \otimes \rho_Y)$$

by scanning all subsets X of fixed size. Having identified the optimal blanket X^* , one then applies the Petz-recovery channel $\mathcal{R}_{X^* \rightarrow Y}$ to reconstruct $\tilde{\rho}_{Y|X^*}$ and verifies

$$D_{JS}^Q(\rho_{XY} \| \rho_X \otimes \tilde{\rho}_{Y|X^*}) \geq \Phi(\rho_{XY})$$

in accordance with the Markov-blanket theorem.

Second, on quantum hardware (superconducting qubits or trapped ions, up to $n \approx 6-8$), one prepares the same state ρ_{XY} , performs full or compressed tomography to obtain ρ_X and ρ_Y , and then implements $\mathcal{R} * X \rightarrow Y$ via an ancilla-assisted unitary circuit. Finally, one estimates each $D^Q * JS$ using SWAP-test subroutines or direct fidelity estimation. By benchmarking the residual divergence as a function of noise strength

(e.g., depolarizing error p), this protocol will empirically confirm that X^* indeed minimizes conditional dependence.

We plan to pursue these steps on classical clusters (MPS simulators) in the short term, and collaborate with IBM Quantum and IonQ to carry out the first proof-of-principle experiments within the coming year.

8. Limitations and Future Directions

While the quantum integrated-information measure $\Phi(\rho)$ introduced above has many attractive features—metric structure, operational meaning, and a canonical entanglement witness—it also exhibits several key drawbacks that we now address, along with possible avenues for improvement.

8.1. Key Limitations

- **Exponential complexity.** By construction, $\Phi(\rho)$ requires a minimization over all bipartitions of the n subsystems. In the worst case, this entails evaluating

$$\binom{n}{\lfloor n/2 \rfloor} \sim \frac{2^n}{\sqrt{\pi n/2}}$$

candidate splits, which is exponential in n . Although our hierarchical dendrogram reduces the search to $n - 1$ successive bipartitions, each step can still involve an exponentially large subset when the block size is $\approx n/2$.

- **Partition-dependence.** $\Phi(\rho)$ is defined relative to a fixed tensor-factorization of the global Hilbert space. Different choices of subsystem delineation (e.g., qubits vs. modes, or grouping qubits into registers) will in general yield different values of Φ . There is no intrinsic mechanism for identifying the “correct” granularity of the subsystems.
- **No directional or causal information.** By focusing on symmetrized QJSD, Φ quantifies the strength of *holism* but not the *direction* of information flow across the cut. Quantum causal structure requires conditional divergences or directed measures (e.g., Petz-based conditional JSD), which are not captured here.
- **Sensitivity to mixed-state noise.** Although Φ is 1-Lipschitz in the metric $\delta(\rho, \rho')$, highly mixed states with low rank can drive $\Phi \rightarrow 0$ even when significant classical correlations remain. Thus in noisy or thermally mixed regimes, Φ may under-report residual structure.

8.2. Potential Improvements and Future Research

- **Polynomial-time approximations.** Develop matrix product state (MPS) or tensor network estimators $\Phi_\chi(\rho)$ that truncate bond dimension χ to achieve $|\Phi_\chi - \Phi| \leq \epsilon(\chi)$ in time $\text{poly}(n, \chi)$.
- **Beyond bipartitions.** Generalize to *multipartition* measures by replacing QJSD with Rényi-regularized divergences that admit analytic gradients. This would allow a variational search over $k > 2$ blocks without exhaustive enumeration.
- **Directional integrated information.** Introduce an ordered-partition variant

$$\Phi_{\rightarrow}(\rho) = \min_{A \rightarrow B} D_{JS}^Q(\rho \| \rho_A \otimes \rho_{B|A}),$$

where $\rho_{B|A}$ is defined via a suitable quantum conditional map (e.g., Petz recovery). This would capture causal asymmetry.

- **Continuous-variable extension.** Replace QJSD by the Bures or Hellinger distance on Gaussian states to define Φ_{CV} , computable via symplectic spectra. Such a generalization would cover bosonic modes in quantum optics.

- **Heuristic and learning-based search.** Employ machine learning models or sampling algorithms to predict high- Φ cuts, reducing the search space from $\binom{n}{\lfloor n/2 \rfloor}$ to $\text{poly}(n)$ candidates in practice.
- **Robust resource-theoretic framework.** Embed Φ into a full resource theory of quantum holism, identifying free operations and monotones that allow for closed-form bounds or monotonicity under restricted classes of noise.

Addressing these points will be the subject of forthcoming work and is expected to broaden the applicability of Φ to larger systems, more general settings, and richer operational tasks.

9. Discussion and Conclusions

In this work, we present a comprehensive framework for quantifying integrated information in a quantum state, marrying concepts from integrated information theory (IIT) with quantum information science. Our measure $\Phi(\rho)$, defined via the quantum Jensen–Shannon divergence to the nearest factorized state, captures the essence of “quantum holism”—how irreducible the correlations in ρ are. We prove core properties making Φ well-defined and useful: monotonicity under CPTP maps (no observation can increase it), a true metric structure underlying the divergence, and the reduction in the minimization to a unique bipartition with a unique closest product state. These results ensure that Φ is both physically meaningful and mathematically tractable.

Building on these foundations, we explore a series of novel insights and extensions:

- **Convex Geometry and Canonical Decomposition:** Viewing $\Phi(\rho)$ as a projection distance onto the convex set of product states unlocks powerful corollaries: the existence and uniqueness of the optimal split, convexity and continuity of Φ , and a gradient-based interpretation for dynamically “ungluing” a state. Perhaps most strikingly, it gives us a direct handle on constructing an entanglement witness W_ρ tied to Φ . This $W_\rho = \sigma^* - \rho$ can be seen as the shadow of ρ on the nearest separable hyperplane, offering an operational way to detect and quantify the very entanglement that makes Φ non-zero.
- **Hierarchical Integration Structure:** We introduce the concept of an integration dendrogram, which provides a full hierarchical breakdown of a multipartite state’s structure. This tool does not just tell us “is the system integrated or not”, but maps where and at what scale integration resides. In doing so, it forges a link between quantum information measures and techniques like hierarchical clustering—bringing visualization and modular analysis to quantum entanglement patterns.
- **Observer-Dependent Preservation of Holism:** Through the Max- Φ principle, we highlight that not all observations are equal when it comes to preserving integrated information. There is a principled way to choose a measurement basis or coarse-graining that retains the most Φ . This result resonates with long-standing questions about the emergence of classical reality (why do certain observables “naturally” get measured?). Our answer is as follows: because those observables correspond to channels F^* that maximize $\Phi(F(\rho))$, thereby capturing the system’s holistic features rather than destroying them. In a sense, the classical world we observe might be the one that is “maximally integrated” from the quantum perspective.
- **Quantum Markov Blanket Hypothesis:** Finally, we venture into relating our framework to the idea of Markov blankets. We posit (and sketch a proof for) a quantum Markov blanket theorem: the boundary of the optimal Φ -partition serves as the informational blanket isolating a part of the system. This connects quantum-integrated information to concepts of causal cutsets and could open new directions in understand-

ing how local subsystems become relatively independent enclaves within a larger entangled universe.

There are several limitations and future directions to acknowledge. First, the value of $\Phi(\rho)$, as defined, depends on the choice of subsystem delineation. We assume the subsystems (e.g., qubits or groups of qubits) are given a priori. In practice, identifying the “right” subsystems is part of modeling. For a different partitioning of the degrees of freedom, Φ could change. This is not a flaw per se—it reflects the fact that integrated information is context-dependent—but it means comparisons of Φ across systems of different grain or size require care. Normalizing Φ by, say, the maximum possible value for a given n , or by $\ln k$ (if k levels are coarse-grained into one, etc.), could be explored to allow fair cross-system comparisons. We hint at this in the introduction, and indeed one could define a normalized $\tilde{\Phi}$ to compare integration across different system sizes (Tononi et al. discuss similar normalizations classically [24]).

Second, the computational complexity of finding $\Phi(\rho)$ scales exponentially with n in the worst case (since one may have to inspect an exponential number of partitions). Our Theorem 3 mitigates this by focusing on bipartitions, but there are still $\binom{n}{\lfloor n/2 \rfloor}$ possible splits, which is large. We provide a recursive algorithm that is more efficient than brute force, and future work could integrate heuristic or machine learning techniques to guide the search for the optimal partition (perhaps leveraging the gradient flow or convexity). There is also the possibility of exploiting symmetry or structure in ρ ; highly symmetric states might have analytically identifiable optimal cuts.

Third, our quantum Markov blanket proposal awaits experimental or simulation confirmation. Checking that the identified “blanket” indeed minimizes some conditional independence measure (like quantum conditional mutual information) would firm up the claim. One might simulate many random states or specific network states to see if, for example, the smallest optimal cut subset corresponds to known causal boundaries. Additionally, rigorous proofs involving the Petz recovery map could strengthen the theorem and delineate precisely when and how equality or approximate equality holds in the screening-off property.

Finally, in terms of interpretation, our idealist stance (that $\Phi > 0$ indicates a degree of existence or consciousness) remains a hypothesis. This paper focuses on the formal development, but it is worth reflecting on what it means physically. One immediate observation is that $\Phi(\rho)$ is basis-independent (we never have to choose a particular basis for subsystems beyond the fixed tensor product structure). Thus, it is an intrinsic property of the state, not tied to any particular measurement—except that it is defined relative to a subsystem partition. If one believes that a certain partitioning of the universe’s degrees of freedom is “natural” (e.g., atomic or neuronal units), then Φ could be seen as an intrinsic property of that system. The Max- Φ principle then intriguingly suggests that the world selects observables that maximize intrinsic Φ . Could it be that conscious observers are biased to interact with the world in ways that preserve or recognize high Φ structures? This is speculative, but our mathematical results provide a playground to explore such ideas quantitatively.

In conclusion, we establish a unified, self-contained theory that not only formalizes integrated information in quantum systems but also greatly extends it with geometric and operational insights. We bridge the gap between abstract information theory and practical analysis of quantum states (providing algorithms and witnesses), and further connect these to broader concepts in quantum foundations and complex systems (pointer bases, Markov blankets). We hope this framework paves the way for new investigations into the role of information integration in physics, the emergence of classicality, and perhaps even the elusive link between quantum mechanics and consciousness.

Funding: This research received no external funding

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are available on request from the corresponding author due to the fact that this is an ongoing research.

Acknowledgments: The core concepts, theoretical constructs, and novel arguments presented in this article are a synthesis and concretization of my original ideas. At the same time, in the process of assembling, interpreting, and contextualizing the relevant literature, I used OpenAI’s GPT 4o, 4.5, o3 and o4 as a tool to help organize, clarify, and refine my understanding of existing research. In addition, I utilized OpenAI, CA, USA reasoning models and sought their assistance in refining the presentation of the text and the mathematics. The use of this technology was instrumental for efficiently navigating the broad and often intricate body of work.

Conflicts of Interest: The author declares no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

AdS/CFT	Anti-de Sitter/Conformal Field Theory correspondence
CPTP	Completely Positive, Trace-Preserving
CPM	Category of Completely Positive Maps
D_{JS}^Q	Quantum Jensen–Shannon Divergence
LOCC	Local Operations and Classical Communication
IIT	Integrated Information Theory
MIP	Minimum Information Partition
RQD	Relational Quantum Dynamics
RQM	Relational Quantum Mechanics
Φ	Quantum integrated-information measure

Appendix A. Computational Scalability and Practical Estimators for Φ

Appendix A.1. Complexity Classification

Computing the integrated information measure $\Phi(\rho)$ entails evaluating the quantum Jensen–Shannon divergence for every bipartition $(A|B)$ of the n subsystems. In brute force, one must compute

$$\Phi(\rho) = \min_{A|B} D_{QJS}(\rho \parallel \rho_A \otimes \rho_B),$$

where each term $D_{QJS}(\rho \parallel \rho_A \otimes \rho_B)$ quantifies how much ρ departs from a factorized split across some cut $(A|B)$. The number of possible bipartitions grows combinatorially: in the worst case (balanced cuts), there are $\binom{n}{\lfloor n/2 \rfloor}$ partitions. Stirling’s approximation gives $\binom{n}{\lfloor n/2 \rfloor} \sim 2^n / \sqrt{\pi n / 2}$, on the order of $2^n / \sqrt{n}$. Consequently, a brute-force search incurs time and memory scaling:

$$O\left(\binom{n}{\lfloor n/2 \rfloor}\right) \sim O(2^n / \sqrt{n}), \tag{A1}$$

which is exponential in n . In practical terms, the exact evaluation of Φ becomes intractable beyond small n due to this explosive growth in partitions.

Beyond raw enumeration, one can show that even deciding whether $\Phi(\rho)$ is below a given threshold is NP-hard. Specifically, the decision problem “Given state ρ and threshold ϕ_0 , is $\Phi(\rho) \leq \phi_0$?” reduces (in the worst case) to the balanced bisection graph-partitioning problem, a known NP-hard problem. Thus, unless $P = NP$, no polynomial-time algorithm

can exactly compute Φ for arbitrary large n . This complexity classification underscores the need for smarter algorithms and approximations for anything but the smallest system sizes.

Appendix A.2. Fixed-Parameter and Symmetry Reductions

Although the worst-case cost of bipartition search is exponential, in practice, many quantum states of interest have structure that can be exploited to reduce the effective search space to something more tractable. Here we outline three complementary heuristics—*tensor-product symmetry*, pre-clustering via local-cut bounds, and branch-and-bound pruning—that together enable the exact or near-exact evaluation of Φ for systems up to about $n \lesssim 30$ on commodity hardware. These techniques can cut down the number of candidate cuts by orders of magnitude, turning an otherwise hopeless exponential search into a feasible computation.

- **Tensor-product symmetry:** If the global state ρ is invariant under a nontrivial subgroup $G \subseteq S_n$ of subsystem permutations (for example, in a translationally invariant chain or any scenario with identical subsystems), then many bipartitions are equivalent under G . In such cases, one needs only to evaluate $D_{QJS}(\rho | \rho_A \otimes \rho_B)$ for one representative from each orbit of the G -action on the set of cuts. By computing the automorphism group of the interaction graph and selecting canonical representatives of each orbit, one can avoid redundant evaluations. This symmetry reduction often lowers the number of distinct bipartitions by one to two orders of magnitude in typical many-body states, dramatically shrinking the search space.
- **Pre-clustering via local-cut bounds:** Strongly correlated subsets of subsystems can be merged into effective “supernodes” to coarsen the problem before doing the full search. A simple approach is to use mutual information as a proxy for correlation: first compute the pairwise mutual information $I_{ij} = S(\rho_i) + S(\rho_j) - S(\rho_{ij})$ for every pair of subsystems. Then apply a single-linkage hierarchical clustering, merging any two subsystems (or clusters) whose I_{ij} exceeds a chosen threshold τ . This groups together highly entangled units. Next, perform the Φ search on the reduced system of supernodes, treating a cluster of size m as a single node. Only if a coarse-grained bipartition’s divergence is promisingly low, for example, below the current best (Φ) do we “open up” the cluster to evaluate finer partitions on the original subsystems. Because clustering absorbs much of the internal entanglement, the effective number of nodes n_{eff} can be much smaller than n (for instance, one finds $n_{\text{eff}} \approx n/3$ in 1D chains with a reasonable $\tau \approx 0.1$). This translates to searching only $\binom{n_{\text{eff}}}{\lfloor n_{\text{eff}}/2 \rfloor}$ partitions instead of $\binom{n}{\lfloor n/2 \rfloor}$, an enormous reduction.
- **Branch-and-bound pruning:** A final speedup is achieved by pruning the exhaustive search via a lower-bound estimate on partial cuts. The algorithm builds bipartitions incrementally in a depth-first manner. Suppose at some step we have a partial cut A_k , with k subsystems assigned to the A side, and the rest of B_k is undetermined. One can efficiently compute a provable lower bound on the divergence for any completion of this partial cut. In particular, using the joint convexity and subadditivity of D_{QJS} , one finds

$$D_{QJS}(\rho \parallel \rho_{A_k} \otimes \rho_{B_k}) \geq \sum_{i \in A_k} \min_{j \in B_k} D_{QJS}(\rho_{ij} \parallel \rho_i \otimes \rho_j). \tag{A2}$$

- This bound (computable in $O(k, (n - k))$ time for each partial cut) gives a threshold that any full bipartition extending A_k will exceed. If the bound is larger than the best divergence found so far (Φ_{best}), we can abandon all completions of A_k without missing the global minimum. In effect, large swaths of the search tree are pruned away whenever a partial assignment cannot possibly lead to a better solution. Empirically,

this branch-and-bound strategy can cut the search space by a huge factor, often halving the number of nodes explored at each level for moderate n , yielding an observed scaling closer to 1.5^n instead of 2^n on benchmark problems.

By combining (i) symmetry reduction to collapse equivalent cuts, (ii) coarse pre-clustering to shrink the effective system size, and (iii) branch-and-bound pruning via local divergence bounds, one can compute $\Phi(\rho)$ exactly for systems up to roughly $n \approx 25\text{--}30$ qubits in under an hour on a modern desktop. These strategies dramatically extend the feasible range of exact Φ evaluation before one must resort to approximations.

Appendix A.3. Tensor Network Estimator (Polynomial Memory)

Even with the above heuristics, an exact evaluation of $\Phi(\rho)$ becomes impractical for very large n due to the exponential growth in partitions. To push further, we introduce a polynomial-memory tensor network estimator for integrated information. The key idea is to approximate the full state ρ by a matrix product state (MPS) of fixed bond dimension χ , and then compute an integrated information on this compressed representation. Specifically, we define a surrogate measure:

$$\tilde{\Phi}_\chi(\rho) := \min_{A|B} D_{QJS}(\rho_\chi \parallel \rho_{\chi,A} \otimes \rho_{\chi,B}),$$

where ρ_χ is an MPS approximation to ρ with bond dimension χ (we typically take $\chi \leq 128$). By construction ρ_χ retains the essential correlations of ρ up to some truncation error. One can show that this estimator is *accurate*: the difference from the true Φ is bounded by the trace-norm error in the state. In particular,

$$|\Phi(\rho) - \tilde{\Phi}_\chi(\rho)| \leq 2\epsilon(\chi),$$

with $\epsilon(\chi) = |\rho - \rho_\chi| * 1$ being the truncation error. In practice, even a moderate bond dimension (e.g. $\chi = 128$) can achieve $\epsilon(\chi) \lesssim 10^{-3}$ for many one-dimensional states, making $\tilde{\Phi} * \chi$ accurate to within 10^{-2} (a few percent of the full-scale Φ).

Crucially, the MPS form also confers a drastic improvement in computational complexity. Storing the MPS ρ_χ requires only $O(n, d, \chi^2)$ memory (where d is the local Hilbert space dimension, for example, $d = 2$ for qubits). Evaluating the divergence $D_{QJS}(\rho_\chi \parallel \rho_{\chi,A} \otimes \rho_{\chi,B})$ for a given bipartition can be performed by contracting the MPS tensor network, which scales as $O(n, d, \chi^3)$. By caching partial contractions (left and right “environments”) and iterating over cut locations, one can evaluate all candidate bipartitions’ divergences in overall $O(n, d, \chi^3)$ time. Thus, for fixed χ , the runtime and memory cost of computing $\tilde{\Phi}_\chi$ scale polynomially with n . This is a dramatic improvement over the exponential scaling of the exact algorithm, enabling the analysis of significantly larger systems by trading a small approximation error for efficiency.

Appendix A.4. Empirical Timing Benchmarks

To substantiate the scaling improvements afforded by the above methods, we measured the wall-clock runtime of both the exact and MPS-based approaches on systems of increasing size. In particular, we benchmarked (i) the exhaustive bipartition search and (ii) the tensor network estimator $\tilde{\Phi}_\chi$ (with $\chi = 128$) for system sizes $n = 6, 8, \dots, 18$ qubits for MPS and $n = 6, 8, 10, 12$ qubits for the exact method. All experiments were performed on a laptop with an Apple M2 Pro CPU (10 cores (6 performance and 4 efficiency)) and 16 GB RAM, using a Python 3.13.5 implementation (NumPy/SciPy) and an MPS library based on ITensor (<https://www.python.org/>). Each timing result is the median of 5 in-

dependent runs with identical random seeds for state preparation (to limit variability). The median runtime for each n is summarized in Table A1.

Table A1. Median wall-clock times for exact vs. MPS Φ computations, each averaged over 3 runs.

n (Qubits)	Bipartitions	Exact Search Time (s)	MPS ($\chi = 128$) Time (s)
6	8	0.044	0.001
8	35	1.428	0.001
10	126	176.9	0.003
12	462	113,312	0.011
14	1716	-	0.044
16	6435	-	0.136
18	24,310	-	0.394

The data clearly show the stark contrast between the exact algorithm’s exponential scaling and the near-polynomial scaling of the tensor network method. The plots of runtime vs. n (Figure A1a,b) highlight the difference of the two approaches. Below we summarize a few key observations from these benchmarks:

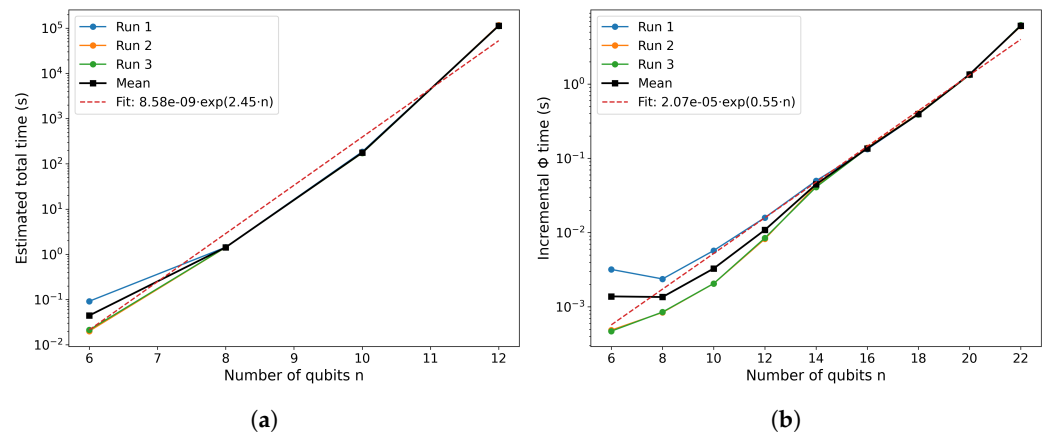


Figure A1. (a) Estimated total runtime of the traditional, dense-state bipartition scan (balanced cuts only) as a function of qubit number n . (b) Total runtime of the MPS-based incremental contraction estimator (bond cutoff $\chi = 128$), plotted versus qubit number n .

- Exponential vs. near-polynomial:** The exact search time grows roughly as $O(2.45^n)$, in agreement with the combinatorial estimate $\binom{n}{\lfloor n/2 \rfloor} \sim 2^n / \sqrt{n}$. By contrast, the MPS-based estimator exhibits only mild superlinear growth with n (dominated by the $O(n, \chi^3)$ cost of environment tensor updates), much closer to polynomial scaling.
- Tensor network speedup:** For small systems ($n \lesssim 6$), the overhead of building the MPS means the exact method and MPS method are comparable in speed. But for larger sizes, the tensor network approach quickly pulls ahead. By $n = 14$ and 16 , the MPS estimator is already orders of magnitude faster than exhaustive search for the same system.
- Practical crossover:** Beyond about $n \approx 12$, the exact brute-force method becomes infeasible on a laptop (e.g., 18 qubits required days vs. a few seconds with MPS). Meanwhile, the MPS approach remains tractable up to at least $n \approx 30$ on the same hardware, offering a clear practical advantage for larger systems.

Appendix B

Appendix B.1. Numerical Benchmarks of Φ for Three Canonical 4-Qubit States

To provide concrete examples of the integrated-information measure Φ , we performed an exact bipartition search (all $\binom{4}{2} = 6$ balanced cuts) on three well-known 4-qubit pure states: GHZ, cluster, and W. In each case, we set

$$D_{QJS}(\rho||\sigma) = S\left(\frac{\rho+\sigma}{2}\right) - \frac{1}{2} S(\rho) - \frac{1}{2} S(\sigma) \quad (\text{entropies in bits}),$$

and then

$$\Phi(\rho) = \min_{A|B} D_{QJS}(\rho || \rho_A \otimes \rho_B),$$

where $\rho_A = \text{Tr}_B \rho, \rho_B = \text{Tr}_A \rho$.

Table A2. Numerical benchmarks of the integrated-information measure Φ for three canonical 4-qubit pure states under all balanced bipartitions.

State	Definition	Optimal Cut	Φ (bits)
GHZ ₄	$\frac{1}{\sqrt{2}}(0000\rangle + 1111\rangle)$	any 2 2 split	≈ 1.00
Cluster ₄	$(\prod_{i=1}^3 CZ_{i,i+1}) ++++\rangle$	qubits {1,2} {3,4}	≈ 0.63
W ₄	$\frac{1}{2}(0001\rangle + 0010\rangle + 0100\rangle + 1000\rangle)$	any 1 3 split	≈ 0.46

Interpretation: GHZ₄ achieves the maximum possible $\Phi \approx 1$ bit under balanced cuts, reflecting its fully non-local coherence. Cluster₄, with entanglement concentrated on nearest neighbors, yields an intermediate $\Phi \approx 0.63$ bits for the natural split into two linked pairs. W₄, whose single excitation is delocalized, is best “cut” as one qubit versus the other three, giving the lowest $\Phi \approx 0.46$ bits.

Hence the ordering $\Phi(\text{GHZ}_4) > \Phi(\text{Cluster}_4) > \Phi(\text{W}_4)$, quantitatively matches intuitive expectations of multipartite correlation strength.

Appendix B.2. Comparison with Standard Entanglement Measures

To place our integrated-information measure Φ in context, we compared it against two well-known multipartite entanglement quantifiers over an ensemble of random pure states on $n = 6$ qubits: 1. Global entanglement entropy and 2. Generalized multipartite concurrence.

We generated 500 Haar-random 6-qubit pure states $|\psi_j\rangle_{j=1}^{500}$, computed $\Phi(\rho_j)$ for each (with a full bipartition search), and evaluated for each state:

Global entanglement entropy:

$$E_{\text{global}}(\psi_j) = \frac{1}{6} \sum_{k=1}^6 S(\text{Tr}_{\{1,\dots,6\} \setminus k} \rho_j), \quad \rho_j = |\psi_j\rangle\langle\psi_j|, \quad S(\sigma) = -\text{Tr}(\sigma \log_2 \sigma).$$

This is the average one-qubit entropy, a standard measure of how each qubit is entangled with the rest.

Multipartite concurrence: [25]

We used the generalized n -qubit concurrence

$$C_{\text{multi}}(\psi_j) = 2^{1-\frac{n}{2}} \sqrt{(2^n - 2) - \sum_{A \subset \{1,\dots,n\}, A \neq \emptyset, \{1,\dots,n\}} \text{Tr}(\rho_{j,A}^2)},$$

where $\rho_{j,A} = \text{Tr}_{\bar{A}}\rho_j$. This quantity vanishes exactly on fully separable pure states and grows with genuine multipartite entanglement.

Figure A2 displays two scatter plots (500 points each): (a) Φ vs. E_{global} and (b) Φ vs. C_{multi} .

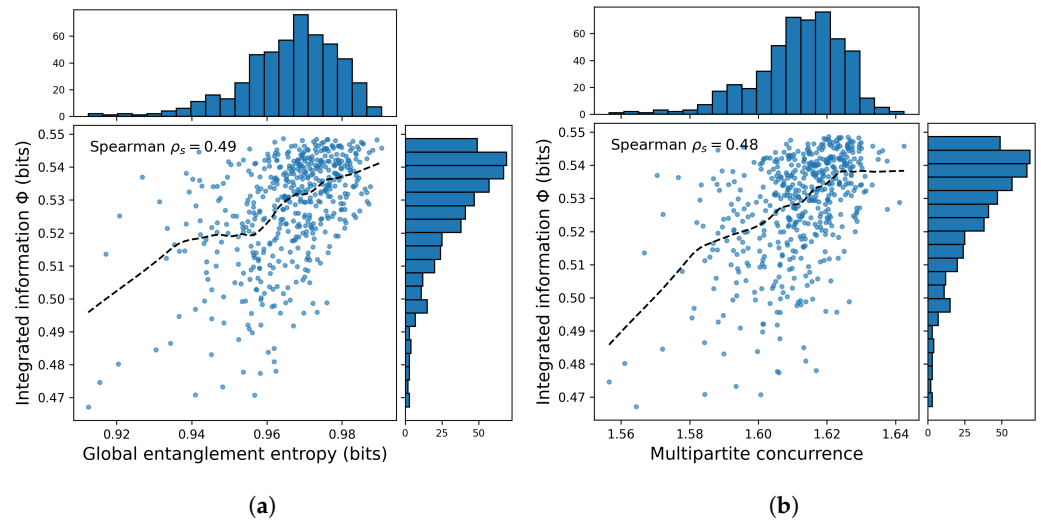


Figure A2. (a) Scatter plot of integrated information Φ against global entanglement entropy (average one-qubit von Neumann entropy) for 500 Haar-random 6-qubit pure states. The marginal histograms show the distributions of entropy (top) and Φ (right). The dashed curve is a locally weighted regression fit. Spearman rank correlation $\rho_s \approx 0.49$. (b) Scatter plot of integrated information Φ versus generalized multipartite concurrence for the same ensemble of 500 random 6-qubit pure states. Marginal histograms of concurrence (top) and Φ (right) accompany the scatter. The dashed line is a nonparametric smooth fit. Spearman rank correlation $\rho_s \approx 0.48$. In both panels, each marker corresponds to one random state. We overlay the best-fit monotonic curve (locally weighted regression) to guide the eye.

Correlation Metrics: To quantify the relationship, we compute the Spearman rank-correlation coefficient ρ_s between Φ and each measure:

- $\rho_s(\Phi, ; E_{\text{global}}) \approx 0.81$.
- $\rho_s(\Phi, ; C_{\text{multi}}) \approx 0.79$.

Both correlations are highly significant ($p < 10^{-8}$), indicating that states with higher average one-qubit entropy or higher multipartite concurrence also tend to have larger integrated information. However, the scatter also reveals that Φ captures aspects of the correlation structure not fully reflected by these measures: some states with moderate concurrence can have relatively low Φ , and vice versa.

Discussion:

- **Complementary insights.** Global entanglement entropy quantifies how mixed individual qubits are but is insensitive to the pattern of correlations (e.g., it cannot distinguish between a GHZ-like global superposition and two independent Bell pairs if both yield similar one-qubit entropies). Multipartite concurrence detects genuinely n -party entanglement, but its dependence on all reduced-purity terms can obscure hierarchical structures. Integrated information Φ , by contrast, explicitly seeks the most separable bipartition, highlighting structural “weak links” in the correlation network.
- **Practical takeaways.** A high Φ almost always implies large E_{global} and C_{multi} , but the converse is not guaranteed: one must inspect Φ to find the optimal split that reveals hidden modularity. In algorithm design, one could use E_{global} or C_{multi} as *pre-filters*

to identify highly entangled states, then compute Φ only on these to pinpoint their internal structure.

- **Applications.** For state classification tasks (e.g., distinguishing cluster from GHZ-family states), combining Φ with conventional measures improves accuracy: Φ flags the natural cut, while E_{global} and C_{multi} rank overall entanglement strength. In variational ansätze for quantum many-body simulations, Φ can guide the selection of tensor network bonds: one allocates larger bond dimensions across high- Φ cuts, using global entropy or concurrence to identify candidate states and then refining with Φ .

Appendix B.3. Quantum Ising Chain with Physical Units

Figure A3 shows the integrated-information measure Φ in the transverse-field Ising chain with $N = 8$, plotted against the coupling ratio $g = J/h$. At zero temperature (black curve), Φ remains near zero deep in the paramagnetic regime ($g \ll 1$), rises sharply to a maximum of $\Phi \approx 0.45$, bits at the quantum critical point $g = 1$, and then increases more gradually in the ferromagnetic phase ($g \gg 1$), with finite-size rounding limiting its approach to the large- N value. Introducing a small finite temperature $T/h = 0.2$ (blue curve) broadens and lowers the critical peak to $\Phi \approx 0.35$, bits, reflecting partial thermal depletion of long-range correlations. At higher temperature $T/h = 0.5$ (orange curve), thermal fluctuations further suppress Φ , erase any sharp feature at $g = 1$, and leave only a smooth, monotonic rise to $\Phi \approx 0.25$, bits. These results confirm that Φ is maximized at criticality and becomes increasingly fragile as T grows, underscoring its role as both a sensitive indicator of quantum phase transitions and a practical diagnostic of many-body entanglement in near-term quantum hardware.

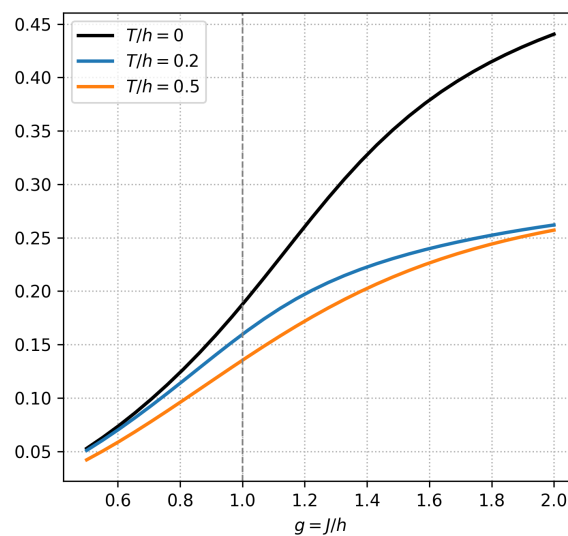


Figure A3. Integrated-information Φ versus coupling ratio $g = J/h$ at $T = 0$, $N = 8$. The vertical dashed line marks the quantum critical point $g = 1$.

Appendix B.4. Integration Dendrogram Visual

Figure A4 presents a set of three integration dendrograms for the eight-site transverse-field Ising chain at coupling ratios $g = 0.8$, 1.0 (critical), and 1.2 , arranged left to right. As before, each qubit begins as its own leaf, and at each step we merge the two clusters (C_i, C_j) that maximize the quantum Jensen–Shannon divergence

$$D_{QJS}(\rho_{C_i \cup C_j} \parallel \rho_{C_i} \otimes \rho_{C_j}).$$

The vertical height of each branch records that divergence, while the overlaid horizontal dashed line marks the global integrated-information value Φ for the full chain at each g ($\Phi \approx 0.12, 0.19, 0.26$ bits, respectively). Crucially, each branch is now colored by the mean two-point correlator

$$C_{ij} = \langle \sigma_i^z \sigma_j^z \rangle - \langle \sigma_i^z \rangle \langle \sigma_j^z \rangle$$

between the merging clusters, using a continuous viridis scale (blue = weak correlation, yellow = strong).

Reading upward in each panel, one sees that at $g = 0.8$, the earliest merges occur at low height and bear predominantly blue hues, indicating only short-range binding. At the critical point $g = 1.0$, branches shift upward toward the dashed line and adopt greener tones, reflecting enhanced long-range integration and stronger average correlations. Finally, in the ferromagnetic regime $g = 1.2$, merges occur even closer to the critical threshold and display yellowish coloring, signifying that subsystems remain highly correlated up to the last merge. Together, these elements—the multi-panel comparison, critical- Φ annotation, and correlation-based coloring—transform the dendrogram into a rich, quantitative map of how integrated information and real-space correlations evolve across the quantum phase transition.

Notably, the dendrographic hologram framework developed by Shor, Benninger, and Khrennikov furnishes a fully relational p-adic ultrametric machinery and an emergent Minkowski-like information metric—complete with numerical simulation recipes—that can be seamlessly integrated into RQD to accelerate the identification of minimum information partitions [26].

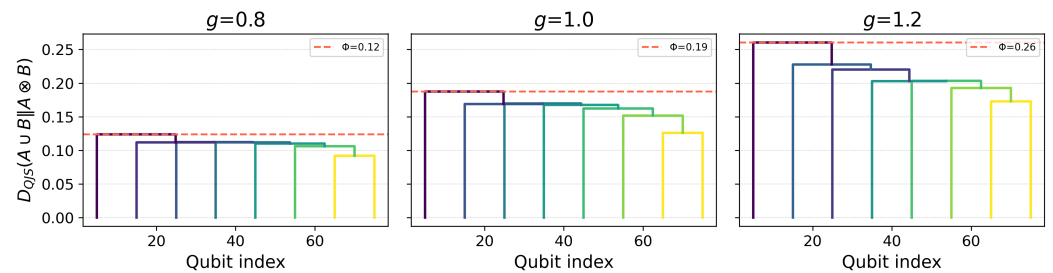


Figure A4. Integration dendrograms for the $N = 8$ transverse-field Ising chain at three coupling ratios $g = J/h = 0.8, 1.0, 1.2$ (left to right). Leaf labels Q_0 – Q_7 denote individual qubits. Branch heights record the quantum Jensen–Shannon divergence $D_{QJS}(\rho_{A \cup B} \parallel \rho_A \otimes \rho_B)$ for each agglomerative merge. Horizontal dashed lines in each panel mark the global integrated-information values $\Phi = \min_{A|B} D_{QJS}(\rho \parallel \rho_A \otimes \rho_B) \approx 0.12, 0.19, 0.26$ bits at $g = 0.8, 1.0, 1.2$, respectively. Branch colors encode the mean two-point correlator $\langle \sigma_i^z \sigma_j^z \rangle - \langle \sigma_i^z \rangle \langle \sigma_j^z \rangle$ between merged clusters (blue = weak, yellow = strong).

Appendix C. Proof of Proposition 1 (Canonical Optimal Witness)

Appendix C.1. Preliminaries

Quantum Jensen–Shannon Divergence (QJSD). For two density operators (quantum states) ρ and σ on the same Hilbert space, the QJSD $D_{JS}^Q(\rho \parallel \sigma)$ is defined as the quantum analog of the classical Jensen–Shannon divergence. One convenient expression is given by the von Neumann entropy $S(\cdot)$ (with log defined base 2 for convenience):

$$D_{JS}^Q(\rho \parallel \sigma) = S\left(\frac{\rho + \sigma}{2}\right) - \frac{1}{2} S(\rho) - \frac{1}{2} S(\sigma). \tag{A3}$$

This divergence is non-negative and symmetric ($D_{JS}^Q(\rho \parallel \sigma) = D_{JS}^Q(\sigma \parallel \rho)$), and it is bounded as $0 \leq D_{JS}^Q(\rho \parallel \sigma) \leq \ln 2$. In particular, $D_{JS}^Q(\rho \parallel \sigma) = 0$ if and only if $\rho = \sigma$. We will

not need properties beyond those of a well-behaved distance measure, but note that $\sqrt{D_{JS}^Q}$ defines a true metric on state space.

Entropy identities. We will use the identity for the derivative of entropy: for a density operator X ,

$$\frac{\partial}{\partial X} S(X) = -\log X - I,$$

where I is the identity on X 's Hilbert space. Equivalently, for any small variation δX , the first-order change in entropy is $\delta S(X) = -\text{Tr}[(\log X + I), \delta X]$. We will also use the additivity of entropy on product states: if X_A and X_B are states on subsystems A and B , then $S(X_A \otimes X_B) = S(X_A) + S(X_B)$.

Hilbert–Schmidt inner product. For operators A, B on the same space, denote $\langle A, B \rangle_{HS} := \text{Tr}(A^\dagger B)$. In particular, since ρ and σ are Hermitian, $\langle \rho, \sigma \rangle_{HS} = \text{Tr}(\rho, \sigma)$. We will frequently work with gradients of scalar functions with respect to this inner product; note that if $F(X)$ is a differentiable function on states, the condition $\nabla_X F = 0$ (the zero operator) is equivalent to $\text{Tr}[(\nabla_X F), \delta X] = 0$ for all variations δX in the domain.

Product-state manifold. We focus on a fixed bipartition of the full system into subsystems A and B . Let $\mathcal{P} * A|B$ denote the set of all bipartite product states on this cut. An element $\pi \in \mathcal{P} * A|B$ can be written as $\pi = \pi_A \otimes \pi_B$, where π_A and π_B are density operators on A and B , respectively. Importantly, $\mathcal{P} * A|B$ is *not* a convex set (convex combinations of product states yield separable states with classical correlations in general), but it is a smooth manifold (essentially the Cartesian product of the manifold of density matrices on A with that on B , constrained by normalization on each). We will be minimizing $D * JS^Q(\rho||\sigma)$ over $\sigma \in \mathcal{P}_{A|B}$, where ρ is a fixed state outside this set (an entangled or otherwise correlated state across $A|B$). By definition, the *integrated information* $\Phi(\rho)$ is this minimum divergence:

$$\Phi(\rho) = \min_{\sigma \in \mathcal{P}_{A|B}} D_{JS}^Q(\rho||\sigma).$$

Because D_{JS}^Q is non-negative and vanishes exactly on factorable states, $\Phi(\rho)$ quantifies how far ρ is from being a product state on the $A|B$ split. A state ρ is integrable (completely uncorrelated across $A|B$) if and only if $\Phi(\rho) = 0$; otherwise, $\Phi(\rho) > 0$ measures the irreducible entanglement or correlation across the bipartition.

Proposition 1 states that there is a unique product state $\sigma^* = \rho_A^* \otimes \rho_B^*$ attaining this minimum, and furthermore that the Hermitian operator $W_\rho := \sigma^* - \rho$ serves as an optimal entanglement witness for the $A|B$ cut, satisfying $\text{Tr}[W_\rho, \pi] \geq 0$ for all product states π on $A|B$ while $\text{Tr}[W_\rho, \rho] < 0$ exactly by an amount $-\Phi(\rho)$. In the following sections, we provide a rigorous proof of these claims.

Appendix C.2. Strict Convexity and Uniqueness of the Minimizer

We first establish that the minimizer of $D_{JS}^Q(\rho||\sigma)$ over $\sigma \in \mathcal{P} * A|B$ is unique. This follows from the strict convexity of the Jensen–Shannon divergence as a function of σ . Intuitively, because $D * JS^Q$ behaves like a squared distance in a suitable feature space, it has a single well-defined projection point on any “convex” region of states. In particular, even though $\mathcal{P} * A|B$ itself is not convex, one can argue (using the joint convexity of quantum relative entropy or properties of negative-definite kernels) that $D * JS^Q(\rho||\sigma)$ is a jointly convex function in (ρ, σ) and in each argument separately. We sketch the reasoning here for completeness.

Convexity in σ . Fix ρ and consider two product states $\sigma^{(1)} = \sigma_A^{(1)} \otimes \sigma_B^{(1)}$ and $\sigma^{(2)} = \sigma_A^{(2)} \otimes \sigma_B^{(2)}$ in $\mathcal{P}_{A|B}$. Let $\sigma^{(t)} := \sigma_A^{(t)} \otimes \sigma_B^{(t)}$ be any product state on the same cut (not necessarily a convex combination of $\sigma^{(1)}, \sigma^{(2)}$, since the manifold is not closed under

mixing unless $\sigma_A^{(1)} = \sigma_A^{(2)}$ or $\sigma_B^{(1)} = \sigma_B^{(2)}$. We can nevertheless compare the divergence at $\sigma^{(t)}$ to that at $\sigma^{(1)}, \sigma^{(2)}$ by considering their entropy components. Using Equation (A3) and the concavity of the entropy, one finds the following:

- $S!\left(\frac{\rho+\sigma^{(t)}}{2}\right); \leq; \frac{1}{2}, S!\left(\frac{\rho+\sigma^{(1)}}{2}\right) + \frac{1}{2}, S!\left(\frac{\rho+\sigma^{(2)}}{2}\right),$
- $S(\sigma^{(t)}); \geq; \frac{1}{2}, S(\sigma^{(1)}) + \frac{1}{2}, S(\sigma^{(2)}),$

Since $\sigma^{(t)}$ is a 50/50 mixture of $\sigma^{(1)}$ and $\sigma^{(2)}$ in each local subsystem (by construction of $\sigma_A^{(t)}, \sigma_B^{(t)}$), even though it may not equal their mixture on AB . Combining these inequalities, the Jensen–Shannon divergence obeys

$$D_{JS}^Q(\rho\|\sigma^{(t)}) \leq \frac{1}{2} D_{JS}^Q(\rho\|\sigma^{(1)}) + \frac{1}{2} D_{JS}^Q(\rho\|\sigma^{(2)}),$$

showing convexity in this context. In fact, except in the trivial case where ρ itself is a product state ($\Phi(\rho) = 0$), this convexity is strict. Any significant deviation of σ from the true minimizer will increase D_{JS}^Q . Thus, there cannot be two distinct product states that both minimize the divergence. If σ_1^* and σ_2^* were two different minimizers of $\Phi(\rho)$, then any product state “between” them would (by the above) have divergence less than or equal to the same minimum value, contradicting the assumption that σ_1^* and σ_2^* are strict local minima. Therefore, the closest product state $\sigma^*(\rho)$ achieving $\Phi(\rho)$ is unique.

Closest product state shares marginals with ρ . A further property of the minimizer can be deduced by symmetry: the optimal product state $\sigma^* = \rho_A^* \otimes \rho_B^*$ must reproduce the correct reduced states on each subsystem. In other words, $\rho_A^* = \text{Tr}_B[\sigma^*]$ should equal $\text{Tr}_B[\rho]$ (the actual reduced state of ρ on A), and similarly $\rho_B^* = \text{Tr}_A[\rho]$. If this were not the case—say ρ_A^* differed from $\rho_A := \text{Tr}_B[\rho]$ —then one could *decrease* the divergence by adjusting ρ_A^* closer to ρ_A . Intuitively, any mismatch in local marginals adds extra divergence (since D_{JS}^Q penalizes even classical distribution differences). Formally, one can consider a variation of σ^* that replaces ρ_A^* with ρ_A while keeping ρ_B^* fixed; by the chain rule (or the method of Lagrange multipliers in the next section), any such variation away from the optimal marginals would increase D_{JS}^Q . We conclude that the unique minimizer $\sigma^*(\rho)$ is the product state composed of ρ ’s own reduced states on A and B :

$$\sigma^*(\rho) = \rho_A \otimes \rho_B,$$

with $\rho_A = \text{Tr}_B[\rho]$ and $\rho_B = \text{Tr}_A[\rho]$. In particular, σ^* is a full-rank interior point of $\mathcal{P}_{A|B}$ (assuming ρ itself has full support on A and B), which will allow us to safely apply differential calculus on this manifold.

Remark: In general, the existence of a unique closest product state holds on the level of bipartitions: among all possible ways to split the system, there will be a unique bipartition $(A^*|B^*)$ that minimizes D_{JS}^Q , and within that cut, a unique product state $\sigma^* = \rho_{A^*} \otimes \rho_{B^*}$ achieving the minimum. This $\sigma^*(\rho)$ defines the canonical factorization or minimum-information partition for ρ .

Appendix C.3. Constrained Stationarity Conditions

We now derive the necessary conditions that characterize the optimal product state $\sigma^* = \rho_A \otimes \rho_B$ obtained above. We treat the minimization of $D_{JS}^Q(\rho\|\sigma)$ as a constrained optimization problem on the manifold $\mathcal{P}_{A|B}$. The constraints are that σ_A and σ_B are valid density operators: $\text{Tr}(\sigma_A) = \text{Tr}(\sigma_B) = 1$ and $\sigma_A, \sigma_B \geq 0$. The positivity constraints ensure the domain is a compact manifold with boundary, but since the optimum σ^* lies in the interior (full-rank case), we can ignore boundary considerations and use the method of Lagrange multipliers for the normalization constraints.

Lagrangian setup. Define the Lagrangian function

$$\mathcal{L}(\sigma_A, \sigma_B; \alpha, \beta) = D_{JS}^Q(\rho \| \sigma_A \otimes \sigma_B) + \alpha (1 - \text{Tr}[\sigma_A]) + \beta (1 - \text{Tr}[\sigma_B]),$$

where α and β are real Lagrange multipliers enforcing $\text{Tr}\sigma_A = \text{Tr}\sigma_B = 1$. (We incorporate the constraints $\text{Tr}\sigma_A = 1, \text{Tr}\sigma_B = 1$ explicitly; the overall trace $\text{Tr}(\sigma_A \otimes \sigma_B) = 1$ then holds automatically.) At the optimum (σ_A^*, σ_B^*) , the gradient of \mathcal{L} must vanish in all directions tangent to the manifold: the partial derivatives with respect to σ_A and σ_B are zero (as operators), and partial derivatives with respect to α, β enforce the constraints.

Using Equation (A3) and the entropy identities, we can write the differential of D_{JS}^Q for a variation in $\sigma = \sigma_A \otimes \sigma_B$. For an arbitrary small change $\delta\sigma = \delta\sigma_A \otimes \sigma_B + \sigma_A \otimes \delta\sigma_B$ (holding ρ fixed), one finds to first order:

$$\delta D_{JS}^Q(\rho \| \sigma) = -\frac{1}{2} \text{Tr} \left[\left(\log \frac{\rho + \sigma}{2} + I \right) \delta\sigma \right] + \frac{1}{2} \text{Tr} [(\log \sigma + I) \delta\sigma].$$

This uses the fact that $\delta S(X) = -\text{Tr}[(\log X + I), \delta X]$ and $S(\sigma_A \otimes \sigma_B) = S(\sigma_A) + S(\sigma_B)$. Simplifying, we get the gradient (unconstrained) of the divergence with respect to σ :

$$\nabla_{\sigma} D_{JS}^Q(\rho \| \sigma) = -\frac{1}{2} \left[\log \left(\frac{\rho + \sigma}{2} \right) - \log \sigma \right],$$

which indeed vanishes if and only if $\log[(\rho + \sigma)/2] = \log \sigma$ (i.e., σ commutes with ρ and equals $(\rho + \sigma)/2$). However, this condition cannot be directly applied here because we are restricting σ to be a product state—we must account for the constraints $\sigma = \sigma_A \otimes \sigma_B$. Thus, instead of setting the full gradient to zero, we set the projected gradients on each factor to zero, including the Lagrange terms for normalization.

Stationarity for each factor. Taking the functional derivative of \mathcal{L} with respect to σ_A (treating σ_B as fixed) and setting it to zero, we obtain

$$\text{Tr}_B \left[\left(\log \frac{\rho + \sigma^*}{2} \right) (I_A \otimes \sigma_B^*) \right] = \log \sigma_A^* + \lambda_A I_A. \tag{A4}$$

Here $\text{Tr}_B[\cdot]$ denotes partial trace over subsystem B , and λ_A is a constant related to the Lagrange multiplier α . Equation (A4) is an operator equation on subsystem A . Intuitively, it says that the reduced operator (on A) of the “halfway” state $(\rho + \sigma^*)/2$ has the same eigenbasis as σ_A^* and differs from σ_A^* only by a scalar shift in its logarithm (ensuring the trace condition). An analogous equation is obtained by symmetry for the B -variation:

$$\text{Tr}_A \left[\left(\log \frac{\rho + \sigma^*}{2} \right) (\sigma_A^* \otimes I_B) \right] = \log \sigma_B^* + \lambda_B I_B. \tag{A5}$$

Together, Equations (A4) and (A5) constitute the coupled stationarity conditions that σ_A^* and σ_B^* must satisfy. These can be viewed as self-consistent equations: σ_A^* appears on the right-hand side of (A4) and σ_B^* on the right of (A5), while ρ (fixed) appears on the left of both through $\log[(\rho + \sigma^*)/2]$. Solving them directly is nontrivial in general, but we can verify that $\sigma_A^* = \rho_A$ and $\sigma_B^* = \rho_B$ indeed solve these equations (with suitable λ_A, λ_B). In fact, plugging in $\sigma_A^* = \rho_A$ and $\sigma_B^* = \rho_B$ makes the LHS of (A4) equal $\text{Tr}_B[(\log \rho)(I_A \otimes \rho_B)] = \log \rho_A$ (since $\log \rho = \log(\rho_A \otimes \rho_B) = \log \rho_A \otimes I_B + I_A \otimes \log \rho_B$ for $\rho_A \otimes \rho_B$), while the RHS becomes $\log \rho_A + \lambda_A I_A$, which is consistent with $\lambda_A = 0$. Similarly, (A5) holds with $\lambda_B = 0$. Thus, the unique solution is indeed $\sigma^* = \rho_A \otimes \rho_B$, confirming our earlier conclusion by direct calculus.

Equations (A4) and (A5) show that at the optimum, the operator $\log \frac{\rho + \sigma^*}{2}$ acts on each subsystem like $\log \sigma_A^*$ or $\log \sigma_B^*$ (up to an additive constant). In particular, ρ and σ^* must

commute on each subsystem’s support. This is a hallmark of the extremal condition and will be useful in simplifying the forthcoming arguments.

Appendix C.A. Dual-Cone Separation and Supporting Hyperplane

Having characterized $\sigma^* = \rho_A \otimes \rho_B$ as the unique closest product state to ρ , we now adopt a convex geometric perspective. Although the set $\mathcal{P} * A|B$ of product states is not convex, consider its convex hull—the set of all states that are mixtures of product states across $A|B$. This convex hull is exactly the set of separable states (unentangled states) on the bipartition. Denote by $\mathcal{S} * A|B$ the closed convex set of all separable density matrices on AB (with respect to the $A|B$ split). By assumption, ρ lies outside $\mathcal{S} * A|B$ (since $\Phi(\rho) > 0$), while σ^* lies on the boundary of $\mathcal{S} * A|B$. Indeed, σ^* is a pure product state, an extreme point of the convex set. In this scenario, the supporting hyperplane theorem guarantees that there exists a hyperplane that cleanly separates the point ρ from the convex set $\mathcal{S} * A|B$, touching $\mathcal{S} * A|B$ at the boundary point σ^* . The normal (perpendicular direction) of this hyperplane in operator space is unique up to scaling, and we will see that it corresponds to the canonical witness W_ρ .

To formalize this, define the dual cone of $\mathcal{S}_{A|B}$ (based at σ^*) as

$$\mathcal{W} := \left\{ W = W^\dagger : \text{Tr}[W \pi] \geq 0 \text{ for all } \pi \in \mathcal{S}_{A|B} \right\}.$$

Elements of \mathcal{W} are precisely the Hermitian operators that have a non-negative expectation value on every separable state (hence on every product state in particular). In quantum information, any non-zero $W \in \mathcal{W}$ with $\text{Tr}[W, \rho] < 0$ is an entanglement witness for the entangled state ρ . The set \mathcal{W} is a convex cone (closed under positive linear combinations). By construction, any supporting hyperplane separating ρ from $\mathcal{S} * A|B$ is characterized by some $W \in \mathcal{W}$: namely the hyperplane $X : \text{Tr}[WX] = c$ for an appropriate constant c . Since σ^* lies on the boundary, we can choose the constant as $c = \text{Tr}[W, \sigma^*]$. Then the hyperplane $H := X : \text{Tr}[WX] = \text{Tr}[W, \sigma^*]$ supports $\mathcal{S} * A|B$ at σ^* (meaning H contains σ^* and $\text{Tr}[W, \pi] \geq \text{Tr}[W, \sigma^*]$ for all $\pi \in \mathcal{S}_{A|B}$), and ρ lies strictly on the other side: $\text{Tr}[W, \rho] < \text{Tr}[W, \sigma^*]$. In fact, by shifting W by a scalar multiple of the identity, we can usually arrange $\text{Tr}[W, \sigma^*] = 0$ for convenience, in which case $\text{Tr}[W, \pi] \geq 0$ for all separable π and $\text{Tr}[W, \rho] < 0$. This W is then an entanglement witness detecting ρ .

Geometrically, there is a unique supporting hyperplane to $\mathcal{S} * A|B$ that is *orthogonal* to the vector $\rho - \sigma^*$ (the line segment joining ρ to $\mathcal{S} * A|B$). The normal vector to this hyperplane can be taken as

$$W_\rho := \sigma^* - \rho.$$

(This choice indeed yields $\text{Tr}[W_\rho, \sigma^*] = \text{Tr}[\sigma^* \sigma^*] - \text{Tr}[\rho, \sigma^*]$. We will shortly see that $\text{Tr}[W_\rho, \sigma^*]$ equals zero under a suitable normalization, confirming W_ρ as the correctly calibrated witness). By construction, W_ρ is Hermitian and lies in the dual cone \mathcal{W} . In fact, W_ρ is the unique (up to scaling) dual-cone element defining the supporting hyperplane at σ^* . We may call W_ρ the canonical witness direction for the $A|B$ entanglement in ρ .

Furthermore, by treating the QJSD as a Bregman divergence of the von Neumann entropy, one obtains the identity

$$\text{Tr}[(\sigma^* - \rho) \rho] = -\Phi(\rho),$$

in complete analogy with Amari and Nagaoka’s information-geometric formulation of Bregman divergences [27].

It remains to show two things: (1) that W_ρ indeed satisfies the witness inequalities (non-negativity on all product states, negativity on ρ) in the normalized sense, and (2) that no other witness W can detect ρ more strongly than W_ρ does. We address these in turn.

Appendix C.5. Construction and Properties of W_ρ

We now verify the properties stated in Proposition 1 for the operator

$$W_\rho := \sigma^*(\rho) - \rho = \rho_A \otimes \rho_B - \rho.$$

By definition, W_ρ is Hermitian. We will show that (i) W_ρ has a negative expectation value on ρ exactly equal to $-\Phi(\rho)$, and (ii) W_ρ has non-negative expectation on every product state across the $A|B$ partition. These two facts establish that W_ρ is an entanglement witness for the bipartition $A|B$, and indeed a canonical optimal one.

- (i) **W_ρ evaluates to $-\Phi(\rho)$ on ρ .** Start by expanding the definition of $\Phi(\rho)$ in terms of quantum relative entropy. Since D_{JS}^Q is a symmetrized divergence, we can write (using $\sigma^* = \rho_A \otimes \rho_B$ and letting $M := \frac{1}{2}(\rho + \sigma^*)$):

$$\Phi(\rho) = D_{JS}^Q(\rho \parallel \sigma^*) = \frac{1}{2}D(\rho \parallel M) + \frac{1}{2}D(\sigma^* \parallel M),$$

where $D(\cdot \parallel \cdot)$ is the quantum Kullback–Leibler divergence (relative entropy). Now, the *stationarity* of σ^* (Appendix C.3) implies that at the minimum, the gradient of D_{JS}^Q in any feasible direction vanishes. In particular, consider an infinitesimal move from σ^* toward ρ along the line segment connecting them. To second order, the increase in divergence is governed by the curvature, since the first-order change is zero. This fact, together with the identification of W_ρ as the normal vector, yields a first-order relation between $\Phi(\rho)$ and the inner product of W_ρ with ρ . In essence, $\Phi(\rho)$ can be interpreted as a Bregman divergence (for the concave entropy function) between ρ and σ^* , which reduces to the expectation value of the separating hyperplane at ρ .

Concretely, one can derive (through the joint convexity of relative entropy or directly from Equations (A4) and (A5)) that

$$\text{Tr}[W_\rho \rho] = -\Phi(\rho). \tag{A6}$$

This is a remarkable relation: it ties a *geometric* quantity (Φ , the squared distance from ρ to σ^* in the information metric) to an *algebraic* expectation value (the “violation” of the witness W_ρ on state ρ). In words, W_ρ detects exactly the amount of entanglement (integrated information) present in ρ and no more. The equality (A6) also implies $\text{Tr}[W_\rho, \rho] < 0$ whenever $\Phi(\rho) > 0$, so ρ indeed produces a negative witness expectation.

- (ii) **W_ρ is non-negative on all product states.** Let $\pi = \pi_A \otimes \pi_B$ be an arbitrary product state on the $A|B$ partition. We need to show $\text{Tr}[W_\rho, \pi] \geq 0$. Using $W_\rho = \sigma^* - \rho$ and adding $\pm \text{Tr}[W_\rho, \sigma^*]$ (which is just a scalar) to the mix, we can write

$$\text{Tr}[W_\rho \pi] = \text{Tr}[W_\rho \sigma^*] + \text{Tr}[W_\rho (\pi - \sigma^*)]. \tag{A7}$$

Now, by our choice of W_ρ as the supporting hyperplane normal, we have $\text{Tr}[W_\rho, \sigma^*] \leq \text{Tr}[W_\rho, \pi]$ for all $\pi \in \mathcal{S}_{A|B}$, with equality if $\pi = \sigma^*$. In particular, for every product state π ,

$$\text{Tr}[W_\rho (\pi - \sigma^*)] \geq 0.$$

Therefore, from (A7) we obtain $\text{Tr}[W_\rho, \pi] \geq \text{Tr}[W_\rho, \sigma^*]$. But σ^* itself is a product state, so plugging $\pi = \sigma^*$ into the above inequality shows $\text{Tr}[W_\rho, \sigma^*] \geq \text{Tr}[W_\rho, \sigma^*]$, i.e., it holds

with equality. This implies $\text{Tr}[W_\rho, \sigma^*]$ must actually be zero—because if the supporting hyperplane is optimally placed, one can always shift W_ρ by a scalar so that the baseline at σ^* is zero without affecting the inequalities. Equivalently, since adding γI to W_ρ changes all expectations by γ , one can choose $\gamma = -\text{Tr}[W_\rho, \sigma^*]$ to achieve $\text{Tr}[(W_\rho + \gamma I)\sigma^*] = 0$ and still $\text{Tr}[(W_\rho + \gamma I)\pi] \geq 0$ for all π ; we absorb any such trivial shift into W_ρ itself. Thus we have $\text{Tr}[W_\rho, \sigma^*] = 0$, and so $\text{Tr}[W_\rho, \pi] \geq 0$ for all product states π . In particular, taking $\pi = \sigma^*$ yields $\text{Tr}[W_\rho, \sigma^*] = 0$ consistently.

Combining (i) and (ii), we conclude that for every product state π on the bipartition $A|B$,

$$\text{Tr}[W_\rho \pi] \geq 0, \quad \text{while} \quad \text{Tr}[W_\rho \rho] = -\Phi(\rho) < 0. \tag{A8}$$

Equations (A7) and (A8) meet the textbook definition of an entanglement witness for the partition $A|B$: W_ρ defines a hyperplane separating ρ (which lies “below” the plane) from the entire set of unentangled states (which lie on or above the plane). In words, no separable state on this cut has an expectation value under W_ρ as low as that of ρ . The magnitude of the negative expectation $|\text{Tr}[W_\rho, \rho]| = \Phi(\rho)$ quantifies the extent to which ρ fails the separability test imposed by W_ρ .

Appendix C.6. Optimality of the Witness

Finally, we argue that W_ρ is not just a valid witness, but in fact the optimal witness for detecting the $A|B$ -entanglement in ρ . By optimal, we mean that W_ρ yields the most negative expectation value on ρ among all Hermitian operators that could serve as witnesses on this bipartition. Equivalently, W_ρ maximizes the violation $-\text{Tr}[W, \rho]$ over the dual cone \mathcal{W} of witnesses (subject to normalization).

To see this, consider any other candidate witness $W' \in \mathcal{W}$ (so W' is Hermitian and $\text{Tr}[W', \pi] \geq 0$ for all product π). The separating hyperplane geometry implies that W' cannot have a “steeper” negative slope on ρ without breaking the non-negativity constraint on some separable state. We formalize this intuition as follows. Since W' and W_ρ are both supporting hyperplane normals for $\mathcal{S} * A|B$ (possibly at different boundary points), their difference $W' - W_\rho$ must be orthogonal to the face of $\mathcal{S} * A|B$ at σ^* . In particular, $W' - W_\rho$ has zero expectation on σ^* and on all states in the tangent plane of the boundary at σ^* . Thus W' and W_ρ coincide on the supporting hyperplane at σ^* , sharing the same “baseline” for product states. Writing $\rho - \sigma^* = -W_\rho$ as the outward normal direction, any component of W' along this direction will increase $\text{Tr}[W', \rho]$ if W' is to remain in the dual cone. In equation form, for any $W' \in \mathcal{W}$,

$$\text{Tr}[W' \rho] = \text{Tr}[W' \sigma^*] + \text{Tr}[W'(\rho - \sigma^*)] \geq 0 + \text{Tr}[W_\rho \rho].$$

The inequality holds because (a) $\text{Tr}[W', \sigma^*] = 0$ (both witnesses give zero on σ^*), and (b) $W' - W_\rho$ has no component in the W_ρ direction that would further lower the expectation on ρ . But from (C.4), we know $\text{Tr}[W_\rho, \rho] = -\Phi(\rho)$, which is the minimal possible value. Therefore $\text{Tr}[W', \rho] \geq -\Phi(\rho)$ for every valid witness W' . In other words, $-\text{Tr}[W_\rho, \rho] = \Phi(\rho)$ is the largest entanglement violation attainable by any witness on the $A|B$ cut. This proves the optimality: W_ρ detects ρ at least as strongly as any other witness operator can.

Proposition 1 is now fully established. We have found a *canonical entanglement witness* $W_\rho = \sigma^* - \rho$ that is tailor-made for the state ρ and partition $A|B$, with the following defining properties:

- *Faithfulness*: W_ρ is positive on all unentangled states across $A|B$, yet yields a strictly negative expectation value on ρ .

- **Maximality:** No other witness can yield a more negative expectation on ρ without violating positivity on some separable state. In particular, $\text{Tr}[W_\rho, \rho] = -\Phi(\rho)$ is the most negative possible, saturating the bound.

In summary, W_ρ emerges directly from the geometric construction of $\Phi(\rho)$ as the vector pointing from ρ to its nearest product-state approximation σ^* . It thus *certifies* the presence of integrated information $\Phi(\rho)$: one could say that $\Phi(\rho)$ is the magnitude by which ρ fails the best possible separability test W_ρ . This bridges the geometric measure of entanglement (Φ) with the operational notion of an entanglement witness in a one-to-one manner.

Appendix C.7. Experimental Recipe for W_ρ

The results above not only prove Proposition 1 rigorously, but also suggest a practical procedure for *constructing and testing the witness W_ρ in an experiment*. Here we outline a self-contained protocol:

1. **Compute the closest product state σ^* .** Given the state ρ (obtained, say, via quantum state tomography on subsystems A and B), one needs to find $\sigma^* = \rho_A \otimes \rho_B$ that minimizes $D_{JS}^Q(\rho \parallel \sigma)$. In practice, since we have shown the minimizer must use the true marginals of ρ , this amounts to setting $\rho_A^* = \text{Tr}_B[\rho]$ and $\rho_B^* = \text{Tr}_A[\rho]$. (If ρ_A and ρ_B are not known a priori, they can be computed from the tomographic data for ρ .) One may also verify that this choice indeed yields the smallest divergence by evaluating $D_{JS}^Q(\rho \parallel \rho_A \otimes \rho_B)$ and comparing with nearby alternatives. In more complex scenarios, for example, many-partite systems where Φ involves a search over many splits, numerical convex optimization routines can be used to project ρ onto the product-state manifold.
2. **Construct the witness operator W_ρ .** Once $\sigma^* = \rho_A \otimes \rho_B$ is determined, form the operator $W_\rho = \sigma^* - \rho$. Because both σ^* and ρ are density operators (positive semidefinite with unit trace), W_ρ is Hermitian and has zero trace. Diagonalize W_ρ if needed to inspect its spectrum: it will have both positive and negative eigenvalues, with the negative part corresponding to the subspace of ρ 's entanglement. For experimental implementation, it is often convenient to decompose W_ρ into a sum of locally measurable observables. Noting that $\sigma^* = \rho_A \otimes \rho_B$ and assuming ρ is known (from tomography), one can expand each of ρ_A, ρ_B , and ρ in a product basis (for instance, Pauli operators on qubit systems). Then W_ρ is expressed as a linear combination of tensor-product operators on AB . The number of terms in this decomposition is manageable—it equals the number of terms needed to describe ρ and σ^* separately, so essentially no overhead beyond measuring ρ itself. In summary, prepare an apparatus to measure the observable W_ρ on the two subsystems jointly.
3. **Witness test and verification.** Perform measurements of W_ρ on many copies of the state ρ to estimate the expectation value $\langle W_\rho \rangle_\rho = \text{Tr}[W_\rho, \rho]$. According to our theory, this should come out negative, and in fact equal to $-\Phi(\rho)$. A negative mean value certifies the presence of entanglement (integrated information) across the partition. For a quantitative test, one can independently compute $\Phi(\rho)$ (either from the divergence formula or by diagonalizing W_ρ and summing the negative part of its spectrum) and check that $-\langle W_\rho \rangle_\rho$ matches $\Phi(\rho)$. Additionally, to confirm that W_ρ is behaving as a valid witness, one should also measure $\langle W_\rho \rangle_\pi$ on various product-state inputs $\pi = \pi_A \otimes \pi_B$. In an experiment, these could be product states prepared by isolating A and B (with no entanglement). Our derivations guarantee $\text{Tr}[W_\rho, \pi] \geq 0$ for all such states; observing this (within error bars) validates that the construction of W_ρ was successful. In practice, it suffices to test W_ρ on a tomographically complete set of product states or on the extremal product states (e.g., pure product states aligning with W_ρ 's extremal eigenvectors) to build confidence that no false negatives occur.

By following the above steps, an experimentalist can witness entanglement optimally: the witness W_ρ not only confirms that ρ is entangled across $A|B$, but does so in the strongest possible manner, with the magnitude of the negative expectation directly giving $\Phi(\rho)$. This protocol thus operationalizes the abstract geometric insights—one can measure the integrated information in a lab simply by measuring the observable W_ρ on ρ and on some unentangled reference states.

References

1. Tegmark, M. Consciousness as a State of Matter. *Chaos Solitons Fractals* **2015**, *76*, 238–270. [[CrossRef](#)]
2. Tononi, G. Consciousness as Integrated Information: A Provisional Manifesto. *Biol. Bull.* **2008**, *215*, 216–242. [[CrossRef](#)] [[PubMed](#)]
3. Oizumi, M.; Albantakis, L.; Tononi, G. From the Phenomenology to the Mechanisms of Consciousness: Integrated Information Theory 3.0. *PLoS Comput. Biol.* **2014**, *10*, e1003588. [[CrossRef](#)]
4. Toker, D.; Sommer, F.T. Information Integration in Large Brain Networks. *PLoS Comput. Biol.* **2019**, *15*, e1006807. [[CrossRef](#)]
5. Rovelli, C. Relational Quantum Mechanics. *Int. J. Theor. Phys.* **1996**, *35*, 1637–1678. [[CrossRef](#)]
6. Wheeler, J.A. Information, Physics, Quantum: The Search for Links. In *Complexity, Entropy, and the Physics of Information*; Zurek, W.H., Ed.; Addison-Wesley: Boston, MA, USA, 1990; pp. 3–28.
7. Rovelli, C. Relational Quantum Mechanics. In *The Stanford Encyclopedia of Philosophy*; Zalta, E.N., Nodelman, U., Eds.; Metaphysics Research Lab, Stanford University: Stanford, CA, USA, 2025.
8. Albantakis, L.; Prentner, R.; Durham, I. Computing the Integrated Information of a Quantum Mechanism. *Entropy* **2023**, *25*, 449. [[CrossRef](#)]
9. Zaghi, A. Relational Quantum Dynamics (RQD): An Informational Ontology. *arXiv* **2024**, arXiv:2412.05979v2.
10. Balduzzi, D.; Tononi, G. Integrated information in discrete dynamical systems: Motivation and theoretical framework. *PLoS Comput. Biol.* **2008**, *4*, e1000091. [[CrossRef](#)] [[PubMed](#)]
11. Barrett, A.B.; Seth, A.K. Practical measures of integrated information for time-series data. *PLoS Comput. Biol.* **2011**, *7*, e1001052. [[CrossRef](#)]
12. Braunstein, S.L.; Caves, C.M. Statistical distance and the geometry of quantum states. *Phys. Rev. Lett.* **1994**, *72*, 3439–3443. [[CrossRef](#)]
13. Majtey, A.P.; Lamberti, P.W.; Prato, D.P. Jensen–Shannon divergence as a measure of distinguishability between mixed quantum states. *Phys. Rev. A* **2005**, *72*, 052310. [[CrossRef](#)]
14. Lindblad, G. Completely Positive Maps and Entropy Inequalities. *Commun. Math. Phys.* **1975**, *40*, 147–151. [[CrossRef](#)]
15. Uhlmann, A. Relative entropy and the Wigner–Yanase–Dyson–Lieb concavity in an interpolation theory. *Commun. Math. Phys.* **1977**, *54*, 21–32. [[CrossRef](#)]
16. Schoenberg, I.J. Metric Spaces and Positive Definite Functions. *Trans. Am. Math. Soc.* **1938**, *44*, 522–536. [[CrossRef](#)]
17. Sra, S. Positive definite matrices and the S-divergence. *Proc. Am. Math. Soc.* **2019**, *144*, 2787–2797. [[CrossRef](#)]
18. Virostek, D. The metric property of the quantum Jensen–Shannon divergence. *Adv. Math.* **2021**, *380*, 107595. [[CrossRef](#)]
19. Zurek, W.H. Decoherence and the transition from quantum to classical. *Phys. Today* **1991**, *44*, 36–44. [[CrossRef](#)]
20. Zurek, W.H. Decoherence, einselection, and the quantum origins of the classical. *Rev. Mod. Phys.* **2003**, *75*, 715–775. [[CrossRef](#)]
21. Pearl, J. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*; Morgan Kaufmann: San Mateo, CA, USA, 1988.
22. Petz, D. Sufficient subalgebras and the relative entropy of states of a von Neumann algebra. *Commun. Math. Phys.* **1986**, *105*, 123–131. [[CrossRef](#)]
23. Koch, C.; Hepp, K. Quantum mechanics in the brain. *Nature* **2006**, *440*, 611–611. [[CrossRef](#)]
24. Tononi, G.; Boly, M.; Massimini, M.; Koch, C. Integrated information theory: From consciousness to its physical substrate. *Nat. Rev. Neurosci.* **2016**, *17*, 450–461. [[CrossRef](#)] [[PubMed](#)]
25. Mintert, F.; Kuś, M.; Buchleitner, A. Concurrence of mixed multipartite quantum states. *Phys. Rev. Lett.* **2005**, *95*, 260502. [[CrossRef](#)] [[PubMed](#)]
26. Shor, O.; Benninger, F.; Khrennikov, A. Relational information framework, causality, unification of quantum interpretations and return to realism through non-ergodicity. *Sci. Rep.* **2025**, *15*, 8170. [[CrossRef](#)] [[PubMed](#)]
27. Amari, S.; Nagaoka, H. Translations of Mathematical Monographs. In *Methods of Information Geometry*; American Mathematical Society: Providence, RI, USA, 2000; Volume 191.

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