IOP Publishing New J. Phys. 27 (2025) 064504 https://doi.org/10.1088/1367-2630/adde7d

New Journal of Physics
The open access journal at the forefront of physics

Deutsche Physikalische Gesellschaft DPG
IOP Institute of Physics
Published in partnership with: Deutsche Physikalische Gesellschaft and the Institute of Physics

OPEN ACCESS

RECEIVED
14 December 2024

REVISED
21 May 2025

ACCEPTED FOR PUBLICATION
29 May 2025

PUBLISHED
9 June 2025

PAPER

Quantum mutual information in time

Zhen Wu[1], Arthur J Parzygnat[2], Vlatko Vedral[3] and James Fullwood[4,*]

[1] School of Mathematical Sciences, MOE-LSC, Shanghai Jiao Tong University, Shanghai 200240, People's Republic of China
[2] Experimental Study Group, Massachusetts Institute of Technology, Cambridge, MA 02139, United States of America
[3] Clarendon Laboratory, University of Oxford, Parks Road, Oxford OX1 3PU, United Kingdom
[4] School of Mathematics and Statistics, Hainan University, Haikou, Hainan Province 570228, People's Republic of China
* Author to whom any correspondence should be addressed.

E-mail: fullwood@hainanu.edu.cn

Keywords: quantum information, quantum correlations, entropy

Abstract

While the quantum mutual information is a fundamental measure of quantum information, it is only defined for spacelike separated quantum systems. Such a limitation is not present in the theory of classical information, where the mutual information between two random variables is well-defined irrespective of whether or not the variables are separated in space or separated in time. Motivated by this disparity between the classical and quantum mutual information, we employ the pseudo-density matrix formalism to define a simple extension of quantum mutual information into the time domain. As in the spatial case, we show that such a notion of quantum mutual information in time serves as a natural measure of correlation between timelike separated systems, while also highlighting ways in which quantum correlations distinguish between space and time. We also show how such quantum mutual information is time-symmetric with respect to quantum Bayesian inversion, and then we conclude by showing how quantum mutual information in time yields a Holevo bound for the amount of classical information that may be extracted from sequential measurements on an ensemble of quantum states.

1. Introduction

The mutual information of two quantum systems A and B whose joint state is represented by a bipartite density matrix ρ_{AB} is the real number $I(A:B)$ given by

$$I(A:B) = S(\rho_A) + S(\rho_B) - S(\rho_{AB}) \,, \tag{1}$$

where $S(\cdot)$ denotes the von Neumann entropy and ρ_A and ρ_B are the reduced density matrices of ρ_{AB} [1, 2]. As ρ_{AB} describes the joint state of two spacelike separated systems A and B, it follows that the quantum mutual information is essentially a *static* measure of information. This non-dynamical nature of quantum mutual information is in stark contrast with the notion of the classical mutual information $I(X:Y)$ of two random variables X and Y, which is defined irrespective of whether X and Y are separated in space or in time [3].

In this work, we make use of the pseudo-density matrix (PDM) formalism [4, 5] to define a temporal extension of quantum mutual information, and we investigate its properties. A PDM is a generalization of a density matrix that encodes correlations across both space and time, thus providing a single mathematical formalism for the study of both static and dynamical aspects of quantum information (see [5–16] for further development and applications of PDMs). In particular, PDMs provide a notion of a joint state R_{AB} associated with timelike separated quantum systems A and B whose reduced density matrices are the individual states ρ_A and ρ_B. However, while R_{AB} is Hermitian and of unit trace, it is not positive in general. Thus, the extension of the density matrix formalism to include non-positive PDMs is akin to the way in which the metric of space is extended into the time domain in special relativity, where such an extension results in a metric of Lorentzian (as opposed to Euclidean) signature [17, 18].

As for defining quantum mutual information in time, note that since the logarithm is not uniquely defined for non-positive matrices, simply replacing ρ_{AB} by R_{AB} in equation (1) will not in general yield a well-defined notion of quantum mutual information in time. To circumvent this issue, we use an extension of von Neumann entropy to Hermitian matrices given by [19–22]

$$S(X) = -\mathrm{Tr}\left[X \log |X|\right]. \tag{2}$$

While there are various justifications for our choice of S over other alternatives (such as those used e.g. in [14, 20]), a primary justification is that S satisfies [19]

- (Additivity) $S(X \otimes Y) = S(X) + S(Y)$ for all Hermitian matrices X and Y.
- (Orthogonal Convexity) If $\{X^i\}$ is a collection of mutually orthogonal Hermitian matrices of unit trace and p_i is a probability distribution, then

$$S\left(\sum_i p_i X^i\right) = H(p) + \sum_i p_i S\left(X^i\right),$$

  where $H(p)$ is the Shannon entropy of the probability distribution p_i.

While the properties of additivity and orthogonal convexity are certainly desirable from a purely mathematical perspective, we show in section 6 that additivity and orthogonal convexity of S are also crucial for establishing a temporal analog of the Holevo bound [23], thus providing an operational justification for the use of S.

Employing such a Hermitian extension S of von Neumann entropy and replacing ρ_{AB} by a PDM R_{AB} in equation (1) yields a well-defined notion of quantum mutual information for timelike separated quantum systems. In what follows, we prove general statements regarding such a notion of mutual information in time and present many examples. As in the spatial case, we find that such a mutual information in time provides a natural measure of correlation between timelike separated systems, while also highlighting the way in which quantum correlations distinguish between space and time. In particular, for sequential measurements performed on a system of qubits at two times, we show in section 3 that the mutual information vanishes for a system which is discarded and re-prepared between measurements, while it is maximized for a system undergoing unitary evolution between measurements. However, contrary to spatial mutual information, which for a pair of spacelike separated qubits can attain a maximum value of 2 in the presence of entanglement, we show in section 4 that the mutual information in time between two timelike separated qubits never exceeds a value of 1. We argue that this is a reflection of the fact that, unlike spatial correlations, maximal correlations in time are non-monogamous [9], as a qubit at a fixed point in time may be maximally correlated with a qubit both at a time in the future and a time in the past.

While for mutual information in space it is manifest from definition 1 that $I(A:B) = I(B:A)$, for mutual information in time the existence of such a symmetry is more subtle, as systems A and B may be related in time by a non-reversible process. However, we show in section 5 that when the quantum channel establishing temporal correlations between A and B is *Bayesian invertible* in the sense defined in [24], then it is fact that case that $I(A:B) = I(B:A)$, in accordance with the spatial case.

2. Pseudo-density matrices

Let σ_i denote the ith Pauli matrix for $i = 0, \ldots, 3$, and let $m > 0$ be a positive integer. For every $\alpha \in \{0, \ldots, 3\}^m$, we let $\alpha_j \in \{0, \ldots, 3\}$ denote the jth component of α for all $j \in \{1, \ldots, m\}$, and then define σ_α as

$$\sigma_\alpha = \sigma_{\alpha_1} \otimes \cdots \otimes \sigma_{\alpha_m}.$$

Suppose a sequential measurement of σ_α followed by σ_β is performed on a system of m-qubits at times t_A and t_B with $t_A < t_B$, and denote the Hilbert spaces of the system at the two times by $\mathcal{H}_A$ and $\mathcal{H}_B$, respectively. If the system evolves according to the quantum channel $\mathcal{N} : \mathcal{L}(\mathcal{H}_A) \to \mathcal{L}(\mathcal{H}_B)$ between measurements (we use $\mathcal{L}(\mathcal{H})$ to denote the algebra of linear operators on a Hilbert space $\mathcal{H}$), then the theoretical *two-time expectation value* of the sequential measurement of σ_α followed by σ_β is the real number $\langle \sigma_\alpha, \sigma_\beta \rangle$ given by

$$\langle \sigma_\alpha, \sigma_\beta \rangle = \mathrm{Tr}\left[\mathcal{N}\left(\Pi_\alpha^+ \rho \Pi_\alpha^+\right) \sigma_\beta\right] - \mathrm{Tr}\left[\mathcal{N}\left(\Pi_\alpha^- \rho \Pi_\alpha^-\right) \sigma_\beta\right],$$

where ρ is the initial state of the system, and $\Pi_\alpha^\pm = \frac{1}{2}(\mathbb{1} \pm \sigma_\alpha)$ is the orthogonal projection operator onto the ± 1-eigenspace of σ_α.

In such a two-time measurement scenario, the associated *PDM* is defined to be the bipartite hermitian operator R_{AB} on $\mathcal{H}_A \otimes \mathcal{H}_B$ given by

$$R_{AB} = \frac{1}{4^m} \sum_{\alpha,\beta} \langle \sigma_\alpha , \sigma_\beta \rangle \sigma_\alpha \otimes \sigma_\beta \,. \tag{3}$$

It follows from the definition of PDM together with properties of Pauli matrices that for all α and β,

$$\langle \sigma_\alpha , \sigma_\beta \rangle = \mathrm{Tr}\left[R_{AB}\left(\sigma_\alpha \otimes \sigma_\beta\right)\right] . \tag{4}$$

Thus, PDMs extend the operational meaning of bipartite density matrices to the temporal domain for Pauli observables.

Although the definition of the PDM R_{AB} given by (3) is conceptually simple, it is not always practical for calculations. In [5, 12, 25], it was shown that R_{AB} may equivalently be given by the formula

$$R_{AB} = \frac{1}{2} \left\{ \rho \otimes \mathbb{1} , \mathscr{J}[\mathcal{N}] \right\} , \tag{5}$$

where $\{\cdot,\cdot\}$ denotes the anticommutator and $\mathscr{J}[\mathcal{N}] = \sum_{i,j} |i\rangle\langle j| \otimes \mathcal{N}(|j\rangle\langle i|)$ is the *Jamiołkowski matrix* of the channel $\mathcal{N}$ [26]. When we wish to emphasize the dependence of the PDM R_{AB} on the initial state ρ and channel $\mathcal{N}$ as part of a two-time measurement scenario, we denote the RHS of equation (5) by $\mathcal{N} \star \rho$ and refer to it as the *spatiotemporal product* of the channel $\mathcal{N}$ and the initial state ρ.

The notion of PDM naturally extends to n-sequential measurements on a system of m-qubits for arbitrary $n > 0$, yielding an operator $R_{A_1 \cdots A_n}$ on the n-fold tensor product $\mathcal{H}_{A_1} \otimes \cdots \otimes \mathcal{H}_{A_n}$, where A_i denotes the system at time t_i, with $t_1 < \cdots < t_n$. We denote such an n-time PDM simply by $R_{1\cdots n}$. If the system evolves according to a channel $\mathcal{N}_i$ between measurements at times t_i and t_{i+1} for $i = 1, \ldots, n-1$, it was shown in [12, 27] that $R_{1\cdots n}$ may be given by the recursive formula

$$R_{1\cdots n} = \left(\mathcal{N}_{n-1} \circ \mathrm{Tr}_{1\cdots(n-2)}\right) \star R_{1\cdots(n-1)} , \tag{6}$$

where $\mathrm{Tr}_{1\cdots(n-2)}$ is the partial trace over the subsystems $A_1 \cdots A_{n-2}$ and $\star$ denotes the spatiotemporal product.

We note that while PDMs associated with multi-time measurement scenarios are not positive in general, they are always Hermitian and of unit trace. Moreover, the reduced marginals onto a single factor $\mathcal{L}(\mathcal{H}_{A_i})$ are always density matrices, representing the state of the system at time t_i. Interestingly, PDMs as in (6) that are positive admit dual interpretations as being extended across space on the one hand, and time on the other. We refer to such PDMs as *dual states*.

3. Mutual information in time

Let $R_{1\cdots n}$ be an n-time PDM as in (6), let A_i denote the associated system at time t_i for $i = 1, \ldots, n$, let $k \in \{2, \ldots, n-1\}$, and let $R_{1\cdots k}$ and $R_{k+1\cdots n}$ be the PDMs corresponding to the first k measurements and final $n-k$ measurements, respectively, which may be obtained from the PDM $R_{1\cdots n}$ by tracing out the associated complementary subsystems. We then define the ***mutual information in time*** between the joint temporal systems $A = A_1 \cdots A_k$ and $B = A_{k+1} \cdots A_n$ to be the element $I(A:B) \in \mathbb{R}$ given by

$$I(A:B) = S(R_{1\cdots k}) + S(R_{k+1\cdots n}) - S(R_{1\cdots n}) ,$$

where $S(\cdot)$ is the Hermitian extension of the von Neumann entropy given by $S(X) = -\mathrm{Tr}\left[X \log |X|\right]$.

We now present some examples of mutual information in time.

Example 1 (a dual state). Let ρ_{AB} be the bipartite density matrix given by

$$\rho_{AB} = \begin{pmatrix} 13/24 & 0 & 0 & 0 \\ 0 & 5/24 & -1/6 & 0 \\ 0 & -1/6 & 5/24 & 0 \\ 0 & 0 & 0 & 1/24 \end{pmatrix} .$$

The density matrix ρ_{AB} is an entangled state [28] that has been shown to also be a 2-time PDM [13], and hence is a dual state. As such, the associated systems A and B may either be viewed as being spacelike separated or timelike separated. Thus, their mutual information in space is equal to their mutual information in time, which is approximately 0.2315.

Example 2 (a single qubit at two times). Let R_{AB} be the 2-time PDM associated with a maximally mixed qubit that undergoes trivial dynamics between measurements, which is given by

$$R_{AB} = \frac{1}{2}\texttt{SWAP} = \frac{1}{2}\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

As the eigenvalues of R_{AB} are $(1/2, 1/2, 1/2, -1/2)$, it follows that $I(A:B) = 1$, which coincides with the von Neumann entropy of the initial maximally mixed state. We note that although the temporal correlations between the qubit at two times in this example can violate CHSH inequalities [29], thus exhibiting maximal correlations in time, the mutual information between the qubit at two times is not 2, as one might expect from the case of a pair of maximally entangled qubits in space [1]. We will further address this disparity between the temporal and spatial cases in section 4.

The previous example is a special case of the following:

Theorem 1. *Suppose $R_{AB} = \mathcal{U} \star \rho$ is a 2-time PDM associated with a multi-qubit system initially in state ρ that evolves according to a unitary channel $\mathcal{U}$ between measurements. Then $I(A:B) = S(\rho)$.*

Proof. Let m be the number of qubits of the system represented by R_{AB}, let $n = 2^m$, let $\mathfrak{mspec}(X)$ denote the multiset of eigenvalues of a matrix X, and suppose $\mathfrak{mspec}(\rho) = \{\lambda_1, \ldots, \lambda_n\}$. It then follows from lemma 5.7 in [19] that

$$\mathfrak{mspec}(\mathcal{U} \star \rho) = \mathfrak{mspec}(\rho) \cup \left\{ \pm\frac{\lambda_i + \lambda_j}{2} \,\middle|\, 0 < i < j \leqslant n \right\}.$$

Since $f(x) = -x\log|x|$ is an odd function, it follows that

$$S(R_{AB}) = S(\mathcal{U} \star \rho) = S(\rho) = S(\mathcal{U}(\rho)), \tag{7}$$

where the final equality follows from the unitary invariance of the entropy function S. We then have

$$I(A:B) = S(\rho) + S(\mathcal{U}(\rho)) - S(\mathcal{U} \star \rho) = S(\rho),$$

as desired. □

The next two examples illustrate cases when the evolution between measurements is non-unitary.

Example 3 (discard and prepare). Let $\mathcal{N}$ be the discard-and-prepare channel given by $\mathcal{N}(\rho) = \mathrm{Tr}[\rho]\sigma$ for some state σ. It then follows that if R_{AB} is a 2-time PDM of the form $R_{AB} = \mathcal{N} \star \rho$, then $R_{AB} = \rho \otimes \sigma$ (and hence is a dual state). In such a case, the timelike separated systems A and B represented by the PDM R_{AB} are such that $I(A:B) = 0$.

Example 4 (decoherence). Let R_{AB} be the 2-time PDM associated with a single qubit in an initial state $\rho = |-\rangle\langle-|$, which between measurements is to evolve according to the decoherence map $\mathcal{D}$ given by

$$\mathcal{D}\begin{pmatrix} \rho_{00} & \rho_{01} \\ \rho_{10} & \rho_{11} \end{pmatrix} = \begin{pmatrix} \rho_{00} & 0 \\ 0 & \rho_{11} \end{pmatrix}.$$

The associated PDM $R_{AB} = \mathcal{D} \star \rho$ has eigenvalues

$$\left(\frac{1+\sqrt{2}}{4}, \frac{1+\sqrt{2}}{4}, \frac{1-\sqrt{2}}{4}, \frac{1-\sqrt{2}}{4}\right),$$

from which it follows that $I(A:B) \approx 0.79824$. In [19] a quantity referred to as 'information discrepancy' was introduced, which turns out to coincide with $-I(A:B)$ in this example. As negative information discrepancy signifies information gain, the fact that $-I(A:B) < 0$ is consistent with the recent results in [30], where it is argued that contrary to conventional wisdom, decoherence is a process of information flowing *into* the system from its environment (as opposed to the other way around). Based on this, it seems plausible that $I(A:B)$ might quantify the information gained by a system due to decoherence.

In the next example, we consider a single qubit at multiple times, with a varying initial state.

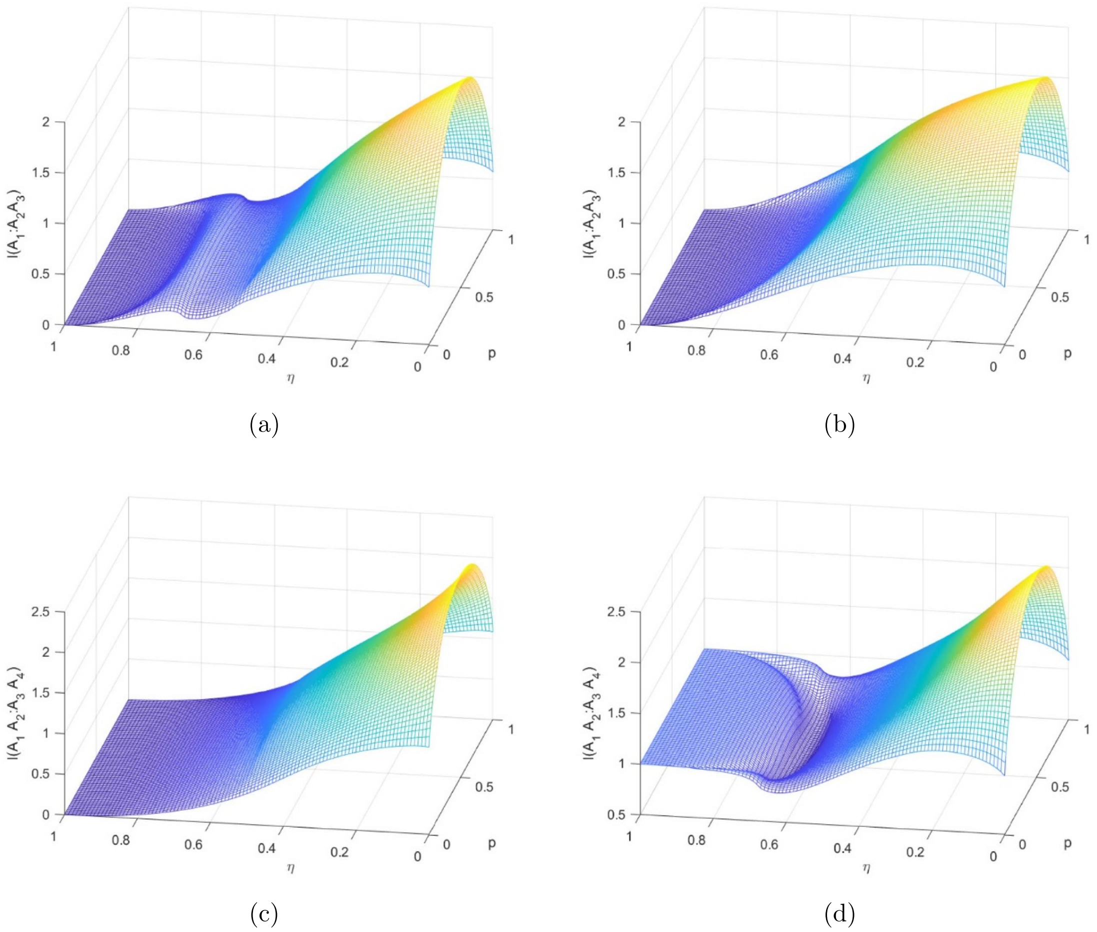

Figure 1. Mutual information between a single qubit at multiple times. In each of the above graphs, η is the depolarizing parameter of the channel $\mathcal{E}_\eta$ and p is the parameter defining the initial state as in example 5. (a) $I(A_1 : A_2A_3)$ for $n=3$ and $\mathcal{N}_1 = \mathcal{N}_2 = \mathcal{E}_\eta$. (b) $I(A_1 : A_2A_3)$ for $n=3, \mathcal{N}_1 = \mathcal{E}_\eta$ and $\mathcal{N}_2 = \mathrm{id}$. (c) $I(A_1A_2 : A_3A_4)$ for $n=4, \mathcal{N}_1 = \mathcal{N}_3 = \mathrm{id}$ and $\mathcal{N}_2 = \mathcal{E}_\eta$ (d) $I(A_1A_2 : A_3A_4)$ for $n=4, \mathcal{N}_1 = \mathcal{N}_3 = \mathcal{E}_\eta$ and $\mathcal{N}_2 = \mathrm{id}$. In all cases except (d), we find that the mutual information is maximized when $\eta=0$ and $p=0.5$, reflecting the fact that in such a case the output of the channel $\mathcal{E}_\eta$ is maximally correlated with the input of the channel (since in such a case the channel is the identity), and that the entropy of the input state is maximized. On the other hand, when $\eta=1$ the channel $\mathcal{E}_\eta$ is completely depolarizing, and all information about the initial state is lost for all p, thus resulting in the mutual information being minimized. In cases (a) and (d), there is some oscillatory behavior as η increases beyond 0.5, reflecting a tradeoff between the noise of the channel and the uncertainty in the initial state.

Example 5 (a single qubit at multiple times). Let $R_{1\cdots n}$ be a multiple time single qubit PDM with initial state $\rho = \mathrm{diag}(p, 1-p)$ with $p \in [0,1]$. Denote the channel responsible for the evolution of the system between times t_i and t_{i+1} by $\mathcal{N}_i$, and let $\mathcal{E}_\eta$ be the depolarizing channel given by

$$\mathcal{E}_\eta(\rho) = (1-\eta)\,\rho + \eta \mathrm{Tr}[\rho]\,\frac{\mathbb{1}}{2},$$

where $\eta \in [0,1]$ is the depolarization parameter. We then consider several cases of mutual information in time where $\mathcal{N}_i$ is either the identity channel or the depolarizing channel $\mathcal{E}_\eta$ for $n=3$ and $n=4$, the results of which are plotted as functions of p and η in figure 1.

Remark 1 (lack of monotonicity). While the spatial quantum mutual information satisfies monotonicity, namely, $I(A:\Gamma(B)) \leqslant I(A:B)$ for every local operation Γ on the system B, local operations can either increase or decrease mutual information in time. In particular, while tracing out a subsystem A_i of a multipartite spatial system $A_1\cdots A_n$ essentially discards the information contained in A_i, tracing out a subsystem A_i of a multipartite *dynamical* system $A_1\cdots A_n$ results in a redistribution of information across time to form a distinct dynamical system $A_1\cdots A_{i-1}A_{i+1}\cdots A_n$, where A_{i-1} then has direct causal influence on A_{i+1}. As such, the local operation of tracing out a subsystem of a dynamical system $A_1\cdots A_n$ may result in an *increase* in the mutual information between different timelike separated subsystems, thus resulting in a failure of monotonicity.

4. A single qubit at two times: the general case

For a single qubit at two-times we obtain the following theorem, whose proof can be found in appendix A.

Theorem 2. *Let R_{AB} be a 2-time PDM associated with a single qubit whose initial state is maximally mixed and which evolves according to a channel that is either unital or has Choi rank no greater than 2 between measurements. Then* $0 \leqslant I(A:B) \leqslant 1$.

While we suspect that $0 \leqslant I(A:B) \leqslant 1$ holds for an arbitrary qubit at two times, our proof of Theorem 2 relies on the hypotheses that the initial state is maximally mixed and that the channel corresponding to the evolution between measurements is either unital or has Choi rank no greater than 2. We note that it already follows from Theorem 1 that the bound $0 \leqslant I(A:B) \leqslant 1$ also holds for arbitrary initial states provided that the evolution between measurements is unitary.

Assuming $0 \leqslant I(A:B) \leqslant 1$ holds for an arbitrary qubit at two times, such a feature of temporal correlations is in stark contrast with the case of two qubits separated in space, which can achieve a maximum mutual information of 2 in the presence of entanglement [1]. This is a reflection of the fact that, unlike the monogamy spatial correlations, a single qubit can be maximally correlated both with a qubit in its immediate future and its immediate past. In particular, given a 3-time measurement scenario, the qubit at time t_2 can violate CHSH inequalities both with the qubits at times t_1 and t_3, which is a consequence of the fact that two-time correlation functions associated with Pauli observables are independent of initial conditions [29, 31].

An interesting consequence of the bound $I(A:B) \leqslant 1$ for a qubit at two times, is that if a pair of spatially entangled qubits has a mutual information greater than 1, then it follows that the two qubits do not admit a dual description of being timelike separated. The converse to this statement however does not hold in general. In particular, a pair of spatially entangled qubits not admitting a dual description of being timelike separated can have mutual information less than one, as in the case of the entangled qubits described by the pure state $\sqrt{p}|00\rangle + \sqrt{1-p}|11\rangle$ for small p.

We also note that it follows from example 3 that the lower bound of 0 for the mutual information $I(A:B)$ in the context of theorem 2 is achieved when a qubit evolves according to a completely depolarizing channel between measurements. This is consistent with the fact that all information about the initial measurement is lost in such a scenario. On the other hand, we know from theorem 1 that the upper bound of 1 for $I(A:B)$ is achieved when a maximally mixed qubit undergoes unitary evolution between the two measurements, which is consistent with the fact that no information is lost (or gained) under unitary evolution.

5. Time-reversal symmetry

For spatial mutual information, it is always the case that $I(A:B) = I(B:A)$, as there is no preferred directionality in space between spacelike separated systems. In particular, applying the swap transformation $\mathscr{S} : \mathcal{L}(\mathcal{H}_A) \otimes \mathcal{L}(\mathcal{H}_B) \to \mathcal{L}(\mathcal{H}_B) \otimes \mathcal{L}(\mathcal{H}_A)$ to a bipartite density matrix ρ_{AB} yields a density matrix ρ_{BA} which is an equivalent mathematical description of the state of the joint system consisting of A and B. Moreover, as the swap transformation is unitary and ρ_{AB} and ρ_{BA} and have the same marginals, it follows that $I(A:B) = I(B:A)$ by (1).

When A and B are timelike separated, the situation is more subtle, as a quantum channel $\mathcal{N}$ establishing temporal correlations between A and B may be non-reversible due to noise. In such a case, the non-reversibility of $\mathcal{N}$ is reflected in the associated PDM $R_{AB} = \mathcal{N} \star \rho$ by the fact that the swap $\mathscr{S}(R_{AB})$ is not a PDM. However, when $\mathscr{S}(R_{AB})$ *is* in fact a PDM, so that $\mathscr{S}(R_{AB}) = R_{BA} = \mathcal{M} \star \sigma$, where $\sigma = \mathcal{N}(\rho)$ and $\mathcal{M}$ is a quantum channel from B to A, then $\mathcal{M}$ is said to be a *Bayesian inverse* of the channel $\mathcal{N}$ *with respect to* the state ρ [24]. As such, the notion of a Bayesian inverse provides a novel form of reversibility for timelike separated quantum systems, even when the systems are temporally correlated via a noisy channel.

As solving for a Bayesian inverse $\mathcal{M}$ involves solving a linear system, the number of solutions is either 0,1 or ∞, with 0 corresponding to no Bayesian inverse existing, 1 corresponding to a unique Bayesian inverse and ∞ corresponding to the notion of Bayesian inverse being non-unique. Fortunately, in the case of non-uniqueness, given two Bayesian inverses $\mathcal{M}$ and $\mathcal{M}'$ of $\mathcal{N}$ with respect to ρ, it follows from the definition of Bayesian inverse that $\mathcal{M} \star \sigma = \mathcal{M}' \star \sigma$. Hence, the notion of a PDM R_{BA} corresponding to a time-reversal of the process described by R_{AB} is well-defined.

In the context of mutual information in time, suppose we have a PDM $R_{AB} = \mathcal{N} \star \rho$ with $\mathcal{N}$ Bayesian invertible with respect to ρ, so that the swap $\mathscr{S}(R_{AB}) = R_{BA} = \mathcal{M} \star \sigma$, where $\sigma = \mathcal{N}(\rho)$ and $\mathcal{M}$ is a quantum channel from B to A corresponding to a Bayesian inverse of $\mathcal{N}$ with respect to ρ. Now since the eigenvalues of a matrix are invariant under the swap map $\mathscr{S}$, it follows that $S(R_{AB}) = S(R_{BA})$. Moreover, since $\sigma = \mathcal{N}(\rho)$

and $\rho = \mathcal{M}(\sigma)$, we have

$$\begin{aligned} I(B:A) &= S(\sigma) + S(\mathcal{M}(\sigma)) - S(R_{BA}) \\ &= S(\rho) + S(\mathcal{N}(\rho)) - S(R_{AB}) \\ &= I(A:B)\,. \end{aligned}$$

Thus, the existence of a Bayesian inverse of a channel establishes a symmetry in time that extends the symmetry $I(A:B) = I(B:A)$ of spatial mutual information into the time domain, as well as the symmetry $I(A:B) = I(B:A)$ of *classical* temporal mutual information, the latter of which follows from Bayes' rule [3].

6. A Holevo bound in time

Suppose Alice sends a state ρ^i to Bob with probability p_i, after which Bob performs a measurement on ρ^i corresponding to a POVM $\{M_j\}$. Then the probability $q_{j|i}$ of the measurement outcome M_j given the state ρ^i is given by $q_{j|i} = \mathrm{Tr}[\rho^i M_j]$. If X is a random variable associated with the probability distribution p_i, and Y is a random variable associated with the probability distribution $q_j = \sum_i p_i q_{j|i}$, then the classical mutual information $I(X:Y)$ is bounded above by the *Holevo quantity* χ given by

$$\chi = S\left(\sum_i p_i \rho^i\right) - \sum_i p_i S\left(\rho^i\right)\,. \tag{8}$$

The inequality $I(X:Y) \leqslant \chi$ is referred to as the *Holevo bound* [23], and it provides a fundamental limit on the amount of classical information that may be extracted from an ensemble of quantum states. Moreover, if we form the joint classical-quantum state ρ_{AB} associated with Alice's classical register and Bob's random acquisition of the state ρ^i, namely,

$$\rho_{AB} = \sum_i p_i \,|i\rangle\langle i| \otimes \rho^i\,, \tag{9}$$

then it is straightforward to show that the mutual information $I(A:B)$ associated with ρ_{AB} is such that $I(A:B) = \chi$. This, together with the Holevo bound, provides an operational interpretation for the quantum mutual information $I(A:B)$. In this section, we obtain analogous results for mutual information in time.

For this, we first note that from a dynamical perspective, the state ρ_{AB} given by (9) is in fact the PDM given by $\rho_{AB} = \mathcal{P} \star \rho_A$, where $\mathcal{P}$ is the channel given by $\mathcal{P}(|i\rangle\langle j|) = \delta_{ij}\rho^i$ and ρ_A is the state given by $\rho_A = \sum_i p_i |i\rangle\langle i|$. In such a 2-time scenario, the channel $\mathcal{P}$ is viewed as a channel from Alice to Bob. If Bob then makes two sequential measurements on ρ^i, with the system evolving according to a channel $\mathcal{N}$ between measurements, then it follows from (6) that the associated 3-time PDM (with $\mathcal{N}_1 = \mathcal{P}$ and $\mathcal{N}_2 = \mathcal{N}$) is then given by

$$R_{AB_1B_2} = \sum_i p_i \,|i\rangle\langle i| \otimes R^i_{12}\,, \tag{10}$$

where $R^i_{12} = \mathcal{N} \star \rho^i$ is the 2-time PDM associated with Bob's sequential measurements. We then obtain the following:

Theorem 3. *Let R^i_{12} be the 2-time PDM associated with Bob's sequential measurements of ρ^i, and let $I(A:B_1B_2)$ be the mutual information associated with the classical-quantum PDM $R_{AB_1B_2}$ given by (10). Then the following statements hold.*

(i) $I(A:B_1B_2) = S\left(\sum_i p_i R^i_{12}\right) - \sum_i p_i S(R^i_{12})$.
(ii) *If Bob's system evolves unitarily between measurements, then $I(A:B_1B_2) = \chi$, where χ is the Holevo quantity (8).*

Proof. First, we have $\mathrm{Tr}_{B_1B_2}\left[R_{AB_1B_2}\right] = \sum_i p_i |i\rangle\langle i|$ and $\mathrm{Tr}_A\left[R_{AB_1B_2}\right] = \sum_i p_i R^i_{12}$, so that

$$I(A:B_1B_2) = H(p) + S\left(\sum_i p_i R^i_{12}\right) - S\left(R_{AB_1B_2}\right)\,, \tag{11}$$

where $H(p)$ denotes the Shannon entropy associated with the probability distribution p_i. Now since S is additive and orthogonally convex (as defined in the introduction), we have

$$\begin{aligned} S(R_{AB_1B_2}) &= S\left(\sum_i p_i |i\rangle\langle i| \otimes R^i_{12}\right) \\ &= H(p) + \sum_i p_i S\left(|i\rangle\langle i| \otimes R^i_{12}\right) \\ &= H(p) + \sum_i p_i \left(S(|i\rangle\langle i|) + S\left(R^i_{12}\right)\right) \\ &= H(p) + \sum_i p_i S\left(R^i_{12}\right) . \end{aligned}$$

Substituting $S(R_{AB_1B_2}) = H(p) + \sum_i p_i S(R^i_{12})$ into equation (11) then yields

$$I(A : B_1B_2) = S\left(\sum_i p_i R^i_{12}\right) - \sum_i p_i S\left(R^i_{12}\right) , \tag{12}$$

which proves item (i). Now if Bob's system evolves according to a unitary channel $\mathcal{U}$ between measurements, so that $R^i_{12} = \mathcal{U} \star \rho^i$, it follows from (7) that $S(R^i_{12}) = S(\rho^i)$. Moreover, as the spatiotemporal product $\star$ is linear in ρ, we have

$$\begin{aligned} S\left(\sum_i p_i R^i_{12}\right) &= S\left(\sum_i p_i \left(\mathcal{U} \star \rho^i\right)\right) = S\left(\mathcal{U} \star \sum_i p_i \rho^i\right) \\ &= S\left(\sum_i p_i \rho^i\right) , \end{aligned}$$

where the final equation follows again from (7). It then follows from equation (12) that $I(A : B_1B_2) = \chi$, which proves item (ii). □

Now let X be the random variable associated with Alice's classical register, and let (Y_1, Y_2) be the bivariate random variable corresponding to the joint statistics of Bob's classical register when performing sequential measurements of the POVMs $\{M_j\}$ and $\{N_k\}$. We assume that after a measurement outcome of M_j when measuring ρ^i that Bob's system updates to $U\sqrt{M_j}\rho^i\sqrt{M_j}U^\dagger$ (suitably normalized) for some fixed unitary U that is independent of i and j. It then follows that the joint distribution p_{jk} corresponding to the bivariate random variable (Y_1, Y_2) is given by $p_{jk} = \sum_i p_{jk|i}$, where the conditional distribution $p_{jk|i}$ is given by

$$p_{jk|i} = \mathrm{Tr}\left[U\sqrt{M_j}\rho^i\sqrt{M_j}U^\dagger N_k\right] .$$

Now, since the mapping $\mathcal{M} : \mathcal{L}(\mathcal{H}_{B_1}) \to \mathcal{L}(\mathcal{H}_{B_1}) \otimes \mathcal{L}(\mathcal{H}_{B_2})$ given by

$$\mathcal{M}(\rho) = \sum_{j,k} \mathrm{Tr}\left[U\sqrt{M_j}\rho\sqrt{M_j}U^\dagger N_k\right] |j\rangle\langle j| \otimes |k\rangle\langle k|$$

is a quantum channel, it follows from lemma 1 of appendix B that the mutual information $I(X : Y_1Y_2)$ between Alice's classical register X and the bivariate register (Y_1, Y_2) corresponding to Bob's sequential measurements is also bounded above by the Holevo quantity χ. Thus, it follows from theorem 3 that

$$I(X : Y_1Y_2) \leqslant I(A : B_1B_2) ,$$

where again $I(A : B_1B_2)$ is the mutual information in time associated with the classical-quantum PDM $R_{AB_1B_2}$ given by (10). By taking a supremum over all possible POVMs $\{M_j\}$ and $\{N_k\}$, it then follows that the amount of classical information that may be extracted from such sequential measurements is bounded above by the mutual information in time $I(A : B_1B_2)$.

7. Concluding remarks

In this work, we have employed the PDM formalism to define an extension of quantum mutual information into the time domain. Such a measure of quantum information naturally quantifies correlations between timelike separated quantum systems, and recovers the usual spatial mutual information for systems that admit a dual description of either being timelike or spacelike separated. We have proved such a notion of mutual information in time satisfies various properties one would expect from such a dynamical measure of information, highlighting the ways in which quantum correlations distinguish between space and time.

We note that while an alternative formulation of quantum mutual information for timelike separated systems has recently appeared in [32], such a quantity is defined in terms of optimizing a relative entropy over all possible system-ancilla coupling schemes, resulting in a quantity which can diverge, even for systems consisting of a single qubit. From the dynamical perspective espoused in this work, the limitations of the aforementioned approach stem from the fact that a spatial notion of relative entropy is being utilized, rather than a dynamical extension of relative entropy using the entropy function $S(X) = -\mathrm{Tr}[X\log|X|]$ [19].

As for the quantum mutual information in time defined in this work, we have shown that a single qubit at two times can never exceed a mutual information of 1. This is consistent with the fact that for two qubits to be correlated in time there must necessarily be a causal transfer of information between the two qubits via a quantum channel, and hence two timelike separated qubits can share at most one bit of information. An interesting consequence of such an upper bound for the mutual information between a qubit at two times is that if two spatially separated qubits have mutual information exceeding 1, then it necessarily follows that the correlations between such qubits may not be realized in a dynamical setting where the two qubits are separated in time.

Another interesting aspect of quantum mutual information in time is that in all known 2-time examples it is non-negative, despite the fact that the entropy function $S(X) = -\mathrm{Tr}[X\log|X|]$ is not subadditive on general hermitian matrices of unit trace. In particular, in [19], an example of a unit trace bipartite Hermitian matrix $\rho_{AB} \in \mathcal{L}(\mathcal{H}_A \otimes \mathcal{H}_B)$ was constructed whose marginals are density matrices, and yet

$$S(\rho_A) + S(\rho_B) - S(\rho_{AB}) < 0. \tag{13}$$

In such a case, ρ_{AB} is of the form $\rho_{AB} = \mathcal{N} \star \rho_A$ with $\mathcal{N}$ *not* a CPTP map, which the entropy function S seems to be detecting via the failure of subadditivity (13).

Finally, it is worth noting that quantum mutual information in time may be used to define a notion of channel capacity for a quantum channel $\mathcal{N}$ which is directly analogous to Shannon's original definition in the classical case, i.e. by taking a supremum of mutual information with respect to all possible input states for the channel $\mathcal{N}$. It would be interesting then to compare such a notion of channel capacity with other notions of quantum capacity [33–35], especially since such a quantity is often computable.

Data availability statement

No new data were created or analysed in this study.

Acknowledgment

JF is supported by a Hainan University startup fund for the project 'Spacetime from Quantum Information'. VV acknowledges support from the Templeton and the Gordon and Betty Moore foundations, and WZ would like to thank Hainan University for hospitality during the course of this project.

Appendix A. Proof of theorem 2

Let A and B denote quantum systems corresponding to a single qubit at times t_A and t_B with $t_A < t_B$, and suppose the system which is initially in the state ρ evolves according to a quantum channel $\mathcal{N} : \mathcal{L}(\mathcal{H}_A) \to \mathcal{L}(\mathcal{H}_B)$ between measurements at times t_A and t_B. In such a case, the mutual information $I(A : B)$ between the qubit at times t_A and t_B is given by

$$I(A:B) = S(\rho) + S(\mathcal{N}(\rho)) - S(R_{AB}), \tag{A.1}$$

where

$$R_{AB} = \mathcal{N} \star \rho \equiv \frac{1}{2}\{\rho \otimes \mathbb{1}, \mathscr{J}[\mathcal{N}]\} \quad , \quad \mathscr{J}[\mathcal{N}] = \sum_{i,j} |i\rangle\langle j| \otimes \mathcal{N}(|j\rangle\langle i|), \tag{A.2}$$

and

$$S(X) = -\mathrm{Tr}\left[X \log |X|\right]$$

for all Hermitian matrices X . In this section, we will prove theorem 2 in main text, which states that if $\rho = \mathbb{1}/2$ is maximally mixed, and if $\mathcal{N}$ is either unital or of Choi rank no more than 2, then $0 \leqslant I(A:B) \leqslant 1$. We note that under the assumption that $\rho = \mathbb{1}/2$ is maximally mixed, it follows from (A.2) that

$$R_{AB} = \frac{1}{2}\mathscr{J}\left[\mathcal{N}\right] \tag{A.3}$$

for every channel $\mathcal{N}: \mathcal{L}(\mathcal{H}_A) \to \mathcal{L}(\mathcal{H}_B)$.

A.1. Reduction to $\mathcal{N}_d$

Let $\mathcal{N}: \mathcal{L}(\mathcal{H}_A) \to \mathcal{L}(\mathcal{H}_B)$ be a quantum channel. As an arbitrary density matrix $\rho \in \mathcal{L}(\mathcal{H}_A)$ may be written with respect to the Pauli basis $\{\sigma_i\}_{i=0}^3$ as

$$\rho = \frac{1}{2}\left(\mathbb{1} + \vec{r}\cdot\vec{\sigma}\right)$$

with $\vec{r} \in \mathbb{R}^3$ and $|\vec{r}| \leqslant 1$, it can be shown that the quantum channel $\mathcal{N}$ takes the form

$$\mathcal{N}(\rho) = \frac{1}{2}\left[\mathbb{1} + \left(\mathbf{t} + \mathbf{T}(\vec{r})\right)\cdot\vec{\sigma}\right]$$

where $\mathbf{t} \in \mathbb{R}^3$ and $\mathbf{T} \in \mathbb{R}^{3\times 3}$ are such that $\left|\mathbf{t} + \mathbf{T}(\vec{r})\right| \leqslant 1$ for all $\vec{r}$ with $|\vec{r}| \leqslant 1$ [36] (more precise conditions for positivity and complete positivity of $\mathcal{N}$ in terms of $\mathbf{t}$ and $\mathbf{T}$ can be found in [36–38]). More generally, if $A = w_0\mathbb{1} + \vec{w}\cdot\vec{\sigma}$, with $w_0 \in \mathbb{C}$ and $\vec{w} \in \mathbb{C}^3$, then

$$\mathcal{N}(A) = w_0\mathbb{1} + \left(w_0\mathbf{t} + \mathbf{T}(\vec{w})\right)\cdot\vec{\sigma}. \tag{A.4}$$

This corrects a typo in equation (2) of [36]. Thus, a matrix representation N of the channel $\mathcal{N}$ with respect to the ordered Pauli basis $(\sigma_0, \sigma_1, \sigma_2, \sigma_3)$ is of the form

$$N = \begin{pmatrix} 1 & \mathbf{0} \\ \mathbf{t} & \mathbf{T} \end{pmatrix}, \tag{A.5}$$

i.e. if $A = \sum_{i=0}^3 \alpha_i\sigma_i$ is an arbitrary 2×2 matrix with $\alpha_i \in \mathbb{C}$, then $\mathcal{N}_d(A) = \sum_{i=0}^3 \beta_i\sigma_i$, where $\vec{\beta} = N\vec{\alpha}$, $\alpha = (\alpha_0, \alpha_1, \alpha_2, \alpha_3)$, and $\beta = (\beta_0, \beta_1, \beta_2, \beta_3)$. Moreover, there exist unitary channels $\mathcal{U}$ and $\mathcal{V}$ such that

$$\mathcal{N} = \mathcal{V}\circ\mathcal{N}_d\circ\mathcal{U}, \tag{A.6}$$

where the matrix representation of $\mathcal{N}_d$ is given by

$$N_d = \begin{pmatrix} 1 & 0 & 0 & 0 \\ t_1 & \lambda_1 & 0 & 0 \\ t_2 & 0 & \lambda_2 & 0 \\ t_3 & 0 & 0 & \lambda_3 \end{pmatrix}. \tag{A.7}$$

Now let $\rho = \mathbb{1}/2$ be the maximally mixed state, and let R_{AB} and $R_{AB}^{(d)}$ be the PDMs given by $R_{AB} = \mathcal{N}\star\rho$ and $R_{AB}^{(d)} = \mathcal{N}_d\star\rho$. Since $\mathcal{N} = \mathcal{V}\circ\mathcal{N}_d\circ\mathcal{U}$, the covariance property of PDMs [39, 40] yields

$$\left(\mathcal{U}^*\otimes\mathcal{V}\right)\left(\mathcal{N}_d\star\rho\right) = \mathcal{N}\star\mathcal{U}^*(\rho), \tag{A.8}$$

where $\mathcal{U}^* = \mathcal{U}^{-1}$ is the Hilbert–Schmidt adjoint of $\mathcal{U}$. If $U \in \mathcal{L}(\mathcal{H}_A)$ is the unitary operator such that $\mathcal{U}(a) = UaU^\dagger$ for all $a \in \mathcal{L}(\mathcal{H}_A)$, then

$$\mathcal{U}^*(\rho) = U^\dagger\rho U = U^\dagger(\mathbb{1}/2)U = (\mathbb{1}/2)U^\dagger U = \rho.$$

Thus, by (A.8), we have

$$\left(\mathcal{U}^*\otimes\mathcal{V}\right)\left(R_{AB}^{(d)}\right) = \left(\mathcal{U}^*\otimes\mathcal{V}\right)\left(\mathcal{N}_d\star\rho\right) = \mathcal{N}\star\rho = R_{AB}.$$

Moreover, since $\mathcal{U}(\rho)=\rho$, we have $\mathcal{N}(\rho)=\mathcal{V}(\mathcal{N}_d(\mathcal{U}(\rho)))=\mathcal{V}(\mathcal{N}_d(\rho))$. It then follows by the unitary-invariance of S that $S(R_{AB})=S(R_{AB}^{(d)})$ and $S(\mathcal{N}(\rho))=\mathcal{S}(\mathcal{N}_d(\rho))$. Now let $I(A:B)$ and $I(A:B)^{(d)}$ denote the mutual information associated with R_{AB} and $R_{AB}^{(d)}$, respectively. We then have

$$I(A:B)=S(\rho)+S(\mathcal{N}(\rho))-S(R_{AB})=S(\rho)+S(\mathcal{N}_d(\rho))-S\left(R_{AB}^{(d)}\right)=I(A:B)^{(d)}\ .$$

Thus, $I(A:B)=I(A:B)^{(d)}$ for any qubit channel $\mathcal{N}$ with initial state $\rho=\mathbb{1}/2$. As such, when proving Theorem 2 in main text, we may compute $I(A:B)$ in terms of $I(A:B)^{(d)}$, which we will refer to as 'reduction to $\mathcal{N}_d$'.

A.2. Unital channels

In this section, we consider the case when $\mathcal{N}$ is a unital channel. Before proceeding with the proof, we recall that a real vector $\vec{y}=(y_1,\cdots,y_n)$ is said to *majorize* a real vector $\vec{x}=(x_1,\cdots,x_n)$, written $\vec{x}\prec\vec{y}$, if and only if $\sum_{i=1}^k x_{[i]}\leqslant\sum_{i=1}^k y_{[i]}$ for $k=1,\cdots,n-1$, where $x_{[1]}\geqslant\cdots\geqslant x_{[n]}$ denotes the components of $\vec{x}$ in decreasing order and $y_{[1]}\geqslant\cdots\geqslant y_{[n]}$ denotes the components of $\vec{y}$ in decreasing order. A function f is then said to be *Schur-concave* if and only if $\vec{x}\prec\vec{y}\implies f(\vec{x})\geqslant f(\vec{y})$ [41].

So now suppose $\mathcal{N}$ is a unital channel, in which case the vector $\mathbf{t}$ in (A.5) is zero. This then implies that $\mathcal{N}_d$ is a Pauli channel, i.e. there exists a probability vector (p_0,p_1,p_2,p_3) such that $\mathcal{N}_d(\rho)=\sum_{i=0}^3 p_i\sigma_i\rho\sigma_i$ for all $\rho\in\mathcal{L}(\mathcal{H}_A)$. Since $\mathcal{N}_d$ is unital, it follows that $\mathcal{N}_d(\rho)=\rho$ for $\rho=\mathbb{1}/2$ maximally mixed, so that the associated mutual information $I(A:B)^{(d)}$ is then given by

$$I(A:B)^{(d)}=2-S\left(R_{AB}^{(d)}\right), \tag{A.9}$$

where $R_{AB}^{(d)}=\mathcal{N}_d\star\rho$. Now since

$$R_{AB}^{(d)}\overset{(A.3)}{=}\frac{1}{2}\mathscr{J}\left[\mathcal{N}_d\right]=\frac{1}{2}\begin{pmatrix} p_0+p_3 & 0 & 0 & p_1-p_2\\ 0 & p_1+p_2 & p_0-p_3 & 0\\ 0 & p_0-p_3 & p_1+p_2 & 0\\ p_1-p_2 & 0 & 0 & p_0+p_3\end{pmatrix}, \tag{A.10}$$

it follows that the eigenvalues of $R_{AB}^{(d)}$ are

$$\left(\frac{p_1+p_2\pm|p_0-p_3|}{2},\frac{p_0+p_3\pm|p_1-p_2|}{2}\right)=\left(\frac{1}{2}-p_0,\frac{1}{2}-p_1,\frac{1}{2}-p_2,\frac{1}{2}-p_3\right).$$

We now consider two cases:

(1) If all $p_i\leqslant 1/2$ for all i, then the eigenvalues of $R_{AB}^{(d)}$ are all non-negative and $R_{AB}^{(d)}$ is a dual state. Hence, $S(R_{AB}^{(d)})\leqslant\log(4)=2$, which proves $0\leqslant I(A:B)^{(d)}$. We now show $S(R_{AB}^{(d)})\geqslant 1$. For this, let $\vec{\lambda}=(\lambda_1,\ldots,\lambda_4)$ denote the probability vector of eigenvalues of $R_{AB}^{(d)}$ with $\lambda_1\geqslant\cdots\geqslant\lambda_4$. Since $p_i\leqslant 1/2$ for $i=0,\ldots,3$ it follows that $\lambda_j\leqslant 1/2$ for $j=1,\ldots,4$. Thus, the probability vector $\vec{y}=(1/2,1/2,0,0)$ majorizes $\vec{\lambda}$. Hence, $1=H(\vec{y})\leqslant H(\vec{\lambda})=S(R_{AB}^{(d)})$ since the Shannon entropy is Schur-concave. It then follows that $I(A:B)^{(d)}\leqslant 1$ by (A.9). Combining the two inequalities yields $0\leqslant I(A:B)^{(d)}\leqslant 1$ in the case that $p_i\leqslant 1/2$ for all i.

(2) If there exists a $p_i>1/2$, without loss of generality, we can assume that $p_0>1/2$ and $p_i<1/2$ for $i=1,2,3$. We then have

$$\begin{aligned} S\left(R_{AB}^{(d)}\right)&=-\left(\frac{1}{2}-p_0\right)\log\left(p_0-\frac{1}{2}\right)-\sum_{i=1}^{3}\left(\frac{1}{2}-p_i\right)\log\left(\frac{1}{2}-p_i\right)\\ &=\left(p_0-\frac{1}{2}\right)\log\left(p_0-\frac{1}{2}\right)-\left(\frac{1}{2}+p_0\right)\log\left(\frac{1}{2}+p_0\right)\\ &\quad-\left(\frac{1}{2}+p_0\right)\sum_{i=1}^{3}\frac{\frac{1}{2}-p_i}{\frac{1}{2}+p_0}\log\left(\frac{\frac{1}{2}-p_i}{\frac{1}{2}+p_0}\right)\\ &=:S\left(R_{AB}^{(d)}\right)_{p_0}, \end{aligned} \tag{A.11}$$

where we use the notion $S(R_{AB}^{(d)})_{p_0}$ when we want to emphasize that we may view $S(R_{AB}^{(d)})$ as a function of p_0. We now wish to bound $S(R_{AB}^{(d)})$ from above and below. To obtain a minimum value of $S(R_{AB}^{(d)})$, we only need compute the minimum value of $S(R_{AB}^{(d)})$ for every fixed p_0, i.e.

$$\min S\left(R_{AB}^{(d)}\right) = \min_{p_0} \min_{p_1,p_2,p_3} S\left(R_{AB}^{(d)}\right)_{p_0}.$$

Now for a fixed p_0, it follows from (A.11) that $S(R_{AB}^{(d)})_{p_0}$ is minimized by finding p_1,p_2,p_3 which minimize the entropy of the probability vector $\vec{v} = (\frac{1/2-p_1}{p_0+1/2}, \frac{1/2-p_2}{p_0+1/2}, \frac{1/2-p_3}{p_0+1/2})$. As $p_i < 1/2$ for $i = 1,2,3$, it follows that the vector $\vec{u} = (\frac{p_0-1/2}{p_0+1/2}, \frac{1/2}{p_0+1/2}, \frac{1/2}{p_0+1/2})$, which is the vector $\vec{v}$ when $p_1 = 1-p_0$ and $p_2 = p_3 = 0$, majorizes $\vec{v}$. Hence,

$$-\frac{p_0-1/2}{p_0+1/2}\log\left(\frac{p_0-1/2}{p_0+1/2}\right) - \frac{1}{p_0+1/2}\log\left(\frac{1/2}{p_0+1/2}\right) = H(\vec{u}) \leqslant H(\vec{v}),$$

so that

$$\begin{aligned} S\left(R_{AB}^{(d)}\right)_{p_0} &= \left(p_0-\frac{1}{2}\right)\log\left(p_0-\frac{1}{2}\right) - \left(\frac{1}{2}+p_0\right)\log\left(\frac{1}{2}+p_0\right) + \left(\frac{1}{2}+p_0\right)H(\vec{v}) \\ &\geqslant \left(p_0-\frac{1}{2}\right)\log\left(p_0-\frac{1}{2}\right) - \left(\frac{1}{2}+p_0\right)\log\left(\frac{1}{2}+p_0\right) + \left(\frac{1}{2}+p_0\right)H(\vec{u}) \\ &= 1. \end{aligned} \tag{A.12}$$

Therefore, for every fixed p_0, we always have $S(R_{AB}^{(d)})_{p_0} \geqslant 1$ so that $S(R_{AB}^{(d)}) \geqslant 1$. To bound $S(R_{AB}^{(d)})$ from above, notice that the vector $\vec{v}$ majorizes the vector $\vec{w} = (1/3, 1/3, 1/3)$ and $\vec{w}$ can always be attained by choosing $p_i = \frac{1-p_0}{3}, i = 1,2,3$ for every fixed $p_0 > 1/2$. Hence,

$$S\left(R_{AB}^{(d)}\right)_{p_0} \leqslant \left(p_0-\frac{1}{2}\right)\log\left(p_0-\frac{1}{2}\right) - \left(\frac{1}{2}+p_0\right)\log\left(\frac{1}{2}+p_0\right) + \left(\frac{1}{2}+p_0\right)\log(3).$$

The maximum value of the right-hand-side of this inequality occurs in the limit $p_0 \to (1/2)^+$. Hence, $S(R_{AB}^{(d)})_{p_0} \leqslant S(R_{AB}^{(d)})_{1/2} \leqslant \log(3)$ for all values of $p_0 > \frac{1}{2}$. Thus, $0 < 2 - \log(3) \leqslant I(A:B)^{(d)}$. Combining this with $(A.9)$ yields $0 \leqslant I(A:B)^{(d)} \leqslant 1$.

By reduction to $\mathcal{N}_d$, it follows that if $\mathcal{N}$ is unital, then $0 \leqslant I(A:B) \leqslant 1$, as desired. We note that in such a case we have $I(A:B) = 0$ if and only if $\mathcal{N}_d$ is the completely depolarizing channel, i.e. when $p_0 = p_1 = p_2 = p_3 = 1/4$, and $I(A:B) = 1$ if and only if there are at least two p_i in $\mathcal{N}_d$ equal to zero, i.e. when the Choi rank of $\mathcal{N}_d$ is no more than 2.

A.3. Channels of Choi rank no greater than 2

In this section, we consider the case when the Choi rank of $\mathcal{N}$ is no greater than 2. First, suppose that a channel $\mathcal{M}: \mathcal{L}(\mathcal{H}_A) \to \mathcal{L}(\mathcal{H}_B)$ has a matrix representation in the same form as that of N_d as given by equation (A.7). Since $\mathcal{M}(\rho) = \frac{1}{2}\left(\mathbb{1} + (\mathbf{t} + \mathbf{T}(\vec{r}))\cdot\vec{\sigma}\right)$ is positive, it follows that $|t_i| + |\lambda_i| \leqslant 1$ for $i = 1,2,3$. Now we compute the Choi matrix of $\mathcal{M}$, which is the matrix $\mathscr{C}[\mathcal{M}] \in \mathcal{L}(\mathcal{H}_A) \otimes \mathcal{L}(\mathcal{H}_B)$ given by

$$\mathscr{C}[\mathcal{M}] = (\mathcal{I}\otimes\mathcal{M})\left(\sum_{i,j}|i\rangle\langle j|\otimes|i\rangle\langle j|\right),$$

where $\mathcal{I}$ denotes the identity map on $\mathcal{L}(\mathcal{H}_A)$. Now, since (by using (A.4))

$$\begin{aligned} (\mathcal{I}\otimes\mathcal{M})(|0\rangle\langle 0|\otimes|0\rangle\langle 0|) &= |0\rangle\langle 0| \otimes \frac{1}{2}\begin{pmatrix} 1+t_3+\lambda_3 & t_1-it_2 \\ t_1+it_2 & 1-t_3-\lambda_3 \end{pmatrix}, \\ (\mathcal{I}\otimes\mathcal{M})(|1\rangle\langle 1|\otimes|1\rangle\langle 1|) &= |1\rangle\langle 1| \otimes \frac{1}{2}\begin{pmatrix} 1+t_3-\lambda_3 & t_1-it_2 \\ t_1+it_2 & 1-t_3+\lambda_3 \end{pmatrix}, \end{aligned}$$

and

$$(\mathcal{I}\otimes\mathcal{M})(|0\rangle\langle 1|\otimes|0\rangle\langle 1| + |1\rangle\langle 0|\otimes|1\rangle\langle 0|) = \frac{1}{2}(\mathcal{I}\otimes\mathcal{M})(\sigma_X\otimes\sigma_X - \sigma_Y\otimes\sigma_Y),$$

the fact that $\mathcal{M}(\sigma_X) = \lambda_1\sigma_X$ and $\mathcal{M}(\sigma_Y) = \lambda_2\sigma_Y$ yields

$$\mathscr{C}[\mathcal{M}] = \frac{1}{2}\begin{pmatrix} 1+t_3+\lambda_3 & t_1-it_2 & 0 & \lambda_1+\lambda_2 \\ t_1+it_2 & 1-t_3-\lambda_3 & \lambda_1-\lambda_2 & 0 \\ 0 & \lambda_1-\lambda_2 & 1+t_3-\lambda_3 & t_1-it_2 \\ \lambda_1+\lambda_2 & 0 & t_1+it_2 & 1-t_3+\lambda_3 \end{pmatrix}.$$

Now suppose that $\mathrm{rank}(\mathscr{C}[\mathcal{M}]) \leqslant 2$. In such a case, it follows that the minors of order 3 and 4 must necessarily vanish. Setting the principal minors of order 3 equal to zero yields

$$M_{11} = (1-t_3-\lambda_3)\left(1-(t_3-\lambda_3)^2-\left(t_1^2+t_2^2\right)\right)-(\lambda_1-\lambda_2)^2(1-t_3+\lambda_3) = 0;$$
$$M_{12} = (t_1+it_2)\left(1-(t_3-\lambda_3)^2-\left(t_1^2+t_2^2\right)\right)+\left(\lambda_1^2-\lambda_2^2\right)(t_1-it_2) = 0;$$
$$M_{13} = (\lambda_1-\lambda_2)(t_1+it_2)(1-t_3+\lambda_3)+(\lambda_1+\lambda_2)(t_1-it_2)(1-t_3-\lambda_3) = 0;$$
$$M_{14} = (\lambda_1-\lambda_2)(t_1+it_2)^2+(\lambda_1+\lambda_2)\left((1-\lambda_3)^2-t_3^2-(\lambda_1-\lambda_2)^2\right) = 0;$$
$$M_{21} = (t_1-it_2)\left(1-(t_3-\lambda_3)^2-(t_1^2+t_2^2)\right)+(\lambda_1^2-\lambda_2^2)(t_1+it_2) = 0;$$
$$M_{22} = (1+t_3+\lambda_3)\left(1-(t_3-\lambda_3)^2-(t_1^2+t_2^2)\right)-(\lambda_1+\lambda_2)^2(1+t_3-\lambda_3) = 0;$$
$$M_{23} = (\lambda_1-\lambda_2)\left((1+\lambda_3)^2-t_3^2\right)+(\lambda_1+\lambda_2)\left((t_1-it_2)^2-(\lambda_1+\lambda_2)(\lambda_1-\lambda_2)\right) = 0;$$
$$M_{24} = (t_1+it_2)(\lambda_1-\lambda_2)(1+t_3+\lambda_3)+(\lambda_1+\lambda_2)(1+t_3-\lambda_3)(t_1-it_2) = 0,$$
$$M_{31} = (t_1-it_2)(\lambda_1-\lambda_2)(1-t_3+\lambda_3)+(\lambda_1+\lambda_2)(1-t_3-\lambda_3)(t_1+it_2) = 0,$$
$$M_{32} = (\lambda_1-\lambda_2)\left((1+\lambda_3)^2-t_3^2\right)+(\lambda_1+\lambda_2)\left((t_1+it_2)^2-(\lambda_1+\lambda_2)(\lambda_1-\lambda_2)\right) = 0;$$
$$M_{33} = (1-t_3+\lambda_3)\left(1-(t_3+\lambda_3)^2-(t_1^2+t_2^2)\right)-(\lambda_1+\lambda_2)^2(1-t_3-\lambda_3) = 0;$$
$$M_{34} = (t_1+it_2)\left(1-(t_3+\lambda_3)^2\right)-(t_1-it_2)\left((t_1+it_2)^2-(\lambda_1+\lambda_2)(\lambda_1-\lambda_2)\right) = 0,$$
$$M_{41} = (\lambda_1-\lambda_2)(t_1-it_2)^2+(\lambda_1+\lambda_2)\left((1-\lambda_3)^2-t_3^2-(\lambda_1-\lambda_2)^2\right) = 0;$$
$$M_{42} = (t_1-it_2)(\lambda_1-\lambda_2)(1+t_3+\lambda_3)+(\lambda_1+\lambda_2)(1+t_3-\lambda_3)(t_1+it_2) = 0,$$
$$M_{43} = (t_1-it_2)\left(1-(t_3+\lambda_3)^2\right)-(t_1+it_2)\left((t_1-it_2)^2-(\lambda_1+\lambda_2)(\lambda_1-\lambda_2)\right) = 0,$$
$$M_{44} = (1+t_3-\lambda_3)\left(1-(t_3+\lambda_3)^2-(t_1^2+t_2^2)\right)-(\lambda_1-\lambda_2)^2(1+t_3+\lambda_3) = 0;$$

The equations $M_{14}-M_{41}=0$ and $M_{32}-M_{23}=0$ yield $(\lambda_1-\lambda_2)t_1t_2=0$ and $(\lambda_1+\lambda_2)t_1t_2=0$. If $t_1, t_2 \neq 0$, then $\lambda_1=\lambda_2=0$, and to ensure $M_{ij}=0$, we have $\lambda_3t_3=0$ (this is immediate from $M_{11}+M_{22}-M_{33}-M_{44}=0$ for example). We now consider various cases for λ_i and t_i.

1. If $t_1\neq 0$, $t_2\neq 0$ (so that $\lambda_1=\lambda_2=0$), and $\lambda_3=0$, then $t_1^2+t_2^2+t_3^2=1$. In this case, the matrix representation is

$$\begin{pmatrix} 1 & 0 & 0 & 0 \\ \cos u & 0 & 0 & 0 \\ \sin u\cos v & 0 & 0 & 0 \\ \sin u\sin v & 0 & 0 & 0 \end{pmatrix},$$

with $\cos u, \sin u\cos v \neq 0$.

2. If $t_1\neq 0$, $t_2\neq 0$ (so that $\lambda_1=\lambda_2=0$), and $t_3=0$, then $t_1^2+t_2^2+\lambda_3^2=1$. In this case, the matrix representation is

$$\begin{pmatrix} 1 & 0 & 0 & 0 \\ \cos u & 0 & 0 & 0 \\ \sin u\cos v & 0 & 0 & 0 \\ 0 & 0 & 0 & \sin u\sin v \end{pmatrix},$$

with $\cos u, \sin u\cos v \neq 0$.

Now, suppose that $t_1=0$ but $t_2\neq 0$. Using $M_{12}-M_{34}=0$, we obtain $\lambda_3t_3=0$. If $\lambda_3=0$, then the condition $M_{11}-M_{22}-M_{33}+M_{44}=0$ yields $\lambda_1\lambda_2=0$.

3. If $t_1 = 0,\ t_2 \neq 0$, and $\lambda_2 = \lambda_3 = 0$, then $1 = t_2^2 + t_3^2 + \lambda_1^2$. In this case, the matrix representation is

$$\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & \sin u \sin v & 0 & 0 \\ \cos u & 0 & 0 & 0 \\ \sin u \cos v & 0 & 0 & 0 \end{pmatrix},$$

with $\cos u \neq 0$.

4. If $t_1 = 0,\ t_2 \neq 0$, and $\lambda_1 = \lambda_3 = 0$, then $M_{12} = 0$ yields $1 = t_2^2 + t_3^2 - \lambda_2^2$, while $M_{11} + M_{22} = 0$ yields $1 = t_2^2 + t_3^2 + \lambda_2^2$. Together, these imply $\lambda_2 = 0$ and $t_2^2 + t_3^2 = 1$, so that the associated matrix representation is

$$\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ \cos u & 0 & 0 & 0 \\ \sin u & 0 & 0 & 0 \end{pmatrix},$$

with $\cos u \neq 0$.

5. If $t_1 = 0$, $t_2 \neq 0$, and $t_3 = 0$, then $M_{12} = 0$ yields $1 - \lambda_3^2 - t_2^2 - \lambda_1^2 + \lambda_2^2 = 0$, and moreover, $M_{13} = 0$ yields $\lambda_2 = \lambda_1 \lambda_3$. Therefore, the matrix representation is

$$\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos u & 0 & 0 \\ \sin u \sin v & 0 & \cos u \cos v & 0 \\ 0 & 0 & 0 & \cos v \end{pmatrix},$$

with $\cos u \cos v \neq 0$.

If $t_1 \neq 0$ but $t_2 = 0$, using $M_{12} = 0\,, M_{34} = 0$, we obtain $\lambda_3 t_3 = 0$. If $\lambda_3 = 0$, by $M_{ii} = 0$ for $i = 1, 2, 3, 4$ implies $\lambda_1 \lambda_2 = 0$.

6. If $\lambda_2 = \lambda_3 = 0$, then $M_{11} = (1 - t_3)(1 - t_3^2 - t_1^2 - \lambda_1^2) = 0\,, M_{22} = (1 + t_3)(1 - t_3^2 - t_1^2 - \lambda_1^2) = 0$, and $M_{12} = t_1(1 - t_3^2 - t_1^2 + \lambda_1^2) = 0$. Since $t_1 \neq 0$, we have $\lambda_1 = 0$ and $t_1^2 + t_3^2 = 1$. In this case, the matrix representation is

$$\begin{pmatrix} 1 & 0 & 0 & 0 \\ \cos u & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ \sin u & 0 & 0 & 0 \end{pmatrix},$$

with $\cos u \neq 0$.

7. If $\lambda_1 = \lambda_3 = 0$, then $1 = t_3^2 + t_1^2 + \lambda_2^2$. In this case, the matrix representation is

$$\begin{pmatrix} 1 & 0 & 0 & 0 \\ \cos u & 0 & 0 & 0 \\ 0 & 0 & \sin u \sin v & 0 \\ \sin u \cos v & 0 & 0 & 0 \end{pmatrix},$$

with $\cos u \neq 0$.

8. If $t_3 = 0$, then $M_{12} = 0$ leads $1 - \lambda_3^2 - t_1^2 + \lambda_1^2 - \lambda_2^2 = 0$ and $M_{13} = 0$ yields $\lambda_1 = \lambda_2 \lambda_3$. Therefore, the associated matrix representation is

$$\begin{pmatrix} 1 & 0 & 0 & 0 \\ \sin u \sin v & \cos u \cos v & 0 & 0 \\ 0 & 0 & \cos u & 0 \\ 0 & 0 & 0 & \cos v \end{pmatrix},$$

with $\sin u \sin v \neq 0$.

9. Finally, if $t_1 = t_2 = 0$, then the associated matrix representation is

$$\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos u & 0 & 0 \\ 0 & 0 & \cos v & 0 \\ \sin u \sin v & 0 & 0 & \cos u \cos v \end{pmatrix}.$$

As cases $1, 2, 3, 4, 6$ and 7 differ by unitary transformations, and cases $5, 8$ and 9 also differ by unitary transformations, we arrive at the following:

Proposition 1. *Suppose the channel $\mathcal{N}_d$ is of Choi rank no more than 2. Then the matrix representation of $\mathcal{N}_d$ is unitarily equivalent to one of following matrices:*

$$\text{(i)} \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos u & 0 & 0 \\ 0 & 0 & \cos v & 0 \\ \sin u \sin v & 0 & 0 & \cos u \cos v \end{pmatrix}, \textit{ with } u \in [0, 2\pi), v \in [0, \pi).$$

$$\text{(ii)} \begin{pmatrix} 1 & 0 & 0 & 0 \\ \cos u & 0 & 0 & 0 \\ 0 & 0 & \sin u \sin v & 0 \\ \sin u \cos v & 0 & 0 & 0 \end{pmatrix}, \textit{ with } u \in [0, 2\pi), v \in [0, \pi) \textit{ and } \cos u \neq 0.$$

We now compute the mutual information of the PDM associated with cases (i) and (ii) as given by proposition 1. In what follows, the initial state $\rho = \mathbb{1}/2$ is fixed.

Case (i): In this case is the matrix N_d is of the form

$$N_d = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos u & 0 & 0 \\ 0 & 0 & \cos v & 0 \\ \sin u \sin v & 0 & 0 & \cos u \cos v \end{pmatrix}, \tag{A.13}$$

with $u \in [0, 2\pi), v \in [0, \pi)$. Since $\mathcal{N}_d$ is non-unital, we have $\sin u \sin v > 0$, so that for $\rho = \mathbb{1}/2$ the PDM $R_{AB} = \mathcal{N}_d \star \rho$ is then given by

$$\begin{aligned} R_{AB} &\overset{(A.3)}{=} \frac{1}{2} \mathscr{J}[\mathcal{N}_d] \\ &= \frac{1}{2} \begin{pmatrix} \cos^2 \frac{v-u}{2} & 0 & 0 & \sin \frac{v-u}{2} \sin \frac{v+u}{2} \\ 0 & \sin^2 \frac{v-u}{2} & \cos \frac{v-u}{2} \cos \frac{v+u}{2} & 0 \\ 0 & \cos \frac{v-u}{2} \cos \frac{v+u}{2} & \sin^2 \frac{v+u}{2} & 0 \\ \sin \frac{v-u}{2} \sin \frac{v+u}{2} & 0 & 0 & \cos^2 \frac{v+u}{2} \end{pmatrix}. \end{aligned} \tag{A.14}$$

Hence, the eigenvalues of R_{AB} are

$$\left(\frac{1}{2}, \frac{1}{2}, \frac{\cos u \cos v}{2}, -\frac{\cos u \cos v}{2} \right).$$

It then follows that $S(R_{AB}) = S(\rho) = 1$, which yields

$$I(A:B) = S(\rho) + S(\mathcal{N}_d(\rho)) - S(R_{AB}) = S(\mathcal{N}_d(\rho)).$$

Case (ii): In this case the matrix N_d is of the form

$$N_d = \begin{pmatrix} 1 & 0 & 0 & 0 \\ \cos u & 0 & 0 & 0 \\ 0 & 0 & \sin u \sin v & 0 \\ \sin u \cos v & 0 & 0 & 0 \end{pmatrix}, \tag{A.15}$$

with $u \in [0, 2\pi), v \in [0, \pi)$ and $\cos u \neq 0$. For $\rho = \mathbb{1}/2$ the PDM $R_{AB} = \mathcal{N}_d \star \rho$ is then given by

$$R_{AB} \overset{(A.3)}{=} \frac{1}{2}\mathscr{J}\left[\mathcal{N}_d\right]$$
$$= \frac{1}{4}\begin{pmatrix} 1+\sin u\cos v & \cos u & 0 & -\sin u\sin v \\ \cos u & 1-\sin u\cos v & \sin u\sin v & 0 \\ 0 & \sin u\sin v & 1+\sin u\cos v & \cos u \\ -\sin u\sin v & 0 & \cos u & 1-\sin u\cos v \end{pmatrix}. \tag{A.16}$$

Hence, the eigenvalues of R_{AB} are

$$\left(0,0,\frac{1}{2},\frac{1}{2}\right).$$

It then follows that $S(R_{AB}) = S(\rho) = 1$, which yields

$$I(A:B) = S(\rho) + S(\mathcal{N}_d(\rho)) - S(R_{AB}) = S(\mathcal{N}_d(\rho)).$$

Since $\mathcal{N}_d(\rho)$ is a density matrix, $0 \leqslant S(\mathcal{N}_d(\rho)) \leqslant 1$, and so $0 \leqslant I(A:B) \leqslant 1$. By reduction to $\mathcal{N}_d$, it follows that if the Choi rank of $\mathcal{N}$ is no greater than 2, then $0 \leqslant I(A:B) \leqslant 1$, as desired.

Appendix B. Holevo bound for sequential measurements

In this section, we prove a lemma which implies that the Holevo bound also holds for *sequential* measurements on an ensemble of quantum states.

Lemma 1. *Let X be a random variable whose values are associated with a probability distribution p_i, let $p_{jk|i}$ be a family of bivariate probability distributions conditioned on i, and suppose there exists a quantum channel $\mathcal{N}: \mathcal{L}(\mathcal{H}_A) \to \mathcal{L}(\mathcal{H}_{Z_1}) \otimes \mathcal{L}(\mathcal{H}_{Z_2})$ and a finite collection of density matrices $\{\rho^i\} \subset \mathcal{L}(\mathcal{H}_A)$, indexed by the same set as the probability distribution p_i, such that $\mathcal{N}(\rho^i) = \sum_{j,k} p_{jk|i}|j\rangle\langle j| \otimes |k\rangle\langle k|$ for all i for some choice of orthonormal bases $\{|j\rangle\}$ and $\{|k\rangle\}$ on $\mathcal{H}_{Z_1}$ and $\mathcal{H}_{Z_2}$, respectively. Then for every bivariate random variable (Y_1, Y_2) whose values are associated with the bivariate distribution $p_{jk} = \sum_i p_i p_{jk|i}$,*

$$I(X:Y_1Y_2) \leqslant S\left(\sum_i p_i\rho^i\right) - \sum_i p_i S\left(\rho^i\right),$$

where $I(X:Y_1Y_2)$ is the mutual information of the random variables X and (Y_1, Y_2).

Proof. By Stinespring's dilation theorem [42, 43], there exists an isometry $\mathcal{V}: \mathcal{L}(\mathcal{H}_A) \to \mathcal{L}(\mathcal{H}_{Z_1}) \otimes \mathcal{L}(\mathcal{H}_{Z_2}) \otimes \mathcal{L}(\mathcal{H}_{Z_3})$ and an operator V such that $\mathcal{V}(\rho) = V\rho V^\dagger$ for all $\rho \in \mathcal{L}(\mathcal{H}_A)$ and $\mathcal{N}(\rho) = \mathrm{Tr}_{Z_3}\left[V\rho V^\dagger\right]$ for all $\rho \in \mathcal{L}(\mathcal{H}_A)$. Now let $\rho_{XZ_1Z_2Z_3}$ be the density matrix given by

$$\rho_{XZ_1Z_2Z_3} = \sum_i p_i|i\rangle\langle i| \otimes V\rho^i V^\dagger,$$

where $\{|i\rangle\}$ is an arbitrary orthonormal basis on a Hilbert space of dimension equal to the index set for the probability distribution p_i. Then, let $\rho_{XZ_1Z_2}$ be the density matrix given by

$$\rho_{XZ_1Z_2} = \sum_{i,j,k} p_{ijk}|i\rangle\langle i| \otimes |j\rangle\langle j| \otimes |k\rangle\langle k|,$$

where $p_{ijk} = p_i p_{jk|i}$. We then have

$$\begin{aligned}\mathrm{Tr}_{Z_3}\left[\rho_{XZ_1Z_2Z_3}\right] = \mathrm{Tr}_{Z_3}\left[\sum_i p_i|i\rangle\langle i| \otimes V\rho^i V^\dagger\right] &= \sum_i p_i|i\rangle\langle i| \otimes \mathrm{Tr}_{Z_3}\left[V\rho^i V^\dagger\right] \\ &= \sum_i p_i|i\rangle\langle i| \otimes \mathcal{N}\left(\rho^i\right) = \sum_i p_i|i\rangle\langle i| \otimes \left(\sum_{j,k} p_{jk|i}|j\rangle\langle j| \otimes |k\rangle\langle k|\right) \\ &= \sum_{i,j,k} p_i p_{jk|i}|i\rangle\langle i| \otimes |j\rangle\langle j| \otimes |k\rangle\langle k| = \sum_{i,j,k} p_{ijk}|i\rangle\langle i| \otimes |j\rangle\langle j| \otimes |k\rangle\langle k| \\ &= \rho_{XZ_1Z_2}.\end{aligned}$$

Hence, monotonicity of (spatial) quantum mutual information implies $I(X:Z_1Z_2) \leqslant I(X:Z_1Z_2Z_3)$, where $I(X:Z_1Z_2)$ is the quantum mutual information associated with the density matrix $\rho_{XZ_1Z_2}$ and $I(X:Z_1Z_2Z_3)$ is the quantum mutual information associated with the density matrix $\rho_{XZ_1Z_2Z_3}$. Moreover, we have

$$\begin{aligned} I(X:Z_1Z_2Z_3) &= S\left(\sum_i p_i|i\rangle\langle i|\right) + S(\rho_{Z_1Z_2Z_3}) - S(\rho_{XZ_1Z_2Z_3}) \\ &= H(p) + S\left(\sum_i p_i V\rho^i V^\dagger\right) - S\left(\sum_i p_i|i\rangle\langle i| \otimes V\rho^i V^\dagger\right) \\ &= H(p) + S\left(\sum_i p_i\rho^i\right) - \left(H(p) + \sum_i p_i S(\rho^i)\right) \\ &= S\left(\sum_i p_i\rho^i\right) - \sum_i p_i S(\rho^i), \end{aligned}$$

where the third equality follows from the isometric invariance of von Neumann entropy. Finally, since $I(X:Y_1Y_2) = I(X:Z_1Z_2)$, we then combine these results to obtain

$$I(X:Y_1Y_2) = I(X:Z_1Z_2) \leqslant I(X:Z_1Z_2Z_3) = S\left(\sum_i p_i\rho^i\right) - \sum_i p_i S(\rho^i),$$

as desired. □

ORCID iD

James Fullwood https://orcid.org/0000-0002-8816-1364

References

[1] Cerf N J and Adami C 1997 *Phys. Rev. Lett.* **79** 5194–7
[2] Nielsen M A and Chuang I L 2011 *Quantum Computation and Quantum Information* 10th edn (Cambridge University Press) (https://doi.org/10.1017/CBO9780511976667)
[3] Cover T M and Thomas J A 2006 *Elements of Information Theory (Wiley Series in Telecommunications and Signal Processing)* (Wiley)
[4] Fitzsimons J F, Jones J A and Vedral V 2015 *Sci. Rep.* **5** 18281
[5] Fullwood J and Parzygnat A J 2024 Operator representation of spatiotemporal quantum correlations (arXiv:2405.17555)
[6] Marletto C, Vedral V, Virzì S, Avella A, Piacentini F, Gramegna M, Degiovanni I P and Genovese M 2021 *Sci. Adv.* **7** eabe4742
[7] Zhao Z, Pisarczyk R, Thompson J, Gu M, Vedral V and Fitzsimons J F 2018 *Phys. Rev. A* **98** 052312
[8] Pisarczyk R, Zhao Z, Ouyang Y, Vedral V and Fitzsimons J F 2019 *Phys. Rev. Lett.* **123** 150502
[9] Marletto C, Vedral V, Virzì S, Rebufello E, Avella A, Piacentini F, Gramegna M, Degiovanni I P and Genovese M 2020 *Entropy* **22** 228
[10] Marletto C, Vedral V, Virzì S, Rebufello E, Avella A, Piacentini F, Gramegna M, Degiovanni I P and Genovese M 2019 *Nat. Commun.* **10** 182
[11] Liu X, Chen Q and Dahlsten O 2024 *Phys. Rev. A* **109** 032219
[12] Liu X, Qiu Y, Dahlsten O and Vedral V 2025 *npj Quantum Inf.* **11** 54
[13] Song M, Narasimhachar V, Regula B, Elliott T J and Gu M 2024 *Phys. Rev. Lett.* **133** 110202
[14] Jia Z, Song M and Kaszlikowski D 2023 *New J. Phys.* **25** 123038
[15] Parzygnat A J and Fullwood J 2024 Time-symmetric correlations for open quantum systems (arXiv:2407.11123)
[16] Utagi S 2021 *Phys. Lett. A* **386** 126983
[17] French A P 1968 *Special Relativity* (W. W. Norton & Company)
[18] Hawking S W and Ellis G F R 1973 *The Large Scale Structure of Space-Time (Cambridge Monographs on Mathematical Physics)* (Cambridge University Press) (https://doi.org/10.1017/CBO9780511524646)
[19] Fullwood J and Parzygnat A J 2023 On dynamical measures of quantum information (arXiv:2306.01831)
[20] Parzygnat A J, Takayanagi T, Taki Y and Wei Z 2023 *J. High Energy Phys.* JHEP12(2023)123
[21] Tu Y T, Tzeng Y C and Chang P Y 2022 *SciPost Phys.* **12** 194
[22] Salek S, Schubert R and Wiesner K 2014 *Phys. Rev. A* **90** 022116
[23] Holevo A S 1973 *Problemy Peredaci Informacii* **9** 3–11
[24] Parzygnat A J and Fullwood J 2023 *PRX Quantum* **4** 020334
[25] Horsman D, Heunen C, Pusey M F, Barrett J and Spekkens R W 2017 *Proc. R. Soc. A* **473** 20170395
[26] Jamiołkowski A 1972 *Rep. Math. Phys.* **3** 275–8
[27] Fullwood J 2023 Quantum dynamics as a pseudo-density matrix (arXiv:2304.03954)
[28] Werner R F 1989 *Phys. Rev. A* **40** 4277–81
[29] Brukner C, Taylor S, Cheung S and Vedral V 2004 Quantum entanglement in time (arXiv:quant-ph/0402127)
[30] Brody D C and Hughston L P 2024 Decoherence implies information gain (arXiv:2402.16740)
[31] Fritz T 2010 *New J. Phys.* **12** 083055
[32] Glorioso P, Qi X L and Yang Z 2024 *J. High Energy Phys.* JHEP05(2024)338
[33] Lloyd S 1997 *Phys. Rev. A* **55** 1613–22

[34] Devetak I 2005 *IEEE Trans. Inf. Theor.* **51** 44–55
[35] Holevo A S 2020 *Quantum Electron* **50** 440–6
[36] Ruskai M B, Szarek S and Werner E 2002 *Linear Algebra Its Appl.* **347** 159–87
[37] Fujiwara A and Algoet P 1999 *Phys. Rev. A* **59** 3290–4
[38] Braun D, Giraud O, Nechita I, Pellegrini C and Žnidarič M 2014 *J. Phys. A: Math. Theor.* **47** 135302
[39] Parzygnat A J, Fullwood J, Buscemi F and Chiribella G 2024 *Phys. Rev. Lett.* **132** 110203
[40] Fullwood J 2023 General covariance for quantum states over time (arXiv:2311.00162)
[41] Marshall A W, Olkin I and Arnold B C 2010 *Inequalities: Theory of Majorization and Its Applications* (Springer) (https://doi.org/10.1007/978-0-387-68276-1)
[42] Stinespring W F 1955 *Proc. Am. Math. Soc.* **6** 211–6
[43] Parzygnat A J 2019 *Compositionality* **1** 2