PAPER • OPEN ACCESS

Channel discord and distortion

To cite this article: Wei-Wei Zhang *et al* 2021 *New J. Phys.* **23** 083025

View the article online for updates and enhancements.

PAPER

Channel discord and distortion

Wei-Wei Zhang[1,2,3,*] , Yuval R. Sanders[4] and Barry C. Sanders[5,6]

1. Max Planck Institute for the Science of Light, Staudtstrae 2, 91058 Erlangen, Germany
2. Institute for Theoretical Physics II, Friedrich-Alexander-Universität Erlangen-Nürnberg, Staudtstrae 7, 91058 Erlangen, Germany
3. Centre for Engineered Quantum Systems, School of Physics, The University of Sydney, Sydney, Australia
4. Department of Physics and Astronomy and ARC Centre of Excellence for Engineered Quantum Systems, Macquarie University, Sydney, Australia
5. Shanghai Branch, National Center for Physical Sciences at Microscale, University of Science and Technology of China, Shanghai 201315, People's Republic of China
6. Institute for Quantum Science and Technology, University of Calgary, Alberta, T2N 1N4, Canada
* Author to whom any correspondence should be addressed.

E-mail: zhangweiwei2021@gusulab.ac.cn

Abstract

Discord, originally notable as a signature of bipartite quantum correlation, in fact can be nonzero classically, i.e. arising from noisy measurements by one of the two parties. Here we redefine classical discord to quantify channel distortion, in contrast to the previous restriction of classical discord to a state, and we then show a monotonic relationship between classical (channel) discord and channel distortion. We show that classical discord is equivalent to (doubly stochastic) channel distortion by numerically discovering a monotonic relation between discord and total-variation distance for a bipartite protocol with one party having a noiseless channel and the other party having a noisy channel. Our numerical method includes randomly generating doubly stochastic matrices for noisy channels and averaging over a uniform measure of input messages. Connecting discord with distortion establishes discord as a signature of classical, not quantum, channel distortion.

1. Introduction

Discord is often touted as a quantifier of quantum correlations in a state, with nonzero discord said to imply that observed correlations transcends non-quantum (i.e. 'classical') limits [1, 2], akin to, but different from, a Bell inequality [3]. Treated as a quantum resource operationalized by state merging [4], discord in quantum computing protocols [5, 6] is believed by many to deliver a quantum advantage to some protocols [7]. However, this quantum nature of discord has been challenged by stochastic information, which shows that discord is due to noisy measurement by one of the two parties [8]. Essentially, discord can be understood in terms of a protocol amenable to a stochastic-information interpretation, and this interpretation fails if and only if (iff) the two parties share bipartite entanglement [8]. Discord thus serves as a fascinating starting point for studying stochastic information.

Previous work analyzed state discord in the context of classical states [8], i.e. analyzing discord as a signature of classical rather than quantum correlations. Here we introduce the concept of classical channel discord, which is based on averaging over all allowed channel input states for a given channel, with the previous definition used in assessing how much discord is added to a state by the given channel. Our expression for channel discord is amenable to numerical evaluation, which shows channel discord is monotonic with respect to channel distortion. We augment this numerical analysis by solving analytically the small but nontrivial two-bit case (each of two parties holds one bit) to confirm our numerics for this case and establish a path for proving discord–distortion monotonicity, which is a challenging calculation as we show. Our monotonicity result establishes meaningfully that channel discord and distortion are essentially equivalent.

Mathematically, discord in a two parties state refers to an apparent discrepancy between two expressions for mutual information obtained by two parties named here as Alice (A) and Bob (B), one quantity depending on joint probability and then other depending on conditional probability. The usual view of the reason for quantum discord in a state is that conditional information must be adapted to the quantum case by introducing measurement, and this incompatibility gives rise to nonzero discord [2]. Quantum discord in a state has been shown to be equivalent to classical discord iff entanglement between the two parties is zero, and this equivalence has been explained by showing that, in the absence of entanglement, discord represents one party, namely Bob, suffering from noisy measurement whereas Alice's measurements are ideal [8]. Our goal is to show that classical channel discord, obtained by averaging channel-added discord to all allowed channel input states, is equivalent to channel distortion by establishing a monotonic relation between discord and total-variation, or Kolmogorov, distance, which quantifies channel distortion. We analyze the general case numerically and solve analytically for the two-bit case.

Our article is structured as follows. In section 2 we summarise essential background on stochastic information, discord, total-variation distance and doubly stochastic channels. Our approach is described in section 3 and our approach specifically elaborates on our model for describing a noisy protocol for creating channel discord. Furthermore, section 3 presents our notation and mathematical expressions, and our methods for solving these expressions numerically. Subsequently, in section 4, we present our numerical results and explain the plots, and we discuss the results thoroughly in section 5. In section 6, we summarise our claims and provide an outlook. In appendices A and B we introduce convenient notation for probability vectors and what we call Hadamard calculus respectively, which fundamental to our approach.

2. Background

In this section we discuss the background and context for our work. In section 2.1 we discuss informational states including what we call stochastic information, which is a probabilistic mixture of definite informational states; specifics regarding probabilistic information states are explained in appendix B, which is based on the Hadamard notation explained in appendix A. Included in this discussion, we review the notions of entropy by considering shared information between parties, and we also review the notions of entropy and mutual information, all in the elegant Hadamard notation elaborated in appendix A, which we apply for the first time to this application. Then, in section 2.2, we explain mappings of information states in terms of channels with special emphasis on doubly stochastic channels. In this subsection we discuss ways to quantify how stochastic a channel is. Finally, in section 2.3, we review the notion of classical discord for states and the concepts of total-variation distance for stochastic information states; quantum discord is explained in appendix C.

2.1. Stochastic information

In this subsection we review the concept and mathematical framework for stochastic information based on the probability-vector representation elaborated in appendix B which uses Hadamard calculus introduced in appendix A. Then we discuss known concepts concerning entropy of a stochastic-information state by using Hadamard calculus introduced in appendix A. Finally, we review bipartite stochastic information including conditional entropy and mutual information.

The joint probability of messages shared between Alice, whose message size is M^A, and Bob, whose message size is M^B, is the bipartite matrix

$$\boldsymbol{p}^{AB} = \sum_{mm'} p_{mm'}^{AB} \boldsymbol{\delta}_{mm'}^{AB} \in \mathrm{mat}_{M^A \times M^B}\left(\mathbb{R}^{\geqslant 0}\right), \quad \left\|\boldsymbol{p}^{AB}\right\| = 1, \tag{1}$$

using the notation that $\mathrm{mat}_{M^A \times M^B}(R)$ refers to matrices with M^A rows and M^B columns whose entries are from any ring R. The norm is defined by equation (A4). Here we have let $\boldsymbol{\delta}_{mm'}^{AB}$ denote a versor for message m' as discussed in appendix B.

In concordance with quantum-information nomenclature, we refer to $\boldsymbol{\delta}_{\tilde{m}}$ as a 'pure state' [9]. Impure states refer to 'mixed' states, which are probabilistic mixtures of pure states and are obtained as a probabilistic mixture of any pure state. We discuss mixed and pure bipartite stochastic-information states in appendix B.

Now we discuss the entropy of the probability vector representing the mixed message. Mixedness of a state $\boldsymbol{p} \in \mathbb{R}^M$ is quantified by entropy [10]

$$0 \leqslant H(\boldsymbol{p}) := -\boldsymbol{p} \odot \log \boldsymbol{p} \leqslant \log M \tag{2}$$

using Hadamard notation explicated in appendix A. A state is pure iff its entropy is zero, which follows from $\boldsymbol{p} \circ \log \boldsymbol{p} = \boldsymbol{0}$ for a versor. The lower bound for the entropy (2) is $H(\boldsymbol{p}) = 0$ for a pure state. The upper bound for entropy is $H(\boldsymbol{p}) = \log M$ for a uniformly mixed state, with M as the size of $\boldsymbol{p}$. A high-entropy state is a state whose entropy is close to this bound. The joint state (1) has joint entropy H^{AB} (2) and total message size M. In our analysis, we always assume, without loss of generality, that

$$M^{\mathrm{A}} \equiv M^{\mathrm{B}} \implies M^{\mathrm{A}} = \sqrt{M} = M^{\mathrm{B}}. \tag{3}$$

Thus, the joint entropy of the bipartite versor (B5) is zero as required for a pure state. The matrix representation of $\boldsymbol{p}^{\mathrm{AB}}$ is an $M^{\mathrm{A}} \times M^{\mathrm{B}}$ matrix, with nonnegative real entries such that the sum of all entries is one.

The marginal distribution is obtained by ignoring the other party's share of the mixed state. Hence, Alice's marginal distribution is the probability vector

$$\boldsymbol{p}^{\mathrm{A}} := \left(\sum_{m'} p_{mm'}^{\mathrm{AB}} \right), \; \|\boldsymbol{p}^{\mathrm{A}}\| = \sum_{m} p_{m}^{\mathrm{A}} \tag{4}$$

using the unit one-norm. Similarly, we construct the marginal distribution $\boldsymbol{p}^{\mathrm{B}}$ by summing over Alice's degree of freedom.

Alice's state conditioned on Bob's state is

$$\boldsymbol{p}^{\mathrm{A|B}} := \boldsymbol{p}^{\mathrm{AB}} \oslash \boldsymbol{p}^{\mathrm{B}} = \left(\frac{p_{mm'}^{\mathrm{AB}}}{p_{m'}^{\mathrm{B}}} \right) \tag{5}$$

with $\oslash$ explained in appendix A. The last term of equation (5) displays row-vector elements $\boldsymbol{p}_{m}^{\mathrm{A}}$ obtained by element-wise division of each matrix element $p_{mm'}^{\mathrm{AB}}$ by respective column-vector elements $p_{m'}^{\mathrm{B}}$. Similarly, the conditional probability distribution for Bob is $\boldsymbol{p}^{\mathrm{B|A}} = \boldsymbol{p}^{\mathrm{AB}} \oslash \boldsymbol{p}_{\mathrm{A}}$ analogous to (5).

The entropy of the conditional probability distribution $\boldsymbol{p}^{\mathrm{A|B}}$ is

$$H^{\mathrm{A|B}} \left(\boldsymbol{p}^{\mathrm{AB}} \right) = H \left(\boldsymbol{p}^{\mathrm{A|B}} \right). \tag{6}$$

Fact 1. *A bipartite stochastic state $\boldsymbol{p}^{\mathrm{AB}}$, which decomposes to $\boldsymbol{p}^{\mathrm{AB}} \oslash \boldsymbol{p}^{\mathrm{B}}$ and to $\boldsymbol{p}^{\mathrm{AB}} \oslash \boldsymbol{p}^{\mathrm{A}}$, is conditionally pure (cp) iff*

$$H^{A|B} \equiv 0 \equiv H^{B|A}. \tag{7}$$

Consequently, $\boldsymbol{p}^{\mathrm{AB}}$ is cp iff it is permutationally equivalent to a diagonal matrix; i.e. in diag$([0, 1])$, which refers to the set of diagonal matrices whose entries are each in the real-number interval $[0, 1]$. Thus,

$$\exists \sigma \in S_{M^{\mathrm{A}}}, \sigma' \in S_{M^{\mathrm{B}}} : \Pi_\sigma \boldsymbol{p}^{\mathrm{AB}} \Pi_{\sigma'} \in \mathrm{diag}_{\min\{M^{\mathrm{A}}, M^{\mathrm{B}}\}}([0, 1]), \tag{8}$$

i.e. is diagonal of size $\min\{M^{\mathrm{A}}, M^{\mathrm{B}}\} \times \min\{M^{\mathrm{A}}, M^{\mathrm{B}}\}$. Furthermore, conditional probability distributions $\boldsymbol{p}^{\mathrm{A|B}}$ and $\boldsymbol{p}^{\mathrm{B|A}}$, which are obtained from a bipartite stochastic pure information state $\boldsymbol{p}^{\mathrm{AB}}$, are necessarily pure.

Operationally speaking, a bipartite state is cp only if Alice's pure state can be known by Bob after he measures his share of the joint stochastic-information state and vice versa. Consequently, the bipartite versor (B5) used in equation (1) is cp.

Mutual information

$$
\begin{aligned}
I^{\mathrm{A;B}} \left(\boldsymbol{p}^{\mathrm{AB}} \right) &:= H^{\mathrm{A}} \left(\boldsymbol{p}^{\mathrm{AB}} \right) + H^{\mathrm{B}} \left(\boldsymbol{p}^{\mathrm{AB}} \right) - H^{\mathrm{AB}} \left(\boldsymbol{p}^{\mathrm{AB}} \right) \\
&= \boldsymbol{p}^{\mathrm{AB}} \odot \log \left(\boldsymbol{p}^{\mathrm{AB}} \oslash \boldsymbol{p}_{\mathrm{A}} \oslash \boldsymbol{p}^{\mathrm{B}} \right)
\end{aligned} \tag{9}
$$

quantifies correlation between two parties, Alice and Bob, with the last part of this expression expressed in a novel way by using Hadamard arithmetic. An equivalent, alternative mutual information definition is

$$J^{\mathrm{A;B}} \left(\boldsymbol{p}^{\mathrm{AB}} \right) := H^{\mathrm{A}} \left(\boldsymbol{p}^{\mathrm{AB}} \right) - H^{\mathrm{A|B}} \left(\boldsymbol{p}^{\mathrm{AB}} \right) = I^{\mathrm{A;B}} \left(\boldsymbol{p}^{\mathrm{AB}} \right). \tag{10}$$

Consequently,

$$\Delta^{\mathrm{A;B}} \left(\boldsymbol{p}^{\mathrm{AB}} \right) := I^{\mathrm{A;B}} \left(\boldsymbol{p}^{\mathrm{AB}} \right) - J^{\mathrm{A;B}} \left(\boldsymbol{p}^{\mathrm{AB}} \right) \equiv 0 \tag{11}$$

so $I^{\mathrm{A;B}} \left(\boldsymbol{p}^{\mathrm{AB}} \right)$ (9) and $J^{\mathrm{A;B}} \left(\boldsymbol{p}^{\mathrm{AB}} \right)$ (10) are equal.

2.2. Stochastic map and stochastic matrix

A noisy channel is any mapping that changes the entropy (or noise, which is monotonically related) of a state in a non-decreasing way and adds noise to at least one state [11]. We are specifically interested in noisy channels that can be represented as stochastic matrices that map probability vectors representing states [8].

Under the action of a channel represented by matrix $\mathcal{E}$, the state, represented by $\boldsymbol{p}$, maps to $\mathcal{E}\boldsymbol{p}$. We require that $\mathcal{E}$ is a square matrix with nonnegative entries such that either rows or columns sum to one. Hence, the norm of the state $\boldsymbol{p}$ is unchanged by the stochastic map by stochastic matrix $\mathcal{E}$. The entropy of this state after passing through the channel is $H(\mathcal{E}\boldsymbol{p}) \geqslant H(\boldsymbol{p})$. A doubly stochastic matrix is a stochastic matrix whose rows and columns both sum to one.

In the bipartite setting, an identity mapping by Alice concomitant with a stochastic map $\mathcal{E}$ by Bob, yields the resultant bipartite state

$$\boldsymbol{p}^{\mathrm{AB}} \mapsto \mathbb{I}\boldsymbol{p}^{\mathrm{AB}}\mathcal{E} = \boldsymbol{p}^{\mathrm{AB}}\mathcal{E}. \tag{12}$$

with the trivial identity map $\mathbb{I}$ on Alice's side and the noise matrix $\mathcal{E}$ only acting on Bob's share. By the Perron–Frobenius theorem, $\mathcal{E}$ being stochastic or doubly stochastic implies that this mapping has at least one stationary vector with all entries being positive real numbers with this vector corresponding to the largest eigenvalue of the matrix representing the mapping [12].

Although we discuss discord and total-variation distance in terms of measurement described by a noisy measurement channel represented by a stochastic matrix, we focus on doubly stochastic matrices due to the abundance of mathematical properties that we can exploit for generating and understanding our results. For stochastic information theory, doubly stochastic matrices are the non-quantum analogue of quantum completely positive trace-preserving maps [13].

Birkhoff's theorem says that any doubly stochastic matrix can be written as a convex hull of permutation matrices, which is known as the Birkhoff polytope [14]. The doubly stochastic matrix thus represents a random permutation of bits in the string. Furthermore, for each strictly positive matrix A, exactly one doubly stochastic matrix T_A exists such that $T_A = DAD'$ with the diagonal matrices D and D' having positive diagonal-elements and themselves unique up to a scalar factor [15, 16]. Here we present key background information on doubly stochastic matrices needed for our study. Specifically, we define doubly stochastic matrices and connect these matrices with the Birkhoff polytope, also known as a permutahedron. A permutation σ is represented by a permutation matrix Π_σ, whose entries are all zeroes and ones such that only one instance of one appears in each row or column. For $\wp$ a length-$M!$ probability vector, a permutahedron $\mathcal{E}$ is the convex sum

$$\mathcal{E} = \sum_{\sigma \in S_{M!}} \wp \cdot \Pi_\sigma, \ \Pi_\sigma \in \mathrm{mat}_{M!}\left(\{0,1\}\right) \implies \mathcal{E} \in \mathrm{mat}_{M!}\left(\mathbb{R}\right). \tag{13}$$

Note that $\mathcal{E}\Pi_\sigma = \Pi_\sigma\mathcal{E}$.

2.3. Classical discord and distortion for a state

In appendix C, we summarise quantum discord; in this subsection, we summarise classical discord for stochastic-information states [8]. Whereas quantum discord is the discrepancy $\Delta^{\mathrm{A;B}}$ between mutual information I (9) and J (10) for quantum states, classical discord (11) is zero in the ideal case. However, discord is nonzero if, analogous to the quantum case, which optimizes over all possible measurements, classical discord also involves noisy measurements.

Alice's measurements are treated as ideal whereas Bob's measurements are treated as being noisy, described by a stochastic or a doubly stochastic mapping $\mathcal{E}$ acting on the state (12) [8]. The resultant state, after Bob's noisy measurement, is $\boldsymbol{p}^{\mathrm{AB}}\mathcal{E}$ (12). Whereas the mutual information I (9) is known, the alternative mutual information (10) is modified to include the effect of Bob's noisy measurements and is consequently described by

$$J_{\mathcal{E}}^{\mathrm{A;B}}\left(\boldsymbol{p}^{\mathrm{AB}}\right) := J^{\mathrm{A;B}}\left(\boldsymbol{p}^{\mathrm{AB}}\mathcal{E}\right) \tag{14}$$

with the subscript $\mathcal{E}$ referring to Bob's noisy channel as we always treat Alice's as ideal: $\mathbb{I}$. In other words, the conditional information inherent in inferring alternative mutual information (14) involves Bob announcing his results to Alice, and Bob's measurement apparatus is noisy: described by stochastic or doubly stochastic channels, as described in section 2.2, prior to ideal measurement and announcement by Bob. Following this definition of alternative mutual information involving noisy measurement (14), discord for stochastic information is [8]

$$\Delta_{\mathcal{E}}^{\mathrm{A;B}}\left(\boldsymbol{p}^{\mathrm{AB}}\right) := I^{\mathrm{A;B}}\left(\boldsymbol{p}^{\mathrm{AB}}\right) - J_{\mathcal{E}}^{\mathrm{A;B}}\left(\boldsymbol{p}^{\mathrm{AB}}\right) \tag{15}$$

for specified noisy channel $\mathcal{E}$, which only affects J and not I. State discord (15) is considered to be classical because states are distributions and because channels are stochastic maps; i.e. all the mathematical objects

are distributions and their mappings and hence does not require a Hilbert-space description. This $\mathcal{E}$-dependent discord is necessarily nonnegative due to the data-processing inequality [11].

By analogy with quantum discord, which minimizes over all measurement, classical discord corresponds to minimizing state discord (15) over all allowed channels $\{\mathcal{E}\}$ [8]. For shared stochastic-information states, classical discord quantifies how much stochasticity is added by a noisy measurement process. If this noise is described by a doubly stochastic channel, this noise corresponds to random permutations, following Birkhoff's theorem, corresponding to instances of measuring some messages incorrectly as other messages, with the identity permutation corresponding to measuring all integers correctly. Non-zero discord can be interpreted as quantifying stochasticity added by measurement only if entanglement is zero; otherwise a quantum model is required to describe correlations [8]. Analogous to quantum state merging operationalizing quantum discord [4], stochastic-information state merging operationalizes classical discord [8].

Channel distortion, which is used in rate-distortion theory [11], quantifies the minimum number of bits required per symbol that could be achieved over a channel so that the input signal can be approximately reconstructed at the output without exceeding a given expected distortion. Mathematically, rate-distortion theory, distortion functions quantify the cost of representing a symbol by an approximate symbol. Typical distortion functions include Hamming distortion, squared-error distortion and total-variation.

Total-variation, or Kolmogorov, distance between probability distributions $\boldsymbol{p}$ and $\boldsymbol{p}'$ (B2), namely [17],

$$\mathcal{D}\left(\boldsymbol{p}, \boldsymbol{p}'\right) := \frac{1}{2} \sum_{m \in [M]} |p_m - p'_m|. \tag{16}$$

has been widely used for extremum problems, such as controlling uncertain stochastic systems [18], approximating a family of probability distributions by a given probability distribution, maximizing or minimizing entropy subject to total-variation distance constraints, quantifying uncertainty of probability distributions by total-variation distance metric, stochastic minimax control, and in many problems of information, decision theory, and minimax theory [19], testing for scale families [20] and distortion of channels [11]. Thus, total-variation distance is well studied and valuable across a broad spectrum of applications, including for us in comparing total-variation distance to discord.

3. Approach

In this section, we begin by explaining our model, which involves three agents: Alice and Bob who share messages and Charlie who provides random messages from a distribution. After describing the model, we develop the mathematics required to analyse the effect of noisy measurement in terms of average discord and average distortion in section 3.2. Finally we elaborate on our methods for solving the expressions and what we plot in section 3.3.

3.1. Model

We describe our model for discord as a three-agent protocol involving Charlie, Alice and Bob. By describing the tasks performed by each of the three agents, we have fully described the protocol and the pertinent quantifiers of discord and distortion. Although this model is implied in a previous study of classical discord, we need to make explicit the agents of this protocol and their actions to be clear in our study of channel discord.

Charlie generates joint distributions

$$\{ \boldsymbol{p}^{\mathrm{AB}} \in \mathrm{mat}_{M^{\mathrm{A}} \times M^{\mathrm{B}}} \left(\mathbb{R}^{\geqslant 0} \right) \} \tag{17}$$

with prior $Q\left(\boldsymbol{p}^{\mathrm{AB}}\right)$ and then computes $I^{\mathrm{A;B}}\left(\boldsymbol{p}^{\mathrm{AB}}\right)$ (9) for each $\boldsymbol{p}^{\mathrm{AB}}$. For given $\boldsymbol{p}^{\mathrm{AB}}$ (1), Charlie generates a length-ς sequence of pairs of integers

$$\left\{ \left(m^{\mathrm{A}} \in [M^{\mathrm{A}}], m^{\mathrm{B}} \in [M^{\mathrm{B}}] \right) \right\} \tag{18}$$

by sampling over $\boldsymbol{p}^{\mathrm{AB}}$. In each instance, the first integer message m^{A} is sent to Alice and the second integer message sent to Bob. As A and B can experience noise in their readout, the resultant messages, $m^{\mathrm{A}'}$ and $m^{\mathrm{B}'}$, can differ from the original messages, m^{A} and m^{B}.

Alice and Bob send this noisy pair, $m^{\mathrm{A}'}$ and $m^{\mathrm{B}'}$, to Charlie. At the end of this part of the protocol, Charlie has stored the length-ς sequence $\{m^{\mathrm{A}'}, m^{\mathrm{B}'}\}$. Charlie then infers the distribution $\boldsymbol{p}^{\mathrm{A}'\mathrm{B}'}$ from these data, with this inferred state represented by $\tilde{\boldsymbol{p}}^{\mathrm{A}'\mathrm{B}'}$.

As Alice's instrument is assumed to be noiseless, $m^{\mathrm{A}'} \equiv m^{\mathrm{A}}$, Charlie's procedure is greatly simplified: he does not send Alice the message but just stores it. The message pair is thus $\{(m^{\mathrm{A}}, m^{\mathrm{B}'})\}$. Charlie's inferred

state is $\widetilde{\boldsymbol{p}^{AB}\mathcal{E}} \in \mathrm{mat}_{M^A \times M^B}\left(\mathbb{R}^{\geqslant 0}\right)$, which approximates the actual state $\boldsymbol{p}^{AB}\mathcal{E}$ after Bob's noisy measurement and equation (1). Charlie computes all permutations of the state $\widetilde{\boldsymbol{p}^{AB}\mathcal{E}}$, with each permuted state denoted $\widetilde{\boldsymbol{p}^{AB}\mathcal{E}}\Pi_\sigma$.

He thence estimates the alternative mutual information $J_\sigma^{A;B}$ (10) with the subscript σ indicating which of the permuted states $\widetilde{\boldsymbol{p}^{AB}\mathcal{E}}\Pi_\sigma$ is being considered. With these results at hand, Charlie estimates the discord for each $\widetilde{\boldsymbol{p}^{AB}\mathcal{E}}\Pi_\sigma$ and computes the minimum discord over all σ. Then he averages over results for many generated states $\boldsymbol{p}^{AB}$ to obtain an estimate for average discord for a specific channel corresponding to Bob's noisy measurement.

For each estimate $\widetilde{\boldsymbol{p}^{AB}\mathcal{E}}\Pi_\sigma$, he computes the distortion, which he quantifies by the total-variation distance $\mathcal{D}_\sigma^{A;B}\left(\widetilde{\boldsymbol{p}^{AB}\mathcal{E}}\Pi_\sigma\right)$, between $\boldsymbol{p}^{AB}$ and the estimate $\widetilde{\boldsymbol{p}^{AB}\mathcal{E}}\Pi_\sigma$. Charlie repeats this task for all permutations σ to obtain the minimum total-variation distance and then averages over all states to obtain average distortion. The mathematical description of this procedure is in section 3.2.4.

In each instance Alice receives noiseless message m^A, which is the versor $\boldsymbol{\delta}_{m^A}$, which she reads and sends the same message back to Charlie. Thus, for our mathematical analysis, Alice's role is superfluous, hence neglected in our protocol.

In each instance Bob receives message m^B. His measurement is noisy, which we describe by a doubly stochastic channel $\mathcal{E}$ described in section 2.3. This noise corresponds to permutations of the message basis so some messages are seen to be different messages incorrectly except in the case of the identity permutation $\mathbb{1}$, which corresponds to reading the message correctly.

To elucidate our model, we consider the specific case of a two-bit channel. Thus, we assume that Charlie generates a single bit each for Alice and Bob; i.e. $M^A = M^B = 2$ and $m^A, m^B \in \{0, 1\}$. Let the noisy channel be the mapping of 0 to 0, i.e. $0 \mapsto 0$, with probability $\frac{2}{3}$. Then, by the doubly stochastic property, $0 \mapsto 1$ with probability $\frac{1}{3}$ and $1 \mapsto 1$ with probability $\frac{2}{3}$ so $1 \mapsto 0$ with probability $\frac{1}{3}$ the matrix describing this mapping is doubly stochastic. Bob sends the message $m^{B'}$ obtained from his measurement back to Charlie. Note that, for the execution of this protocol, the same stochastic matrix, which describes Bob's measurement noise, is applied once per instance, i.e. each time that Bob reports each measurement outcome.

3.2. Mathematics

In this subsection, we describe in section 3.2.1 how Charlie generates states as randomly chosen joint distributions to be sent to Alice and Bob, and then we describe how random channels are generated in section 3.2.2 for Bob. In section 3.2.3 we describe mathematically how the channel is applied to the joint state. Permutations are applied to states, and both discord and distortion are minimized over all permutations, as described in section 3.2.4. Finally, in section 3.2.4, we explain how average discord and average distortion are estimated.

3.2.1. Generating joint distributions

In this subsection, we explain mathematically how Charlie generates $\boldsymbol{p}^{AB}$. Charlie constructs a prior $Q\left(\boldsymbol{p}^{AB}\right)$, which is heavily weighted over high-entropy states, meaning that state entropy (2) is close to the upper bound. By sampling this prior, Charlie obtains states with high, low and medium entropy by a linear interpolation between states drawn randomly from $Q\left(\boldsymbol{p}^{AB}\right)$ and the state represented by the identity matrix denoted $\mathbb{1}$. This linear interpolation generates a continuum of interpolated states for each of the N_{rand} states. Thus, Charlie generates random states from which he draws messages to send to Alice and Bob.

To sample from $Q\left(\boldsymbol{p}^{AB}\right)$, we generate a random $\boldsymbol{p} \in \mathrm{mat}_{M^A \times M^B}(\mathbb{R})$ with each of the $M = M^A M^B$ entries chosen uniformly from $[0, 1]$. The matrix $\boldsymbol{p}$ is then normalized by dividing all entries by their sum. This method of selecting random states leads to states that are mostly high-entropy with the maximum entropy being $\log\left(M^A M^B\right)$. For $M^A = M^B$, which corresponds to a square matrix, the maximum entropy is $\log M$ for M the square (3) of M^A.

Alternatively, Charlie can sample from a joint-state prior $Q\left(\boldsymbol{p}_{\mathrm{cp}}^{AB}\right)$ for cp states, with cp states satisfying fact 1. Conditionally pure states are permutationally equivalent to diagonal matrices $\mathrm{diag}_{\min\{M^A, M^B\}}([0, 1])$, as explained in equation (8). Thus, For $M^A = M^B$, a cp state is constructed by generating a random $\boldsymbol{p}_{\mathrm{cp}} \in \mathrm{diag}_{M^B}([0, 1])$ with each of the diagonal entries chosen uniformly from $[0, 1]$ and then normalized such that the sum of diagonal elements is one. As for general states, such cp states also tend to have high entropies.

Sampling either $Q\left(\boldsymbol{p}^{\mathrm{AB}}\right)$ or $Q\left(\boldsymbol{p}_{\mathrm{cp}}^{\mathrm{AB}}\right)$ yields a candidate state $\boldsymbol{p}_{\mathrm{cp}}^{\mathrm{AB}}$, which is then mapped to a family according to

$$\boldsymbol{p}_{\mathrm{cp}}^{\mathrm{AB}} \mapsto \frac{a\boldsymbol{p}_{\mathrm{cp}}^{\mathrm{AB}} + b\mathbb{1}}{\mathrm{norm}\ a\boldsymbol{p}_{\mathrm{cp}}^{\mathrm{AB}} + b\mathbb{1}} \quad \forall\, a \in [0,1],\, b \in [B], \tag{19}$$

where $B \gg 1$ to ensure sufficiently many medium- and low-entropy states. The channel representing Bob's noisy measurement then acts on Bob's message share, and we explain how to generate these channels in the next subsection.

3.2.2. Generating channels

In this subsection, we explain how to generate a random doubly stochastic channel for Bob. As the doubly stochastic channel is a permutahedron discussed in section 2.2, generating a random channel is equivalent to constructing the length $M^{\mathrm{B}}!$ weight, or probability, vector $\wp$, which corresponds to the probability coefficients for reverse lexicographic ordering of permutations $\{\sigma\}$, similarly to reverse lexicographical ordering of the vector of permutation matrices (B4). A given weight vector $\wp$ has associated entropy

$$0 \leqslant H(\wp) \leqslant \log M^{\mathrm{B}}!, \tag{20}$$

which is the entropy of the corresponding channel $\mathcal{E}$. Now we explain how to generate $\wp$ from a distribution $\mathcal{P}$ that is an equal weighting of a uniform prior $\mathcal{P}_{\uparrow}$, resulting in high-entropy weight vectors $\wp_{\uparrow}$ such that its entry $H(\wp_{\uparrow})$ (2) is the maximum entropy $\log M^{\mathrm{B}}!$, and another prior $\mathcal{P}_{\downarrow}$ that generates states $\wp_{\downarrow}$ with low entropy $H(\wp_{\downarrow})$ (2).

To generate $\mathcal{P}_{\uparrow}$, we first set each entry of $\wp_{\uparrow}$ be 1 and then normalize this length $M^{\mathrm{B}}!$ weight vector by dividing each element by $\|\wp\|_1$. In contrast, we generate $\mathcal{P}_{\downarrow}$ by uniformly randomly generating the first element of $\wp_{\downarrow}$, namely $\wp_{\downarrow}^1$, from the interval $[0,1]$ and replace

$$\wp_{\downarrow}^1 \leftarrow \frac{\wp_{\downarrow}^1}{M^{\mathrm{B}}!}. \tag{21}$$

The next element of the weight vector, namely, $\wp_{\downarrow}^2$, is drawn uniformly from the interval

$$\left[1 - \wp_{\downarrow}^1, 1\right],$$

and we continue according to the rule that $\wp_{\ell}$ is drawn uniformly from

$$\left[1 - \sum_{j=1}^{\ell-1} \wp_{\downarrow}^j, 1\right].$$

After randomly generating all these $M^{\mathrm{B}}!$ elements, we normalise this weight vector to obtain $\wp_{\downarrow}$.

Now that we have generated an instance of a low-entropy weight vector $\wp_{\downarrow}$ and a high-entropy weight vector $\wp_{\uparrow}$ we generate numerous vectors in between the two by linear interpolation, thereby sampling the continuous set

$$\wp\left(\wp_{\downarrow}, \wp_{\uparrow}, a\right) := \frac{a\wp_{\downarrow} + (1-a)\wp_{\uparrow}}{\|a\wp_{\downarrow} + (1-a)\wp_{\uparrow}\|} \quad \forall\, a \in [0,1], \tag{22}$$

This interpolation (22) yields medium-entropy channels to round out the sampling.

As we have generated many random instances of $\wp$ (22), we can construct corresponding descriptions of doubly stochastic channels. The elements of $\wp$ are coefficients of reverse lexicographically ordered permutation matrices, and this weighted sum is then the permutahedron that describes random doubly stochastic channel $\mathcal{E}$ with entropy given by the entropy of its representative weight vector. Mathematically, the matrix description of the channel is

$$\mathcal{E} = \wp \cdot \boldsymbol{\Pi}, \quad \boldsymbol{\Pi} := \left(\Pi_{\sigma}; \sigma \in S_{M^{\mathrm{B}}!}\right), \tag{23}$$

for $\wp$ and $\boldsymbol{\Pi}$ length-$M^{\mathrm{B}}!$ vectors of real numbers and permutation matrices (B4), respectively, and $\sigma \in S_{M^{\mathrm{B}}!}$ is drawn in reverse lexicographical order.

3.2.3. Applying the channel to the joint distributions

To describe Bob's noisy measurement mathematically, we apply the generated random channel, discussed in section 3.2.2, to Bob's share of the entire message $\boldsymbol{p}^{\mathrm{AB}}$ sent by Charlie. First suppose that Alice and Bob each have noisy measurements described by channels $\mathcal{E}^{\mathrm{A}}$ and $\mathcal{E}^{\mathrm{B}}$, respectively. Then the state sent back to Charlie from Alice and Bob is

$$\mathcal{E}^{\mathrm{A}}\boldsymbol{p}^{\mathrm{AB}}\mathcal{E}^{\mathrm{B}} \in \mathrm{mat}_{M^{\mathrm{A}}\times M^{\mathrm{B}}}(\mathbb{C}), \boldsymbol{p}^{\mathrm{AB}} \in \mathrm{mat}_{M^{\mathrm{A}}\times M^{\mathrm{B}}}(\mathbb{C}), \mathcal{E}^{\mathrm{A}} \in \mathrm{mat}_{M^{\mathrm{A}}\times M^{\mathrm{A}}}(\mathbb{C}), \mathcal{E}^{\mathrm{B}} \in \mathrm{mat}_{M^{\mathrm{B}}\times M^{\mathrm{B}}}(\mathbb{C}), \tag{24}$$

corresponding to one application of a noisy channel for each of Alice's and Bob's noisy measurements but noiseless transmission back to Charlie. As Alice's measurement is noiseless, we assign

$$\mathcal{E}^{\mathrm{A}} \equiv \mathbb{1}, \; \mathcal{E}^{\mathrm{B}} =: \mathcal{E}. \tag{25}$$

Remark 1. In section 3.1, Charlie estimates $\boldsymbol{p}^{\mathrm{AB}}\mathcal{E}$ to be $\widetilde{\boldsymbol{p}^{\mathrm{AB}}\mathcal{E}}$ by repeated sampling but here, in the mathematical description, we work with exact descriptions of the state (24).

3.2.4. Estimating discord and distortion

After Charlie receives $\boldsymbol{p}^{\mathrm{AB}}\mathcal{E}$ (24), he computes all permutations of this matrix. Charlie generates each of the $M^{\mathrm{B}}!$ instances of $M^{\mathrm{B}} \times M^{\mathrm{B}}$ permutation matrices Π_σ (B3), for each σ drawn from permutation group $S_{M^{\mathrm{B}}}$ in reverse lexicographical order. For each instance σ, Charlie obtains the permuted state by multiplying $\boldsymbol{p}^{\mathrm{AB}}\mathcal{E}$ by each permutation matrix Π_σ.

Here we redefine average discord to quantify channel distortion, in contrast to the earlier definition of classical discord in terms of fluctuations for a state [8], and we define average distortion. This redefinition of average discord, in terms of channels, then allows us to show numerically the monotonic relationship between classical (channel) discord and channel distortion. Average discord is obtained first by minimizing discord (15) over all state permutations and then by averaging over all states according to the prior $Q\left(\boldsymbol{p}^{\mathrm{AB}}\right)$ in section 3.2.1. Similarly, average distortion is obtained by averaging state-dependent distortion (16) over the prior of states $Q\left(\boldsymbol{p}^{\mathrm{AB}}\right)$.

For average discord, we first extend state-dependent discord (15) to be the minimized state-dependent discord over all permutations of the state

$$\min_\sigma \Delta_{\mathcal{E}}^{\mathrm{A;B}}\left(\boldsymbol{p}^{\mathrm{AB}}\Pi_\sigma\right). \tag{26}$$

Averaging over the prior $Q\left(\boldsymbol{p}^{\mathrm{AB}}\right)$ yields channel discord

$$\Delta^{\mathrm{A;B}}\left(\mathcal{E}\right) := \int \mathrm{d}Q\left(\boldsymbol{p}^{\mathrm{AB}}\right) \min_\sigma \Delta_{\mathcal{E}}^{\mathrm{A;B}}\left(\boldsymbol{p}^{\mathrm{AB}}\Pi_\sigma\right), \tag{27}$$

which quantifies the discord due to noisy measurement in a state-independent but of course prior-dependent way. Average discord is obtained by sampling the integral (27) to obtain the estimate $\tilde{\Delta}^{\mathrm{A;B}}(\mathcal{E})$.

Similarly, we extend the definition of distortion $\mathcal{D}$ (16) first by minimizing over permutations and then by averaging over the prior $Q\left(\boldsymbol{p}^{\mathrm{AB}}\right)$. The state- and channel-dependent distortion is

$$\mathcal{D}_{\mathcal{E}}^{\mathrm{A;B}}\left(\boldsymbol{p}^{\mathrm{AB}}\right) := \mathcal{D}\left(\boldsymbol{p}^{\mathrm{AB}}\mathcal{E}, \boldsymbol{p}^{\mathrm{AB}}\right) \tag{28}$$

and its minimization over all permutations is

$$\min_\sigma \mathcal{D}_{\mathcal{E}}^{\mathrm{A;B}}\left(\boldsymbol{p}^{\mathrm{AB}}\Pi_\sigma\right). \tag{29}$$

Then, analogous to average discord (27), we integrate to obtain channel distortion

$$\mathcal{D}^{\mathrm{A;B}}\left(\mathcal{E}\right) := \int \mathrm{d}Q\left(\boldsymbol{p}^{\mathrm{AB}}\right) \min_\sigma \mathcal{D}_{\mathcal{E}}^{\mathrm{A;B}}\left(\boldsymbol{p}^{\mathrm{AB}}\Pi_\sigma\right) \tag{30}$$

for a given channel. Average discord is obtained by sampling the integral (27) to obtain the estimate $\tilde{\mathcal{D}}^{\mathrm{A;B}}(\mathcal{E})$.

3.2.5. Example: two-bit channel

We now elucidate our model by considering the special case where Charlie generates a single bit for Alice and a single bit for Bob: $M^{\mathrm{A}} = M^{\mathrm{B}} = 2$. Thus, equation (1) implies that the joint probability of messages shared between Alice and Bob is specified by the matrix

$$\boldsymbol{p}^{\mathrm{AB}} = \begin{pmatrix} p_{00} & p_{01} \\ p_{10} & p_{11} \end{pmatrix}, \quad p_{ij} \geqslant 0, \; p_{00} + p_{01} + p_{10} + p_{11} = 1. \tag{31}$$

The two-bit state (31) is subjected to a distortion channel whose form is dictated by equation (23), which implies

$$\mathcal{E} = \begin{pmatrix} 1-\mu \\ \mu \end{pmatrix} \cdot \begin{pmatrix} \mathbb{1} \\ X \end{pmatrix} = (1-\mu)\mathbb{1} + \mu X = \begin{pmatrix} 1-\mu & \mu \\ \mu & 1-\mu \end{pmatrix}, \quad 0 \leqslant \mu \leqslant 1. \tag{32}$$

Consequently, the entropy is

$$H = -(1-\mu)\log(1-\mu) - \mu \log \mu \tag{33}$$

according to equation (20).

Applying the channel to the two-bit state yields

$$\boldsymbol{p}^{\mathrm{AB}}\mathcal{E} = \begin{pmatrix} p_{00} & p_{01} \\ p_{10} & p_{11} \end{pmatrix} \cdot \begin{pmatrix} 1-\mu & \mu \\ \mu & 1-\mu \end{pmatrix} = \begin{pmatrix} (1-\mu)p_{00}+\mu p_{01} & \mu p_{00}+(1-\mu)p_{01} \\ (1-\mu)p_{10}+\mu p_{11} & \mu p_{10}+(1-\mu)p_{11} \end{pmatrix}. \tag{34}$$

We then compute the state discord per equation (26), and we could perform a similar calculation for state distortion per equation (16). The explicit expression for discord (26) is

$$
\begin{aligned}
&\big((1-\mu)p_{00}+\mu p_{01}+(1-\mu)p_{10}+\mu p_{11}\big)\log\big((1-\mu)p_{00}+\mu p_{01}+(1-\mu)p_{10}+\mu p_{11}\big) \\
&+ \big(\mu p_{00}+(1-\mu)p_{01}+\mu p_{10}+(1-\mu)p_{11}\big)\log\big(\mu p_{00}+(1-\mu)p_{01}+\mu p_{10}+(1-\mu)p_{11}\big) \\
&- \big((1-\mu)p_{00}+\mu p_{01}\big)\log\big((1-\mu)p_{00}+\mu p_{01}\big) - \big(\mu p_{00}+(1-\mu)p_{01}\big)\log\big(\mu p_{00}+(1-\mu)p_{01}\big) \\
&- \big((1-\mu)p_{10}+\mu p_{11}\big)\log\big((1-\mu)p_{10}+\mu p_{11}\big) - \big(\mu p_{10}+(1-\mu)p_{11}\big)\log\big(\mu p_{10}+(1-\mu)p_{11}\big) \\
&- \big(p_{00}+p_{10}\big)\log\big(p_{00}+p_{10}\big) - \big(p_{01}+p_{11}\big)\log\big(p_{01}+p_{11}\big) \\
&+ p_{00}\log\big(p_{00}\big) + p_{01}\log\big(p_{01}\big) + p_{10}\log\big(p_{10}\big) + p_{11}\log\big(p_{11}\big)
\end{aligned} \tag{35}
$$

as a function of state $\boldsymbol{p}^{\mathrm{AB}}$ and channel parameter μ.

If we could integrate discord and distortion over all possible joint states, i.e. over the formal variables $p_{00}, p_{01}, p_{10}, p_{l1}$ with respect to an appropriate prior distribution, then we could obtain channel discord and distortion via equations (27) and (30), respectively. Note that the resulting expressions depend only on μ. On the other hand, if discord is monotonic with respect to entropy for all states $\{\boldsymbol{p}^{\mathrm{AB}}\}$, i.e. for all points on the tetrahedron, then discord would be monotonic with respect to this average.

3.3. Methods

In this subsection we discuss how we average discord and average distortion for many randomly chosen channels, with each of these channels corresponding to a different noisy measurement process implemented by Bob. First, we explain in section 3.3.1 how many instances of states we generate and how to create those instances. Then, in section 3.3.2, we explain how we generate all permutation matrices and thence all permuted states. Then we explain how we generate random channels $\{\mathcal{E}\}$ numerically in section 3.3.3 and how many such channels. Finally, in section 3.3.4 we explain how we calculate the states sent back to Charlie, how to use these states to compute average discord and average distortion, and then study their relations.

3.3.1. Numerically generating states

We begin by generating joint states from either prior $Q\left(\boldsymbol{p}^{\mathrm{AB}}\right)$ or $Q\left(\boldsymbol{p}^{\mathrm{AB}}_{\mathrm{cp}}\right)$ as described in section 3.2.1. In practice, we generate random states as follow. First we choose $B = 99$ in equation (19) as we have discovered empirically that this value of B yields a good spread of low-, medium- and high-entropy states. Then we step through values of linear interpolation parameter a by increasing in step sizes that grow quadratically: the coefficient a of $\mathbb{1}$ increases in steps of $([\varpi-1]0.0101)^2$ for $\varpi \in [100]$.

For each choice of a and fixing $b = (1-a)B$, we choose a new random instance of $\boldsymbol{p}^{\mathrm{AB}}$ according to the random-matrix construction method described in section 3.2.1. We insert a, b and $\boldsymbol{p}^{\mathrm{AB}}$ into equation (19) to obtain one instance of a state for calculating average discord and average distortion.

3.3.2. Numerically generating all permutation matrices

Our analysis considers different messages sizes M such that $M^{\mathrm{A}} = M^{\mathrm{B}} = \sqrt{M}$, where we have chosen to study only cases for which Alice's and Bob's messages are the same size. For each message size all permutation matrices are generated according to the mathematical description in section 2. We need to construct $M^{\mathrm{B}}!$ permutation matrices, each of size $M^{\mathrm{B}} \times M^{\mathrm{B}}$, representing permutations $\sigma \in S_{M^{\mathrm{B}}}$. Each row of the matrix corresponds to a permuted version of the length-M^{B} initial row vector, which is the vector comprising entries that are the column numbers themselves. These columns correspond to reverse lexicographical ordering of the permutations, with reverse lexicographical ordering discussed in section 2.

We apply perms from MATLAB$^{\circledR}$ to the initial vector, and the output is the $M^{\mathrm{B}}! \times M^{\mathrm{B}}$ matrix Θ, whose elements are column indices for nonzero entries of permutation matrices Π_σ. For each of the $M^{\mathrm{B}}!$ rows, labelled by σ in reverse lexicographical order, we construct the corresponding $M^{\mathrm{B}} \times M^{\mathrm{B}}$ permutation matrix Π_σ, whose entries are all zeroes and ones such that only one instance of one appears in each row or column as described in and around equation (B3).

Specifically, to generate Π_σ, we pick row σ from the matrix Θ, denoted as a vector Θ_σ. The value of the entry in the first column of Θ_σ indicates which element of the first row of Π_σ is one, with the rest of the elements in that row being zero. Then we proceed to the second entry of the vector Θ_σ, and its value indicates which element of the second row of Π_σ is one. We continue for all M^B rows of Π_σ and then repeat for all $\sigma \in S_{M^B}$. In this way we have constructed the full set of permutation matrices for message m^B.

3.3.3. Numerically generating random channels

We generate 6000 doubly stochastic channels, each channel represented by some weight vector $\wp\left(\wp_\downarrow, a\right)$ (22), regardless of message size. Construction of each of these weight vectors proceeds according to the mathematical description in section 3.2.2. We choose to generate 6000 instances of random channels because we allow for 100 equally spaced values of a in equations (22) and 60 randomly chosen $\wp_\downarrow$ for each a.

A doubly stochastic channel is a permutahedron, and we have generated the set of all permutation matrices $\Pi_\sigma \in S_{M^B}$ in section 3.3.2. Specifically, we generate random weight vectors according to equation (22), thereby yielding low-, medium- and high-entropy (20) weight vectors. The resultant set of 6000 randomly generated weight vectors faithfully represents 6000 randomly generated channels, and our interpolation (22) ensures good sampling of a wide range of channel entropies. In addition, we manually add the noiseless channel $\mathbb{1}$ to our simulations to include the instance of zero average discord and zero average distortion.

3.3.4. Relating average discord to average distortion

Now that state $\boldsymbol{p}^{AB}$ is generated numerically according to the procedure described in section 3.3.1, for both $Q\left(\boldsymbol{p}^{AB}\right)$ corresponding to random initial states and $Q\left(\boldsymbol{p}^{AB}_{cp}\right)$ for random cp states, we calculate the corresponding $\boldsymbol{p}^{AB}\mathcal{E}$ for each permutation σ. These permuted returned states $\{\boldsymbol{p}^{AB}\mathcal{E}\Pi_\sigma\}$ are used to calculate both average discord and average distortion. Mathematical expressions for average discord and average distortion are given in section 3.2.4.

We begin with how we calculate average discord $\Delta^{A;B}(\mathcal{E})$ (27). First we calculate each $\Delta^{A;B}_{\mathcal{E}}\left(\boldsymbol{p}^{AB}\Pi_\sigma\right)$ for each $\boldsymbol{p}^{AB}$ and then minimize over all $\sigma \in S_{M^B}$ according to equation (26), thereby obtaining the minimized $\Delta^{A;B}(\mathcal{E})$. The next step is to average over all $\boldsymbol{p}^{AB}$. As explained in section 3.3.1, we generate a random state $\boldsymbol{p}^{AB}$ for each choice of linear interpolation parameter a (22), which suffices to sample the integral (27) fairly and thus obtain a good estimate $\tilde{\Delta}^{A;B}(\mathcal{E})$ of the actual average discord $\Delta^{A;B}(\mathcal{E})$.

The procedure for calculating average distortion $\mathcal{D}^{A;B}(\mathcal{E})$ (30) is similar. First we calculate each $\mathcal{D}^{A;B}_{\mathcal{E}}\left(\boldsymbol{p}^{AB}\Pi_\sigma\right)$ for each $\boldsymbol{p}^{AB}$ and then minimize over all $\sigma \in S_{M^B}$ according to equation (29), thereby obtaining the minimized $\mathcal{D}^{A;B}(\mathcal{E})$. The next step is to average over all $\boldsymbol{p}^{AB}$. As explained in section 3.3.1, we generate a random state $\boldsymbol{p}^{AB}$ for each choice of linear interpolation parameter a (22), which suffices to sample the integral (30) fairly and thus obtain a good estimate $\tilde{\mathcal{D}}^{A;B}(\mathcal{E})$ of the actual average distortion $\mathcal{D}^{A;B}(\mathcal{E})$.
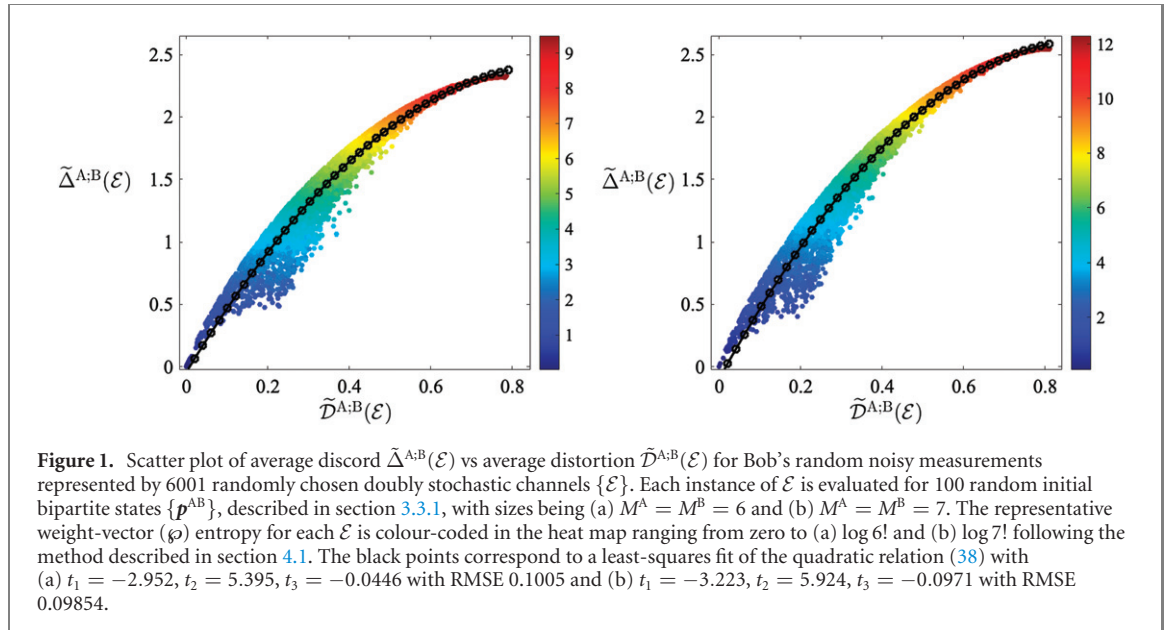
Finally, we relate estimated average discord to estimated average distortion by plotting $\tilde{\Delta}^{A;B}(\mathcal{E})$ against $\tilde{\mathcal{D}}^{A;B}(\mathcal{E})$. Specifically, we choose sufficiently large yet tractable message sizes, namely, $M^B \in \{6, 7\}$ and make distinct plots for each M^B. For each randomly chosen channel, the resultant single point on the graph corresponding to $\tilde{\Delta}^{A;B}(\mathcal{E})$ and $\tilde{\mathcal{D}}^{A;B}(\mathcal{E})$ is marked, and we thereby obtain a scatter plot. We create plots for two cases, random initial states and random initial cp states, and compare these two cases.

3.3.5. Example: two-bit channel

We explicitly analyze the relationship between average channel discord (27) and channel distortion (30) in the case of a two-bit channel as described in section 3.2.5. By doing so, we establish the monotonicity of discord as a function of channel entropy. Our approach follows.

(a) We first recognize that the set of possible channels $\mathcal{E}$ can be parameterized by a single number $0 \leqslant \mu \leqslant 1$ per equation (32). Hence, the channel discord can be expressed as a function of μ. Furthermore, the channel discord is defined to be an integral with respect to some measure over an integrand that itself depends on μ. If we prove that the integrand is monotonic in μ over some interval, we will also have proved that the integral is monotonic in μ over the same interval.

(b) We can therefore assess the monotonicity of the channel discord (26), which we write here simply as Δ, as a function of H (33) by examining the derivative of Δ with respect to H and applying the chain rule:

$$\frac{d\Delta}{dH} = \frac{\frac{d\Delta}{d\mu}}{\frac{dH}{d\mu}} = \left(\log(1-\mu) - \log\mu\right)^{-1} \cdot \frac{d\Delta}{d\mu}. \tag{36}$$

Figure 1. Scatter plot of average discord $\tilde{\Delta}^{A;B}(\mathcal{E})$ vs average distortion $\tilde{\mathcal{D}}^{A;B}(\mathcal{E})$ for Bob's random noisy measurements represented by 6001 randomly chosen doubly stochastic channels $\{\mathcal{E}\}$. Each instance of $\mathcal{E}$ is evaluated for 100 random initial bipartite states $\{p^{AB}\}$, described in section 3.3.1, with sizes being (a) $M^A = M^B = 6$ and (b) $M^A = M^B = 7$. The representative weight-vector $(\wp)$ entropy for each $\mathcal{E}$ is colour-coded in the heat map ranging from zero to (a) $\log 6!$ and (b) $\log 7!$ following the method described in section 4.1. The black points correspond to a least-squares fit of the quadratic relation (38) with (a) $t_1 = -2.952$, $t_2 = 5.395$, $t_3 = -0.0446$ with RMSE 0.1005 and (b) $t_1 = -3.223$, $t_2 = 5.924$, $t_3 = -0.0971$ with RMSE 0.09854.

This expression is well-defined except when $\mu = \frac{1}{2}$, but this point corresponds to the maximum value of H and so is not relevant for monotonicity arguments. Thus, monotonicity of Δ with respect to H can be proven by demonstrating monotonicity on the intervals $0 < \mu < \frac{1}{2}$ and $\frac{1}{2} < \mu < 1$, which can be proven by demonstrating monotonicity of the integrand as described in the previous point.

The above two points imply that we can prove the monotonicity of channel discord as a function of channel entropy by showing that the state discord is a monotonically increasing (decreasing) function of μ for $0 < \mu < \frac{1}{2}$ $(\frac{1}{2} < \mu < 1)$. We prove that this is so in section 4.4.

4. Results

In this section we present our results, which are numerical in nature. We choose tractable message-size values, namely $M^B \in \{6, 7\}$, to study the relation between average discord $\tilde{\Delta}^{A;B}(\mathcal{E})$ and average distortion $\tilde{\mathcal{D}}^{A;B}(\mathcal{E})$. Specifically, we plot $\tilde{\Delta}^{A;B}(\mathcal{E})$ vs $\tilde{\mathcal{D}}^{A;B}(\mathcal{E})$ for many generated channels as described in section 3.3.3 averaged over randomly generated states. We plots for two cases: randomly generated joint distributions in section 4.1 and randomly generated cp states in section 4.2. Then we explain our best-fit quadratic relation between $\tilde{\Delta}^{A;B}(\mathcal{E})$ and $\tilde{\mathcal{D}}^{A;B}(\mathcal{E})$ in section 4.3. Our methods for generating these plots are described in section 3.3.

4.1. Plots for randomly generated joint states

In figure 1, we have plotted estimated average discord $\tilde{\Delta}^{A;B}(\mathcal{E})$ (27) and estimated average distortion $\tilde{\mathcal{D}}^{A;B}(\mathcal{E})$ (30) for two cases of total message length $M \in \{6, 7\}$ as discussed in section 3.3.4. This scatter plot represents 6001 instances of randomly chosen channels for Bob and randomly chosen initial states by Charlie, and the points are colour-coded by the Shannon entropy of the weight vector representing the channel (20).
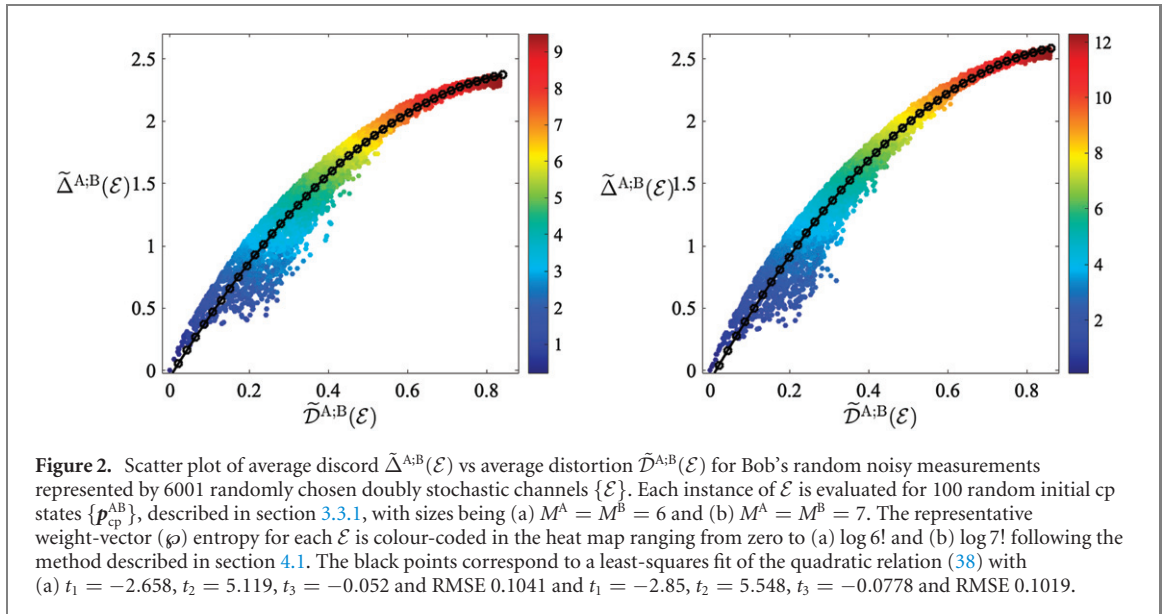
The origin of the plot corresponds to zero average discord and zero average distortion and arises for Bob's measurement being noiseless, i.e. for a zero-entropy weight vector $\wp$. We observe a monotonic trend of increasing average discord with respect to increasing average distortion. This monotonicity inference is reinforced in section 4.3 where we explain the best-fit curve, which is certainly monotonic. Furthermore, based on the colour-coded heat map in figure 1, we see monotonically increasing of all three: average discord, average distortion and channel entropy.

The scatter plot shows more features. The highest point of the curve has the maximum allowed entropy $\log M^B!$ (20) for the channel. For the two chosen messages sizes, the maximum entropies are

$$\log 6! = 9.492, \qquad \log 7! = 12.299, \tag{37}$$

respectively. Also the scatter plot is narrow for low- and high-entropy cases of channels and wide for medium-low choices of channel entropy.

We have provided figures 1(a) and (b) showing scatter plots for $M^B = 6$ and $M^B = 7$, respectively. The two scatter plots are similar. The differences are that the maximum entropy for the second scatter plot is

Figure 2. Scatter plot of average discord $\tilde{\Delta}^{A;B}(\mathcal{E})$ vs average distortion $\tilde{\mathcal{D}}^{A;B}(\mathcal{E})$ for Bob's random noisy measurements represented by 6001 randomly chosen doubly stochastic channels $\{\mathcal{E}\}$. Each instance of $\mathcal{E}$ is evaluated for 100 random initial cp states $\{\boldsymbol{p}_{cp}^{AB}\}$, described in section 3.3.1, with sizes being (a) $M^A = M^B = 6$ and (b) $M^A = M^B = 7$. The representative weight-vector ($\wp$) entropy for each $\mathcal{E}$ is colour-coded in the heat map ranging from zero to (a) log 6! and (b) log 7! following the method described in section 4.1. The black points correspond to a least-squares fit of the quadratic relation (38) with (a) $t_1 = -2.658$, $t_2 = 5.119$, $t_3 = -0.052$ and RMSE 0.1041 and $t_1 = -2.85$, $t_2 = 5.548$, $t_3 = -0.0778$ and RMSE 0.1019.

higher due to the larger message size, with the increase in maximum entropy given by the ratio of the numbers in (37). Both estimated average discord and estimated average distortion are increased slightly for increased message size.

4.2. Plots for randomly generated conditionally pure states

In this subsection we obtain scatter plots of estimated average discord vs estimated average discord for the case that Charlie generates random cp states (8) instead of random joint states (1) as was done in section 4.1. Other than using cp states here, we follow exactly the same procedure used to obtain figure 1. The purpose of this subsection is to verify or refute that the two cases of initial random joint distributions vs initial random cp states show the same or different features.

The scatter plot for estimated average discord vs estimated average distortion is shown in figure 2 for initial cp states. Similarly to figure 1, the scatter shows monotonically increasing average discord with respect to average distortion, monotonicity of both with respect to channel entropy represented by the heat map, and wider scatter for medium entropy compared to narrow scatter width for low and high entropy. The differences are only with respect to randomness of generated channels in the plots, suggesting that figures 1 and 2 are identical up to random-sampling variability.

4.3. Quadratic best fit to the plots

Now we explain how we fit a curve to the scatter plots of figures 1 and 2. Our numerical results fit well with a quadratic curve

$$\tilde{\Delta}^{A;B}(\mathcal{E}) = t_1 \left(\tilde{\mathcal{D}}^{A;B}(\mathcal{E}) \right)^2 + t_2 \tilde{\mathcal{D}}^{A;B}(\mathcal{E}) + t_3 \tag{38}$$

with $\{t_i\}$ chosen differently for each plot to minimize root-mean-square error (RMSE). In all four cases, RMSE ~ 0.1, which indicates a good fit as the RMSE is much smaller than the range of $\tilde{\Delta}^{A;B}(\mathcal{E})$.

4.4. Monotonicity of two-bit channel discord as function of channel entropy

Now we show that the discord of a two-bit noise channel varies monotonically with the entropy of that channel. As we explain in section 3.3.5, it suffices to show that the *state* discord Δ for an arbitrary two-bit state (31) varies monotonically in the parameter μ that specifies the channel (32). This monotonicity relation can in turn be proven by showing $\frac{d\Delta}{d\mu} \lessgtr 0$ for $\mu \gtrless \frac{1}{2}$ by exploiting equation (36). For ease of calculation, we substitute $\mu \mapsto \frac{1+\alpha}{2}$. Thus, we accomplish our aim of proving $\frac{d\Delta}{d\mu} \lessgtr 0$ when $\mu \gtrless \frac{1}{2}$ by instead proving that

$$\mathrm{sgn}\left(\frac{d\Delta}{d\alpha} \right) = -\,\mathrm{sgn}\left(\alpha \right) \tag{39}$$

for $-1 < \alpha < 1$.

First we derive an expression for $\frac{d\Delta}{d\alpha}$. A tedious-but-straightforward calculation starting from equation (35) reveals that

$$
\frac{d\Delta}{d\alpha} = (p_{00} - p_{01} + p_{10} - p_{11}) \log \left(\frac{1 + \alpha (p_{00} - p_{01} + p_{10} - p_{11})}{1 - \alpha (p_{00} - p_{01} + p_{10} - p_{11})} \right)
$$

$$
- (p_{00} - p_{01}) \log \left(\frac{1 + \alpha \left(\frac{p_{00} - p_{01}}{p_{00} + p_{01}} \right)}{1 - \alpha \left(\frac{p_{00} - p_{01}}{p_{00} + p_{01}} \right)} \right) - (p_{10} - p_{11}) \log \left(\frac{1 + \alpha \left(\frac{p_{10} - p_{11}}{p_{10} + p_{11}} \right)}{1 - \alpha \left(\frac{p_{10} - p_{11}}{p_{10} + p_{11}} \right)} \right). \tag{40}
$$

If we assign

$$
\begin{aligned}
\gamma_0 &\leftarrow \frac{p_{00} - p_{01}}{p_{00} + p_{01}}, \\
\gamma_1 &\leftarrow \frac{p_{10} - p_{11}}{p_{10} + p_{11}}, \\
w_0 &\leftarrow p_{00} + p_{01}, \\
w_1 &\leftarrow p_{10} + p_{11}, \\
f_\alpha(x) &\leftarrow x \log \left(\frac{1 + \alpha x}{1 - \alpha x} \right),
\end{aligned} \tag{41}
$$

we can rewrite equation (40) as

$$
\frac{d\Delta}{d\alpha} = f_\alpha (w_0 \gamma_0 + w_1 \gamma_1) - w_0 f_\alpha(\gamma_0) - w_1 f_\alpha(\gamma_1). \tag{42}
$$

Noting that $-1 < \gamma_{0,1} < 1$ and $0 < w_{0,1} < 1$ when none of $p_{00}, p_{01}, p_{10}, p_{11}$ are equal to zero, we have $\frac{d\Delta}{d\alpha}$ is positive/negative/zero when f_α is concave/convex/linear on the interval $(-1, 1)$. If one or more of $p_{00}, p_{01}, p_{10}, p_{11}$ are zero, we can have the logarithm argument become of the 0/0 type, in which case we apply l'Hôpital's rule to obtain the limit.

We now show that f_α is concave/linear/convex depending on whether $\operatorname{sgn} \alpha = -1, 0, 1$. It is easy to see that $f_0 \equiv 0$, which is a linear function, so we focus instead on the case $\alpha \neq 0$. In that case, the identity $\log \left(\frac{1+y}{1-y} \right) = \operatorname{arctanh} y$ shows $f_\alpha(x) = \frac{1}{2\alpha} \times g(\alpha x)$, where $g(y) = y \operatorname{arctanh} y$. Hence we need only show that $g(y)$ is convex for $-1 < y < 1$. This convexity can be seen directly from the Maclaurin series for arctanh,

$$
g(y) = \sum_{k=1}^{\infty} \frac{y^{2k+2}}{2k + 1}, \tag{43}
$$

which is a sum of convex functions and hence is convex. Thus, f_α is convex/concave depending on the sign of α. Hence, $\frac{d\Delta}{d\alpha}$ has the appropriate sign; hence, $\frac{d\Delta}{dH}$ is monotonic as required.

5. Discussion

We have developed a full classical (i.e. non-quantum) theory for channel discord, which establishes the meaning of nonzero discord in the context of stochastic information theory. Although classical discord and stochastic information theory were introduced in 2015 [8], key notions were sketched rather than fully developed. Here we have given a detailed theory of non-quantum discord all the way from the context of a three-party protocol with noisy measurement to building in Hadamard notation for making expressions clear and elegant to the unprecedented connection between average channel discord, average channel distortion and entropy of the noisy measurement (noisy channel), including monotonic relations between them, which thereby show that channel discord, in a classical setting, is a form of channel distortion arising due to one party's noisy measurement.

Scatter plots for average discord vs average distortion in figures 1 and 2 show this monotonic relation between average discord and average distortion and, through the heat maps, also the monotonicity between average distortion and channel entropy. These results are purely numerical but show a simple quadratic relationship for two choices of message sizes. A general mathematical relation connecting average discord to average distortion is beyond the scope of this work but we derive mathematical relations for the two-bit case both to illustrate how analytical results can be obtained and also to lend support to our conjectures based on numerical analysis; our analysis has focused on developing the protocol, making clear the problem,

defining appropriate quantities and tackling numerically. Mathematically proving the general case is challenging so we instead solve the special two-bit case, i.e. the case that each of Alice and Bob hold one bit and Bob's measurement is noisy, and there we prove that channel discord is a monotonic function of channel entropy. Thus, the numerical results are backed up by a closed-expression analysis, and, furthermore, this analysis points the way to general proofs, likely using the Hadamard calculus elaborated in appendix A.

The plots in section 4 display a high level of scatter for medium-entropy cases and much less scatter for low- and high-entropy cases. Although the spread is large, monotonicity and quadratic scaling is clearly evident in these plots and the root-mean-squared error (RMSE) of each plot hovering around 0.1 is testament to the quality of the quadratic fit and hence the inference of monotonicity.

Our analysis has focused only on noise represented by doubly stochastic maps, which correspond to permutahedrons. In other words, we have concentrated on noise that would arise from random permutations of classical message measurements correspondingly measuring some messages as different messages incorrectly. In the spirit of quantum discord, our average discord and average distortion calculations are built on minimizing over all such permutations. Future work should involve generalization from doubly stochastic to stochastic maps; in the quantum context, this generalization would be akin to extending from completely positive trace-preserving mappings to completely positive mappings.

6. Conclusions

Discord has emerged as one of the most significant quantum resources [7] but not without controversy [8, 21]. Separating quantum and non-quantum aspects of discord is vital to determine quantum resourcefulness and otherwise for discord. The connection between state discord and entanglement is known, but discord for channels has been unexplored under the treatment of noisy measurement as being manifested by a noisy channel. Here we establish and elucidate connections between classical channel discord and channel distortion and entropy. To this end, we have developed a protocol, mathematical framework and numerical analysis of average channel discord, with averaging being over random shared message states (for random initial joint distributions and, to check consistency, over cp states) and over random doubly stochastic channels representing noisy measurement by one party. Note that we have defined classical channel discord to quantify channel 'fluctuations' (i.e. 'noise'), in contrast to earlier work on classical discord, which quantifies fluctuations for a state [8]. Our notion of classical channel discord then leads to our numerical demonstration of monotonicity between classical (channel) discord and channel distortion. Thus, our results show numerically that average discord, in the non-quantum setting, is equivalent to average distortion of a channel with channel distortion based on total-variation distance. Furthermore, we show numerically that this distortion measure is monotonic in channel entropy, which builds confidence that total-variation distance is a reasonable way to quantify distortion.

Akin to quantum discord, we have incorporated minimization of average discord and average distortion over all permutations, which permutations referring to permuting messages. The identity permutation corresponds to reading each message correctly and other permutations cause some messages to be read as other messages. Given that the noisy measurement is modelled as a doubly stochastic channel, which is a permutahedron, the idea of minimizing over all permutations is to identify which permutation minimizes channel discord and minimizes channel distortion averaged over all states. This minimization is key to connecting our notion of classical distortion to the quantum version.

We have created a full framework for studying the connection between average discord and average distortion for a noisy channel and have shown numerically a monotonic relation between the two. This monotonic relation is satisfying as we can now regard discord, in the classical setting, as an alternative measure of channel distortion, manifested as noisy readout by one party. We augment our numerical analysis of channel discord vs channel distortion by studying the two-bit example analytically and provide results for discord corresponding to certain special but important two-bit states. We can see that analytic methods are challenging, even in the two-bit case, but our analytical study shows a path forward for general analytical work, which benefits from using the Hadamard calculus we discuss in appendix A.

Acknowledgments

Arshed in the early stages of this project. The authors acknowledge University of Sydney for providing high-performance computing used to obtain early informative results.

Data availability statement

The data that support the findings of this study are available upon reasonable request from the authors.

Appendix A. Hadamard calculus

We review the convenient Hadamard notation [22], and introduce what we call Hadamard calculus, which is novel but useful for studying stochastic information. Specifically, we explain the Hadamard product and summing over elements of a Hadamard product, the Hadamard logarithm and the entropy in this notation. Also we introduce vector calculus based on Hadamard notation principles.

The Hadamard product between two rank-t tensors

$$\boldsymbol{a} = \left(a_{\imath_1 \imath_2 \ldots \imath_t} \right), \qquad \boldsymbol{b} = \left(b_{\imath_1 \imath_2 \ldots \imath_t} \right) \tag{A1}$$

is

$$\boldsymbol{a} \circ \boldsymbol{b} := \left(a_{\imath_1 \imath_2 \ldots \imath_t} b_{\imath_1 \imath_2 \ldots \imath_t} \right), \tag{A2}$$

which is simply the rank-t tensor comprising element-wise products of the elements of each of the two rank-t tensors in the product. For $t = 1$, $\boldsymbol{a}$ and $\boldsymbol{b}$ are vectors, and $\circ$ is just the vector obtained by element-wise products of the corresponding vector elements. We define the sum over all elements in the Hadamard product (A2) by

$$\boldsymbol{a} \odot \boldsymbol{b} := \sum_{\imath_1 \imath_2 \ldots \imath_t} (\boldsymbol{a} \circ \boldsymbol{b})_{\imath_1 \imath_2 \ldots \imath_t} = \sum_{\imath_1 \imath_2 \ldots \imath_t} a_{\imath_1 \imath_2 \ldots \imath_t} b_{\imath_1 \imath_2 \ldots \imath_t}. \tag{A3}$$

Typically, in the literature, $\circ$ and $\odot$ are both employed to refer to our $\circ$, but here we use $\circ$ and $\odot$ as distinct operations as we have explained here. The norm of a vector, matrix or tensor is

$$\text{norm } \boldsymbol{a} = \sqrt{\boldsymbol{a} \odot \boldsymbol{a}}, \tag{A4}$$

which is nonnegative.

For $\boldsymbol{a}$ (A1) restricted by

$$0 < a_{\imath_1 \ldots \imath_t} \leqslant 1, \tag{A5}$$

and introducing $\mathbb{J} := (1)$ as the tensor with every entry being 1 (in contrast to $\mathbb{I}$ being the matrix such that only diagonals are 1 and off-diagonal elements are all 0) of equal size to tensor $\boldsymbol{a}$, the element-wise logarithm is

$$\log \boldsymbol{a} := -\sum_{\ell=1}^{\infty} \frac{1}{\ell} (\boldsymbol{1} - \boldsymbol{a})^{\circ \ell} = \left(\log a_{\imath_1 \ldots \imath_t} \right). \tag{A6}$$

Here we use the notation $\bullet^{\circ \ell}$ to refer to the ℓ-fold element-wise product of the tensor $\bullet$ with itself.

For constructing conditional states, we employ Hadamard division $\oslash$ [23]. Hadamard division for two same-dimensional tensors (including vectors and matrices) is simply their element-wise division. Another definition of Hadamard division applies for a matrix divided by a vector, where the length of the vector equals the number of rows (or columns) of the matrix; in this case Hadamard division of the matrix by the vector corresponds to division of the row (or column) vectors of the matrix by the elements of a vector.

Appendix B. Probability vector and entropy

We introduce stochastic information, which is essentially already known [11], but we employ the elegant notation of appendix A in a novel way to convey the concepts with simple, easy-to-grasp expressions once Hadamard arithmetic [23] is clear. Below we explain the concept of a stochastic-information state. Then we have set the stage for the subsequent discussion on bipartite stochastic-information states.

Our stochastic-information construct is a distribution of messages, with each message labelled by integer

$$m \in [M] := \{1, 2, \ldots, M\} \subset \mathbb{Z}^+ \tag{B1}$$

so the distribution of such messages is

$$\boldsymbol{p} := \left(p_m; m \in [M]\right) \in \mathbb{R}^L, p_m \geqslant 0\, \forall\, m, \sum_{m=1}^{M} p_m = 1 \tag{B2}$$

represented here as a probability vector [24]. The probability vector can be permuted in the sense of rearranging the probabilities of various messages.

A given permutation is represented by some $\sigma \in S_M$, for S_M the permutation (or symmetry) group over all M messages. The cardinality of S_M is $M!$, and the permutation group can be ordered in various ways, and we adopt reverse lexicographical ordering [25]. A permutation of the probability vector is represented by permutation matrix Π_σ for permutation $\sigma \in M_M\left(\{0,1\}\right)$ (i.e. $M \times M$ matrices whose entries are only 0 or 1) given by [26]

$$\Pi_\sigma \boldsymbol{p} = \left(p_{\sigma(b)}\right) \quad \forall\, \boldsymbol{p} \tag{B3}$$

so Π_σ contains exactly one entry of 1 in each row or column and the rest of the entries are 0. We represent the sequence of all permutation matrices by the vector

$$\boldsymbol{\Pi} := (\Pi_\sigma; \sigma \in S_M) \tag{B4}$$

with the sequence of σ drawn from the permutation group in reverse lexicographical order.

A specific message of interest is a versor, which is a vector whose entries all zero except one element whose entry is one [27]. Any permutation of a versor corresponds to just replacing a given message by a new message so a permutation of a versor is just another versor. A versor is equivalent to a stochastic-information state $p_m = \delta_{m\check{m}}$ for $\check{m}$ the given message. We write this versor, corresponding to specific message $\check{m}$, as $\boldsymbol{\delta}_{\check{m}}$. Versors form a basis for stochastic-information states, which we call the message basis. A permutation matrix (B3) is actually a tensor product of a versor and a coversor, with a coversor defined to be a covector version of a versor.

The Cartesian product of versors forms a basis for bipartite stochastic-information states, from which a joint distribution can be constructed. Suppose the two parties, Alice and Bob, each hold an information state $\boldsymbol{\delta}_{\check{m}}^{A}$ and $\boldsymbol{\delta}_{\check{m}'}^{B}$, respectively, where we use superscripts ^{A} and ^{B} to denote who owns which of the vector spaces in the tensor product. Alice's and Bob's joint state is the bipartite versor

$$\boldsymbol{\delta}_{mm'}^{AB} := \boldsymbol{\delta}_m^A \boldsymbol{\delta}_{m'}^B \in M_{M^A \times M^B}\left(\{0,1\}\right), \tag{B5}$$

which is pure over the Cartesian product. For $M^A = M^B$, the state represented by the $M^A \times M^B$ identity matrix $\mathbb{1}/\sqrt{M}$, for $M := M^A M^B$, corresponds to all messages to Alice and Bob being identical but all message instances are equally likely. Thus, the conditional entropy of this mixture of states is zero whereas the entropy of this set of messages pairs is maximum; i.e. the maximum entropy is $\log M^A$.

Appendix C. Quantum discord

We now explain quantum discord at a high level instead of delving into a full mathematical description of quantum discord, which requires Hilbert space. We begin with a brief discussion of the history of quantum discord, both theoretical and experimental, and typical interpretations of what quantum discord means such as quantum correlations and resources.

The concept of quantum discord was proposed to separate total correlations of a bipartite quantum state into purely quantum and classical parts [1, 2]. Quantum discord per se is the difference between two classically identical expressions for mutual information but adapted for a quantum system and was described as a measure of quantumness of correlations. Discord appeared to be a more general way of quantifying quantum correlations [21, 28–30] compared to entanglement as vanishing entanglement does not ensure vanishing discord, and the absence of entanglement does not imply classicality. Operationally, quantum discord has been interpreted in the context of quantum state merging, if pertinent prior information is discarded [31, 32], the quantum–classical separation associated with discord has been cast in terms of negative conditional entropy [8]. The value of discord as a quantum resource has been debated vigorously, sometimes as a powerful quantum resource [7] but also critically, for example that pure states of nonzero discord have zero measure [21] and that nonzero discord is classically explainable iff entanglement is not present, as discussed in remark 4 of reference [8]. Discord has been generalised for different types of quantum measurements [33], for Rényi entropy [34, 35] and for higher dimension [36].

Quantum discord has been experimentally studied in optical systems in which quantum discord was shown to be a resource for quantum remote-state preparation, specifically showing that separable states with non-zero quantum discord can outperform entangled states [37]. In continuous-variable Gaussian

optics, the experimental Gaussian quantum discord has been studied for a two-mode squeezed thermal state [38]. Experimentally encoding information within the discordant correlations of two separable Gaussian states shows that bipartite discord can be consumed to encode information that is only accessible by coherent quantum interactions [39]. A flexible two-photon setup has realized a three-qubit system with programmable degrees of initial correlations, measurement interaction, and characterization processes, thereby yielding the demonstration that local observation in an activation protocol for converting discord into distillable entanglement [40]. A trapped-ion experiment has shown that quantum discord inference of open-system dynamics detects system-environment quantum correlations without accessing the environment [41].

In contrast to stochastic information states, which are probability distributions (B2) or joint distributions for the bipartite case (1), the quantum state is a trace-class bounded completely positive operator ρ on Hilbert space $\mathscr{H}$, or on the tensor product $\mathscr{H} \otimes \mathscr{H}$ for the bipartite case [9]. The quantum state's entropy is $H(\rho) = -tr\left(\rho \log \rho\right)$ for tr the trace operation. In quantum information theory, measurement is described by positive operator-valued measures [9], but, for quantum discord, only projective-valued measures $\{P\}$ [9], which are self-adjoint projections on $\mathscr{H}$, are considered. Each projective-valued measure P comprises a set of projective operators P_ι with $P_\iota P_j = P_\iota \delta_{\iota j}$. Measurement of a state yields a real-valued outcome, and the state is subsequently described by the jth projection $P_j \rho$ corresponding to that outcome. This projection is expressed as a conjugation of ρ in the literature [2], but this way of expressing is superfluous for our purposes and hence not employed.

The conditional quantum state, first defined by Cerf and Adami [42], is now typically defined as [2]

$$\rho_j^{\mathrm{A|B}} := \left(\mathbb{1} \otimes P_j\right) \rho^{\mathrm{AB}\left(\mathbb{1} \otimes P_j\right)}/p_j, \; p_j := tr\left(\left(\mathbb{1} \otimes P_j\right) \rho^{\mathrm{AB}\left(\mathbb{1} \otimes P_j\right)}\right) \tag{C1}$$

with p_j the probability of Bob obtaining jth outcome after he has applied projection-valued measures (PVM) P. The conditional quantum entropy [42] is

$$H^{\mathrm{A|B}} = \sum_j p_j H_j^{\mathrm{A|B}}, \quad H_j^{\mathrm{A|B}}\left(\rho^{\mathrm{AB}}\right) := H\left(\rho_j^{\mathrm{A|B}}\right), \tag{C2}$$

which is a probability-weighted average of conditional entropy.

Mutual information $I^{\mathrm{A;B}}(\rho^{\mathrm{AB}})$ is the same as for equation (9) except that H now corresponds to the quantum entropy, and the last term is replaced by $H^{\mathrm{A:B}} \mapsto H\left(\rho^{\mathrm{AB}}\right)$. The alternative mutual information is

$$J^{\mathrm{A;B}} = H^{\mathrm{A}} - \sup_{P^{\mathrm{B}}} H^{\mathrm{A|B}}, \quad H^{\mathrm{A}} := H\left(\rho^{\mathrm{A}}\right), \quad \rho^{\mathrm{A}} = tr_{\mathrm{B}} \rho^{\mathrm{AB}}, \tag{C3}$$

which is the supremum over all Bob's PVMs. Quantum discord, analogous to classical discord (11), which was actually defined later than quantum discord [8], is

$$\Delta^{\mathrm{A;B}}\left(\rho^{\mathrm{AB}}\right) := I^{\mathrm{A;B}}\left(\rho^{\mathrm{AB}}\right) - J^{\mathrm{A;B}}\left(\rho^{\mathrm{AB}}\right), \tag{C4}$$

which quantifies the difference between the two mutual information quantities for quantum discord.

Quantum discord is interpreted as correlations that remain after classical correlations are subtracted from total correlation and recognised to quantify the non-classical correlations in a quantum system, including entanglement and therefore identified as a quantum resource [30]. The primary feature encapsulated by its quantum property is how a state is affected by local measurements and seen as a form of classical correlation aided with quantum coherence (superpositions) at the level of individual subsystems [21, 28, 43]. The presence of discord in quantum computing protocols [5, 6], motivates the assertion that discord is a quantum resource, operationalized by state merging [4], which can deliver the quantum advantage.

ORCID iDs

Wei-Wei Zhang ⓘ https://orcid.org/0000-0002-8164-9527
Yuval R. Sanders ⓘ https://orcid.org/0000-0001-8003-0039
Barry C. Sanders ⓘ https://orcid.org/0000-0002-8326-8912

References

[1] Henderson L and Vedral V 2001 Classical, quantum and total correlations *J. Phys. A: Math. Gen.* **34** 6899–905
[2] Ollivier H and Zurek W H 2001 Quantum discord: a measure of the quantumness of correlations *Phys. Rev. Lett.* **88** 017901

[3] Brunner N, Cavalcanti D, Pironio S, Scarani V and Wehner S 2014 Bell nonlocality *Rev. Mod. Phys.* **86** 419–78

[4] Horodecki M, Oppenheim J and Winter A 2007 Quantum state merging and negative information *Commun. Math. Phys.* **269** 107–36

[5] Lanyon B P, Barbieri M, Almeida M P and White A G 2008 Experimental quantum computing without entanglement *Phys. Rev. Lett.* **101** 200501

[6] Datta A, Shaji A and Caves C M 2008 Quantum discord and the power of one qubit *Phys. Rev. Lett.* **100** 050502

[7] Merali Z 2011 Quantum computing: the power of discord *Nature* **474** 24–6

[8] Gheorghiu V, Marcos C, de Oliveira , Barry C and Sanders 2015 Nonzero classical discord *Phys. Rev. Lett.* **115** 030403

[9] Nielsen M A and Chuang I 2011 *Quantum Computation and Quantum Information* (Cambridge: Cambridge University Press)

[10] Shannon C E 1948 A mathematical theory of communication *Bell Syst. Tech. J.* **27** 379–423

[11] Cover T M and Thomas J A 2012 *Elements of Information Theory* (New York: Wiley)

[12] Milonni P W 1976 Semiclassical and quantum-electrodynamical approaches in nonrelativistic radiation theory *Phys. Rep.* **25** 1–81

[13] Gasbarri G and Ferialdi L 2018 Stochastic unravelings of non-Markovian completely positive and trace-preserving maps *Phys. Rev.* A **98** 042111

[14] Budish E, Che Y-K, Kojima F and Paul M 2009 Implementing random assignments: a generalization of the Birkhoff–von Neumann theorem *2009 Cowles Summer Conf.*

[15] Sinkhorn R 1966 A relationship between arbitrary positive matrices and stochastic matrices *Can. J. Math.* **18** 303–6

[16] Johnson C R 1981 Row stochastic matrices similar to doubly stochastic matrices *Linear and Multilinear Algebra* **10** 113–30

[17] Levin D A and Peres Y 2017 *Markov Chains and Mixing Times* (*MBK*) (Providence, RI: American Mathematical Society)

[18] Rezaei F, Charalambous C D and Ahmed N U 2012 Optimal control of uncertain stochastic systems subject to total variation distance uncertainty *SIAM J. Control Optim.* **50** 2683–725

[19] Charalambous C D, Tzortzis I, Loyka S and Charalambous T 2014 Extremum problems with total variation distance and their applications *IEEE Trans. Autom. Control* **59** 2353–68

[20] Gulati S and Mi J 2006 Testing for scale families using total variation distance *J. Stat. Comput. Simul.* **76** 773–92

[21] Ferraro A, Aolita L, Cavalcanti D, Cucchietti F M and Acín A 2010 Almost all quantum states have nonclassical correlations *Phys. Rev.* A **81** 052318

[22] Bocci C, Carlini E and Kileel J 2015 Hadamard products of linear spaces *J. Algebra* **448** 595–617

[23] Wetzstein G, Lanman D, Hirsch M and Raskar R 2012 Tensor displays: compressive light field synthesis using multilayer displays with directional backlighting *ACM Trans. Graph.* **31** 1–11

[24] Wang X M 2007 Probability bracket notation, probability vectors, Markov chains and stochastic processes (arXiv:cs/0702021)

[25] Ziabicki A 1992 The theory of ordering lexicographic entries: principles, algorithms and computer implementation *Comput. Hum.* **26** 119–37

[26] Najnudel J and Nikeghbali A 2013 The distribution of eigenvalues of randomized permutation matrices *Assoc. Ann. Inst. Fourier* **63** 773–838

[27] Loomis L H and Sternberg S 2014 *Advanced Calculus* Revised edn (Singapore: World Scientific)

[28] Vedral V 2017 Foundations of quantum discord *Lectures on General Quantum Correlations and their Applications* (Berlin: Springer) pp 3–7

[29] Egloff D, Matera J M, Theurer T and Plenio M B 2018 Of local operations and physical wires *Phys. Rev.* X **8** 031005

[30] Chitambar E and Gour G 2019 Quantum resource theories *Rev. Mod. Phys.* **91** 025001

[31] Madhok V and Datta A 2011 Interpreting quantum discord through quantum state merging *Phys. Rev.* A **83** 032323

[32] Cavalcanti D, Aolita L, Boixo S, Modi K, Piani M and Winter A 2011 Operational interpretations of quantum discord *Phys. Rev.* A **83** 032324

[33] Xu J 2011 Generalizations of quantum discord *J. Phys. A: Math. Theor.* **44** 445310

[34] Hou X-W, Huang Z-P and Chen S 2014 Quantum discord through the generalized entropy in bipartite quantum states *Eur. Phys. J.* D **68** 87

[35] Bellomo G, Plastino A, Majtey A P and Plastino A R 2014 Comment on quantum discord through the generalized entropy in bipartite quantum states *Eur. Phys. J.* D **68** 337

[36] Radhakrishnan C, Laurière M and Byrnes T 2020 Multipartite generalization of quantum discord *Phys. Rev. Lett.* **124** 110401

[37] Dakić B *et al* 2012 Quantum discord as resource for remote state preparation *Nat. Phys.* **8** 666–70

[38] Blandino R, Genoni M G, Etesse J, Barbieri M, Paris M G A, Grangier P, Tualle-Brouri R and Tualle-Brouri R 2012 Homodyne estimation of Gaussian quantum discord *Phys. Rev. Lett.* **109** 180402

[39] Gu M, Chrzanowski H M, Assad S M, Symul T, Modi K, Ralph T C, Vedral V and Lam P K 2012 Observing the operational significance of discord consumption *Nat. Phys.* **8** 671–5

[40] Adesso G, D'Ambrosio V, Nagali E, Piani M and Sciarrino F 2014 Experimental entanglement activation from discord in a programmable quantum measurement *Phys. Rev. Lett.* **112** 140501

[41] Gessner M, Ramm M, Pruttivarasin T, Buchleitner A, Breuer H-P and Häffner H 2014 Local detection of quantum correlations with a single trapped ion *Nat. Phys.* **10** 105–9

[42] Cerf N J and Adami C 1997 Negative entropy and information in quantum mechanics *Phys. Rev. Lett.* **79** 5194–7

[43] Wiseman H M 2013 Quantum discord is Bohr's notion of non-mechanical disturbance introduced to counter the Einstein–Podolsky–Rosen argument *Ann. Phys., NY* **338** 361–74