

# Quantum Rate-Distortion Coding of Relevant Information

Sina Salek, Daniela Cadamuro, Philipp Kammerlander and Karoline Wiesner

**Abstract**—Rate-distortion theory provides bounds for compressing data produced by an information source to a specified encoding rate that is strictly less than the source’s entropy. This necessarily entails some loss, or distortion, between the original source data and the best approximation after decompression. The so-called Information Bottleneck Method is designed to compress only ‘relevant’ information. Which information is relevant is determined by the correlation between the data being compressed and a variable of interest, so-called side information. In this paper, an Information Bottleneck Method is introduced for the compression of quantum data. The channel communication picture is used for compression and decompression. The rate of compression is derived using an entanglement assisted protocol with classical communication, and under an unproved conjecture that the rate function is convex in the distortion parameter. The optimum channel achieving this rate for a given input state is characterised. The conceptual difficulties arising due to differences in the mathematical formalism between quantum and classical probability theory are discussed and solutions are presented.

## I. INTRODUCTION

One of the most central results in classical information theory is Shannon’s data compression theorem [13] which gives a fundamental limit on lossless compressibility of data. Due to statistical redundancies, data can be compressed at a rate bounded below by the source entropy, such that after decompression the full information is recovered without loss. Rate-distortion theory (RDT) is the branch of information theory that compresses the data produced by an information source down to a specified encoding rate that is strictly less than the source entropy [14]. This necessarily entails some loss, or distortion, between the original source data and the best approximation after decompression, according to some distortion measure. RDT is frequently used in multimedia data compression where a large amount of data can be discarded without any noticeable change to a listener or viewer.

Whilst RDT is an important and widely used concept in information theory, there are cases where only part of the information in the data to be compressed is relevant. For instance in speech signal processing one might be interested

only in information about the spoken words in audio data. The Information Bottleneck Method (IBM), introduced by Tishby *et al.*, achieves a lossy compression rate even lower than the rate given by RDT by keeping only ‘relevant’ information [16]. Which information is relevant is determined by the correlation between the data being compressed and a variable of interest. The information to be recovered after decoding is only the relevant part of the source data. For example, one might have access to the transcript of an audio recording which has an entropy by orders of magnitude lower than the original audio data. The transcript here can be used as side information to compress the audio data further than what can be achieved by RDT, without increasing the distortion of the relevant information.

Loss of information in the context of a compression-decompression scheme is mathematically equivalent to transmission of information through a noisy channel. Rather than characterising the information lost by encoding, one characterises the information lost during transmission. The Information Bottleneck Method is formulated as a communication problem with the relevant variable acting as side information. Iterative algorithms to compute the optimum channel achieving the task are also provided in [16].

In [12], a first quantum extension of the Information Bottleneck Method was developed and applied to predictive filtering. In this paper we extend the Information Bottleneck Method to the quantum case by considering the transmission of quantum information through a quantum channel with side information. We derive the compression rate for an entanglement-assisted protocol with classical communication for our quantum extension of the IBM. Our derivation relies on the conjecture that the rate function is convex in distortion. The optimum quantum channel that can achieve this rate is also characterised.

A quantum extension to RDT was introduced by Barnum [3]. However, the results were unsatisfactory since the bound on the rate was given in terms of coherent information which can be negative. The results were improved and a complete characterisation of quantum channels achieving rate-distortion coding in a number of settings was given by Datta *et al.* [8]. Various settings of quantum RDT in the presence of auxiliary information were discussed in the work of Wilde *et al.* [18]. However, the specific question of transmitting relevant information asked in the IBM with its choice of distance measure and the specifics of the optimisation problem have not been considered yet.

The setting of the classical IBM is as follows. A correlated pair of random variables  $X$  and  $Y$  is given with a joint probability distribution  $P(X, Y)$ . The task is to find the

Sina Salek is with the Department of Computer Science, University of Oxford, Wolfson Building, Parks Road, UK. Email: salek.sina@gmail.com. He was with the University of Hong Kong, Hong Kong SAR, and the University of Bristol, UK, for part of this research.

Daniela Cadamuro is with Zentrum Mathematik, Technische Universität München, 85748 Garching, Germany. She was with the University of Bristol for part of this research.

Philipp Kammerlander is with Institut für Theoretische Physik, ETH Zürich, 8093 Zürich, Switzerland.

Karoline Wiesner is with School of Mathematics, University of Bristol, Bristol BS8 1TW, United Kingdom.

optimum channel with input  $X$  and output  $\tilde{X}$  such that  $\tilde{X}$  retains a fixed amount of correlation,  $C$ , with variable  $Y$ . The amount of correlation is quantified by the Shannon mutual information  $I(\tilde{X}; Y) := H(\tilde{X}) + H(Y) - H(\tilde{X}Y)$ , where  $H(\cdot)$  is the Shannon entropy. For a successful completion of the IBM task it is required that  $I(\tilde{X}; Y) \geq C$ . Representing the channel by the conditional probability distribution  $P(\tilde{X}|X)$ , one can show that the classical rate of compression for a given minimum amount of correlation  $C$ ,  $R_{cls}(C)$ , is given by

$$R_{cls}(C) = \min_{P(\tilde{X}|X): I(\tilde{X}; Y) \geq C} I(X; \tilde{X}), \quad (1)$$

as it was first proposed in [16] and proved in [11].

The IBM, however, is concerned not with an output distribution close to the input distribution but with an output characterised by its information about some other variable  $Y$ . The task of the IBM is to find the lowest value of  $I(X; \tilde{X})$  such that  $I(\tilde{X}; Y)$  is still above some given threshold. The value of  $I(X; \tilde{X})$  can be reinterpreted as a communication rate, namely the number of transmitted bits needed to specify an element of  $\tilde{X}$ , per element of  $X$  [16, Sec. 2]. Minimising the mutual information with respect to all channels that satisfy the threshold criterion achieves the task.<sup>1</sup>

The channel that achieves the rate in Eq. (1) can be found by the Lagrangian technique. The Lagrangian is defined as

$$\mathcal{L}_{cls} = I(X; \tilde{X}) - \beta I(\tilde{X}; Y) - \sum_{x, \tilde{x}} \lambda(x) P(\tilde{x}|x), \quad (2)$$

where  $\beta$  is the Lagrange multiplier for the information constraint and  $\lambda(x)$  are the Lagrange multipliers for the normalisation constraint of the conditional distribution  $P(\tilde{X}|X)$ . Taking the derivative of the Lagrangian with respect to the channel and setting it to 0 gives the expression for the channel as

$$P(\tilde{x}|x) = P(\tilde{x}) \frac{e^{-\beta D(P(Y|x)||P(Y|\tilde{x}))}}{Z}. \quad (3)$$

$D(\cdot||\cdot)$  in the exponent on the right hand side is the Kullback-Leibler divergence of two probability distributions  $P(Y|x)$  and  $P(Y|\tilde{x})$ , and  $Z$  is the normalising factor.

The setting of the quantum IBM is as follows. The input to the channel is the system  $X$  which is described by a state  $\rho_{XY}$  with side information  $Y$ . The channel acts on the  $X$  part of this state,  $\rho_X$ . The output of the quantum channel is the system  $\tilde{X}$  which is also an entangled state with the side information  $Y$ ,  $\tilde{\rho}_{\tilde{X}Y}$ . Correlations in the state  $\tilde{\rho}_{\tilde{X}Y}$  are measured by the quantum mutual information  $I(\tilde{X}; Y)_{\tilde{\rho}} := S(\tilde{X})_{\tilde{\rho}} + S(Y)_{\tilde{\rho}} - S(\tilde{X}Y)_{\tilde{\rho}}$  where  $S(\cdot)$  is the von Neumann entropy to base  $e$ . The Bottleneck constraint in the quantum case is  $I(\tilde{X}; Y)_{\tilde{\rho}} \geq J$ , that is, a minimum amount of correlation  $J$  between systems  $\tilde{X}$  and  $Y$  must be retained. This choice of measure for correlation between  $\tilde{X}$  and  $Y$  naturally generalizes the setup of the classical Bottleneck method; it includes both classical correlation and quantum entanglement between the two systems.

<sup>1</sup>Note that the analogy between IBM and RDT is in spirit only. The technical difference is the distortion measure, which is a function of the output alphabet in the case of RDT while it is a function of the probability distribution of the outputs in the case of IBM.

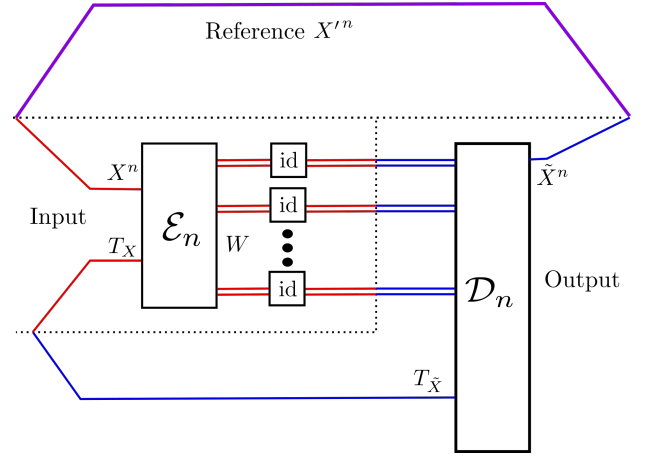


Fig. 1. The protocol for the entanglement assisted Quantum Information Bottleneck Method. The compression encoding  $\mathcal{E}_n$  acts on  $n$  copies of the state of the system  $X$ . The  $n$  copies of the state of the system  $Y$  are used as the *relevance variable* and are entangled with the system  $X$ . However, as the system  $Y$  is not transmitted through the channel, we do not depict the system in the figure.  $T_X$  and  $T_{\tilde{X}}$  are the two entangled systems that the input and the output of the protocol share to assist the transmission. The output of the compressing map  $\mathcal{E}_n$  is classical system  $W$  which is transferred to the output section for decompression,  $\mathcal{D}_n$ , via the noiseless channels “id”. The reference, labeled  $X'$ , is what purifies the state  $\rho_X$ . The state  $\rho_X$  is the reduced density operator of the given initial state  $\rho_{XY}$ .

In the next section we provide the protocol for the quantum IBM which results in a conjecture for the quantum counterpart of the compression rate in Eq. (1). Section III contains the channel optimisation, resulting in the quantum counterpart of Eq. (3).

## II. A PROTOCOL FOR THE QUANTUM INFORMATION BOTTLENECK METHOD

In this section we explicitly describe a protocol that simulates the action of the channel in the setting discussed in the introduction by compression and decompression channels and classical communication with the assistance of entangled resources. We then define the compression rate for this protocol, before proving our main result.

Following the usual approach in RDT we call a rate-distortion pair  $(r, J)$  achievable, where  $r \geq 0$  is the rate and  $J \geq 0$  is the distortion, if, roughly speaking, a lossy channel exists such that it can transmit a message, i.e., a state of the given input system, by sending  $r$  bits with at the amount of distortion not exceeding  $J$ . The rate function  $R(J)$  is then given as an infimum over achievable pairs. Of course, it is crucial to specify what exactly is meant by “lossy channel” and “achievable” in our context, and this is what we will discuss now.

Let us consider the entanglement-assisted protocol with noiseless classical communication illustrated in Fig. 1. The information of system  $X$  to be compressed is represented by  $n$  independently and identically distributed (i.i.d.) copies,  $\rho_X^{\otimes n}$ , of the density operator  $\rho_X$ , with a purification  $\tau_{X'X}$ . The input, however, is correlated with a system  $Y$  which contains

our relevant information, and the state is denoted  $\rho_{XY}$ . The sender and receiver of the information share an entangled resource  $\Phi_{T_X T_{\tilde{X}}}$ , where the system  $T_X$  is with the sender and the system  $T_{\tilde{X}}$  is with the receiver. The sender acts on the input state,  $\rho_X^{\otimes n}$ , and the state of half of the entangled pair,  $T_X$ , with the compression map  $\mathcal{E}_n := \mathcal{E}_{X^n T_X \rightarrow W}$ , where  $W$  is a classical system of size at most  $e^{nr}$ , and  $r$  is the communication rate. Then, the receiver acts on  $W$  and the state of the other half of the entangled pair,  $T_{\tilde{X}}$ , with the decompression channel  $\mathcal{D}_n := \mathcal{D}_{W T_{\tilde{X}} \rightarrow \tilde{X}^n}$ . The overall action  $\mathcal{F}_n := \mathcal{D}_n \circ \mathcal{E}_n$  of the compression-decompression channel, as specified above, defines an  $(n, r)$  quantum rate-distortion code.

Given such  $\mathcal{F}_n$ , we consider the marginal operation,  $1 \leq i \leq n$ ,

$$\mathcal{F}_n^{(i)}(\xi_X) := \text{Tr}_{\tilde{X}_1, \dots, \tilde{X}_{i-1}, \tilde{X}_{i+1}, \dots, \tilde{X}_n} [\mathcal{F}_n(\rho_X^{\otimes(i-1)} \otimes \xi_X \otimes \rho_X^{\otimes(n-i)})]. \quad (4)$$

Then, for any  $i$ , we can define

$$\tilde{\sigma}_{\tilde{X}_i Y_i} := \mathcal{F}_n^{(i)} \otimes \mathcal{I}_Y(\rho_{XY}), \quad (5)$$

where  $\mathcal{I}_Y$  is the identity channel acting on the system  $Y$ , and its partial traces are

$$\tilde{\sigma}_{\tilde{X}_i} = \mathcal{F}_n^{(i)}(\rho_X), \quad \tilde{\sigma}_{Y_i} = \rho_Y. \quad (6)$$

Using Eqs. (5) and (6) we can define the mutual information

$$I_i(\tilde{X}; Y) := S(\tilde{\sigma}_{\tilde{X}_i}) + S(\tilde{\sigma}_{Y_i}) - S(\tilde{\sigma}_{\tilde{X}_i Y_i}), \quad (7)$$

and its average over many uses of the channel,

$$\bar{I}_n(\tilde{X}; Y) := \frac{1}{n} \sum_{i=1}^n I_i(\tilde{X}; Y). \quad (8)$$

Now, for any  $r, J \geq 0$ , we call  $(r, J)$  an *achievable rate-distortion pair* (for the Quantum Information Bottleneck Method) if there exists a sequence of  $(n, r)$  quantum rate-distortion codes such that

$$\lim_{n \rightarrow \infty} \bar{I}_n(\tilde{X}; Y) > J. \quad (9)$$

Finally, the rate function  $R(J)$  is defined as

$$R(J) := \inf\{r : (r, J) \text{ is achievable}\}. \quad (10)$$

Given these definitions, we can now express  $R(J)$  in a more direct way in terms of *one* copy of the input state  $\rho_{XY}$ .

**Conjecture 1.** Consider a memoryless quantum information source  $\rho_{XY}$ , where system  $X$  is to be compressed such that after decompression it retains an amount of correlation  $0 \leq J \leq 1$  with system  $Y$ . The asymptotic Bottleneck rate function for entanglement-assisted lossy source coding with noiseless classical communication is given by

$$R(J) = \min_{\mathcal{N}_{X \rightarrow \tilde{X}} : I(\tilde{X}; Y)_{\tilde{\rho}} \geq J} I(X'; \tilde{X})_{\tilde{\tau}}, \quad (11)$$

under the conjecture that the expression on the right hand side of Eq. (11) is convex in  $J$ . Here, the quantum mutual information  $I(X'; \tilde{X})$  is evaluated over the state

$$\tilde{\tau}_{X' \tilde{X}} := (\mathcal{I}_{X'} \otimes \mathcal{N}_{X \rightarrow \tilde{X}})(\tau_{X' X}), \quad (12)$$

such that  $\tau_{X' X}$  is the purification of the reduced state  $\rho_X$  of the state  $\rho_{XY}$ , and

$$\tilde{\rho}_{\tilde{X} Y} := (\mathcal{N}_{X \rightarrow \tilde{X}} \otimes \mathcal{I}_Y)(\rho_{XY}) \quad (13)$$

While we are currently unable to remove the extra convexity assumption, we have verified it in numerical examples which are discussed in Appendix B.

*Proof.* Here we follow the approach of [8, Theorem 2]. We temporarily denote the right-hand side of Eq. (11) as

$$M(J) := \min_{\mathcal{N}_{X \rightarrow \tilde{X}} : I(\tilde{X}; Y)_{\tilde{\rho}} \geq J} I(X'; \tilde{X})_{\tilde{\tau}}. \quad (14)$$

We need to show achievability of the rate,  $R(J) \leq M(J)$ , as well as optimality,  $R(J) \geq M(J)$ . We start with optimality. Let  $(r, J)$  be an achievable rate-distortion pair and  $\mathcal{F}_n$  a corresponding sequence of codes. We have for large  $n$ ,

$$\begin{aligned} nr &\geq S(W) \\ &\geq S(W|T_{\tilde{X}}) \\ &\geq S(W|T_{\tilde{X}}) - S(W|X'^n T_{\tilde{X}}) \\ &= I(W; X'^n | T_{\tilde{X}}) \\ &= I(W; X'^n | T_{\tilde{X}}) + I(X'^n; T_{\tilde{X}}) \\ &= I(W T_{\tilde{X}}; X'^n) \\ &\geq I(\tilde{X}^n; X'^n) \\ &\geq \sum_{i=1}^n I(\tilde{X}_i; X'_i) \\ &\geq \sum_{i=1}^n M(I_i(\tilde{X}; Y)) \\ &= n \sum_{i=1}^n \frac{1}{n} M(I_i(\tilde{X}; Y)) \\ &\geq n M(\bar{I}_n(\tilde{X}; Y)) \\ &\geq n M(J). \end{aligned} \quad (15)$$

The first inequality follows from the fact that the entropy of the uniform distribution,  $nr$ , is the upper bound of  $S(W)$ . The second inequality follows because entropy is nondecreasing under conditioning. The third inequality follows because the state of the system  $W X'^n T_{\tilde{X}}$  is separable with respect to the classical system  $W$  and therefore  $S(W|X'^n T_{\tilde{X}}) \geq 0$  [7, footnote 10]. The first equality follows from the definition of mutual information. The second equality follows since the state  $\Phi_{T_X T_{\tilde{X}}}$  is in a tensor product with the state of the remaining input and therefore  $I(X'^n; T_{\tilde{X}}) = 0$ . In the third equality we use the chain rule for mutual information. The fourth inequality follows from the data processing inequality. The fifth inequality follows from superadditivity of quantum mutual information [8, Lemma 15]. The sixth inequality follows from the definition of  $M(J)$ , where we use the channel  $\mathcal{N}_{X \rightarrow \tilde{X}} = \mathcal{F}_n^{(i)}$ . In the seventh inequality we have used conjecture of convexity of  $M$ . Finally the last inequality follows for large  $n$  from Eq. (9), using that  $M(J)$  is a nondecreasing function of  $J$ . The rate function is nondecreasing in  $J$ , because for any  $J' > J$  the domain of

minimisation in Eq. (11) becomes smaller, which implies that the rate function can only become larger. Thus, since  $(r, J)$  was arbitrary, (15) implies  $R(J) \geq M(J)$ . □

Achievability follows from an application of the Quantum Reverse Shannon Theorem (QRST) [4], [5], [10]. The general form of the QRST states that a quantum channel can be simulated by an unlimited amount of shared entanglement and an amount of classical communication equal to the channel's entanglement-assisted classical capacity. Specifically, in our case, fix  $J > 0$  and let  $\mathcal{N}_{X \rightarrow \tilde{X}}$  be the optimum channel at which the minimum in Eq. (14) is attained. For a given  $\epsilon > 0$ , we use the QRST to construct a sequence of channels  $\mathcal{F}_n = \mathcal{D}_n \circ \mathcal{E}_n$  such that they are close to an  $n$ -fold application of  $\mathcal{N}_{X \rightarrow \tilde{X}}$ , in the sense that

$$\|\tilde{\sigma}_{X^n \tilde{X}^n} - \tilde{\tau}_{X^n \tilde{X}^n}^{\otimes n}\|_1 \leq \epsilon, \quad (16)$$

where  $\tau_{X' \tilde{X}}$  is defined in Eq. (12) and

$$\tilde{\sigma}_{X^n \tilde{X}^n} := \mathcal{I}_{X^n} \otimes \mathcal{F}_n(\tau_{X' \tilde{X}}^{\otimes n}). \quad (17)$$

According to the QRST we can construct such a sequence of channels,  $\mathcal{F}_n$ , using classical communication at rate  $r = I(X'; \tilde{X})_{\tilde{\tau}}$ , given an unlimited amount of entanglement for a known tensor power input, which is our setting, *c.f.* [4, Fig. 2, top right]<sup>2</sup>. From Eq. (16) and the fact that  $I(\tilde{X}; Y) \geq J$  for the channel  $\mathcal{N}_{X \rightarrow \tilde{X}}$ , one can show (see Lemma 1 in Appendix A) that Eq. (9) is fulfilled with  $J - \delta$  instead of  $J$ , where  $\delta \rightarrow 0$  as  $\epsilon \rightarrow 0$ . Hence,  $(r, J - \delta)$  is achievable and  $R(J - \delta) \leq M(J)$ .

We have now shown that

$$\forall \delta > 0 : M(J - \delta) \leq R(J - \delta) \leq M(J). \quad (18)$$

From this it follows that

$$\lim_{\delta \searrow 0} R(J - \delta) = M(J). \quad (19)$$

Since  $R$  is nondecreasing, by Eqs. (9) and (10), and  $M$  is continuous (a property that follows from the conjecture of convexity),  $M(J) = R(J)$  for all  $J$ .

<sup>2</sup>Notice that the QRST in Fig. 2 of [4] was done in two ways, a so-called *feedback* and a *non-feedback* simulation, respectively. The former means that the environment of the simulated channel is in the possession of the sender, while the latter means some part of the environment of the channel will end up on the receiver's side. In both cases, if unlimited entanglement is available, the rate is given by the single-letter formula, the above mutual information. In the case of the non-feedback simulation of the channel  $\mathcal{N}_{X \rightarrow \tilde{X}}$ , the feedback simulation protocol [4, Theorem 3(a)] is used, namely the isometry  $\mathcal{U}_{\mathcal{N}}^{X \rightarrow E \tilde{X}}$  composed with the isometry  $V^{E^n \rightarrow E_A E_B}$ , is simulated by QRST, such that systems  $E_A$  and  $X$  are with the sender and the systems  $E_B$  and  $\tilde{X}$  are with the receiver. Tracing out the environment from this channel gives the desired channel,  $\mathcal{N}_{X \rightarrow \tilde{X}}$ . The result listed in Fig. 2 of [4] can be obtained from Theorem 3(b) there as follows. In the presence of an unlimited shared entanglement, condition (21) in [4] trivially holds and the maximum in Eq. (23) there is always attained by the first of the two terms. To obtain the desired expression of the mutual information from (20) and (23) there, it is enough to choose  $E^n, E_A$  and  $E_B$  such that the isometry  $V$  acts trivially and we have  $I(R^n : B^n E_B) = I(R^n : B^n)$ . Finally, the single letter formula  $I(R : B)$  that we use follows from the fact that we have a known tensor-power input.

<sup>3</sup>Notice that the rates shown in Fig. 2 of [4] are quantum communication rates which is why they are different by a factor of 1/2 from what we use. The extra factor drops because 1-qubit channels can be simulated with 2 bit classical channels in the presence of entanglement using the teleportation protocol.

*Remark 1:* In Eq. (11) the Quantum Reverse Shannon Theorem is used, which in addition uses shared entanglement (“entanglement assistance”) to generate a protocol that achieves the rate. However, the requirements for the Quantum Reverse Shannon Theorem are much more stringent than those of the Bottleneck method. Therefore, it might be possible to find a rate function without entanglement assistance. This is still an open problem.

*Remark 2:* The systems  $XY$  and  $XX'$  are understood to be in separate setups; the states  $\tilde{\rho}_{\tilde{X}Y}$  and  $\tilde{\tau}_{X'\tilde{X}}$  are not partial traces of a common parent state, but they involve an identical channel  $\mathcal{N}$  (“two separate experiments using the same apparatus”). The QRST applies only to the system  $XX'$ . To prove Eq. (11), these two setups need to be connected, which is done in Lemma 1 in Appendix A.

*Remark 3:* Generally, rate distortion functions obtained from the definition in Eq. (10) are non-increasing functions of the distortion, whereas the RHSs of Eqs. (11) and (1) are both non-decreasing functions. This is not a fundamental difference between RDT and the Bottleneck method. It is merely due to the fact that in the IBM the constraint of minimisation is chosen to be the amount of correlation preserved,  $I(\tilde{X}; Y) \geq J$ , while in RDT the constraint of minimisation is the average loss of information,  $\langle d(X; \tilde{X}) \rangle \leq D$ , for some fixed  $D$  and some distortion measure  $d(X; \tilde{X})$ , characterising the noise introduced by a channel  $\mathcal{N}_{X \rightarrow \tilde{X}}$  to a state  $\rho_X$ .

The formulation of the IBM can easily be changed to using a constraint on the loss of correlation such as  $I(X; Y) - I(\tilde{X}; Y) \leq D$ , in which case the rate function is a non-increasing function of  $D$ . Rather than changing it, the structure of the minimisation constraint is kept in line with the classical IBM.

*Remark 4:* As discussed above, in a lossy compression-decompression protocol the minimisation is performed over all channels satisfying a certain criterion, see Eq. (11). Although the functions  $I(\tilde{X}; X')$  and  $I(\tilde{X}; Y)$  are convex in the channel, the optimisation problem in Eq. (11) is not of convex type due to the sign of the inequality in the constraint ( $\geq$  rather than  $\leq$ ). Just as in the classical case, it is a so-called “reverse convex problem” (see, e.g., [15]), for which many of the standard results, such as strong duality or convexity of the resulting function in  $J$ , are not known to hold or apply only in a weaker form. Nevertheless, we present numerical evidence for convexity of the RHS in Appendix B. We also show that the minimum in Eq. (11) is actually attained at the boundary  $I(\tilde{X}; Y)_{\tilde{\rho}} = J$  (see Lemma 2 in Appendix C), which makes it useful to search for local optima with the method of Lagrange multipliers, as we show in the next section.

### III. THE OPTIMISATION PROBLEM FOR THE QUANTUM INFORMATION BOTTLENECK METHOD

We now proceed to the optimisation problem. In order to formulate the Lagrangian corresponding to the quantum counterpart of Eq. (2), we need to choose a suitable way of representing the channel  $\mathcal{N}$ , e.g., the Kraus operators or the Choi-Jamiołkowski representation. It turns out that indeed the

most compact and convenient way to compute the derivatives of the Lagrangian is with respect to the Choi-Jamiołkowski representation defined as

$$\Psi_{X'\tilde{X}} := (\mathcal{I}_{X'} \otimes \mathcal{N}_{X \rightarrow \tilde{X}})(\tilde{\Phi}_{X'X}), \quad (20)$$

where  $\tilde{\Phi}_{X'X} := \sum_{i,j=0}^{d-1} |i\rangle\langle j|_{X'} \otimes |i\rangle\langle j|_X$ . For the rate function given in Eq. (11) one can write the Lagrangian

$$\mathcal{L} := I(X'; \tilde{X})_{\tilde{\tau}} - \beta I(\tilde{X}; Y)_{\tilde{\rho}} - \text{Tr}_{X\tilde{X}}(\Psi_{X\tilde{X}}^{t_X}(\Lambda_X \otimes I_{\tilde{X}})), \quad (21)$$

where  $t_X$  denotes the partial transpose with respect to the basis  $\{|i\rangle_X\}_i$  on  $X$ ; here  $\beta$  is the Lagrange multiplier for the constraint of minimisation and the Hermitian operator  $\Lambda_X$  is the Lagrange multiplier to guarantee that the channel is a completely positive trace preserving map. The states in Eq. (21) can be written as functions of the Choi-Jamiołkowski state of the channel  $\Psi_{X\tilde{X}}$ . The joint state  $\tau_{X'X}$  in Eq. (21) can be written as

$$\begin{aligned} \tilde{\tau}_{X'\tilde{X}} &= \text{Tr}_X \{ \Psi_{X\tilde{X}}^{t_X} \tau_{X'X} \} \\ &= (\rho_{X'}^X \otimes I_{\tilde{X}})^{1/2} \Psi_{X'\tilde{X}}^{t_{X'}} (\rho_{X'}^X \otimes I_{\tilde{X}})^{1/2}, \end{aligned} \quad (22)$$

where  $\rho_{X'}^X$  is the same state as  $\rho_X$  acting on the Hilbert space  $\mathcal{H}_{X'}$  of the system  $X'$ .

By similar considerations, one can show that the joint state

$$\tilde{\rho}_{\tilde{X}Y} := (\mathcal{N}_{X \rightarrow \tilde{X}} \otimes \mathcal{I}_Y)(\rho_{XY}) \quad (23)$$

can be written as

$$\begin{aligned} \tilde{\rho}_{\tilde{X}Y} &= \text{Tr}_X \{ \Psi_{X\tilde{X}}^{t_X} \rho_{XY} \} \\ &= \text{Tr}_{X'Y'} \left\{ (\rho_{X'Y'}^{XY} \otimes I_{\tilde{X}Y})^{1/2} (\Psi_{X'\tilde{X}}^{t_{X'}} \otimes \Phi_{Y'Y'}^{t_{Y'}}) \right. \\ &\quad \left. (\rho_{X'Y'}^{XY} \otimes I_{\tilde{X}Y})^{1/2} \right\}, \end{aligned}$$

where we have chosen a maximally entangled state  $\Phi_{Y'Y}$  and a corresponding transpose  $t_{Y'}$ . In the third term of Eq. (21) the dependence on the channel state is already explicit.

Let

$$\begin{aligned} D_{X\tilde{X}}^{\beta Y} &:= \beta \log \tilde{\rho}_{\tilde{X}} \otimes I_X \\ &\quad - \beta \text{Tr}_Y \{ (\rho_X^{-1/2} \rho_{XY} \rho_X^{-1/2} (\log \tilde{\rho}_{\tilde{X}Y} \otimes I_X)) \}, \end{aligned} \quad (24)$$

and

$$\tilde{\Lambda}_X := \rho_X^{-1/2} \Lambda_X \rho_X^{-1/2}. \quad (25)$$

Taking the derivative of the Lagrangian in Eq. (21) with respect to the channel and setting it to zero (for details, see Appendix C) gives the optimum channel as

$$\begin{aligned} \Psi_{X\tilde{X}}^{t_X} &= (\rho_X \otimes I_{\tilde{X}})^{-1/2} \\ &\quad e^{\log \tilde{\rho}_{\tilde{X}} \otimes I_X - D_{X\tilde{X}}^{\beta Y} + \tilde{\Lambda}_X \otimes I_{\tilde{X}}} (\rho_X \otimes I_{\tilde{X}})^{-1/2}. \end{aligned} \quad (26)$$

Note that this determines  $\Psi_{X\tilde{X}}^{t_X}$  implicitly since it also appears on the right hand side of this equation in  $\tilde{\rho}_{\tilde{X}}$  and in the definition of  $D_{X\tilde{X}}^{\beta Y}$  (for details, see Appendix C). In order to find the optimum channel, Eq. (26) needs to be solved iteratively for  $\Psi_{X\tilde{X}}^{t_X}$ . The solution also depends on the unknown Lagrange multipliers  $\Lambda_X, \beta$  associated with the constraints of the problem; these need to be determined in a further

optimisation step. We comment on a possible algorithm to that end in Appendix C.

Eq. (26) reduces to its classical counterpart in Eq. (3) in the case of diagonal density operators. To see this, consider the diagonal case where the density operators reduce to probability distributions. From Eq. (26) it follows that

$$\begin{aligned} P(\tilde{x}|x) &= \frac{1}{P(x)} \exp \left\{ \log P(\tilde{x}) - \beta (\log P(\tilde{x}) \right. \\ &\quad \left. - \sum_y P(y|x) \log P(\tilde{x}y)) + \frac{\lambda(x)}{P(x)} \right\}, \end{aligned} \quad (27)$$

with  $\lambda(x)$  being the same normalisation Lagrange multiplier as in Eq. (2). Notice that since  $H(Y|X = x) = -\sum_y P(y|x) \log P(y|x)$  depends only on  $x$  but not on  $\tilde{x}$ , it can be absorbed into  $\lambda(x)$ . Defining

$$\tilde{\lambda}(x) := \frac{\lambda(x)}{P(x)} - \beta H(Y|X = x) - \log P(x), \quad (28)$$

Eq. (27) becomes

$$P(\tilde{x}|x) = P(\tilde{x}) e^{-\beta D(P(Y|x)||P(Y|\tilde{x})) + \tilde{\lambda}(x)}, \quad (29)$$

which is the same classical channel as Eq. (3), with all the extra terms being absorbed into the normalisation factor. This also shows that  $D_{X\tilde{X}}^{\beta Y}$  is a quantum operator corresponding to the distance measure in the classical Bottleneck method. The idea of distance operators has been used in a number of quantum information processing tasks [17], [9], however the  $D_{X\tilde{X}}^{\beta Y}$  is particular to the present setting. Eq. (26) can be used in principle to compute numerical values of quantum channels using iterative algorithms, akin to their classical counterparts by methods introduced by Blahut and Arimoto [6], [2].

Finally, we would like to remark that a previous result in the domain of quantum information theory was inspired by the classical Information Bottleneck Method [12], and indeed some of the technical steps in Section III of our paper are very similar. Here we clarify the difference between our contribution and that of [12]. To reflect on our contribution, consider that in classical communication, Shannon's channel capacity theorem gives a single-letter formula that fully characterises a communication channel. This allows one to interpret the mutual information between the input and the output of a channel as a communication rate, as was done in [16]. In quantum communication theory the situation is different since quantum channels (except for very few cases) are typically not characterised by such single-letter quantities. Therefore, in order to give a communication theoretic interpretation to an entropic quantity such as the one that we (and [12]) use, one has to prove that the quantity corresponds to a rate in a communication scenario by giving a coding scenario to which the rate refers. We have done this, up to the previously stated conjecture, by designing a setup that is compatible with the Quantum Reverse Shannon Theorem (notice, e.g., the unlimited amount of entanglement), and by proving the lemma given in Appendix A. This lemma shows that in our setting the Quantum Reverse Shannon Theorem can be invoked. In Ref. [12], the focus there was on applying a quantum extension of the Information Bottleneck Method to predictive filtering, whereas we focus here on the communication scenario.

#### IV. CONCLUSION AND OUTLOOK

This paper introduced the quantum extension of the Information Bottleneck Method. This method compresses the data such that only the information relevant with respect to some given variable is preserved. We derive a lower bound to the compression rate of relevant quantum information. The problem was formulated as a communication channel problem and the rate was shown to be achievable by explicitly constructing a channel achieving it. Just like in the classical case, the compression rate of the quantum Information Bottleneck Method is lower than that given by quantum rate distortion theory. Several conceptual issues arose from the structural differences between the mathematical formalism of quantum theory and classical probability theory which were discussed and solutions were presented.

Some open questions remain. Our proof of Eq. (11) relied on an unproven conjecture (convexity of the expression on the right hand side of Eq. (11) in  $J$ ). While this seems to be fulfilled in examples (cf. Appendix B), a proof of this property is currently missing.

In Appendix B a simple algorithm is used to compute the optimum channel and thus the rate function  $R(J)$  in low dimensional systems; but for systems of realistic size a more efficient algorithm would be required. This might be based on numerically solving the implicit equation (26).

#### ACKNOWLEDGMENT

The authors are grateful to Henning Bostelmann for discussions on the subject and for assistance with the numerics in Appendix B. Further, D. C. would like to thank Roger Colbeck and Robert Koenig for helpful discussions. P. K. would like to thank David Sutter and Renato Renner for comments on earlier versions of the manuscript.

S. S. is supported by the National Science Foundation of China through Grant No. 11675136, the Hong Kong Research Grant Council through Grant No. 17326616, the EPSRC National Quantum Technology Hub in Networked Quantum Information Technologies. This publication was made possible through the support of a grant from the John Templeton Foundation. The opinions expressed in this publication are those of the authors and do not necessarily reflect the views of the John Templeton Foundation. P. K. acknowledges funding from the Swiss National Science Foundation (through the National Centre of Competence in Research “Quantum Science and Technology” and through Project No. 200020\_165843), from the European Research Council (grant 258932) and through COST MP1209 “Thermodynamics in the quantum regime”.

K.W. acknowledges support through EPSRC grant I013717/1.

#### REFERENCES

- [1] The code is supplied with this article as a Maple worksheet `randomchannels.mw` and a PDF copy as `randomchannels.pdf`.
- [2] Suguru Arimoto. An algorithm for computing the capacity of arbitrary discrete memoryless channels. *IEEE Transactions on Information Theory*, 18(1):14–20, 1972.
- [3] Howard Barnum. Quantum rate-distortion coding. *Physical Review A*, 62(4):042309, 2000.

- [4] Charles H Bennett, Igor Devetak, Aram W Harrow, Peter W Shor, and Andreas Winter. The quantum reverse shannon theorem and resource tradeoffs for simulating quantum channels. *IEEE Transactions on Information Theory*, 60(5):2926–2959, 2014.
- [5] Mario Berta, Matthias Christandl, and Renato Renner. The quantum reverse shannon theorem based on one-shot information theory. *Communications in Mathematical Physics*, 306(3):579, 2011.
- [6] Richard E Blahut. Computation of channel capacity and rate-distortion functions. *IEEE Transactions on Information Theory*, 18(4):460–473, 1972.
- [7] N. J. Cerf and C. Adami. Negative entropy and information in quantum mechanics. *Physical Review Letters*, 79(26):5194–5197, 1997.
- [8] Nilanjana Datta, Min-Hsiu Hsieh, and Mark M Wilde. Quantum rate distortion, reverse shannon theorems, and source-channel separation. *IEEE Transactions on Information Theory*, 59(1):615–630, 2013.
- [9] Nilanjana Datta, Joseph M Renes, Renato Renner, and Mark M Wilde. One-shot lossy quantum data compression. *IEEE Transactions on Information Theory*, 59(12):8057–8076, 2013.
- [10] Igor Devetak. Triangle of dualities between quantum communication protocols. *Physical review letters*, 97(14):140503, 2006.
- [11] Ran Gilad-Bachrach, Amir Navot, and Naftali Tishby. An information theoretic tradeoff between complexity and accuracy. In Bernhard Schölkopf and Manfred K. Warmuth, editors, *Learning Theory and Kernel Machines*, pages 595–609, Berlin, Heidelberg, 2003. Springer Berlin Heidelberg.
- [12] A. L. Grimsmo and S. Still. Negative entropy and information in quantum mechanics. *Physical Review A*, 94:012338, 2016.
- [13] Claude E Shannon. A mathematical theory of communication. *Bell System Technical Journal*, 27:379–423, 1948.
- [14] Claude E Shannon. Coding theorems for a discrete source with a fidelity criterion. *IRE Nat. Conv. Rec.*, 4(142-163):1, 1959.
- [15] Ivan Singer. *Duality for Nonconvex Approximation and Optimization*. Springer Science & Business Media, 2007.
- [16] N Tishby, F C Pereira, and W Bialek. The information bottleneck method. *Proceedings of the 37th Annual Allerton Conference on Communication, Control and Computing*, pages 368–377, 1999.
- [17] Mark M Wilde. *Quantum information theory*. Cambridge University Press, 2013.
- [18] Mark M Wilde, Nipu Datta, Min-Hsiu Hsieh, and Andreas Winter. Quantum rate-distortion coding with auxiliary resources. *IEEE Transactions on Information Theory*, 59(10):6755–6773, 2013.
- [19] Andreas Winter. Compression of sources of probability distributions and density operators. *arXiv preprint quant-ph/0208131*, 2002.

#### APPENDIX A

##### LEMMA FOR THE PROOF OF THE BOTTLENECK RATE FUNCTION

The following lemma is relevant for the proof of achievability of the communication rate. It has the same application as in Theorem 19 of [19] and Lemma 1 of [8], but has been adapted to the distortion criterion for the quantum Bottleneck method.

**Lemma 1.** *Let  $\eta(\lambda) := -\lambda \log \lambda$ . There exists a constant  $k > 0$  depending only on the dimension of  $\mathcal{H}_{\tilde{X}Y}$  such that the following holds:*

*Let  $0 < J \leq I(X; Y)_\rho$  be fixed. Let a quantum channel  $\mathcal{N}_{X \rightarrow \tilde{X}}$  be such that if we apply the channel to the system  $X$  and an identity channel  $\mathcal{I}_Y$  on the system  $Y$  the effect will meet the condition  $I(\tilde{X}; Y)_{\tilde{\rho}} \geq J$ , where  $\tilde{\rho}_{\tilde{X}Y}$  is given by (23). Further, let  $\mathcal{F}_n$  be a sequence of quantum channels from the space of density matrices  $\mathcal{D}(\mathcal{H}_{X'X}^{\otimes n})$  to  $\mathcal{D}(\mathcal{H}_{X'\tilde{X}}^{\otimes n})$  such that*

$$\|\tilde{\sigma}_{X'n\tilde{X}^n} - \tilde{\tau}_{X'\tilde{X}}^{\otimes n}\|_1 \leq \epsilon \quad (30)$$

*for some  $0 \leq \epsilon < \frac{1}{e}$  and large enough  $n$ , where  $\tilde{\tau}_{X'\tilde{X}}$  and  $\tilde{\sigma}_{X'n\tilde{X}^n}$  are given by (12) and (17), respectively.*

*Then, for large enough  $n$  and  $\delta := k\eta(\epsilon)$ , we have*

$$\tilde{I}_n(\tilde{X}; Y)_{\tilde{\sigma}} \geq J - \delta. \quad (31)$$

*Proof.* Adding and subtracting  $I(\tilde{X}; Y)_{\tilde{\rho}}$  to the left hand side of Eq. (31) and using the triangle inequality, we obtain

$$\begin{aligned} \bar{I}_n(\tilde{X}; Y)_{\tilde{\sigma}} &= |I(\tilde{X}; Y)_{\tilde{\rho}} - (I(\tilde{X}; Y)_{\tilde{\rho}} - \bar{I}_n(\tilde{X}; Y)_{\tilde{\sigma}})| \\ &\geq |I(\tilde{X}; Y)_{\tilde{\rho}}| - |I(\tilde{X}; Y)_{\tilde{\rho}} - \bar{I}_n(\tilde{X}; Y)_{\tilde{\sigma}}|. \end{aligned} \quad (32)$$

Since we assumed that  $I(\tilde{X}; Y)_{\tilde{\rho}} \geq J$ , what we need to show is that

$$|I(\tilde{X}; Y)_{\tilde{\rho}} - \bar{I}_n(\tilde{X}; Y)_{\tilde{\sigma}}| \leq k\eta(\epsilon) =: \delta \quad (33)$$

with some  $k > 0$  and  $\eta$  as above. Inserting  $\bar{I}_n$  given in Eq. (8), we have

$$\begin{aligned} |I(\tilde{X}; Y)_{\tilde{\rho}} - \bar{I}_n(\tilde{X}; Y)_{\tilde{\sigma}}| &= \left| \frac{1}{n} \left( S(\tilde{\rho}_{\tilde{X}}) + S(\rho_Y) \right. \right. \\ &\quad \left. \left. - S(\tilde{\rho}_{\tilde{X}Y}) \right) - \frac{1}{n} \sum_{i=1}^n \left( S(\tilde{\sigma}_{\tilde{X}_i}) \right. \right. \\ &\quad \left. \left. + S(\tilde{\sigma}_{Y_i}) - S(\tilde{\sigma}_{\tilde{X}_i Y_i}) \right) \right|. \end{aligned}$$

Hence, it suffices to show that for all  $1 \leq i \leq n$ ,

$$\begin{aligned} |S(\tilde{\rho}_{\tilde{X}}) + S(\rho_Y) - S(\tilde{\rho}_{\tilde{X}Y}) - S(\tilde{\sigma}_{\tilde{X}_i}) \\ - S(\tilde{\sigma}_{Y_i}) + S(\tilde{\sigma}_{\tilde{X}_i Y_i})| \leq k\eta(\epsilon). \end{aligned}$$

In particular, we will prove bounds of the above type on  $|S(\tilde{\rho}_{\tilde{X}}) - S(\tilde{\sigma}_{\tilde{X}_i})|$ ,  $|S(\tilde{\rho}_{\tilde{X}Y}) - S(\tilde{\sigma}_{\tilde{X}_i Y_i})|$  and  $|S(\rho_Y) - S(\tilde{\sigma}_{Y_i})|$  for all  $i$ . We start with the first of these. To prove this inequality, we recall that  $\tau_{X'X}$  is a purification of  $\rho_X$ . Now let  $\rho_{X'Y'XY}$  be a purification of  $\rho_{XY}$ , then there is a Hilbert space  $\mathcal{H}$  and a unitary  $U : \mathcal{H}_{X'} \oplus \mathcal{H} \rightarrow \mathcal{H}_{X'Y'Y}$  such that

$$(I_X \otimes U)\tau_{X'X}(I_X \otimes U)^\dagger = \rho_{X'Y'XY}, \quad (34)$$

where  $\tau_{X'X}$  is extended to the orthogonal complement of  $\mathcal{H}_{X'}$  by zeros. Then, (30) implies

$$\|\tilde{\rho}_{X'Y'\tilde{X}Y}^{\otimes n} - (I_X \otimes U)^\dagger \tilde{\sigma}_{X'Y'\tilde{X}Y}^{\otimes n} (I_X \otimes U)^{\otimes n}\|_1 \leq \epsilon, \quad (35)$$

where  $\tilde{\rho}_{X'Y'\tilde{X}Y} = (\mathcal{N}_{X \rightarrow \tilde{X}} \otimes \mathcal{I}_{X'Y'Y})(\rho_{X'Y'XY})$ . Further, one computes that

$$(I_X \otimes U)^\dagger \tilde{\sigma}_{X'Y'\tilde{X}Y}^{\otimes n} (I_X \otimes U)^{\otimes n} = \tilde{\sigma}_{X'Y'Y'Y}^{\otimes n}, \quad (36)$$

$$\tilde{\sigma}_{X'Y'Y'Y}^{\otimes n} := (\mathcal{I}_{X'Y'Y'Y} \otimes \mathcal{F}_n)(\rho_{X'Y'XY}^{\otimes n}). \quad (37)$$

To summarize, we found that

$$\|\tilde{\rho}_{X'Y'\tilde{X}Y}^{\otimes n} - \tilde{\sigma}_{X'Y'Y'Y}^{\otimes n}\|_1 \leq \epsilon. \quad (38)$$

Using monotonicity of the trace norm under partial trace, we find that

$$\|\tilde{\rho}_{X'Y'\tilde{X}Y} - \tilde{\sigma}_{X'Y'\tilde{X}Y}\|_1 \leq \|\tilde{\rho}_{X'Y'\tilde{X}Y}^{\otimes n} - \tilde{\sigma}_{X'Y'Y'Y}^{\otimes n}\|_1. \quad (39)$$

Moreover,

$$\|\tilde{\rho}_{\tilde{X}} - \tilde{\sigma}_{\tilde{X}_i}\|_1 \leq \|\tilde{\rho}_{X'Y'\tilde{X}Y} - \tilde{\sigma}_{X'Y'\tilde{X}_i Y_i}\|_1. \quad (40)$$

again using the monotonicity of the trace norm under partial trace. This implies that  $\|\tilde{\rho}_{\tilde{X}} - \tilde{\sigma}_{\tilde{X}_i}\|_1 \leq \epsilon$ .

Now, by the Fannes Inequality the following bound holds for all  $1 \leq i \leq n$ :

$$\begin{aligned} |S(\tilde{\rho}_{\tilde{X}}) - S(\tilde{\sigma}_{\tilde{X}_i})| &\leq \log(k') \|\tilde{\rho}_{\tilde{X}} - \tilde{\sigma}_{\tilde{X}_i}\|_1 \\ &\quad + \frac{1}{\log(2)} \eta(\|\tilde{\rho}_{\tilde{X}} - \tilde{\sigma}_{\tilde{X}_i}\|_1), \end{aligned}$$

where  $k'$  is the dimension of  $\mathcal{H}_{\tilde{X}}$ . Then, using the bound  $\|\tilde{\rho}_{\tilde{X}} - \tilde{\sigma}_{\tilde{X}_i}\|_1 \leq \epsilon$ , we find for all  $1 \leq i \leq n$ ,

$$\begin{aligned} |S(\tilde{\rho}_{\tilde{X}}) - S(\tilde{\sigma}_{\tilde{X}_i})| &\leq \epsilon \log(k') + \frac{\eta(\epsilon)}{\log(2)} \\ &\leq \hat{k}' \eta(\epsilon), \end{aligned} \quad (41)$$

where the last inequality uses the fact that  $\eta(\epsilon) \geq \epsilon$  for  $0 \leq \epsilon < \frac{1}{e}$ , and where  $\hat{k}'$  is defined in terms of the constants in the first inequality, including  $k'$ .

With a similar method, one can prove that for all  $1 \leq i \leq n$ ,

$$|S(\tilde{\rho}_{\tilde{X}Y}) - S(\tilde{\sigma}_{\tilde{X}_i Y_i})| \leq \hat{k}'' \eta(\epsilon), \quad (42)$$

$$|S(\rho_Y) - S(\tilde{\sigma}_{Y_i})| \leq \hat{k}''' \eta(\epsilon), \quad (43)$$

where  $\hat{k}'', \hat{k}'''$  also depend on the dimensions of  $\mathcal{H}_{\tilde{X}Y}$  and  $\mathcal{H}_Y$ , respectively.

Combining Eq. (41), (42) and (43) we find for all  $1 \leq i \leq n$  that Eq. (34) holds for a constant  $k$  that includes the constants in the three estimates above and depends only on the dimension of  $\mathcal{H}_{\tilde{X}Y}$ .

Hence, we obtain for large enough  $n$  that  $\bar{I}_n(\tilde{X}; Y)_{\tilde{\sigma}} \geq J - k\eta(\epsilon)$ .  $\square$

## APPENDIX B NUMERICAL EXAMPLES

The aim of this appendix is to compute the communication rate as a function of  $J$  for some examples, using a numerical optimisation algorithm for evaluating the right-hand side of Eq. (11). In particular, in all these examples the rate function turns out to be convex in  $J$ . Consider the following normalized version of the rate

$$\hat{R}(J) := \min_{\mathcal{N}_{X \rightarrow \tilde{X}} : \frac{I(\tilde{X}; Y)_{\tilde{\rho}}}{I(\tilde{X}; X)_{\tilde{\rho}}} \geq J} \frac{I(X'; \tilde{X})_{\tilde{\tau}}}{I(X'; X)_{\tilde{\tau}}}, \quad (44)$$

where now  $0 < J < 1$ . In the following examples the systems  $X$  and  $Y$  are described by two-dimensional Hilbert spaces spanned by the basis  $|\uparrow\rangle, |\downarrow\rangle$ .

To find the optimum numerically, a simple random search algorithm is used [1]. It initially chooses a number of channels at random (in terms of their Kraus operators) and computes the related mutual information, then randomly varies those channels with the lowest  $I(X', \tilde{X})$  further until a stable optimum is reached.

This algorithm is applied to three classes of input states  $\rho_{XY}$ : The first example is a ‘‘classical’’ state, i.e., a state without entanglement between the systems  $X$  and  $Y$ , given by the density matrix

$$\rho_{XY}^{(1)} := p_1 |\uparrow\uparrow\rangle\langle\uparrow\uparrow| + p_2 |\downarrow\downarrow\rangle\langle\downarrow\downarrow| + p_3 |\uparrow\downarrow\rangle\langle\uparrow\downarrow| + p_4 |\downarrow\uparrow\rangle\langle\downarrow\uparrow|,$$

where  $p_1, p_2, p_3, p_4$  are nonnegative numbers with  $p_1 + p_2 + p_3 + p_4 = 1$ . The second example is a state with entanglement between  $X$  and  $Y$ , namely,

$$\rho_{XY}^{(2)} = \frac{1}{2} |\uparrow\uparrow\rangle\langle\uparrow\uparrow| + \frac{1}{4} |\uparrow\uparrow\rangle\langle\downarrow\downarrow| + \frac{1}{4} |\downarrow\downarrow\rangle\langle\uparrow\uparrow| + \frac{1}{2} |\downarrow\downarrow\rangle\langle\downarrow\downarrow|.$$

Finally, the third example is again a state with entanglement defined as

$$\rho_{XY}^{(3)} = p_1 |v\rangle\langle v| + p_2 |w\rangle\langle w|$$

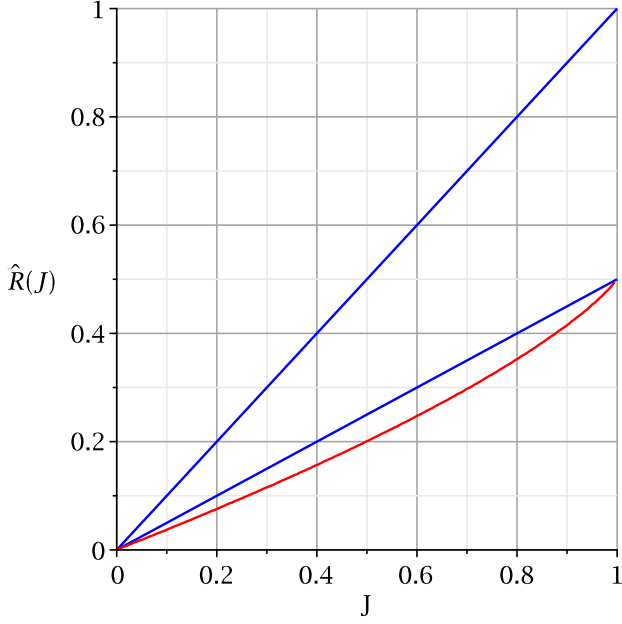


Fig. 2. The function  $\hat{R}(J)$  (in red) for the initial state  $\rho_{XY}^{(1)}$  with  $(p_1, p_2, p_3, p_4) = (0.1, 0.2, 0.3, 0.4)$ .

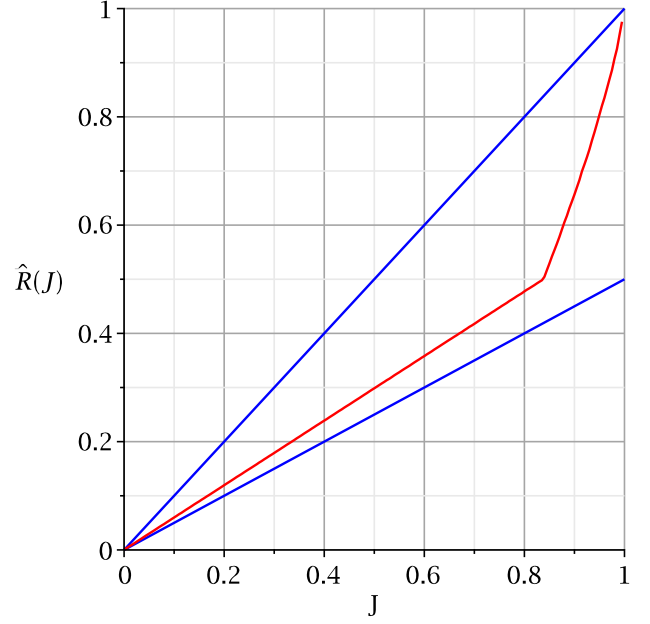


Fig. 4. The function  $\hat{R}(J)$  (in red) for the initial state  $\rho_{XY}^{(3)}$  with  $(p_1, p_2) = (0.4, 0.6)$ .

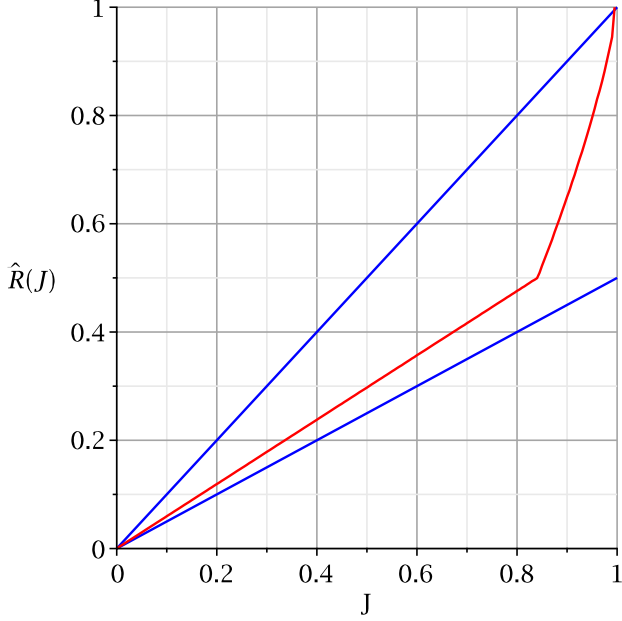


Fig. 3. The function  $\hat{R}(J)$  (in red) for the initial state  $\rho_{XY}^{(2)}$ .

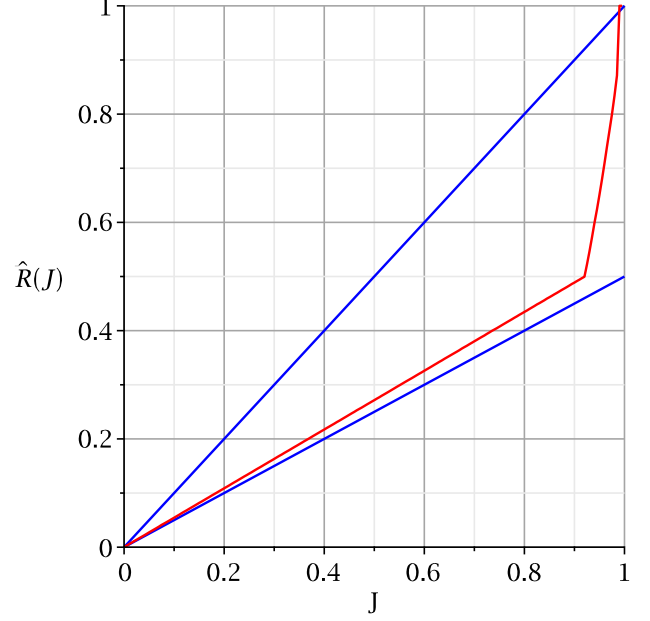


Fig. 5. The function  $\hat{R}(J)$  (in red) for the initial state  $\rho_{XY}^{(3)}$  with  $(p_1, p_2) = (0.2, 0.8)$ .

with the normalized vectors

$$v = \frac{1}{\sqrt{2}}(|\uparrow\uparrow\rangle + |\downarrow\downarrow\rangle), \quad w = |\downarrow\downarrow\rangle$$

and nonnegative numbers  $p_1, p_2$  with  $p_1 + p_2 = 1$ . The plots presented in Figs. 2–5 show the rate as a function of distortion. The blue lines correspond to the curves  $\hat{R}(J) = J$  and  $\hat{R}(J) = \frac{1}{2}J$  and are introduced for comparison with the actual result (red line).

The four plots show that the function  $\hat{R}(J)$ , given by (44), is indeed a convex function in  $J$  for the specific choices of initial state  $\rho_{XY}$  and within the limits of the numerical

approximation. Note that in Figs. 3–5, the graph does not appear to be differentiable at the point  $\hat{R} = \frac{1}{2}$ . This seems to be a common feature of the examples  $\rho_{XY}^{(2)}, \rho_{XY}^{(3)}$ , but we do not currently have an analytic explanation for this behaviour. Note that in Fig. 2, one has  $M(1) = \frac{1}{2}$ , while in Figs. 3–5, one has  $M(1) = 1$ . In other words, in the case of a “classical” (non entangled) state  $\rho_{XY}$ , there is a channel such that  $I(\tilde{X}; Y) = I(X; Y)$  and  $I(\tilde{X}; X') = \frac{1}{2}I(X; X')$ . To obtain an analytic expression of this channel we proceed as follows.

The initial state is  $\rho_{XY}^{(1)}$ . Then,  $\rho_X^{(1)} = \text{Tr}_Y \rho_{XY}^{(1)} = (p_1 +$



$p_3)|\uparrow\rangle\langle\uparrow| + (p_2 + p_4)|\downarrow\rangle\langle\downarrow|$  and a purification is given by  $\tau_{X'X}^{(1)} = |w\rangle\langle w|$  with  $|w\rangle = \sqrt{p_1 + p_3}|\uparrow\uparrow\rangle + \sqrt{p_2 + p_4}|\downarrow\downarrow\rangle$ . Our ansatz for the channel is

$$\mathcal{N}_{X \rightarrow \tilde{X}}(\rho) = \frac{1}{2} \sum_{i=1}^2 K_i \rho K_i^\dagger \quad (45)$$

with  $K_1 := |\uparrow\rangle\langle\uparrow| - |\downarrow\rangle\langle\downarrow|$  and  $K_2 := |\uparrow\rangle\langle\uparrow| + |\downarrow\rangle\langle\downarrow|$ . Applying this channel to  $\rho_{XY}^{(1)}$  and  $\tau_{X'X}^{(1)}$ , we obtain  $\tilde{\rho}_{\tilde{X}Y}^{(1)} = \rho_{XY}^{(1)}$  and

$$\tilde{\tau}_{X'\tilde{X}}^{(1)} = (p_1 + p_3)|\uparrow\uparrow\rangle\langle\uparrow\uparrow|_{X'\tilde{X}} + (p_2 + p_4)|\downarrow\downarrow\rangle\langle\downarrow\downarrow|_{X'\tilde{X}}.$$

Note that here  $\tilde{\tau}_{\tilde{X}}^{(1)} = \tau_{X'}^{(1)} = \tau_X^{(1)} = \rho_X^{(1)}$  are all matrices, but they act on different Hilbert spaces. With this choice of channel and initial state it is clear that  $I(\tilde{X}; Y)_{\tilde{\rho}^{(1)}} = I(X; Y)_{\rho^{(1)}}$ . In order to show that  $I(\tilde{X}; X')_{\tilde{\tau}^{(1)}} = \frac{1}{2}I(X; X')_{\tau^{(1)}}$  we compute the von Neumann entropies in  $I(\tilde{X}; X') = S(\tilde{X}) + S(X') - S(\tilde{X}X')$  and in  $I(X; X') = S(X) + S(X') - S(XX')$  using the fact that  $S(XX') = 0$ , since  $\tau_{X'X}$  is pure, and that  $\rho_X^{(1)}, \tilde{\tau}_{X'\tilde{X}}^{(1)}$  are diagonal. After evaluating the matrix functions, we obtain

$$\begin{aligned} I(\tilde{X}; X') &= - (p_1 + p_3) \log(p_1 + p_3) \\ &\quad - (p_2 + p_4) \log(p_2 + p_4) \\ &= \frac{1}{2} I(X; X') \end{aligned} \quad (46)$$

as desired. Therefore,  $\hat{R}(1) \leq 1/2$  for this class of states, consistent with the graph.

This somewhat unexpected feature may be understood as follows: In order to transmit the  $X$  part of the non-entangled state  $\rho_{XY}^{(1)}$  perfectly, a *classical* channel of 1 bit capacity is sufficient. By the usual quantum teleportation result, this corresponds to a quantum channel of only  $\frac{1}{2}$  qbit capacity, if shared entanglement is available in abundance.

While the simplistic random search algorithm that we used in this appendix is sufficient for demonstrating the essential quantum features of the system, it would be too inefficient to compute the rate function  $R(J)$  for systems with larger Hilbert space dimension. More sophisticated numerical schemes would be required to that end, e.g., taking the gradient of the objective function into account. Appendix C describes a Lagrangian method that may be used to this end.

## APPENDIX C THE OPTIMAL MAP

In this appendix we describe a method for finding the optimum channel in Eq. (11). We first remark that the optimum is actually attained at the boundary of the constraint region,  $I(\tilde{X}; Y) = J$ . This is heuristically clear, as for a channel in the interior ( $I(\tilde{X}; Y) > J$ ) one could always reduce the channel's communication rate by a small amount at the expense of a slightly decreased quality measure  $I(\tilde{X}; Y)$ , staying inside the region of the constraint. We give a more formal argument to this end.

**Lemma 2.** *The minimum in Eq. (11) is attained on the hypersurface  $I(\tilde{X}; Y) = J$ .*

*Proof.* Suppose that the minimum was attained at a channel  $\mathcal{N}_1$  such that  $I(\tilde{X}; Y)_{\mathcal{N}_1} > J$  (here the subscript denotes that the state  $\tilde{\rho}_{\tilde{X}Y}$  is computed with respect to the channel  $\mathcal{N}_1$ ). Let  $\mathcal{N}_0$  be the channel  $\mathcal{N}_0(\sigma) := (\dim \mathcal{H}_{\tilde{X}})^{-1} \text{Tr}(\sigma) \cdot \mathbf{1}$ ; we note that  $I(\tilde{X}; Y)_{\mathcal{N}_0} = 0 = I(\tilde{X}; X')_{\mathcal{N}_0}$ . For  $0 < \lambda < 1$ , let us consider a new channel  $\mathcal{N}_\lambda := \lambda \mathcal{N}_1 + (1 - \lambda) \mathcal{N}_0$ . From convexity of the mutual information in the channel, it follows that

$$\begin{aligned} I(\tilde{X}; X')_{\mathcal{N}_\lambda} &\leq \lambda I(\tilde{X}; X')_{\mathcal{N}_1} + (1 - \lambda) I(\tilde{X}; X')_{\mathcal{N}_0} \\ &< I(\tilde{X}; X')_{\mathcal{N}_1}. \end{aligned} \quad (47)$$

On the other hand, from continuity considerations we know that  $I(\tilde{X}; Y)_{\mathcal{N}_\lambda} > J$  for  $\lambda$  close enough to 1, so that also  $\mathcal{N}_\lambda$  fulfills the constraint in this region. Hence  $M(J) < I(\tilde{X}; X')_{\mathcal{N}_1}$ , and  $\mathcal{N}_1$  is not the position of the minimum.  $\square$

Therefore, we deal with an optimisation problem of a function over a constraint hypersurface, which makes it useful to look for local extrema with the method of Lagrange multipliers. We use the Lagrangian in Eq. (21),

$$\begin{aligned} \mathcal{L} := & I(X'; \tilde{X})_{\tilde{\tau}_{X'\tilde{X}}} \\ & - \beta I(\tilde{X}; Y)_{\tilde{\rho}_{\tilde{X}Y}} - \text{Tr}_{X\tilde{X}}(\Psi_{X\tilde{X}}^{t_X}(\Lambda_X \otimes I_{\tilde{X}})). \end{aligned}$$

and look for zeros of its derivatives. We use the Choi-Jamiołkowski representation

$$\Psi_{X'\tilde{X}} := (\mathcal{I}_{X'} \otimes \mathcal{N}_{X \rightarrow \tilde{X}})(\Phi_{X'X}) \quad (48)$$

of the channel in order to compute the derivative of the Lagrangian, where  $\Phi_{X'X} := \sum_{i,j=0}^{d-1} |i\rangle\langle j|_{X'} \otimes |i\rangle\langle j|_X$  is the Choi-Jamiołkowski matrix corresponding to the identity channel from the Hilbert space  $\mathcal{H}_{X'}$  to  $\mathcal{H}_X$ , and  $\mathcal{N}_{X \rightarrow \tilde{X}}$  is the channel that simulates the compression-decompression process.  $\Lambda_X$ , an operator on the Hilbert space  $\mathcal{H}_X$ , is the Lagrange multiplier introduced for the normalisation of the channel  $\Psi_{X\tilde{X}}^{t_X}$ . Considering the definition of the mutual information, to compute the derivative  $\frac{\delta \mathcal{L}}{\delta \Psi_{X\tilde{X}}^{t_X}}$  of the Lagrangian, we need to compute the following derivatives,

$$\begin{aligned} & \frac{\delta S(X')_{\tilde{\tau}}}{\delta \Psi_{X\tilde{X}}^{t_X}}, \quad \frac{\delta S(Y)_{\tilde{\rho}}}{\delta \Psi_{X\tilde{X}}^{t_X}}, \\ & \frac{\delta S(\tilde{X})_{\tilde{\tau}}}{\delta \Psi_{X\tilde{X}}^{t_X}}, \quad \frac{\delta S(\tilde{X})_{\tilde{\rho}}}{\delta \Psi_{X\tilde{X}}^{t_X}}, \quad \frac{\delta S(X'\tilde{X})_{\tilde{\tau}}}{\delta \Psi_{X\tilde{X}}^{t_X}}, \\ & \frac{\delta S(\tilde{X}Y)_{\tilde{\rho}}}{\delta \Psi_{X\tilde{X}}^{t_X}}, \quad \frac{\delta \text{Tr}_{X\tilde{X}}(\Psi_{X\tilde{X}}(\Lambda_X \otimes I_{\tilde{X}}))}{\delta \Psi_{X\tilde{X}}^{t_X}}. \end{aligned} \quad (49)$$

Notice that the functions in the numerator of the expressions in the first two equations in (49) are independent of the channel and, hence, the derivatives are zero. For the five remaining ones we note that for a Hermitian operator,  $A$ , and a function,  $f$ , which is analytic on the spectrum of  $A$ , the directional derivative of  $\text{Tr}[f(A)]$  is given by

$$\frac{\delta \text{Tr}[f(A)]}{\delta A}[B] = \text{Tr}[f'(A)B], \quad (50)$$

with the direction given by the operator  $B$  and  $f'$  being the first derivative of the function  $f$ . (This follows from analytic functional calculus, expanding  $\text{Tr} f(A + \epsilon B)$  in a Taylor series

around  $\epsilon = 0$ .) Specifically, let us define  $f(z) := z \log(z)$ . Since the derivative of our function  $f$  is given by  $f'(\cdot) = (1 + \log)(\cdot)$ , using (50) and (22), we have

$$\frac{\delta S(\tilde{X})_{\tilde{\tau}}}{\delta \Psi_{X\tilde{X}}^{t_X}}[B_{X\tilde{X}}] = -\text{Tr}_{X'\tilde{X}} \left\{ [(I_{\tilde{X}} + \log \tilde{\tau}_{\tilde{X}}) \otimes I_{X'}] \mathbb{E} \right\}, \quad (51)$$

where

$$\mathbb{E} := \text{Tr}_X (B_{X\tilde{X}} \tau_{X'X}). \quad (52)$$

Likewise, we can compute

$$\frac{\delta S(X'\tilde{X})_{\tilde{\tau}}}{\delta \Psi_{X\tilde{X}}^{t_X}}[B_{X\tilde{X}}] = -\text{Tr}_{X'\tilde{X}} \left\{ [(I_{X'\tilde{X}} + \log \tilde{\tau}_{X'\tilde{X}})] \mathbb{E} \right\}. \quad (53)$$

In order to compute the derivative for  $S(\tilde{X})_{\tilde{\rho}}$  and  $S(\tilde{X}Y)_{\tilde{\rho}}$  we use Eqs. (24) and (50), and we find

$$\frac{\delta S(\tilde{X})_{\tilde{\rho}}}{\delta \Psi_{X\tilde{X}}^{t_X}}[B_{X\tilde{X}}] = -\text{Tr}_{\tilde{X}Y} \left\{ [(I_{\tilde{X}} + \log \tilde{\rho}_{\tilde{X}}) \otimes I_Y] \mathbb{G} \right\} \quad (54)$$

and

$$\frac{\delta S(\tilde{X}Y)_{\tilde{\rho}}}{\delta \Psi_{X\tilde{X}}^{t_X}}[B_{X\tilde{X}}] = -\text{Tr}_{\tilde{X}Y} \left\{ (I_{\tilde{X}Y} + \log \tilde{\rho}_{\tilde{X}Y}) \mathbb{G} \right\}, \quad (55)$$

where

$$\mathbb{G} := \text{Tr}_X (B_{X\tilde{X}} \rho_{XY}). \quad (56)$$

For the last derivative we have

$$\frac{\delta \text{Tr}_{X\tilde{X}} \left\{ \Psi_{X\tilde{X}}^{t_X} (\Lambda_X \otimes I_{\tilde{X}}) \right\}}{\delta \Psi_{X\tilde{X}}^{t_X}}[B_{X\tilde{X}}] = \text{Tr}_{X\tilde{X}} \left\{ (\Lambda_X \otimes I_{\tilde{X}}) B_{X\tilde{X}} \right\}.$$

Putting all the terms together we have

$$\begin{aligned} \frac{\delta \mathcal{L}}{\delta \Psi_{X\tilde{X}}^{t_X}}[B_{X\tilde{X}}] &= \frac{\delta S(\tilde{X})_{\tilde{\tau}}}{\delta \Psi_{X\tilde{X}}^{t_X}}[B_{X\tilde{X}}] - \frac{\delta S(X'\tilde{X})_{\tilde{\tau}}}{\delta \Psi_{X\tilde{X}}^{t_X}}[B_{X\tilde{X}}] \\ &\quad - \beta \frac{\delta S(\tilde{X})_{\tilde{\rho}}}{\delta \Psi_{X\tilde{X}}^{t_X}}[B_{X\tilde{X}}] + \beta \frac{\delta S(\tilde{X}Y)_{\tilde{\rho}}}{\delta \Psi_{X\tilde{X}}^{t_X}}[B_{X\tilde{X}}] \\ &\quad - \frac{\delta \text{Tr}_{X\tilde{X}} (\Psi_{X\tilde{X}}^{t_X} (\Lambda_X \otimes I_{\tilde{X}}))}{\delta \Psi_{X\tilde{X}}^{t_X}}[B_{X\tilde{X}}] \\ &= \text{Tr}_{X'\tilde{X}} \left\{ \left[ -I_{X'} \otimes (I_{\tilde{X}} + \log \tilde{\tau}_{\tilde{X}}) \right. \right. \\ &\quad \left. \left. + (I_{X'\tilde{X}} + \log \tilde{\tau}) + \beta I_{X'} \right. \right. \\ &\quad \left. \left. \otimes (I_{\tilde{X}} + \log \tilde{\tau}_{\tilde{X}})] \mathbb{E} \right\} \\ &\quad - \beta \text{Tr}_{\tilde{X}Y} \left\{ (I_{\tilde{X}Y} + \log \tilde{\rho}_{\tilde{X}Y}) \mathbb{G} \right\} \\ &\quad - \text{Tr}_{X\tilde{X}} \left\{ (\Lambda_X \otimes I_{\tilde{X}}) B_{X\tilde{X}} \right\}. \end{aligned} \quad (57)$$

Setting this expression to zero ( $\frac{\delta \mathcal{L}}{\delta \Psi_{X\tilde{X}}^{t_X}}[B_{X\tilde{X}}] = 0$ ), we find

$$\begin{aligned} \text{Tr}_{X'\tilde{X}} \left\{ \log \tilde{\tau}_{X'\tilde{X}} \mathbb{E} \right\} &= \\ \text{Tr}_{X'\tilde{X}} \left\{ \left[ I_{X'} \otimes \log \tilde{\tau}_{\tilde{X}} - \beta I_{X'} \otimes \log \tilde{\tau}_{\tilde{X}} \right] \mathbb{E} \right\} & \\ + \beta \text{Tr}_{\tilde{X}Y} \left\{ \log \tilde{\rho}_{\tilde{X}Y} \mathbb{G} \right\} + \text{Tr}_{X\tilde{X}} \left\{ (\Lambda_X \otimes I_{\tilde{X}}) B_{X\tilde{X}} \right\}. & \end{aligned} \quad (58)$$

Rearranging left and right hand sides of this equation, we find

$$\begin{aligned} \text{Tr}_{X\tilde{X}} \left\{ B_{X\tilde{X}} \text{Tr}_{X'} \left\{ \tau_{X'X} (\log \tilde{\tau}_{X'\tilde{X}} \otimes I_X) \right\} \right\} &= \\ \text{Tr}_{X\tilde{X}} \left\{ B_{X\tilde{X}} \text{Tr}_{X'} \left\{ \tau_{X'X} [(I_{X'} \otimes \log \tilde{\tau}_{\tilde{X}} - \beta I_{X'} \otimes \log \tilde{\tau}_{\tilde{X}}) \right. \right. \right. \\ \left. \left. \otimes I_X] \right\} \right\} + \beta \text{Tr}_{X\tilde{X}} \left\{ B_{X\tilde{X}} \text{Tr}_Y \left\{ \rho_{XY} (\log \tilde{\rho}_{\tilde{X}Y} \otimes I_X) \right\} \right\} & \\ + \text{Tr}_{X\tilde{X}} \left\{ B_{X\tilde{X}} (\Lambda_X \otimes I_{\tilde{X}}) \right\}. & \end{aligned}$$

This holds for all directions  $B_{X\tilde{X}}$ , which implies

$$\begin{aligned} \text{Tr}_{X'} \left\{ \tau_{X'X} (\log \tilde{\tau}_{X'\tilde{X}} \otimes I_X) \right\} &= \\ \text{Tr}_{X'} \left\{ \tau_{X'X} [(I_{X'} \otimes \log \tilde{\tau}_{\tilde{X}} - \beta I_{X'} \otimes \log \tilde{\tau}_{\tilde{X}}) \right. \\ \left. - \beta I_{X'} \otimes \log \tilde{\tau}_{\tilde{X}}] \otimes I_X \right\} & \\ + \beta \text{Tr}_Y \left\{ \rho_{XY} (\log \tilde{\rho}_{\tilde{X}Y} \otimes I_X) \right\} & \\ + \Lambda_X \otimes I_{\tilde{X}}. & \end{aligned} \quad (59)$$

By performing the partial trace on the left hand side of this expression, we obtain

$$\begin{aligned} \rho_X^{1/2} (\log \tilde{\tau}_{X\tilde{X}})^{t_X} \rho_X^{1/2} &= \tau_X (\log \tilde{\tau}_{\tilde{X}} - \beta \log \tilde{\tau}_{\tilde{X}}) \\ &\quad + \beta \text{Tr}_Y \left\{ \rho_{XY} (\log \tilde{\rho}_{\tilde{X}Y} \otimes I_X) \right\} \\ &\quad + \Lambda_X \otimes I_{\tilde{X}}. \end{aligned} \quad (60)$$

Simplifying this expression further, we find

$$\begin{aligned} (\log \tilde{\tau}_{X\tilde{X}})^{t_X} &= I_X \otimes (\log \tilde{\tau}_{\tilde{X}} - \beta \log \tilde{\tau}_{\tilde{X}}) \\ &\quad + \beta \text{Tr}_Y \left\{ \rho_X^{-1/2} \rho_{XY} \rho_X^{-1/2} (\log \tilde{\rho}_{\tilde{X}Y} \otimes I_X) \right\} \\ &\quad + \rho_X^{-1/2} \Lambda_X \rho_X^{-1/2} \otimes I_{\tilde{X}}. \end{aligned} \quad (61)$$

Let us denote

$$\begin{aligned} D_{X\tilde{X}}^{\beta Y} &:= \beta I_X \otimes \log \tilde{\tau}_{\tilde{X}} \\ &\quad - \beta \text{Tr}_Y \left\{ \rho_X^{-1/2} \rho_{XY} \rho_X^{-1/2} (\log \tilde{\rho}_{\tilde{X}Y} \otimes I_X) \right\}, \end{aligned} \quad (62)$$

and the normalisation term  $\tilde{\Lambda}_X := \rho_X^{-1/2} \Lambda_X \rho_X^{-1/2}$ . Exponentiating both sides of Eq. (61), we obtain

$$\tilde{\tau}_{X\tilde{X}}^{t_X} = e^{\log \tilde{\tau}_{\tilde{X}} \otimes I_X - D_{X\tilde{X}}^{\beta Y} + \tilde{\Lambda}_X \otimes I_{\tilde{X}}}. \quad (63)$$

Using Eq. (22), we arrive at the expression for the Choi-Jamiołkowski matrix corresponding to the channel,

$$\begin{aligned} \Psi_{X\tilde{X}}^{t_X} &= (\rho_X \otimes I_{\tilde{X}})^{-1/2} e^{\log \tilde{\tau}_{\tilde{X}} \otimes I_X - D_{X\tilde{X}}^{\beta Y} + \tilde{\Lambda}_X \otimes I_{\tilde{X}}} \\ &\quad (\rho_X \otimes I_{\tilde{X}})^{-1/2}. \end{aligned} \quad (64)$$

Note that this is an implicit equation in  $\Psi_{X\tilde{X}}$  since it also appears on the right hand side of this expression. To find the optimum channel, Eq. (64) needs to be solved iteratively for  $\Psi_{X\tilde{X}}$ . Note that the unknown Lagrange multiplier  $\Lambda_X$ , which is associated with the normalisation constraint, is still contained in this equation. An algorithm that recursively computes the channel might work as follows. Starting with a guess for the channel  $\Psi_{X\tilde{X}}$  and normalising it, this guess is inserted into Eq. (60) to compute a self-consistent value for  $\Lambda_X$ . This would allow to compute  $\tilde{\rho}_{\tilde{X}}$  and  $\tilde{\rho}_{\tilde{X}Y}$  from the channel, and

hence give an approximation for all quantities that enter the right hand side of Eq. (26). Thus, a new approximation for the left-hand side is obtained, i.e., the channel  $\Psi_{X\tilde{X}}$ . When this procedure is repeated iteratively, an optimal channel  $\Psi_{X\tilde{X}}$  is obtained at a given value for  $\beta$ . Repeating this procedure for different values of  $\beta$  and optimising under the constraint  $I(\tilde{X}; Y)_{\tilde{\rho}} \geq J$  yields the minimum in (11).

**Sina Salek** After completing his PhD at the University of Bristol, Sina Salek became a postdoctoral research fellow at the University of Hong Kong, before joining the Department of Computer Science, the University of Oxford as a postdoctoral researcher. His research interest is in quantum information theory and quantum foundations.

**Daniela Cadamuro** Daniela Cadamuro leads an Emmy Noether junior research group at the Institute for Theoretical Physics at the University of Leipzig. She received a PhD in theoretical physics from the Georg-August-Universität Göttingen in Germany in 2012. She held postdoctoral positions at the University of Bristol and at the Technische Universität München (TUM). Her research interests lie in mathematical aspects of quantum theory, in particular quantum field theory.

**Philipp Kammerlander** Philipp Kammerlander received the B.Sc. and M.Sc. degrees in Physics from ETH Zurich in 2012 and 2013. He is currently pursuing a Ph.D. degree at the Institute for Theoretical Physics at ETH Zurich. His research interests include classical and quantum thermodynamics and quantum information theory.

**Karoline Wiesner** Karoline Wiesner is Associate Professor in the School of Mathematics at the University of Bristol. She received a PhD in physics from Uppsala University in Sweden in 2004. She was a postdoctoral fellow at the Santa Fe Institute and the University of California, Davis. In 2007 she joined the faculty of the School of Mathematics at the University of Bristol. Her research interests lie in quantum and classical information theory and complex systems, ranging from physics to the social sciences.