No-Signalling Assisted Zero-Error Capacity of Quantum Channels and an Information Theoretic Interpretation of the Lovász Number

Runyao Duan[1,2,3,*] and Andreas Winter[4,5,†]

[1]*Centre for Quantum Computation and Intelligent Systems (QCIS), Faculty of Engineering and Information Technology, University of Technology, Sydney, NSW 2007, Australia*

[2]*State Key Laboratory of Intelligent Technology and Systems, Tsinghua National Laboratory for Information Science and Technology, Department of Computer Science and Technology, Tsinghua University, Beijing 100084, China*

[3]*UTS-AMSS Joint Research Laboratory for Quantum Computation and Quantum Information Processing, Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing 100190, China*

[4]*ICREA & Física Teòrica: Informació i Fenòmens Quàntics, Universitat Autònoma de Barcelona, ES-08193 Bellaterra (Barcelona), Spain*

[5]*School of Mathematics, University of Bristol, Bristol BS8 1TW, United Kingdom*

(Dated: 5 February 2015)

We study the one-shot zero-error classical capacity of a quantum channel assisted by quantum no-signalling correlations, and the reverse problem of exact simulation of a prescribed channel by a noiseless classical one. Quantum no-signalling correlations are viewed as two-input and two-output completely positive and trace preserving maps with linear constraints enforcing that the device cannot signal. Both problems lead to simple semidefinite programmes (SDPs) that depend only the Kraus operator space of the channel. In particular, we show that the zero-error classical simulation cost is precisely the conditional min-entropy of the Choi-Jamiołkowski matrix of the given channel. The zero-error classical capacity is given by a similar-looking but different SDP; the asymptotic zero-error classical capacity is the regularization of this SDP, and in general we do not know of any simple form.

Interestingly however, for the class of classical-quantum channels, we show that the asymptotic capacity is given by a much simpler SDP, which coincides with a semidefinite generalization of the fractional packing number suggested earlier by Aram Harrow. This finally results in an operational interpretation of the celebrated Lovász ϑ function of a graph as the zero-error classical capacity of the graph assisted by quantum no-signalling correlations, the first information theoretic interpretation of the Lovász number.

* runyao.duan@uts.edu.au
† andreas.winter@uab.cat

CONTENTS

I. INTRODUCTION

We choose as the starting point of the present work the fundamental problem of channel simulation. Roughly speaking, this problem asks when a communication channel $\mathcal{N}$ from Alice (A) to Bob (B) can be used to simulate another channel $\mathcal{M}$, also from A to B? [1] This problem has many variants according to the resources available to A and B. In particular, the case when A and B can access unlimited amount of shared entanglement has been completely solved. Let $C_E(\mathcal{N})$ denote the e ntanglement-assisted classical capacity of $\mathcal{N}$ [2]. It was shown that, in the asymptotic setting, to optimally simulate $\mathcal{M}$, we need to apply $\mathcal{N}$ at rate $\frac{C_E(\mathcal{M})}{C_E(\mathcal{N})}$ [3, 4]. In other words, the entanglement-assisted classical capacity uniquely determines the properties of the channel in the simulation process. Furthermore, even with stronger resources such as no-signalling correlations or feedback, this rate cannot be improved – otherwise we would violate causality, see [3] for a discussion.

Here we are interested in the zero-error setting [5]. It is well known that the zero-error communication problem is extremely difficult, already for classical channels. Indeed, the single-shot zero-error classical communication capability of a classical noisy channel equals the independence number of the (classical) confusability graph induced by the channel, and the latter problem is known to be NP-complete. The behaviour of quantum channels in zero-error communication is even more complex as striking effects such as super-activation are possible [6–9]. The most general zero-error simulation problem remains wide open. To overcome this difficulty, many variants of this problem have been proposed. The most natural way is to introduce some additional resources and see how this changes the capacity. Indeed, extra resources such as classical feedback [5], entanglement [7, 10, 11], and even a small (constant) amount of forward communication [12], have been introduced. It has been shown these extra resources can increase the capacity, and generally simplify the problem. In particular, it was shown that even for classical communication channel, shared entanglement can strictly increase the asymptotic zero-error classical capacity [15]. However, determining the entanglement-assisted zero-error classical capacity remains an open problem even for classical channels. More powerful resources are actually required in order to simplify the problem. Cubitt *et al.* [12] introduced classical no-signalling correlations into the zero-error communication for classical channels, and showed that the well-known fractional packing number of the bipartite graph induced by the channel, gives precisely the zero-error classical capacity of the channel. Previously, it was known by Shannon that this fractional packing number corresponds to the zero-error classical capacity of the channel when assisted with a feedback link from the receiver to the sender and when the unassisted zero-error classical capacity is not vanishing [5]. For general background on graph theory see [13], and for "fractional graph theory" the delightful book [14].

Another major motivation for this work is to further explore the connection between quantum information theory and the so-called "non-commutative graph theory" suggested in [11]. Such a connection has been well known in classical information theory. In [5], Shannon realized that the zero-error capacity of a classical noisy channel only depends on the confusability graph induced by the channel. He further pointed out that in the presence of classical feedback, the zero-error capacity is completely determined by the bipartite graph of possible input-output transitions associated with the channel. Thus it makes sense to talk about the zero-error capacity of a (bipartite) graph. The notion of non-commutative graph naturally occurs when we use quantum channels for zero-error communication. For any quantum channel, the non-commutative graph associated with the channel captures the zero-error communication properties, thus playing a similar role to confusability graph. Most notably, this notion also makes it possible to introduce a quantum Lovász ϑ function to upper bound the entanglement-assisted zero-error capacity that has properties quite similar to its classical analogue [11]. Very recently, it was shown that the zero-error classical capacity of a quantum channel in the presence of quantum feedback only depends on the Kraus operator space of the channel [16]. In other words, the Kraus operator space plays a role that is quite similar to the bipartite graph. Now it becomes clear that any classical channel induces a bipartite graph as well as a confusability graph, while a quantum channel induces a non-commutative bipartite graph and a non-commutative graph. The new insight is that we can simply regard a non-commutative (bipartite) graph as a high-level abstraction of all underlying quantum channels, and study its information-theoretic properties, not limited to zero-error setting. This leads us to a very general viewpoint: graphs as communication channels. For instance, we can define the entanglement-assisted classical capacity of a non-commutative bipartite graph as the minimum of the entanglement-assisted classical capacity of quantum channels that induce the given Kraus operator space. It was shown that this quantity enjoys a number of interesting properties including additivity under tensor product and an operational interpretation as a sort of entanglement-assisted conclusive capacity of the bipartite graph [16]. It remains a great challenge to find tractable forms of various capacities for non-commutative (bipartite) graphs.

In this paper we consider a more general class of quantum no-signalling correlations described by two-input and two-output quantum channels with the no-signalling constraints. This kind of correlations naturally arises in the study of the relativistic causality of quantum operations [17–

19]; see also the more recent [20]. Distinguishability of these correlations from an information theoretic viewpoint has also been studied [21]. We provide a number of new properties of these correlations, and establish several structural theorems of these correlations. Then we generalize the approach of [12] to study the zero-error classical capacity of a noisy quantum channel assisted by quantum no-signalling correlations, and the reverse problem of perfect simulation. We show that both problems can be completely solved in the one-shot scenario, revealing some nice structure:

1. The answers are given by semidefinite programmes (SDPs, cf. [22]);

2. At the same time they generalize the results of Cubitt *et al.* [12];

3. For the simulation, the question is really how to form a constant channel by a convex combination of the one we want to simulate and an arbitrary other quantum channel, and the number of bits needed is just $-\log p$, where p is the probability weight of the target channel in the convex combination (throughout this paper, $\log$ denotes the binary logarithm);

4. For assisted communication, there is an analogous problem of convex-combing a certain channel from B to A which has some kind of orthogonality relation with the given channel from A to B, with another one to form a constant channel. If the target channel has weight p, then the number of bits sent is again $-\log p$.

Most interestingly, the solution to the communication problem only depends on the Kraus operator space of the channel, not directly on the channel itself. For the simulation problem, the solution is given by the conditional min-entropy [24, 25] of the channel's Choi-Jamiołkowski matrix, and is actually additive, thus also gives the asymptotic cost of simulating the channel. If we are interested in simulating the cheapest channel contained in the Kraus operator space, we obtain a SDP in terms of the projection of the Choi-Jamiołkowski matrix. Both the capacity and the simulation SDPs are in general not known to be multiplicative under the tensor product of channels, thus we do not know the optimal asymptotic simulation cost.

We then focus on the asymptotic zero-error classical capacity and simulation cost assisted with quantum no-signalling correlations. This requires determining the asymptotic behaviour of a sequence of SDPs. In general the one-shot solution does not give the asymptotic result, since the corresponding SDP is not multiplicative with respect to the tensor product of channels. A simple formula for the asymptotic channel capacity remains unknown. However, for the special cases of classical-quantum (cq) channels, we find that the zero-error capacity is given by the solution of a rather simple SDP suggested earlier by Harrow as a natural generalization of the classical fractional packing number [26], which we call *semidefinite packing number*. This result has two interesting corollaries. First, it implies that the zero-error classical capacity of cq-channels assisted by quantum no-signalling correlations is additive. Second, and more importantly, we show that for a classical graph G, the celebrated Lovász number $\vartheta(G)$ [27], is actually the minimum zero-error classical capacity of any cq-channel that has the given graph as its confusability graph. In other words, Lovász' ϑ function is the zero-error classical capacity of a graph assisted by quantum no-signalling correlations. To the best of our knowledge, this is the first information theoretic operational interpretation of the Lovász number since its introduction in 1979. Previously, it was known that it is an upper bound on the entanglement-assisted zero-error classical capacity of a graph [11, 29]. It remains unknown whether the use of quantum no-signalling correlations could be replaced by shared entanglement. The asymptotic simulation cost for Kraus operator spaces associated with cq-channels is rather simpler, and is actually given by the one-shot simulation cost.

Before we proceed to the technical details, it may be helpful to present an overview of our main results. Let $\mathcal{N}$ be a quantum channel from $\mathcal{L}(A')$ to $\mathcal{L}(B)$, with a Kraus operator sum representation $\mathcal{N}(\rho) = \sum_k E_k \rho E_k^\dagger$, where $\sum_k E_k^\dagger E_k = \mathbb{1}_A$. Let $K = K(\mathcal{N}) = \mathrm{span}\{E_k\}$ denote the Kraus operator space of $\mathcal{N}$. The Choi-Jamiołkowski matrix of $\mathcal{N}$ is given by $J_{AB} = \sum_{ij} |i\rangle\langle j|_A \otimes \mathcal{N}(|i\rangle\langle j|_{A'}) = (\mathrm{id}_A \otimes \mathcal{N})\Phi_{AA'}$, A and A' are isomorphic Hilbert spaces, and $\{|i\rangle\}$ (or $\{|j\rangle\}$) is orthonormal basis over A (A'), $\Phi_{AA'}$ is the unnormalized maximally entangled state over $A \otimes A'$. Let P_{AB} denote the projection onto the support of J_{AB}, which is the subspace $(\mathbb{1} \otimes K)|\Phi\rangle$.

It is worth noting that many results below can be defined on any matrix subspace K, not just those corresponding to a quantum channel $\mathcal{N}$. However, we have to make sure that K is actually corresponding to some quantum channel $\mathcal{N}$. This puts additional constraint on K. More precisely, suppose $K = \mathrm{span}\{E_k\}$ for some orthonormal basis $\{E_k\}$ such that $\mathrm{Tr}\, E_j^\dagger E_k = \delta_{jk}$. Then we should be able to find a quantum channel $\mathcal{N}' = \sum_j F_j \cdot F_j^\dagger$ such that $F_j = \sum_k a_{jk} E_k$, for some invertible matrix $a = [a_{jk}]$ and $\sum_j F_j^\dagger F_j = \mathbb{1}_{A'}$. This is equivalent to

$$\sum_{kl} r_{kl} E_k^\dagger E_l = \mathbb{1}_{A'}, \text{ where } r_{kl} = \sum_j a_{jk}^* a_{jl}.$$

If such a positive definite matrix $r = [r_{kl}]$ cannot be found, K will not correspond to a quantum channel (nor any $K^{\otimes n}$ for $n \geq 1$). In this case one might still be able to find $K' \subset K$ that is a Kraus operator space for some quantum channel $\mathcal{N}'$. For instance, $K = \mathrm{span}\{\mathbb{1}, |0\rangle\langle 1|\}$, and $K' = \mathrm{span}\{\mathbb{1}\}$. (Note however that K is the limit of the Kraus subspaces of genuine quantum channels, namely amplitude damping channels with damping parameter going to 0; hence it might still be considered as admissible Kraus space of an infinitesimally amplitude damping channel.) We will, therefore, always assume that K corresponds to some quantum channel $\mathcal{N}$ such that $K = K(\mathcal{N})$. From now on, any such Kraus operator space K will be alternatively called "non-commutative bipartite graph" – in fact, below we shall argue why it is a natural generalization of bipartite graphs.

Theorem 1 *The one-shot zero-error classical capability (quantified as the largest number of messages) of $\mathcal{N}$ assisted by quantum no-signalling correlations depends only on the non-commutative graph K, and is given by the integer part of the following SDP:*

$$\begin{aligned}\Upsilon(\mathcal{N}) &= \Upsilon(K) \\ &= \max \mathrm{Tr}\, S_A \;\; s.t. \;\; 0 \leq U_{AB} \leq S_A \otimes \mathbb{1}_B, \mathrm{Tr}_A U_{AB} = \mathbb{1}_B, \mathrm{Tr}\, P_{AB}(S_A \otimes \mathbb{1}_B - U_{AB}) = 0.\end{aligned}$$

Hence we are motivated to call $\Upsilon(K)$ the no-signalling assisted independence number of K.

The proof of this theorem will be given in Section III A, where we also explore other properties of the above SDP. For another direction of investigation, looking at the specific channel $\mathcal{N}$ with Choi-Jamiołkowski matrix J and trying to minimize the error probability for given number of messages, we refer the reader to the recent and highly relevant work of Leung and Matthews [23], where exactly this is done in an environment with free no-signalling resources subject other semidefinite constraints.

It is evident from this theorem that the one-shot zero-error classical capacity of $\mathcal{N}$ only depends on the Kraus operator space K. That is, any two quantum channels $\mathcal{N}$ and $\mathcal{M}$ will have the same capacity if they have the same Kraus operator space. For this reason, we usually use $\Upsilon(K)$ to denote $\Upsilon(\mathcal{N})$. Furthermore, we can talk about the capacity of the Kraus operator space K directly without referring to the underling quantum channel. Notice that a classical channel $N = (X, p(y|x), Y)$ naturally induces a bipartite graph $\Gamma(N) = (X, E, Y)$, where the input and output alphabets X and Y are the two sets of vertices, and $E \subset X \times Y$ is the set of edges such that

$(x,y) \in E$ if and only if $p(y|x) > 0$. We shall also use the notation $\Gamma(y|x) = 1$ if $(x,y) \in E$, and $\Gamma(y|x) = 0$ otherwise. In this case, we have

$$K(N) = \operatorname{span}\{|y\rangle\langle x| : N(y|x) > 0\},$$

and our notion $K(\mathcal{N})$ generalizes this to arbitrary quantum channels.

Similarly, the simulation cost is given as follows.

Theorem 2 *The one-shot zero-error classical simulation cost (quantified as the minimum number of messages) of a quantum channel $\mathcal{N}$ with Choi-Jamiołkowski matrix $J_{AB} = (\mathrm{id}_A \otimes \mathcal{N})(\Phi_{AA'})$ is given by $\lceil 2^{-H_{\min}(A|B)_J} \rceil$. Here, $H_{\min}(A|B)_J$ is the conditional min-entropy defined as follows [24, 25]:*

$$2^{-H_{\min}(A|B)_J} = \Sigma(\mathcal{N}) = \min \operatorname{Tr} T_B, \text{ s.t. } J_{AB} \leq \mathbb{1}_A \otimes T_B.$$

For example, the asymptotic zero-error classical simulation cost of the cq-channel $0 \to \rho_0$ and $1 \to \rho_1$, is given by $\log(1 + D(\rho_0, \rho_1))$, where $D(\rho_0, \rho_1) = \frac{1}{2}\|\rho_0 - \rho_1\|_1$ is the trace distance between ρ_0 and ρ_1. This gives a new operational interpretation of the trace distance between ρ_0 and ρ_1 as the asymptotic exact simulation cost for the above cq-channel.

Since there might be more than one channel with Kraus operator space included in K, we are interested in the exact simulation cost of the "cheapest" among these channels. More precisely, the one-shot zero-error classical simulation cost of a Kraus operator space K is defined as

$$\Sigma(K) = \min\{\Sigma(\mathcal{N}) : \mathcal{N} \text{ is quantum channel and } K(\mathcal{N}) < K\},$$

where $K(\mathcal{N}) < K$ means that $K(\mathcal{N})$ is a subspace of K. Then it follows immediately from Theorem 2 that

Theorem 3 *The one-shot zero-error classical simulation cost of a Kraus operator space K is given by the integer ceiling of*

$$\Sigma(K) = \min \operatorname{Tr} T_B \textit{ s.t. } 0 \leq V_{AB} \leq \mathbb{1}_A \otimes T_B,\ \operatorname{Tr}_B V_{AB} = \mathbb{1}_A,\ \operatorname{Tr} V_{AB}(\mathbb{1}_{AB} - P_{AB}) = 0,$$

where P_{AB} is the projection onto the Choi-Jamiołkowski support of K.

We will prove these two theorems in Section III B.

We introduce the asymptotic zero-error channel capacity of K by considering the number of bits that can communicated over n copies of the channel $\mathcal{N}$, i.e. $\mathcal{N}^{\otimes n}$, having Kraus operator space $K^{\otimes n}$, per channel use as $n \to \infty$; we denote it as $C_{0,\text{NS}}(K)$. Likewise, the asymptotic number of bits needed per channel use to simulate $\mathcal{N}^{\otimes n}$ as $n \to \infty$, denoted $S_{0,\text{NS}}(\mathcal{N})$, and the same minimized over all channels with Kraus operator space $K^{\otimes n}$ (not necessarily product channels!), which we denote $S_{0,\text{NS}}(K)$.

From these definitions, it is clear that they are given by the regularizations of the respective one-shot quantities:

$$S_{0,\text{NS}}(\mathcal{N}) = -H_{\min}(A|B)_J, \tag{1}$$

$$C_{0,\text{NS}}(K) = \sup_{n\geq 1} \frac{1}{n} \log \Upsilon\left(K^{\otimes n}\right), \quad S_{0,\text{NS}}(K) = \inf_{n\geq 1} \frac{1}{n} \log \Sigma\left(K^{\otimes n}\right), \tag{2}$$

the first one because $H_{\min}(A|B)$ is additive under tensor products.

So far, we are unable to determine closed formulas for the latter two in general. Interestingly, the special case of K coming from a cq-channel can be solved completely. Note that if K corresponds to a cq-channel $\mathcal{N}: i \longmapsto \rho_i$, with P_i the projection over the support of ρ_i, then K can be uniquely identified by a set of projections $\{P_i\}$ (up to a permutation over inputs). In this case,

$$K = \text{span}\big\{|i\rangle\langle\psi| : |\psi\rangle \in \text{supp}\ \rho_i\big\}, \quad P = \sum_i |i\rangle\langle i| \otimes P_i.$$

Any such Kraus operator space will be called "non-commutative bipartite cq-graph" or simply "cq-graph".

The case of assisted communication seems very complicated, and most interesting. We show that the zero-error classical capacity of cq-graphs is given by the solution of the following SDP:

$$\mathsf{A}(K) := \max \sum_i s_i \text{ s.t.} 0 \le s_i,\ \sum_i s_i P_i \le \mathbb{1}. \tag{3}$$

This number was suggested by Harrow as a natural generalization of the Shannon's classical fractional packing number [26], and we will refer to it as *semidefinite (fractional) packing number* associated with a set of projections $\{P_i\}$.

Our result can be summarized as

Theorem 4 *The zero-error classical capacity of a cq-channel $\mathcal{N}: i \to \rho_i$ assisted by quantum no-signalling correlations is given by the logarithm of the semidefinite packing number $\mathsf{A}(K)$, i.e.,*

$$C_{0,NS}(K) = \log \mathsf{A}(K).$$

To be precise,

$$\frac{1}{\text{poly}(n)} \mathsf{A}(K)^n \le \Upsilon\left(K^{\otimes n}\right) \le \mathsf{A}(K)^n.$$

The proof of this theorem is given in Section IV B.

The asymptotic zero-error classical simulation cost for cq-graphs K is relatively easy and straightforward. Indeed, we show that the one-shot zero-error classical simulation cost for cq-channels is multiplicative, *i.e.*

$$\Sigma(K_1 \otimes K_2) = \Sigma(K_1)\,\Sigma(K_2), \tag{4}$$

for cq-graphs K_1 and K_2. The equality is proved by simply combining the sub-multiplicativity of the primal SDP and the super-multiplicativity of the dual problem, and then applying strong duality of SDPs. It readily follows that the asymptotic simulation cost for any Kraus operator K corresponding to a cq-channel is given by the one-shot simulation cost, namely

$$S_{0,\text{NS}}(K) = \log \Sigma(K). \tag{5}$$

It is worth noting that the above equality is valid for many other cases. In particular, it holds when K corresponds to a quantum channel that is an extreme point in the set of all quantum channels. It remains unknown whether this is true for more general K.

As an unexpected byproduct of our general analysis we obtain the following. To understand it, note that any cq-channel $\mathcal{N}: i \longmapsto \rho_i$ naturally induces a confusability graph G on the vertices i, by letting $\{i, j\} \in E$ if and only if $\rho_i \not\perp \rho_j$, i.e. inputs i and j are confusable.

Theorem 5 *For any classical graph G, the Lovász number $\vartheta(G)$ [27] is the minimum zero-error classical capacity assisted by quantum no-signalling correlations of any cq-channels that have G as non-commutative graph,* i.e.

$$\log \vartheta(G) = \min\{C_{0,NS}(K) : K^\dagger K \subseteq G\},$$

where the minimization is over cq-graphs K.

In particular, equality holds for any cq-channel $i \to |\psi_i\rangle\langle\psi_i|$ such that $\{|\psi_i\rangle\}$ is an optimal orthogonal representation for G in the sense of Lovász' original definition [27].

The above result gives the first clear information theoretic interpretion of the Lovász ϑ function of a graph G, as the zero-error classical capacity of G assisted by quantum no-signalling correlations, in the sense of taking the worst cq-channel with confusability graph G. Its proof will be given in Section V.

In Section VI, we provide results which completely characterize the feasibility (i.e., positivity) of zero-error communication with a non-commutative bipartite graph, or a quantum channel, assisted by no-signalling correlations. Finally, in Section VII, we conclude and propose several open problems for further study.

II. STRUCTURE OF QUANTUM NO-SIGNALLING CORRELATIONS

A. Do quantum no-signalling correlations naturally occur in communication?

We will provide an intuitive explanation for how the quantum no-signalling correlations naturally arise from the information-theoretic viewpoint. Let $\mathcal{M}$ and $\mathcal{N}$ be two quantum channels both from A to B, and assume that A and B can access any quantum resources that cannot be used for communicating between them directly. An interesting question is to ask when $\mathcal{N}$ can exactly simulate $\mathcal{M}$? This class of resources clearly includes shared entanglement, and actually has some other more important members that are the central interest of this paper. We will derive some general constraints that all these resources should be satisfied. Let us start with the one-shot case, that is, A and B can establish quantum channel $\mathcal{M}$ by using all "allowable resources" and one use of channel $\mathcal{N}$. We can abstractly represent the whole procedure as in Fig. 1; note that it falls within the formalism of "quantum combs" [28], but as it is a rather special case we can understand it without explicitly invoking that theory.

Here is how this simulation works. First, A performs some pre-processing on the input quantum system A_i together with all possible resources at her hand, and outputs quantum system A_o as the input of channel $\mathcal{N}$. The channel $\mathcal{N}$ then outputs a quantum system B_i. B will do some post-processing on this system together with all possible resources he has, and finally generates an output B_o. If we remove $\mathcal{N}$, we are left with a network with two inputs A_i and B_i, and two outputs A_o and B_o. Clearly this represents all possible pre- and/or post-processing that A and B have done and all resources that are available to A and B. In the framework of quantum mechanics, this network can be formulated as a quantum channel Π with two inputs and two outputs. Thus we can redraw the simulation procedure as Fig. 1.b. However, as $\mathcal{N}$ is the only communicating device from A to B, we must have that Π cannot be used to communicate from A to B. Furthermore, the output A_o represents the input of A to the channel $\mathcal{N}$, and thus can be uniquely determined by A, but not B_i, which is the output of $\mathcal{N}$. We will see this constraint is equivalent to B cannot communicate to A using Π. These constraints have led us to a fruitful class of resources that A and B could use in communication.

As before, we don't attempt to solve the most general channel simulation problem. Instead, we will focus on two simpler but most interesting cases: i). $\mathcal{M}$ is a noiseless classical channel and

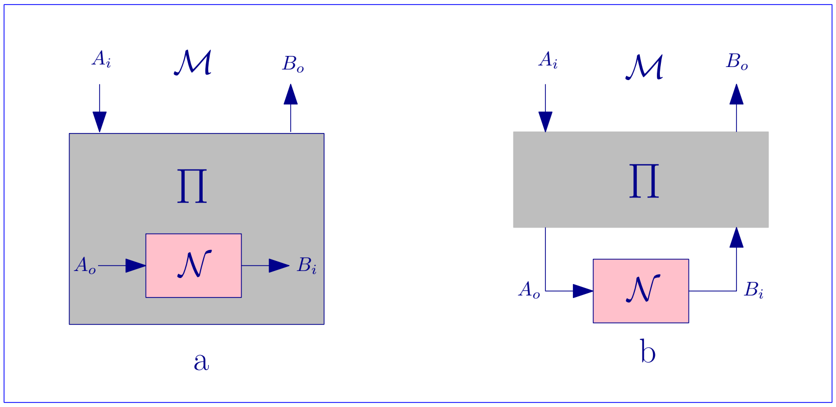

FIG. 1. A general simulation network: a). We have abstractly represented the general simulation procedure for implementing a channel $\mathcal{M}$ using another channel $\mathcal{N}$ just once, and the correlations between A and B; b). This is just an equivalent way to redraw a), and we have highlighted all correlations between A an B, and their pre- and/or post- processing as Π.

$\mathcal{N}$ is the given noisy channel. The optimal solution to this problem will lead us to the notion of zero-error classical capacity of $\mathcal{N}$; ii). $\mathcal{N}$ is a noiseless classical channel and $\mathcal{M}$ is the given noisy channel. The optimal solution will lead us to the notion of zero-error classical simulation cost of $\mathcal{M}$. In the communication problem, we want to maximize the number of messages we can send exactly by the given channel; while in the simulation problem, we want to minimize the amount of the noiseless classical communication to simulate the given channel. In the rest of this section, we will study the mathematical structures of quantum no-signalling correlations in detail.

B. Mathematical definition of quantum no-signalling correlations

As discussed before, no-signalling correlations are linear maps

$$\Pi : \mathcal{L}(A_i) \otimes \mathcal{L}(B_i) \to \mathcal{L}(A_o) \otimes \mathcal{L}(B_o)$$

with additional constraints. First, Π is required to be completely positive (CP) and trace-preserving (TP). This makes Π a physically realizable quantum operation. Furthermore, Π is A to B no-signalling (A$\not\to$B). That is, A cannot send classical information to B by using Π. More precisely, for any density operators $\rho_{A_i}^{(0)}, \rho_{A_i}^{(1)} \in \mathcal{L}(A_i)$ and $\sigma_{B_i} \in \mathcal{L}(B_i)$, we have

$$\operatorname{Tr}_{A_o} \Pi(\rho_{A_i}^{(0)} \otimes \sigma_{B_i}) = \operatorname{Tr}_{A_o} \Pi(\rho_{A_i}^{(1)} \otimes \sigma_{B_i}).$$

Or equivalently,

$$\operatorname{Tr}_{A_o} \Pi(X_{A_i} \otimes Y_{B_i}) = 0 \,\forall X,\ Y \text{ s.t. } \operatorname{Tr} X = 0.$$

Likewise, Π is required to be B to A no-signalling (B$\not\to$A). That is, B cannot send classical information to A by using Π. This constraint can be formulated as the following

$$\operatorname{Tr}_{B_o} \Pi(X_{A_i} \otimes Y_{B_i}) = 0 \,\forall X,\ Y \text{ s.t. } \operatorname{Tr} Y = 0.$$

Let the Choi-Jamiołkowski matrix of Π be

$$\Omega_{A_i' A_o B_i' B_o} = (\mathrm{id}_{A_i'} \otimes \mathrm{id}_{B_i'} \otimes \Pi)(\Phi_{A_i A_i'} \otimes \Phi_{B_i B_i'}),$$

where $\mathrm{id}_{A_i'}$ is the identity operator over $\mathcal{L}(A_i')$, $\Phi_{A_iA_i'} = |\Phi_{A_iA_i'}\rangle\langle\Phi_{A_iA_i'}|$, and $|\Phi_{A_iA_i'}\rangle = \sum_k |k_{A_i}\rangle|k_{A_i'}\rangle$ the un-normalized maximally entangled state. We now show that all above constraints on Π can be easily reformulated into the semidefinite programming constraints in terms of the Choi-Jamiołkowski matrix Ω. For convenience, we often use unprimed letters such as A_i and B_i to denote the quantum systems inputting to quantum channels, and the primed letters A_i' and B_i' for the reference systems which are isomorphic to A_i and B_i, respectively. The constraints on Π can be equivalently formulated in terms of Ω as follows:

$$\begin{aligned} \Omega \geq 0, &\quad (CP) \\ \operatorname{Tr}_{A_oB_o}\Omega = \mathbb{1}_{A_i'B_i'}, &\quad (TP) \\ \operatorname{Tr}_{A_oA_i'}\Omega X^T_{A_i'} = 0 \;\forall \operatorname{Tr} X = 0, &\quad (A \not\to B) \\ \operatorname{Tr}_{B_oB_i'}\Omega Y^T_{B_i'} = 0 \;\forall \operatorname{Tr} Y = 0, &\quad (B \not\to A) \end{aligned}$$

where X and Y are arbitrary Hermitian operators, so the transpose is not really necessary. The first two constraints guarantee that Ω corresponds to a CPTP map Π, while the latter two make sure that Ω cannot be used for communicating from A to B and B to A, respectively. Both two constraints need only be verified on a Hermitian matrix basis of A_i', B_i', respectively.

The key to deriving the above constraints is the following useful fact:

$$X_A = \operatorname{Tr}_{A'}\Phi_{AA'}(\mathbb{1}_A \otimes {X_{A'}}^T) = \operatorname{Tr}_{A'}\Phi_{AA'}{X_{A'}}^T,$$

where A' is an isomorphic copy of A.

It is worth noting that the class of quantum no-signalling correlations are closed under the convex combination. That is, if Π_0 and Π_1 are quantum no-signalling correlations and $0 \leq p \leq 1$, then $p\Pi_0 + (1-p)\Pi_1$ are also no-signalling correlations. Furthermore, this class is also stable under the pre- or post-processing by A or B. That is, if Π is a no-signalling correlation from $\mathcal{L}(A_i' \otimes B_i')$ to $\mathcal{L}(A_o \otimes B_o)$. Then $\Pi' = (\mathcal{A}_1 \otimes \mathcal{B}_1)\Pi(\mathcal{A}_0 \otimes \mathcal{B}_0)$ is also no-signalling, where $\mathcal{A}_0, \mathcal{A}_1, \mathcal{B}_0, \mathcal{B}_1$ are CPTP maps on suitable Hilbert spaces.

It is instructive to compare the quantum no-signalling correlations and the classical no-signalling correlations. Recall that any classical no-signalling correlations can be described as a memoryless classical channel $Q = \{A \times B, Q(ab|xy), X \times Y\}$ with two classical inputs $x \in X$ and $y \in Y$, and two classical outputs $a \in A$ and $b \in B$, where

$$Q(ab|xy) \geq 0,\ \forall x \in X, y \in Y, a \in A,\ b \in B, \tag{6}$$

$$\sum_{ab} Q(ab|xy) = 1,\ \forall x \in X,\ y \in Y, \tag{7}$$

$$\sum_a Q(ab|xy) = \sum_a Q(ab|x'y),\ \forall x, x' \in X,\ y \in Y, \tag{8}$$

$$\sum_b Q(ab|xy) = \sum_b Q(ab|xy'),\ \forall x \in X,\ y, y' \in Y. \tag{9}$$

Evidently, Q can also be represented as a quantum channel in the following way

$$Q(\rho) = \sum_{a,b,x,y} Q(ab|xy)|ab\rangle\langle xy|\rho|xy\rangle\langle ab|.$$

One can easily verify the above constraints are exactly the same as treating Q a quantum no-signalling correlation. From this viewpoint, quantum no-signalling correlations are natural generalizations of their classical correspondings.

Finally, we would like to mention another interesting fact. That is, any two-input and two-output quantum channel Π can be reduced to a two-input and two-output classical channel Q that has the same signalling property by simply doing pre- or post-processing, and all inputs and outputs of Q are binary, i.e., if Π is A to B (and/or B to A) signalling then Q is also A to B (resp. B to A) signalling. Due to its significance, we formulate it as

Proposition 6 *For any CPTP map Π from $\mathcal{L}(A_i \otimes B_i)$ to $\mathcal{L}(A_o \otimes B_o)$ such that Π is B to A (and/or A to B) signalling, one can obtain a classical channel $Q = \{A \times B, Q(ab|xy), X \times Y\}$ with all $|A| = |B| = |X| = |Y| = 2$, by doing suitable local pre- and post-processing on Π, such that Q is also B to A signalling (resp. A to B).*

Proof Assume Π is both way signalling (the case that it is one-way signalling is similar, and in fact simpler). Then we can find a pair of states $\rho_{A_i}^{(0)}$ and $\sigma_{B_i}^{(0)}$, such that the following two maps $\mathcal{A} : \mathcal{L}(A_i) \to \mathcal{L}(B_o)$ and $\mathcal{B} : \mathcal{L}(B_i) \to \mathcal{L}(A_o)$ are *non-constant* CPTP maps:

$$\mathcal{A}(\rho_{A_i}) = \operatorname{Tr}_{A_o} \Pi(\rho_{A_i} \otimes \sigma_{B_i}^{(0)}),$$
$$\mathcal{B}(\sigma_{B_i}) = \operatorname{Tr}_{B_o} \Pi(\rho_{A_i}^{(0)} \otimes \sigma_{B_i}).$$

I.e., we can find another two states $\rho_{A_i}^{(1)}$ and $\sigma_{B_i}^{(1)}$ such that

$$\mathcal{A}(\rho_{A_i}^{(0)}) \neq \mathcal{A}(\rho_{A_i}^{(1)}), \;\; \mathcal{B}(\sigma_{B_i}^{(0)}) \neq \mathcal{B}(\sigma_{B_i}^{(1)}).$$

By Helstrom's theorem, for any two different quantum states $\{\rho_0, \rho_1\}$, we can find a projective measurement $\{P_0, P_1\}$ to distinguish the given states with equal prior probably with an average success probability

$$p = \frac{1}{2} \operatorname{Tr} \rho_0 P_0 + \frac{1}{2} \operatorname{Tr} \rho_1 P_1 = \frac{1}{2} + \frac{1}{4} \|\rho_0 - \rho_1\|_1 > 1/2.$$

As a direct consequence, the CPTP map

$$\mathcal{M}(\rho) = \operatorname{Tr} \rho P_0 |0\rangle\langle 0| + \operatorname{Tr} \rho P_1 |1\rangle\langle 1|$$

will produce two different classical binary probability distributions $\mathcal{M}(\rho_0)$ and $\mathcal{M}(\rho_1)$. We will need this fact below.

Let $\big(P_{A_o}^{(0)}, P_{A_o}^{(1)}\big)$ and $\big(Q_{B_o}^{(0)}, Q_{B_o}^{(1)}\big)$ be the measurements to optimally distinguish $\big\{\mathcal{B}(\sigma_{B_i}^{(0)}), \mathcal{B}(\sigma_{B_i}^{(1)})\big\}$ and $\big\{\mathcal{A}(\rho_{A_i}^{(0)}), \mathcal{A}(\rho_{A_i}^{(1)})\big\}$, respectively. Then we can define four CPTP maps as follows:

$$\mathcal{A}_0(\rho) = (\operatorname{Tr} |0\rangle\langle 0|\rho)\rho_{A_i}^{(0)} + (\operatorname{Tr} |1\rangle\langle 1|\rho)\rho_{A_i}^{(1)}, \quad \mathcal{A}_1(\rho) = (\operatorname{Tr} P_{A_o}^{(0)}\rho)|0\rangle\langle 0| + (\operatorname{Tr} P_{A_o}^{(1)}\rho)|1\rangle\langle 1|,$$
$$\mathcal{B}_0(\sigma) = (\operatorname{Tr} |0\rangle\langle 0|\sigma)\sigma_{B_i}^{(0)} + (\operatorname{Tr} |1\rangle\langle 1|\sigma)\sigma_{B_i}^{(1)}, \quad \mathcal{B}_1(\sigma) = (\operatorname{Tr} Q_{B_o}^{(0)}\sigma)|0\rangle\langle 0| + (\operatorname{Tr} Q_{B_o}^{(1)}\sigma)|1\rangle\langle 1|.$$

Using these as pre- and post-processing on Π, we obtain the desired channel

$$Q = (\mathcal{A}_1 \otimes \mathcal{B}_1) \circ \Pi \circ (\mathcal{A}_0 \otimes \mathcal{B}_0).$$

In Q, if B inputs 0 or 1, and A inputs 0, then by the above construction, A must output two binary probability distributions that are different, hence can be used for signalling from B to A. Similarly, if A inputs 0 or 1, and B inputs 0, then B must output another two different binary probability distributions that can be used for signalling from A to B. $\square$

C. Structure theorems for quantum no-signalling correlations

We will establish several structure theorems regarding quantum no-signalling correlations. Note that Π is a two-input and two-output quantum channel. So there are two natural ways to think of Π according to the relation between the inputs and outputs. The first way is to partition Π as $A_iA_o : B_iB_o$, so the output of A_i and B_i would be A_o and B_o, respectively; this is perhaps the most standard way. The second way is to focus on the communication between A and B and partition Π as $A_iB_o : B_iB_o$. In this case the output of A_i and B_i will be regarded as B_o and A_o, respectively. This kind of partition is quite useful when we are interested in the communication between A and B.

Let us start with the bipartition $A_iA_o : B_iB_o$ of Π. In this case, we can have a full characterization of no-signalling maps. (Bear in mind that we have assumed Π to be a CPTP map in the following discussion).

Proposition 7 *Let $\Pi : \mathcal{L}(A_i) \otimes \mathcal{L}(B_i) \to \mathcal{L}(A_o) \otimes \mathcal{L}(B_o)$ be a bipartite CPTP map with a decomposition $\Pi = \sum_k \lambda_k \mathcal{A}_k \otimes \mathcal{B}_k$ according to bipartition $A_iA_o : B_iB_o$, where $\mathcal{A}_k : \mathcal{L}(A_i) \to \mathcal{L}(A_o)$, $\mathcal{B}_k : \mathcal{L}(B_i) \to \mathcal{L}(B_o)$, and λ_k are complex numbers. Then we have the following:*

i) Π is B to A no-signalling iff $\mathcal{B}_k$ can be chosen as CPTP for any k. In this case $\sum_k \lambda_k \mathcal{A}_k$ will be also CPTP.

ii) Π is A to B no-signalling iff $\mathcal{A}_k$ can be chosen as CPTP for any k. In this case $\sum_k \lambda_k \mathcal{B}_k$ will be also CPTP.

iii) Π is no-signalling between A and B iff both $\mathcal{A}_k$ and $\mathcal{B}_k$ can be chosen as CPTP maps, $\sum_k \lambda_k = 1$, and λ_k are real for all k.

The most interesting part is item iii). Intuitively, any no-signalling correlation Π between A and B can be written as a real affine combination of product CPTP maps $\mathcal{A}_k \otimes \mathcal{B}_k$. The proofs are relatively straightforward and we simply leave them to interested readers as exercises.

Now we turn to the bipartition $A_iB_o : B_iA_o$ of Π. We don't have a full characterization of no-signalling correlations in this setting. Nevertheless, a simple but extremely useful class of no-signalling correlations can be constructed using the following facts.

Proposition 8 *Let $\Pi : \mathcal{L}(A_i) \otimes \mathcal{L}(B_i) \to \mathcal{L}(A_o) \otimes \mathcal{L}(B_o)$ be a bipartite CPTP map with a decomposition $\Pi = \sum_k \mu_k \mathcal{E}_k \otimes \mathcal{F}_k$ according to bipartition $A_iB_o : B_iA_o$, where $\mathcal{E}_k : \mathcal{L}(A_i) \to \mathcal{L}(B_o)$, $\mathcal{F}_k : \mathcal{L}(B_i) \to \mathcal{L}(A_o)$, and μ_k are complex numbers. Then we have the following:*

i) If $\sum_k \mu_k \mathcal{E}_k$ is a constant map and $\mathcal{F}_k$ is TP for any k, then Π is A to B no-signalling.

ii) If $\sum_k \mu_k \mathcal{F}_k$ is a constant map and $\mathcal{E}_k$ is TP for any k, then Π is B to A no-signalling.

iii) If all $\mathcal{E}_k$ and $\mathcal{F}_k$ are TP and both $\sum_k \mu_k \mathcal{E}_k$ and $\sum_k \mu_k \mathcal{F}_k$ are constant maps, then Π is no-signalling between A and B.

For example, we can choose $\Pi = \sum_k \mu_k \mathcal{E}_k \otimes \mathcal{F}_k$, where $\mathcal{E}_k$ and $\mathcal{F}_k$ are CPTP maps from $\mathcal{L}(A_i)$ to $\mathcal{L}(B_o)$ and $\mathcal{L}(B_i)$ to $\mathcal{L}(A_o)$, respectively, and $\{\mu_k\}$ is a probability distribution. If we further have $\sum_k \mu_k \mathcal{E}_k$ and $\sum_k \mu_k \mathcal{F}_k$ are constant maps, then we obtain a no-signalling correlation Π. Clearly, for any such Π, neither of A_i and B_i can send classical information to A_o or B_o. To emphasize this special feature, this class of no-signalling correlations are said to be *totally no-signalling*. We show see later even this class of correlations could be very useful in assisting zero-error communication and simulation. In particular, we need the following technical result:

Lemma 9 *Let $\mathcal{E}_0$ and $\mathcal{E}_1$ be two CPTP maps from $\mathcal{L}(A_i)$ to $\mathcal{L}(B_o)$, and let $\mathcal{F}_0, \mathcal{F}_1$ be two CP maps from $\mathcal{L}(B_i)$ to $\mathcal{L}(A_o)$. Furthermore, assume there is unique $0 \leq p \leq 1$ such that $p\mathcal{E}_0 + (1-p)\mathcal{E}_1$ is a constant channel. Then $\Pi = p\mathcal{E}_0 \otimes \mathcal{F}_0 + (1-p)\mathcal{E}_1 \otimes \mathcal{F}_1$ is a no-signalling correlation if and only if both $\mathcal{F}_0$ and $\mathcal{F}_1$ are CPTP maps, and $p\mathcal{F}_0 + (1-p)\mathcal{F}_1$ is a constant map.*

Proof This can be proven by directly applying the definition of quantum no-signalling correlations. Two key points are: First, the uniqueness of p such that $p\mathcal{E}_0 + (1-p)\mathcal{E}_1$ is constant; secondly, if $\mathcal{E}(X) = 0$ for any any traceless Hermitian operator X, $\mathcal{E}$ is a constant map. □

A third result about the structure of no-signalling correlations is the following [18] (see also [19] for an alternate proof, and [21] for more discussions):

Proposition 10 *Π is $B\not\to A$ iff $\Pi = \mathcal{G} \circ \mathcal{F}$ for two CPTP maps $\mathcal{F} : \mathcal{L}(A_i) \to \mathcal{L}(A_o \otimes R)$ and $\mathcal{G} : \mathcal{L}(B_i \otimes R) \to \mathcal{L}(B_o)$, where R is an internal memory.* □

This result is quite intuitive. It says that in the case that B cannot communicate to A, we can actually have the output of A before we input to B.

D. Composing no-signalling correlations with quantum channels

Suppose now we are given a no-signalling map $\Pi : \mathcal{L}(A_i \otimes B_i) \to \mathcal{L}(A_o \otimes B_o)$ and a CPTP map $\mathcal{N} : \mathcal{L}(A) \to \mathcal{L}(B)$. When $A = A_o$ and $B = B_i$, we can construct a new and unique map $\mathcal{M} : \mathcal{L}(A_i) \to \mathcal{L}(B_o)$ by feeding the output A_o of Π into $\mathcal{N}$, and use the output B of $\mathcal{N}$ as the input B_i to Π. Such a map can be abstractly represented as follows

$$\mathcal{M}^{A_i \to B_o} = \Pi^{A_i \otimes B_i \to A_o \otimes B_o} \circ \mathcal{N}^{A_o \to B_i}. \tag{10}$$

Next we will show how to derive more explicit forms of this map.

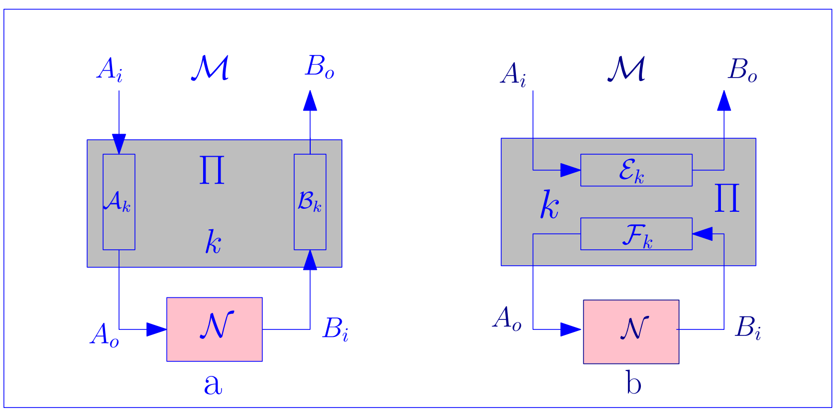

FIG. 2. Two different partitions of the input-output registers of Π

Let us start by considering the partition according to $A_iA_o : B_iB_o$ to look at Π. If Π is of product form $\mathcal{A} \otimes \mathcal{B}$ where $\mathcal{A} : \mathcal{L}(A_i) \to \mathcal{L}(A_o)$ and $\mathcal{B} : \mathcal{L}(B_i) \to \mathcal{L}(B_o)$, then clearly the composition must be given by $\mathcal{M}(X) = \mathcal{B} \circ \mathcal{N} \circ \mathcal{A}(X)$. We can simply extend this construction to the superpositions of these product maps by linearity. I.e., for a general $\Pi = \sum_k \mathcal{A}_k \otimes \mathcal{B}_k$, we define the new map by composing $\mathcal{N}$ and Π as the following

$$\mathcal{M}^{A_i \to B_o} = \sum_k \mathcal{B}_k^{B_i \to B_o} \circ \mathcal{N}^{A_o \to B_i} \circ \mathcal{A}_k^{A_i \to A_o}, \tag{11}$$

and this is clearly well-defined.

We can also formulate $\mathcal{M}$ in another useful way by considering the bipartition $A_iB_o : B_iA_o$. Suppose Π is given in the form of $\sum_k \mathcal{E}_k \otimes \mathcal{F}_k$ where $\mathcal{E}_k : \mathcal{L}(A_i) \to \mathcal{L}(B_o)$, $\mathcal{F}_k : \mathcal{L}(B_i) \to \mathcal{L}(A_o)$.

Then to compose Π and $\mathcal{N}$, the only thing we need to do is to compose $\mathcal{N}$ and $\mathcal{F}_k$ directly and take trace. That is, we have

$$\mathcal{M} = \sum_k \mathcal{E}_k^{A_i \to B_o} (\operatorname{Tr} \mathcal{F}_k \circ \mathcal{N}), \tag{12}$$

where the trace of super-operator $\mathcal{F}_k \circ \mathcal{N}$ on $\mathcal{L}(A_o)$ is given by

$$\operatorname{Tr} \mathcal{F}_k \circ \mathcal{N} = \sum_j \operatorname{Tr} C_j^\dagger [(\mathcal{F}_k \circ \mathcal{N}) C_j]$$

for any orthonormal basis $\{C_j\}$ on $\mathcal{L}(A_o)$ in the sense of $\operatorname{Tr} C_j^\dagger C_l = \delta_{jl}$. This is just the natural generalization of the trace function of the square matrices to super-operators. It is easy to verify that $\operatorname{Tr} \mathcal{F} \circ \mathcal{N} = \operatorname{Tr} \mathcal{N} \circ \mathcal{F}$ holds for any $\mathcal{F}$ and $\mathcal{N}$, and the trace is independent of the choice of the orthonormal basis $\{C_j\}$.

For instance, for any bipartite classical channel $Q = (X \times Y, Q(ab|xy), A \times B)$, and another classical channel $N = (A, N(y|a), Y)$, we can easily show that the map $M = (X, M(b|x), B)$ constructed by composing Q and N is given by

$$M(b|x) = \sum_{y,a} Q(ab|xy) N(y|a),$$

which coincides with the discussions in [12].

Note that the above two constructions of the map $\mathcal{M}$ from Π and $\mathcal{N}$ are quite general, and purely mathematical. Indeed, we can form a map $\mathcal{M}$ from any $\mathcal{N}$ and Π. An interesting and important question is to ask whether $\mathcal{M}$ is a CPTP map when both $\mathcal{N}$ and Π are.

It turns out that whenever $\mathcal{N}$ and Π are CP maps, $\mathcal{M}$ should also be CP. This fact is actually a simple corollary of the teleportation protocol. Indeed, we can easily see from the Fig. 3 that

$$\mathcal{M}(X_{A_i}) = \operatorname{Tr}_{BB_i'}[(\mathrm{id}_{B_o B_i'} \otimes \mathcal{N}) \circ (\mathrm{id}_{B_i'} \otimes \Pi)(X_{A_i} \otimes \Phi_{B_i B_i'})(\mathbb{1}_{B_o} \otimes \Phi_{BB_i'})], \tag{13}$$

where $A = A_o$, and B, B_i, and B_i' are all isomorphic. The way to get the above equality is to first apply the teleportation protocol to the special case that $\Pi = \mathcal{A} \otimes \mathcal{B}$, and then extend the result to to the general case by linearity. The above expression not only provides a proof for the complete positivity of $\mathcal{M}$, but also the uniqueness of such $\mathcal{M}$.

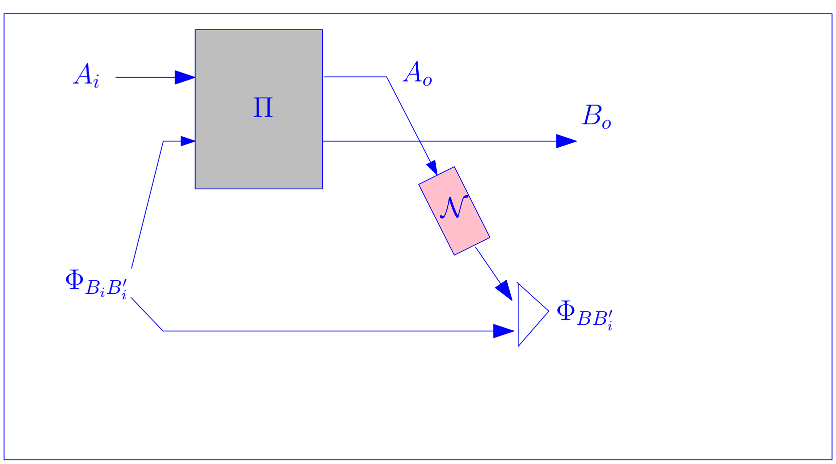

FIG. 3. Composing $\mathcal{N}$ and Π from the viewpoint of teleportation.

So we have obtained that for any two maps $\Pi : \mathcal{L}(A_i \otimes B_i) \to \mathcal{L}(A_o \otimes B_o)$ and $\mathcal{N} : \mathcal{L}(A_o) \to \mathcal{L}(B_i)$, we can uniquely construct a map $\mathcal{M} : \mathcal{L}(A_i) \to \mathcal{L}(B_o)$ by composing Π and $\mathcal{N}$, and trace out A_o and B_i (refer to Eq. (10)). An explicit construction can be equivalently given by any of Eqs. (11), (12), and (13). Furthermore, if Π and $\mathcal{N}$ are CP, so is $\mathcal{M}$.

To guarantee that $\mathcal{M}$ is TP, we have to put further constraints on Π, but not on $\mathcal{N}$ except that it is an arbitrary CPTP map. We will show that such constraints are precisely the same as the constraints that B to A no-signalling (B$\not\to$ A).

Proposition 11 *Let $\Pi : \mathcal{L}(A_i \otimes B_i) \to \mathcal{L}(A_o \otimes B_o)$ and $\mathcal{N} : \mathcal{L}(A_o) \to \mathcal{L}(B_i)$ be two CPTP maps, and let $\mathcal{M} : \mathcal{L}(A_i) \to \mathcal{L}(B_o)$ be the unique CP constructed by composing Π and $\mathcal{N}$. If Π is CPTP and B to A no-signalling, then $\mathcal{M}$ is also CPTP for any CPTP map $\mathcal{N}$. Conversely, if $\mathcal{M}$ is CPTP for any $\mathcal{N}$, then Π has to be B to A no-signalling.*

Proof The sufficiency is just a direct corollary of item ii) of Proposition 7 or Proposition 10. The later actually provides a physical realization of this map $\mathcal{M}$. This constraint is logically reasonable as B to A no-signalling simply means that the input of the channel $\mathcal{N}$ cannot depend on its output (otherwise we will have a closed loop, and against the causality principle).

Now we will prove that the constraint of no-signalling from B to $A$is also necessary for $\mathcal{M}$ to be CPTP for any CPTP $\mathcal{N}$. Our strategy is to first prove this fact for two-input and two-output binary classical channel Q. Then we can apply Proposition 6 for the general case.

We can actually show that if Q is signalling from B to A, and the composition of Q and one-bit noiseless classical channel is not a legal classical channel anymore (the output for some input is not a legal probability distribution).

By contradiction, assume that the composition of Π and any CPTP map $\mathcal{N}$ is always CPTP, but that Π is B to A signalling. Applying Proposition 6, we can find a two-input and two-output binary classical channel Q such that i). Q is constructed from Π by simply performing pre- and/or post-processing such that $Q = (\mathcal{A}_1 \otimes \mathcal{B}_1) \circ \Pi \circ (\mathcal{A}_0 \otimes \mathcal{A}_1)$; ii). Q is also B to A signalling; iii) the composition of Q and any classical channel A to Y is also a legal classical channel. We will show such Q does not exist, hence complete the proof. Actually, ii) implies there are $x \in A, a \in A, y, y' \in Y$, such that

$$\sum_{b=0}^{1} Q(ab|xy) \neq \sum_{b=0}^{1} Q(ab|xy').$$

w.o.l.g, we can assume $x = 0$, $a = 0$, $y = 0$, and $y' = 1$. Then the above inequality can be rewritten explicitly as

$$Q(00|00) + Q(01|00) \neq Q(00|01) + Q(01|01).$$

On the other hand, the composition of Q and a one-bit noiseless classical channel $N(y|a) = \delta_{ya}$ is give by

$$M(b|x) = \sum_{ya} Q(ab|xy)N(y|a) = \sum_{a=0}^{1} Q(ab|xa) = Q(0b|x0) + Q(1b|x1).$$

In particular, taking $x = 0$ and applying the above inequality, we have

$$\begin{aligned}\sum_{b=0}^{1} M(b|0) &= Q(00|00) + Q(01|00) + Q(10|01) + Q(11|01) \\ &\neq Q(00|01) + Q(01|01) + Q(10|01) + Q(11|01) = 1.\end{aligned}$$

This contradicts the fact that $M(b|x)$ is a classical channel. □

Thus, to guarantee that a CPTP map $\mathcal{M}$ can always be obtained by composing a two-input and two-output channel Π and any quantum channel from A to B, Π is only required to be B to A no-signalling. However, the constraints of no-signalling from A to B is a natural requirement as we have assumed that A cannot communicate to B directly, and the given channel $\mathcal{N}$ is the only directed resource from A to B we can use.

III. SEMIDEFINITE PROGRAMMES FOR ZERO-ERROR COMMUNICATION AND SIMULATION ASSISTED BY QUANTUM NO-SIGNALLING CORRELATIONS

A. Zero-error assisted communication capacity

For any integer M, we denote by M_a and M_b classical registers with size M. We use $\mathcal{I}_M : M_a \to M_b$ to denote the noiseless classical channel that can send M messages from A to B, *i.e.* $\mathcal{I}_M(|m\rangle\langle m'|_{M_a}) = \delta_{mm'}|m\rangle\langle m|_{M_b}$, or equivalenyly,

$$\mathcal{I}_M(\rho) = \sum_{m=1}^{M} \left(\operatorname{Tr}\rho|m\rangle\langle m|_{M_a}\right)|m\rangle\langle m|_{M_b}.$$

If a channel $\mathcal{N}$ can simulate $\mathcal{I}_M$, we say that M messages can be perfectly transmitted by one use of $\mathcal{N}$. The problem we are interested in is to determine the largest possible M, which will be called one-shot zero-error classical transmission capability of $\mathcal{N}$ assisted with no-signalling correlations. We restate here the result that we will prove in this subsection.

Theorem 1 *The one-shot zero-error classical capability (quantified as the largest number of messages) of $\mathcal{N}$ assisted by quantum no-signalling correlations depends only on the non-commutative graph K, and is given by the integer part of the following SDP:*

$$\begin{aligned}\Upsilon(\mathcal{N}) &= \Upsilon(K)\\ &= \max \operatorname{Tr} S_A \ \ s.t.\ \ 0 \le U_{AB} \le S_A \otimes \mathbb{1}_B, \operatorname{Tr}_A U_{AB} = \mathbb{1}_B, \operatorname{Tr} P_{AB}(S_A \otimes \mathbb{1}_B - U_{AB}) = 0.\end{aligned}$$

Hence we are motivated to call $\Upsilon(K)$ the no-signalling assisted independence number of K.

Proof From now on, to make the indices of the Choi-Jamiołkowski matrices more readable, we reserve the unprimed letters such as A_i and B_i for the reference systems (thus as the indexes of Choi-Jamiołkowski matrices), and using the primed versions A_i' and B_i' as the inputs to the quantum channels.

We will use a general no-signalling correlation Π with Choi-Jamiołkowski matrix Ω and the channel $\mathcal{N}$, to exactly simulate a noiseless classical channel $\mathcal{I}_M$. Our goal is to determine the maximum integer M when such simulation is possible. In this case both $A_i' = M_a'$ and $B_o = M_b$ are classical, and $A_o = A'$, $B_i' = B$. We will show that Ω can be chosen as the following form:

$$\Omega_{M_aM_bAB} = \frac{1}{M} D_{M_aM_b} \otimes U_{AB} + \frac{1}{M}(\mathbb{1} - D)_{M_aM_b} \otimes V_{AB}, \tag{14}$$

where $D = \sum_m |mm\rangle\langle mm|$ is the Choi-Jamiołkowski matrix of the noiseless classical channel $\mathcal{I}_M$, and U and V are positive semidefinite. We will show that the no-signalling conditions are translated into

$$\operatorname{Tr}_A U_{AB} = \operatorname{Tr}_A V_{AB} = \mathbb{1}_B, \quad (A \not\to B), \tag{15}$$

$$\frac{1}{M} U_{AB} + \left(1 - \frac{1}{M}\right) V_{AB} = \sigma_A \otimes \mathbb{1}_B, \quad (B \not\to A), \tag{16}$$

with some state σ. In other words, U and V are proportional to Choi-Jamiołkowski matrices of channels from $\mathcal{L}(B)$ to $\mathcal{L}(A)$, whose weighted sum is equal to a constant channel mapping quantum states in B to a fixed state σ_A.

Let us first have a closer look at the the mathematical structure of Ω. Rewrite Ω into the following form,

$$\Omega_{M_aM_bAB} = \frac{1}{M}D_{M_aM_b} \otimes U_{AB} + \left(1 - \frac{1}{M}\right)\tilde{D}_{M_aM_b} \otimes V_{AB},$$

where $\tilde{D}_{M_aM_b} = \frac{1}{M-1}(\mathbb{1} - D)$ is the Choi matrix of the classical channel $\widetilde{\mathcal{I}}_M$ that sends each m into a uniform distribution of $m' \neq m$. In other words, the no-signalling correlation Π should have the following form:

$$\Pi = \frac{1}{M}\mathcal{I}_M \otimes \mathcal{E} + \left(1 - \frac{1}{M}\right)\widetilde{\mathcal{I}}_M \otimes \mathcal{F}, \tag{17}$$

and $\mathcal{E}$ and $\mathcal{F}$ are CP maps from B to A corresponding to Choi-Jamiołkowski matrices U_{AB} and V_{AB}.

Now we can directly apply Lemma 9 to Π in Eq. (17) to obtain the no-signalling constraints in Eqs. (15) and (16). First, the constraints for $A \not\to B$ are automatically satisfied due to the special forms of $\mathcal{I}_M$ and $\widetilde{\mathcal{I}}_M$, and $p = 1/M$ is the unique number such that $\frac{1}{M}\mathcal{I}_M + \left(1 - \frac{1}{M}\right)\widetilde{\mathcal{I}}_M$ is constant. The no-signalling constraint $B \not\to A$ is equivalent to $\mathcal{E}$ and $\mathcal{F}$ are CPTP maps, and $\frac{1}{M}\mathcal{E} + \left(1 - \frac{1}{M}\right)\mathcal{F}$ is some constant channel.

When composing $\mathcal{N}$ with Π in Eq. (17), we have the following channel

$$\mathcal{M} = \frac{1}{M}\mathcal{I}_M\big(\mathrm{Tr}\,\mathcal{N} \circ \mathcal{E}\big) + \left(1 - \frac{1}{M}\right)\widetilde{\mathcal{I}}_M\big(\mathrm{Tr}\,\mathcal{N} \circ \mathcal{F}\big).$$

The zero-error constraint requires $\mathcal{M} = \mathcal{I}_M$, which is equivalent to $\mathrm{Tr}(\mathcal{N} \circ \mathcal{F}) = 0$, or

$$\mathrm{Tr}\, V_{AB} J_{AB} = 0,$$

where $J_{AB} = (\mathcal{N}^\dagger \otimes \mathrm{id}_B)(\Phi_{B'B})$ is the Choi-Jamiołkowski matrix of $\mathcal{N}$ (strictly speaking, this is the complex conjugate of the Choi-Jamiołkowski matrix of $\mathcal{N}$, but this does not make any difference to the problem we are studying).

Since U_{AB} and V_{AB} depend on each other, we can eliminate one of them. For instance, we only keep U_{AB}. Then the existence of V_{AB} will be equivalent to

$$\mathrm{Tr}_A\, U_{AB} = \mathbb{1}_B, 0 \leq U_{AB} \leq M\sigma_A \otimes \mathbb{1}_B, \mathrm{Tr}\, P_{AB}(M\sigma_A \otimes \mathbb{1}_B - U_{AB}) = 0.$$

By absorbing M into σ_A and introducing $S_A = M\sigma_A$, we get M as the integer part of

$$\max \mathrm{Tr}\, S_A \text{ s.t. } 0 \leq U_{AB} \leq S_A \otimes \mathbb{1}_B, \mathrm{Tr}_A\, U_{AB} = \mathbb{1}_B,\ \mathrm{Tr}\, P_{AB}(S_A \otimes \mathbb{1}_B - U_{AB}) = 0, \tag{18}$$

which ends the proof. □

Now we provide a detailed derivation of the form (14) of the no-signalling correlation. Assume the channel is $\mathcal{N} : \mathcal{L}(A) \to \mathcal{L}(B'')$. The no-signalling correlation we will use is $\Pi : \mathcal{L}(M_a' \otimes B') \to \mathcal{L}(A \otimes M_b)$. Suppose we can send M messages exactly by one use of the channel $\mathcal{N}$ when assisted by Π. Then we should have (refer to Fig. 3)

$$\mathrm{Tr}_{BB''}[\Phi_{BB''}(\mathcal{N} \otimes \mathrm{id}_{M_bB}) \circ (\Pi \otimes \mathrm{id}_B)(|m\rangle\langle m|_{M_a'} \otimes \Phi_{B'B})] = |m\rangle\langle m|_{M_b}\ \forall m \in \{1, \ldots, M\}. \tag{19}$$

The Choi-Jamiołkowski matrix of Π, say $\Omega_{M_aM_bAB}$, should satisfy the following constraints

$$\Omega_{M_aM_bAB} \geq 0,\ \operatorname{Tr}_{AM_b} \Omega = \mathbb{1}_{M_aB}.$$

Furthermore, noticing that

$$|m\rangle\langle m|_{M_a'} = \operatorname{Tr}_{M_a} \Phi_{M_a'M_a}(\mathbb{1}_{M_a'} \otimes |m\rangle\langle m|_{M_a}^T) \text{ and } |m\rangle\langle m| = |m\rangle\langle m|^T,$$

we have

$$(\Pi \otimes \mathrm{id}_B)(|m\rangle\langle m|_{M_a'} \otimes \Phi_{B'B}) = \operatorname{Tr}_{M_a} \Omega \cdot (|m\rangle\langle m|_{M_a} \otimes \mathbb{1}_{M_bAB}).$$

Thus the left hand side (l.h.s) of Eq. (19) gives, for all $m \in \{1, \ldots, M\}$,

$$\operatorname{Tr}_{BB''M_a}[((\mathcal{N} \otimes \mathrm{id}_{M_aM_bB})\Omega_{M_aM_bAB})(\Phi_{BB''} \otimes |m\rangle\langle m|_{M_a} \otimes \mathbb{1}_{M_b})] = |m\rangle\langle m|_{M_b}.$$

In other words, for any $m \neq m'$, we have

$$\operatorname{Tr}\big((\mathcal{N} \otimes \mathrm{id})\Omega(\Phi_{BB''} \otimes |m\rangle\langle m|_{M_a} \otimes |m'\rangle\langle m'|_{M_b})\big) = 0.$$

The next crucial step is to simplify the form of Ω. We will study all the possible forms of Ω satisfying the above equation. (Any such operator is said to be feasible). Since both M_a and M_b are classical registers, we can perform the dephasing operation on them and assume Ω has the following form

$$\Omega = \sum_{m=1}^{M} \sum_{m'=1}^{M} |mm'\rangle\langle mm'|_{M_aM_b} \otimes \Omega_{AB}^{(mm')},$$

where $\Omega_{AB}^{(mm')}$ might not be identical for different pairs (m, m').

To further simplify the form of Ω, we next exploit the permutation invariance of $\mathcal{I}_M$. More precisely, for any $M \times M$ permutation $\tau \in S_M$, if $\Omega_{M_aM_bAB}$ is feasible, then

$$\Omega'_{M_aM_bAB} = (\tau_{M_a} \otimes \tau_{M_b} \otimes \mathbb{1}_{AB})\Omega_{M_aM_bAB}(\tau_{M_a} \otimes \tau_{M_b} \otimes \mathbb{1}_{AB})^\dagger$$

is also feasible. Furthermore, if Ω' and Ω'' are feasible, so is any convex combination $\lambda\Omega' + (1-\lambda)\Omega''$ for $0 \leq \lambda \leq 1$. With these two observations, from any feasible Ω (has been dephased), we can always constructing a new feasible $\tilde{\Omega}$ by performing the following twirling operation

$$\tilde{\Omega}_{M_aM_bAB} = \frac{1}{M!} \sum_{\tau \in S_M} (\tau_{M_a} \otimes \tau_{M_b} \otimes \mathbb{1}_{AB})\Omega_{M_aM_bAB}(\tau_{M_a} \otimes \tau_{M_b} \otimes \mathbb{1}_{AB})^\dagger.$$

By applying Schur's Lemma to the symmetry group $\{\tau_{M_a} \otimes \tau_{M_b} : \tau \in S_M\}$, we can see that $\tilde{\Omega}$ can be chosen as the following form

$$\tilde{\Omega}_{M_aM_bAB} = \sum_{m} |mm\rangle\langle mm|_{M_aM_b} \otimes \Omega_{AB}^{=} + \sum_{m \neq m'} |mm'\rangle\langle mm'| \otimes \Omega_{AB}^{\neq},$$

which has exactly the same form as Eq. (14).

The SDP characterization of Theorem 1,

$$\begin{aligned} \Upsilon(K) = \max \operatorname{Tr} S_A \ \text{ s.t. }\ & 0 \leq U_{AB} \leq S_A \otimes \mathbb{1}_B, \\ & \operatorname{Tr}_A U_{AB} = \mathbb{1}_B, \\ & \operatorname{Tr} P_{AB}(S_A \otimes \mathbb{1}_B - U_{AB}) = 0, \end{aligned} \tag{20}$$

has a dual. It is given by

$$
\begin{aligned}
\Upsilon(K) = \min \operatorname{Tr} T_B \ \text{ s.t. } & z \geq 0,\ W_{AB} \geq 0, \\
& zP_{AB} - W_{AB} \leq \mathbb{1}_A \otimes T_B, \\
& \operatorname{Tr}_B (zP - W)_{AB} = \mathbb{1}_A \\
= \min \operatorname{Tr} T_B \ \text{ s.t. } & V_{AB} \leq \mathbb{1}_A \otimes T_B, \\
& \operatorname{Tr}_B V_{AB} = \mathbb{1}_A, \\
& (\mathbb{1} - P)_{AB} V_{AB} (\mathbb{1} - P)_{AB} \leq 0,
\end{aligned}
\tag{21}
$$

which can be derived by the usual means; by strong duality, which applies here, the values of both the primal and the dual SDP coincide.

For an easy example of the evaluation of these SDPs, $\Upsilon(\Delta_\ell) = \ell$ for the cq-graph Δ_ℓ of the noiseless classical channel $\mathcal{I}_\ell$ of ℓ symbols. The Kraus subspace of the noiseless ℓ-level channel id_ℓ is $\mathbb{C}\mathbb{1}$, and has $\Upsilon(\mathbb{C}\mathbb{1}) = \ell^2$.

Remark By inspection of the primal SDP, for any K_1 and K_2, $\Upsilon(K_1 \otimes K_2) \geq \Upsilon(K_1)\Upsilon(K_2)$, because the tensor product of feasible solutions of the SDP (20) is feasible for $K_1 \otimes K_2$. In particular, $\Upsilon(K \otimes \Delta_\ell) \geq \ell\, \Upsilon(K)$. We do not know whether equality holds here (sometimes it does); the most natural way for proving this would be to use the dual SDP (21), but to use it to show "$\leq$" by tensoring together dual feasible solutions would require that $T \geq 0$.

Leaving this aside, this last observation is why $\Upsilon(K)$ is rightfully called the no-signalling assisted independence number, and not its integer part. Indeed, the number of messages we can send via $K \otimes \Delta_\ell$ is $\lfloor \Upsilon(K \otimes \Delta_\ell) \rfloor > \ell \lfloor \Upsilon(K) \rfloor$, for non-integer $\Upsilon(K)$ and sufficiently large ℓ. □

B. Zero-error assisted simulation cost

For convenience we restate here the two theorems already announced in the introduction.

Theorem 2 *The one-shot zero-error classical simulation cost (quantified as the minimum number of messages) of a quantum channel $\mathcal{N}$ with Choi-Jamiołkowski matrix $J_{AB} = (\mathrm{id}_A \otimes \mathcal{N})(\Phi_{AA'})$ is given by $\lceil 2^{-H_{\min}(A|B)_J} \rceil$. Here, $H_{\min}(A|B)_J$ is the conditional min-entropy defined as follows [24, 25]:*

$$
2^{-H_{\min}(A|B)_J} = \Sigma(\mathcal{N}) = \min \operatorname{Tr} T_B, \ \text{s.t. } J_{AB} \leq \mathbb{1}_A \otimes T_B.
$$

Theorem 3 *The one-shot zero-error classical simulation cost of a Kraus operator space K is given by the integer ceiling of*

$$
\Sigma(K) = \min \operatorname{Tr} T_B \ \textit{s.t.}\ 0 \leq V_{AB} \leq \mathbb{1}_A \otimes T_B,\ \operatorname{Tr}_B V_{AB} = \mathbb{1}_A,\ \operatorname{Tr} V_{AB}(\mathbb{1}_{AB} - P_{AB}) = 0,
$$

where P_{AB} is the projection onto the Choi-Jamiołkowski support of K.

Proof In this case we have $A_o = M_a$ and $B_i = M_b$ are classical, $A_i = A$, $B_o = B$. We will show that w.l.o.g.

$$
\Omega_{ABM_aM_b} = \frac{1}{M} U_{AB} \otimes D_{M_aM_b} + \left(1 - \frac{1}{M}\right) V_{AB} \otimes \tilde{D}_{M_aM_b},
\tag{22}
$$

with positive semidefinite U_{AB} and V_{AB}. Then according to Lemma 9, the no-signalling conditions are equivalent to

$$\mathrm{Tr}_B U_{AB} = \mathrm{Tr}_B V_{AB} = \mathbb{1}_A,$$
$$\frac{1}{M}U_{AB} + \left(1 - \frac{1}{M}\right) V_{AB} = \mathbb{1}_A \otimes \gamma_B,$$

for a state γ.

Now, there are two variants of the problem. First, to simulate the precise channel $\mathcal{N}$, U_{AB} has to be equal to J_{AB}. By identifying $T_B = M\gamma_B$ and eliminating V_{AB}, we get the solution for the minimal M as smallest integer

$$\geq 2^{-H_{\min}(A|B)_J} = \min \mathrm{Tr}\, T_B \text{ s.t. } J_{AB} \leq \mathbb{1}_A \otimes T_B,$$

proving Theorem 2. The latter is then also asymptotic cost of simulating many copies of N because the conditional min-entropy is additive, Eq. (1).

Furthermore, to simulate the "cheapest" $\mathcal{N}$ with Choi-Jamiołkowski matrix supporting on P_{AB}, we get

$$\min \mathrm{Tr}\, T_B \text{ s.t. } 0 \leq V_{AB} \leq \mathbb{1}_A \otimes T_B,\ \mathrm{Tr}_B V_{AB} = \mathbb{1}_A,\ \mathrm{Tr}\, V_{AB}(\mathbb{1}_{AB} - P_{AB}) = 0,$$

which is the claim of Theorem 3. □

Now we provide a more detailed derivation of the form of no-signalling correlations. Suppose we can use a noiseless classical channel $\mathcal{I}_M$ to simulate a quantum channel $\mathcal{N} : \mathcal{L}(A') \to \mathcal{L}(B)$. The no-signalling correlation we will use is $\Pi : \mathcal{L}(A' \otimes M_b') \to \mathcal{L}(M_a \otimes B)$. Then we have

$$\mathcal{N}(\rho_{A'}) = \sum_{m=1}^{M} \mathrm{Tr}_{M_a}[(|m\rangle\langle m|_{M_a} \otimes \mathbb{1}_{B_o})\Pi(\rho_{A'} \otimes |m\rangle\langle m|_{M_b'})].$$

Denote the Choi-Jamiołkowski matrix of Π as

$$\Omega_{M_a M_b AB} = (\mathrm{id}_{M_b A} \otimes \Pi)(\Phi_{M_b M_b'} \otimes \Phi_{AA'}).$$

Thus the Choi-Jamiołkowski matrix of $\mathcal{N}$ is given by

$$\begin{aligned} J_{AB} = (\mathrm{id}_A \otimes \mathcal{N})(\Phi_{AA'}) &= \sum_m \mathrm{Tr}_{M_a}[|m\rangle\langle m|_{M_a} (\mathrm{id} \otimes \Pi)(\Phi_{AA'} \otimes |m\rangle\langle m|_{M_b'})] \\ &= \sum_m \mathrm{Tr}_{M_a M_b}[|mm\rangle\langle mm|_{M_a M_b} \Omega_{M_a M_b AB}] \\ &= \mathrm{Tr}_{M_a M_b} D_{M_a M_b} \Omega_{M_a M_b AB}, \end{aligned}$$

where $D_{M_a M_b} = \sum_m |mm\rangle\langle mm|_{M_a M_b}$ is the Choi-Jamiołkowski matrix of the noiseless classical channel $\mathcal{I}_M$, as before.

In summary, to simulate $\mathcal{N}$ exactly, we have

$$J_{AB} = \mathrm{Tr}_{M_a M_b} D_{M_a M_b} \Omega_{AB M_a M_b} \text{ s.t. } \Omega \geq 0,\ \mathrm{Tr}_{M_a B}\, \Omega = \mathbb{1}_{AM_b},$$

and the no-signalling constraints on Ω.

By depasing and twirling classical registers M_a and M_b (refer to the case of zero-error communication), we can choose w.l.o.g $\Omega_{M_a M_b AB}$ to have the form in Eq. (22).

We end this subsection, like the previous one, recording the primal and dual SDP form of $\Sigma(K)$, again equal by strong duality:

$$\begin{aligned}\Sigma(K) = \min \operatorname{Tr} T_B \ \text{ s.t. } \ & 0 \leq V_{AB} \leq \mathbb{1}_A \otimes T_B, \\ & \operatorname{Tr}_B V_{AB} = \mathbb{1}_A, \\ & \operatorname{Tr}(\mathbb{1} - P)_{AB} V_{AB} = 0.\end{aligned} \tag{23}$$

Its dual SDP is

$$\begin{aligned}\Sigma(K) = \max \operatorname{Tr} S_A \ \text{ s.t. } \ & 0 \leq U_{AB},\ \operatorname{Tr}_A U_{AB} = \mathbb{1}_B, \\ & P_{AB}(S_A \otimes \mathbb{1}_B - U_{AB})P_{AB} \leq 0.\end{aligned} \tag{24}$$

Remark Just as for the no-signalling assisted independence number $\Upsilon(K)$, $\Sigma(\Delta_\ell) = \ell$ and hence $\Sigma(K \otimes \Delta_\ell) \leq \ell\, \Sigma(K)$. For cq-graphs K we can say a bit more: since $\mathcal{I}_\ell$ is a cq-graph, too, we find $\Sigma(K \otimes \Delta_\ell) = \ell\, \Sigma(K)$, by Proposition 18 in Section IV C below. □

IV. TOWARDS ASYMPTOTIC CAPACITY AND COST

For a classical channel with bipartite graph Γ, such that

$$K = \operatorname{span}\{|y\rangle\langle x| : (x, y) \text{ edge in } \Gamma\}$$

is a special type of cq-graph, it was shown in [12] that

$$C_{0,\text{NS}}(K) = S_{0,\text{NS}}(K) = \log \mathsf{A}(K) = \log \alpha^*(\Gamma),$$

where $\alpha^*(\Gamma)$ is the *fractional packing number* [5] (which is equal to its *fractional covering number*):

$$\begin{aligned}\alpha^*(\Gamma) &= \max \sum_x p_x \ \text{ s.t. } \ \sum_x p_x \Gamma(y|x) \leq 1\ \forall y,\ 0 \leq p_x \leq 1\ \forall x, \\ &= \min \sum_j q_j \ \text{ s.t. } \ \sum_y q_y \Gamma(y|x) \geq 1\ \forall x,\ q_y \geq 1\ \forall y.\end{aligned} \tag{25}$$

In fact, there it was shown (and one can check immediately from Theorems 1 and 3) that in this case

$$\Upsilon(K) = \Sigma(K) = \mathsf{A}(K) = \alpha^*(\Gamma).$$

Furthermore, to attain the simulation cost $S_{0,\text{NS}}(\mathcal{N})$, as well as $S_{0,\text{NS}}(K)$, asymptotically no non-local resources beyond shared randomness are necessary [12]. We will see in the following that quantum channels exhibit more complexity. Indeed, while $2^{-H_{\min}(A|B)_J}$ is clearly multiplicative in the channel (or equivalently in J), cf. [24, 25], and $\alpha^*(\Gamma)$ is well-known to be multiplicative under direct products of graphs, cf. [12], by contrast Υ and Σ are only super- and sub-multiplicative, respectively:

$$\begin{aligned}\Upsilon(K_1 \otimes K_2) &\geq \Upsilon(K_1)\Upsilon(K_2), \\ \Sigma(K_1 \otimes K_2) &\leq \Sigma(K_1)\Sigma(K_2).\end{aligned}$$

We know that the first inequality can be strict (see subsection IV D below), and suspect that the second can be strict, too. Thus we are facing a regularization issue to compute the zero-error capacity and the zero-error simulation cost, assisted by no-signalling correlations:

$$\begin{aligned} C_{0,\mathrm{NS}}(K) &= \lim_{n\to\infty} \frac{1}{n} \log \Upsilon(K^{\otimes n}) = \sup_n \frac{1}{n} \log \Upsilon(K^{\otimes n}), \\ S_{0,\mathrm{NS}}(K) &= \lim_{n\to\infty} \frac{1}{n} \log \Sigma(K^{\otimes n}) = \inf_n \frac{1}{n} \log \Sigma(K^{\otimes n}). \end{aligned}$$

A standard causality argument, together with entanglement-assisted coding for and simulation of quantum channels [2], implies also

$$C_{0,\mathrm{NS}}(K) \leq C_{\min \mathrm{E}}(K) \leq S_{0,\mathrm{NS}}(K), \tag{26}$$

where $C_{\min \mathrm{E}}(K)$ is the minimum of the entanglement-assisted classical capacity of quantum channels $\mathcal{N}$ such that $K(\mathcal{N}) < K$. With the quantum mutual information $I(\rho;\mathcal{N}) := I(A:B)_\sigma$ of the state $\sigma^{AB} = (\mathrm{id} \otimes \mathcal{N})\phi_\rho$, where ϕ_ρ is a purification of ρ:

$$\begin{aligned} C_{\min \mathrm{E}}(K) &= \min C_E(\mathcal{N}) \text{ s.t. } \mathcal{K}(\mathcal{N}) < K \\ &= \min_{\substack{\mathcal{N} \text{ s.t.} \\ \mathcal{K}(\mathcal{N}) < K}} \max_\rho \; I(\rho;\mathcal{N}) \\ &= \max_\rho \min_{\substack{\mathcal{N} \text{ s.t.} \\ \mathcal{K}(\mathcal{N}) < K}} I(\rho;\mathcal{N}). \end{aligned} \tag{27}$$

For more properties, including the above minimax formulas, and an operational interpretation of $C_{\min \mathrm{E}}(K)$ as the entanglement-assisted capacity of K (i.e. the maximum rate of block codings adapted simultaneously to all channels $\mathcal{N}^{(n)}$ such that $K(\mathcal{N}^{(n)}) < K^{\otimes n}$), we refer the reader to [16].

To put better and easier to use bounds on $C_{0,\mathrm{NS}}$ and $S_{0,\mathrm{NS}}$, we introduce the semidefinite packing number:

$$\begin{aligned} \mathsf{A}(K) &= \max \operatorname{Tr} S_A \;\text{ s.t. }\; 0 \leq S_A,\; \operatorname{Tr}_A P_{AB}(S_A \otimes \mathbb{1}_B) \leq \mathbb{1}_B \\ &= \min \operatorname{Tr} T_B \;\text{ s.t. }\; 0 \leq T_B,\; \operatorname{Tr}_B P_{AB}(\mathbb{1}_A \otimes T_B) \geq \mathbb{1}_A, \end{aligned} \tag{28}$$

which we have given in primal and dual form; this generalizes the form given in Eq. (3) for cq-graphs. It was suggested to us in the past by Aram Harrow [26] for its nice mathematical properties.

A slightly modified and more symmetric form is given by

$$\begin{aligned} \widetilde{\mathsf{A}}(K) &= \max \operatorname{Tr} S_A \;\text{ s.t. }\; 0 \leq S_A,\; \operatorname{Tr}_A P_{AB}(S_A \otimes \mathbb{1}_B) P_{AB} \leq \mathbb{1}_B \\ &= \min \operatorname{Tr} T_B \;\text{ s.t. }\; 0 \leq T_B,\; \operatorname{Tr}_B P_{AB}(\mathbb{1}_A \otimes T_B) P_{AB} \geq \mathbb{1}_A, \end{aligned} \tag{29}$$

again both in primal and dual form.

From these it is straightforward to see that $\mathsf{A}(K)$ and $\widetilde{\mathsf{A}}(K)$ are both sub- and super-multiplicative, and so

$$\begin{aligned} \mathsf{A}(K_1 \otimes K_2) &= \mathsf{A}(K_1)\mathsf{A}(K_2), \\ \widetilde{\mathsf{A}}(K_1 \otimes K_2) &= \widetilde{\mathsf{A}}(K_1)\widetilde{\mathsf{A}}(K_2). \end{aligned}$$

Both definitions reduce to the familiar notion of fractional packing number $\alpha^*(\Gamma)$ when K is associated to a bipartite graph Γ, coming from a classical channel $\mathcal{N}$. Furthermore, $\mathsf{A}(K)$ and $\widetilde{\mathsf{A}}(K)$ are equal for cq-graphs. However, we will see later that in general they are different.

A. Revised semidefinite packing number and simulation

Proposition 12 *For any non-commutative bipartite graph $K < \mathcal{L}(A \to B)$,*

$$\Sigma(K) \geq \widetilde{\mathsf{A}}(K).$$

Consequently, $S_{0,\mathrm{NS}}(K) \geq \log \widetilde{\mathsf{A}}(K)$.

Proof From the SDP (23) for $\Sigma(K)$, we get operators T_B and V_{AB} such that $0 \leq V_{AB} \leq \mathbb{1}_A \otimes T_B$, with $\mathrm{Tr}_B V_{AB} = \mathbb{1}_A$ and $V(\mathbb{1} - P) = 0$. Hence $P(\mathbb{1} \otimes T)P \geq PVP = V$ and so

$$\mathrm{Tr}_B P(\mathbb{1} \otimes T)P \geq \mathrm{Tr}_B V = \mathbb{1}_A,$$

i.e. T is feasible for the dual formulation of $\widetilde{\mathsf{A}}(K)$, hence the result follows. □

Below we will see that this bound (both in its one-shot and regularized form) is in general strict, indeed already for cq-channels it is not an equality.

B. Asymptotic assisted zero-error capacity of cq-graphs

We do not know whether $\Upsilon(K)$ is in general related to $\mathsf{A}(K)$, but we will show bounds in either direction for cq-channels.

Suppose that the cq-channel $\mathcal{N}$ acts as $|i\rangle\langle i| \longmapsto \rho_i$, with support K_i and support projection P_i of ρ_i. Then, the non-commutative bipartite graph K associated with $\mathcal{N}$ is given by $K = \bigoplus_i |i\rangle \otimes K_i$, and the projection for the Choi-Jamiołkowski matrix is $P = \sum_i |i\rangle\langle i|^A \otimes P_i^B$. The SDP (20) easily simplifies to

$$\begin{aligned} \Upsilon(K) = \max \sum_i s_i \ \text{ s.t. } \ 0 \leq s_i,\ 0 \leq R_i \leq s_i(\mathbb{1} - P_i), \\ \sum_i (s_i P_i + R_i) = \mathbb{1}. \end{aligned} \tag{30}$$

The semidefinite packing number (28), on the other hand, simplifies to

$$\mathsf{A}(K) = \max \sum_i s_i \ \text{ s.t. } \ 0 \leq s_i,\ \sum_i s_i P_i \leq \mathbb{1}. \tag{31}$$

As this is an SDP relaxation of the problem (30), we obtain:

Lemma 13 *For a non-commutative bipartite cq-graph $K < \mathcal{L}(A \to B)$,*

$$\Upsilon(K) \leq \mathsf{A}(K).$$

Consequently, $C_{0,\mathrm{NS}}(K) \leq \log \mathsf{A}(K)$. □

In general, $\Upsilon(K)$ can be strictly smaller than $\mathsf{A}(K)$, see Subsection IV D below, but we shall prove equality for the regularization, by exhibiting a lower bound

$$\Upsilon(K^{\otimes n}) \geq n^{-O(1)} \mathsf{A}(K^{\otimes n}) = n^{-O(1)} \mathsf{A}(K)^n.$$

The way to do this is to take a feasible solution s_i of the SDP (31); w.l.o.g. its value $\sum_i s_i > 1$, otherwise the above statement is trivial. On strings $\underline{i} = i_1 \ldots i_n$ of length n this gives a feasible solution $s_{\underline{i}} = s_{i_1} \cdots s_{i_n}$ for $K^{\otimes n}$, with value

$$\sum_{\underline{i}} s_{\underline{i}} = \left(\sum_i s_i \right)^n.$$

Now for a different symbols $i = 1, \dots, a$ there are at most $(n+1)^a$ many types of strings, hence there is one type τ such that

$$\sum_{\underline{i}\in\tau} s_{\underline{i}} \geq (n+1)^{-a} \left(\sum_i s_i \right)^n .$$

Restricting the input of the channel to $\underline{i} \in \tau$ (while the output is still B^n), we thus loose only a polynomial factor of the semidefinite packing number. What we gain is that the inputs are all of the same type, which means that the output projectors $P_{\underline{i}} = P_{i_1} \otimes \cdots \otimes P_{i_n}$ are related to each other by unitaries U^π permuting the n B-systems. Note that then also all of the $s_{\underline{i}}$ on the left hand side above are the same, say s_τ, and hence the left hand side is $s_\tau |\tau|$.

Abstractly, we are thus in the following situation: Assume that there exists a transitive group action by unitary conjugation on the P_i, i.e. we have a finite group G acting transitively on the labels i (running over a set of size N), and a unitary representation U^g, such that $P_{i^g} = (U^g)^\dagger P_i U^g$ for $g \in G$. In other words, the entire set of $\{P_i\}$ is the orbit $\left\{(U^g)^\dagger P_0 U^g\right\}$ of a fiducial element P_0 under the group action. Then we can twirl the SDP (31) and w.l.o.g. assume that all s_i are identical s, so the constraint reduces to $s \sum_i P_i \leq \mathbb{1}$, meaning that the largest admissible s is $\|\sum_i P_i\|_\infty^{-1}$, and the semidefinite packing number $\mathsf{A}(K) = sN = \left\|\frac{1}{N}\sum_i P_i\right\|_\infty^{-1}$.

From this we see that the representation theory of U^g has a bearing on the semidefinite packing number $\mathsf{A}(K)$; cf. [30] for some basic facts that we shall invoke in the following. Indeed, it also governs $\Upsilon(K)$ since we can do the same twirling operation and find that in the SDP (30) we have w.l.o.g. that all s_i are equal to the same s, but also that for $P_j = (U^g)^\dagger P_i U^g$, w.l.o.g. $R_j = (U^g)^\dagger R_i U^g$. In particular, in this case

$$\begin{aligned} \Upsilon(K) = \max sN \;\text{ s.t. }\; 0 \leq s,\; 0 \leq R_0 \leq s(\mathbb{1} - P_0), \\ \frac{1}{|G|}\sum_g (U^g)^\dagger (sP_0 + R_0) U^g = \frac{1}{N}\mathbb{1}. \end{aligned} \tag{32}$$

Let us be a little more explicit in the reduction of the SDP (30) to the above SDP: Let s and R_0 be feasible as above, and denote by G_0 the subgroup of G leaving 0 invariant, $G_0 = \{g \in G : 0^g = 0\}$. By Lagrange's Theorem, $N = |G/G_0|$. Then,

$$\overline{R}_0 := \frac{1}{|G_0|}\sum_{g\in G_0} (U^g)^\dagger R_0 U^g$$

is also feasible with the same s, using $(U^g)^\dagger P_0 U^g = P_0$ for all $g \in G_0$. Letting $\overline{R}_i := (U^g)^\dagger \overline{R}_0 U^g$ for any g such that $i = 0^g$, and $s_i = s$, then yields a feasible solution for (30) – and this is well-defined. Thus $\Upsilon(K)$ is not smaller than the above SDP. In the other direction, let s_i and R_i be feasible for (30), i.e. $0 \leq R_i \leq s_i(\mathbb{1} - P_i)$. Letting $\overline{s} = \frac{1}{N}\sum_i s_i$ and

$$\overline{R}_0 := \frac{1}{G}\sum_{g\in G} U^g R_{0^g} (U^g)^\dagger$$

yields a feasible solution for (32).

If the representation U^g happens to be irreducible, we are lucky because then the group average in the second line in Eq. (32) is automatically proportional to the identity, by Schur's Lemma. Hence the optimal choice is $R_0 = R_i = 0$ and we find $\Upsilon(K) = \mathsf{A}(K)$. In general this won't be the case, but if the representation U^g is "not too far" from being irreducible, in a sense made precise in the following proposition, then $\Upsilon(K)$ is not too much smaller than $\mathsf{A}(K)$:

Proposition 14 *For a set of projections P_i on B with a transitive group action by conjugation under U^g, let*

$$B = \bigoplus_\lambda \mathcal{Q}_\lambda \otimes \mathcal{R}_\lambda$$

be the isotypical decomposition of B into irreps $\mathcal{Q}_\lambda$ of U^g, with multiplicity spaces $\mathcal{R}_\lambda$. Denote the number of terms λ by L, and the largest occurring multiplicity by $M = \max_\lambda |\mathcal{R}_\lambda|$. Then, for the corresponding cq-graph K,

$$\Upsilon(K) \geq \frac{1}{4L^2 M^{9/2}} \mathsf{A}(K),$$

if $\mathsf{A}(K) \geq 64 L^6 M^{14}$.

Proof Assume that we have a feasible s^* for $\mathsf{A}(K)$ such that $s^* N \geq 64 L^6 M^{14}$.

Our point of departure is the SDP (32): for given $s \geq 0$ and $R_0 \geq 0$, Schur's Lemma tells us

$$\frac{1}{|G|} \sum_g (U^g)^\dagger (sP_0 + R_0) U^g = \frac{1}{N} \sum_\lambda Q_\lambda \otimes \zeta_\lambda,$$

where Q_λ is the projection onto the irrep $\mathcal{Q}_\lambda$, ζ_λ is a semidefinite operator on $\mathcal{R}_\lambda$. Feasibility of s and R_0 (to be precise: the equality constraints) is equivalent to $\zeta_\lambda = \Pi_\lambda$, the projection onto $\mathcal{R}_\lambda$, for all λ.

Now, for each λ choose an orthogonal basis $\{Z_\mu^{(\lambda)}\}$ of Hermitians over $\mathcal{R}_\lambda$, with $Z_0^{(\lambda)} = \frac{1}{\operatorname{Tr}\Pi_\lambda}\Pi_\lambda$ and $\|Z_\mu^{(\lambda)}\|_2 = 1$ for $\mu \neq 0$. Then the $\frac{1}{\operatorname{Tr} Q_\lambda} Q_\lambda \otimes Z_\mu^{(\lambda)}$ form a basis of the U^g-invariant operators, hence our SDP can be rephrased as

$$\begin{aligned} \Upsilon(K) = \max sN \ \text{ s.t. } \ 0 \leq s,\ 0 \leq R_0 \leq s(\mathbb{1} - P_0), \\ \forall \lambda\, \mu \quad \operatorname{Tr}(sP_0 + R_0)\left(\frac{Q_\lambda}{\operatorname{Tr} Q_\lambda} \otimes Z_\mu^{(\lambda)} \right) = \frac{1}{N}\delta_{\mu 0}. \end{aligned} \tag{33}$$

Notice that here, the semidefinite constraints on R_0 leave quite some room, whereas we have "only" LM^2 linear conditions to satisfy. Given s^* satisfying the constraint of $\mathsf{A}(K)$, our strategy now will be to show that we can construct a $0 \leq R_0 \leq \frac{2\beta}{N}(\mathbb{1} - P_0)$ such that the above equations are true with $s = s^*$ on the left hand side, and with a factor β on the right hand side. We will choose $\beta = 4L^2 M^{9/2}$ and thus there is a feasible solution with $s = s^*/\beta$ to (33), hence $\Upsilon(K) \geq s^* N/\beta$ as claimed.

In detail, introduce a new variable $X \geq 0$, with

$$R_0 = \frac{\beta}{N}(\mathbb{1} - P_0) X (\mathbb{1} - P_0),$$

which makes sure that R_0 is automatically supported on the complement of P_0. Rewrite the equations

$$\operatorname{Tr}(s^* P_0 + R_0)\left(\frac{Q_\lambda}{\operatorname{Tr} Q_\lambda} \otimes Z_\mu^{(\lambda)} \right) = \frac{\beta}{N}\delta_{\mu 0}.$$

in terms of X, introducing the notation

$$C_{\lambda\mu} = \frac{1}{\operatorname{Tr} Q_\lambda} Q_\lambda \otimes Z_\mu^{(\lambda)}, \quad D_{\lambda\mu} = (\mathbb{1} - P_0) C_{\lambda\mu} (\mathbb{1} - P_0).$$

This gives, noting $\mathrm{Tr}\, P_0 C_{\lambda\mu} = \mathrm{Tr}\, P_i C_{\lambda\mu}$ for all i because of the U^g invariance of $C_{\lambda\mu}$,

$$
\begin{aligned}
\mathrm{Tr}\, X D_{\lambda\mu} &= \delta_{\mu 0} - \frac{1}{\beta} \mathrm{Tr}\ s^* N P_0 C_{\lambda\mu} \\
&= \delta_{\mu 0} - \frac{1}{\beta} \mathrm{Tr} \left(\sum_i s^* P_i \right) C_{\lambda\mu} \\
&=: \delta_{\mu 0} - \frac{1}{\beta} t_{\lambda\mu}.
\end{aligned}
\tag{34}
$$

What we need of the coefficients $t_{\lambda\mu}$ is that they cannot be too large: from $0 \le \sum_i s^* P_i \le \mathbb{1}$ we get

$$
|t_{\lambda\mu}| \le \|C_{\lambda\mu}\|_1 = \|Z_\mu^{(\lambda)}\|_1 \le \sqrt{M}. \tag{35}
$$

Our goal will be to find a "nice" dual set $\{\widehat{D}_{\lambda\mu}\}$ to the $\{D_{\lambda\mu}\}$, i.e. $\mathrm{Tr}\, D_{\lambda\mu} \widehat{D}_{\lambda'\mu'} = \delta_{\lambda\lambda'} \delta_{\mu\mu'}$, with which we can write a solution $X = \sum_{\lambda\mu} \left(\delta_{\mu 0} - \frac{1}{\beta} t_{\lambda\mu} \right) \widehat{D}_{\lambda\mu}$. To this end, we construct first the dual set $\widehat{C}_{\lambda\mu}$ of the $\{C_{\lambda_\mu}\}$, which is easy:

$$
\widehat{C}_{\lambda\mu} = Q_\lambda \otimes \widehat{Z}_\mu^{(\lambda)} = \begin{cases} Q_\lambda \otimes \Pi_\lambda & \text{for } \mu = 0, \\ Q_\lambda \otimes Z_\mu^{(\lambda)} & \text{for } \mu \neq 0, \end{cases}
$$

so that indeed $\mathrm{Tr}\, C_{\lambda\mu} \widehat{C}_{\lambda'\mu'} = \delta_{\lambda\lambda'} \delta_{\mu\mu'}$. Now, consider the $LM^2 \times LM^2$-matrix T,

$$
\begin{aligned}
T_{\lambda\mu,\lambda'\mu'} &= \mathrm{Tr}\, D_{\lambda\mu} \widehat{C}_{\lambda'\mu'} \\
&= \mathrm{Tr} (\mathbb{1} - P_0) C_{\lambda\mu} (\mathbb{1} - P_0) \widehat{C}_{\lambda'\mu'} \\
&= \delta_{\lambda\lambda'} \delta_{\mu\mu'} - \Delta_{\lambda\mu,\lambda'\mu'},
\end{aligned}
$$

with the deviation

$$
\Delta_{\lambda\mu,\lambda'\mu'} = \mathrm{Tr}\, P_0 C_{\lambda\mu} (\mathbb{1} - P_0) \widehat{C}_{\lambda'\mu'} + \mathrm{Tr}\, C_{\lambda\mu} P_0 \widehat{C}_{\lambda'\mu'}.
$$

Here,

$$
\begin{aligned}
|\Delta_{\lambda\mu,\lambda'\mu'}| &\le 2 \|P_0 C_{\lambda\mu}\|_1 \|\widehat{C}_{\lambda'\mu'}\|_\infty \le 2 \|P_0 C_{\lambda\mu}\|_1 \\
&= 2 \big\| P_0 |C_{\lambda\mu}| \big\|_1 \\
&\le 2 \sqrt{\mathrm{Tr}\, P_0 |C_{\lambda\mu}|} \sqrt{\|C_{\lambda\mu}\|_1},
\end{aligned}
$$

using $\|\widehat{C}_{\lambda'\mu'}\|_\infty \le 1$, the unitary invariance of the trace norm, and Lemma 15 below. Since $|C_{\lambda\mu}| = \frac{1}{\mathrm{Tr}\, Q_\lambda} Q_\lambda \otimes \big| Z_\mu^{(\lambda)} \big|$ is invariant under the action of U^g, we have $\mathrm{Tr}\, P_0 |C_{\lambda\mu}| = \mathrm{Tr}\, P_i |C_{\lambda\mu}|$ for all i, and using $\sum_i s^* P_i \le \mathbb{1}$ we get

$$
|\Delta_{\lambda\mu,\lambda'\mu'}| \le 2 \sqrt{\frac{1}{s^* N} \|C_{\lambda\mu}\|_1^2} \le 2\sqrt{M} (s^* N)^{-1/2}. \tag{36}
$$

With this we get that

$$
\begin{aligned}
\|T - \mathbb{1}\|_\infty \le \|T - \mathbb{1}\|_2 &= \sqrt{\sum_{\lambda\mu\lambda'\mu'} |\Delta_{\lambda\mu,\lambda'\mu'}|^2} \\
&\le \sqrt{L^2 M^4 4M (s^* N)^{-1}} \le \frac{1}{\beta},
\end{aligned}
\tag{37}
$$

if $s^*N \geq 4\beta^2 L^2 M^5$. Assuming $\beta \geq 2$ (which will be the case with our later choice), we thus know that T is invertible; in fact, we have $T = \mathbb{1} - \Delta$ with $\|\Delta\| \leq \frac{1}{\beta} \leq \frac{1}{2}$, hence $T^{-1} = \sum_{k=0}^{\infty} \Delta^k$ and so

$$\left\|T^{-1} - \mathbb{1}\right\|_\infty = \left\|\sum_{k=1}^{\infty} \Delta^k\right\|_\infty \leq \sum_{k=1}^{\infty} \|\Delta\|_\infty^k = \frac{1}{\beta - 1} \leq \frac{2}{\beta}.$$

I.e., writing $T^{-1} = \mathbb{1} + \widetilde{\Delta}_{\lambda\mu,\lambda'\mu'}$ we get

$$|\widetilde{\Delta}_{\lambda\mu,\lambda'\mu'}| \leq \|\widetilde{\Delta}\|_\infty \leq \frac{2}{\beta}. \tag{38}$$

The invertibility of T implies that there is a dual set to $\{D_{\lambda\mu}\}$ in $\operatorname{span}\{\widehat{C}_{\lambda\mu}\}$. Indeed, from the definition of $T_{\lambda\mu,\lambda'\mu'}$ and the dual sets,

$$\begin{aligned} \widehat{C}_{\lambda'\mu'} &= \sum_{\lambda\mu} T_{\lambda\mu,\lambda'\mu'} \widehat{D}_{\lambda\mu}, \quad \text{which can be rewritten as} \\ \widehat{D}_{\lambda\mu} &= \sum_{\lambda'\mu'} (T^{-1})_{\lambda'\mu',\lambda\mu} \widehat{C}_{\lambda'\mu'}. \end{aligned}$$

Now we can finally write down our candidate solution to Eqs. (34):

$$\begin{aligned} X &= \sum_{\lambda\mu} \left(\delta_{\mu 0} - \frac{1}{\beta} t_{\lambda\mu}\right) \widehat{D}_{\lambda\mu} \\ &= \sum_{\lambda\mu} \left(\delta_{\mu 0} - \frac{1}{\beta} t_{\lambda\mu}\right) \sum_{\lambda'\mu'} (T^{-1})_{\lambda'\mu',\lambda\mu} \widehat{C}_{\lambda'\mu'} \\ &= \sum_{\lambda} \widehat{C}_{\lambda 0} - \frac{1}{\beta} \sum_{\lambda\mu} t_{\lambda\mu} \sum_{\lambda'\mu'} (T^{-1})_{\lambda'\mu',\lambda\mu} \widehat{C}_{\lambda'\mu'} + \sum_{\lambda\lambda'\mu'} \widetilde{\Delta}_{\lambda'\mu',\lambda 0} \widehat{C}_{\lambda'\mu'} \\ &= \mathbb{1} + \text{Rest}. \end{aligned}$$

The remainder term "Rest" can be bounded as follows:

$$\begin{aligned} \|\text{Rest}\|_\infty &\leq \frac{1}{\beta} \sum_{\lambda\mu\lambda'\mu'} 2\sqrt{M} + \sum_{\lambda\lambda'\mu'} \frac{2}{\beta} \\ &= \frac{2}{\beta} \left(L^2 M^{9/2} + L^2 M^2\right) \leq \frac{4}{\beta} L^2 M^{9/2}, \end{aligned}$$

using Eqs. (35) and (38). Thus we find $\|\text{Rest}\|_\infty \leq 1$ if $\beta \geq 4L^2 M^{9/2}$ and $s^*N \geq 4\beta^2 L^2 M^5 \geq 64 L^6 M^{14}$. In this case, $0 \leq X \leq 2$ and we can wrap things up: $R_0 := \frac{\beta}{N}(\mathbb{1} - P_0) X (\mathbb{1} - P_0)$ satisfies

$$0 \leq R_0 \leq \frac{2\beta}{N}(1 - P_0) \leq s^*(\mathbb{1} - P_0),$$

as well as

$$\frac{1}{|G|} \sum_{g \in G} (U^g)^\dagger (s^* P_0 + R_0) U^g = \frac{\beta}{N} \mathbb{1}.$$

I.e., we get a feasible solution $\sum_i \left(\frac{s^*}{\beta} P_i + \frac{1}{\beta} R_i\right) = \mathbb{1}$ for $\Upsilon(K)$. $\square$

Lemma 15 *Let ρ be a state and P a projection in a Hilbert space $\mathcal{H}$. Then,*

$$\operatorname{Tr}\rho P \leq \|\rho P\|_1 \leq \sqrt{\operatorname{Tr}\rho P}.$$

More generally, for $X \geq 0$ and a POVM element $0 \leq E \leq \mathbb{1}$,

$$\operatorname{Tr} XE \leq \|XE\|_1 \leq \sqrt{\operatorname{Tr} X}\sqrt{\operatorname{Tr} XE}.$$

Proof We start with the first chain of inequalities. The left hand one follows directly from the definition of the trace norm. For the right hand one, choose a purification of $\rho = \operatorname{Tr}_{\mathcal{H}'} |\psi\rangle\langle\psi|$ on $\mathcal{H} \otimes \mathcal{H}'$. Now, $\|\rho P\|_1 = \operatorname{Tr} \sqrt{P\rho^2 P}$ and $\rho P = \operatorname{Tr}_{\mathcal{H}'} |\psi\rangle\langle\psi|(P \otimes \mathbb{1})$. Thus, by the monotonicity of the trace norm under partial trace,

$$\begin{aligned}
\|\rho\Pi\|_1 &\leq \big\||\psi\rangle\langle\psi|(P \otimes \mathbb{1})\big\|_1 \\
&= \operatorname{Tr} \sqrt{(P \otimes \mathbb{1})|\psi\rangle\langle\psi|(P \otimes \mathbb{1})} \\
&= \sqrt{\operatorname{Tr} |\psi\rangle\langle\psi|(P \otimes \mathbb{1})} \\
&= \sqrt{\operatorname{Tr}\rho P}.
\end{aligned}$$

The second chain is homogenous in X, so we may w.l.o.g. assume that $\operatorname{Tr} X = 1$, i.e. $X = \rho$ is a state. For a general POVM element E there is an embedding U of the Hilbert space $\mathcal{H}$ into a larger Hilbert space $\mathcal{H}_0$ and a projection P in $\mathcal{H}_0$ such that $E = U^\dagger P U$. Then, $\operatorname{Tr}\rho E \leq \|\rho E\|_1$ as before by the definition of the trace norm, and using the invariance of the trace number under unitaries and the first part,

$$\|\rho E\|_1 = \|\rho U^\dagger P U\|_1 = \|U\rho U^\dagger P\|_1 \leq \sqrt{\operatorname{Tr} U\rho U^\dagger P} = \sqrt{\operatorname{Tr}\rho E},$$

which concludes the proof. □

For the permutation action of S_n on B^n, the irreps λ are labelled by Young diagrams with at most $b = |B|$ rows, hence $L \leq (n+1)^b$, and it is well-known that $M \leq (n+b)^{\frac{1}{2}b^2}$ [31, Sec. 6.2], [32, 33]. Thus the previous proposition yields directly the following result, observing that L and M are polynomially bounded in n, whilst $\mathsf{A}(K^{\otimes n}) = \mathsf{A}(K)^n$ grows exponentially.

Proposition 16 *Let K be a non-commutative bipartite cq-graph with $a = |A|$ inputs and output dimension $b = |B|$. Then for sufficiently large n,*

$$\Upsilon(K^{\otimes n}) \geq \frac{1}{4(n+1)^{a+2b}(n+b)^{9b^2/4}} \mathsf{A}(K)^n.$$

Consequently, $C_{0,\mathrm{NS}}(K) = \log \mathsf{A}(K)$. □

Lemma 13 and Proposition 16 together prove Theorem 4.

Corollary 17 *For any two non-commutative bipartite cq-graphs K_1 and K_2, $C_{0,\mathrm{NS}}(K_1 \otimes K_2) = C_{0,\mathrm{NS}}(K_1) + C_{0,\mathrm{NS}}(K_2)$.* □

C. Asymptotic assisted zero-error simulation cost of cq-graphs

Here we study the asymptotic zero-error simulation cost of a non-commutative bipartite cq-graph $K = \bigoplus_i |i\rangle \otimes K_i$, where the subspace K_i is the support of the projection P_i. Thus, $P = \sum_i |i\rangle\langle i|^A \otimes P_i^B$ and the SDP (23) easily simplifies to

$$\Sigma(K) = \min \operatorname{Tr} T \;\text{ s.t. }\; T \geq V_i,\; 0 \leq V_i \leq P_i,\; \operatorname{Tr} V_i = 1. \tag{39}$$

Similarly, the dual SDP (24) simplifies to

$$\Sigma(K) = \max \sum_i s_i \;\text{ s.t. }\; s_i P_i \leq P_i U_i P_i, 0 \leq U_i,\; \sum_i U_i = \mathbb{1}. \tag{40}$$

Proposition 18 *For non-commutative bipartite cq-graphs K, $\Sigma(K)$ is multiplicative under tensor products, i.e.*

$$\Sigma(K_1 \otimes K_2) = \Sigma(K_1)\Sigma(K_2),$$

where K_1 and K_2 are arbitrary non-commutative bipartite cq-graphs.

Proof The sub-multiplicativity of $\Sigma(K)$ is evident from (39). We will show that the super-multiplicativity follows from the dual SDP (40). Indeed, let K_1 and K_2 correspond to $\{P_i\}$ and $\{Q_j\}$, respectively, and assume that (s_i, U_i) and (t_j, V_j) are optimal solutions to $\Sigma(K_1)$ and $\Sigma(K_2)$ in dual SDPs, respectively. Then we have

$$s_i = \lambda_{\min}(P_i U_i P_i), \text{ with } \sum_i U_i = \mathbb{1}_1,\; U_i \geq 0, \text{ and } \Sigma(K_1) = \sum_i s_i.$$

where $\lambda_{\min}(\cdot)$ denotes the minimal eigenvalue of the linear operator $P_i U_i P_i$ in the support of P_i. Similarly, we have

$$t_j = \lambda_{\min}(Q_j V_j Q_j), \text{ with } \sum_j V_j = \mathbb{1}_2,\; V_j \geq 0, \text{ and } \Sigma(K_2) = \sum_j t_j.$$

Clearly, we have

$$s_i t_j = \lambda_{\min}\big((P_i \otimes Q_j)(U_i \otimes V_j)(P_i \otimes Q_j)\big), \text{ and } \sum_{ij} U_i \otimes V_j = \mathbb{1}_1 \otimes \mathbb{1}_2, \text{ and } U_i \otimes V_j \geq 0.$$

So $(s_i t_j, U_i \otimes V_j)$ is a feasible solution to the dual SDP for $\Sigma(K_1 \otimes K_2)$. Since the dual SDP takes maximization, we have

$$\Sigma(K_1 \otimes K_2) \geq \sum_{ij} s_i t_j = \left(\sum_i s_i\right)\left(\sum_j t_j\right) = \Sigma(K_1)\Sigma(K_2).$$

□

From the above result we can read off directly

Theorem 19 *For any non-commutative bipartite cq-graph $K < \mathcal{L}(A \to B)$,*

$$S_{0,NS}(K) = \log \Sigma(K).$$

In fact,

$$S_{0,NS}(K_1 \otimes K_2) = S_{0,NS}(K_1) + S_{0,NS}(K_2),$$

for any two non-commutative bipartite cq-graphs K_1 and K_2. □

This theorem motivates us to call $\Sigma(K)$ the *semidefinite covering number*, at least for cq-graphs K, in analogy to a result from [10] which states that the zero-error simulation rate if a bipartite graph is given by its fractional packing number. Note however that while fractional packing and fractional covering number are dual linear programmes and yield the same value, the semidefinite versions $\mathsf{A}(K) \geq \Sigma(K)$ are in general distinct; already in the following Subsection IV D we will see a simple example for strict inequality.

For general non-commutative bipartite graph K, we do not know whether the one-shot simulation cost also gives the asymptotic simulation cost. However, this is true when $K = \operatorname{span}\{E_i\}$ corresponds to an extremal channel $\mathcal{N}(\rho) = \sum_i E_i \rho E_i^\dagger$, which is well-known to be equivalent to the set of linear operators $\{E_i^\dagger E_j\}$ being linearly independent [34]. Actually, in this case, there can only be a unique channel $\mathcal{N}$ such that $K = K(\mathcal{N})$, and furthermore, the unique channel $\mathcal{N}^{\otimes n}$ such that $K^{\otimes n} = K(\mathcal{N}^{\otimes n})$. Hence

$$\Sigma(K^{\otimes n}) = \Sigma(\mathcal{N}^{\otimes n}) = n\Sigma(\mathcal{N}) = n\Sigma(K).$$

Thus we have the following result:

Theorem 20 *Let $K = \operatorname{span}\{E_i\}$ be an extremal non-commutative bipartite graph in the sense that $\{E_i^\dagger E_j\}$ is linearly independent. Then*

$$S_{0,NS}(K) = \log \Sigma(K) = -H_{\min}(A|B)_J,$$

for the Choi-Jamiołkowski state of the unique channel $\mathcal{N}$ with $K(\mathcal{N}) < K$. Furthermore,

$$S_{0,NS}(K_1 \otimes K_2) = S_{0,NS}(K_1) + S_{0,NS}(K_2),$$

if both K_1 and K_2 are extremal non-commutative bipartite graphs. □

The fact that the set of extremal non-commutative bipartite graphs has a one-to-one correspondence to the set of extremal quantum channels has greatly simplified the simulation problem. How to use this property to simplify the assisted-communication problem is still unclear.

D. Example: Non-commutative bipartite cq-graphs with two output states

Here we will examine our above findings of one-shot and asymptotic capacities and simulation costs for the simplest possible cq-channel, which has only two inputs and two pure output states $P_i = |\psi_i\rangle\langle\psi_i|$, w.l.o.g.

$$|\psi_0\rangle = \alpha|0\rangle + \beta|1\rangle,$$
$$|\psi_1\rangle = \alpha|0\rangle - \beta|1\rangle,$$

with $\alpha \geq \beta = \sqrt{1-\alpha^2}$. In fact, we shall assume $\alpha > \beta > 0$ since the two equality cases are trivial (noiseless classical channel and completely noisy channel, respectively). Note $|\langle\psi_0|\psi_1\rangle| = \alpha^2 - \beta^2 = 2\alpha^2 - 1$; the non-commutative bipartite cq-graph $K = \operatorname{span}\{|\psi_0\rangle\langle 0|, |\psi_1\rangle\langle 1|\}$. We can easily work out all the optimization problems introduced before:

$$\Upsilon(K) = 1, \tag{41}$$

$$\Upsilon(K \otimes K) \geq \frac{1}{\alpha^4 + \beta^4}, \tag{42}$$

$$\Upsilon(K^{\otimes n}) \geq \frac{1}{\alpha^{2n} + \beta^{2n}} \geq \frac{1}{2\alpha^{2n}} \text{ for sufficiently large } n, \tag{43}$$

$$\mathsf{A}(K) = \frac{1}{\alpha^2} = \frac{2}{1 + |\langle\psi_0|\psi_1\rangle|}, \tag{44}$$

$$C_{\min \mathrm{E}}(K) = H(\alpha^2, \beta^2), \tag{45}$$

$$S_{0,\mathrm{NS}}(K) = 1 + \frac{1}{2}\|P_0 - P_1\|_1 = 1 + 2\alpha\beta. \tag{46}$$

Indeed, for pure state cq-channels $\mathcal{N}$, by dephasing the input the simulation of any channel with Kraus operators in $K = \mathrm{span}\{|\psi_i\rangle\langle i| : i = 0, 1\}$, we get a simulation of $\mathcal{N}$ itself, hence

$$\Sigma(K) = 1 + \frac{1}{2}\|\psi_0 - \psi_1\|_1 = 1 + 2\alpha\beta,$$

proving Eq. (41).

On the other hand, for Eqs. (44) and (46) we get

$$\begin{aligned} C_{0,\mathrm{NS}}(K) = \log \mathsf{A}(K) = \log \frac{2}{1 + |\langle\psi_0|\psi_1\rangle|} \\ &< H\left(\frac{1 + |\langle\psi_0|\psi_1\rangle|}{2}, \frac{1 - |\langle\psi_0|\psi_1\rangle|}{2}\right) = C_{\min \mathrm{E}}(K) \\ &< \log\left(1 + \frac{1}{2}\|\psi_0 - \psi_1\|_1\right) = S_{0,\mathrm{NS}}(K). \end{aligned}$$

Also $C_{\min \mathrm{E}}(K)$ is easy to compute, yielding Eq. (45).

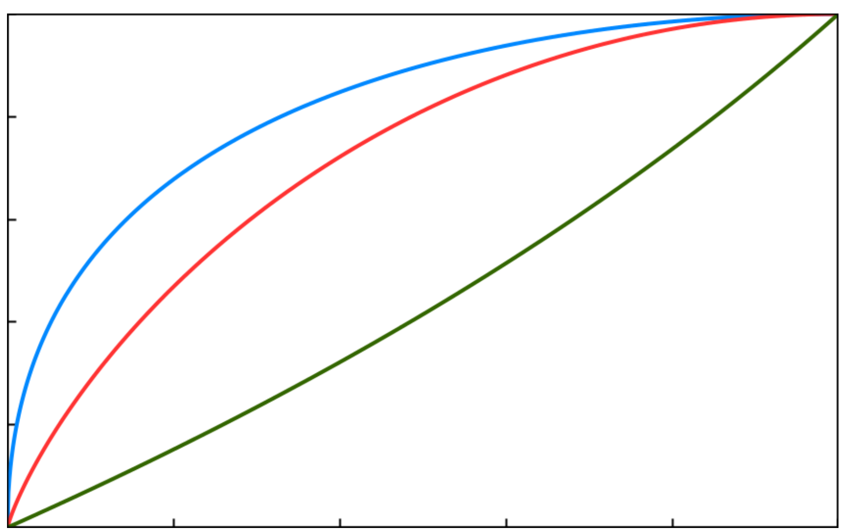

FIG. 4. Comparison between $C_{0,\mathrm{NS}}$ (green), $C_{\min \mathrm{E}}$ (red) and $S_{0,\mathrm{NS}}$ (blue) for the cq-channel of two pure states, as a function of $0 \leq \beta^2 \leq \frac{1}{2}$.

The largest effort goes into calculating or bounding the numbers $\Upsilon(K^{\otimes n})$. In this case we can obtain much better lower bounds (only at most one bit less than optimal value by n uses) compared to Theorem 4. Since the signal ensemble is symmetric under the Pauli Z unitary (which exchanges two output states), it is easy to evaluate $\mathsf{A}(K) = 1/\alpha^2$, Eq. (44).

One copy $n = 1$: Let $Q_i = \mathbb{1} - P_i$ be the projection orthogonal to P_i (which is rank-one), so that $R_i = r_i Q_i$ with a number $r_i \leq s_i$. Because of the Z-symmetry,

$$ZP_0Z = P_1, ZQ_0Z = Q_1.$$

We can symmetrize any solution and assume $s_0 = s_1 = s$ and $r_0 = r_1 = r$. Then the normalization condition reads

$$\begin{aligned}\mathbb{1} &= s(P_0 + P_1) + r(Q_0 + Q_1) \\ &= s(\alpha^2|0\rangle\langle 0| + \beta^2|1\rangle\langle 1|) + r(\beta^2|0\rangle\langle 0| + \alpha^2|1\rangle\langle 1|),\end{aligned}$$

which implies $r = s = 1/2$. Hence the maximum value of $\Upsilon(K) = 1$.

Many copies $n > 1$: In this case it is difficult to find the optimum, but it is enough that we show achievability of $1/(\alpha^{2n} + \beta^{2n})$. Note that this already implies that the SDP for the zero-error number of messages is not multiplicative! Somehow, what's happening is that the normalization condition of $\sum_i (s_i P_i + R_i) = \mathbb{1}$ is a non-trivial constraint because R_i has to be supported on the orthogonal complement of P_i; this hurt us in the case $n = 1$. Now in the case of many copies, P_i is a tensor product of single-system projectors, hence the orthogonal complement is asymptotically dominating. We have seen that how these considerations help in understanding the general case.

We have 2^n states $P_{i^n} = P_{i_1} \otimes P_{i_2} \otimes \cdots \otimes P_{i_n}$, indexed by n-bit strings i^n, which are related by qubit-wise Z-symmetry:

$$P_{i^n} = Z^{i^n} P_{0^n} Z^{i^n},$$

which motivates that we find R_{0^n} (orthogonal to P_{0^n}) and define

$$R_{i^n} := Z^{i^n} R_{0^n} Z^{i^n}.$$

Furthermore we set all s_{i^n} equal to s, and propose the following ansatz:

$$R_{0^n} = sQ_{0^n} + \sum_{w=1}^{n-1} c_w X_w, \tag{47}$$

where $X_w \geq 0$ will have non-zero matrix elements only on strings of weight w (=number of 1's), and $c_w \geq 0$ are parameters to be determined. Concretely, X_w is going to a suitable S_n-symmetrization of a singlet, tensored with a string of weight $w - 1$:

$$X_w := \frac{1}{w(n-w)} \sum_{I=\{k,l\}\subset[n]} \Psi_I^- \otimes \sum_{\text{weight}(j^{n-2})=w-1} |j^{n-2}\rangle\langle j^{n-2}|_{[n]-I},$$

with the singlet projector Ψ^-, and the index indicating the system(s) where the state sits. Note that $\|X_w\| =< n/2$, the worst case being $w = 1$. Looking at Eq. (47), observe that Q_{0^n} is orthogonal to P_{0^n} and the X_w are orthogonal to both of them because each term has a singlet and $\langle\phi|\langle\phi|\Psi^-\rangle = 0$ for arbitrary vector $|\phi\rangle$.

Now, we can calculate, using the Z^n symmetry:

$$\sum_{i^n} (sP_{i^n} + R_{i^n}) = 2^n s \left((\alpha^2|0\rangle\langle 0| + \beta^2|1\rangle\langle 1|)^{\otimes n} + (\beta^2|0\rangle\langle 0| + \alpha^2|1\rangle\langle 1|)^{\otimes n} \right) + 2^n \sum_w c_w Y_w,$$

with

$$Y_w = \frac{1}{w(n-w)} \sum_{I=\{k,l\}\subset[n]} \frac{1}{2} \left(|01\rangle\langle 01| + |10\rangle\langle 10|\right)_I \otimes \sum_{\mathrm{wt}(j^{n-2})=w-1} |j^{n-2}\rangle\langle j^{n-2}|_{[n]\setminus I}$$
$$= \sum_{\mathrm{wt}(i^n)=w} |i^n\rangle\langle i^n|.$$

Hence, letting

$$s := \frac{2^{-n}}{\alpha^{2n}+\beta^{2n}},$$
$$c_w := s\left(\alpha^{2n}+\beta^{2n}-\alpha^{2w}\beta^{2n-2w}-\beta^{2w}\alpha^{2n-2w}\right),$$

we can meet the normalization constraint $\sum_{i^n}(sP_{i^n}+R_{i^n}) = \mathbb{1}$, and the sum of the s_i comes out as

$$\sum_i s_i = 2^n s = \frac{1}{\alpha^{2n}+\beta^{2n}},$$

as advertised. It remains to check that R_{0^n} is feasible (then all the other R_{i^n} follow): the largest eigenvalue of R_{0^n} is the maximum of the numbers s and $c_w\|X_w\|$, $w = 1, \cdots, n-1$. Noting that the largest c_w is attained at $w = n/2$, we have

$$c_w\|X_w\| \le s(\alpha^n - \beta^n)^2 n/2,$$

which is smaller than s for large enough n. [For $n = 2$, the above argument shows by inspection $\Upsilon(K \otimes K) \ge \frac{1}{\alpha^4+\beta^4}$.] □

Let us consider now the more general case with two general output states ρ_0 and ρ_1 with projections P_0 and P_1, respectively. Denote

$$F_{\max}(\rho_0,\rho_1) = F_{\max}(P_0,P_1) := \max\{|\langle\psi_0|\psi_1\rangle| : \psi_i \in K_i, i = 0,1\}.$$

$F_{\max}(\rho_0,\rho_1)$ is known as the maximal fidelity between ρ_0 and ρ_1, but depends only on the supports K_i. A key property of the maximal fidelity is the following [35]:

Proposition 21 *There exists a CPTP map $\mathcal{T}$ such that $\mathcal{T}(\rho_0) = \psi_0$ and $\mathcal{T}(\rho_1) = \psi_1$ if and only if $F_{\max}(\rho_0,\rho_1) \le |\langle\psi_0|\psi_1\rangle|$.* □

Applying this result, we can show that the case of two general output states ρ_0 and ρ_1 is simply equivalent to the case of two pure output states ψ_0 and ψ_1 such that $|\langle\psi_0|\psi_1\rangle| = F_{\max}(\rho_0,\rho_1)$. Thus we have

$$C_{0,\mathrm{NS}}(K) = \log \mathsf{A}(K) = \log \frac{2}{1+F_{\max}},$$
$$C_{\min \mathrm{E}}(K) = H\left(\frac{1+F_{\max}}{2}, \frac{1-F_{\max}}{2}\right),$$
$$S_{0,\mathrm{NS}}(K) = \log \Sigma(K) = 1 + \sqrt{1-F_{\max}^2}.$$

V. AN OPERATIONAL INTERPRETATION OF THE LOVÁSZ NUMBER

As we have seen, different non-commutative bipartite graphs K, even cq-graphs, having the same confusability graph G, can have different assisted zero-error capacity $C_{0,\text{NS}}(K)$ and simulation cost $S_{0,\text{NS}}(K)$; c.f. the last subsection IV D in the previous section.

A classical undirected graph is given by $G = (V, E)$, where $V = \{1, ..., n\}$ is the set of vertices, and $E \subset V \times V$ is the set of edges. As shown in previous work [11], G is naturally associated with a non-commutative graph, denoted S, via the following way:

$$G = \operatorname{span}\{|i\rangle\langle j| : i \sim j\},$$

where $i \sim j$ means confusability: $i = j$ or $\{i, j\} \in E$ is an edge of the graph [27].

Hence the questions we are facing are the maximum and minimum $C_{0,\text{NS}}(K)$ and $S_{0,\text{NS}}(K)$ over all cq-graphs K with $K^\dagger K \subseteq G$. While the maximum is clearly $\log |B|$ for both quantities, the minima turn out to be much more interesting. We restate here the main result we will go on to prove in this section.

Theorem 5 *For any classical graph G, the Lovász number $\vartheta(G)$ is the minimum zero-error classical capacity assisted by quantum no-signalling correlations of any cq-channels that have G as non-commutative graph,* i.e.

$$\log \vartheta(G) = \min\big\{C_{0,NS}(K) : K^\dagger K \subseteq G\big\},$$

where the minimization is over cq-graphs K.

In particular, equality holds for any cq-channel $i \to |\psi_i\rangle\langle\psi_i|$ such that $\{|\psi_i\rangle\}$ is an optimal orthogonal representation for G in the sense of Lovász' original definition [27].

The proof of this result is achieved by combining two facts about the semidefinite packing number for cq-channels: 1) $\mathsf{A}(K)$ gives the zero-error classical capacity assisted with no-signaling correlations for a cq-channel; 2) the Lovász number $\vartheta(G)$ of a graph is given by the minimization of $\mathsf{A}(K)$ for all non-commutative bipartite graph K that generate the same confusability graph G. The first fact has been proven in Theorem 16, so we focus on the second for the rest of the section.

Let $K = \operatorname{span}\{E_1, \dots, E_n\}$ be a Kraus operator space with $\operatorname{Tr}(E_i^\dagger E_j) = \delta_{ij}$, and let $|\Phi\rangle = \sum_{i=1}^d |i\rangle|i\rangle$ be the non-normalized maximally entangled state. Then P, the projection on the support of the Choi-Jamiołkowski state, can be written as

$$P_{AB} = \sum_{i=1}^n (\mathbb{1} \otimes E_i)|\Phi\rangle\langle\Phi|(\mathbb{1} \otimes E_i)^\dagger. \tag{48}$$

We can rewrite the semidefinite packing number $\mathsf{A}(K)$ using these Kraus operators $\{E_i\}$:

$$\mathsf{A}(K) = \max \operatorname{Tr} R \text{ s.t. } \sum_{i=1}^n E_i R E_i^\dagger \leq \mathbb{1}_A,\ R \geq 0. \tag{49}$$

Note that $\{E_i\}$ spans a valid Kraus space of a quantum channel. So $\sum_i E_i^\dagger E_i > 0$ (positive definite). The dual SDP is

$$\mathsf{A}(K) = \min \operatorname{Tr} T \ \text{ s.t. } \ \sum_{i=1}^n E_i^\dagger T E_i \geq \mathbb{1}_A,\ T \geq 0, \tag{50}$$

and we can easily verify that both the primal and the dual are strictly feasible by choosing $R = 0$ and $T = \lambda\mathbb{1}$ (Here $\lambda > 0$ is sufficiently large), respectively. Hence strong duality holds.

We will start by deriving some minimax representations of $\mathsf{A}(K)$. Let us introduce

$$\begin{aligned}\widehat{\mathsf{A}}(K) := \min_\rho \lambda_{\max}(\mathcal{N}(\rho)) &= \min_\rho \max_\sigma \operatorname{Tr} \mathcal{N}(\rho)\sigma \\ &= \max_\sigma \min_\rho \operatorname{Tr} \mathcal{N}^\dagger(\sigma)\rho \\ &= \max_\sigma \lambda_{\min}\big(\mathcal{N}^\dagger(\sigma)\big).\end{aligned}$$

where $\lambda_{\max}(A)$ and $\lambda_{\min}(A)$ represent the maximal and the minimal eigenvalues of a Hermitian operator A, respectively, and $\mathcal{N}(\rho) = \sum_i E_i \rho E_i^\dagger$, $\mathcal{N}^\dagger(\sigma) = \sum_i E_i^\dagger \sigma E_i$ are CP maps (but not necessarily trace or unit preserving); ρ and σ range over all density operators, i.e. $\rho, \sigma \geq 0$ and $\operatorname{Tr}\rho = \operatorname{Tr}\sigma = 1$.

In the second and the fourth equalities above, we have employed the following well-known characterizations:

$$\lambda_{\max}(A) = \max_\rho \operatorname{Tr} \rho A \text{ and } \lambda_{\min}(A) = \min_\sigma \operatorname{Tr} \sigma A.$$

In the third equality we have employed the obvious equality $\operatorname{Tr} \mathcal{N}(\rho)\sigma = \operatorname{Tr} \rho N^\dagger(\sigma)$, and von Neumann's minimax theorem [36], since $\operatorname{Tr} \mathcal{N}(\rho)\sigma$ is a linear functional with respect to ρ and σ, and ρ and σ range over convex compact sets.

Lemma 22 *Under the above definitions,*

$$\mathsf{A}(K) = \frac{1}{\widehat{\mathsf{A}}(K)},$$

for any non-commutative bipartite graph K.

Proof Suppose that $\mathsf{A}(K) = \operatorname{Tr} R_0$ for some $R_0 \geq 0$. Let us construct a density operator $\rho_0 = \frac{R_0}{\operatorname{Tr} R_0}$. By the assumption $\sum_i E_i R_0 E_i^\dagger \leq \mathbb{1}_B$, we have

$$\sum_i E_i \rho_0 E_i^\dagger \leq \frac{1}{\operatorname{Tr} R_0} \mathbb{1}_B,$$

or equivalently

$$\mathcal{N}(\rho_0) \leq \frac{1}{\mathsf{A}(K)} \mathbb{1}_B.$$

By the definition of $\widehat{\mathsf{A}}(K)$, we have

$$\widehat{\mathsf{A}}(K) \leq \lambda_{\max}(\mathcal{N}(\rho_0)) \leq \frac{1}{\mathsf{A}(K)}.$$

Conversely, suppose that $\widehat{\mathsf{A}}(K) = \lambda_{\max}(\mathcal{N}(\rho_0))$ for some density operator ρ_0. Then we have $\mathcal{N}(\rho_0) \leq \widehat{\mathsf{A}}(K)\mathbb{1}_B$, and thus

$$\mathcal{N}\big(\rho_0/\widehat{\mathsf{A}}(K)\big) \leq \mathbb{1}_B.$$

That means $\rho_0/\widehat{\mathsf{A}}(K)$ is a feasible solution of the SDP defining the semidefinite packing number. By the definition of $\mathsf{A}(K)$, we know that

$$\mathsf{A}(K) \geq \operatorname{Tr} \rho_0/\widehat{\mathsf{A}}(K) = \frac{1}{\widehat{\mathsf{A}}(K)},$$

concluding the proof. □

Focussing on the special class of cq-channels, which are in some sense the direct quantum generalizations of classical channels, we are now ready for

Proof of Theorem 5 It will be more convenient to study $\widehat{\mathsf{A}}(K)$ instead of $\mathsf{A}(K)$. Actually, applying the above minimax representation to this special case, we have

$$\widehat{\mathsf{A}}(K) = \min_{\{t_i\}} \max_{\sigma} \operatorname{Tr} \sigma \left(\sum_i t_i P_i \right) = \max_{\sigma} \min_{\{t_i\}} \operatorname{Tr} \sigma \left(\sum_i t_i P_i \right) = \max_{\sigma} \min_{i} \operatorname{Tr} \sigma P_i,$$

where $\{t_i\}$ ranges over probability distributions, and σ ranges over density operators. The right-most expression motivates us to introduce some notations.

Let P_i be the projection on the support of ρ_i. Then $\{P_i\}$ is an orthogonal representation (OR) of the confusability graph G. The value of an OR $\{P_i\}$ is defined as follows:

$$\eta(\{P_i\}) = \max_{\sigma} \min_{i} \operatorname{Tr} P_i \sigma.$$

We introduce the following function of a graph G,

$$\eta(G) = \max_{\{P_i\}} \eta(\{P_i\}),$$

where the maximization ranges over all possible ORs of G. Clearly, if we require that an OR consists of only rank-one projections and σ takes only rank-one projection, then $\eta(G) = \vartheta(G)^{-1}$, the reciprocal of the Lovász number of G [27]. It has been shown in [38] that even allowing P_i to be general projection but σ to be rank-one projection, there is no difference between $\eta(G)$ and $\vartheta(G)^{-1}$. However, if σ is a mixed state, we can only have $\eta(G) \geq \vartheta(G)^{-1}$. Interestingly, we can show that equality does hold. In fact, it is evident that if $\{P_i\}$ is an OR for a graph G, then $\{P_i \otimes \mathbb{1}_B\}$ remains an OR for the same graph, where B is any auxiliary system. Now the value of $\{P_i\}$ with respect to general mixed states, is the same as the value of $\{P_i \otimes \mathbb{1}_B\}$ with respect to pure states. That is,

$$\max_{\sigma} \min_{i} \operatorname{Tr} P_i \sigma = \max_{\Psi} \min_{i} \operatorname{Tr}(P_i \otimes \mathbb{1}_B) \Psi^{AB},$$

where $\sigma = \operatorname{Tr}_B \Psi^{AB}$. The above equality follows directly from the fact that

$$\operatorname{Tr} P_i \sigma = \operatorname{Tr}(P_i \otimes \mathbb{1}_B) \Psi^{AB},$$

where Ψ^{AB} is any purification of σ.

Summarizing, we have

$$\min_K \mathsf{A}(K) = \min_K \frac{1}{\widehat{\mathsf{A}}(K)} = \frac{1}{\eta(G)} = \vartheta(G),$$

and we are done. □

As a final comment on the above proof, note that for a fixed OR $\{P_i\}$, we cannot always choose a pure state as its optimal handle – the restriction to rank-one projectors and pure state handle only emerges as we optimize over both elements.

We would like to interpret Theorem 5 to say intuitively that the zero-error capacity of a graph G assisted by no-signalling correlations is $\vartheta(G)$. The problematic part of such a manner of speaking is that there are many cq-graphs K with the same confusability graph G, but the no-signalling assisted capacity may vary with these K.

However, note that for any (finite) family of cq-graphs $K^{(\alpha)}$ such that ${K^{(\alpha)}}^\dagger K^{(\alpha)} = G$, $K_\alpha = \bigoplus_i |i\rangle \otimes K_i^{(\alpha)}$, we can construct a cq-graph K that "dominates" all of the $K^{(\alpha)}$ in the sense that any no-signalling assisted code for K can be used directly for $K^{(\alpha)}$ because actually $K^{(\alpha)} < K$:

$$K = \bigoplus_i |i\rangle \otimes K_i, \quad K_i = \bigoplus K_i(\alpha).$$

By going from direct sums to direct integrals, we can thus construct a universal cq-graph

$$\widetilde{K}(G) = \bigoplus_i |i\rangle \otimes \widetilde{K}_i$$

which dominates *all* $K^{(\alpha)}$ with confusability graph G, in fact contains them up to isomorphism. Any no-signalling assisted code for this object will deal in particular with every eligible channel simultaneously. The only caveat is that the $\widetilde{K}_i$ are subspaces in an a priori infinite dimensional Hilbert space, and all of our proofs (in particular that of Theorem 4) require finite dimension as a technical condition. We conjecture however that the capacity result of Theorem 4 still holds in that setting.

Minimum simulation cost of a confusability graph. Just as we were looking at the smallest zero-error capacity over all cq-channels with a given confusability graph G in this section, we can study the minimum simulation cost over all cq-graphs K with $K^\dagger K \subseteq G$. To be precise, we are interested in

$$\Sigma(G) := \inf\{\Sigma(K) : K \text{ cq-graph with } K^\dagger K \subseteq G\},$$

and the asymptotic simulation cost (regularization)

$$S_{0,\mathrm{NS}}(G) := \lim_{n\to\infty} \frac{1}{n} \log \Sigma(G^n),$$

where $G^n = G \times \cdots \times G$ denotes the n-fold strong graph product. The latter limit exists and equals the infimum because evidently $\log \Sigma(G \times H) \leq \log \Sigma(G) + \log \Sigma(H)$.

For a cq-graph K, we have

$$\log \mathsf{A}(K) \leq C_{\min E}(K) \leq \log \Sigma(K), \tag{51}$$

in fact for every eligible cq-channel $\mathcal{N}$ with non-commutative bipartite graph K,

$$\log \mathsf{A}(K) \leq C(\mathcal{N}) = C_E(\mathcal{N}) \leq \log \Sigma(\mathcal{N}).$$

The reason is that $\log \mathsf{A}(K) = C_{0,\mathrm{NS}}(K)$ is the zero-error capacity assisted by no-signalling correlations (Theorem 4); while $C_E(\mathcal{N}) = C(\mathcal{N})$ is the Holevo (small-error) capacity, which is the same as the entanglement-assisted capacity for cq-channels and which is not increased by any other available no-signalling resources because of the Quantum Reverse Shannon Theorem [3, 4]; and $\log \Sigma(\mathcal{N}) = -H_{\min}(X|B)_J$ is the perfect simulation cost of the channel when assisted by no-signalling resources, with the Choi-Jamiołkowski matrix $J = \sum_i |i\rangle\langle i| \otimes P_i$, Eq. (5).

In Eq. (51), $C_{\min E}(K) = C_{\min}(K)$ because K is a cq-graph, so we only need to consider cq-channels $\mathcal{N} : i \mapsto \rho_i$ in the minimization, for which

$$C_E(\mathcal{N}) = C(\mathcal{N}) = \max_{P_X} I(X:B) = \max_{P_X} S\left(\sum_i P_X(i)\rho_i\right) - \sum_i P_X(i) S(\rho_i).$$

Letting now

$$C_{\min}(G) := \inf\{C_{\min}(K) : K \text{ cq-graph with } K^\dagger K \subseteq G\},$$

we then have the following additivity result.

Lemma 23 *For any two graphs G and H,*

$$C_{\min}(G \times H) = C_{\min}(G) + C_{\min}(H).$$

Proof The subadditivity, $C_{\min}(G \times H) \leq C_{\min}(G) + C_{\min}(H)$, is evident from the definition, because if $K^\dagger K \subseteq G$ and $L^\dagger L \subseteq H$, then $(K \otimes L)^\dagger (K \otimes L) \subseteq G \times H$.

It remains to show the opposite inequality "$\geq$". This relies crucially on the minimax identity

$$C_{\min}(G) = \inf_{\mathcal{N}} \max_{P_X} I(X:B) = \max_{P_X} \inf_{\mathcal{N}} I(X:B), \tag{52}$$

where the maximum is over probability distributions P_X and the infimum is over cq-channels $\mathcal{N}$ with confusability graph contained in G. This is a special case of Sion's minimax theorem [36], since the Holevo mutual information is well-known to be concave in P_X and convex in $\mathcal{N}$, while the domain of P_X is the convex compact simplex of finite probability distributions and the domain of $\mathcal{N}$ is an infinite-dimensional convex set.

Now, for a cq-channel $\mathcal{N} : ij \mapsto \rho_{ij} \in \mathcal{S}(B)$ with confusability graph contained in $G \times H$, and an arbitrary distribution P_{XY} of the two input variables X and Y, we have

$$I(XY:B) = I(X:B) + I(Y:B|X) = I(X:B) + \sum_i P_X(i) I(Y:B|X=i). \tag{53}$$

Here, the first term refers to the cq-channel

$$\overline{\mathcal{N}} : i \mapsto \overline{\rho}_i = \sum_j P_{Y|X}(j|i)\rho_{ij},$$

while the i-th summand in the second term sum refers to the cq-channel

$$\mathcal{N}_i : j \mapsto \rho_{ij}.$$

Note that $\overline{\mathcal{N}}$ is eligible for G (since $i \not\sim i'$ implies $ij \not\sim i'j'$ for all j, j', hence $\rho_{ij} \perp \rho_{i'j'}$), while similarly for all i, $\mathcal{N}_i$ is eligible for H. Thus, in Eq. (53), we can take the infimum over eligible cq-channels, to obtain

$$\inf_{\mathcal{N}} I(XY:B) \geq \inf_{\mathcal{M}_1} I(X:C_1) + \inf_{\mathcal{M}_2} I(Y:C_2) = \inf_{\mathcal{M}_1,\mathcal{M}_2} I(X'Y':C_1C_2),$$

where the minimizations are over cq-channels $\mathcal{N}$ eligible for $G \times H$, $\mathcal{M}_1$ eligible for G and $\mathcal{M}_2$ eligible for H, whereas X' and Y' are independent copies of X and Y, i.e. they are jointly distributed according to $P_X \times P_Y$. Now, taking the maximum over distributions P_{XY} completes the proof because of the minimax formula (52). $\square$

As a corollary, we get the following chain of inequalities:

$$\log \vartheta(G) \leq C_{\min}(G) \leq S_{0,\mathrm{NS}}(G) \leq \log \Sigma(G) \leq \log \alpha^*(G). \tag{54}$$

Note that it may be true that $S_{0,\mathrm{NS}}(G) = \log \Sigma(G)$, but to prove this we would need to show the additivity relation $\log \Sigma(G \times H) = \log \Sigma(G) + \log \Sigma(H)$, which remains unknown. Another observation is that if in the respective minimizations, the channels are restricted to classical channels, then the results of [12] show that

$$\min_K \log \mathsf{A}(K) = \min_{\mathcal{N}} C(\mathcal{N}) = \min_{\mathcal{N}} \Sigma(\mathcal{N}) = \log \alpha^*(G).$$

We now demonstrate by example that the rightmost inequality in (54) can be strict, when quantum channels are considered. Namely, for even n let $G = \overline{H_n}$ the complement of the Hadamard graph H_n, whose vertices are the vectors $\{\pm 1\}^n$ and two vectors v, w are adjacent in H_n if and only if they are orthogonal in the Euclidean sense, $v^\top w = 0$. In other words, the cq-channel $\mathcal{N} : v \mapsto \frac{1}{n}|v\rangle\langle v|$ has confusability graph G. Since the output dimension is n, we see from this that $\Sigma(G) \leq \Sigma(\mathcal{N}) \leq n$, which happens to coincide with the Lovász number, $\vartheta(G) = n$, hence $\log \vartheta(G) = C_{\min}(G) = S_{0,\text{NS}}(G) = \log \Sigma(G) = \log n$. On the other hand, the clique number of G is known to be upper bounded $\omega(G) \leq 1.99^n$ [37], hence $\alpha^*(G) \geq \frac{|G|}{\omega(G)} \geq 1.005^n$, meaning $\log \alpha^*(G) \geq \Omega(n)$.

We think that also the leftmost inequality in (54) can be strict. But although the pentagon $G = C_5$ seems to be a candidate, for which we conjecture (based on some ad hoc calculations) that $C_{\min}(G) = S_{0,\text{NS}}(G) = \log \Sigma(G) = \log \alpha^*(G) = \log \frac{5}{2}$, whereas $\log \vartheta(G) = \frac{1}{2} \log 5$, a rigorous proof of this has so far eluded us.

Whether the other two inequalities can be strict remains an open question.

VI. FEASIBILITY OF ZERO-ERROR COMMUNICATION VIA GENERAL NON-COMMUTATIVE BIPARTITE GRAPH ASSISTED BY QUANTUM NO-SIGNALLING CORRELATIONS

Given a non-commutative bipartite graph K, it is important to know when K is able to send classical information exactly in the presence of quantum no-signalling correlations. It turns out that these channels can be precisely characterized; we will start with cq-graphs.

Theorem 24 *Let K be a non-commutative bipartite cq-graph specified by a set of projections P_i with supports $K_i = \operatorname{supp} P_i$. Then the following are equivalent:*

i. $C_{0,NS}(K) > 0$;

ii. $\mathsf{A}(K) > 1$;

iii. $\bigcap_i K_i = 0$.

Proof The equivalence of i) and ii) follows directly from Theorem 16.

The equivalence of ii) and iii) is only a simple application of the SDP of $\mathsf{A}(K)$ for bipartite cq-graphs. First, we show that iii) implies ii). By contradiction, assume that the intersection of supports of P_i is empty while $\mathsf{A}(K) = 1$. Recall that the dual SDP of $\mathsf{A}(K)$ is given by

$$\mathsf{A}(K) = \min \operatorname{Tr} T \text{ s.t. } \operatorname{Tr} P_i T \geq 1,\ T \geq 0.$$

Then $\mathsf{A}(K) = 1$ implies that we can find $T_0 \geq 0$ such that $\operatorname{Tr} T_0 = 1$ and $\operatorname{Tr} P_i T_0 \geq 1$ for any i. Clearly T_0 is a density operator, and we should have $\operatorname{Tr} P_i T_0 = 1$ for any i. The only possibility is that T_0 is in the intersection of the supports of P_i, which is a contradiction. Now we turn to show that ii) implies iii). Again by contradiction, assume that $\mathsf{A}(K) > 1$ while the intersection of supports is nonempty. Then we can find a pure state $|\psi\rangle$ from the intersection such that $P_i \geq |\psi\rangle\langle\psi|$. So

$$\mathbb{1} \geq \sum_i s_i P_i \geq \left(\sum_i s_i \right) |\psi\rangle\langle\psi|,$$

thus any feasible solution should have $\sum_i s_i \leq 1$, which indeed implies that $\mathsf{A}(K) = 1$. □

Theorem 25 *Let K be a non-commutative bipartite graph with Choi-Jamiołkowski projection P_{AB}, and let $Q_{AB} = \mathbb{1}_{AB} - P_{AB}$ be the orthogonal complement of P_{AB}. Then the following are equivalent:*

i. $C_{0,\mathrm{NS}}(K) > 0$;

ii. $\mathsf{A}(K) > 1$;

iii. $\operatorname{Tr}_A P_{AB} < d_A \mathbb{1}_B$;

iv. $\operatorname{Tr}_A Q_{AB}$ *is positive definite.*

As a matter of fact, we have

$$C_{0,\mathrm{NS}}(K) \geq \log \frac{d_A}{\|\operatorname{Tr}_A P_{AB}\|_\infty} \quad \text{and} \quad \mathsf{A}(K) \geq \frac{d_A}{\|\operatorname{Tr}_A P_{AB}\|_\infty}.$$

Proof The meaning of i) and ii) are very clear, while iii) and iv) need some explanation.

Essentially, iv) means we can find a CP map from B to A with Choi-Jamiołkowski matrix V_{AB} supporting on some subset of Q_{AB}. Note that in the one-shot SDP formulation of $\Upsilon(K)$, we need V_{AB} to be a CPTP map. However, the trace-preserving condition is not necessary for asymptotic case but only $\operatorname{Tr}_A V_{AB}$ is positive definite, which is the most nontrivial part of this theorem.

iii) is directly equivalent to iv) as we have $P_{AB} + Q_{AB} = \mathbb{1}_{AB}$, hence

$$\operatorname{Tr}_A P_{AB} + \operatorname{Tr}_A Q_{AB} = d_A \mathbb{1}_B.$$

In the following we only focus on i), ii), and iii). The equivalence of ii) and iii) is straightforward, simply noticing that $\mathbb{1}_A / \|P_B\|$ is a feasible solution to the primal SDP for $\mathsf{A}(K)$, where $P_B = \operatorname{Tr}_A P_{AB}$.

The equivalence of i) and iii) is much more difficult and non-trivial. In the following we want to explain a little bit more about this equivalence as there is some tricky points.

First, let's see how to use iii) to derive i). We can apply the standard super-dense coding protocol, and obtain a cq-channel with d_A^2 outputs $\{(U_m \otimes \mathbb{1}_B) J_{AB} (U_m \otimes \mathbb{1}_B)^\dagger\}$, and the projections are given by $\{(U_m \otimes \mathbb{1}_B) P_{AB} (U_m \otimes \mathbb{1}_B)^\dagger\}$, where U_m are generalized Pauli matrices acting on A. So we can compute the semidefinite packing number as

$$\frac{d_A^2}{\sum_{m=1}^{d_A^2} (U_m \otimes \mathbb{1}_B) P_{AB} (U_m \otimes \mathbb{1}_B)^\dagger} = \frac{d_A}{\|P_B\|_\infty}. \tag{55}$$

This is also the zero-error no-signalling assisted classical capacity of this cq-channel. Noticing that when $P_B < d_A \mathbb{1}_B$ strictly holds, the right-hand side of the above equation is strictly larger than 1.

The fact that i) implies iii) can be proven by contradiction together with the one-shot SDP formulation for $\Upsilon(K)$. Assume i) holds but Q_B does not have full rank. Then we can find a non-zero vector $|x\rangle_B$ such that $Q_B|x\rangle = 0$. Or equivalently, $Q_{AB}(\mathbb{1}_A \otimes |x\rangle\langle x|_B) = 0$.

i) means for some $n > 1$ we have $\Upsilon(K^{\otimes n}) > 1$. By the one-shot SDP formulation of $\Upsilon(K^{\otimes n})$, we can find positive U_{AB}, S_A, such that

$$\operatorname{Tr} S > 1, S \otimes \mathbb{1} \geq U \geq 0, \operatorname{Tr}_A U_{AB} = \mathbb{1}_B, \text{ and } \operatorname{Tr} P^{\otimes n}(S \otimes \mathbb{1} - U) = 0.$$

So we can find $V_{AB} = S \otimes \mathbb{1} - U$ with $\operatorname{Tr}_B V_{AB} = (\operatorname{Tr} S - 1)\mathbb{1}_B$, with full rank. On the other hand, we also have V_{AB} supported on $\mathbb{1}_{A^n B^n} - P_{AB}^{\otimes n}$, and the later is the summation of product terms such as $P_{AB} \otimes Q_{AB} \otimes \cdots \otimes P_{AB}$, containing at least one factor Q_{AB} each. So V_{AB} is vanishing on the product vector $|x\rangle^{\otimes n}$, which contradicts the fact that V_B has full rank. □

The simple bound $\frac{d_A}{\|\operatorname{Tr}_A P_{AB}\|_\infty}$ is very interesting, and in some important cases it is tight, such as the cq-channels with symmetric outputs, and the class of Pauli channels. It would be interesting to know whether this kind of "entanglement-assisted coding" could provide a possible way to resolve our puzzle between $C_{0,\mathrm{NS}}(K)$ and $\mathsf{A}(K)$, eventually.

VII. CONCLUSION AND OPEN PROBLEMS

We have shown that there is a meaningful theory of zero-error communication via quantum channels when assisted by quantum no-signalling correlations.

In the terminology of non-commutative graph theory, both the one-shot zero-error classical capacity and simulation cost for non-commutative bipartite graphs assisted by quantum no-signalling correlations have been formulated into feasible SDPs. The asymptotic problems for non-commutative bipartite cq-graphs have also been successfully solved, where the capacity turns out to involve a nontrivial regularization of super-multiplicative SDPs, which nevertheless leads to another SDP, the semidefinite packing number. We found analogously that the zero-error simulation cost of a cq-graph is given by a semidefinite covering number, which in contrast to the classical case is in general larger than the packing number.

The zero-error classical capacity of a classical graph assisted by quantum no-signalling correlations is given precisely by the celebrated Lovász number. For the most general non-commutative bipartite graphs, we are able to provide a necessary and sufficient condition for when these graphs have positive zero-error classical capacity assisted with quantum no-signalling correlations

We know rather little about the asymptotic capacity and simulation cost for non-commutative bipartite graphs that are not classical-quantum, however. A very interesting candidate is the non-commutative bipartite graph $K(r) = \mathrm{span}\{E_0, E_1\}$ associated with the amplitude damping channel $\mathcal{N} = \sum_{i=0}^{1} E_i \cdot E_i^\dagger$, with $E_0 = |0\rangle\langle 0| + \sqrt{1-r}|1\rangle\langle 1|$, $E_1 = \sqrt{r}|0\rangle\langle 1|$ and $0 \leq r \leq 1$. The extremes $r = 0$ and $r = 1$ correspond to the noiseless qubit channel and a constant channel, respectively, and are trivial; we will hence assume $0 < r < 1$. This channel is quite interesting because it gives a separation between the semidefinite packing number and the modified version, and other quantities. It is also interesting because there is only one unique channel $\mathcal{N}_r$ which as $K(r)$ as the Kraus operator space. Actually it is a non-unital extreme point of the convex set of CPTP maps. This channel is also able to communicate classical information without error when assisted with no-signalling correlations. By applying the above super-dense coding bound, we have

$$C_{0,\mathrm{NS}}(K_r) \geq \log \frac{4-2r}{3-r}.$$

By some routine calculation, we can show that the semidefinite packing number is $\mathsf{A}(K_r) = 2 - r$, the revised version $\widetilde{\mathsf{A}}(K_r) = (2-r)^2$, and

$$C_{\min \mathrm{E}}(K_r) = \max_{0 \leq p \leq 1} H_2(p) + H_2(rp) - H_2((1-r)p),$$

where $H_2(x) = -x \log x - (1-x)\log(1-x)$ is the binary entropy. Clearly, $C_{\min \mathrm{E}}(K_{0.5}) = 1$, while $\widetilde{\mathsf{A}}(K_{0.5}) = 2.25 > 2$. This means that the revised semidefinite packing number cannot be equal to the zero-error classical capacity assisted by quantum no-signalling correlations. Whether the original form has the same problem remains unknown.

There are many other interesting open problems, of which we highlight a few here. First, it would be very interesting to explore the mathematical structures of quantum no-signalling correlations in greater detail, and to characterize quantitatively how much non-locality is contained in a quantum no-signalling correlation.

Second, it is of great importance to solve the general asymptotic capacity and cheapest simulation problems in the zero-error setting when assisted with quantum no-signalling correlations. In particular, are there examples of strict sub-multiplicativity of $\Sigma(K)$?

The third problem of great interest is to explore the relationship between the quantum Lovász ϑ function introduced in Ref. [11] for a non-commutative graph S, and the zero-error classical capacity of this non-commutative graph assisted with quantum no-signalling correlations, *i.e.* of Kraus spaces K with $K^\dagger K < S$. Likewise, what is the minimum simulation cost $\Sigma(K)$ or $S_{0,\text{NS}}(K)$ over all such K? This latter is open already for a classical graph G and cq-graphs K with confusability graph G, see the end of section V. In any case, it seems that $\Sigma(G)$ and $2^C_{\min}(G)$ are two new interesting, multiplicative graph parameters, different from both $\vartheta(G)$ and $\alpha^*(G)$.

Fourth: Are no-signalling correlations really necessary to achieve the asymptotic simulation cost $S_{0,\text{NS}}(K)$, or is perhaps entanglement enough? The motivation for this question comes from [12] where it was shown that for classical channels, optimal asymptotic simulation is possible using only shared randomness, no other non-local resources are needed. A nice test case for this question is provided by cq-graphs, for which it is easy to see that $S_{0,\text{NS}}(K) \leq \log |B|$. On the other hand, the entanglement-assisted simulation of cq-channels, also known as *remote state preparation* [39, 40] is far harder to understand. Indeed, for generic channels the best known protocol requires $2 \log |B|$ bits of communication, which is the communication cost of teleportation [39, 41].

Fifth, which no-signalling resources are actually required to achieve the asymptotic zero-error capacity $C_{0,\text{NS}}(K)$? This seems an innocuous question – after all, for each channel the SDP $\Upsilon(K)$ tells us precisely which no-signalling correlation achieves the maximum. But looking at the case of classical channels and bipartite graphs [10], in the light of a classical result of Elias [42], shows that there only a very specific resource is required, which is universal for all channels. Namely, according to [42, Prop. 4], a list size-L list code for a bipartite graph Γ can achieve a rate $R = \left(1 - \frac{1}{L}\right) \log \alpha^*(\Gamma) - O\left(\frac{1}{L}\right)$, which is arbitrarily close to $\log \alpha^*(\Gamma)$ for sufficiently large L. Using such a code, Alice can send a message i out of $M = \lfloor 2^{nR} \rfloor$ over n uses of the channel, and Bob will end up with a list, i.e. a subset $I \subset 2^{[M]}$ of $|I| = L$ possible messages such that $i \in I$. To resolve the remaining ambiguity, Alice and Bob now require a no-signalling resource $S_{\binom{M}{L}} = S(\alpha\beta|iI)$ that we call *subset correlation*, defined for $i \in [M]$, $I \in \binom{[M]}{L}$, $t, u \in [L]$, with

$$S(tu|iI) = \begin{cases} \frac{1}{L} & \text{if } i = i_s \in I = \{i_1 < \ldots < i_L\} \;\&\; u - t = s \mod L, \\ \frac{1}{L} & \text{if } i = i_s \in I = \{i_1 < \ldots < i_L\} \;\&\; u - t \neq s \mod L, \\ \frac{1}{L^2} & \text{if } i \notin I. \end{cases}$$

Alice will input i into the box, Bob the set I, and then Alice will send her output t to Bob; since $i \in I$, Bob knows that he can recover the index s of $i = i_s \in I = \{i_1 < \ldots < i_L\}$ as $u - t \mod L$. Do these or other universal quantum no-signalling correlations allow to achieve the $C_{0,\text{NS}}(K)$ (or at least $\log \mathsf{A}(K)$ for cq-graphs)?

Finally, in [43], the observation was made that the optimal orthogonal representation of a graph G in the sense of Lovász (see Theorem 5) seems to satisfy $\Upsilon(K) = \vartheta(G)$, which was confirmed by direct calculation for several graph families. If this held in general, it would mean that $\vartheta(G)$ can in a certain sense be achieved by a single channel use, rather than requiring a many-copy limit.

ACKNOWLEDGMENTS

It is a pleasure to thank many people for their interests and feedback on the present work, including Salman Beigi, Mick Bremner, Eric Chitambar, Min-Hsiu Hsieh, Aram Harrow, Masahito Hayashi, Richard Jozsa, Debbie Leung, Will Matthews, Jonathan Oppenheim, and Stephanie Wehner. Further thanks are due to Giulio Chiribella, Marco Piani, and John Watrous for delightful discussions on quantum no-signalling correlations.

RD was supported in part by the Australian Research Council (ARC) under Grant DP120103776 (with AW) and by the National Natural Science Foundation of China under grants no. 61179030. He was also supported in part by an ARC Future Fellowship under Grant FT120100449. AW was supported by the European Commission (STREPs "QCS" and "RAQUEL"), the European Research Council (Advanced Grant "IRQUAT") and the Philip Leverhulme Trust. Furthermore, by the Spanish MINECO, projects FIS2008-01236 and FIS2013-40627-P, with the support of FEDER funds, as well as the Generalitat de Catalunya CIRIT, project no. 2014 SGR 966.

---

[1] D. Kretschmann, R. F. Werner, "Tema con variazioni: quantum channel capacity", *New J. Phys.* **6**:26 (2004).

[2] C. H. Bennett, P. W. Shor, J. A. Smolin, A. V. Thapliyal, "Entanglement-assisted capacity of a quantum channel and the reverse Shannon theorem", *IEEE Trans. Inf. Theory* **48**(10):2637-2655 (2002).

[3] C. H. Bennett, I. Devetak, A. W. Harrow, P. W. Shor, A. Winter, "The Quantum Reverse Shannon Theorem and Resource Tradeoffs for Simulating Quantum Channels", *IEEE Trans. Inf. Theory* **60**(3):2926-2959 (2014); arXiv[quant-ph]:0912.5537.

[4] M. Berta, M. Christandl, R. Renner, "The Quantum Reverse Shannon Theorem Based on One-Shot Information Theory", *Commun. Math. Phys.* **306**(3):579-615 (2011).

[5] C. E. Shannon, "The zero-error capacity of a noisy channel", *IRE Trans. Inf. Theory* **2**(3):8-19 (1956).

[6] R. Duan, Y. Shi, "Entanglement between two uses of a noisy multipartite quantum channel enables perfect transmission of classical information", *Phys. Rev. Lett.* **101**:020501 (2008).

[7] R. Duan, "Super-activation of zero-error capacity of noisy quantum channels", arXiv[quant-ph]:0906.2526 (2009).

[8] T. S. Cubitt, J. Chen, A. W. Harrow, "Superactivation of the asymptotic zero-error classical capacity of a quantum channel", *IEEE Trans. Inf. Theory* **57**(12):8114-8126 (2011). arXiv[quant-ph]:0906.2547.

[9] T. S. Cubitt, G. Smith, "An extreme form of super-activation for quantum zero-error capacities", *IEEE Trans. Inf. Theory* **58**(3):1953-1961 (2012).

[10] T. S. Cubitt, D. Leung, W. Matthews, A. Winter, "Improving zero-error classical communication with entanglement", *Phys. Rev. Lett.* **104**:230503 (2010).

[11] R. Duan, S. Severini, A. Winter, "Zero-error communication via quantum channels, non-commutative graphs and a quantum Lovász number", *IEEE Trans. Inf. Theory* **59**(2):1164-1174 (2013). arXiv[quant-ph]:1002.2514.

[12] T. S. Cubitt, D. Leung, W. Matthews, A. Winter, "Zero-error channel capacity and simulation assisted by non-local correlations", *IEEE Trans. Inf. Theory* **57**(8):5509-5523 (2011).

[13] C. Berge, *Graphs and Hypergraphs*, North-Holland (Elsevier), Amsterdam, 1973.

[14] E. R. Scheinerman, D. H. Ullman, *Fractional Graph Theory: A Rational Approach to the Theory of Graphs*, Vol. 46 of Wiley Series in Discrete Mathematics and Optimization, John Wiley & Sons, 1997.

[15] D. Leung, L. Mančinska, W. Matthews, M. Ozols, A. Roy, "Entanglement can increase asymptotic rates of zero-error classical communication over classical channels", *Commun. Math. Phys.* **311**:97-111 (2012).

[16] R. Duan, S. Severini, A. Winter, "On zero-error communication via quantum channels in the presence of noiseless feedback", arXiv[quant-ph]:1502.02987 (2015).

[17] D. Beckman, D. Gottesman, M. A. Nielsen, J. Preskill, "Causal and localizable quantum operations", *Phys. Rev. A* **64**:052309 (2001).

[18] T. Eggeling, D. Schlingemann, R. F. Werner, "Semicausal operations are semilocalizable", *Europhys. Lett.* **57**(6):782-788 (2002).

[19] M. Piani, M. Horodecki, P. Horodecki, R. Horodecki, "Properties of quantum nonsignaling boxes", *Phys. Rev. A* **74**:012305 (2006).

[20] O. Oreshkov, F. Costa, Č. Brukner, "Quantum correlations with no causal order", *Nature Comm.* **3**(10):1092 (2012).

[21] G. Chiribella, "Perfect discrimination of non-signalling channels via quantum superposition of causal structures", *Phys. Rev. A* **86**:040301 (**R**) (2012).

[22] L. Vandenberghe, S. Boyd, "Semidefinite Programming", *SIAM Rev.* **38**(1):49-95 (1996).

[23] D. Leung, W. Matthews, "On the power of PPT-preserving and non-signalling codes", arXiv[quant-ph]:1406.7142 (2014).

[24] R. König, R. Renner, C. Schaffner,"The operational meaning of min- and max-entropy", *IEEE Trans.*

Inf. Theory **55**(9):4337-4347 (2009).
[25] M. Tomamichel, *A Framework for Non-Asymptotic Quantum Information Theory*, PhD thesis, ETH Zürich, 2012, arXiv[quant-ph]:1203.2142 (2012).
[26] A. Harrow, personal communication (December 2011).
[27] L. Lovász, "On the Shannon capacity of a graph", *IEEE Trans. Inf. Theory* **25**(1):1-7 (1979).
[28] G. Chiribella, G. M. D'Ariano, P. Perinotti, "Quantum Circuit Architecture", *Phys. Rev. Lett.* **101**:060401 (2008).
[29] S. Beigi, "Entanglement-assisted zero-error capacity is upper bounded by the Lovász theta function", *Phys. Rev. A* **82**:010303(R) (2010).
[30] W. Fulton, J. Harris, *Representation Theory: A First Course*, Springer Verlag, Berlin Heidelberg New York, 1991.
[31] A. W. Harrow, *Applications of Coherent Classical Communication and the Schur transform to quantum information theory*, PhD thesis, MIT, 2005.
[32] M. Christandl, *The Structure of Bipartite Quantum States — Insights from Group Theory and Cryptography*, PhD thesis, University of Cambridge, 2006. arXiv:quant-ph/0604183.
[33] M. Hayashi, "Exponents of quantum fixed-length pure state source coding", arXiv:quant-ph/0202002 (2002).
[34] M.-D. Choi, "Completely Positive Linear Maps on Complex Matrices", *Lin. Alg. Appl.* **10**(3):285-290 (1975).
[35] R. Duan, Y. Feng, M. Ying, "Perfect distinguishability of quantum operations", *Phys. Rev. Lett.* **103**:210501 (2009).
[36] M. Sion, "On General Minimax Theorems", *Pacific J. Math.* **8**(1):171-176 (1958).
[37] P. Frankl, V. Rödl, "Forbidden Intersections", *Trans. Amer. Math. Soc.* **300**(1):259-286 (1987).
[38] A. Cabello, S. Severini, A. Winter, "(Non-)Contextuality of Physical Theories as an Axiom", arXiv[quant-ph]:1010.2163 (2010).
[39] H.-K. Lo, "Classical-communication cost in distributed quantum-information processing: A generalization of quantum-communication complexity", *Phys. Rev. A* **62**:012313 (2000).
[40] C. H. Bennett, D. P. DiVincenzo, P. W. Shor, J. A. Smolin, B. M. Terhal, W. K. Wootters, "Remote State Preparation", *Phys. Rev. Lett.* **87**:077902 (2001).
[41] C. H. Bennett, G. Brassard, C. Crépeau, R. Jozsa, A. Peres, W. K. Wootters, "Teleporting an Unknown Quantum State via Dual Classical and Einstein-Podolsky-Rosen Channels", *Phys. Rev. Lett.* **70**(13):1895-1899 (1993).
[42] P. Elias, "Zero Error Capacity Under List Decoding", IEEE Trans. Inf. Theory **34**(5):1070-1074 (1987).
[43] C.-Y. Lai, R. Duan, in preparation.