

Zero-error channel capacity and simulation assisted by non-local correlations

Toby S. Cubitt, Debbie Leung, William Matthews and Andreas Winter

Abstract—The theory of zero-error communication is re-examined in the broader setting of using one classical channel to simulate another exactly in the presence of various classes of non-signalling correlations between sender and receiver i.e. shared randomness, shared entanglement and arbitrary non-signalling correlations. When the channel being simulated is noiseless, this is zero-error coding assisted by correlations. When the resource channel is noiseless, it is the reverse problem of simulating a noisy channel exactly by a noiseless one, assisted by correlations. In both cases, separations between the power of the different classes of assisting correlations are exhibited for finite block lengths. The most striking result here is that entanglement can assist in zero-error communication. In the large block length limit, shared randomness is shown to be just as powerful as arbitrary non-signalling correlations for exact simulation, but not for asymptotic zero-error coding. For assistance by arbitrary non-signalling correlations, linear programming formulas for the asymptotic capacity and simulation rates are derived, the former being equal (for channels with non-zero unassisted capacity) to the feedback-assisted zero-error capacity derived by Shannon. Finally, a kind of reversibility between non-signalling-assisted zero-error capacity and exact simulation is observed, mirroring the usual reverse Shannon theorem.

I. INTRODUCTION

Much of classical and quantum information theory is concerned with the use of one resource (a channel,

Toby Cubitt is at the University of Bristol. William Matthews (corresponding author: will@northala.net) and Debbie Leung are with the Institute for Quantum Computing at the University of Waterloo. Andreas Winter is at the University of Bristol and the National University of Singapore. TSC is supported by a Leverhulme early-career fellowship and the EC project “QAP” (contract no. IST-2005-15848). DL was funded by CRC, CFI, ORF, CIFAR, NSERC, and QuantumWorks. WM acknowledges the support of NSERC and QuantumWorks. AW is supported by the EC, the U.K. EPSRC, the Royal Society, and a Philip Leverhulme Prize. The CQT is funded by the Singapore MoE and the NRF as part of the Research Centres of Excellence programme. We are grateful for the hospitality of the Kavli Institute for Theoretical Physics at UCSB, where a large part of this research was performed. This research was supported in part by the NSF under Grant No. PHY05-51164.

an entangled state, etc.) to simulate another. Typically errors are allowed in the simulation protocol if they vanish asymptotically as the number of resources involved grows. One then asks for the asymptotic rates of resource exchange: Shannon’s channel coding theorem [2] tells us the asymptotic rate at which we need to make use of a given discrete memoryless channel to simulate a perfect bit channel. The quantum reverse Shannon theorem [5] shows that a single number associated to quantum channels, the entanglement assisted classical capacity C_E , determines the rate at which it can simulate another when entanglement is a free resource. Since C_E reduces to the Shannon capacity for classical channels, the availability of entanglement does not affect the rate at which one classical channel can simulate another, in the setting where errors which vanish in the large block length limit are tolerated.

Since it is often unrealistic to assume that arbitrarily long block lengths can be used in encoding and decoding, an alternative, idealised, task of *zero-error coding* [18] has been considered since the seminal 1956 paper of Shannon [1] and more recently in quantum information theory [7], [8], [9], [10].

For a suitable definition of decoding error probability p_e , both asymptotic and zero-error coding theory make statements about the region of triples (n, k, p_e) which can be achieved by codes which use n channel uses to transmit k bits (or, equivalently, one of 2^k symbols). The full characterisation of this achievable region is normally far from tractable. Whereas the freedom granted by demanding only that $p_e \rightarrow 0$ as $n \rightarrow \infty$ admits simplification via random coding arguments (for example) in the asymptotic theory, the zero-error theory (which studies the restriction of the region to the plane $p_e = 0$) is attractive because the problem becomes essentially combinatorial. Nevertheless, it is a source of hard mathematical problems: In Shannon’s groundbreaking work on the subject [1] he made a conjecture (on the zero-error capacity of

the pentagon channel) which had to wait over twenty years before it was proven by Lovász [19]. Many related open problems remain [18].

In this paper we consider both zero-error coding and the “reverse” problem of exact simulation of noisy channels when various types of correlations between sender and receiver are freely available. This leads to various relaxations of the combinatorial problems posed by the unassisted theory, some of which have complete and general solutions.

II. OVERVIEW

This section introduces the central concepts and quantities dealt with in the rest of the paper (please note that an index of notations is provided as an appendix). $\mathbf{C}(X \rightarrow Y)$ denotes the set of discrete, memoryless, classical channels (i.e. conditional probability distributions) with inputs in X and outputs in Y , X and Y being finite sets. $\mathbf{C}(A \rightarrow S, B \rightarrow T)$ means the set of bipartite conditional probability distributions, with inputs in the set A and outputs in S for Alice, and inputs in B and outputs in T for Bob. We will frequently consider bipartite distributions that are non-signalling, which we will refer to as *correlations*. A *class* Ω of correlations is a subset of all possible bipartite conditional probability distributions defined by some property such that the set is closed under local operations by either party, in addition to all distributions in Ω being non-signalling. We denote the subset of $\mathbf{C}(A \rightarrow S, B \rightarrow T)$ which is in the class Ω by $\Omega(A \rightarrow S, B \rightarrow T)$.

Here we deal with the following classes of correlations: A bipartite channel is in NC if it can be implemented by local operations alone — there are No Correlations between the two parties at all. Correlations belong to SR if they can be obtained using (classical) Shared Randomness (and local operations); to SE (Shared Entanglement) if they can be obtained from local operations on a shared quantum state; and to NS if the correlation is Non-Signalling in both directions: That is, the marginal distribution of Alice’s output is independent of Bob’s input and vice versa. Each class in this list has a strictly weaker defining property than the last, so we have $\text{NC}(A \rightarrow S, B \rightarrow T) \subset \text{SR}(A \rightarrow S, B \rightarrow T) \subset \text{SE}(A \rightarrow S, B \rightarrow T) \subset \text{NS}(A \rightarrow S, B \rightarrow T)$.

If Alice and Bob are connected by a classical channel $\mathcal{N} \in \mathbf{C}(X \rightarrow Y)$ and have access to any correlation in class Ω (shared randomness, entanglement

etc.) then they can *exactly simulate* $\mathcal{M} \in \mathbf{C}(Q \rightarrow R)$ if there is a *local* protocol whereby Alice takes an input $q \in Q$ and, through local operations and a single use of \mathcal{N} and any use of Ω , Bob produces an output $r \in R$, such that the conditional probability of r given q is exactly $\mathcal{M}(r|q)$. To say that n uses of \mathcal{N} can exactly simulate m uses of \mathcal{M} means that $\mathcal{N}^{\otimes n}$ can exactly simulate $\mathcal{M}^{\otimes m}$.

On pairs consisting of a bipartite correlation $P \in \Omega(Q \rightarrow X, Y \rightarrow R)$ and a classical channel $\mathcal{N} \in \mathbf{C}(X \rightarrow Y)$ we define a bilinear map W which corresponds to ‘wiring’ Alice’s output of P to the input of \mathcal{N} and the output of \mathcal{N} to Bob’s input to P to produce a new classical channel $\mathcal{M} = W[P, \mathcal{N}]$, with

$$\mathcal{M}(r|q) := \sum_{x \in X, y \in Y} P(x, r|q, y) \mathcal{N}(y|x).$$

Because of the time ordering involved, this only makes operational sense if Ω is non-signalling from Bob to Alice, and if it is not then \mathcal{M} may not be a valid conditional distribution. See Figure 1 for a diagram of the operational meaning. Set valued arguments to W are given the natural interpretation as yielding the image sets of classical channels.

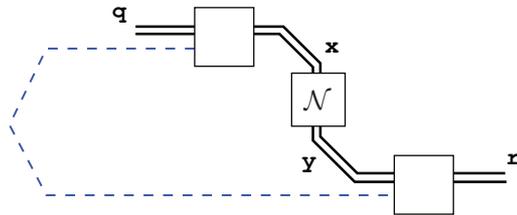


Fig. 1. Schematic representation of $\mathcal{M} = W[P, \mathcal{N}]$: A pre-shared resource (randomness, entanglement or maybe something non-physical) is portrayed by a dotted line. Alice and Bob interact with this, resulting in the non-signalling correlation $P(x, r|q, y) \in \mathbf{C}(Q \rightarrow X, Y \rightarrow R)$: Alice goes first, and obtains x which she inputs into the channel $\mathcal{N} \in \mathbf{C}(X \rightarrow Y)$. Then, based on the channel output y , Bob interacts with the correlation resource, obtaining an outcome r . For example, if the resource is an entangled system, then w.l.o.g. both parties’ interactions consist in choosing from a set of generalised measurements to perform on their local system. This all results in the channel $\mathcal{M} = W[P, \mathcal{N}] \in \mathbf{C}(Q \rightarrow R)$.

Since classes of correlations are closed under local operations, a channel $\mathcal{M} \in \mathbf{C}(Q \rightarrow R)$ can be exactly simulated by a single use of $\mathcal{N} \in \mathbf{C}(X \rightarrow Y)$ and

correlations in Ω if and only if

$$\mathcal{M} \in W[\Omega(Q \rightarrow X, Y \rightarrow R), \mathcal{N}].$$

Now, we can ask for the optimal use of one channel to simulate another one in the presence of some class of correlations Ω . In this paper, we shall concentrate on the simulation of perfect (i.e. identity) channels by noisy ones (“zero-error capacity”) and the reverse (“exact simulation cost”).

Definition 1. For a classical channel $\mathcal{N} \in \mathbf{C}(X \rightarrow Y)$, and free correlations from the class Ω , let $c_0^\Omega(\mathcal{N})$ denote the maximum alphabet size c such that one symbol from the alphabet can be sent without error using Ω and a single use of \mathcal{N} :

$$c_0^\Omega(\mathcal{N}) := \max\{c : \text{id}_c \in W[\Omega([c] \rightarrow X, Y \rightarrow [c]), \mathcal{N}]\},$$

where id_c is the classical identity channel on c symbols.

Since clearly $c_0^\Omega(\mathcal{N}_1 \otimes \mathcal{N}_2) \geq c_0^\Omega(\mathcal{N}_1)c_0^\Omega(\mathcal{N}_2)$, Fekete’s lemma guarantees existence of the Ω -assisted zero-error capacity of a channel \mathcal{N} defined by

$$C_0^\Omega(\mathcal{N}) := \lim_{n \rightarrow \infty} \frac{1}{n} \log c_0^\Omega(\mathcal{N}^{\otimes n}).$$

In this paper “log” is base 2 so this is the capacity in bits. We use “ln” for the natural logarithm.

Definition 2. For a classical channel $\mathcal{N} \in \mathbf{C}(X \rightarrow Y)$, and free correlations in class Ω , let $k_0^\Omega(\mathcal{N})$ denote the minimum alphabet size k such that perfect transmission of one symbol of the alphabet allows exact simulation of one use of the channel:

$$k_0^\Omega(\mathcal{N}) := \min\{k : \mathcal{N} \in W[\Omega(X \rightarrow [k], [k] \rightarrow Y), \text{id}_k]\}.$$

Similarly, we define

$$K_0^\Omega(\mathcal{N}) := \lim_{n \rightarrow \infty} \frac{1}{n} \log k_0^\Omega(\mathcal{N}^{\otimes n})$$

as the asymptotic rate at which perfect classical bits must be transmitted to perfectly simulate \mathcal{N} , if correlations in class Ω are free. The existence of the limit is once more guaranteed by Fekete’s lemma, because k_0^Ω is clearly submultiplicative:

$$k_0^\Omega(\mathcal{N}_1 \otimes \mathcal{N}_2) \leq k_0^\Omega(\mathcal{N}_1)k_0^\Omega(\mathcal{N}_2).$$

A. Structure of the paper

The classical reverse Shannon theorem [4] assures us that in a setting of asymptotically vanishing simulation errors, all channels can reversibly simulate each other when shared randomness between sender and receiver is freely available: the rate at which \mathcal{N}_1 can simulate \mathcal{N}_2 being the ratio of their Shannon capacities, $C(\mathcal{N}_1)/C(\mathcal{N}_2)$. This remains true when entanglement and even more general non-signalling resources are shared by sender and receiver.

The exact coding and simulation problem will be shown to have a much more complex structure. In the next section we review some of the classical theory of zero-error coding, discuss the correlation assisted zero-error quantities c_0^Ω and C_0^Ω , and then show several separations between them for different classes Ω of assisting correlation. The most striking results here are a complete solution for the non-signalling assisted case and the construction of channels where entanglement assists for the one-shot scenario (i.e. where $c_0^{\text{SE}} > c_0$).

In section IV we explore the quantities k_Ω , which we show to be all different (in general) for $\Omega \in \{\text{NC}, \text{SR}, \text{SE}, \text{NS}\}$; and the simulation rates K_0^Ω , which turn out to be all the same for $\Omega \in \{\text{SR}, \text{SE}, \text{NS}\}$, and indeed are given by a simple formula. We even find a kind of combinatorial reverse Shannon theorem for zero-error communication/noisy channel simulation in the presence of general non-signalling correlations: the simulation rate minimised over all channels with the same pattern of zeroes as the matrix $\mathcal{N}(y|x)$ is the same the non-signalling assisted zero-error capacity of \mathcal{N} .

We conclude with some open questions.

III. ASSISTED ZERO-ERROR CAPACITIES

A. Local operations and shared randomness

We start with the least powerful resources, NC and SR. The former simply describes arbitrary encoding and decoding maps. Shared randomness doesn’t change anything since any value of the shared randomness will have to yield a zero-error coding if the randomised protocol does. For the same reason nothing is lost by requiring deterministic encoding and decoding maps, so the coding can be fully specified by giving a subset of input symbols to use as codewords. Thus we are in Shannon’s original zero-error setting [1], and we shall write $c_0 = c_0^{\text{NC}} = c_0^{\text{SR}}$ and $C_0 = C_0^{\text{NC}} = C_0^{\text{SR}}$.

The fundamental observation is that, for zero-error coding over a channel $\mathcal{N} \in \mathbf{C}(X \rightarrow Y)$, two symbols can both be used as codewords only if they are not *confusable*, that is, only if the corresponding output distributions have disjoint support. Therefore, a zero-error code is just a set of pairwise non-confusable input symbols in X , and $c_0(\mathcal{N})$ is the largest size of such a set.

In general, it is not hard to see that for any of our resources Ω , only the pattern of zeroes in $\mathcal{N}(y|x)$ can affect c_0^Ω and C_0^Ω , so that the zero/one matrix $[\mathcal{N}(y|x)]$ encodes all the relevant information. This motivates the introduction of the following combinatorial representations of channels.

Definition 3. *The hypergraph $H(\mathcal{N})$ of a channel $\mathcal{N} \in \mathbf{C}(X \rightarrow Y)$ has vertex set X and hyperedges*

$$E(H(\mathcal{N})) := \{e_y := \{x : \mathcal{N}(y|x) > 0\} : \forall y \in Y\}$$

capturing the equivocation of each output symbol $y \in Y$.

Note that different output symbols can give rise to the same hyperedge, so that the number of hyperedges may be less than the number of output symbols.

Looking back at Definition 1, let $P(\hat{z}|z, y; x)$ denote the probability distribution on Bob's output from the correlation conditional on Alice having input z , Bob having input y , and Alice having obtained output x . This is not well defined if z never occurs for x , and in this case we set $P(\hat{z}|z, y; x) = 0$ (so it is not in fact a distribution). When, Bob obtains an output y he knows that there is non-zero probability that Alice made input x iff it belongs to the hyperedge e_y . The correlation P yields a zero-error coding iff $\sum_{\hat{z}} P(\hat{z}|z, y; x)P(\hat{z}|z', y; x) = 0$ for all $x \in e_y$ whenever $z \neq z'$ for every hyperedge e_y in $E(H(\mathcal{N}))$. Therefore, a correlation assisted zero-error capacities depends only on the channel hypergraph.

To compute the unassisted zero-error capacity an even coarser representation of the channel will suffice:

Definition 4. *The confusability graph $G(\mathcal{N})$ of a channel $\mathcal{N} \in \mathbf{C}(X \rightarrow Y)$ has vertices X and an edge between input symbols x and x' iff they are confusable, i.e. $\sum_{y \in Y} \mathcal{N}(y|x)\mathcal{N}(y|x') > 0$.*

With this notation, $c_0(\mathcal{N})$ is simply $\alpha(G(\mathcal{N}))$: the *independence number* of $G(\mathcal{N})$.

Clearly $G(\mathcal{N})$ can be obtained from $H(\mathcal{N})$ by taking the vertex set of H as the vertex set of G and joining vertices with an edge iff there is a hyperedge of H containing both. On the other hand, given a graph G with vertex set X , there are generally many hypergraphs on X which are mapped to G by this rule. Hypergraphs on a given vertex set form a lattice when ordered by inclusion of their sets of hyperedges. The supremum of the set of hypergraphs with confusability graph G is the clique hypergraph of G , $\chi(G)$, whose hyperedges are all of the cliques in G . From the point of view of zero-error coding, extra hyperedges can only be a bad thing, and this represents the worst case: For all hypergraphs H with a given confusability graph G , $c_0^\Omega(H) \geq c_0^\Omega(\chi(G))$.

For two graphs G_1, G_2 with vertex sets X_1, X_2 their strong product $G_1 \otimes G_2$ is the graph on $X_1 \times X_2$ with an edge $\{(x_1, x_2), (z_1, z_2)\}$ iff $(\{x_1, x_2\} \in E(G_1)) \wedge (\{z_1, z_2\} \in E(G_2))$ or $(x_1 = x_2) \wedge (\{z_1, z_2\} \in E(G_2))$ or $(\{x_1, x_2\} \in E(G_1)) \wedge (z_1 = z_2)$. In terms of confusability graphs, $G(\mathcal{N}_1 \otimes \mathcal{N}_2) = G(\mathcal{N}_1) \otimes G(\mathcal{N}_2)$. For two hypergraphs H_i with vertex sets X_i and edges E_i , $(i = 1, 2)$, we define the product $H_1 \otimes H_2$ on vertex set $X_1 \times X_2$ to have the hyperedges $\{e \times f : \forall e \in E_1, f \in E_2\}$. The hypergraph of a product channel is the product of the individual hypergraphs, and the clique hypergraph of a strong graph product is the product of the individual clique hypergraphs.

The *Shannon capacity* of a graph is the asymptotic behaviour of the independence number of the strong product of n copies

$$\Theta(G) := \lim_{n \rightarrow \infty} \sqrt[n]{\alpha(G^{\otimes n})}.$$

The zero-error capacity of \mathcal{N} is the same quantity but measured in bits per channel use

$$C_0(\mathcal{N}) = \log \Theta(G(\mathcal{N})).$$

The smallest example where the supermultiplicity of $c_0 (= \alpha)$ is strict is the pentagon graph C_5 , for which $c_0(C_5) = 2$ but $c_0(C_5^{\otimes 2}) = 5$. Shannon conjectured that $\Theta(C_5) = \sqrt{5}$, which was only shown to true by Lovász [19].

Determining whether $c_0(\mathcal{N})$ is greater than a given integer k is NP-complete (indeed it is trivially equivalent to k -CLIQUE). Whether Θ is larger than some number is not even known to be decidable.

Shannon [1] found an upper bound on the zero-error capacity by considering *feedback* assistance. In

this scenario, as soon as Bob receives an output y from the channel, Alice gets to know this y with perfect reliability. While this is no advantage if only a single use of the channel is made, it is sometimes useful given multiple uses (an observation Shannon attributes to Elias [1]). Shannon goes on to give a general formula for the asymptotic feedback assisted zero-error capacity $C_{0\text{FB}}$. It is zero whenever $C_0(\mathcal{N})$ is zero (i.e. whenever the confusability graph of the channel is complete) but otherwise is precisely the *fractional packing number* of the channel hypergraph.

Definition 5. A fractional packing of a hypergraph H with vertex set $V(H) = X$ is an assignment of non-negative weights $v(x) \leq 1$ to all vertices x such that

$$\forall e \in E(H) \quad \sum_{x \in e} v(x) \leq 1.$$

A fractional covering of a hypergraph H with vertex set $V(H) = X$ is an assignment of non-negative weights $w(e) \leq 1$ to all hyperedges $e \in E(H)$ such that

$$\forall x \in X \quad \sum_{e \ni x} w(e) \geq 1.$$

(For weights in $\{0, 1\}$ we recover the combinatorial notions of packing and covering.)

The fractional packing number $\alpha^*(H)$ is the maximum total weight allowed in fractional packing of H and the fractional covering number $\omega^*(H)$ is the minimum total weight required for a fractional covering of H . These are clearly dual linear programs, which for a channel hypergraph $H(\mathcal{N})$ have the formulation

$$\begin{aligned} \alpha^*(H(\mathcal{N})) &= \max \left\{ \sum_{x \in X} v(x) : \forall x \in X, v(x) \geq 0, \right. \\ &\quad \left. \sum_{y \in Y} [\mathcal{N}(y|x)] v(x) \leq 1 \right\} \\ \omega^*(H(\mathcal{N})) &= \min \left\{ \sum_{y \in Y} w(y) : \forall y \in Y, w(y) \geq 0, \right. \\ &\quad \left. \sum_{x \in X} [\mathcal{N}(y|x)] w(y) \geq 1 \right\}. \end{aligned}$$

Note that the possibility of redundant hyperedges in this representation (as compared with the purer one in terms of sets) has no effect on either quantity.

The fractional packing problem is always feasible. On the other hand, the fractional covering problem is feasible if and only if the union of all hyperedges

covers X . Where both are feasible $\alpha^*(H) = \omega^*(H)$ by the strong duality theorem for linear programs. In particular, this holds for a channel hypergraph, since the fractional covering problem is always feasible (as every input symbol always results in *some* output symbol occurring).

From the definition of α^* , we have $\alpha(G) \leq \alpha_\chi^*(G) := \alpha^*(\chi(G)) \leq \alpha^*(H)$ for any hypergraph with confusability graph G . But it turns out that α^* is even an upper bound on $\Theta(G)$:

Proposition 6. α^* is multiplicative with respect to the direct hypergraph product: $\alpha^*(H_1 \otimes H_2) = \alpha^*(H_1) \alpha^*(H_2)$.

Proof: To show multiplicativity, the strong duality means that it suffices to show supermultiplicativity of α^* , and submultiplicativity of ω^* , i.e.

$$\begin{aligned} \alpha^*(H_1 \otimes H_2) &\geq \alpha^*(H_1) \alpha^*(H_2) \\ \omega^*(H_1 \otimes H_2) &\leq \omega^*(H_1) \omega^*(H_2). \end{aligned}$$

These are easy, because it is straightforward to confirm that the tensor product of two feasible vectors v_1 and v_2 (dual feasible vectors w_1 and w_2) for H_1 and H_2 , respectively, is feasible (dual feasible) for $H_1 \times H_2$. ■

Therefore, for any integer n , $\alpha(G^{\otimes n}) \leq \alpha_\chi^*(G^{\otimes n}) = \alpha^*(\chi(G)^{\otimes n}) = (\alpha_\chi^*(G))^n$, so $\Theta(G) \leq \alpha_\chi^*(G)$. But this bound is often not tight. For example for the pentagon C_5 , it yields $\Theta(C_5) \leq \frac{5}{2}$; the above two-copy consideration shows on the other hand that $\Theta(C_5) \geq \sqrt{5}$. The celebrated result of Lovasz [19] says that the lower bound is tight, $\Theta(C_5) = \sqrt{5}$. He proved this by introducing another, tighter, but still multiplicative relaxation for $\alpha(G)$, denoted $\vartheta(G)$.

B. Assistance by non-signalling correlations

Now that we have reviewed the state of the art regarding c_0 and C_0 , we go on to present our complete solution for c_0^{NS} and C_0^{NS} .

Theorem 7. For a classical channel $\mathcal{N} \in \mathbf{C}(X \rightarrow B)$ with hypergraph $H(\mathcal{N})$

$$c_0^{\text{NS}}(\mathcal{N}) = \lfloor \alpha^*(H(\mathcal{N})) \rfloor$$

where $\alpha^*(H(\mathcal{N}))$ is the fractional packing number of $H(\mathcal{N})$. Being a linear program, this can be efficiently computed from the channel.

Furthermore, since α^* is multiplicative, the NS-assisted zero-error capacity of a channel is

$$C_0^{\text{NS}}(\mathcal{N}) = \log \alpha^*(H(\mathcal{N})).$$

Proof: $\mathcal{N} \in \mathbf{C}(X \rightarrow Y)$ can exactly simulate a g -message identity channel with non-signalling correlations if and only if there exists P in $\text{NS}([g] \rightarrow X, Y \rightarrow [g])$ such that

$$\sum_{x \in X, y \in Y} P(x, \hat{z}|z, y) \mathcal{N}(y|x) = \begin{cases} 1 & \text{if } \hat{z} = z, \\ 0 & \text{if } \hat{z} \neq z. \end{cases}$$

Without loss of generality, we can assume that a simplified form of non-signalling correlation is used: Suppose some P satisfies the above condition. Then the symmetry of the identity channel under simultaneous permutation of the input and output alphabets means that we can always construct a new P' which is symmetrised by the following ‘twirling’ procedure

$$P'(x, \hat{z}|z, y) = \frac{1}{|S_g|} \sum_{\pi \in S_g} P(x, \pi(\hat{z})|\pi(z), y)$$

where S_g is the symmetric group of order g and $\pi(z)$ is the image of z under the permutation π . This clearly simulates the same channel as P , but it is highly symmetric in that

$$P'(x, \hat{z}|z, y) = \begin{cases} D_{xy} & \text{if } \hat{z} = z, \\ Q_{xy} & \text{if } \hat{z} \neq z. \end{cases}$$

With this simplification in mind, we maximize g such that a valid non-signalling correlation P' allows the simulation of a g -message identity channel. We enumerate the constraints on P' in terms of D and Q .

(1) P' is a valid conditional probability distribution iff

$$\forall x, y : D_{xy} \geq 0, Q_{xy} \geq 0$$

and

$$\forall y : \sum_{x \in X} (D_{xy} + (g-1)Q_{xy}) = 1.$$

(2) The non-signalling condition from Bob to Alice is given by:

$$\forall y : D_{xy} + (g-1)Q_{xy} = u_x$$

for some u_x , whereas the condition that Alice cannot signal to Bob is

$$\forall y : \sum_{x \in X} D_{xy} = \sum_{x \in X} Q_{xy}.$$

(3) The resulting channel is the g -message identity iff

$$\sum_{x \in X, y \in Y} D_{xy} \mathcal{N}(y|x) = 1 \quad \text{and} \\ \sum_{x \in X, y \in Y} Q_{xy} \mathcal{N}(y|x) = 0.$$

Eliminating D using condition (2), the full set of constraints (in terms of Q and u) can be simplified:

$$\forall x, y : Q_{xy} \geq 0, \quad u_x \geq (g-1)Q_{xy}, \\ \forall y : \sum_{x \in X} Q_{xy} = \frac{1}{g}, \quad \sum_{x \in X} u_x = 1 \\ \text{and} \quad \sum_{x \in X, y \in Y} Q_{xy} \mathcal{N}(y|x) = 0.$$

$c_0^{\text{NS}}(\mathcal{N})$ is the largest integer smaller than the largest real number g satisfying these constraints, which we now show is the $\alpha^*(H(\mathcal{N}))$ of the theorem. Defining $T_{xy} := (g-1)Q_{xy}$, the largest feasible value of g is

$$g = \max \left\{ \frac{1}{1-s} : \sum_{x \in X} T_{xy} = s, T_{xy} \geq 0, u_y \geq T_{xy}, \right. \\ \left. \sum_{x \in X} u_x = 1, \sum_{x \in X, y \in Y} T_{xy} \mathcal{N}(y|x) = 0 \right\}.$$

By a simple application of the linear-fractional programming technique [6] this optimisation can be recast as a linear program: Making the substitutions $t := 1/(1-s)$, $T'_{xy} := tT_{xy}$, $v(x) = tu_x$ yields

$$g = \max \left\{ t : v(x) \geq T'_{xy} \geq 0, \right. \\ \left. \sum_{x \in X} v(x) = t, \sum_{x \in X} T'_{xy} = t-1, \right. \\ \left. \sum_{x \in X, y \in Y} T'_{xy} \mathcal{N}(y|x) = 0 \right\}.$$

This is equivalent to the linear program

$$g = \max \left\{ \sum_{x \in X} v(x) : v(x) \geq T'_{xy} \geq 0, \right. \\ \left. \sum_{x \in X} (v(x) - T'_{xy}) \leq 1, \right. \\ \left. \sum_{x \in X, y \in Y} T'_{xy} \mathcal{N}(y|x) = 0 \right\}.$$

The would just be a reorganisation, except that we have also replaced the equality constraints on the second line with inequalities. This doesn't change the value of the linear program: In any optimal solution with the inequalities, the sum over $v(x)$ will be at least one. Therefore, lowering the values of the T'_{xy} , a solution to the LP where the equalities hold can be found which has the same objective value.

Finally, note that the T'_{xy} are redundant in the above formulation. Indeed, we may always set $T'_{xy} = v(x)$ unless we are forced to take $T'_{xy} = 0$ due to $\mathcal{N}(y|x) > 0$ or, equivalently, due to $\lceil \mathcal{N}(y|x) \rceil = 1$. Therefore,

$$g = \max \left\{ \sum_{x \in X} v(x) : v(x) \geq 0 \ \forall x \in X, \right. \\ \left. \sum_x \lceil \mathcal{N}(y|x) \rceil v(x) \leq 1 \ \forall y \in Y \right\},$$

precisely the fractional packing number of $H(\mathcal{N})$. ■

Corollary 8. *If a channel \mathcal{N} with n inputs has at most m non-zero entries $\mathcal{N}(y|x)$ for each y (i.e. the hyperedges of the equivocation graph are all of size $\leq m$). Then,*

$$c_0^{\text{NS}}(\mathcal{N}) \geq \left\lfloor \frac{n}{m} \right\rfloor \text{ and } C_0^{\text{NS}}(\mathcal{N}) \geq \log \frac{n}{m}.$$

The proof is by checking that the assignment $v(x) = \frac{1}{m}$ is feasible. □

We now show that C_0^{NS} can be arbitrarily larger than C_0 . In fact, there are channels for which the latter is 0 while the former is positive!

Let $\binom{[n]}{m}$ denote the set of all size- m subsets of $[n]$. For all $n > m \geq 2$, define the channels $\mathcal{S}_{n,m} \in \mathbf{C}([n] \rightarrow \binom{[n]}{m})$, each maps $x \in [n]$ to a random subset of $[n]$ of cardinality m containing x , i.e.

$$\mathcal{S}_{n,m}(y|x) = \begin{cases} 0 & \text{if } x \notin y, \\ \frac{1}{\binom{n-1}{m-1}} & \text{if } x \in y. \end{cases}$$

For all these channels, $c_0(\mathcal{S}_{n,m}) = C_0(\mathcal{S}_{n,m}) = 0$ because any two inputs x and x' are contained in a common set, hence they are confusable. (In other words, the confusability graph is the complete graph K_n .) On the other hand, by Corollary 8, all of these channels have $C_0^{\text{NS}}(\mathcal{S}_{n,m}) \geq \log \frac{n}{m}$, a strictly positive non-signalling-assisted capacity.

The smallest parameters for which this effect can be seen on the single-shot level are $n = 4$ and $m = 2$: $\mathcal{S}_{4,2}$ is a channel with 4 inputs and 6 outputs, and

Theorem 7 gives $c_0^{\text{NS}}(\mathcal{S}_{4,2}) = 1$. How can this be? Define a non-signalling correlation $P \in \text{NS}(\{0,1\} \rightarrow [4], \binom{[4]}{2} \rightarrow \{0,1\})$ as follows. Alice's input is a bit z , her output x' is a random element of $[4]$. Bob's input is a subset $y' \in \binom{[4]}{2}$. If $x' \in y'$ then Bob's output bit \hat{z} is z and otherwise is $\neg z$. Clearly, Bob's output is independent of Alice's input and vice versa so it is indeed non-signalling.

Suppose Alice wires her output into the channel $\mathcal{S}_{4,2}$ (so $x' = x$) and Bob uses the output of $\mathcal{S}_{4,2}$ as his input to P (so $y' = y$). The behaviour of the channel ensures that y' will always contain x' and therefore Bob's output \hat{z} will always be equal to z . A bit is transmitted from Alice to Bob with perfect reliability — and that despite the fact that any two inputs of the channel cannot be told apart with certainty by Bob!

Whenever $C_0(\mathcal{N}) > 0$, the non-signalling assisted zero-error capacity $C_0^{\text{NS}}(\mathcal{N})$ is precisely the same as the feedback assisted zero-error capacity. This is especially remarkable because the corresponding quantities for a finite number of channel uses are not necessarily the same, and the proofs of the capacity formulas are very different [1]. Also interesting is the fact that non-signalling proves strictly more powerful than feedback. In fact, when the capacities differ, the feedback assisted capacity must be zero.

C. Assistance by entanglement

In [15] we show that, like $c_0(\mathcal{N})$ and unlike $c_0^{\text{NS}}(\mathcal{N})$, the one-shot (and hence also, asymptotic) entanglement assisted zero-error capacity depends only on $G(\mathcal{N})$. An immediate corollary of this is that if $c_0(\mathcal{N}) = 0$ then $c_0^{\text{SE}} = 0$. Proposition 30 of the appendix, shows that these facts hold for assistance by any class of correlations with a certain operational property, which is possessed by SE but not NS. Here, for the reader's convenience we repeat the proof of [15].

Theorem 9. *For any channel \mathcal{N} with inputs X and outputs Y , $c_0^{\text{SE}}(\mathcal{N}) = \max c$ subject to the constraint that there exists a density matrix ρ_B and positive semidefinite operators $\beta_x^{(z)}$ for all $z \in [c]$, $x \in X$, on some Hilbert space such that,*

$$\forall z : \sum_{x \in X} \beta_x^{(z)} = \rho_B$$

$$\forall z \neq z', \{x, x'\} \in E(G(\mathcal{N})) : \text{Tr } \beta_x^{(z)} \beta_{x'}^{(z')} = 0.$$

Consequently, $c_{\text{SE}}(\mathcal{N})$ depends only on $G(\mathcal{N})$.

Proof: We call the shared entangled state ρ_{AB} . Without loss of generality, to send message z , Alice performs a measurement with POVM elements $\{M_x^{(z)} : x \in X\}$, and with probability $p_x^{(z)} (= \text{Tr}[M_x^{(z)}(\text{Tr}_B \rho_{AB})])$, obtains outcome x . Conditional on the knowledge of z and x , the residual state of Bob's system is $\rho_x^{(z)} = (\text{Tr}_A M_x^{(z)} \otimes \mathbb{1}_{\rho_{AB}}) / p_x^{(z)}$. Letting $\beta_x^{(z)} := p_x^{(z)} \rho_x^{(z)}$, for all messages z

$$\sum_x \beta_x^{(z)} = \text{Tr}_A \rho_{AB} =: \rho_B$$

reflecting the fact that without information from the classical channel, Bob has no idea which message Alice sent (i.e. causality is respected). Conversely, any set of positive operators $\beta_x^{(z)}$ which satisfy this condition for some ρ_B can be realised by a suitable choice of ρ_{AB} and generalised measurements.

Alice puts the outcome x into the channel \mathcal{N} . Bob obtains the channel output y , in addition to a quantum state left in his half of the entangled system. This bipartite state on Bob's side is given by:

$$\sigma_z := \sum_{x \in X, y \in Y} \mathcal{N}(y|x) |y\rangle\langle y| \otimes \beta_x^{(z)}.$$

The encoding works if and only if Bob can distinguish perfectly between all the σ_z , i.e. for all distinct $z, z' \in [c]$

$$\begin{aligned} 0 &= \text{Tr} \sigma_z \sigma_{z'} \\ &= \sum_{x, x' \in X, y, y' \in Y} \mathcal{N}(y|x) \mathcal{N}(y'|x') \delta_{yy'} \text{Tr} \beta_x^{(z)} \beta_{x'}^{(z')} \\ &= \sum_{x, x' : \{x, x'\} \in E(G)} \sum_y \mathcal{N}(y|x) \mathcal{N}(y|x') \text{Tr} \beta_x^{(z)} \beta_{x'}^{(z')}. \end{aligned}$$

■

Entanglement can still help, though. As shown in Theorem 13 (and previously in [15]) there are channels with $c_0^{\text{SE}}(\mathcal{N}) > c_0(\mathcal{N}) > 0$.

Whether there are channels \mathcal{N} exhibiting an asymptotic separation $C_0^{\text{SE}}(\mathcal{N}) > C_0(\mathcal{N})$ remains an open question at this time. The efficiently computable formulae for c_0^{NS} and C_0^{NS} derived in the previous section provide upper bounds on entanglement assistance in both the one shot and asymptotic cases, but a tighter bound is known: Duan *et al.* [17] have defined a generalisation of the Lovász theta function [19] for quantum channels. It is multiplicative for tensor products of channels and reduces to the

classical Lovász theta function when the channel is classical. They show that this function is an upper bound on the entanglement assisted one-shot zero-error capacity (for sending classical messages) for any *quantum* channel. Therefore, the classical theta function is an upper bound on c_0^{SE} for classical channels. A short and direct proof of this fact was derived independently by Beigi [16]. The bound is a strict improvement over the fractional packing bound and since ϑ , like α^* , is multiplicative, it too can be immediately applied to the asymptotic rate:

$$C_0^{\text{SE}}(\mathcal{N}) \leq \log \vartheta(G(\mathcal{N})).$$

In terms of trying to decide whether separations exist between C_0^{SE} and C_0 , this is a rather frustrating result because $\vartheta(G(\mathcal{N}))$ is typically also the best bound we have on C_0 ! Exceptions to this have been found, by Haemers [11], but then the problem is to determine whether entanglement assisted protocols exist which beat the best known upper bound on C_0 for those special cases, and even then, only a positive answer would settle the general problem.

Another intriguing corollary of the result is that for the channels with $c_0 < c_0^{\text{SE}}$ made according to our construction from [15], the Lovász theta function coincides exactly with the *lower* bound on c_0^{SE} provided by the explicit protocol we give, so for these channels we know the precise value of $C_0^{\text{SE}}(\mathcal{N})$ and furthermore that it is achieved by repeating the optimal protocol for a single use of the channel.

We now review the proof of the above statement as well as the construction from [15] to which it applies.

Definition 10. Let G be a graph with vertex set X . An orthonormal representation Γ of G in \mathbb{C}^d is an assignment of unit vectors in \mathbb{C}^d to the vertices of G such that if two vertices connected by an edge then their assigned vectors are orthogonal (where orthogonality is with respect to the usual inner product $\langle \cdot, \cdot \rangle$):

$$\forall x, x' \in X : \langle \Gamma(x), \Gamma(x') \rangle = 0 \iff \{x, x'\} \in E(G).$$

Theorem 11. Suppose that G is a graph with an orthonormal representation in \mathbb{C}^d whose vertices can be partitioned into exactly q cliques $\{\mathcal{K}_1, \dots, \mathcal{K}_q\}$ each of size d . Then there is a one-shot zero-error communication protocol assisted by a rank- d maximally entangled state, which shows that $c_0^{\text{SE}}(G) \geq q$. Also, $\vartheta(G) = q$, and since [16], [17] proved that $c_0^{\text{SE}}(G) \leq \vartheta(G)$. Therefore, $c_0^{\text{SE}}(G) = q$.

Proof: First, we describe the entanglement assisted protocol. Alice and Bob share $\frac{1}{\sqrt{d}} \sum_{j=1}^d |j\rangle_A \otimes |j\rangle_B$, with $|j\rangle$ the computational basis vectors for each party. The q cliques of size d which partition the vertices of the graph correspond to q complete orthonormal bases for \mathbb{C}^d given by $\mathcal{B}_z = \{\Gamma(x) : \forall x \in \mathcal{K}_z\}$ for $z = 1$ to q . To encode the message z , Alice measures her half of the shared state along the basis \mathcal{B}_z^c (obtained by conjugating each state in \mathcal{B}_z). If the outcome corresponds to x , Bob's subsystem is left in the state $\Gamma(x)$.¹

Alice inputs x to the channel. Bob's output y from \mathcal{N} tells him a clique e_y in G that contains x , which is not necessarily one of cliques in the partition. So Bob's subsystem must be in one of the corresponding set of *mutually orthogonal* states $\Gamma(e_y)$. Therefore, he can perform a projective measurement on his subsystem to determine exactly which state he has, from which he can deduce x and, *a fortiori*, the symbol $z \in [q]$ which Alice chose, with certainty.

Second, to obtain $\vartheta(G)$, note that it can only increase if edges are removed, and it is multiplicative under strong graph product we have

$$\vartheta(G) \leq \vartheta(\bar{K}_q \otimes K_d) = \vartheta(\bar{K}_q)\vartheta(K_d) = q,$$

where K_n and \bar{K}_n are the complete and empty graphs on n , which have Lovász theta values of 1 and n , respectively.

Using the result from [16], [17] that $c_0^{\text{SE}}(G) \leq \vartheta(G)$, and putting both parts together,

$$c_0^{\text{SE}}(G) = \vartheta(G) = q. \quad \blacksquare$$

Definition 12. We call a set $Z = \{B_m\}_{m=1}^q$ of q complete orthogonal bases $B_m = \{|b_{mj}\rangle : j = 1, \dots, d\}$ for \mathbb{C}^d a KS basis set, if it is impossible to pick one vector from each basis so that no two are orthogonal.

That such sets exist is a simple corollary of the Kochen-Specker theorem [13]. An example of a KS basis set with 6 bases for \mathbb{C}^4 taken from a proof of the Kochen-Specker theorem by Peres [14] is given in [15].

Theorem 13. For any KS basis set $Z = \{B_m\}_{m=1}^q$ in \mathbb{C}^d of q bases, one can construct a classical channel

¹If $\Gamma(x) = \sum_i a_i |i\rangle$, then, the postmeasurement state is $(\sum_i a_i |i\rangle_A \otimes I_B) \sum_{j=1}^d |j\rangle_A \otimes |j\rangle_B = \sum_j a_j |j\rangle_B = \Gamma(x)_B$.

\mathcal{N}_Z (with qd input symbols) with $c_0(\mathcal{N}_Z) < q$ and $c_{\text{SE}}(\mathcal{N}_Z) = q$.

Proof: Construct the graph G_Z on $[q] \times [d]$ with (m, j) connected to (m', j') iff $|b_{mj}\rangle$ and $|b_{m'j'}\rangle$ are orthogonal. Clearly, G_Z partitions into q cliques corresponding to the q bases in Z , so $\alpha(G) \leq q$, and if there was an independent set in G_Z of size q , it would have to have one element in each of the q cliques. But this would correspond to a selection of one vector in each basis in Z such that no two are orthogonal, in contradiction to the fact that Z is a KS basis set.

Letting \mathcal{N}_Z be a channel with confusability graph G_Z , we have just shown that $c_0(\mathcal{N}_Z) < q$. On the other hand, since $\Gamma((m, j)) := |b_{mj}\rangle$ clearly defines an orthonormal representation of G_Z in \mathbb{C}^d , Theorem 11 tells us that $c_0^{\text{SE}}(\mathcal{N}_Z) = q$ (and that this can be achieved using a rank- d maximally entangled state). \blacksquare

IV. EXACTLY SIMULATING NOISY CHANNELS WITH PERFECT COMMUNICATION

This section concerns the “reverse” problem to the zero-error channel coding problem: How much zero-error communication is *required* to exactly simulate a *noisy* channel. It will turn out that the one-shot communication cost can differ wildly between availability of no correlation, shared randomness, entanglement and non-signalling resources. However, in the many-copy limit they all turn out to give the same rate, as long as shared randomness is available. Under a relaxed (namely: combinatorial) notion of channel simulation, we find complete reversibility between non-signalling assisted zero-error channel coding and channel simulation.

A. Without any assistance

What does it mean to simulate a channel $\mathcal{N} \in \mathbf{C}(X \rightarrow Y)$? With a k symbol identity channel and no other correlations between sender and receiver the most general protocol is simply this: Alice applies a local channel $\mathcal{Q} \in \mathbf{C}(X \rightarrow [k])$ and sends to Bob the result, on which he applies a channel $\mathcal{R} \in \mathbf{C}([k] \rightarrow Y)$. The composition should be the desired channel $\mathcal{N} = \mathcal{R} \circ \mathcal{Q}$.

Theorem 14. For a channel $\mathcal{N} \in \mathbf{C}(X \rightarrow Y)$, $k_0(\mathcal{N})$ equals the positive-rank [12] of the transition probability matrix $\mathcal{N}(y|x)$, i.e. the smallest number

k of probability distributions on Y such that their convex hull contains all of the output distributions $\mathcal{N}(\cdot|x)$.

Since positive-rank is lower bounded by linear rank, we get the following lower bounds:

$$k_0(\mathcal{N}) \geq \text{rank } \mathcal{N}, \quad K_0(\mathcal{N}) \geq \log \text{rank } \mathcal{N},$$

the latter because the rank is multiplicative. \square

For instance, the channel $\mathcal{N}_{\text{NOT}} \in \mathbf{C}([n] \rightarrow [n])$ with $\mathcal{N}_{\text{NOT}}(y|x) = 0$ if $y = x$ and $\frac{1}{n-1}$ if $y \neq x$, will have $K_0(\mathcal{N}_{\text{NOT}}) = \log n$, the same as the perfect channel, even though both its Shannon and zero-error capacities are much lower.

B. With shared randomness

The set of channels one can perfectly simulate by sending one of k symbols when arbitrary shared randomness is available is simply the convex hull of the set (just described) that can be achieved without shared randomness:

Theorem 15. For a channel $\mathcal{N} \in \mathbf{C}(X \rightarrow Y)$, $k_0^{\text{SR}}(\mathcal{N})$ is the minimum integer k such that

$$\mathcal{N} \in \text{conv} \left(\bigcup_{Z \subseteq Y, |Z| \leq k} \mathbf{C}(X \rightarrow Z) \right),$$

where we view $\mathbf{C}(X \rightarrow Z)$ naturally as a subset of $\mathbf{C}(X \rightarrow Y)$. In fact, on the right hand side, we may replace the sets $\mathbf{C}(X \rightarrow Z)$ with their corresponding subsets of deterministic channels. As matrices, these are zero/one stochastic matrices, with rank $\leq k$ and a channel \mathcal{N} has $k_0^{\text{SR}}(\mathcal{N}) \leq k$ iff its matrix is a convex combination of these.

Proof: Any protocol to simulate \mathcal{N} exactly using shared randomness and k messages amounts to writing \mathcal{N} as a convex (probability) combination of product channels,

$$\mathcal{N} = \sum_i p_i \mathcal{R}_i \circ \mathcal{Q}_i,$$

with $\mathcal{Q}_i \in \mathbf{C}(X \rightarrow [k])$, $\mathcal{R}_i \in \mathbf{C}([k] \rightarrow Y)$, and $p_i \geq 0$, $\sum_i p_i = 1$. Since the extreme points in the set of channels are the deterministic channels, we may push the randomness involved in forming a stochastic map into the shared randomness. But since \mathcal{Q}_i has only k outputs symbols and \mathcal{R}_i is deterministic, also the composition $\mathcal{R}_i \circ \mathcal{Q}_i$ can have only k possible output symbols forming a subset $Z \subseteq Y$, $|Z| \leq k$. In other

words, \mathcal{N} is a convex combination of deterministic channels in $\mathbf{C}(A \rightarrow Z)$, for k -element subsets $Z \subseteq Y$. \blacksquare

C. Non-signalling correlations

Just as in the case of zero-error communication, making non-signalling correlations freely available gives a very tractable structure to the problem of perfectly simulating noisy channels and the one-shot communication cost $k_0^{\text{NS}}(\mathcal{N})$ has a correspondingly simple form: It is the smallest integer greater than or equal to a certain simple norm of the conditional probability matrix. This norm is multiplicative under tensor products, so the corresponding asymptotic rate $K_0^{\text{NS}}(\mathcal{N})$ is just its logarithm.

Theorem 16. For a channel $\mathcal{N} \in \mathbf{C}(X \rightarrow Y)$,

$$k_0^{\text{NS}}(\mathcal{N}) = \left\lceil \sum_y \max_x \mathcal{N}(y|x) \right\rceil,$$

and since $\sum_y \max_x \mathcal{N}^{\otimes n}(y|x) = (\sum_y \max_x \mathcal{N}(y|x))^n$, (note that this function is a norm on stochastic matrices), the corresponding asymptotic rate is just

$$K_0^{\text{NS}}(\mathcal{N}) = \log \left(\sum_y \max_x \mathcal{N}(y|x) \right).$$

Proof: If \mathcal{N} is in $\mathbf{C}(X \rightarrow Y)$, it can be simulated with a k -input identity channel and non-signalling correlations if and only if there exists P in $\text{NS}(X \rightarrow [k], [k] \rightarrow Y)$ such that

$$\sum_{z=\hat{z}} P(z, y|x, \hat{z}) = \mathcal{N}(y|x). \quad (1)$$

Again the twirling procedure in the proof of Theorem 7 simplifies things, but now the symmetry is in the identity channel used for the simulation. Defining

$$P'(z, y|x, \hat{z}) = \frac{1}{|S_k|} \sum_{\pi \in S_k} P(\pi(z), y|x, \pi(\hat{z}))$$

where S_k is the symmetric group of order k and $\pi(z)$ is the image of z under the permutation π yields a correlation which simulates the same channel when used in place of P (summing over $\pi(z) = \pi(\hat{z})$ is the same as summing over $\hat{z} = z$), but where

$$P'(z, y|x, \hat{z}) = \begin{cases} D_{yx} & \text{if } \hat{z} = z, \\ Q_{yx} & \text{if } \hat{z} \neq z. \end{cases}$$

We now list the conditions on P' (in terms of D and Q):

(1) The correctness of the simulation is given by Eq. (1):

$$D_{yx} = \mathcal{N}(y|x)/k.$$

(2a) The conditions for no signalling from Alice to Bob are

$$\sum_{z \in [k]} P'(z, y|x, \hat{z}) = \sum_{z \in [k]} P'(z, y|x', \hat{z})$$

for all x, x' , which reduce to

$$D_{yx} + (k-1)Q_{yx} = D_{yx'} + (k-1)Q_{yx'},$$

so we write $D_{yx} + (k-1)Q_{yx} = u_y$. Clearly we require that $\sum_y u_y = 1$ and $u_y \geq 0$ (in fact, u_y is just the marginal distribution of the output y which is independent of both inputs, like in the PR box).

(2b) The conditions for no signalling from Bob to Alice,

$$\forall \hat{z}, \hat{z}', x : \sum_y P'(z, y|x, \hat{z}) = \sum_y P'(z, y|x, \hat{z}'),$$

which reduce to $\sum_y D_{yx} = \sum_y Q_{yx} \forall x$, are already ensured by the condition that $D_{yx} = \mathcal{N}(y|x)/k$ and $\sum_y D_{yx} + (k-1)Q_{yx} = 1$ for all x (these mean that $\sum_y D_{yx} = \sum_y Q_{yx} = 1/k$ for all x).

(3) The only other constraint is that the entries of Q are positive.

Putting these constraints together, we see that a suitable P' (and hence P) exists if and only if there is a probability vector u such that the resulting Q matrix has positive entries, i.e.

$$u_y - \mathcal{N}(y|x)/k \geq 0$$

for all y, x . Such a u is possible if and only if $\sum_y \max_x \mathcal{N}(y|x)/k \leq 1$. ■

Remark 17. *It is not hard to verify directly that the bit rate needed to perfectly simulate \mathcal{N} with free non-signalling correlations is greater than the Shannon capacity of \mathcal{N} . If the channel input is the random variable R_x where $\Pr(R_x = x) = p(x)$ and the resulting channel output is the random variable R_y*

then

$$\begin{aligned} I(R_x : R_y) &= \sum_{x,y} \mathcal{N}(y|x)p(x) \log \frac{\mathcal{N}(y|x)}{\sum_z \mathcal{N}(y|z)p(z)} \\ &\leq \log \sum_{x,y} \frac{\mathcal{N}(y|x)^2 p(x)}{\sum_z \mathcal{N}(y|z)p(z)} \\ &\leq \log \sum_{x,y} \frac{\mathcal{N}(y|x)p(x) \max_r \mathcal{N}(y|r)}{\sum_z \mathcal{N}(y|z)p(z)} \\ &= \log \sum_y \max_r \mathcal{N}(y|r). \end{aligned}$$

The Shannon capacity of \mathcal{N} is obtained by maximising the left-hand side over all input distributions p .

D. Arbitrarily large gap between k_0 , $k_0^{\text{SR}}(\mathcal{N})$, and $k_0^{\text{NS}}(\mathcal{N})$

Shared randomness is one type of non-signalling correlation so it is clear that $k_0^{\text{SR}}(\mathcal{N}) \geq k_0^{\text{NS}}(\mathcal{N})$. It turns out that there can be an arbitrarily large gap between these two costs. This is the case for the “universal channels” to be defined below.

Definition 18. *Recall that the set of all size- m subsets of $[n]$ is denoted by $\binom{[n]}{m}$. The universal channel $\mathcal{U}_{n,m}$ is the channel in $\mathcal{C}(\binom{[n]}{m} \rightarrow [n])$ with,*

$$\mathcal{U}_{n,m}(y|x) = \begin{cases} \frac{1}{m} & \text{if } y \in x, \\ 0 & \text{if } y \notin x. \end{cases}$$

In words, the channel takes as input a set $x \in \binom{[n]}{m}$ and outputs a random element of that set.

Note that \mathcal{N}_{NOT} introduced earlier in this section is a special case of the universal channel with $m = n-1$.

The universal channels have a great deal of symmetry. The symmetric group S_n acts on both the input and the output alphabet of $\mathcal{U}_{n,m}$: on the latter naturally as permutations of the symbols (written $\pi(y)$), on the former as simultaneous permutations of all elements in the sets $x \subseteq [n]$ (written x^π). With these actions, $\mathcal{U}_{n,m}$ is S_n -covariant:

$$\forall y, x \quad \mathcal{U}_{n,m}(y|x) = \mathcal{U}_{n,m}(\pi(y)|x^\pi). \quad (2)$$

A beautiful consequence of the covariance is that it specifies $\mathcal{U}_{n,m}$ almost uniquely: $\mathcal{U}_{n,m}$ is the only channel satisfying eq. (2) and in addition $\mathcal{U}_{n,m}(y|x) = 0$ if $y \notin x$.

To simulate $\mathcal{U}_{n,m}$ with zero error when assisted by arbitrary non-signalling correlations, Theorem 16 shows one needs a noiseless channel of

$$k_0^{\text{NS}}(\mathcal{U}_{n,m}) = \left\lceil \sum_y \max_x N(y|x) \right\rceil = \left\lceil \frac{n}{m} \right\rceil$$

many symbols, and this is sufficient. The minimal asymptotic rate of communication needed given free non-signalling correlations is $\log \frac{n}{m}$. On the other hand, when only shared randomness is available, the communication cost can be much higher:

Proposition 19. *For any $n \geq m \geq 1$,*

$$k_0^{\text{SR}}(\mathcal{U}_{n,m}) = n - m + 1.$$

Proof: We first show that $k = n - m$ is not sufficient by contradiction. Recall Theorem 15 and consider an element \mathcal{N} from $\mathbf{C}(X \rightarrow Z)$, with $Z \subseteq Y$, $|Z| = n - m$ in the convex decomposition of $\mathcal{U}_{n,m}$ that comes with a strictly positive weight p . That means, for any input x ,

$$p\mathcal{N}(\cdot|x) \leq \mathcal{U}_{n,m}(\cdot|x)$$

in the sense of element-wise ordering of the probability vectors. Choosing $x = [n] \setminus Z$ — which has cardinality m — leads to the desired contradiction: restricted to Z , $\mathcal{U}_{n,m}(\cdot|x)$ is the zero vector (see the definition), whereas $\mathcal{N}(\cdot|x)$ has all of its probability mass in Z .

On the other hand, there is a protocol that uses only $n - m + 1$ messages: The shared randomness is a uniformly distributed subset $T \in \binom{[n]}{n-m+1}$. On input x , Alice selects a uniformly random element $y \in x \cap T$, which is non-empty by the pigeonhole principle. To send y to Bob, she needs only a number from 1 through $n - m + 1$ to specify where y occurs in T . Clearly, this protocol, and hence the simulated channel, has the same S_n -covariance as $\mathcal{U}_{n,m}$. Furthermore, the simulated channel assigns zero conditional probability to all $y \notin a$ for input x . Thus, the simulated channel must be $\mathcal{U}_{n,m}$. ■

These universal channels provide simple and highly structured examples for separating k_0 , k_0^{SR} , and k_0^{NS} . We have already seen that for $m = n - 1$, $k_0(\mathcal{U}_{n,m}) = n$ but Prop. 19 says that $k_0^{\text{SR}}(\mathcal{U}_{n,m}) = 2$. In a different regime, for example when n is even and $m = \frac{n}{2}$, $k_0^{\text{SR}}(\mathcal{U}_{n,m}) = (\frac{n}{2} + 1)$ but $k_0^{\text{NS}}(\mathcal{U}_{n,m}) = 2$. In both cases, the separation is of order n which is maximal given the size of the input alphabet.

E. Shared entanglement

The possibility of large separations between $k_0^{\text{NS}}(\mathcal{N})$ and $k_0^{\text{SR}}(\mathcal{N})$, raises the question of where the power of the intermediate shared entanglement class fits in between the two. While we do not have a general understanding of this matter yet, we can at least give examples where entanglement beats shared randomness and cases where general non-signalling correlations can beat entanglement.

Let \mathcal{T}_p denote the ternary erasure channel with transmission probability p :

$$\mathcal{T}_p = \begin{pmatrix} p & 0 & 0 \\ 0 & p & 0 \\ 0 & 0 & p \\ 1-p & 1-p & 1-p \end{pmatrix}.$$

We can use the ideas of the appendix to show that non-signalling correlations can beat shared entanglement for one-shot simulation of $\mathcal{T}_{1/2}$.

Proposition 20. *Whereas $k_0^{\text{NS}}(\mathcal{T}_{1/2}) = 2$, $k_0^{\text{SE}}(\mathcal{T}_{1/2}) = 3$. Therefore, using a single (perfect) bit of communication, strictly more channels can be simulated if generalised non-signalling correlations are available rather than entanglement.*

Proof: By using the twirling procedure of Theorem 16 (which can be assumed w.l.o.g. whenever shared randomness is available), and considering the simplified non-signalling constraints which result, it is not hard to see that exact simulation of $\mathcal{T}_{1/2}$ with a single bit is equivalent to the ability to realise a particular non-signalling correlation P^* defined, in the notation of Theorem 16, by $k = 2$,

$$D_{yx} = \frac{1}{4} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 1 & 1 \end{pmatrix}$$

and $Q_{yx} = 1/4 - D_{yx}$. This P^* must be in the class of available correlations. Applying the observations of the appendix, and looking at the 8 conditional channels P_{xz}^* , one finds that they are all pair-wise distinguishable and therefore P^* cannot be in the class SE, by Proposition 31. ■

Proposition 21. *Strictly more classical channels can be simulated using shared entanglement than can with shared randomness. In particular, there are channels \mathcal{N} for which $k_0^{\text{SE}}(\mathcal{N}) = 2$ but $k_0^{\text{SR}}(\mathcal{N}) \geq 3$.*

Proof: We construct these channels and demonstrate the separation in three steps. First, we show that $\mathcal{T}_{1/2}$ can be simulated using one bit of communication and a certain non-signalling correlation called the ‘‘PR-box’’. Then, using the same protocol, but replacing the PR-box by a weaker correlation obtainable via shared entanglement, we write down the channel \mathcal{N} that is being simulated. Finally, we explain why \mathcal{N} cannot be simulated with shared randomness and one bit of communication.

The PR-Box (introduced by Popescu and Rohrlich [20]) is a particular correlation $P(s, t|a, b)$ given by:

$$\begin{array}{cccc} & (0, 0) & (0, 1) & (1, 0) & (1, 1) \\ \begin{array}{l} (0, 0) \\ (0, 1) \\ (1, 0) \\ (1, 1) \end{array} & \begin{pmatrix} 1/2 & 1/2 & 1/2 & 0 \\ 0 & 0 & 0 & 1/2 \\ 0 & 0 & 0 & 1/2 \\ 1/2 & 1/2 & 1/2 & 0 \end{pmatrix} \end{array}$$

In other words, the outputs s, t are random bits except for the constraint $s \oplus t = a \cdot b$. We let z to be the message sent via the classical channel.

We first give an explicit method for simulating $\mathcal{T}_{1/2}$ using a bit of communication and a single use of a PR box. Bob always chooses his PR-box input b to be the channel output z and outputs (z, t) for the simulation. Alice has a message $x \in \{0, 1, 2\}$. If $x = 0$, she sets her PR-box input to be $a = 0$. The PR-box outputs a random bit s . She sends $z = s$ to Bob, who puts it in the PR-box and obtains $t = s$. Thus, Bob outputs $(0, 0)$ or $(1, 1)$ randomly. If $x = 1$, Alice sets $a = 1$ and her PR-box outputs a random bit s . She chooses $z = s$. For either value of s , $t = 0$. Thus, Bob outputs $(0, 0)$ or $(1, 0)$ randomly. If $x = 2$, Alice sets $a = 0$ and $z = 0$. Thus, Bob outputs $(0, 0)$ or $(0, 1)$ randomly. Identifying the outputs $(0, 0)$, $(1, 1)$, $(1, 0)$ and $(0, 1)$ with the erasure symbol E , 0 , 1 , and 2 respectively, the above simulates $\mathcal{T}_{1/2}$ perfectly.

In the second step, we generalize the PR-Box to correlations P_λ given by:

$$\begin{array}{cccc} & (0, 0) & (0, 1) & (1, 0) & (1, 1) \\ \begin{array}{l} (0, 0) \\ (0, 1) \\ (1, 0) \\ (1, 1) \end{array} & \begin{pmatrix} \lambda/2 & \lambda/2 & \lambda/2 & \bar{\lambda}/2 \\ \bar{\lambda}/2 & \bar{\lambda}/2 & \bar{\lambda}/2 & \lambda/2 \\ \bar{\lambda}/2 & \bar{\lambda}/2 & \bar{\lambda}/2 & \lambda/2 \\ \lambda/2 & \lambda/2 & \lambda/2 & \bar{\lambda}/2 \end{pmatrix} \end{array}$$

where $\bar{\lambda} = 1 - \lambda$. Note that P_1 is the PR-Box. The PR box can be approximated using a maximally entangled pair of qubits. An optimal approximation, in terms of the CHSH violation, yields the correlation P_λ with $\lambda = (1 + 1/\sqrt{2})/2 \approx 0.85$. If this

entanglement based approximation of the PR box is substituted into the protocol given above, the resulting channel \mathcal{N} is given by

$$\begin{array}{ccc} & 0 & 1 & 2 \\ \begin{array}{l} \text{E} \leftrightarrow (0, 0) \\ 0 \leftrightarrow (1, 1) \\ 1 \leftrightarrow (1, 0) \\ 2 \leftrightarrow (0, 1) \end{array} & \begin{pmatrix} \alpha & \alpha & 1/2 \\ \alpha & \beta & 0 \\ \beta & \alpha & 0 \\ \beta & \beta & 1/2 \end{pmatrix}, \end{array}$$

where $\alpha = (1 + 1/\sqrt{2})/4 \approx 0.43$ and $\beta = 1/2 - \alpha \approx 0.07$.

Finally, we check with a computer that \mathcal{N} is not a convex combination of rank-two, zero-one, stochastic matrices and so, according to Theorem 15, can't be exactly simulated using one bit of communication if only shared randomness is available. ■

F. Asymptotic equality of correlation assisted communication costs

Among the results of this section so far are channels proving separations between the communication costs of NS-, SE- and SR-assisted channel simulation for a single channel use. In the case of NS vs. SR, the universal channels of Definition 18 show that this gap can be arbitrarily large. Despite this, we will prove in Theorem 24 that when simulating many uses of a channel, a protocol using shared randomness can achieve an *asymptotic* rate of communication as low as the optimal rate with non-signalling assistance derived in Theorem 16. Since the rate with entanglement assistance is sandwiched between these two rates, it follows that $K_0^{\text{SR}}(\mathcal{N}) = K_0^{\text{SE}}(\mathcal{N}) = K_0^{\text{NS}}(\mathcal{N}) = \log \sum_y \max_x \mathcal{N}(y|x)$, for all channels \mathcal{N} .

The proof is structured into two steps. First, we show that the asymptotic equality discussed above holds for all the universal channels. Then, roughly speaking, any channel can be exactly simulated by a universal channel with the same value of K_0^{NS} . It is this ability of the channels $\mathcal{U}_{n,m}$ that earns the name ‘‘universal.’’ We need a lemma for this proof:

Lemma 22. *We call a set $T \subseteq [n]^q$ m -touching if*

$$\forall x_1, \dots, x_q \in \binom{[n]}{m} \quad T \cap (x_1 \times x_2 \times \dots \times x_q) \neq \emptyset.$$

There is an m -touching set of cardinality $\min\{n^q, 2nq \binom{n}{m}^q\}$.

Proof: If a set is populated by picking $r = 2nq\left(\frac{n}{m}\right)^q$ elements of $[n]^q$ picked uniformly at random (with replacement), the probability that it is not m -touching is bounded above by

$$P_{\text{fail}} \leq \left(\frac{n}{m}\right)^q \left(1 - \left(\frac{m}{n}\right)^q\right)^r.$$

With the simple estimates $\left(\frac{n}{m}\right) \leq 2^n$ and $\ln(1-x) \leq -x$,

$$\begin{aligned} \ln P_{\text{fail}} &\leq qn \ln 2 - r \left(\frac{m}{n}\right)^q \\ &= (\ln 2 - 2)qn < 0, \end{aligned}$$

so a set with the desired property and cardinality must exist. Indeed, the probability that a set chosen in the way described above isn't m -touching is exponentially small in qn . ■

Proposition 23. *For any universal channel $\mathcal{U}_{n,m}$,*

$$K_0^{\text{SR}}(\mathcal{U}_{n,m}) = K_0^{\text{NS}}(\mathcal{U}_{n,m}) = \log \frac{n}{m}.$$

Proof: By definition, $K_0^{\text{SR}}(\mathcal{U}_{n,m}) \geq K_0^{\text{NS}}(\mathcal{U}_{n,m})$, so it suffices to exhibit a protocol using only shared randomness that achieves this bound. To be precise, for q copies of the channel, we prove the existence of such a protocol which uses the transmission of one of

$$k = \min \left\{ n^q, \left\lceil 2qn \left(\frac{n}{m}\right)^q \right\rceil \right\}$$

symbols. The rate is $\leq \log \frac{n}{m} + \frac{1}{q} \log 2qn$, which approaches $\log \frac{n}{m}$ as $q \rightarrow \infty$.

The protocol works as follows: Alice and Bob agree on an m -touching set T of size k (see Lemma 22). They share randomness in the form of q uniformly random permutations $\pi_1, \dots, \pi_q \in S_n$. On input (x_1, \dots, x_q) Alice picks a uniformly random element $(y_1, \dots, y_q) \in T^{\pi_1, \dots, \pi_q} \cap (x_1 \times x_2 \times \dots \times x_q)$, where

$$\begin{aligned} &T^{\pi_1, \dots, \pi_q} \\ &= \{(\pi_1(z_1), \dots, \pi_q(z_q)) : \forall (z_1, \dots, z_q) \in T\} \end{aligned}$$

is the set T with its elements permuted according to π_j in coordinate $j = 1, \dots, q$. The intersection is guaranteed to exist because T^{π_1, \dots, π_q} is also m -touching. To send y , she only needs a number from 1 through k to specify the location within T^{π_1, \dots, π_q} since the latter is known to Bob.

This protocol evidently simulates an $S_n^{\times q}$ -covariant channel with the property

that $\mathcal{N}(y_1 \dots y_q | x_1 \dots x_q) = 0$ whenever $y_1 \dots y_q \notin x_1 \times x_2 \times \dots \times x_q$. As discussed before, this means that the simulated channel must be $\mathcal{N}^{\otimes q}$. ■

Theorem 24. *For any channel $\mathcal{N} \in \mathbf{C}(X \rightarrow Y)$,*

$$K_0^{\text{SR}}(\mathcal{N}) = K_0^{\text{NS}}(\mathcal{N}) = \log \sum_y \max_x \mathcal{N}(y|x).$$

Proof: First, suppose all the entries of $\mathcal{N}(y|a)$ are rational numbers, with common denominator M , so that $\mathcal{N}(y|x) = \frac{1}{M}t(y|x)$ for integers $t(y|x)$. Split up each output symbol y into $t_y := \max_x t(y|x)$ many, denoted (y, j) , with $j = 1, \dots, t_y$. Now define a new channel by letting $\tilde{\mathcal{N}}((y, j)|x)$ be either 0 or $1/M$, in such a way that $\mathcal{N} = \Pi \circ \tilde{\mathcal{N}}$ with the projection map/channel $\Pi : (y, j) \mapsto y$. Clearly $\tilde{\mathcal{N}}$ is a sub-channel (i.e. a restriction on the input alphabet) of the universal channel $\mathcal{U}_{N,M}$ with $N = \sum_y t_y$. It can therefore be exactly simulated using shared randomness by the protocol of Proposition 23. This requires asymptotic communication rate

$$\log \left(\frac{1}{M} \sum_y t_y \right) = \log \left(\sum_y \max_x \mathcal{N}(y|x) \right),$$

which is precisely the lower bound set by $K_0^{\text{NS}}(\mathcal{N})$, so the claim holds for rational \mathcal{N} .

For the general case, pick a large integer M and let, for all $x \in X$, $y \in Y$, $t(x|y) := \lfloor M\mathcal{N}(y|x) \rfloor$. Now adjoin new elements $x' (x \in X)$ to the output alphabet, i.e. define $\tilde{Y} := Y \cup X'$ and a new channel $\tilde{\mathcal{N}} : X \rightarrow \tilde{Y}$ with

$$\begin{aligned} \tilde{\mathcal{N}}(y|x) &:= \frac{1}{M}t(y|x), \\ \tilde{\mathcal{N}}(x'|x) &:= 1 - \sum_y \tilde{\mathcal{N}}(y|x). \end{aligned}$$

Now, \mathcal{N} can be simulated by $\tilde{\mathcal{N}}$ using post-processing by Bob only: if $y \in Y$ is obtained, then it is left alone; if $x' \in X'$ is seen, then Bob uses local randomness to output y with probability

$$Q(y|x') = \frac{1}{\tilde{\mathcal{N}}(x'|x)} (\mathcal{N}(y|x) - \tilde{\mathcal{N}}(y|x)).$$

So, extending Q to a proper channel by letting $Q(y|y') = 1$ for $y' \in Y$ iff $y = y'$, we have $\mathcal{N} = Q \circ \tilde{\mathcal{N}}$.

Second, the cost of simulating $\tilde{\mathcal{N}}$ is

$$\begin{aligned} & \log \left(\sum_{x \in \tilde{Y}} \max_x \tilde{\mathcal{N}}(x|x) \right) \\ &= \log \left(\sum_{y \in Y} \max_x \tilde{\mathcal{N}}(x|y) + \sum_{x \in X} \tilde{\mathcal{N}}(x'|x) \right) \\ &\leq \log \left(\sum_{y \in Y} \max_x \mathcal{N}(y|x) + \frac{1}{M} |X||Y| \right). \end{aligned}$$

Letting $M \rightarrow \infty$, this rate approaches $\log \left(\sum_y \max_x \mathcal{N}(y|x) \right) = K_0^{\text{NS}}(\mathcal{N})$. ■

To illustrate the idea, if $\mathcal{N}_0 \prec \mathcal{N}_1$ denotes the partial order on channels “ \mathcal{N}_0 is equal to $\mathcal{R} \circ \mathcal{N}_1 \circ \mathcal{Q}$, for some channels \mathcal{R}, \mathcal{Q} ”, the proof uses

$$\mathcal{N} = \frac{1}{5} \begin{pmatrix} 4 & 2 \\ 1 & 3 \end{pmatrix} \prec \frac{1}{5} \begin{pmatrix} 1 & 1 \\ 1 & 1 \\ 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \\ 1 & 1 \end{pmatrix} \prec \mathcal{N}_{7,5}$$

to show that $K_0^{\text{SR}}(\mathcal{N}) \leq K_0^{\text{SR}}(\mathcal{N}_{7,5}) = \log 7/5 = K_0^{\text{NS}}(\mathcal{N}) \leq K_0^{\text{SR}}(\mathcal{N})$.

Remark 25. *The classical reverse Shannon theorem [4] also yields an exact simulation of the noisy channel using shared randomness. The difference to our result is explained by the different way to account for the communication: whereas [4] shows that the expected rate of communication (with respect to the shared randomness used in the protocol) is the normal Shannon capacity of the channel, here we consider the much more stringent worst case communication cost.*

G. Weak simulation and reversibility

Looking over the formulas for C_0^{NS} and K_0^{NS} of a channel \mathcal{N} , we notice that the former only depends on the channel hypergraph, while the latter actually involves the transition probabilities. Hence it is not surprising that the former is typically strictly smaller than the latter. However if we are content with the simulation of any channel that has the same hypergraph, we recover reversibility:

Proposition 26. *Let $\mathcal{N} \in \mathbf{C}(X \rightarrow Y)$ with channel hypergraph $H(\mathcal{N})$ (having hyperedges $\{e_y : y \in Y\}$). Then,*

$$\inf \{ K_0^{\text{NS}}(\mathcal{M}) : H(\mathcal{M}) = H(\mathcal{N}) \} = \log \omega^*(H(\mathcal{N}))$$

where $\omega^*(H(\mathcal{N}))$ is the fractional covering number of the hypergraph of the channel. Since the fractional covering number is equal to the fractional packing number $\alpha^*(H)$, this minimum rate is also equal to $C_0^{\text{NS}}(\mathcal{N})$.

Proof: Recall the formula for $K_0^{\text{NS}}(\mathcal{N})$: it is the logarithm of the value of the following linear program (all variables understood as non-negative):

$$\min \left\{ \sum_{y \in Y} w(y) : \forall x, y \ w(y) \geq \mathcal{N}(y|x) \right\}.$$

The additional minimisation over channels with prescribed hypergraph H is also a linear program:

$$\begin{aligned} & \min \left\{ \sum_{y \in Y} w(y) : w(y) \geq \mathcal{N}(y|x), \right. \\ & \quad \mathcal{N}(y|x) = 0 \text{ if } x \notin e_y, \\ & \quad \left. \sum_y \mathcal{N}(y|x) = 1 \right\}. \end{aligned}$$

But this is evidently equivalent to

$$\min \left\{ \sum_{y \in Y} w(y) : \forall x \in X, \sum_{y \text{ with } e_y \ni x} w(y) \geq 1 \right\},$$

which is exactly the fractional covering number $\omega^*(H)$. For the other statements see Proposition 6. ■

V. CONCLUSION

Let us summarise the results (both our own, and others) discussed in this paper. For zero-error communication, we found both the one-shot and asymptotic non-signalling assisted capacities, upper bounding the chains of operationally obvious inequalities:

$$\begin{aligned} \alpha(G(\mathcal{N})) &= c_0(\mathcal{N}) = c_0^{\text{SR}}(\mathcal{N}) \\ &\leq c_0^{\text{SE}}(\mathcal{N}) \leq c_0^{\text{NS}}(\mathcal{N}) = \lfloor \alpha^*(H(\mathcal{N})) \rfloor \end{aligned}$$

$$\begin{aligned} \log(\Theta(\mathcal{N})) &= C_0(\mathcal{N}) = C_0^{\text{SR}}(\mathcal{N}) \leq C_0^{\text{SE}}(\mathcal{N}) \\ &\leq C_0^{\text{NS}}(\mathcal{N}) = \log \alpha^*(H(\mathcal{N})) \end{aligned}$$

These upper bounds on the entanglement assisted capacities from non-signalling are improved upon by the results of [16] and [17] which show that the Lovász theta bound applies even in the entanglement assisted case:

$$c_0^{\text{SE}}(\mathcal{N}) \leq \lfloor \vartheta(\mathcal{N}) \rfloor, C_0^{\text{SE}}(\mathcal{N}) \leq \log \vartheta(\mathcal{N}).$$

While we proved that $c_0(\mathcal{N}) \leq c_0^{\text{SE}}(\mathcal{N})$ can be strict, we don't yet know whether the same can be said of the asymptotic rates, and regard this as one of the main open problems.

In the reverse problem of exactly simulating noisy channels, the non-signalling assisted case was again completely soluble, providing *lower* bounds on the chain

$$\left[\sum_y \max_x \mathcal{N}(y|x) \right] = k_0^{\text{NS}}(\mathcal{N}) \leq k_0^{\text{SE}}(\mathcal{N}) \\ \leq k_0^{\text{SR}}(\mathcal{N}) \leq k_0^{\text{NC}}(\mathcal{N}).$$

For each inequality in this chain of one-shot costs, a channel showing that it can be strict was exhibited. Some open questions remain regarding the potential sizes of these separations (see section IV).

For the asymptotic rates of communication things were shown to be simpler: While large gaps can exist between the costs with free shared randomness and without ($K_0^{\text{NC}}(\mathcal{N}) \geq \log \text{rank } \mathcal{N}$), given free correlations from *any* class of non-signalling correlations which contains shared randomness the rates are equal:

$$\log \left(\sum_y \max_x \mathcal{N}(y|x) \right) = K_0^{\text{NS}}(\mathcal{N}) = K_0^{\text{SE}}(\mathcal{N}) \\ = K_0^{\text{SR}}(\mathcal{N}) \leq K_0^{\text{NC}}(\mathcal{N}).$$

ACKNOWLEDGMENTS

We would like to thank Nicolas Brunner, Runyao Duan, Tsuyoshi Ito, Ashley Montanaro, Marcin Pawłowski, Paul Skrzypczyk and Stephanie Wehner for useful discussions.

REFERENCES

- [1] C. E. Shannon, "The zero-error capacity of a noisy channel", IRE Trans. Inform. Theory, vol. IT-2(3):8-19 (1956).
- [2] C. E. Shannon, "A mathematical theory of communication", Bell Syst. Tech. J. **27**:379-423, 623-656 (1948).
- [3] J. Barrett, N. Linden, S. Massar, S. Pironio, S. Popescu, D. Roberts, "Nonlocal correlations as an information-theoretic resource", Phys. Rev. A **71**(2):022101 (2005).

- [4] C. H. Bennett, P. Shor, J. Smolin, A. V. Thapliyal, "Entanglement-assisted capacity of a quantum channel and the reverse Shannon theorem", IEEE Trans. Info. Theory **48**, 2637-2655, (2002) arXiv:quant-ph/0106052.
- [5] C. H. Bennett, I. Devetak, A. W. Harrow, P. W. Shor, A. Winter, "Quantum Reverse Shannon Theorem", arXiv:0912.5537 (2009).
- [6] S. Boyd, L. Vandenberghe, *Convex Optimization*, Cambridge University Press (Cambridge, U.K., 2004).
- [7] R. A. C. Medeiros, R. Alleaume, G. Cohen, F. M. de Assis, "Quantum states characterization for the zero-error capacity", arXiv:quant-ph/0611042 (2006).
- [8] R. Duan, "Super-Activation of Zero-Error Capacity of Noisy Quantum Channels", arXiv:0906.2527 (2009).
- [9] T. S. Cubitt, J. Chen, A. W. Harrow, "Superactivation of the Asymptotic Zero-Error Classical Capacity of a Quantum Channel", arXiv:0906.2547 (2009).
- [10] T. S. Cubitt, G. B. Smith, "Super-Duper-Activation of Quantum Zero-Error Capacities", arXiv:0912.2737 (2009).
- [11] W. Haemers, "On Some Problems of Lovász Concerning the Shannon Capacity of a Graph", IEEE Trans. Inf. Theory **25**(2):231-232 (1979). W. Haemers, "An upper bound for the Shannon capacity of a graph", Coll. Math. Soc. János Bolyai **25**:267-272 (1978).
- [12] J. E. Cohen, U. G. Rothblum, "Nonnegative Ranks, Decompositions, and Factorizations of Nonnegative matrices", Linear Algebra and its Applications, **190**, 1, (1993).
- [13] S. Kochen and E. P. Specker, "The problem of hidden variables in quantum mechanics", Journal of Mathematics and Mechanics **17**, 59-87 (1967).
- [14] A. Peres, "Two simple proofs of the Kochen-Specker theorem", Journal of Physics A: Mathematical and General, Volume 24, Issue 4, pp. L175-L178 (1991).
- [15] T. S. Cubitt, D. Leung, W. Matthews, A. Winter, "Improving zero-error classical communication with entanglement", Phys. Rev. Lett. **104**, 230503 (2010) arXiv:0911.5300.
- [16] S. Beigi, "Entanglement-assisted zero-error capacity is upper bounded by the Lovasz theta function", arXiv:1002.2488 (2010).
- [17] R. Duan, S. Severini, A. Winter, "Zero-error communication via quantum channels, non-commutative graphs and a quantum Lovász ϑ function", arXiv:1002.2514 (2010).
- [18] J. Körner, A. Orlitsky, "Zero-Error Information Theory", IEEE Trans. Inf. Theory **44**(6):2207-2229 (1998).
- [19] L. Lovász, "On the Shannon Capacity of a Graph", IEEE Trans. Inf. Theory **25**(1):1-7 (1979).
- [20] S. Popescu, D. Rohrlich, "Quantum nonlocality as an axiom", Found. Phys. **24**, 379-385 (1994).

APPENDIX NOTATIONS

$[n]$: The set $\{1, \dots, n\}$.

$\mathbf{C}(X \rightarrow Y)$: The set of classical channels with input alphabet X and output alphabet Y .

$\mathbf{C}(A \rightarrow S, B \rightarrow T)$: The set of bipartite classical channels with input alphabets A (for Alice) and B (for Bob) and respective output alphabets S and T .

Ω : Some class of correlations: one of NS = non-signalling, SE = shared entanglement, SR = shared randomness, NC (or omitted) = no correlation.

$\Omega(A \rightarrow S, B \rightarrow T)$: The subset of $\mathbf{C}(A \rightarrow S, B \rightarrow T)$ in the class Ω .

$\mathcal{N}(y|x)$: The probability that the channel \mathcal{N} outputs symbol y when symbol x is input.

$E(G)$: Edges of the graph G .

$E(H)$: Hyperedges of the hypergraph H .

$G(\mathcal{N})$: Confusability graph of the channel \mathcal{N} .

$H(\mathcal{N})$: Hypergraph of the channel \mathcal{N} .

$\chi(G)$: The clique hypergraph of the graph G .

$\alpha(G)$: The independence number of the graph G .

$\alpha^*(H)$: The fractional packing number of the hypergraph H .

$\omega^*(H)$: The fractional covering number of the hypergraph H .

$c_0^\Omega(\mathcal{N})$: One-shot zero-error capacity of \mathcal{N} assisted by Ω .

$C_0^\Omega(\mathcal{N})$: Zero-error capacity of \mathcal{N} assisted by Ω .

$k_0^\Omega(\mathcal{N})$: One-shot simulation cost of \mathcal{N} assisted by Ω .

$K_0^\Omega(\mathcal{N})$: Simulation cost of \mathcal{N} assisted by Ω .

APPENDIX

PAIR-WISE VERSUS MUTUAL DISTINGUISHABILITY FOR SETS OF LOCAL RESIDUAL STATES OF CORRELATIONS

Definition 27. We say that two classical channels \mathcal{N} and \mathcal{M} in $\mathbf{C}(X \rightarrow Y)$ are pair-wise distinguishable, and write $\mathcal{N} \bowtie \mathcal{M}$ if there is an input $x^* \in X$ such that

$$\sum_{y \in Y} \mathcal{N}(y|x^*) \mathcal{M}(y|x^*) = 0.$$

If Alice makes an input a to her side of a bipartite correlation $P \in \mathbf{C}(A \rightarrow X, B \rightarrow Y)$, and obtains the output x then the conditional distribution on Bob's side is a classical channel $P_{ax}(y|b) \in \mathbf{C}(B \rightarrow Y)$ where $P_{ax}(y|b)$ is simply the conditional distribution $P(y|b, a, x)$ given by Bayes rules, but written differently to emphasise the fact that we regard a and x as fixed. Similarly, there are $|B||Y|$ such conditional channels $P_{by}(x|a)$ on Alice's side.

Definition 28. We say that a class of bipartite correlations Ω has property PW_A if the existence of a correlation $P(x, y|a, b) \in \Omega(A \rightarrow X, B \rightarrow Y)$, and $S \subseteq A \times X$, satisfying

$$P_{ax} \bowtie P_{a'x'} \quad \forall (a, x), (a', x') \in S$$

implies the existence of another correlation, $P'(x, y|a, b) \in \Omega(A \rightarrow X, B \cup \{b^*\} \rightarrow Y \cup Y')$,

which is identical to P when restricted to the input alphabets of P ,

$$\forall a \in A, x \in X, b \in B, y \in Y :$$

$$P'(x, y|a, b) = P(x, y|a, b),$$

but which has an extra input b^* on Bob's side such that

$$\forall (a, x), (a', x') \in S :$$

$$\sum_{y \in Y \cup Y'} P'_{ax}(y|b^*) P'_{a'x'}(y|b^*) = 0.$$

Ω has property PW_B if it satisfies the same condition with the roles of the parties reversed. If Ω has property PW_A and property PW_B then we simply say it has property PW .

In other words, if a correlation P belongs to a PW_A class Ω , and the graph induced on $A \times X$ by the pair-wise distinguishability relation between the conditional states P_{ax} associated with the vertices contains a clique S , then there is another correlation P' in Ω which behaves like P except that Bob has some extra output symbols (possibly), and one new input symbol b^* , which when input, yields pair-wise orthogonal output distributions on Y for all elements of S so that they can be perfectly distinguished simultaneously.

To illustrate this idea, we now show that the class NS of generalised non-signalling correlations is not PW : If Alice and Bob's shared correlation P is the PR -box, then the conditional channels P_{ax} are given by

$$P_{ax}(y|b) = [x \oplus y = (a \wedge b)].$$

These channels are all pair-wise distinguishable, but of course, the required input on Bob's side depends on the pair. If NS were PW , then the existence of $P \in \text{NS}$ would imply the existence of another correlation in NS where a single input on Bob's side would suffice to distinguish the 4 residual states. But obviously this would allow Bob to determine Alice's input, so this is a contradiction: Put another way, if a class is PW and contains the PR -box then it also contains signalling correlations.

On the other hand,

Proposition 29. The class of bipartite correlations which can be implemented as local measurements on entangled quantum states (SE) is PW .

Proof: Assuming w.l.o.g. that Alice measures first: Alice inputs a (corresponding to her measuring of some POVM on her side) and obtains outcome x , leaving a residual state ρ_{ax} on Bob's side. The conditional channel $P_{ax}(y|b)$ is given by

$$P_{ax}(y|b) = \text{Tr } \rho_{ax} B_y^{(b)},$$

so if $P_{ax} \bowtie P_{a'x'}$ then there must be some input b on Bob's side, corresponding to a POVM with elements $\{B_y\}_{y \in Y}$ say, such that

$$\begin{aligned} \forall y : P_{ax}(y|b)P_{a'x'}(y|b) \\ = (\text{Tr } B_y \rho_{ax})(\text{Tr } B_y \rho_{a'x'}) = 0 \end{aligned}$$

which implies that the residual states ρ_{ax} and $\rho_{a'x'}$ are orthogonal (i.e. have disjoint support).

A clique of pair-wise distinguishable conditional channels on Bob's side therefore corresponds to a clique of *mutually* orthogonal residual states on his side. Therefore, there is a single measurement which perfectly distinguishes all members of the clique, which we can obviously use to construct a correlation P' in the class of correlations SE with the required properties. ■

From this result and the previous example, it is clear that the PR-box cannot be perfectly implemented by shared entanglement (a fact which can alternatively be proved by the Tsirelson bound).

Proposition 30. *If Ω is a PW class of correlations, then the one-shot (and hence asymptotic) Ω -assisted zero-error capacities $c_0^\Omega(\mathcal{N})$ and $C_0^\Omega(\mathcal{N})$ of a channel \mathcal{N} only depend on the confusability graph $G(\mathcal{N})$.*

Proof: Let P be a correlation in $\Omega([c] \rightarrow X, Y \rightarrow [c])$ such that the standard 'wiring' yields the largest possible identity channel i.e. one with $c = c_0^\Omega(\mathcal{N})$ symbols.

$$\sum_{x,y} P(x, \hat{z}|z, y) = \begin{cases} 1 & \text{if } z = \hat{z} \\ 0 & \text{otherwise.} \end{cases}$$

Write $\delta_y(x, x')$ if there is a single input y on Bob's side such that $\forall z \neq z', \forall x : P_{zx:y} \perp P_{z'x:y}$. In words, $\delta_y(x, x')$ means that if Bob knows that the channel input was one of x or x' then he can distinguish which z Alice chose by making input y to his side of the correlation.

When Bob gets output symbol y from the channel, let e_y denote the set of possible inputs. He can decode z perfectly iff $\delta_y(x, x') \forall x \neq x' \in e_y$. If we draw a

graph on X with edges labelled by outputs Y , with a y -edge between x and x' iff $\delta_y(x, x')$, then (ignoring edge labels and multiplicities) this graph must contain $G(\mathcal{N})$.

Recalling the discussion after Definition 4, we know that $c_0^\Omega(\chi(G(\mathcal{N}))) \leq c_0^\Omega(H(\mathcal{N}))$. By the PW property of Ω it must be possible to find a new correlation in Ω such that if Bob knows that x was in *any* clique in this graph, then he can still determine z . So the Ω -assisted zero-error capacity of the clique hypergraph of G is at least c .

Therefore, if Ω is PW then

$$c_0^\Omega(\chi(G(\mathcal{N}))) = c_0^\Omega(H(\mathcal{N})),$$

so the zero-error capacity depends only on the confusability graph. ■

Proposition 31. *For a correlation $P \in \mathbf{C}(A \rightarrow X, B \rightarrow Y)$, let Δ_B (Δ_A) be the graph on the $|A||X|$ ($|B||Y|$) conditional channels on Bob's (Alice's) side where edges denote pair-wise distinguishability. If P belongs to a class which is both PW and non-signalling, then*

$$\bar{\chi}(\Delta_B) \geq |A|$$

and

$$\bar{\chi}(\Delta_A) \geq |B|$$

where $\bar{\chi}$ denotes the clique covering number. In particular, these bounds apply to the correlation class SE: The set of bipartite correlations which can be implemented using entanglement.

Proof: Let A^* (B^*) be a minimal clique covering of Δ_A (Δ_B). Suppose that P is in Ω which is PW and non-signalling. By repeated use of the definition of the PW property, Ω contains a $P \in \mathbf{C}(A \cup A^* \rightarrow X \cup X', B \cup B^* \rightarrow Y \cup Y')$ such that for $q \in B^*$

$$\begin{aligned} \forall y \in Y \cup Y', (a, x) \in q, (a', x') \in q : \\ P'_{ax}(y|q)P'_{a'x'}(y|q) = 0. \end{aligned}$$

This means that if Alice inputs $a \in A$ to P' and obtains output x , and then tell's Bob a clique in B^* which contains (a, x) , Bob can determine (a, x) exactly. In particular, he discovers Alice's choice of input in A without error, by Alice transmitting one of $|B^*| = \bar{\chi}(\Delta_B)$ messages. Since non-signalling correlations can't increase the zero-error capacity of identity channels (a simple consequence of Theorem 7), if $\bar{\chi}(\Delta_B) < |A|$ then P' cannot be non-signalling (and similarly if $\bar{\chi}(\Delta_A) < |B|$). ■