



PAPER

Optimal quantum networks and one-shot entropies

OPEN ACCESS

RECEIVED

14 June 2016

REVISED

5 August 2016

ACCEPTED FOR PUBLICATION

6 September 2016

PUBLISHED

29 September 2016

Original content from this work may be used under the terms of the [Creative Commons Attribution 3.0 licence](#).

Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.

Giulio Chiribella^{1,2,3} and Daniel Ebler¹¹ Department of Computer Science, The University of Hong Kong, Pokfulam Road, Hong Kong² Canadian Institute for Advanced Research, CIFAR Program in Quantum Information Science, Toronto, ON M5G 1Z8, Canada³ Author to whom any correspondence should be addressed.E-mail: Giulio@cs.hku.hk**Keywords:** quantum networks, quantum causal structures, optimal controlization of unitary gates, conditional-min entropy, semidefinite programming, max relative entropy, quantum noncausal games**Abstract**

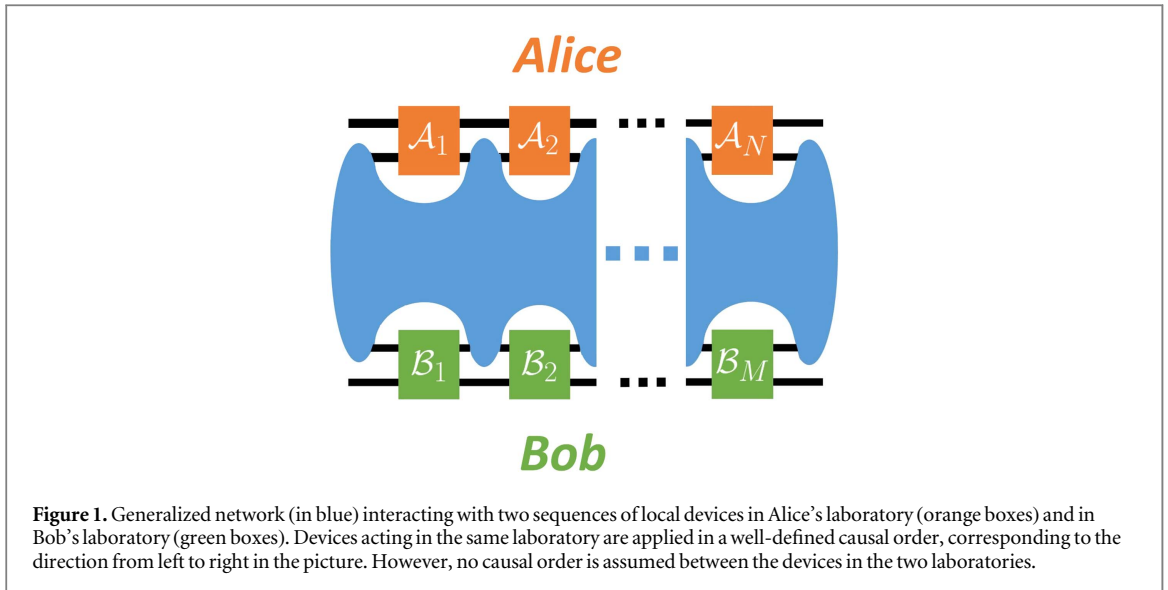
We develop a semidefinite programming method for the optimization of quantum networks, including both causal networks and networks with indefinite causal structure. Our method applies to a broad class of performance measures, defined operationally in terms of interactive tests set up by a verifier. We show that the optimal performance is equal to a max relative entropy, which quantifies the informativeness of the test. Building on this result, we extend the notion of conditional min-entropy from quantum states to quantum causal networks. The optimization method is illustrated in a number of applications, including the inversion, charge conjugation, and controlization of an unknown unitary dynamics. In the non-causal setting, we show a proof-of-principle application to the maximization of the winning probability in a non-causal quantum game.

1. Introduction

Advances in quantum communication [1–3] and in the integration of quantum hardware [4–8] are pushing towards the realization of networked quantum information systems, such as quantum communication networks [9–13] and distributed quantum computing [14–16]. Networks of interacting quantum devices are attracting interest also at the theoretical level, providing a framework for quantum games [17] and protocols [18–20], insights on the foundations of quantum mechanics [18, 21–23], a starting point for a general theory of Bayesian inference [24–31] and for the development of models of higher-order quantum computation [32–34].

The network scenario motivates a new set of optimization problems, where the goal is not to optimize individual devices, but rather to optimize how different devices interact with one another. In many situations, the devices operate in a well-defined causal order—this is the case, for example, in the circuit model of quantum computing, where computations are implemented by sequences of gates [35, 36]. Recently, researchers have started to investigate more general situations, where the causal order can be in a quantum superposition [20, 33, 34, 37–39] or can be indefinite in other more exotic ways, in principle compatible with quantum mechanics [34, 40–45]. In these new situations, optimizing quantum networks is important, for at least three reasons: First, in order to establish an advantage, one has to first find the optimal performances achievable in a definite causal order. Second, finding the maximum advantage requires an optimization over all non-causal networks. This is an essential step for assessing the power of the new, non-causal models of information processing. Third, identifying the ultimate performances achieved in the absence of pre-defined causal structure is expected to shed light on the interplay between quantum mechanics and spacetime.

In this paper we develop a semidefinite programming (SDP) approach to the optimization of quantum networks. We start by analyzing scenarios with definite causal order, choosing an operational measure of performance, quantified by how much the network scores in a given test. The test consists in sending inputs to the devices, performing local computations, and finally measuring the outputs. Tests of this type are also important in the theory of quantum interactive proof systems [46], wherein they are used to model the interaction between a prover and a verifier. The input-output behavior of a quantum causal network is described in the framework of *quantum combs* [18, 47] (also known as *quantum strategies* [17]), which associates a positive



operator to any given sequence of quantum operations. In this framework, the optimization is a SDP. We work out the dual optimization problem, showing that the maximum score is quantified by a one-shot entropic quantity that characterizes the informativeness of the test. This quantity extends to networks the notion of *max relative entropy* [48–51] (see also the monograph [52]). Building on the connection with the max relative entropy, we define a measure of the amount of correlations that a causal network can generate over time. This quantity is based on the notion of *conditional min-entropy* [50, 51], originally defined for quantum states and extended here to quantum causal networks.

After discussing the causal case, we turn our attention to quantum networks with indefinite causal order. Some of these networks arise when multiple quantum devices are connected in a way that is controlled by the state of a quantum system [18, 20, 38]. Some other networks are not built by linking up individual devices [41]. They are ‘networks’ in a generalized sense: they are spatially distributed objects that can interact with a set of local devices. The description of these generalized networks is trickier, because we cannot specify their behavior in terms of the behavior of individual quantum devices. Instead, we must characterize them through the way they respond to external inputs. More specifically, a general quantum network is specified by a map that accepts as input the operations taking place in local laboratories and returns as output an operation, as figure 1. Maps that transform quantum operations are known as *quantum supermaps*. They were originally introduced in the causal scenario [18, 53] and later generalized to the case of networks with indefinite causal structure [18, 34, 41]. These maps can be represented by positive operators, subject to a set of constraints that guarantee that valid operations are transformed into valid operations. Again, the form of these constraints leads to SDPs. In this case, we find that the maximum score can be expressed in terms of a max relative entropy, here named the *max relative entropy of signaling*, which quantifies the deviation from the set of no-signaling channels. In addition, we characterize the max relative entropy between two non-causal network, showing that it is equal to the maximum of the max relative entropy over all the states that can be generated by interacting with the two networks. This result opens the way to the definition of hypothesis testing protocols to probe the fundamental structure of spacetime, by testing the possibility of exotic non-causal networks against the null hypothesis that events have a well-defined causal structure.

To illustrate the general method, we provide a number of applications to concrete tasks, involving the optimization of both causal and non-causal networks. For the optimization in the causal setting, we consider the tasks of inverting an unknown unitary dynamics, simulating the evolution of a charge conjugate particle, and adding control to an unknown unitary gate. Looking at these tasks in terms of network optimization is a relatively new approach and here we provide the first optimized solutions. For the optimization in the non-causal setting, we illustrate our method by analyzing the non-causal game introduced by Oreshkov *et al* [41]. In this case, we fix the operations performed by the players (as in [41]) and we search for the non-causal network that offers the largest advantage for these operations. Using the SDP approach, we obtain a simple proof of the optimality of the network presented in [41]. Optimality can also be derived from a recent result of Brukner [54], who considered a more general scenario where the players’ operations are not fixed, but rather subject to optimization. When the operations are fixed as in [41], however, our SDP technique yields a significantly shorter optimality proof. The simplification in this restricted scenario suggests that SDP may prove useful also for the

broader scope of identifying a non-causal analogue of the Tsirelson bound, which was the motivating problem of [54].

The paper is organized as follows. In section 2 we introduce the framework of quantum combs and the characterization of quantum causal networks. In section 3 we review the basic facts about SDP and establish a general relation with the max relative entropy. The general result is applied to quantum causal networks in section 4 and is then used to define a suitable extension of the conditional min-entropy (section 5) and of the max relative entropy (section 6). In sections 7 and 8 we extend the results to quantum networks without predefined causal structure. Our techniques are illustrated in section 9, where we present applications to the tasks of inverting unknown evolutions, simulating charge conjugation, controlling unitary gates, and maximizing the winning probability in a non-causal quantum game. Finally, the conclusions are drawn in section 10.

2. The framework of quantum combs

In this section we introduce the concepts required for the optimization of quantum causal networks. First of all, we review the connection between quantum channels and operators. Then, we present the basics of the framework of quantum combs.

2.1. Quantum operations, quantum channels, and the Choi isomorphism

Quantum operations [55] describe the most general transformations of quantum systems, including both the reversible transformations associated to unitary gates and the irreversible transformations due to measurements. A quantum operation with input system A and output system B is a completely positive trace non-increasing map \mathcal{C} , transforming operators on the input Hilbert space \mathcal{H}_A into operators on the output Hilbert space \mathcal{H}_B . We will often use the diagrammatic notation

$$\text{---} \overset{A}{\square} \boxed{\mathcal{C}} \text{---} \overset{B}{\square} \text{---} . \tag{1}$$

We say that the quantum operation \mathcal{C} in the above diagram is of type $A \rightarrow B$.

When system A is trivial—that is, when its Hilbert space is one-dimensional—the quantum operation \mathcal{C} corresponds to the preparation of a *state* of system B , diagrammatically represented as $\text{---} \overset{B}{\square} \text{---}$. When system B is trivial, the quantum operation \mathcal{C} in equation (1) corresponds to a measurement *effect* on system A and is represented as $\text{---} \overset{A}{\square} \boxed{P} \text{---}$. Measurement effects are positive (semidefinite) operators P satisfying $P \leq I$, where I is the identity operator on the system's Hilbert space. Effects are associated to the outcomes of measurements and the probability of the outcome corresponding to the effect P is given by the Born rule

$$\text{---} \overset{A}{\square} \boxed{\rho} \text{---} \overset{A}{\square} \boxed{P} \text{---} = \text{Tr}[P\rho], \tag{2}$$

where ρ is the state of the system before the measurement. In the special case where P is the identity operator, we represent the corresponding effect as $\text{---} \overset{A}{\square} \boxed{\text{Tr}} \text{---}$.

In general, quantum measurement processes are described by *quantum instruments*. A quantum instrument with input A and output B is a collection of quantum operations $\{\mathcal{C}_x\}_{x \in X}$ of type $A \rightarrow B$, subject to the condition that the sum $\sum_{x \in X} \mathcal{C}_x$ is trace-preserving. Each quantum operation corresponds to one alternative outcome x and the probability that the quantum operation \mathcal{C}_x takes places on a given input state ρ is given by

$$\text{---} \overset{A}{\square} \boxed{\rho} \text{---} \overset{A}{\square} \boxed{\mathcal{C}_x} \text{---} \overset{B}{\square} \boxed{\text{Tr}} \text{---} = \text{Tr}[\mathcal{C}_x(\rho)], \tag{3}$$

When the instrument $\{\mathcal{C}_x\}_{x \in X}$ has a single outcome, say x_0 , the corresponding process is *deterministic*, meaning that one can predict in advance that the outcome will be x_0 . In this case, the quantum operation \mathcal{C}_{x_0} is trace preserving. Trace preserving quantum operations are also known as *quantum channels*.

Completely positive maps can be represented by positive operators. Let $\text{Lin}(\mathcal{H})$ be the space of linear operators on the Hilbert space \mathcal{H} and let \mathcal{C} be a completely positive map transforming operators in $\text{Lin}(\mathcal{H}_0)$ into operators on $\text{Lin}(\mathcal{H}_1)$. Then, the map \mathcal{C} can be represented by a positive operator $C \in \text{Lin}(\mathcal{H}_1 \otimes \mathcal{H}_0)$, defined as

$$C = (C \otimes \mathcal{I}_0)(|I\rangle\rangle\langle\langle I|), \tag{4}$$

where \mathcal{I}_0 denotes the identity map on $\text{Lin}(\mathcal{H}_0)$ and $|I\rangle\rangle$ is the unnormalized maximally entangled state $|I\rangle\rangle = \sum_i |i\rangle |i\rangle \in \mathcal{H}_0 \otimes \mathcal{H}_0$. The operator C is known as the *Choi operator* [56].

Quantum operations and quantum channels can be characterized in terms of their Choi operators: a positive operator $Q \in \text{Lin}(\mathcal{H}_1 \otimes \mathcal{H}_0)$ corresponds to a quantum operation if and only if it satisfies the condition

$$\mathrm{Tr}_1[Q] \leq I_0, \quad (5)$$

where Tr_1 denotes the partial trace over the Hilbert space \mathcal{H}_1 , I_0 denotes the identity operator on the Hilbert space \mathcal{H}_0 , and \leq denotes the standard operator order: $A \leq B$ iff $\langle \varphi | A | \varphi \rangle \leq \langle \varphi | B | \varphi \rangle$, $\forall |\varphi\rangle \in \mathcal{H}_0$. A positive operator $C \in \mathrm{Lin}(\mathcal{H}_1 \otimes \mathcal{H}_0)$ corresponds to a quantum channel if and only if it satisfies the condition

$$\mathrm{Tr}_1[C] = I_0. \quad (6)$$

2.2. The link product

Two quantum operations can be connected with each other, as long as the output of the first operation matches the input of the second. At the level of Choi operators, the connection is implemented by the operation of *link product* [47], denoted as $*$. To define the link product, it is convenient to introduce the following shorthand notation: if A is an operator on $\mathcal{H}_X \otimes \mathcal{H}_Y$ and B is an operator on $\mathcal{H}_Y \otimes \mathcal{H}_Z$, then we use the notation AB for the product

$$AB := (A \otimes I_Z)(I_X \otimes B). \quad (7)$$

With this notation, the link product of A and B is the operator $A * B$ defined as

$$A * B := \mathrm{Tr}_Y[A B^{T_Y}], \quad (8)$$

where B^{T_Y} denotes the partial transpose of B with respect to the Hilbert space \mathcal{H}_Y . Note that the definition of the link product presupposes that the Hilbert spaces have been labeled: in order to compute the link product, one needs to take the partial transpose and the trace on the Hilbert space in common between A and B .

Mathematically, the partial transpose in the rhs of equation (8) is essential to guarantee that the link product of two positive operators is a positive operator [47]. As a counterexample, think of the case where \mathcal{H}_X , \mathcal{H}_Y , and \mathcal{H}_Z are two-dimensional and A and B are projectors on a maximally entangled state: in this case, removing the partial transpose results in a non-positive $A * B$. Physically, the role of the partial transpose can be understood in terms of entanglement swapping [57, 58]. Thanks to the partial transpose, the link product can be expressed as

$$A * B = \mathrm{Tr}_Y \mathrm{Tr}_{Y'} [(A_{XY} \otimes B_{Y'Z})(I_X \otimes |I\rangle\langle I| \otimes I_Z)], \quad (9)$$

where $|I\rangle := \sum_{n=1}^{d_Y} |n\rangle |n\rangle$ is the unnormalized maximally entangled state on $\mathcal{H}_Y \otimes \mathcal{H}_{Y'}$, $\mathcal{H}_{Y'}$ being an identical copy of \mathcal{H}_Y . This means that, up to normalization, the link product $A * B$ is the state obtained when a Bell measurement, performed on the states $A/\mathrm{Tr} A$ and $B/\mathrm{Tr} B$, yields the outcome corresponding to the projector $|I\rangle\langle I|/d_Y$. At the fundamental level, the possibility of representing operations as states and their composition as entanglement swapping follows from the Purification Principle—the property that every state can be obtained as the marginal of a pure state, unique up to reversible transformations [59].

The link product is associative, namely

$$A * (B * C) = (A * B) * C,$$

for all operators A , B , and C . Moreover, the link product is commutative, up to re-ordering of the Hilbert spaces: in formula,

$$A * B \simeq B * A,$$

having used the notation $A * B \simeq B * A$ to mean $A * B = \mathrm{SWAP}_{XZ} (B * A) \mathrm{SWAP}_{XZ}$, where SWAP_{XZ} is the unitary operator that swaps the spaces \mathcal{H}_X and \mathcal{H}_Z . From now on we will omit the swaps, implicitly understanding that the Hilbert spaces have been reordered in the right way wherever needed.

Using the above notation, we have the following

Proposition 1 [47]. *Let \mathcal{A} be a quantum operation transforming operators on \mathcal{H}_0 to operators on \mathcal{H}_1 , let \mathcal{B} be a quantum operation transforming operators on \mathcal{H}_1 to operators on \mathcal{H}_2 , and let $C = \mathcal{B}\mathcal{A}$ be the quantum operation resulting from the composition of \mathcal{A} and \mathcal{B} . Then, one has*

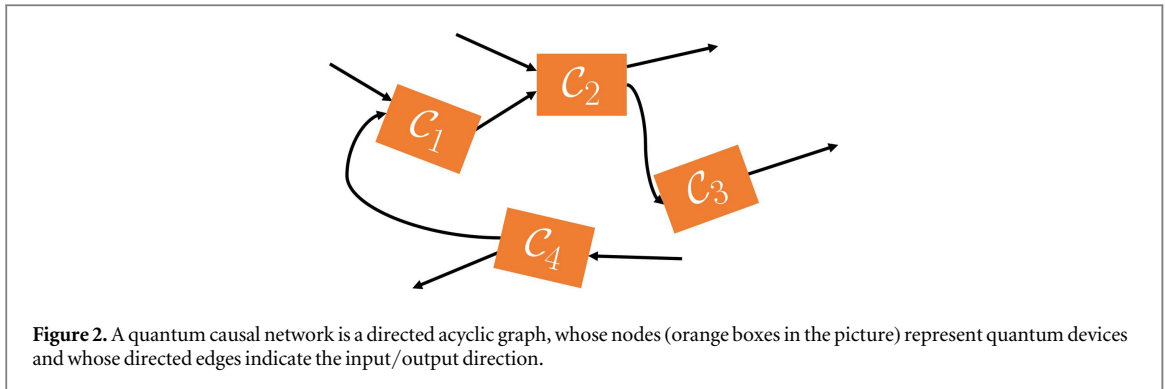
$$C := A * B,$$

where A , B , and C are the Choi operators of \mathcal{A} , \mathcal{B} , and C , respectively.

In the next paragraph we will use the link product to construct the Choi operator of quantum networks consisting of multiple interconnected quantum operations.

2.3. Quantum causal networks and quantum combs

A quantum network is a collection of quantum devices connected with each other. We will call the network *causal* if there are no loops connecting the output of a device to the output of the same device. Mathematically, a quantum causal network can be represented by a direct acyclic graph, where each vertex of the graph corresponds to a quantum device—see figure 2. For every DAG, one can always define a total ordering of



the vertices, through a procedure known as topological sorting [60]. Using this fact, one can always represent the a quantum causal network as an ordered sequence of quantum devices, such as

$$\begin{array}{c}
 A_1^{\text{in}} \quad A_1^{\text{out}} \quad A_2^{\text{in}} \quad A_2^{\text{out}} \quad \dots \quad A_N^{\text{in}} \quad A_N^{\text{out}} \\
 \left[\begin{array}{c} \text{---} \\ \text{---} \\ \text{---} \end{array} \right] C_1 \left[\begin{array}{c} \text{---} \\ \text{---} \\ \text{---} \end{array} \right] C_2 \left[\begin{array}{c} \text{---} \\ \text{---} \\ \text{---} \end{array} \right] \dots \left[\begin{array}{c} \text{---} \\ \text{---} \\ \text{---} \end{array} \right] C_N \left[\begin{array}{c} \text{---} \\ \text{---} \\ \text{---} \end{array} \right]
 \end{array}
 , \tag{10}$$

where A_j^{in} (A_j^{out}) denotes the input (output) system of the network at the j th time step.

We say that a network is *deterministic* if all devices in the network are deterministic, i.e. if they are described by quantum channels. Using the link product, we associate a Choi operator to the network: specifically, if the individual channels in the network have Choi operators C_1, C_2, \dots, C_N , then the network has Choi operator

$$C = C_1 * C_2 * C_3 * \dots * C_N. \tag{11}$$

The Choi operator of a deterministic network is called a *quantum comb* [18, 47], or also a *quantum strategy* [17]. The quantum comb C is a positive operator on $\bigotimes_{j=1}^N (\mathcal{H}_j^{\text{out}} \otimes \mathcal{H}_j^{\text{in}})$, where $\mathcal{H}_j^{\text{in}}$ ($\mathcal{H}_j^{\text{out}}$) is the Hilbert space of system A_j^{in} (A_j^{out}). Quantum combs can be characterized as follows:

Proposition 2 [17, 18, 47]. *A positive operator C is a quantum comb if and only if it satisfies the linear constraints*

$$\text{Tr}_{A_n^{\text{out}}}[C^{(n)}] = I_{A_n^{\text{in}}} \otimes C^{(n-1)} \quad \forall n \in \{1, \dots, N\}, \tag{12}$$

where Tr_A is the partial trace over the Hilbert space \mathcal{H}_A , $C^{(n)}$ is a suitable operator on $\mathcal{H}_n := \bigotimes_{j=1}^n (\mathcal{H}_j^{\text{out}} \otimes \mathcal{H}_j^{\text{in}})$, $C^{(N)} := C$, and $C^{(0)} := 1$.

The constraints in equation (12) are a direct consequence of the normalization condition of quantum channels, expressed by equation (6). Physically, the positive operator $C^{(n)}$ represents the subnetwork transforming the first n inputs to the first n outputs. We denote by

$$\text{Comb}(A_1^{\text{in}} \rightarrow A_1^{\text{out}}, A_2^{\text{in}} \rightarrow A_2^{\text{out}}, \dots, A_N^{\text{in}} \rightarrow A_N^{\text{out}})$$

the set of positive semidefinite operators satisfying the constraint (12). When there is no ambiguity, we will simply write Comb .

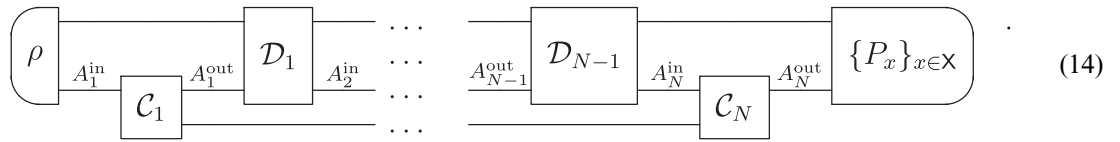
2.4. Quantum testers and the generalized Born rule

So far we considered deterministic networks, resulting from the connection of quantum channels. However, it is also useful to consider networks containing measurement devices, which may generate random outcomes. We call such networks *non-deterministic*. Non-deterministic quantum networks can be thought as the quantum version of classical electric networks containing measurement devices, such as voltmeters and ammeters. Like these classical relatives are useful for testing the behavior of electrical circuits, quantum non-deterministic networks are useful for testing the behavior of quantum circuits, or, slightly more broadly, physical processes consisting of multiple time steps.

An example of non-deterministic network is the following

$$\left(\rho \begin{array}{c} \text{---} \\ \text{---} \\ \text{---} \end{array} \right) A_1^{\text{in}} \quad A_1^{\text{out}} \quad \left[\begin{array}{c} \text{---} \\ \text{---} \\ \text{---} \end{array} \right] D_1 \left[\begin{array}{c} \text{---} \\ \text{---} \\ \text{---} \end{array} \right] A_2^{\text{in}} \quad \dots \quad A_{N-1}^{\text{out}} \quad \left[\begin{array}{c} \text{---} \\ \text{---} \\ \text{---} \end{array} \right] D_{N-1} \left[\begin{array}{c} \text{---} \\ \text{---} \\ \text{---} \end{array} \right] A_N^{\text{in}} \quad A_N^{\text{out}} \quad \left\{ P_x \right\}_{x \in X} \tag{13}$$

where ρ is a quantum state, (D_1, \dots, D_{N-1}) is a sequence of quantum channels, and $\{P_x\}_{x \in X}$ is a positive operator-valued measure (POVM), describing a quantum measurement on the last output system. Networks of the type (13) can be used to probe quantum networks of the type (10), as follows



When the two networks are wired together, the final measurement produces one of the outcomes in the set X . Using proposition 1, the probability of the outcome x is computed as

$$\begin{aligned}
 p_x &= \rho * C_1 * D_1 * C_2 * D_2 * \cdots * D_{N-1} * C_N * P_x^T \\
 &= (\rho * D_1 * D_2 * \cdots * D_{N-1} * P_x^T) * (C_1 * C_2 * \cdots * C_N) \\
 &= T_x * C \\
 &= \text{Tr}[T_x C^T],
 \end{aligned} \tag{15}$$

where C is the Choi operator of the tested network, C^T is the transpose of C , and $\{T_x\}_{x \in X}$ is the collection of operators defined by

$$T_x := \rho * D_1 * D_2 * \cdots * D_{N-1} * P_x^T \tag{16}$$

(here the transpose of P_x is needed because, according to definition 4, the Choi operator of the quantum operation $\mathcal{Q}_x(\cdot) = \text{Tr}[P_x \cdot]$ is P_x^T instead of P_x).

We call the set of operators $\mathbf{T} = \{T_x\}_{x \in X}$ a *quantum tester* and equation (15) the *generalized Born rule* [18, 61, 62]. The quantum tester \mathbf{T} describes the response of the non-deterministic network (13) when connected to external devices. Quantum testers are a useful abstraction in many applications, such as quantum games [17] and cryptographic protocols [19, 20], quantum interactive proof systems [46], quantum learning of gates [63–65], quantum channel discrimination [61, 66, 67], incompatibility of multitime quantum measurements [68], tomography of quantum channels [62, 69], non-Markovian processes [70, 71], and causal models [28].

Quantum testers can be characterized as follows:

Proposition 3 [61]. *Let \mathbf{T} be a collection of positive operators on $\bigotimes_{j=1}^N (\mathcal{H}_j^{\text{out}} \otimes \mathcal{H}_j^{\text{in}})$. \mathbf{T} is a quantum tester if and only if*

$$\begin{aligned}
 \sum_{x \in X} T_x &= I_{A_N^{\text{out}}} \otimes \Gamma^{(N)} \\
 \text{Tr}_{A_n^{\text{in}}}[\Gamma^{(n)}] &= I_{A_{n-1}^{\text{out}}} \otimes \Gamma^{(n-1)}, \quad n = 2, \dots, N \\
 \text{Tr}_{A_1^{\text{in}}}[\Gamma^{(1)}] &= 1,
 \end{aligned} \tag{17}$$

where each $\Gamma^{(n)}$, $n = 1, \dots, N$ is a positive operator on $\mathcal{H}_n^{\text{in}} \otimes [\bigotimes_{j=1}^{n-1} (\mathcal{H}_j^{\text{out}} \otimes \mathcal{H}_j^{\text{in}})]$.

2.5. Assessing the performance of a quantum network

Suppose that we are given black box access to a quantum network, whose internal functioning is unknown to us. Our goal is to assess how well the network fares in a desired task, such as solving a desired computational problem [72], estimating an unknown parameter [64, 65], emulating a sequence of gates [63, 73], or replicating the action of a desired gate [74–78].

For example, suppose that a manufacturer provides us with a special-purpose computer, designed to implement a quantum search algorithm. How can we test the performance of our device? Since the computer is claimed to find the location of an item in a list, a natural approach is to place the item in a set of random positions and then to check whether the answer provided by the computer is correct. A simple measure of performance is given the number of inputs on which the computer gives the right answer. More generally, one can assign different scores depending on the distance between the correct answer and the output of the computer. Let us consider this example in more detail, as a concrete illustration of what it means to test a quantum network. Suppose that the computer attempts at reproducing Grover's algorithm [79], by interacting with unitary gates $U_i = 2|i\rangle\langle i| - I$ that encode the position of an item i in a list of K items. A possible test, illustrated in figure 3, is as follows:

- (i) Prepare a 'position register' in the maximally mixed state $\rho = I/K$.
- (ii) Upon receiving an input from the computer, apply the control-unitary gate $W = \sum_{i=1}^K |i\rangle\langle i| \otimes U_i$ to the position register and the input.
- (iii) Repeat the previous operation until the computer returns an output. In this way, the input provided by the computer is processed by a gate U_i , with i chosen at random.

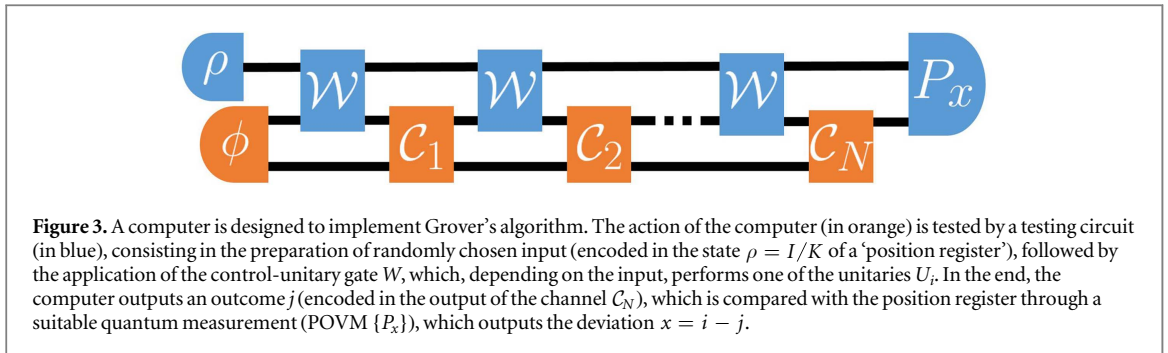


Figure 3. A computer is designed to implement Grover’s algorithm. The action of the computer (in orange) is tested by a testing circuit (in blue), consisting in the preparation of randomly chosen input (encoded in the state $\rho = I/K$ of a ‘position register’), followed by the application of the control-unitary gate W , which, depending on the input, performs one of the unitaries U_i . In the end, the computer outputs an outcome j (encoded in the output of the channel C_N), which is compared with the position register through a suitable quantum measurement (POVM $\{P_x\}$), which outputs the deviation $x = i - j$.

- (iv) Compare the output with the actual position, by performing a joint measurement to the position register and the output register. The measurement is described by the POVM $\{P_x\}_{x=-K}^K$ with operators given by

$$P_x = \sum_{i=\max\{0,-x\}}^{\min\{K,K-x\}} |i\rangle\langle i| \otimes |i+x\rangle\langle i+x|.$$

In this way, the measurement outcome returns the deviation x from the correct position

- (v) If the deviation is x , assign score $\omega_x = 1 - |x|/K$.

Mathematically, the test is represented by the quantum tester $\{T_x\}_{x=-K}^K$ with

$$T_x = \rho * |W\rangle\rangle\langle\langle W| * \dots * |W\rangle\rangle\langle\langle W| * P_x^T.$$

The sequence of operations performed by the computer is represented by the quantum comb

$$C = |\phi\rangle\langle\phi| * C_1 * \dots * C_N,$$

and the probability of finding a deviation x is given by

$$p_x = T_x * C.$$

The average score obtained by the computer can be expressed as

$$\begin{aligned} \omega &= \sum_{i,j=0}^K \left(1 - \frac{|i-j|}{K}\right) p_{i-j} \\ &= \sum_{x=-K}^K \left(1 - \frac{|x|}{K}\right) T_x * C \\ &= \Omega * C, \end{aligned}$$

where Ω is the operator $\Omega := \sum_x \left(1 - \frac{|x|}{K}\right) T_x$.

Generalizing the above example, we assess the performance of an unknown quantum network by referring to experiments where the unknown network is connected to a ‘testing network’, containing measuring devices. The testing network will return an outcome x , to which one can assign a ‘score’ ω_x . In this way, the expected score serves as an operational measure of performance. Specifically, let C be the quantum comb describing the tested network and let $T = \{T_x, x \in X\}$ be the quantum tester describing the testing network. Then, the average score is given by

$$\begin{aligned} \omega &= \sum_x \omega_x (T_x * C) \\ &= \Omega * C \quad \Omega := \sum_x \omega_x T_x. \end{aligned} \tag{18}$$

Note that the performance of the network C is completely determined by the operator Ω , which we call the *performance operator*.

For a given performance operator Ω , the maximum expected score is given by

$$\begin{aligned} \omega_{\max} &:= \max_{C \in \text{Comb}} \Omega * C \\ &= \max_{C \in \text{Comb}} \text{Tr}[\Omega C^T] \\ &= \max_{C \in \text{Comb}} \text{Tr}[\Omega C]. \end{aligned} \tag{19}$$

The third equality comes from the fact that the set of quantum combs is closed under transposition and, therefore, we can omit the transpose in equation (15). Using the notation

$$\langle A, B \rangle := \text{Tr}[AB], \quad (20)$$

we express the maximum score as

$$\omega_{\max} := \max_{C \in \text{Comb}} \langle \Omega, C \rangle. \quad (21)$$

The above equation shows that the search for the maximum score is a SDP. The basic tools needed to address it will be reviewed in the next section.

3. SDP and the max relative entropy

3.1. Basic facts about SDP

Here we review the background about SDP. For further details, we refer the reader to Watrous' lecture notes [80].

Let \mathcal{X} and \mathcal{Y} be two Hilbert spaces and let $\text{Herm}(\mathcal{X})$ be the space of Hermitian operators on \mathcal{X} and \mathcal{Y} , respectively.

Definition 1. A SDP is a triple (ϕ, A, B) , where A and B are operators in $\text{Herm}(\mathcal{X})$ and $\text{Herm}(\mathcal{Y})$, respectively, and ϕ is a linear map from $\text{Herm}(\mathcal{X})$ to $\text{Herm}(\mathcal{Y})$.

A SDP is associated to an optimization problem in the standard form

$$\begin{aligned} & \text{maximize } \langle A, X \rangle \\ & \text{subject to } \phi(X) = B \\ & X \geq 0. \end{aligned} \quad (22)$$

This problem is known as the *primal*. The *dual* problem is

$$\begin{aligned} & \text{minimize } \langle B, Y \rangle \\ & \text{subject to } \phi^\dagger(Y) \geq A \\ & Y \in \text{Herm}(\mathcal{Y}), \end{aligned} \quad (23)$$

where ϕ^\dagger is the adjoint of ϕ , namely the linear map defined by the relation

$$\langle X, \phi^\dagger(Y) \rangle = \langle \phi(X), Y \rangle, \quad \forall X \in \text{Herm}(\mathcal{X}), \forall Y \in \text{Herm}(\mathcal{Y}).$$

The optimal values of the primal and dual problems, denoted as

$$\omega_{\text{primal}} := \sup \langle A, X \rangle \quad \text{and} \quad \omega_{\text{dual}} := \inf \langle B, Y \rangle,$$

are related by duality: for every SDP, one has the weak duality $\omega_{\text{primal}} \leq \omega_{\text{dual}}$. The strong duality $\omega_{\text{primal}} = \omega_{\text{dual}}$ holds under suitable conditions, provided by Slater's theorem [81]. In this paper we will use the following.

Proposition 4. Let (ϕ, A, B) be a SDP. If there exists a positive operator X satisfying $\phi(X) = B$ and an Hermitian operator Y satisfying $\phi^\dagger(Y) > A$, then $\omega_{\text{primal}} = \omega_{\text{dual}}$.

For the proof, see e. g. [80].

3.2. The max relative entropy

An important quantity in one-shot quantum information theory is the *max relative entropy*, introduced by Datta in [82]:

Definition 2. Let A and B be two positive operators on \mathcal{X} . The max entropy of A relative to B is given by

$$D_{\max}(A \parallel B) := -\log \max \{w \mid wA \leq B\}, \quad (24)$$

with the convention $\log 0 := -\infty$.

The max relative entropy provides one way to quantify the deviation of A from B . More generally, it is useful to consider the deviation between A and a *set* of operators:

Definition 3. Let A be a positive operator on \mathcal{X} and let $\mathbf{S} \subset \text{Herm}(\mathcal{X})$ be a set of positive operators. The max entropy of A relative the set \mathbf{S} , denoted as $D_{\max}(A \parallel \mathbf{S})$, is the quantity defined by

$$D_{\max}(A \parallel \mathbf{S}) := \inf_{B \in \mathbf{S}} D_{\max}(A \parallel B). \quad (25)$$

The max relative entropy between a quantum state and a set of quantum states plays a central role in

entanglement theory [83], where relative entropies are used to quantify the deviation from the set of separable states, and in quantum thermodynamics [84, 85], where relative entropies are used to quantify the deviation from the set of Gibbs states. In this paper we will extend the application of the max relative entropy to dynamical scenarios, where \mathbf{S} represents a set of quantum networks. This extension is promising, e. g. for applications to hypothesis testing. Indeed, it is natural to consider scenarios where one has a null hypothesis on the input-output behavior of a quantum network and one wants to test the null hypothesis against an alternative hypothesis. In the case of quantum states, the minimum probability of a type II error (failing to accept the alternative hypothesis) can be estimated in terms of the max relative entropy [86]. In the case of quantum networks, it is natural to expect that the max relative entropy defined here will yield similar bounds—a result in this direction will be provided in sections 6 and 8.

3.3. From SDPs to the max relative entropy

In this section we provide a general bound on the primal value of an arbitrary SDP. The bound can always be attained and its value can be expressed in terms of a max relative entropy whenever the operator A in the SDP (ϕ, A, B) is positive. To state the result, we need some basic notation, provided in the following:

For a vector space \mathcal{V} , we denote by \mathcal{V}^* the dual space, i. e. the space of linear functionals on \mathcal{V} . Given a subset $\mathbf{S} \subseteq \mathcal{V}$, we define the *dual affine space* $\bar{\mathbf{S}}$ as

$$\bar{\mathbf{S}} := \{\Gamma \in \mathcal{V}^* \mid \langle \Gamma, X \rangle = 1, \forall X \in \mathbf{S}\}.$$

Regarding \mathcal{V} as a subspace of \mathcal{V}^{**} , one has the inclusion $\mathbf{S} \subseteq \bar{\bar{\mathbf{S}}}$. When \mathcal{V} is finite dimensional and \mathbf{S} is an affine set, one has the equality $\mathbf{S} = \bar{\bar{\mathbf{S}}}$.

Given a SDP (ϕ, A, B) , we define the *primal affine space* as

$$\mathbf{S} := \{X \in \text{Herm}(\mathcal{X}) \mid \phi(X) = B\}. \quad (26)$$

Simply, \mathbf{S} is the set of operators that satisfy the equality constraint of the primal problem. The *dual affine space* is given by

$$\bar{\mathbf{S}} = \{\Gamma \in \text{Herm}(\mathcal{X}) \mid \langle \Gamma, X \rangle = 1, \forall X \in \mathbf{S}\}, \quad (27)$$

having used the identification of $\text{Herm}(\mathcal{X})$ with its dual space. With this notation, we have

Theorem 1. *Let (ϕ, A, B) be a SDP. The optimal solution of the primal problem is upper bounded as*

$$\omega_{\text{primal}} \leq \inf_{\Gamma \in \bar{\mathbf{S}}} \min \{\lambda \in \mathbb{R} \mid \lambda \Gamma \geq A\}, \quad (28)$$

where $\bar{\mathbf{S}}$ is the dual affine space defined in equation (27). If \mathbf{S} contains a positive operator and $\bar{\mathbf{S}}$ contains a strictly positive operator, then equation (28) holds with the equality sign. If, in addition, the operator A is positive, then one has the expression

$$\omega_{\text{primal}} = 2^{D_{\max}(A \parallel \bar{\mathbf{S}}_+)}, \quad (29)$$

where $\bar{\mathbf{S}}_+$ is the dual convex set $\bar{\mathbf{S}}_+ := \{\Gamma \in \bar{\mathbf{S}} \mid \Gamma \geq 0\}$.

The proof can be found in appendix A.

We call the quantity $D(A \parallel \bar{\mathbf{S}}_+)$ the *max divergence from normalization*. This quantity measures how much the operator A deviates from the set of positive functionals that are normalized on every element of the primal set.

The connection between SDP and the max relative entropy has previously appeared in the special case where the task is to optimize quantum channels [51, 87]. A related result was obtained by Jenčová in the framework of base norms [88]. In the next sections we will elaborate on the physical meaning of theorem 1, which will be applied to the optimization of quantum networks, both with definite and indefinite causal structure. Before specializing ourselves to quantum networks, however, it is worth emphasizing a simple connection between the max relative entropy arising in generic SDPs and the max relative entropy of quantum states.

Proposition 5. *Let C_0 and C_1 be two elements of the convex set*

$$\mathbf{S}_+ = \{X \in \text{Herm}(\mathcal{X}) \mid \phi(X) = B, X \geq 0\}.$$

Then, one has the bound

$$D_{\max}(\sqrt{\Gamma} C_0 \sqrt{\Gamma} \parallel \sqrt{\Gamma} C_1 \sqrt{\Gamma}) \leq D_{\max}(C_0 \parallel C_1), \quad \forall \Gamma \in \bar{\mathbf{S}}_+.$$

The bound holds with the equality if the dual convex set $\bar{\mathbf{S}}_+$ contains a full-rank operator.

The proof can be found in appendix C. Note that, by construction the operators $\sqrt{\Gamma} C_i \sqrt{\Gamma}$, $i = 0, 1$ are density matrices: indeed, they are positive and $\text{Tr}[\sqrt{\Gamma} C_i \sqrt{\Gamma}] = \text{Tr}[\Gamma C_i] = 1$, since, by definition Γ is a positive function normalized on the primal set S . Proposition will be used to show that the relative entropy of two quantum networks is equal to the maximum relative entropy between the output states generated by the networks.

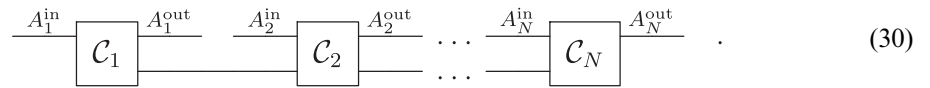
4. Optimizing quantum causal networks

Here we consider the scenario where a network of quantum devices, arranged in a definite causal order, is required to perform a desired task, such as implementing a distributed algorithm. What is the maximum performance that the network can attain? In this section we answer the question, measuring the performance through the score obtained in a suitable test (depending on the task at hand) and providing a close form expression for the maximum score.

4.1. The dual networks

Following section 2.5, the mathematical description of the test is provided by a performance operator Ω , acting on the Hilbert spaces of the input and output systems of the tested network. The maximum performance achieved by an arbitrary causal network is determined by the following

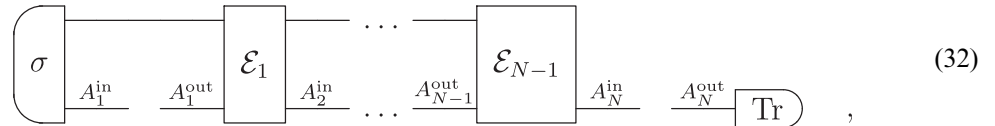
Theorem 2. Let Ω be an operator on $\bigotimes_{j=1}^N (\mathcal{H}_j^{\text{out}} \otimes \mathcal{H}_j^{\text{in}})$ and let ω_{max} be the maximum of $\langle \Omega, C \rangle$ over all operators C representing quantum networks of the form



Then, ω_{max} is given by

$$\omega_{\text{max}} = \min_{\Gamma \in \text{DualComb}} \min \{ \lambda \in \mathbb{R} \mid \lambda \Gamma \geq \Omega \}, \tag{31}$$

where DualComb denotes the set of dual combs, that is, positive operators Γ representing networks of the form



where σ is a quantum state, $(\mathcal{E}_1, \mathcal{E}_2, \dots, \mathcal{E}_{N-1})$ is a sequence of quantum channels, and $\text{Tr}_{A_N^{\text{out}}}$ represents the trace over the last system. Explicitly, DualComb is the set of all positive operators Γ satisfying the linear constraint

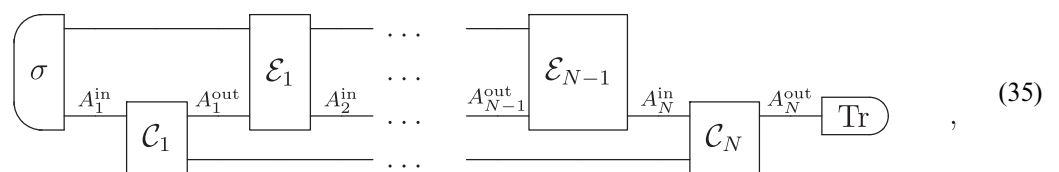
$$\begin{aligned} \Gamma &= I_{A_N^{\text{out}}} \otimes \Gamma^{(N)} \\ \text{Tr}_{A_n^{\text{in}}}[\Gamma^{(N)}] &= I_{A_{n-1}^{\text{out}}} \otimes \Gamma^{(N-1)}, \quad n = 2, \dots, N \\ \text{Tr}_{A_1^{\text{in}}}[\Gamma^{(1)}] &= 1, \end{aligned} \tag{33}$$

for suitable positive operators $\Gamma^{(n)}$ acting on $\mathcal{H}_n^{\text{in}} \otimes [\bigotimes_{j=1}^{n-1} (\mathcal{H}_j^{\text{out}} \otimes \mathcal{H}_j^{\text{in}})]$. When Ω is positive, the maximum performance can be expressed as

$$\omega_{\text{max}} = 2^{D_{\text{max}}(\Omega \parallel \text{DualComb})}. \tag{34}$$

The proof can be found in appendix B.

Theorem 2 has an intuitive interpretation. The dual networks (32) and the primal networks (30) ‘deterministically complement each other’: when two such networks are connected, one obtains the closed circuit



which yields no information about the primal network and makes any such information inaccessible to further tests. Hence, the dual networks represent the non-informative tests. The max relative entropy quantifies how much the test with performance operator Ω deviates from the set of non-informative tests.

4.2. The case of binary testers

Consider a binary test, described by the tester $\{T_{\text{yes}}, T_{\text{no}}\}$ and assume that the test is passed if and only if the testing network yields the outcome ‘yes’. Binary testers have applications in the theory of quantum interactive proof systems [46], where they can be used to compute the probability that the verifier accepts the token provided by the prover through a sequence of operations. In this scenario, the performance operator is given by $\Omega = T_{\text{yes}}$ and the probability that the prover passes the test, optimized over all possible quantum strategies, is

$$p_{\max} = \left(\max_{\Gamma \in \text{DualComb}} \max \{w T_{\text{yes}} \leq \Gamma\} \right)^{-1}, \quad (36)$$

having used equation (31) with λ replaced by its inverse $w = 1/\lambda$. In words, the problem is to find the maximum weight for which one can squeeze the tester operator T_{yes} under some dual comb Γ .

This maximization has an intuitive interpretation:

Corollary 1. *The maximum probability that a quantum causal network passes the test defined by the operator T_1 is equal to the inverse of the maximum weight w for which there exists a two-outcome tester $\{T'_{\text{yes}}, T'_{\text{no}}\}$ satisfying $T'_{\text{yes}} = w T_1$.*

Proof. Suppose that the relation $wT_1 \leq \Gamma$ holds for some weight w and some dual comb Γ . Then, define $T'_{\text{yes}} := wT_{\text{yes}}$ and $T'_{\text{no}} := \Gamma - T'_{\text{yes}}$. By construction, the operators $\{T'_{\text{yes}}, T'_{\text{no}}\}$ form a tester: they are positive and their sum satisfies equation (17). \square

In other words, the dual problem amounts to finding the binary tester $\{T_{\text{yes}}^*, T_{\text{no}}^*\}$ that assigns the maximum possible probability to the outcome 1, subject to the condition that T_{yes}^* is proportional to T_{yes} . The content of the duality is that the maximum is attained when there exists a primal network that triggers deterministically the outcome 1:

Corollary 2. *Let $\{T_{\text{yes}}^*, T_{\text{no}}^*\}$ be the optimal tester for the dual problem and let C^* be the optimal quantum comb for the primal problem. Then, one has*

$$\langle T_{\text{yes}}^*, C^* \rangle = 1.$$

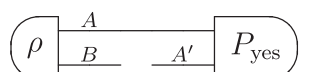
Proof. Let w^* be the optimal weight in the dual problem, Then, one has $T_{\text{yes}}^* = w^* T_{\text{yes}}$ and $\langle T_{\text{yes}}^*, C^* \rangle = 1/w^*$. Combining these two equations, one gets $\langle T_{\text{yes}}^*, C^* \rangle = w^* \langle T_{\text{yes}}, C^* \rangle = 1$. \square

5. The conditional min-entropy of quantum causal networks

Theorem 2 allows us to extend the notion of conditional min-entropy [50] from quantum states to quantum causal networks. Let us first review the basic properties of the conditional min-entropy of quantum states: For a quantum state $\rho \in \text{St}(AB)$, the conditional min-entropy of system A , conditional on system B , is defined as [50]

$$H_{\min}(A|B)_\rho := -\log \left[\min_{\gamma \in \text{St}(B)} \min \{ \lambda \in \mathbb{R} \mid \lambda(I_A \otimes \gamma_B) \geq \rho_{AB} \} \right]. \quad (37)$$

König *et al* [51] clarified the operational meaning of $H_{\min}(A|B)_\rho$ in terms of the following task: given the state ρ_{AB} , find the quantum channel \mathcal{C} that produces the best approximation of the maximally entangled state $|\Phi\rangle_{AA'} := \sum_{n=1}^{d_A} |n\rangle |n\rangle / \sqrt{d_A}$, by acting locally on system B . Here the quality of the approximation is measured by the fidelity, namely the probability that the output state passes a binary test with POVM $\{P_{\text{yes}}, P_{\text{no}}\}$, defined by $P_{\text{yes}} := |\Phi\rangle\langle\Phi|$. Overall, we can jointly regard the preparation of the state ρ and the measurement of the binary POVM $\{P_{\text{yes}}, P_{\text{no}}\}$ as a test performed on the channel \mathcal{C} . Diagrammatically, the successful instance of the test is represented by the network



$$, \quad (38)$$

whose Choi operator is given by

$$\begin{aligned} T_{\text{yes}} &:= \rho * P_{\text{yes}}^T \\ &= \rho/d_A, \end{aligned}$$

(with a slight abuse of notation, in the second equality we regard ρ as an operator on $A'B$, instead of AB). Hence, the probability that the channel passes the test is

$$p = T_{\text{yes}} * C = \frac{\text{Tr}[\rho C^T]}{d_A},$$

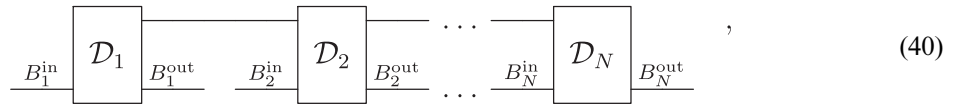
where C is the Choi operator of \mathcal{C} . König, Renner, and Schaffner showed that the maximum probability over all possible channels is

$$p_{\text{max}} = \frac{2^{-H_{\min}(A|B)_\rho}}{d_A}. \tag{39}$$

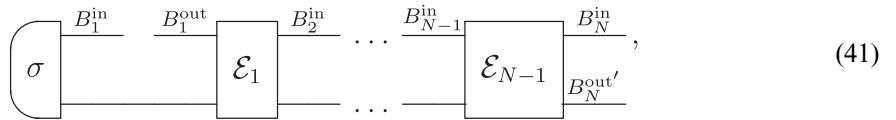
We now extend the notion of conditional min-entropy from states to networks with a definite causal structure. This can be done in two slightly different ways, illustrated in the following subsections.

5.1. The conditional min-entropy of a quantum causal network

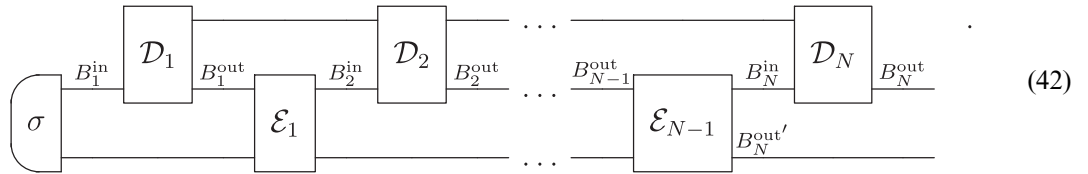
The first way to generalize the conditional min-entropy from states is to regard $H_{\min}(A|B)_\rho$ as a measure of the correlations that can be extracted from the state ρ_{AB} by acting on system B alone. A natural generalization to the network scenario arises if we consider a quantum network of the form



and ask how much correlation can be generated by interacting with the network in the first $N - 1$ time steps. To generate the correlations, we can connect the network (40) with a second network that processes all input/output systems before B_N^{out} . Graphically, the second network can be described as



where $B_N^{\text{out}'}$ is a quantum system of the same dimension as B_N^{out} . When the two networks are connected, they generate the bipartite state



A measure of the correlations generated by the interaction of the two networks is then provided by the fidelity between the state (42) and the maximally entangled state.

Explicitly, the fidelity is given by

$$F := \langle \Phi | (\sigma * D_1 * E_1 * \dots * D_{N-1} * E_{N-1} * D_N) | \Phi \rangle = \frac{\text{Tr}[D E^T]}{d_{B_N^{\text{out}}}}, \tag{43}$$

with

$$D := D_1 * \dots * D_N \quad \text{and} \quad E := \sigma * E_1 * \dots * E_{N-1}$$

(with a little abuse of notation, in the second equality we regard E as an operator on $\mathcal{H}_{B_1^{\text{in}}} \otimes \mathcal{H}_{B_1^{\text{out}}} \otimes \dots \otimes \mathcal{H}_{B_N^{\text{in}}} \otimes \mathcal{H}_{B_N^{\text{out}'}}$ instead of $\mathcal{H}_{B_1^{\text{in}}} \otimes \mathcal{H}_{B_1^{\text{out}}} \otimes \dots \otimes \mathcal{H}_{B_N^{\text{in}}} \otimes \mathcal{H}_{B_N^{\text{out}'}}$).

The maximum of the fidelity over all networks of the form (41) can be computed via theorem 2, which yields the expression

$$F_{\text{max}} = \frac{\min_{\Gamma_{t_1 \dots t_{N-1}}} \min \{ \lambda \in \mathbb{R} \mid \lambda (I_{t_N} \otimes \Gamma_{t_1 \dots t_{N-1}}) \geq R \}}{d_{B_N^{\text{out}}}}, \tag{44}$$

where $\Gamma_{t_1 \dots t_{N-1}}$ is a generic element of $\text{Comb}(B_1^{\text{in}} \rightarrow B_1^{\text{out}}, \dots, B_{N-1}^{\text{in}} \rightarrow B_{N-1}^{\text{out}'})$ and $I_{t_N} := I_{B_N^{\text{out}}} \otimes I_{B_N^{\text{in}}}$.

Equation (44) motivates the following.

Definition 4. Let $D \in \text{Comb}(B_1^{\text{in}} \rightarrow B_1^{\text{out}}, \dots, B_N^{\text{in}} \rightarrow B_N^{\text{out}'})$ be a quantum comb and let $t_j := B_j^{\text{in}} \rightarrow B_j^{\text{out}'}$ be the type corresponding to the j th time step. The network min-entropy of the N th time step, conditionally on the first $N - 1$ time steps is the quantity

$$H_{\min}(t_N | t_1 \cdots t_{N-1})_D := -\log \left[\min_{\Gamma_{t_1 \dots t_{N-1}}} \min \{ \lambda \in \mathbb{R} \mid \lambda(I_{t_N} \otimes \Gamma_{t_1 \dots t_{N-1}}) \geq D \} \right], \tag{45}$$

where the first minimum is over the elements of $\text{Comb}(B_1^{\text{in}} \rightarrow B_1^{\text{out}}, \dots, B_{N-1}^{\text{in}} \rightarrow B_{N-1}^{\text{out}})$ and $I_{t_N} := I_{B_N^{\text{out}}} \otimes I_{B_N^{\text{in}}}$.

The above definition is a compelling generalization of the conditional min-entropy for states. First of all, it comes with a natural operational interpretation, as the maximum amount of correlations between the last output of the network and all the system involved in the previous history. Moreover, the conditional min-entropy of quantum networks is consistent with the conditional min-entropy of quantum states: Concretely, one can interpret the conditional min-entropy (45) as the maximum conditional min-entropy of the output state of the network, conditionally on an external reference system generated through the intermediate time steps. This interpretation is based on the following.

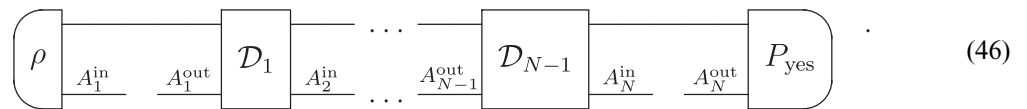
Proposition 6. For a causal network with Choi operator $D \in \text{Comb}(B_1^{\text{in}} \rightarrow B_1^{\text{out}}, \dots, B_N^{\text{in}} \rightarrow B_N^{\text{out}})$, the min-entropy $H_{\min}(t_N | t_1 \cdots t_{N-1})_D$ is equal to the min-entropy of the output state $\rho \in \text{St}(\mathcal{H}_{B_N^{\text{out}}} \otimes \mathcal{H}_{B_N^{\text{in}}})$ in equation (42) maximized over the input state σ and over the sequence of intermediate operations $\mathcal{E}_1, \dots, \mathcal{E}_{N-1}$.

The proof is given in appendix D. We expect that the network min-entropy defined in equation (45) will play a role in the study non-Markovian quantum evolutions, along the lines of the entropic characterization of Markovianity provided in [89, 90]. Intuitively, the idea is that one can evaluate how the correlations build up from one step to the next and use this information to infer properties of the internal memory used by the network.

5.2. The conditional min-entropy of a test

An alternative way to extend the notion of conditional min-entropy is to regard $H_{\min}(A|B)_\rho$ as a quantity associated to a test—specifically, the test depicted in equation (38). From this point of view, it is natural to extend the definition to tests consisting of multiple time steps, as follows

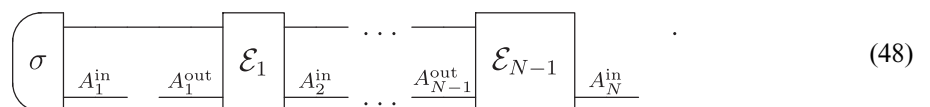
Definition 5. Let T_{yes} be a positive operator associated to a test of the form



The conditional min-entropy of the output system A_N^{out} , conditionally on all the previous systems is

$$H_{\min}(A_N^{\text{out}} | A_1^{\text{in}} A_1^{\text{out}} A_2^{\text{in}} \dots A_N^{\text{in}})_{T_{\text{yes}}} := -\log \left[\min_{\Gamma^{(N)}} \min \{ \lambda \in \mathbb{R} \mid \lambda(I_{A_N^{\text{out}}} \otimes \Gamma^{(N)}) \geq T_{\text{yes}} \} \right], \tag{47}$$

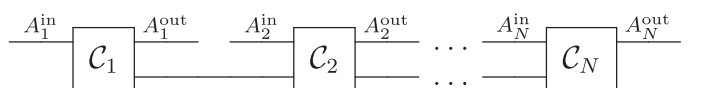
where $\Gamma^{(N)}$ is a generic element of $\text{Comb}(I \rightarrow A_1^{\text{in}}, A_1^{\text{out}} \rightarrow A_2^{\text{in}} \dots, A_{N-1}^{\text{out}} \rightarrow A_N^{\text{in}})$, corresponding to a network of the form



The conditional min-entropy for ‘states’ (or, more precisely, for tests of the form (38)) can be retrieved as a special case of this definition, by setting $N = 1$, $A_1^{\text{in}} = B$, $A_1^{\text{out}} = A$, and $T_{\text{yes}} = \rho/d_A$. The appeal of the above definition is that it extends the definition of min-entropy to a class of probabilistic operations.

The conditional min-entropy can be interpreted operationally as the (negative logarithm of the) maximum probability that a quantum causal network passes the test T_{yes} . This interpretation follows from theorem 2, which yields the following.

Corollary 3. The maximum probability that a quantum causal network of the form



passes the test with operator T_{yes} is

$$p_{\max} = 2^{-H_{\min}(A_N^{\text{out}} | A_1^{\text{in}} A_1^{\text{out}} A_2^{\text{in}} \dots A_N^{\text{in}})_{T_{\text{yes}}}}$$

This result secures an operational interpretation for the conditional min-entropy defined in equation (47). Quite

intuitively, the conditional min-entropy of the test is a measure of how of the first $N - 1$ time steps can be used to predict the outcome of the measurement performed in the last step.

6. The max relative entropy of causal networks

We conclude our study of causal networks with a result relating the max relative entropy of quantum networks to the max relative entropy of quantum states:

Proposition 7. Let $C^{(0)}$ and $C^{(1)}$ be the Choi operators of two networks

$$\begin{array}{c}
 A_1^{\text{in}} \quad A_1^{\text{out}} \quad A_2^{\text{in}} \quad A_2^{\text{out}} \quad \dots \quad A_N^{\text{in}} \quad A_N^{\text{out}} \\
 \boxed{C_1^{(x)}} \quad \boxed{C_2^{(x)}} \quad \dots \quad \boxed{C_N^{(x)}}
 \end{array}, \quad x = 0, 1, \quad (49)$$

and let E be the Choi operator of a network of the form

$$\begin{array}{c}
 \sigma \quad A_1^{\text{in}} \quad A_1^{\text{out}} \quad \mathcal{E}_1 \quad A_2^{\text{in}} \quad \dots \quad A_{N-1}^{\text{in}} \quad \mathcal{E}_{N-1} \quad S \\
 \dots \quad \dots \quad \dots \quad \dots \quad \dots \quad \dots \quad \dots \quad \dots \quad \dots
 \end{array}, \quad (50)$$

where S is a generic quantum system. Then, one has

$$D_{\max}(C^{(0)} \parallel C^{(1)}) = \max_E D_{\max}(C^{(0)} * E \parallel C^{(1)} * E). \quad (51)$$

where the maximum runs over all networks of the form (50), with arbitrary system S .

The proof is provided in appendix E. In words, equation (70) states that the max relative entropy between two quantum networks is equal to the max relative entropy between the output states one can generate from them. Diagrammatically, the output states are

$$\begin{array}{c}
 \sigma \quad A_1^{\text{in}} \quad A_1^{\text{out}} \quad \mathcal{E}_1 \quad A_2^{\text{in}} \quad \dots \quad A_{N-1}^{\text{out}} \quad \mathcal{E}_{N-1} \quad S \\
 \boxed{C_1^{(x)}} \quad \dots \quad \boxed{C_N^{(x)}}
 \end{array} \quad x = 0, 1. \quad (52)$$

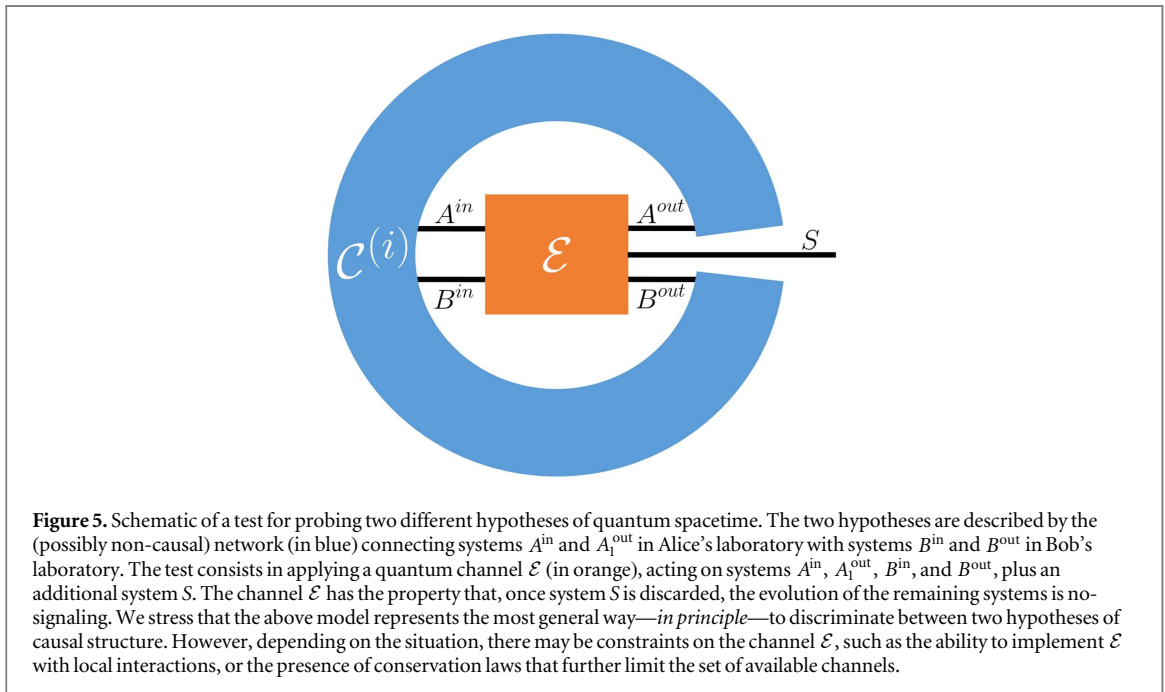
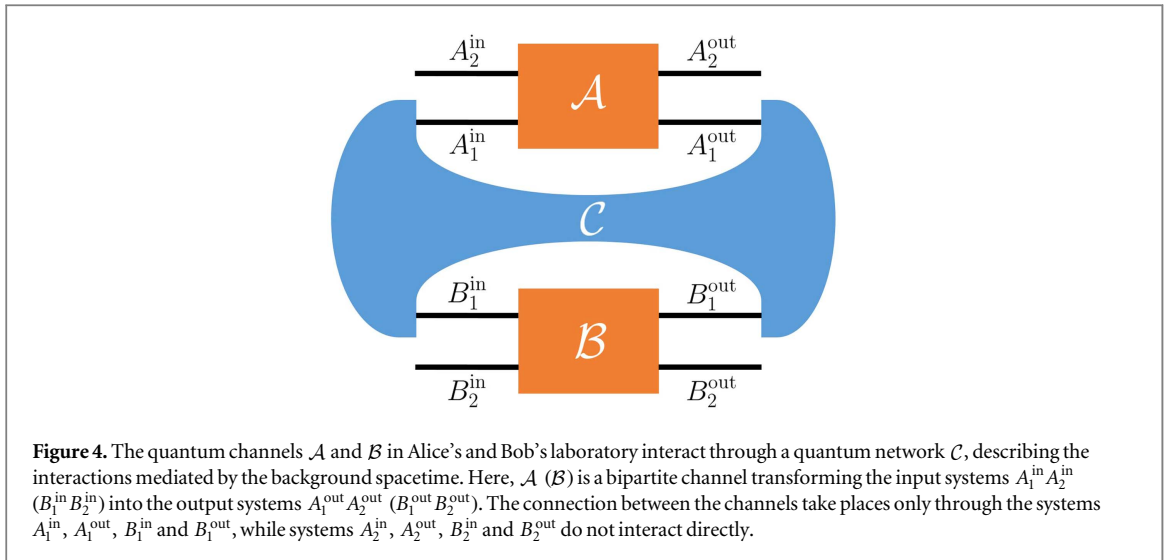
Proposition 7 has an application to problems of hypothesis testing where the task is to distinguish between two quantum networks. Here one has access to an quantum network, that is promised to have quantum comb $C^{(0)}$ or $C^{(1)}$. In order to determine which of these two hypotheses is correct, one has to interact with the network, by sending inputs to it and processing its outputs. In the end, these operations will result in the preparation of a quantum state, as in equation (52). At this point, the problem is to distinguish between two states $\rho^{(0)}$ and $\rho^{(1)}$ corresponding to the two hypotheses. One-shot hypothesis testing of quantum states has been studied by Datta *et al* in [86], where they provided bounds on the type II error probability in terms of the max relative entropy. Proposition 7 then allows to relate the max relative entropy of the output states to the max relative entropy of the networks, opening a route to adapting the results of [86] to the study of hypothesis testing to the more general scenario.

7. Non-causal networks

In the previous sections we restricted our focus to causal networks. We will address the general scenario, concerning networks that are not compatible with any pre-defined causal order [33, 34, 37–42, 44]. Some of these networks arise when multiple quantum devices are connected in a way that is controlled by the state of a quantum system [33, 38]. Some other networks are not built from individual devices [34, 41] but may possibly arise in exotic quantum gravity scenarios. These generalized quantum networks are characterized by the way in which they interact with external quantum devices.

7.1. A bipartite example

The characterization of the non-causal networks is not as simple as in the case of causally ordered networks. We first illustrate the idea in a simple example, inspired by the work of Oreshkov *et al* [41]. Imagine two laboratories, A and B , where two parties, Alice and Bob perform local experiments. In each laboratory, ordinary quantum theory holds and, in particular, one can describe the time evolution by quantum channels. Specifically, let \mathcal{A} and \mathcal{B} be the quantum channels describing the evolution of the systems in laboratories A and B , respectively. Now,



one can model the interactions between one laboratory and the other by a generalized quantum network, which describes the background structure of spacetime.

Concretely, suppose that, at some earlier time, system A_1^{in} in the first laboratory has been prepared jointly with system B_1^{in} in the second laboratory, and that, at a later time, system A_1^{out} and system B_1^{out} are discarded. Indulging into a bit of science fiction, one could imagine a scenario where systems A_1^{in} and B_1^{in} emerge from a wormhole at time t_0 and system A_1^{out} and B_1^{out} enter the same wormhole at time t_1 . Between times t_0 and t_1 the systems A_1^{in} and B_1^{in} can interact with the other systems in Alice's and Bob's laboratories, here denoted as $A_2^{\text{in}}, A_2^{\text{out}}$ and $B_2^{\text{in}}, B_2^{\text{out}}$, respectively. The interaction is controlled locally by Alice and Bob, who implement the channels \mathcal{A} and \mathcal{B} , as illustrated in figure 4. The connection of Alice's and Bob's laboratories through the background spacetime structure can be described as a map

$$\mathcal{S}: \mathcal{A} \otimes \mathcal{B} \mapsto \mathcal{S}(\mathcal{A} \otimes \mathcal{B}), \tag{53}$$

which transforms the quantum channels \mathcal{A} and \mathcal{B} into a new quantum channel $\mathcal{S}(\mathcal{A} \otimes \mathcal{B})$. Maps that transform channels into channels are known as *quantum supermaps* [18, 53]. The basic requirements for quantum supermaps are linearity, complete positivity, and normalization. In this setting, linearity means that one has

$$\mathcal{S}\left(\sum_i p_i \mathcal{A}_i \otimes \mathcal{B}_i\right) = \sum_i p_i \mathcal{S}(\mathcal{A}_i \otimes \mathcal{B}_i), \quad (54)$$

for every choice of coefficients $\{p_i\}$. The standard motivation for linearity comes from the requirement that convex combinations of input channels (generated by Alice and Bob by sharing random bits) be mapped into convex combinations of the corresponding outputs.

Regarding complete positivity, it can be motivated by the local form of the interactions. Since the interaction between Alice's and Bob's laboratory takes place only through systems A_1 and B_1 , it is natural to assume that the supermap \mathcal{S} acts non-trivially only on these systems, as

$$\mathcal{S}(\mathcal{A} \otimes \mathcal{B}) = (\mathcal{I}_{A_2 \rightarrow A_2} \otimes \mathcal{C} \otimes \mathcal{I}_{B_2 \rightarrow B_2})(\mathcal{A} \otimes \mathcal{B}), \quad (55)$$

where $\mathcal{I}_{A_2 \rightarrow A_2}$ ($\mathcal{I}_{B_2 \rightarrow B_2}$) is the *identity supermap*, acting trivially on the channels with input A_2^{in} (B_2^{in}) and output A_2^{out} (B_2^{out}), and \mathcal{C} is a supermap that annihilates channels with input A_1^{in} B_1^{in} and output A_1^{out} B_1^{out} . Physically, the map \mathcal{C} represents the piece of spacetime connecting A_1^{in} and B_1^{in} with A_1^{out} and B_1^{out} .

7.2. Choi operator formulation

Since all the maps \mathcal{A} , \mathcal{B} , and \mathcal{C} are completely positive, one can represent them with Choi operators A , B , and C , respectively. In terms of Choi operators, equation (55) can be expressed as

$$C' = \text{Tr}_{A_1^{\text{in}}, A_1^{\text{out}}, B_1^{\text{in}}, B_1^{\text{out}}}[I_{A_2^{\text{out}}} \otimes I_{A_2^{\text{in}}} \otimes C \otimes I_{B_2^{\text{out}}} \otimes I_{B_2^{\text{in}}}(A^T \otimes B^T)], \quad (56)$$

where C' the Choi operator of the channel $\mathcal{S}(\mathcal{A} \otimes \mathcal{B})$. Oreshkov *et al* refer to the Choi operator C as a *process matrix* [41]. For the supermaps that can be implemented by connecting quantum devices in a fixed causal structure, Choi operators C are the same as the quantum combs considered in the previous sections.

Here the operator C acts on the tensor product Hilbert space $\mathcal{H}_{A_1^{\text{in}}} \otimes \mathcal{H}_{A_1^{\text{out}}} \otimes \mathcal{H}_{B_1^{\text{in}}} \otimes \mathcal{H}_{B_1^{\text{out}}}$. In order to be the Choi operator of a valid quantum network, the operator C must be positive semidefinite and satisfy a suitable normalization condition—specifically, C should satisfy the condition

$$\text{Tr}_{A_1^{\text{in}}, A_1^{\text{out}}, B_1^{\text{in}}, B_1^{\text{out}}}[C(\tilde{A} \otimes \tilde{B})] = 1 \quad (57)$$

for every operators \tilde{A} and \tilde{B} satisfying the conditions

$$\text{Tr}_{A_1^{\text{out}}}[\tilde{A}] = I_{A_1^{\text{in}}} \quad \text{and} \quad \text{Tr}_{B_1^{\text{out}}}[\tilde{B}] = I_{B_1^{\text{in}}}. \quad (58)$$

(See appendix F for the derivation). Physically, this means that the non-causal network \mathcal{C} deterministically annihilates every pair of local channels \tilde{A} and \tilde{B} , acting on systems A_1^{in} , A_1^{out} and B_1^{in} , B_1^{out} , respectively.

Equivalently, the valid networks can be characterized as in the following:

Proposition 8. *An operator C is the Choi operator of a non-causal network as in figure 4 if and only if C is positive and $\text{Tr}[CD] = 1$ for every operator D satisfying the conditions*

$$\text{Tr}_{A_1^{\text{out}}}[D] = I_{A_1^{\text{in}}} \otimes \tilde{B}, \quad \text{Tr}_{B_1^{\text{out}}}[\tilde{B}] = I_{B_1^{\text{in}}} \quad (59)$$

and

$$\text{Tr}_{B_1^{\text{out}}}[D] = I_{B_1^{\text{in}}} \otimes \tilde{A}, \quad \text{Tr}_{A_1^{\text{out}}}[\tilde{A}] = I_{A_1^{\text{in}}}, \quad (60)$$

with suitable operators \tilde{A} and \tilde{B} .

For the proof, see theorem 2 of [34]. The operator D represents the Choi operator of a no-signaling channel [91–93], that is, a channel that prevents the transmission of information from Alice to Bob and from Bob to Alice. The intuitive idea is that whenever a network can be connected with two local channels, it can also be connected with a no-signaling channel.

In the following we will denote by $\text{NoSig}(A_1^{\text{in}} \rightarrow A_1^{\text{out}} | B_1^{\text{in}} \rightarrow B_1^{\text{out}})$ the set of positive operators satisfying the no-signaling conditions (59) and (60). With this notation, proposition 8 can be reformulated as

Corollary 4. *An operator C is the Choi operator of a non-causal network as in figure 4 if and only if*

$$C \geq 0 \quad (61)$$

$$C \in \overline{\text{NoSig}(A_1^{\text{in}} \rightarrow A_1^{\text{out}} | B_1^{\text{in}} \rightarrow B_1^{\text{out}})}, \quad (62)$$

where $\overline{\text{NoSig}(A_1^{\text{in}} \rightarrow A_1^{\text{out}} | B_1^{\text{in}} \rightarrow B_1^{\text{out}})}$ is the dual affine space of the set of no-signalling channels.

We will denote by $\text{DualNoSig}(A_1^{\text{in}} \rightarrow A_1^{\text{out}} | B_1^{\text{in}} \rightarrow B_1^{\text{out}})$ the set of operators satisfying conditions (61) and (62). The set DualNoSig is the set containing all the Choi operators of the non-causal networks of actin on pairs of local operators.

7.3. The max relative entropy of signaling

In some situations, such as the study of non-causal games [41], it is natural to search for the non-causal networks that maximize a certain figure of merit. For example, consider an experiment where Alice and Bob probe a non-causal network as in figure 4. In their local laboratories, Alice and Bob measure the output systems of the network with the POVMs $\{P_i\}_{i=1}^K$ and $\{Q_j\}_{j=1}^L$, respectively, and prepare inputs for the systems, say ρ and σ , respectively. The outcomes i and j are assigned a score $\omega(i, j)$, which quantifies the performance of the non-causal network. For example, Alice and Bob may want to quantify how much the network correlates their outcomes, corresponding to the score $\omega(i, j) = \delta_{ij}$. More generally, Alice and Bob can probe the network by preparing correlated states, applying local interactions, and performing local measurements.

Describing the test with a performance operator Ω , the maximum score is achievable by quantum non-causal networks is

$$\omega_{\max} = \max_{C \in \text{DualNoSig}(A_1^{\text{in}} \rightarrow A_1^{\text{out}} | B_1^{\text{in}} \rightarrow B_1^{\text{out}})} \langle \Omega, C \rangle. \quad (63)$$

Finding the network that achieves maximum score is similar to finding the entangled state that maximizes the violation of a Bell inequality. The optimization task can be tackled with our theorem 1, which provides a dual expression for the maximum score:

Proposition 9. *Let $\Omega \in \text{Herm}(\mathcal{H}_{A_1^{\text{out}}} \otimes \mathcal{H}_{A_1^{\text{in}}} \otimes \mathcal{H}_{B_1^{\text{out}}} \otimes \mathcal{H}_{B_1^{\text{in}}})$ be a generic performance operator and ω_{\max} be the maximum score defined in equation (63). Then, one has*

$$\omega_{\max} = \min_{\Gamma \in \text{NoSig}(A_1^{\text{in}} \rightarrow A_1^{\text{out}} | B_1^{\text{in}} \rightarrow B_1^{\text{out}})} \min \{ \lambda \in \mathbb{R} \mid \lambda \Gamma \geq \Omega \}.$$

When Ω is positive, the maximum score is given by

$$\omega_{\max} = 2^{D_{\max}(\Omega \parallel \text{NoSig}(A_1^{\text{in}} \rightarrow A_1^{\text{out}} | B_1^{\text{in}} \rightarrow B_1^{\text{out}}))}. \quad (64)$$

In words: the maximum score achieved by quantum non-causal networks is determined by the deviation of the performance operator from set of (Choi operators of) no-signaling channels. We call $D_{\max}(A \parallel \text{NoSig}(A_1^{\text{in}}, A_1^{\text{out}} | B_1^{\text{in}}, B_1^{\text{out}}))$ the *max relative entropy of signaling*, in analogy with the *relative entropy of entanglement* of a state ρ [94–96].

7.4. Optimizing multipartite non-causal networks

The results presented in the bipartite case can be easily generalized to multipartite non-causal networks.

Consider a quantum network that can interact with k local devices, by providing an input system to each device and annihilating its output system. As in the bipartite case, the network can be represented by its Choi operator C , which will have to satisfy the condition

$$\text{Tr}[C(\tilde{A}_1 \otimes \tilde{A}_2 \otimes \cdots \otimes \tilde{A}_k)] = 1,$$

for every set of Choi operators $(\tilde{A}_1, \tilde{A}_2, \dots, \tilde{A}_k)$ representing local quantum channels. Equivalently, the normalization condition can be expressed as

$$\text{Tr}[C D] = 1,$$

for every Choi operator D representing a k -partite no-signaling channel. Specifically, the set of Choi operators representing k -partite no-signaling channels is defined as follows:

Definition 6. An operator D , acting on $\bigotimes_{i=1}^k (\mathcal{H}_{A_i^{\text{out}}} \otimes \mathcal{H}_{A_i^{\text{in}}})$, is the Choi operator of a no-signaling channel iff for every subset $J \subseteq \{1, \dots, k\}$ one has

$$\text{Tr}_{A_J^{\text{out}}}[D] = I_{A_J^{\text{in}}} \otimes D_J^c,$$

where $\text{Tr}_{A_J^{\text{out}}}$ is the partial trace over the Hilbert space $\mathcal{H}_{A_J^{\text{out}}} := \bigotimes_{i \in J} \mathcal{H}_{A_i^{\text{out}}}$, $I_{A_J^{\text{in}}}$ is the identity operator on the Hilbert space $\mathcal{H}_{A_J^{\text{in}}} := \bigotimes_{i \in J} \mathcal{H}_{A_i^{\text{in}}}$, and D_J^c is the Choi operator of a quantum channel transforming density matrices on $\mathcal{H}_{A_J^{\text{in}}} := \bigotimes_{i \in J} \mathcal{H}_{A_i^{\text{in}}}$ into density matrices on $\mathcal{H}_{A_J^{\text{out}}} := \bigotimes_{i \in J} \mathcal{H}_{A_i^{\text{out}}}$.

We denote the set of k -partite no-signaling channels as NoSig_k , keeping implicit the specification of the Hilbert spaces.

Like in the bipartite case, it is natural to consider tasks where one has to find the non-causal network that maximizes a score of the form $\omega = \text{Tr}[\Omega C]$ for some performance operator Ω . The maximum score is then given by

$$\omega_{\max} = \max \{ \text{Tr}[\Omega C] \mid C \in \overline{\text{NoSig}_k} \}. \quad (65)$$

In general, characterizing the dual affine space of the set of no signaling channels is a rather laborious task. Using theorem 1 we can circumvent the problem and express the maximum score as

$$\omega_{\max} = \min_{D \in \text{NoSig}_k} \min \{ \lambda \in \mathbb{R} \mid \lambda D \geq \Omega \}, \tag{66}$$

or, when Ω is positive

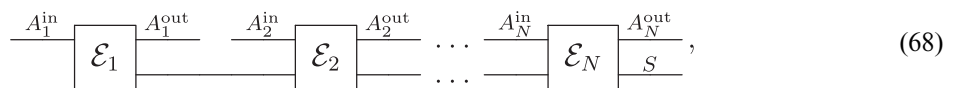
$$\omega_{\max} = 2^{D_{\max}(\Omega \parallel \text{NoSig}_k)}. \tag{67}$$

Again, the performance is determined by the deviation of the performance operator from the set of (Choi operators of) no-signaling channels.

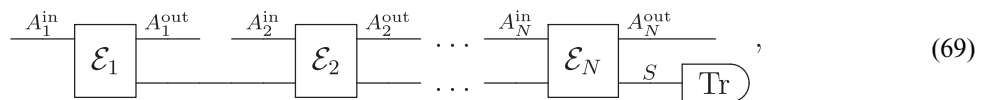
8. The max relative entropy of non-causal networks

Like in the case of causal networks, the max relative entropy between two quantum networks can be related to the max relative entropy of their output states:

Proposition 10. *Let $C^{(0)}$ and $C^{(1)}$ be the Choi operators of two non-causal networks in $\overline{\text{NoSig}}_k$ and let E be the Choi operator of a network of the form*



where S is a generic quantum system and the reduced channel



is no-signaling. Then, one has

$$D_{\max}(C^{(0)} \parallel C^{(1)}) = \max_E D_{\max}(C^{(0)} * E \parallel C^{(1)} * E), \tag{70}$$

where the maximum runs over all networks of the form (69), with arbitrary system S .

The proof is the same as the proof of proposition 7. The above result shows that the max relative entropy between two non-causal networks is equal to the max relative entropy between the output states generated by connecting the networks to the ‘no-signaling part’ of a quantum channel, as in figure 5.

Like in the causal case, there is a nice connection to one-shot hypothesis testing. Here one can consider the problem of distinguishing between two alternative models of spacetime, resulting into different ways to connect the operations performed in N local laboratories. For example, $C^{(0)}$ could describe a null hypothesis of space time where all the events are causally ordered, while $C^{(1)}$ could describe an exotic, non-causal space time. Proposition 10 tells us that, in terms of max relative entropy, the distinguishability of two models of spacetime is quantified by the max relative entropy of the corresponding non-causal networks.

9. Applications

In the following we apply our results to four optimization problems involving quantum networks. We will start from the causal case, considering networks that approximately transform a given set of input channels into a target set of output channels. Then, we will move the case of non-causal networks.

9.1. Transforming quantum channels

Consider the following scenario: A black box implements a quantum channel in the set $\{\mathcal{E}_x\}_{x \in X}$, where X is an arbitrary index set. The task is to simulate another channel \mathcal{F}_x using the channel \mathcal{E}_x as a subroutine. For example, the black box could implement a unitary gate U_x and the task could be to build the control-unitary gate [97–100].

$$\text{ctrl} - U_x = I \otimes |0\rangle\langle 0| + U_x \otimes |1\rangle\langle 1|.$$

To simulate the desired channel \mathcal{F}_x , we insert the input channel \mathcal{E}_x into a quantum causal network, as in the following diagram

$$\begin{array}{c}
 A_0 \quad A_1 \quad A_2 \quad A_3 \\
 \boxed{C_1} \quad \boxed{\mathcal{E}_x} \quad \boxed{C_2} \\
 \hline
 A_0 \quad A_3 \\
 \boxed{\mathcal{E}'_x}
 \end{array} = \quad , \tag{71}$$

where C_1 and C_2 are suitable quantum channels. The Choi operator of the output channel \mathcal{E}'_x is then given by

$$E'_x = C * E_x, \tag{72}$$

where C is the Choi operator of the network and $*$ denotes the link product.

Let us focus on the case where the target channel \mathcal{F}_x is an isometry, namely $\mathcal{F}_x = V_x \cdot V_x^\dagger$, with $V_x^\dagger V_x = I$. To measure how close the channel \mathcal{E}'_x is to the target, we use the *channel fidelity* [101–103], given by

$$F(\mathcal{E}'_x, \mathcal{F}_x) := \frac{1}{d_0^2} \langle\langle V_x | E'_x | V_x \rangle\rangle, \tag{73}$$

where d_0 is the dimension of the input system A_0 and the notation $|V\rangle\rangle$ denotes the unnormalized state

$$|V\rangle\rangle := (V \otimes I) |I\rangle\rangle, \quad |I\rangle\rangle := \sum_{n=1}^d |n\rangle |n\rangle.$$

In this case, the fidelity can be interpreted as the probability that the network passes a test, where the channel \mathcal{E}'_x is applied locally on one part of an entangled state and the output is tested with a POVM containing the projector on the entangled state $|V\rangle\rangle/\sqrt{d_0}$. The fidelity can be expressed as

$$F(\mathcal{E}'_x, \mathcal{F}_x) := \frac{1}{d_0^2} \text{Tr}[C(|V_x\rangle\rangle\langle\langle V_x| \otimes E_x^T)].$$

Now, if the input channel \mathcal{E}_x is given with prior probability $p(x)$, the average channel fidelity is given by

$$\begin{aligned}
 F &= \sum_x p(x) F(\mathcal{E}'_x, \mathcal{F}_x) \\
 &= \text{Tr}[\Omega C], \quad \Omega := \sum_x p(x) (|V_x\rangle\rangle\langle\langle V_x| \otimes E_x^T).
 \end{aligned} \tag{74}$$

Thanks to theorem (2), the maximum fidelity can be expressed as

$$F_{\max} = \min_{\Gamma \in \text{DualComb}} \{ \lambda \in \mathbb{R} \mid \lambda \Gamma \geq \Omega \}, \tag{75}$$

where DualComb is the set of positive operators on $\mathcal{H}_3 \otimes \mathcal{H}_2 \otimes \mathcal{H}_1 \otimes \mathcal{H}_0$ satisfying the conditions

$$\begin{aligned}
 \Gamma &= I_3 \otimes T_{210}, \\
 \text{Tr}_2[T_{210}] &= I_1 \otimes T_0, \\
 \text{Tr}[T_0] &= 1.
 \end{aligned} \tag{76}$$

In the following we illustrate the use of this expression in a few examples.

9.2. Optimal inversion of an unknown unitary dynamics

Unitary quantum dynamics is, by definition, invertible: given a classical description of a unitary gate U , in principle one can always engineer the gate U^\dagger implementing the inverse physical process. However, the situation is different when the gate U is unknown. Can we devise a *physical* inversion mechanism, which transforms every unknown unitary dynamics U , given as a black box, into its inverse? Classically, the analogue of inverting a unitary dynamics is inverting a permutation. Inverting a permutation with a single evaluation is clearly impossible, because evaluating the permutation allows us to know its action on one input at most, and this information is not sufficient to perform an inversion on the other inputs. In the quantum domain, the situation is more interesting, because one use of a unitary gate is enough to store it faithfully into a quantum memory, by applying U on one side of a maximally entangled state. A first question is whether the information stored in the memory can be extracted and used to implement the inverse gate U^\dagger . Interestingly, this possibility is barred by Nielsen’s and Chuang’s no-programming theorem [104], which states that only orthogonal states can be used to program unitary gates deterministically and without error. As an alternative, one can try to think of protocols that simulate U^\dagger with one use of U , without storing U in a quantum memory. Protocols of this form are implemented by quantum networks as in equation (71). We now show that even such protocols cannot implement a perfect inversion. More specifically, we now show that the best way to generate the inverse of an unknown dynamics is simply to estimate it and to use the estimate to implement an approximate inversion. Our result highlights an analogy between the optimal inversion of an unknown unitary dynamics and the optimal universal NOT (UNOT) gate [105, 106], the quantum channel that attempts to transform every pure quantum state into its orthogonal complement. A known fact is that no quantum channel can approximate the ideal UNOT gate better than a channel that measures the input state and produces an orthogonal state based on the measurement outcome [105, 106]. Considering this feature, one can think of the unitary inversion as the

analogue of the UNOT: they are both involutions and they both are implemented optimally by measure-and-prepare strategies.

Let us assume that the unknown unitary gate U is drawn at random according to the normalized Haar measure dU . Then, the performance operator in equation (74) takes the form

$$\Omega = \frac{1}{d^2} \int dU |U^\dagger\rangle\rangle\langle\langle U^\dagger|_{30} \otimes |\bar{U}\rangle\rangle\langle\langle \bar{U}|_{21}, \quad (77)$$

with $d = d_0 = d_1 = d_2 = d_3$. The evaluation of the fidelity is provided in appendix G, where we obtain the value

$$F_{\max} = \frac{2}{d^2}. \quad (78)$$

Now, it turns out that the maximum fidelity can be achieved through the estimation of the gate U . Indeed, the optimal strategy for gate estimation is to prepare a maximally entangled state, to apply the unknown gate U on one side, and to perform the POVM $P_{\hat{U}} = d |\hat{U}\rangle\rangle\langle\langle \hat{U}|$ [107]. This strategy leads to the conditional probability distribution

$$p(\hat{U}|U) = |\text{Tr}[U^\dagger \hat{U}]|^2,$$

normalized with respect to the Haar measure. Averaging the channel fidelity $F(\hat{U}, U) = |\text{Tr}[U^\dagger \hat{U}]|^2 / d^2$, we then obtain the value

$$\begin{aligned} F_{\text{est}}(U) &= \int d\hat{U} |\text{Tr}[U^\dagger \hat{U}]|^4 / d^2 \\ &= 2/d^2 \\ &\equiv F_{\max}, \quad \forall U \in \text{SU}(d). \end{aligned} \quad (79)$$

The continuous POVM with $P_{\hat{U}} = d |\hat{U}\rangle\rangle\langle\langle \hat{U}|$ can also be replaced by a discrete Bell measurement, with d^2 outcomes, without affecting the fidelity in the worst case scenario, or equivalently, the average fidelity over all unitaries. One way or another, the above discussion proves that no quantum network can invert a gate better than a classical network that generates the inverse by using gate estimation as an intermediate step.

9.3. Simulating the evolution of a charge conjugate particle

In quantum mechanics, complex conjugation implements the symmetry between particles and antiparticles. If the evolution of a quantum particle is described by the unitary transformation U , then the evolution of the corresponding antiparticle will be described by the unitary transformation \bar{U} , where each matrix element is replaced by its complex conjugate. Consider the scenario where one is given a black box that performs a unitary transformation on a certain particle. Can we use this black box to simulate the evolution of the corresponding antiparticle? Physically, the most general simulation strategy is described by a quantum network as in equation (71).

For the charge conjugation problem, the performance operator Ω reads

$$\begin{aligned} \Omega &= \frac{1}{d^2} \int dU |\bar{U}\rangle\rangle\langle\langle \bar{U}|_{30} \otimes |\bar{U}\rangle\rangle\langle\langle \bar{U}|_{21} \\ &\quad \frac{1}{d^2} \left(\frac{P_{+,32} \otimes P_{+,10}}{d_+} + \frac{P_{-,32} \otimes P_{-,10}}{d_-} \right), \end{aligned} \quad (80)$$

where P_+ and P_- (d_+ and d_-) are the projectors on (the dimensions of) the symmetric and antisymmetric subspaces, respectively. In appendix H we evaluate the dual expression equation (75), obtaining the maximum fidelity

$$F_{\max} = \frac{2}{d(d-1)}.$$

Note that the fidelity is equal to 1 in the case of two-dimensional quantum systems. This is consistent with the fact that, for $d = 2$, the matrices U and \bar{U} are unitarily equivalent—specifically, $\bar{U} = YUY$, where Y is the Pauli matrix $Y := \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}$. Therefore, one can implement the complex conjugation by sandwiching the original unitary between two Pauli gates.

For systems of large dimension, the fidelity converges to $2/d^2$, the value achieved by gate estimation (see equation (79) in the previous paragraph). This means that gate estimation is asymptotically the optimal strategy, but, remarkably, it is not the optimal strategy when d is finite. The optimal simulation of the charge conjugate dynamics is achieved by the network with Choi operator

$$C = \frac{d P_{-,32}}{d_-} \otimes \frac{d P_{-,10}}{d_-}.$$

It is immediate to verify that, indeed, the operator C satisfies the normalization constraints and that one has $\text{Tr}[\Omega C] = 1/d_- = F_{\max}$. Physically, C represents a ‘disconnected network’ of the form

$$\begin{array}{c} A_0 \\ \hline \boxed{\mathcal{K}} \\ \hline A_1 \end{array} \quad \begin{array}{c} A_2 \\ \hline \boxed{\mathcal{K}} \\ \hline A_3 \end{array},$$

consisting of two subsequent uses of the channel \mathcal{K} with Choi operator $K := d P_-/d_-$. When the input gate U is inserted in the open slot, the overall evolution from system A_0 to system A_3 is given by the channel $\mathcal{F} = \mathcal{K} U \mathcal{K}$, which optimally simulates the charge conjugate evolution \bar{U} .

It is interesting to further elaborate on the physical meaning of the operations in the network. At first, one may guess that the optimal way to conjugate an unknown unitary U is to approximate the sequence of transformations

$$\rho \xrightarrow{\text{transpose}} \rho^T \xrightarrow{U} U \rho^T U^\dagger \xrightarrow{\text{transpose}} \bar{U} \rho U^T. \quad (81)$$

As the transpose is not a physical operation, one may try to use the optimal transpose channel [108–112], which has Choi operator $T = d P_+/d_+$. However, this choice would be suboptimal, leading to the fidelity

$$F_{\text{transpose}} = 1/d_+ = 2/[d(d+1)] < F_{\max}.$$

Instead, the optimal strategy is to approximate the *transpose NOT*, i.e. the impossible transformation that maps every projector into its orthogonal complement. In the Heisenberg picture, the transpose NOT maps every observable A into the observable $I - A^T$, allowing us to reproduce the charge conjugate dynamics as

$$A \xrightarrow{\text{transpose NOT}} I - A^T \xrightarrow{U} I - U^\dagger A^T U \xrightarrow{\text{transpose NOT}} U^T A \bar{U}.$$

It turns out that the optimal approximation of the transpose NOT is exactly the channel \mathcal{K} used in our network: in summary, the optimal simulation of the charge conjugate dynamics employs the optimal transpose NOT instead of the optimal transpose. Some intuition to justify this bizarre fact comes from the observation that the optimal transpose can be implemented via state estimation and, therefore, approximating the sequence (81) would lead to a classical, estimation-based strategy. Instead, the transpose NOT cannot be achieved via state estimation. For example, the transpose NOT for qubits is a unitary transformation, corresponding to the Pauli matrix Y .

9.4. Optimal controlization of unknown gates

Given a unitary gate U , the corresponding control unitary gate is

$$\text{ctrl} - U := I \otimes |0\rangle\langle 0| + U \otimes |1\rangle\langle 1|,$$

where $|0\rangle$ and $|1\rangle$ are the states of a qubit acting as control system. Controlization is the task of transforming an unknown gate U , accessed as a black box, into the corresponding gate $\text{ctrl} - U$.

When U is an arbitrary unitary, perfect controlization is impossible, as it was recently shown in [97, 98]. Like the no-cloning Theorem, this ‘no-controlization’ result establishes the impossibility of a perfect functionality. But what about approximate controlization? *A priori*, nothing forbids that one could engineer an approximate controlization protocol that achieves high-fidelity, almost circumventing the no-go Theorem. In the following we show that this is not the case. For a completely unknown unitary gate U , we show that not only is perfect controlization impossible, but also that every quantum strategy for controlization will be at most as good as a classical strategy that measures the control qubit and performs the gate U or the identity depending on the measurement outcome.

For the controlization task, the performance operator Ω reads

$$\Omega = \frac{1}{2d^2} \int dU |\text{ctrl} - U\rangle\langle\langle \text{ctrl} - U| \otimes |\bar{U}\rangle\rangle\langle\langle \bar{U}|. \quad (82)$$

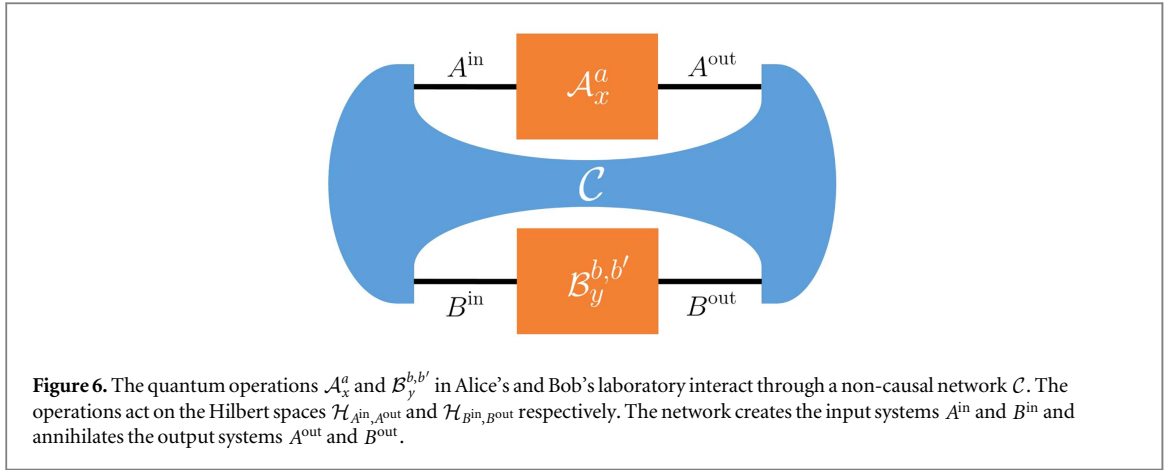
The evaluation of the maximum fidelity, carried out in appendix I, yields the optimal fidelity

$$F_{\max} = \frac{1}{2}.$$

By direct inspection, one can check that this is the same fidelity achieved by a network that measures the control qubit in the computational basis $\{|0\rangle, |1\rangle\}$ and applies the unknown gate U when the outcome is 1. Specifically, such strategy turns the input gate U into the classically-controlled channel \mathcal{C}_U defined by

$$\mathcal{C}_U(\rho \otimes \sigma) := \langle 0|\sigma|0\rangle \rho + \langle 1|\sigma|1\rangle U \rho U^\dagger,$$

where ρ is an arbitrary state of the system and σ is an arbitrary state of the control qubit. It is immediate to check that the fidelity between the classically-controlled channel \mathcal{C}_U and the control-unitary gate is $1/2$ for every



unitary. The above argument shows that no quantum circuit can perform better than a classical circuit where the control qubit is decohered by a measurement.

9.5. Maximization of the payoff in a non-causal quantum game

Here we consider the non-causal game introduced by Oreshkov *et al* in [41]. The game involves two spatially separated parties, Alice and Bob, and a referee, who sends inputs to and receives outputs from the players. Specifically, the referee sends an input bit a to Alice and two input bits b and b' to Bob. Then, the referee demands one output bit x from Alice and one output bit y from Bob. The referee assigns a score $\omega(x, y|a, b, b')$, given by

$$\omega(x, y|a, b, 0) = \begin{cases} 1 & x = b, \\ 0 & x \neq b, \end{cases} \quad \text{and} \quad \omega(x, y|a, b, 1) = \begin{cases} 1 & y = a, \\ 0 & y \neq a. \end{cases} \quad (83)$$

In this game, Alice and Bob are not subject to the no-signaling constraint. In principle, Alice may be able to communicate to Bob, or vice-versa. The only constraint is that Alice and Bob can interact only through a fixed network, which allows for communication at most in one-way: either from Alice to Bob, or from Bob to Alice.

It is interesting to see how quantum resources can help Alice and Bob. The most general quantum resource is described by a network that connects Alice's operations to Bob's operations. The network will provide inputs A^{in} and B^{in} to Alice and Bob, respectively. Alice and Bob then perform local operations, transforming systems A^{in} and B^{in} them into systems A^{out} and B^{out} . The local operations depend on the inputs a and (b, b') and will generate the outputs x and y , respectively. Diagrammatically, this scenario is depicted in figure 6.

Mathematically, the operations are described by two quantum instruments $\{\mathcal{M}_x^a\}_{x=0,1}$ and $\{\mathcal{N}_y^{b,b'}\}_{y=0,1}$. With these settings, the probability distribution of the outputs is given by

$$p(x, y|a, b, b') = \text{Tr}[(\mathcal{M}_x^a \otimes \mathcal{N}_y^{b,b'}) C],$$

where $\{\mathcal{M}_x^a\}_{x=0,1}$ and $\{\mathcal{N}_y^{b,b'}\}_{y=0,1}$ are the Choi operators of Alice's and Bob's instruments, respectively, and C is the Choi operator of the network that mediates the interaction.

With this settings, the average score is given by

$$\begin{aligned} \omega &= \frac{1}{8} \sum_{a,b,b',x,y} \omega(x, y|a, b, b') p(x, y|a, b, b') \\ &= \text{Tr}[\Omega C], \end{aligned}$$

where Ω is the performance operator

$$\Omega := \frac{1}{8} \sum_{a,b,b',x,y} \omega(x, y|a, b, b') (\mathcal{M}_x^a \otimes \mathcal{N}_y^{b,b'}). \quad (84)$$

The main result by Oreshkov *et al* is that the average score is upper bounded as $\omega \leq 3/4$ whenever the network C has a definite causal order, whereas there exists a non-causal network C_* and local operations $\{\mathcal{M}_{x*}^a\}_{x=0,1}$ and $\{\mathcal{N}_{y*}^{b,b'}\}_{y=0,1}$ that achieve score

$$\omega_* = \frac{1}{2} \left(1 + \frac{1}{\sqrt{2}} \right). \quad (85)$$

Specifically, the score ω_* is achieved by choosing systems A^{in} , B^{in} , A^{out} , B^{out} to be qubits and by choosing the local operations with Choi operators

$$\begin{aligned}
M_{x*}^a &= \frac{1}{4} [I + (-1)^x \sigma_z]_{A^{\text{in}}} \otimes [I + (-1)^a \sigma_z]_{A^{\text{out}}}, \\
N_{y*}^{b,b'} &= b' \left\{ \frac{1}{2} [I + (-1)^y \sigma_z]_{B^{\text{in}}} \otimes \rho_{B^{\text{out}}} \right\} \\
&\quad + (b' \oplus 1) \left\{ \frac{1}{4} [I + (-1)^y \sigma_x]_{B^{\text{in}}} \otimes [I + (-1)^{b+y} \sigma_z]_{B^{\text{out}}} \right\}.
\end{aligned} \tag{86}$$

where \oplus denotes the addition modulo 2 and $\rho_{B^{\text{out}}}$ is a fixed quantum state on Bob's output, which can be chosen to be the maximally mixed state without loss of generality.

The score ω can be regarded as a measure of the non-causality of the network mediating the interactions between Alice and Bob. An interesting question is whether ω_* is the maximum score attainable when Alice's and Bob's instruments (86) are connected by an arbitrary non-causal network. This question has been indirectly answered by Brukner [54], who considered a more general scenario, wherein Alice's and Bob's local operations are also subject to optimization. Brukner showed that the payoff $\omega_* = (1 + 1/\sqrt{2})/2$ is maximum over all non-causal networks and over a certain class of two-outcome instruments on Alice's and Bob's side, allowing Alice's and Bob's systems to have generic dimensions. When Alice's and Bob's operations are fixed to the qubit operations (86) used in the original paper [41], we now present an alternative (and comparatively shorter) optimality proof for the value $\omega_* = (1 + 1/\sqrt{2})/2$. This result serves as an illustration of the SDP method, which provides here a nice and straightforward solution.

Inserting equation (86) into equation (84) we obtain the performance operator

$$\Omega = \sum_{i,j,k} |i\rangle \langle i|_{A^{\text{in}}} \otimes |j\rangle \langle j|_{A^{\text{out}}} \otimes \Omega_{ijk} \otimes |k\rangle \langle k|_{B^{\text{out}}},$$

where Ω_{ijk} are operators acting on B_1 and are defined as

$$\begin{aligned}
\Omega_{000} &= \frac{1}{8} (|+\rangle \langle +| + |0\rangle \langle 0|), & \Omega_{001} &= \frac{1}{8} (|-\rangle \langle -| + |0\rangle \langle 0|), \\
\Omega_{010} &= \frac{1}{8} (|+\rangle \langle +| + |1\rangle \langle 1|), & \Omega_{011} &= \frac{1}{8} (|-\rangle \langle -| + |1\rangle \langle 1|), \\
\Omega_{100} &= \frac{1}{8} (|-\rangle \langle -| + |0\rangle \langle 0|), & \Omega_{101} &= \frac{1}{8} (|+\rangle \langle +| + |0\rangle \langle 0|), \\
\Omega_{110} &= \frac{1}{8} (|-\rangle \langle -| + |1\rangle \langle 1|), & \Omega_{111} &= \frac{1}{8} (|+\rangle \langle +| + |1\rangle \langle 1|).
\end{aligned}$$

Now, the dual optimization problem is to find the minimum λ such that $\lambda \Gamma \geq \Omega$, for some Choi operator Γ representing a no-signaling channel. The key observation is that all the Ω_{ijk} have the same maximum eigenvalue, equal to $e_{\text{max}} = 1/8(1 + 1/\sqrt{2})$. As a result, we can satisfy the dual constraint by setting $\lambda = 1/2(1 + 1/\sqrt{2})$ and $\Gamma = I_{A^{\text{in}}A^{\text{out}}B^{\text{in}}B^{\text{out}}}/4$. Note that Γ is the Choi operator of a no-signaling channel, as it satisfies equations (59) and (60). Hence, we obtain the bound

$$\omega \leq \frac{1}{2} \left(1 + \frac{1}{\sqrt{2}} \right), \tag{87}$$

valid for every non-causal network. The bound can be achieved, since r.h.s. matches the value in equation (85).

10. Conclusions

We developed a SDP method for the optimization of quantum networks. The method can be applied to causal networks as well as more general networks with indefinite causal structure. For a large class of optimization problems, we observed that the maximum performance can be expressed in terms of a max relative entropy. Building on this fact, we extended the notions of conditional min-entropy and max relative entropy from quantum states to quantum networks. Specifically, the relative entropy between two networks can be characterized as the maximum of the relative entropy between the states that can be generated by the two networks. Similarly, the min-entropy of a quantum causal network can be characterized as the maximum min-entropy that the network can build up by interacting over time with a sequence of quantum devices. Intuitively, the network min-entropy can be regarded as a measure of the amount of quantum correlations generated over a sequence of time steps.

Our results have applications to a number of scenarios, including e. g. the optimization of algorithms for quantum causal discovery [28], tomography of quantum channels and causal networks [18, 69, 113, 114], and quantum machine learning [115–117]. Another stimulating avenue of future research is on the quantum engineering side, where our method can be adapted to deal with optimization tasks in the presence of limited energy resources. For example, it is interesting to explore the causal networks that can be implemented at zero-

energy cost, extending to the network scenario the results obtained in [118] for individual state transitions. The interesting aspect here is the possibility to borrow energy resources at a certain time and to return them at later times, resulting in an overall zero energy balance. As a further step, the extension from quantum networks working in the zero-energy regime to network using bounded energy resources is even more compelling in view of future applications. Exploring how energy and coherence across energy eigenstates can be optimally allocated within a distributed system is expected to unveil new quantum advantages, leading to a new layer of optimization in the design of quantum technologies.

Acknowledgments

We acknowledge the referees of this paper for useful comments that helped improving the presentation and strengthening our results. The research of our group is supported by the Foundational Questions Institute (FQXi-RFP3-1325 and FQXi-MGA-1502), the National Natural Science Foundation of China through Grant No. 11675136, the Hong Kong Research Grant Council through Grant No.17326616, the Canadian Institute for Advanced Research, the HKU Seed Funding for Basic Research, and the John Templeton Foundation. GC is grateful to F Buscemi and YC Liang for useful discussions and to A Acín, M Hoban, and R Chavez for organizing the workshop ‘Quantum Networks’ Barcelona, 30 March–1 April 2016, which offered the occasion for a stimulating exchange of ideas that benefitted this paper.

Appendix A. Proof of theorem 1

Proof. By definition, the value of the primal problem is given by

$$\begin{aligned}\omega_{\text{primal}} &= \sup \{ \langle A, X \rangle \mid X \geq 0, X \in \mathbf{S} \} \\ &= \sup \{ \langle A, X \rangle \mid X \geq 0, X \in \bar{\mathbf{S}} \} \\ &= \sup \{ \langle A, X \rangle \mid X \geq 0, \langle \Gamma, X \rangle = 1, \forall \Gamma \in \bar{\mathbf{S}} \},\end{aligned}$$

having used the relation $\mathbf{S} = \bar{\mathbf{S}}$. Now, let us pick an affine basis for $\bar{\mathbf{S}}$, say $(\Gamma_i)_{i=1}^K$ and re-write the value of the primal problem as

$$\omega_{\text{primal}} = \sup \{ \langle A, X \rangle \mid X \geq 0, \langle \Gamma_i, X \rangle = 1, \forall i \in \{1, \dots, K\} \}.$$

Weak duality then yields the relation

$$\omega_{\text{primal}} \leq \inf \left\{ \sum_{i=1}^K \lambda_i \mid \lambda_i \in \mathbb{R}, \sum_i \lambda_i \Gamma_i \geq A \right\} \quad (\text{A.1})$$

$$\leq \inf \left\{ \sum_{i=1}^K \lambda_i \mid \lambda_i \in \mathbb{R}, \sum_i \lambda_i \Gamma_i \geq A, \sum_i \lambda_i \neq 0 \right\} \quad (\text{A.2})$$

$$= \inf_{\Gamma \in \bar{\mathbf{S}}} \min \{ \lambda \in \mathbb{R} \mid \lambda \Gamma \geq A \}, \quad (\text{A.3})$$

having defined $\lambda := \sum_i \lambda_i$ and $\Gamma := \sum_i \lambda_i \Gamma_i / \lambda$.

Now, suppose that \mathbf{S} contains a positive operator X_0 and $\bar{\mathbf{S}}$ contains a strictly positive operator Γ_0 , then Slater’s theorem implies the equality: indeed, one can choose the affine basis $(\Gamma_i)_{i=1}^K$ to contain the operator Γ_0 . Since Γ_0 is strictly positive, one can find strictly positive coefficients $(\lambda_i)_{i=1}^K$ such that $\sum_i \lambda_i \Gamma_i \geq A$. This means that the dual problem in the rhs of equation (A.1) admits a strictly positive solution. Hence, proposition 4 implies the equality in equation (A.1). The equality holds also in equation (A.2), because every solution with $\sum_i \lambda_i = 0$ can be replaced by a new solution with $\sum_i \lambda'_i = \epsilon$, by substituting λ_1 with $\lambda_1 + \epsilon$, $\epsilon > 0$. Since ϵ can be arbitrarily small, this substitution does not change the value of the infimum. If A is positive, then one has the lower bound $\omega_{\text{primal}} \geq \langle A, X_0 \rangle \geq 0$. Equation (A.3) then implies that every λ satisfying $\lambda \Gamma \geq A$, $\Gamma \in \bar{\mathbf{S}}$ must be non-negative. If λ is strictly positive, the operator Γ must be positive. If $\lambda = 0$, the operator Γ can be chosen to be positive without loss of generality. In conclusion, the infimum in equation (A.3) can be restricted to $\bar{\mathbf{S}}_+$. Setting $w := 1/\lambda$ one finally obtains the desired expression. \square

Appendix B. Proof of theorem 2

Proof. The maximum performance is given by equation (21). The expression can be re-written as

$$\omega_{\text{max}} := \max \{ \langle \Omega, C \rangle \mid C \in \mathbf{S}, C \geq 0 \}, \quad (\text{B.1})$$

where \mathbf{S} is the affine space of all the operators on $\bigotimes_{j=1}^N (\mathcal{H}_j^{\text{out}} \otimes \mathcal{H}_j^{\text{in}})$ that are Hermitian and satisfy the linear constraint (12). Note that \mathbf{S} contains the strictly positive operator

$$C_0 = I_0 \otimes I_1 \otimes \cdots \otimes I_{2N-1} / (d_1 d_3 \dots d_{2N-1}) \quad (\text{B.2})$$

and the dual affine space $\bar{\mathbf{S}}$ contains the strictly positive operator

$$\Gamma_0 = I_0 \otimes I_1 \otimes \cdots \otimes I_{2N-1} / (d_0 d_2 \dots d_{2N-2}). \quad (\text{B.3})$$

Since the sets \mathbf{S} and $\bar{\mathbf{S}}$ contain strictly positive operators, the expression in theorem 1 holds with the equality. Moreover, one can choose the performance operator Ω to be positive without loss of generality: if Ω is not positive, one can define $\Omega' = \Omega + c\Gamma_0$, where c is a positive constant and Γ_0 is the operator in equation (B.3). This substitution only shifts the primal and dual values by the constant c , while preserving the optimal solutions. For the shifted problem, theorem 1 guarantees that the dual optimization can be restricted to the positive operators in $\bar{\mathbf{S}}_+$, namely

$$\omega'_{\max} = \inf_{\Gamma \in \bar{\mathbf{S}}_+} \min \{ \lambda \in \mathbb{R} \mid \lambda \Gamma \geq \Omega' \}.$$

Now, the set $\bar{\mathbf{S}}_+$ has been characterized in [18]: precisely, $\bar{\mathbf{S}}^+$ is the set of all positive operators Γ satisfying the linear constraint

$$\begin{aligned} \Gamma &= I_{A_N^{\text{out}}} \otimes \Gamma^{(N)} \\ \text{Tr}_{A_n^{\text{in}}}[\Gamma^{(N)}] &= I_{A_{n-1}^{\text{out}}} \otimes \Gamma^{(N-1)}, \quad n = 2, \dots, N \\ \text{Tr}_{A_1^{\text{in}}}[\Gamma^{(1)}] &= 1, \end{aligned} \quad (\text{B.4})$$

for suitable positive operators $\Gamma^{(n)}$ acting on $\mathcal{H}_n^{\text{in}} \otimes [\bigotimes_{j=1}^{n-1} (\mathcal{H}_j^{\text{out}} \otimes \mathcal{H}_j^{\text{in}})]$. Observing that $I_{A_N^{\text{out}}}$ is the Choi operator of the trace channel $\text{Tr}_{A_N^{\text{out}}}$ and comparing equation (B.4) with equation (12) we then obtain that every operator Γ in $\bar{\mathbf{S}}^+$ is the Choi operator of a network of the form (32). Hence, $\bar{\mathbf{S}}^+ = \text{DualComb}$. Finally, note that the set DualComb is compact and therefore the infimum is a minimum. \square

Appendix C. Proof of proposition 5

Proof. By definition, the max relative entropies are given by $D_{\max}(C_0 \| C_1) = -\log \max W$ and $D_{\max}(\sqrt{\Gamma} C_0 \sqrt{\Gamma} \| \sqrt{\Gamma} C_1 \sqrt{\Gamma}) = -\log \max W(\Gamma)$, with

$$\begin{aligned} W &:= \{w \in \mathbb{R} \mid w C_0 \leq C_1\}, \\ W(\Gamma) &:= \{w \in \mathbb{R} \mid w \sqrt{\Gamma} C_0 \sqrt{\Gamma} \leq \sqrt{\Gamma} C_1 \sqrt{\Gamma}\}. \end{aligned}$$

By construction, one has $W \subseteq W(\Gamma)$ for every Γ , and therefore

$$D_{\max}(C_0 \| C_1) \geq D_{\max}(\sqrt{\Gamma} C_0 \sqrt{\Gamma} \| \sqrt{\Gamma} C_1 \sqrt{\Gamma}).$$

On the other hand, if $\bar{\mathbf{S}}_+$ contains a full-rank element Γ_* , then $W(\Gamma_*) = W$. \square

Appendix D. Proof of proposition 6

Proof. Let us compute the conditional min-entropy of the output state

$$\rho = \sigma * D_1 * E_1 * D_2 * E_2 * \cdots * E_{N-1} * D_N \in \text{St}(B_N^{\text{out}} \otimes B_N^{\text{out}'})$$

(See equation (42)). By the operational characterization of the conditional min-entropy (equation (5)), we have

$$H_{\min}(B_N^{\text{out}} | B_N^{\text{out}'})_{\rho} = \max_{\substack{C \geq 0, \\ \text{Tr}_{B_N^{\text{out}}}[C] = I_{B_N^{\text{out}'}}} \frac{\text{Tr}[\rho C^T]}{d_{B_N^{\text{out}}}}, \quad (\text{D.1})$$

where C is the Choi operator of a recovery channel \mathcal{C} , which attempts to turn ρ into the maximally entangled state $|\Phi\rangle$. Substituting the expression for ρ and maximizing over the sequence $(\sigma, E_1, \dots, E_{N-1})$, we then obtain

$$\begin{aligned} & \max_{\sigma, E_1, E_2, \dots, E_{N-1}} H_{\min}(B_N^{\text{out}} | B_N^{\text{out}'})_{\rho} \\ &= \max_{\sigma, E_1, E_2, \dots, E_{N-1}, C} \frac{\text{Tr}[(\sigma * D_1 * E_1 * D_2 * E_2 * \dots * E_{N-1} * D_N) C^T]}{d_{B_N^{\text{out}}}} \\ &= \max_{\sigma, E_1, E_2, \dots, E_{N-1}, C} \frac{(\sigma * D_1 * E_1 * D_2 * E_2 * \dots * E_{N-1} * D_N) * C}{d_{B_N^{\text{out}}}} \end{aligned} \quad (\text{D.2})$$

$$= \max_{\sigma, E_1, E_2, \dots, E_{N-1}, C} \frac{E' * R}{d_{B_N^{\text{out}}}}, \tag{D.3}$$

having defined

$$E' = \sigma * E_1 \cdots * E_{N-1} * C \quad \text{and} \quad R = D_1 * \cdots * D_N.$$

Now, note that E' is the Choi operator of a network of the form of equation (41). Moreover, since the channel C can be chosen to be the identity, E' is the Choi operator of an *arbitrary* network of the form of equation (41). Using equations (43) and (44) we finally obtain

$$\max_{\sigma, E_1, E_2, \dots, E_{N-1}} H_{\min}(B_N^{\text{out}} | B_N^{\text{out}'})_\rho = H_{\min}(t_N | t_1 \dots t_{N-1})_R.$$

□

Appendix E. Proof of proposition 7

Proof. The proof is based on proposition 5. Take an operator $\Gamma \in \text{DualComb}(\mathcal{H}_{A_1}^{\text{in}}, \mathcal{H}_{A_1}^{\text{out}}, \dots, \mathcal{H}_{A_N}^{\text{in}}, \mathcal{H}_{A_N}^{\text{out}})$ and diagonalize it as $\Gamma = \sum_i g_i |\phi_i\rangle \langle \phi_i|$. Choose S to be the composite system $A_1^{\text{in}} A_1^{\text{out}} \cdots A_N^{\text{in}} A_N^{\text{out}}$ and define the vector $|\Psi\rangle = \sum_i \sqrt{g_i} |\phi_i\rangle |\bar{\phi}_i\rangle \in \mathcal{H}_{A_1^{\text{in}}} \otimes \mathcal{H}_{A_1^{\text{out}}} \otimes \cdots \otimes \mathcal{H}_{A_N^{\text{in}}} \otimes \mathcal{H}_{A_N^{\text{out}}} \otimes \mathcal{H}_S$. Then, the positive operator $E = |\Psi\rangle \langle \Psi|$ is the Choi operator of a network of the form of equation (50), as one can check from equation (12). Then, explicit calculation gives

$$C^{(x)} * E = \sqrt{\Gamma} C^{(x)} \sqrt{\Gamma}.$$

Using proposition (5) we then conclude the equality

$$D_{\max}(C^{(0)} \| C^{(1)}) = \max_{\Gamma} D_{\max}(C^{(0)} * E \| C^{(1)} * E).$$

□

Appendix F. Normalization condition for supermaps on product channels

Equation (56) gives us the Choi operator C . In order for C to be the Choi operator of a channel, we must have

$$\text{Tr}_{A_2^{\text{out}}, B_2^{\text{out}}}[C] = I_{A_2^{\text{in}}} \otimes I_{B_2^{\text{in}}}. \tag{F.1}$$

Inserting equation (56), we then obtain the condition

$$\text{Tr}_{A_2^{\text{out}}, B_2^{\text{out}}, A_1^{\text{in}}, A_1^{\text{out}}, B_1^{\text{in}}, B_1^{\text{out}}}[(I_{A_2^{\text{out}}} \otimes I_{A_2^{\text{in}}} \otimes N \otimes I_{B_2^{\text{out}}} \otimes I_{B_2^{\text{in}}})(A \otimes B)] = I_{A_2^{\text{in}}} \otimes I_{B_2^{\text{in}}}, \tag{F.2}$$

which must be satisfied whenever A and B satisfy the conditions

$$\text{Tr}_{A_1^{\text{out}}, A_2^{\text{out}}}[A] = I_{A_1^{\text{in}}} \otimes I_{A_2^{\text{in}}} \quad \text{and} \quad \text{Tr}_{B_1^{\text{out}}, B_2^{\text{out}}}[B] = I_{B_1^{\text{in}}} \otimes I_{B_2^{\text{in}}}. \tag{F.3}$$

Now, we have the following

Proposition 11. For every operator N , the following conditions are equivalent:

- (i) N satisfies the condition (F.2) for every operators A and B satisfying the condition (F.3)
- (ii) N satisfies the condition

$$\text{Tr}_{A_1^{\text{in}}, A_1^{\text{out}}, B_1^{\text{in}}, B_1^{\text{out}}}[N(\tilde{A} \otimes \tilde{B})] = 1 \tag{F.4}$$

for every operators $\tilde{A} \in \text{Lin}(\mathcal{H}_{A_1^{\text{out}}} \otimes \mathcal{H}_{A_1^{\text{in}}})$ and $\tilde{B} \in \text{Lin}(\mathcal{H}_{B_1^{\text{out}}} \otimes \mathcal{H}_{B_1^{\text{in}}})$ satisfying the conditions

$$\text{Tr}_{A_1^{\text{out}}}[\tilde{A}] = I_{A_1^{\text{in}}} \quad \text{and} \quad \text{Tr}_{B_1^{\text{out}}}[\tilde{B}] = I_{B_1^{\text{in}}}. \tag{F.5}$$

Proof. Suppose that the operators \tilde{A} and \tilde{B} satisfy the trace conditions (F.5). By defining the operators A and B as $A = \tilde{A} \otimes I_{A_2^{\text{in}}, A_2^{\text{out}}}/d_{A_2^{\text{in}}}$ and $B = \tilde{B} \otimes I_{B_2^{\text{in}}, B_2^{\text{out}}}/d_{B_2^{\text{in}}}$, we see that equation (F.3) is satisfied. Then, equation (F.2) becomes

$$\begin{aligned} & \text{Tr}_{A_2^{\text{out}}, B_2^{\text{out}}, A_1^{\text{in}}, A_1^{\text{out}}, B_1^{\text{in}}, B_1^{\text{out}}}[(I_{A_2^{\text{out}}} \otimes I_{A_2^{\text{in}}} \otimes N \otimes I_{B_2^{\text{out}}} \otimes I_{B_2^{\text{in}}})(A \otimes B)] \\ &= \text{Tr}_{A_2^{\text{out}}, B_2^{\text{out}}, A_1^{\text{in}}, A_1^{\text{out}}, B_1^{\text{in}}, B_1^{\text{out}}}\left\{ \frac{I_{A_2^{\text{out}}}}{d_{A_2^{\text{in}}}} \otimes I_{A_2^{\text{in}}} \otimes [N(\tilde{A} \otimes \tilde{B})] \otimes \frac{I_{B_2^{\text{out}}}}{d_{B_2^{\text{in}}}} \otimes I_{B_2^{\text{in}}} \right\} = I_{A_2^{\text{in}}} \otimes I_{B_2^{\text{in}}}. \end{aligned} \tag{F.6}$$

The above equation holds if and only if condition (F.4) is satisfied. Conversely, if the operator N satisfies condition (F.4) and \bar{A} and \bar{B} the trace conditions (F.5), we obtain

$$\begin{aligned} & \text{Tr}_{A_2^{\text{out}}, B_2^{\text{out}}, A_1^{\text{in}}, A_1^{\text{out}}, B_1^{\text{in}}, B_1^{\text{out}}} [(I_{A_2^{\text{out}}} \otimes I_{A_2^{\text{in}}} \otimes N \otimes I_{B_2^{\text{out}}} \otimes I_{B_2^{\text{in}}})(A \otimes B)] \\ &= \text{Tr}_{A_1^{\text{in}}, A_1^{\text{out}}, B_1^{\text{in}}, B_1^{\text{out}}} [(I_{A_2^{\text{in}}} \otimes N \otimes I_{B_2^{\text{in}}})(\text{Tr}_{A_2^{\text{out}}, B_2^{\text{out}}}[A \otimes B])] \\ &= \text{Tr}_{A_1^{\text{in}}, A_1^{\text{out}}, B_1^{\text{in}}, B_1^{\text{out}}} [(I_{A_2^{\text{in}}} \otimes N \otimes I_{B_2^{\text{in}}})(\bar{A} \otimes \bar{B})], \end{aligned} \quad (\text{F.7})$$

where we defined $\text{Tr}_{A_2^{\text{out}}}[A] = \bar{A}$ and $\text{Tr}_{B_2^{\text{out}}}[B] = \bar{B}$. Hence, equation (F.2) holds if and only if

$$\text{Tr}_{A_1^{\text{in}}, A_1^{\text{out}}, B_1^{\text{in}}, B_1^{\text{out}}} [(I_{A_2^{\text{in}}} \otimes N \otimes I_{B_2^{\text{in}}})(\bar{A} \otimes \bar{B})] = I_{A_2^{\text{in}}} \otimes I_{B_2^{\text{in}}}. \quad (\text{F.8})$$

In turn, the above equation holds if and only if

$$\text{Tr}_{A_1^{\text{in}}, A_1^{\text{out}}, B_1^{\text{in}}, B_1^{\text{out}}} [N(\bar{A}_\rho \otimes \bar{B}_\sigma)] = 1, \quad \forall \rho \in \text{St}(\mathcal{H}_{A_2^{\text{in}}}), \forall \sigma \in \text{St}(\mathcal{H}_{B_2^{\text{in}}}), \quad (\text{F.9})$$

where \bar{A}_ρ and \bar{B}_σ are defined as

$$\bar{A}_\rho := \text{Tr}_{A_2^{\text{in}}}[(\rho \otimes I_{A_1^{\text{out}} A_1^{\text{in}}})\bar{A}] \quad \text{and} \quad \bar{B}_\sigma := \text{Tr}_{B_2^{\text{in}}}[(\rho \otimes I_{B_1^{\text{out}} B_1^{\text{in}}})\bar{B}]. \quad (\text{F.10})$$

Now, the normalization condition (F.9) is nothing but equation (F.4). The condition is satisfied because the operators \bar{A}_ρ and \bar{B}_σ satisfy condition (F.5). \square

Appendix G. Maximum fidelity for the inversion of an unknown dynamics

The performance operator Ω reads

$$\begin{aligned} \Omega &= \frac{1}{d^2} \int dU |U^\dagger\rangle\langle U^\dagger|_{30} \otimes |\bar{U}\rangle\langle \bar{U}|_{21} \\ &= \frac{1}{d^2} \int dU (I_3 \otimes \bar{U}_0 \otimes \bar{U}_2 \otimes I_1) (|I\rangle\langle I|_{30} \otimes |I\rangle\langle I|_{21}) (I_3 \otimes \bar{U}_0 \otimes \bar{U}_2 \otimes I_1)^\dagger. \end{aligned} \quad (\text{G.1})$$

Explicit calculation using Schur's lemma yields the relations

$$[\Omega, I_3 \otimes U_2 \otimes I_1 \otimes U_0] = 0, \quad (\text{G.2})$$

$$[\Omega, U_3 \otimes I_2 \otimes U_1 \otimes I_0] = 0, \quad (\text{G.3})$$

required to hold for every unitary U . Explicitly, the operator Ω is given by

$$\Omega = \frac{1}{d^2} \left(\frac{P_{+,31} \otimes P_{+,20}}{d_+} + \frac{P_{-,31} \otimes P_{-,20}}{d_-} \right), \quad (\text{G.4})$$

P_+ and P_- are the projectors on the symmetric and antisymmetric subspace, respectively.

The problem is to find the minimum λ such that $\lambda\Gamma \geq \Omega$, for Γ satisfying the conditions (76). The first condition requires Γ to be of the form $\Gamma = I_3 \otimes T_{210}$. Now, equation (G.3) implies that, without loss of generality, the operator T_{210} can be chosen to satisfy the condition

$$[T_{210}, I_2 \otimes U_1 \otimes I_0] = 0 \quad \forall U \in \text{SU}(d) \quad (\text{G.5})$$

which in turn implies

$$T_{210} = Q_{20} \otimes I_1, \quad (\text{G.6})$$

where Q_{20} is some positive operator on \mathcal{H}_{20} . Similarly, equation (G.2) implies that we can choose T_{210} to satisfy the condition

$$[T_{210}, U_2 \otimes I_1 \otimes U_0] = 0, \quad \forall U \in \text{SU}(d). \quad (\text{G.7})$$

Combined with equation (G.6), the above relation implies

$$[Q_{20}, U_2 \otimes U_0] = 0 \quad \forall U \in \text{SU}(d) \quad (\text{G.8})$$

and therefore

$$Q_{20} = \alpha P_+ + \beta P_-. \quad (\text{G.9})$$

Finally, the last condition in equation (76) gives $\text{Tr}[Q_{20}] = 1$ and, therefore,

$$\alpha d_+ + \beta d_- = d.$$

The dual constraint $\lambda\Gamma \geq \Omega$ then reads

$$\lambda [\alpha (I_{31} \otimes P_{+,20}) + \beta (I_{31} \otimes P_{-,20})] \geq \frac{1}{d^2} \left(\frac{P_{+,31} \otimes P_{+,20}}{d_+} + \frac{P_{-,31} \otimes P_{-,20}}{d_-} \right). \quad (\text{G.10})$$

Pinching both sides with the projectors $P_{+,31} \otimes P_{+,20}$ and $P_{-,31} \otimes P_{-,20}$, one obtains

$$\lambda \geq \frac{1}{d_+ d^2 \alpha} \quad \text{and} \quad \lambda \geq \frac{1}{d^2 (d - \alpha d_+)}. \quad (\text{G.11})$$

By separately considering the cases $d_+ \alpha d^2 \geq (d - d_+ \alpha) d^2$ and $d_+ \alpha d^2 < (d - d_+ \alpha) d^2$, we find that the minimum λ is $\lambda_{\min} = 2/d^2$.

Appendix H. Maximum fidelity for the charge conjugation of an unknown unitary evolution

The maximization of the fidelity proceeds in the same way as for gate inversion. The only difference is that now the performance operator Ω is given by equation (80), namely

$$\Omega = \frac{1}{d^2} \left(\frac{P_{+,32} \otimes P_{+,10}}{d_+} + \frac{P_{-,32} \otimes P_{-,10}}{d_-} \right). \quad (\text{H.1})$$

The form of Ω implies the relations

$$[\Omega, U_3 \otimes U_2 \otimes I_{10}] = 0, \quad (\text{H.2})$$

$$[\Omega, I_{32} \otimes U_1 \otimes U_0] = 0, \quad (\text{H.3})$$

valid for every U in $\text{SU}(d)$. Now, one has to find the minimum λ such that $\lambda(I_3 \otimes T_{210}) \geq \Omega$, with some Γ satisfying equations (76). Equation (H.2) implies that, without loss of generality, one has

$$[T_{210}, U_2 \otimes I_{10}] = 0 \quad \forall U \in \text{SU}(d), \quad (\text{H.4})$$

and therefore $T_{210} = I_2 \otimes Q_{10}$. Moreover, the second condition in equation (76) reads

$$\text{Tr}_2[T_{210}] = I_1 \otimes \rho_0$$

and implies that Q_{10} has the form $Q_{10} = I_1 \otimes \rho_0/d$. Finally, equation (H.3) implies that one can choose $\rho_0 = I/d$ without loss of generality. Summing everything up, Γ can be chosen to be of the form $\Gamma = I_3 \otimes T_{210} = I_{3210}/d^2$. The dual constraint $\lambda \Gamma \geq \Omega$ then becomes

$$\lambda \frac{I_{3210}}{d^2} \geq \frac{1}{d^2} \left(\frac{P_{+,32} \otimes P_{+,10}}{d_+} + \frac{P_{-,32} \otimes P_{-,10}}{d_-} \right) \quad (\text{H.5})$$

yielding the minimal value $\lambda_{\min} = 1/d_- = 2/d(d-1)$.

Appendix I. Maximum fidelity for unitary controlization

The performance operator for the controlization problem is

$$\begin{aligned} \Omega &= \frac{1}{4d^2} \int dg |\text{ctrl} - U\rangle\langle\langle \text{ctrl} - U |_{30Q'Q} \otimes |\bar{U}\rangle\langle\langle \bar{U} |_{21} \\ &= \Omega_{3210}^{(0)} \otimes |0\rangle\langle 0|_{Q'} \otimes |0\rangle\langle 0|_Q + \Omega_{3210}^{(1)} \otimes |1\rangle\langle 1|_{Q'} \otimes |1\rangle\langle 1|_Q, \end{aligned} \quad (\text{I.1})$$

where Q and Q' denote the control qubit before and after the interaction, respectively, and

$$\Omega_{3210}^{(0)} := \frac{1}{4d^2} (E_{30} \otimes I_{21}), \quad (\text{I.2})$$

$$\Omega_{3210}^{(1)} := \frac{1}{4d^2} \left(E_{32} \otimes E_{10} + \frac{E_{32}^\perp \otimes E_{10}^\perp}{d_\perp} \right). \quad (\text{I.3})$$

Here E denotes the projector on the maximally entangled state $|\Phi^+\rangle = |I\rangle/\sqrt{d}$, E_\perp is the orthogonal projector $E_\perp := I^{\otimes 2} - E$, and $d_\perp := d^2 - 1$. Note that the operators $\Omega_{3210}^{(0)}$ and $\Omega_{3210}^{(1)}$ satisfies the conditions

$$[\Omega_{3210}^{(1)}, U_3 \otimes I_{21} \otimes \bar{U}_0] = 0, \quad (\text{I.4})$$

$$[\Omega_{3210}^{(1)}, U_3 \otimes \bar{U}_2 \otimes I_{10}] = 0, \quad (\text{I.5})$$

$$[\Omega_{3210}^{(1)}, I_{32} \otimes \bar{U}_1 \otimes U_0] = 0, \quad (\text{I.6})$$

for every group element $U \in \text{SU}(d)$.

To solve the dual problem, we have to find the minimum λ satisfying the relation $\lambda \Gamma \geq \Omega$ for some dual comb Γ . By equation (76), we have $\Gamma = I_b \otimes I_3 \otimes T_{210a}$, for some suitable operator T_{210a} satisfying the conditions

$$\begin{aligned}\mathrm{Tr}_2[T_{210a}] &= I_1 \otimes \rho_{0a}, \\ \mathrm{Tr}_{0a}[\rho_{0a}] &= 1.\end{aligned}$$

Without loss of generality, T_{210a} can be chosen of the form

$$T_{210a} = T_{210}^{(0)} \otimes |0\rangle\langle 0|_Q + T_{210}^{(1)} \otimes |1\rangle\langle 1|_Q, \quad (\text{I.7})$$

with the operators $T_{210}^{(0)}$ and $T_{210}^{(1)}$ satisfying the conditions

$$\mathrm{Tr}_2[T_{210}^{(0)}] = p_0 [I_1 \otimes \rho_0^{(0)}] \quad \text{and} \quad \mathrm{Tr}_2[T_{210}^{(1)}] = p_1 [I_1 \otimes \rho_0^{(1)}], \quad (\text{I.8})$$

where $\rho_0^{(0)}$ and $\rho_0^{(1)}$ are two density matrices and p_0 and p_1 are probabilities. The dual constraint is then reduced to

$$\lambda [I_3 \otimes T_{210}^{(k)}] \geq \Omega_{3210}^{(k)}, \quad \forall k \in \{0, 1\}. \quad (\text{I.9})$$

At this point, equation (I.4) implies that, without loss of generality, one can choose $T_{210}^{(0)}$ to satisfy the relation

$$[T_{210}^{(0)}, I_{21} \otimes \bar{U}_0] = 0, \quad \forall U \in \mathrm{SU}(d),$$

which implies $T_{210}^{(0)} = Q_{21}^{(0)} \otimes I_0$ for some suitable operator $Q_{21}^{(0)}$. Moreover, equation (I.2) implies that, without loss of generality, one can choose $Q_{21}^{(0)}$ to be proportional to the identity, so that, eventually one has

$$T_{210}^{(0)} = p_0 \frac{I_2 \otimes I_1 \otimes I_0}{d^2}. \quad (\text{I.10})$$

Similarly, equation (I.3) implies that, without loss of generality, one can choose $T_{210}^{(1)}$ to satisfy the relations

$$[T_{210}^{(1)}, \bar{U}_2 \otimes I_{10}] = 0, \quad (\text{I.11})$$

$$[T_{210}^{(1)}, I_2 \otimes \bar{U}_1 \otimes U_0] = 0, \quad (\text{I.12})$$

for every unitary $U \in \mathrm{SU}(d)$. Now, equation (I.11) implies that $T_{210}^{(1)}$ has the form

$$T_{210}^{(1)} = I_2 \otimes Q_{10}^{(1)} \quad (\text{I.13})$$

and equation (I.8) implies the condition

$$\begin{aligned}d Q_{10}^{(1)} &= \mathrm{Tr}_2[T_{210}^{(1)}] \\ &= p_1 [I_1 \otimes \rho_0^{(1)}]\end{aligned}$$

for some probability p_1 and some quantum state $\rho_0^{(1)}$. Combining equations (I.13) and (I.12) one finally obtains $Q_{10}^{(1)} = p_1 I_1 \otimes I_0/d^2$, and therefore

$$T_{210}^{(1)} = p_1 \frac{I_2 \otimes I_1 \otimes I_0}{d^2}. \quad (\text{I.14})$$

Inserting the above relations into the dual constraint, we then obtain

$$\lambda p_0 \frac{I_{3210}}{d^2} \geq \frac{1}{4d^2} (E_{30} \otimes I_{21}) \quad (\text{I.15})$$

$$\lambda p_1 \frac{I_{3210}}{d^2} \geq \frac{1}{4d^2} \left(E_{32} \otimes E_{10} + \frac{E_{32}^\perp \otimes E_{10}^\perp}{d_\perp} \right), \quad (\text{I.16})$$

having used equations (I.10), and (I.14). To satisfy the constraint, the parameters λ , p_0 , and p_1 must satisfy $\lambda p_0 \geq 1/4$ and $\lambda p_1 \geq 1/4$, leading to the minimum value $\lambda_{\min} = 1/2$.

References

- [1] Aspelmeyer M, Zeilinger A, Lindenthal M, Molina-Terriza G, Poppe A, Resch K, Ursin R, Walther P and Jennewein T D 2005 Advanced quantum communications experiments with entangled photons *Quantum Communications and Cryptography* (Boca Raton, FL: CRC) p 45
- [2] Pirandola S, Eisert J, Weedbrook C, Furusawa A and Braunstein S L 2015 Advances in quantum teleportation *Nat. Photon.* **9** 641–52
- [3] Wang X-L, Cai X-D, Su Z-E, Chen M-C, Wu D, Li L, Liu N-L, Lu C-Y and Pan J-W 2015 Quantum teleportation of multiple degrees of freedom of a single photon *Nature* **518** 516–9
- [4] Politi A, Matthews J C F, Thompson M G and O'Brien J L 2009 Integrated quantum photonics *IEEE J. Sel. Top. Quantum Electron.* **15** 1673–84
- [5] Crespi A, Ramponi R, Osellame R, Sansoni L, Bongioanni I, Sciarrino F, Vallone G and Mataloni P 2011 Integrated photonic quantum gates for polarization qubits *Nat. Commun.* **2** 566
- [6] Barends R *et al* 2013 Coherent josephson qubit suitable for scalable quantum integrated circuits *Phys. Rev. Lett.* **111** 080502
- [7] Devoret M H and Schoelkopf R J 2013 Superconducting circuits for quantum information: an outlook *Science* **339** 1169–74
- [8] Zwanenburg F A, Dzurak A S, Morello A, Simmons M Y, Hollenberg L C L, Klimeck G, Rogge S, Coppersmith S N and Eriksson M A 2013 Silicon quantum electronics *Rev. Mod. Phys.* **85** 961
- [9] Cirac J I, Zoller P, Kimble H J and Mabuchi H 1997 Quantum state transfer and entanglement distribution among distant nodes in a quantum network *Phys. Rev. Lett.* **78** 3221

- [10] Acín A, Cirac J I and Lewenstein M 2007 Entanglement percolation in quantum networks *Nat. Phys.* **3** 256–9
- [11] Kimble H J 2008 The quantum internet *Nature* **453** 1023–30
- [12] Perseguers S, Lewenstein M, Acín A and Cirac J I 2010 Quantum random networks *Nat. Phys.* **6** 539–43
- [13] Yin J *et al* 2013 Experimental quasi-single-photon transmission from satellite to Earth *Opt. Express* **21** 20032–40
- [14] Broadbent A, Fitzsimons J and Kashefi E 2009 universal blind quantum computation *50th Annual IEEE Symp. on Foundations of Computer Science, 2009, FOCS'09* (Piscataway, NJ: IEEE) pp 517–26
- [15] Beals R, Brierley S, Gray O, Harrow A W, Kutin S, Linden N, Shepherd D and Stather M 2013 Efficient distributed quantum computing *Proc. R. Soc. A* **469** 20120686
- [16] Barz S, Kashefi E, Broadbent A, Fitzsimons J F, Zeilinger A and Walther P 2012 Demonstration of blind quantum computing *Science* **335** 303–8
- [17] Gutoski G and Watrous J 2007 Toward a general theory of quantum games *Proc. 39th Annual ACM Symp. on Theory of Computing* (ACM) pp 565–574
- [18] Chiribella G, D'Ariano G M and Perinotti P 2009 Theoretical framework for quantum networks *Phys. Rev. A* **80** 022339
- [19] Chiribella G, D'Ariano G M, Perinotti P, Schlingemann D and Werner R 2013 A short impossibility proof of quantum bit commitment *Phys. Lett. A* **377** 1076–87
- [20] Portmann C, Matt C, Maurer U, Renner R and Tackmann B 2015 Causal boxes: quantum information-processing systems closed under composition (arXiv:1512.02240)
- [21] Chiribella G, D'Ariano G M and Perinotti P 2011 Informational derivation of quantum theory *Phys. Rev. A* **84** 012311
- [22] Hardy L 2011 Reformulating and reconstructing quantum theory (arXiv:1104.2066)
- [23] Hardy L 2012 The operator tensor formulation of quantum theory *Phil. Trans. R. Soc. A* **370** 3385–417
- [24] Tucci R R 1995 Quantum bayesian nets *Int. J. Mod. Phys. B* **9** 295–337
- [25] Leifer M S and Poulin D 2008 Quantum graphical models and belief propagation *Ann. Phys., NY* **323** 1899–946
- [26] Leifer M S and Spekkens R W 2013 Towards a formulation of quantum theory as a causally neutral theory of bayesian inference *Phys. Rev. A* **88** 052130
- [27] Henson J, Lal R and Pusey M F 2014 Theory-independent limits on correlations from generalized bayesian networks *New J. Phys.* **16** 113043
- [28] Ried K, Agnew M, Vermeyden L, Janzing D, Spekkens R W and Resch K J 2015 A quantum advantage for inferring causal structure *Nat. Phys.* **11** 414–20
- [29] Chaves R, Majenz C and Gross D 2015 Information-theoretic implications of quantum causal structures *Nat. Commun.* **6** 5766
- [30] Pienaar J *et al* 2015 A graph-separation theorem for quantum causal models *New J. Phys.* **17** 073020
- [31] Costa F and Shrapnel S 2015 Quantum causal modelling arXiv:1512.07106
- [32] Selinger P 2004 Towards a semantics for higher-order quantum computation *Proc. 2nd Int. Workshop on Quantum Programming Languages, TUCS General Publication* vol 33 pp 127–43
- [33] Chiribella G, D'Ariano G M, Perinotti P and Valiron B 2009 Beyond quantum computers (arXiv:0912.0195)
- [34] Chiribella G, D'Ariano G M, Perinotti P and Valiron B 2013 Quantum computations without definite causal structure *Phys. Rev. A* **88** 022318
- [35] Nielsen M and Chuang I 2000 Quantum information and computation *Quantum Information and Computation* (Cambridge: Cambridge University Press)
- [36] Kitaev A Y, Shen A and Vyalı M N 2002 *Classical and Quantum Computation* vol 47 (Providence, RI: American Mathematical Society)
- [37] Chiribella G 2012 Perfect discrimination of no-signalling channels via quantum superposition of causal structures *Phys. Rev. A* **86** 040301
- [38] Colnaghi T, D'Ariano G M, Facchini S and Perinotti P 2012 Quantum computation with programmable connections between gates *Phys. Lett. A* **376** 2940–3
- [39] Araújo M, Fabio Costa and Brukner Č 2014 Computational advantage from quantum-controlled ordering of gates *Phys. Rev. Lett.* **113** 250402
- [40] Hardy L 2009 Quantum gravity computers: on the theory of computation with indefinite causal structure *Quantum Reality, Relativistic Causality, and Closing the Epistemic Circle* (Berlin: Springer) pp 379–401
- [41] Oreshkov O, Costa F and Brukner Č 2012 Quantum correlations with no causal order *Nat. Commun.* **3** 1092
- [42] Baumeler Á and Wolf S 2014 Perfect signaling among three parties violating predefined causal order *IEEE Int. Symp. on Information Theory* pp 526–30
- [43] Morimae T 2014 Acausal measurement-based quantum computing *Phys. Rev. A* **90** 010101
- [44] Baumeler Á and Wolf S 2016 Non-causal computation avoiding the grandfather and information antinomies (arXiv:1601.06522)
- [45] Akibue S, Owari M, Kato G and Muraō M 2016 Entanglement assisted classical communication simulates classical communication without causal order arXiv:1602.08835
- [46] Kitaev A and Watrous J 2000 Parallelization, amplification, and exponential time simulation of quantum interactive proof systems *Proc. 32nd Annual ACM Symposium on Theory of Computing* (ACM) pp 608–617
- [47] Chiribella G, D'Ariano G M and Perinotti P 2008 Quantum circuit architecture *Phys. Rev. Lett.* **101** 060401
- [48] Renner R and Wolf S 2004 Smooth rényi entropy and applications *IEEE Int. Symp. on Information Theory* p 233
- [49] Renner R 2008 Security of quantum key distribution *Int. J. Quantum Inf.* **6** 1–27
- [50] Datta N and Renner R 2009 Smooth entropies and the quantum information spectrum *IEEE Trans. Inf. Theory* **55** 2807–15
- [51] König R, Renner R and Schaffner C 2009 The operational meaning of min- and max-entropy *IEEE Trans. Inf. Theory* **55** 4337–47
- [52] Tomamichel M 2015 *Quantum Information Processing with Finite Resources: Mathematical Foundations* vol 5 (Berlin: Springer)
- [53] Chiribella G, D'Ariano G M and Perinotti P 2008 Transforming quantum operations: quantum supermaps *Europhys. Lett.* **83** 30004
- [54] Brukner Č 2015 Bounding quantum correlations with indefinite causal order *New J. Phys.* **17** 083034
- [55] Kraus K, Böhm A, Dollard J D and Wootters W H 1983 States, effects, and operations fundamental notions of quantum theory
- [56] Choi M-D 1975 Completely positive linear maps on complex matrices *Linear Algebr. Appl.* **10** 285–90
- [57] Yurke B and Stoler D 1992 Einstein-podolsky-rosen effects from independent particle sources *Phys. Rev. Lett.* **68** 1251
- [58] Zukowski M, Zeilinger A, Horne M A and Ekert A K 1993 event-ready-detectors" bell experiment via entanglement swapping *Phys. Rev. Lett.* **71** 4287–90
- [59] Chiribella G, D'Ariano G M and Perinotti P 2010 Probabilistic theories with purification *Phys. Rev. A* **81** 062348
- [60] Cormen T H, Leiserson C E, Rivest R L and Stein C 2001 *Introduction to Algorithms* vol 6 (Cambridge: MIT Press)
- [61] Chiribella G, D'Ariano G M and Perinotti P 2008 Memory effects in quantum channel discrimination *Phys. Rev. Lett.* **101** 180501

- [62] Ziman M 2008 Process positive-operator-valued measure: a mathematical framework for the description of process tomography experiments *Phys. Rev. A* **77** 062112
- [63] Bisio A, Chiribella G, D'Ariano G M, Facchini S and Perinotti P 2010 Optimal quantum learning of a unitary transformation *Phys. Rev. A* **81** 032324
- [64] Giovannetti V, Lloyd S and Maccone L 2006 Quantum metrology *Phys. Rev. Lett.* **96** 010401
- [65] Giovannetti V, Lloyd S and Maccone L 2011 Advances in quantum metrology *Nat. Photon.* **5** 222–9
- [66] Gutoski G 2012 On a measure of distance for quantum strategies *J. Math. Phys.* **53** 032202
- [67] Jenčová A and Plávala M 2016 Conditions for optimal input states for discrimination of quantum channels (arXiv:1603.01437)
- [68] Sedláč M, Reitzner D, Chiribella G and Ziman M 2016 Incompatible measurements on quantum causal networks *Phys. Rev. A* **93** 052323
- [69] Bisio A, Chiribella G, D'Ariano G M, Facchini S and Perinotti P 2009 Optimal quantum tomography of states, measurements, and transformations *Phys. Rev. Lett.* **102** 010404
- [70] Pollock F A, Rodríguez-Rosario C, Frauenheim T, Paternostro M and Modi K 2015 Complete framework for efficient characterisation of non-markovian processes (arXiv:1512.00589)
- [71] Modi K *et al* 2012 Unification of witnessing initial system–environment correlations and witnessing non-markovianity *Europhys. Lett.* **99** 20010
- [72] Deutsch D 1989 Quantum computational networks *Proc. R. Soc. A* **425** 73–90
- [73] Marvian I and Lloyd S 2016 universal quantum emulator arXiv:1606.02734)
- [74] Chiribella G, D'Ariano G M and Perinotti P 2008 Optimal cloning of unitary transformation *Phys. Rev. Lett.* **101** 180504
- [75] Dür W, Sekatski P and Skotiniotis M 2015 Deterministic superreplication of one-parameter unitary transformations *Phys. Rev. Lett.* **114** 120503
- [76] Chiribella G, Yang Y and Huang C 2015 universal superreplication of unitary gates *Phys. Rev. Lett.* **114** 120504
- [77] Mičuda M, Stárek R, Straka I, Miková M, Sedláč M, Ježek M and Fiurášek J 2016 Experimental replication of single-qubit quantum phase gates *Phys. Rev. A* **93** 052318
- [78] Chiribella G and Yang Y 2016 Quantum superreplication of states and gates *Front. Phys.* **11** 1–9
- [79] Grover L K 1996 A fast quantum mechanical algorithm for database search *Proc. Twenty-eighth Annual ACM Symp. on Theory of Computing* pp 212–9
- [80] Watrous J 2011 *Lecture Notes on Semidefinite Programming* (University of Waterloo)
- [81] Boyd S and Vandenberghe L 2004 *Convex Optimization* (Cambridge: Cambridge University Press)
- [82] Datta N 2009 Min- and max-relative entropies and a new entanglement monotone *Information Theory, IEEE Transactions on* **55** 2816–26
- [83] Brandao F G S L and Datta N 2011 One-shot rates for entanglement manipulation under non-entangling maps *IEEE Trans. Inf. Theory* **57** 1754–60
- [84] Horodecki M and Oppenheim J 2013 Fundamental limitations for quantum and nanoscale thermodynamics *Nat. Commun.* **4** 2059
- [85] Brandão F, Horodecki M, Ng N, Oppenheim J and Wehner S 2015 The second laws of quantum thermodynamics *Proc. Natl Acad. Sci.* **112** 3275–9
- [86] Datta N, Mosonyi M, Hsieh M-H and Brandao F G S L 2013 A smooth entropy approach to quantum hypothesis testing and the classical capacity of quantum channels *IEEE Trans. Inf. Theory* **59** 8014–26
- [87] Chiribella G and Xie J 2013 Optimal design and quantum benchmarks for coherent state amplifiers *Phys. Rev. Lett.* **110** 213602
- [88] Jenčová A 2014 Base norms and discrimination of generalized quantum channels *J. Math. Phys.* **55** 022201
- [89] Buscemi F and Datta N 2016 Equivalence between divisibility and monotonic decrease of information in classical and quantum stochastic processes *Phys. Rev. A* **93** 012101
- [90] Bae J and Chruscinski D 2016 Operational characterization of divisibility of dynamical maps (arXiv:1601.05522)
- [91] Beckman D, Gottesman D, Nielsen M A and Preskill J 2001 Causal and localizable quantum operations *Phys. Rev. A* **64** 052309
- [92] Piani M, Horodecki M, Horodecki P and Horodecki R 2006 Properties of quantum nonsignaling boxes *Phys. Rev. A* **74** 012305
- [93] D'Ariano G M, Facchini S and Perinotti P 2011 No signaling, entanglement breaking, and localizability in bipartite channels *Phys. Rev. Lett.* **106** 010501
- [94] Bennett C, Bernstein H, Popescu S and Schumacher B 1996 Concentrating partial entanglement by local operations *Phys. Rev. A* **53** 2046–52
- [95] Popescu S and Rohrlich D 1997 Thermodynamics and the measure of entanglement *Phys. Rev. A* **56** R3319
- [96] Vedral V, Plenio M B, Rippin M A and Knight P L 1997 Quantifying entanglement *Phys. Rev. Lett.* **78** 2275
- [97] Friis N, Dunjko V, Dür W and Briegel H J 2014 Implementing quantum control for unknown subroutines *Phys. Rev. A* **89** 030303
- [98] Araújo M, Feix A, Costa F and Brukner Č 2014 Quantum circuits cannot control unknown operations *New J. Phys.* **16** 093026
- [99] Nakayama S, Soeda A and Murao M 2015 Quantum algorithm for universal implementation of the projective measurement of energy *Phys. Rev. Lett.* **114** 190501
- [100] Bisio A, Dall'Arno M and Perinotti P 2015 The quantum conditional statement arXiv:1509.01062)
- [101] Raginsky M 2001 A fidelity measure for quantum channels *Phys. Lett. A* **290** 11–8
- [102] Nielsen M A 2002 A simple formula for the average gate fidelity of a quantum dynamical operation *Phys. Lett. A* **303** 249–52
- [103] Horodecki M, Horodecki P and Horodecki R 1999 General teleportation channel, singlet fraction, and quasidistillation *Phys. Rev. A* **60** 1888–98
- [104] Nielsen M and Chuang I 1997 Programmable quantum gate arrays *Phys. Rev. Lett.* **79** 321
- [105] Bužek V, Hillery M and Werner R F 1999 Optimal manipulations with qubits: universal-not gate *Phys. Rev. A* **60** R2626
- [106] Rungta P, Bužek V, Caves C M, Hillery M and Milburn G J 2001 universal state inversion and concurrence in arbitrary dimensions *Phys. Rev. A* **64** 042315
- [107] Chiribella G, D'Ariano G M and Sacchi M F 2005 Optimal estimation of group transformations using entanglement *Phys. Rev. A* **72** 042338
- [108] Horodecki P and Ekert A 2002 Method for direct detection of quantum entanglement *Phys. Rev. Lett.* **89** 127902
- [109] Horodecki P 2003 From limits of quantum operations to multicopy entanglement witnesses and state-spectrum estimation *Phys. Rev. A* **68** 052101
- [110] Buscemi F, D'Ariano G M, Perinotti P and Sacchi M F 2003 Optimal realization of the transposition maps *Phys. Lett. A* **314** 374–9
- [111] Kalev A and Bae J 2013 Optimal approximate transpose map via quantum designs and its applications to entanglement detection *Phys. Rev. A* **87** 062314

- [112] Lim H-T, Kim Y-S, Ra Y-S, Bae J and Kim Y-H 2011 Experimental realization of an approximate partial transpose for photonic two-qubit systems *Phys. Rev. Lett.* **107** 160401
- [113] Scott A James 2008 Optimizing quantum process tomography with unitary 2-designs *J. Phys. A: Math. Theor.* **41** 055308
- [114] Ringbauer M, Wood C J, Modi K, Gilchrist A, White A G and Fedrizzi A 2015 Characterizing quantum dynamics with initial system-environment correlations *Phys. Rev. Lett.* **114** 090402
- [115] Aïmeur E, Brassard G and Gambs S 2006 Machine learning in a quantum world *Advances in Artificial Intelligence* (Berlin: Springer) pp 431–42
- [116] Bisio A, Chiribella G, D’Ariano G M, Facchini S and Perinotti P 2010 Optimal quantum learning of a unitary transformation *Phys. Rev. A* **81** 032324
- [117] Paparo G D, Dunjko V, Makmal A, Martin-Delgado M A and Briegel H J 2014 Quantum speedup for active learning agents *Phys. Rev. X* **4** 031002
- [118] Chiribella G and Yang Y 2015 arXiv:1502.00259