Optimal quantum networks and one-shot entropies

Giulio Chiribella and Daniel Ebler
Department of Computer Science, The University of Hong Kong, Pukfulam Road, Hong Kong

Abstract. We present a general-purpose 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 tests set up by a verifier. We show that the optimal performance can be quantified by the max relative entropy between a suitable operator, associated to the test, and the set of operators associated to non-informative tests. Building on the connection with the max relative entropy, we extend the notion of conditional min-entropy from quantum states to quantum causal networks, providing a measure of the total correlations generated by ordered sequences of quantum channels with memory. 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 an application to the maximization of the winning probability in a non-causal quantum game.
1. Introduction

Quantum technologies are laying the foundations of a new generation of information-processing devices, with enhanced computational power security and sensitivity. This process is leading to advances in engineering and to the exploration of the ultimate limits to the performances of information-processing devices. An operational way to assess the performance of a device is to present it with inputs and to perform measurements on the outputs. For example, suppose that a device is designed to search for an item in a database. A possible test consists in putting the item in a known, randomly chosen position and checking if the device returns the correct answer. The operational approach to quantifying performances applies not only to individual devices, but also to networks consisting of multiple interconnected devices. The study of quantum networks is becoming increasingly more important with advancements in quantum communication [1] and, more generally, in the integration of quantum hardware [2, 3, 4], which pave the way towards the realization of new information systems, such as satellite quantum communication [5], distributed quantum computing [6, 7, 8], and quantum internet systems [9].

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 the computation is implemented by a sequence of gates. More recently, researchers have started to investigate more general situations, where the causal order is not definite [10, 11, 12, 13, 14, 15, 16, 17, 18, 19]. Allowing for the gates to act in an indefinite causal order offers several information-processing advantages with respect to the conventional scenario, including an improved ability to win non-local games [12, 20, 21] and to distinguish quantum channels [13, 18]. Some of these advantages can be already demonstrated in table-top experiments, like the recent experiment of Procopio et al [22], which demonstrated the enhanced the distinguishability of quantum channels due to the quantum superposition causal structures [13]. Other advantages are not directly accessible with known physics, but may become accessible in exotic quantum gravity scenarios where the causal structure becomes a dynamical variable. In all these situations, optimizing quantum networks is important, for at least three reasons: First, proving an advantage from indefinite causal order requires finding the ultimate limits achievable with a definite causal order. Second, finding the maximum advantage offered by non-causal networks is essential for assessing the advantages of the new, non-causal models of computation and communication. 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 general approach to the optimization of quantum networks. We start by analyzing scenarios with definite causal order. We choose 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 [23], wherein they are used to model the interaction between a prover and a verifier. The input-output behavior of a quantum causal network is in the framework of quantum combs [24, 25], which associates a positive operator to any given sequence of quantum operations. In this framework, the optimization is a semidefinite program. 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 [26, 27, 28, 29] (see also the monography [30]). 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 [28, 29], 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 [31, 15, 19]. Some other networks are not built by linking up individual devices [12]. 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 behaviour in terms of the behaviour 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 takes as input the operations taking place in local laboratories and returns as output an operation, as Figure 1. Maps that transform quantum operations were originally introduced in the causal scenario in Refs. [32, 31] and later generalized to the case of networks with indefinite causal structure [31, 12, 14]. 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 semidefinite programs. In this case, we find that the maximum score can be expressed in terms of a max relative entropy, which we call the max relative entropy of signalling, as it quantifies the deviation from the set of no-signalling channels.

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, Costa, and Brukner [12]. In this case, we fix the operations performed by the players and search for the non-causal network that offers the largest advantage. This optimization problem was recently solved by Brukner [20],
who showed the optimality of the network presented in Ref. [12]. Using our semidefinite programming technique, we provide a significantly shorter optimality proof.

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 semidefinite programming 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). In Section 6 we extend the results to quantum networks without predefined causal structure. Our techniques are illustrated in Section 7, where we present applications to the tasks of inverting unknown evolutions, simulating charge conjugation, controlling unitary gates, and to the study of a quantum non-causal game. Finally, the conclusions are drawn in Section 8.

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 [33] 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 $\mathcal{H}_A$ to operators on the output Hilbert
Optimal quantum networks and one-shot entropies

The diagrammatic notation

\[
\begin{array}{c}
A \quad C \\
B
\end{array}
\]  

(1)

will often be convenient. We say that the quantum operation \( C \) in the above diagram is of type \( A \rightarrow B \). We include the special case where system \( A \) is the trivial system \( I \), corresponding to the one-dimensional Hilbert space \( \mathcal{H}_I = \mathbb{C} \). Quantum operations with trivial input correspond will be represented as

\[
\begin{array}{c}
B \\
\rho
\end{array} := \begin{array}{c}
I \\
\rho
\end{array}
\]  

(2)

while quantum operations with trivial output will be represented as

\[
\begin{array}{c}
A \\
P
\end{array} := \begin{array}{c}
P \\
I
\end{array}
\]  

(3)

Deterministic processes (processes happening with unit probability) correspond to trace preserving quantum operations. Mathematically, they are represented by completely positive trace preserving (CPTP) maps, also known as quantum channels.

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

\[
C = (C \otimes I_0)(|I\rangle \langle I|)
\]  

(4)

where \( I_0 \) denotes the identity map on \( \text{Lin}(\mathcal{H}_0) \) and \( |I\rangle \) is the unnormalized maximally entangled state \( |I\rangle = \sum_i |i\rangle |i\rangle \in \mathcal{H}_0 \otimes \mathcal{H}_0 \). The operator \( C \) is known as the Choi operator [34].

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

\[
\text{Tr}_1[Q] \leq I_0,
\]

where \( \text{Tr}_1 \) 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 \( C \in \text{Lin}(\mathcal{H}_1 \otimes \mathcal{H}_0) \) corresponds to a quantum channel if and only if it satisfies the condition

\[
\text{Tr}_1[C] = I_0.
\]

2.2. The link product

Two quantum operations can be connected with each other, as long as the output of one operation matches the input of the other. At the level of Choi operators, the connection is implemented by the operation of link product [24], denoted as \( \ast \). To define the link product, it is convenient to introduce the following shorthand notation: if \( A \) is
an operator on $\mathcal{H}_{XY}$ and $B$ is an operator on $H_{YX}$, when we use the notation $AB$ for the product
\begin{equation}
AB := (A \otimes I_Z)(I_X \otimes B).
\end{equation}
With this notation, the link product of $A$ and $B$ is the operator $A \ast B$ defined as
\begin{equation}
A \ast B := \text{Tr}_Y [A^{T_Y} B]
\end{equation}
where $A^{T_Y}$ denotes the partial transpose of $A$ with respect to the Hilbert space $\mathcal{H}_Y$. Note that the definition of the link product presupposes that the Hilbert spaces have been labelled: 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 $B$.

It is not hard to see that the link product of two positive operators is positive [24]. Moreover, the link product is associative, namely
\begin{equation}
A \ast (B \ast C) = (A \ast B) \ast C,
\end{equation}
for all operators $A$, $B$, and $C$. Finally, the link product is commutative, up to re-ordering of the Hilbert spaces: in formula,
\begin{equation}
A \ast B \simeq B \ast A,
\end{equation}
having used the notation $A \ast B \simeq B \ast A$ to mean $A \ast B = \text{SWAP}_{XZ} (B \ast A) \text{SWAP}_{XZ}$, where $\text{SWAP}_{XZ}$ is the unitary operator that swaps the spaces $\mathcal{H}_0$ and $\mathcal{H}_1$. From now on we will omit the swaps, implicitly understanding that the Hilbert spaces have been reordered in the right way.

Using the above notation, we have the following

**Proposition 1** ([24]). Let $A$ be a quantum operation transforming operators on $\mathcal{H}_0$ to operators on $\mathcal{H}_1$, let $B$ be a quantum operation transforming operators on $\mathcal{H}_1$ to operators on $\mathcal{H}_2$, and let $C = BA$ be the quantum operation resulting from the composition of $A$ and $B$. Then, one has
\begin{equation}
C := A \ast B,
\end{equation}
where $A$, $B$, and $C$ are the Choi operators of $A$, $B$, and $C$, respectively.

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

### 2.3. 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—cf. Figure 2. For every DAG, one can always define a total ordering of the vertices, through the procedure of topological sorting [35]. Using this fact, one
can always represent the a quantum causal network as an ordered sequence of quantum
device, such as
\[ A_{1}^{\text{in}} \rightarrow C_{1} \rightarrow A_{2}^{\text{out}} \rightarrow C_{2} \rightarrow \ldots \rightarrow A_{N}^{\text{in}} \rightarrow C_{N} \rightarrow A_{N}^{\text{out}}, \]  
\( (7) \)
where \( A_{j}^{\text{in}} \) (\( A_{j}^{\text{out}} \)) 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 \( C_{1}, C_{2}, \ldots, C_{N} \), then the network has Choi operator
\[ C = C_{1} \ast C_{2} \ast C_{3} \ast \ldots \ast C_{N}. \]  
\( (8) \)

The Choi operator of a deterministic network is called a quantum comb \([24, 31]\). The quantum comb \( C \) is an positive operator on \( \bigotimes_{j=1}^{N} (H_{j}^{\text{out}} \otimes H_{j}^{\text{in}}) \), where \( H_{j}^{\text{in}} \) (\( H_{j}^{\text{out}} \)) is the Hilbert space of system \( A_{j}^{\text{in}} \) (\( A_{j}^{\text{out}} \)). The condition for a positive operator \( C \) to be a quantum comb is that \( C \) satisfies the linear constraints
\[ \text{Tr}_{A_{n}^{\text{out}}}[C^{(n)}] = I_{A_{n}^{\text{in}}} \otimes C^{(n-1)} \quad \forall n \in \{1, \ldots, N\}, \]  
\( (9) \)
where \( \text{Tr}_{A} \) is the partial trace over the Hilbert space \( H_{A} \), \( C^{(n)} \) is a suitable operator on \( \bigotimes_{j=1}^{n} (H_{j}^{\text{out}} \otimes H_{j}^{\text{in}}) \), \( C^{(N)} := C \) and \( C^{(0)} := 1 \). Physically, the positive operator \( C^{(n)} \) represents the subnetwork transforming the first \( n-1 \) inputs to the first \( n-1 \) outputs. We denote by
\[ \text{Comb}\left(A_{1}^{\text{in}} \rightarrow A_{1}^{\text{out}}, A_{2}^{\text{in}} \rightarrow A_{2}^{\text{out}}, \ldots, A_{N}^{\text{in}} \rightarrow A_{N}^{\text{out}}\right) \]
the set of positive semidefinite operators satisfying the constraint (9). When there is no ambiguity, we will simply write \( \text{Comb} \).

2.4. Quantum testers and the generalized Born rule

So far we considered deterministic networks. However, it is also useful to consider networks containing measurement devices, such as the following
\[ \rho \rightarrow A_{1}^{\text{in}} \rightarrow D_{1} \rightarrow \ldots \rightarrow D_{N-1} \rightarrow A_{N}^{\text{in}} \rightarrow A_{N}^{\text{out}} \rightarrow \{P_{x}\}_{x \in X}, \]  
\( (10) \)
where $\rho$ is a quantum state, $(D_1, \ldots, D_{N-1})$ is a sequence of quantum channels, and \( \{P_x\}_{x \in X} \) is a positive operator-valued measure (POVM).

We associate every outcome $x$ with the Choi operator $T_x$ defined by

$$T_x := \rho \ast D_1 \ast D_2 \ast \cdots \ast D_{N-1} \ast P_x.$$ 

The set of operators $T = \{T_x\}_{x \in X}$ is called quantum tester [36, 31] and describes all the possible ways in which the network (10) can respond when connected with external devices. Quantum testers can be characterized as follows:

**Proposition 2** ([36]). Let $T$ be a collection of positive operators on $\bigotimes_{j=1}^N (\mathcal{H}_{\text{out}}^j \otimes \mathcal{H}_{\text{in}}^j)$. $T$ is a quantum tester if and only if

$$\sum_{x \in X} T_x = I_{A_{\text{out}}^N} \otimes \Gamma^{(N)},$$

$$\text{Tr}_{A_{\text{in}}^n} \left[ \Gamma^{(n)} \right] = I_{A_{\text{out}}^{n-1}} \otimes \Gamma^{(n-1)}, \quad n = 2, \ldots, N,$$

$$\text{Tr}_{A_{\text{in}}^1} \left[ \Gamma^{(1)} \right] = 1,$$

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

Networks of the type (10) can be used to probe networks of the type (7), by connecting the open wires as follows:

$$\rho \xrightarrow{A_{\text{in}}^1} D_1 \xrightarrow{A_{\text{out}}^1} \cdots \xrightarrow{A_{\text{in}}^{N-1}} D_{N-1} \xrightarrow{A_{\text{out}}^{N-1}} \xrightarrow{A_{\text{in}}^N} P_x \xrightarrow{A_{\text{out}}^N} \{P_x\}_{x \in X}.$$  

(12)

When the two networks are connected, the measurement produces one of the outcomes in the set $X$. The probability of the outcome $x$ is given by

$$p_x = \rho \ast C_1 \ast D_1 \ast C_2 \ast D_2 \ast \cdots \ast D_{N-1} \ast C_N \ast P_x$$

$$= \left( \rho \ast D_1 \ast D_2 \ast \cdots \ast D_{N-1} \ast P_x \right) \ast \left( C_1 \ast C_2 \ast \cdots \ast C_N \right)$$

$$= T_x \ast C$$

$$= \text{Tr} \left[ T_x C^T \right],$$

where $\{T_x\}_{x \in X}$ is the tester, $C$ is the Choi operator of the tested network, and $C^T$ is the transpose of $C$. We call Eq. (13) the generalized Born rule [36, 31].

**2.5. Assessing the performance of a quantum network**

Suppose that we are given access to an unknown quantum network, but we have no information about its internal functioning. Our goal is to assess how well the network fares in a desired task. An operational way to assess the performance is to set up an experiment where the unknown network is connected to another network consisting of measuring devices. Then, we can assign a “score” to every outcome, so that the expected score can serve as a performance indicator.
Precisely, 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. Assigning score $\omega_x$ to the outcome $x$, we have that the average score is given by

$$\omega = \sum_x \omega_x (T_x * C) = \Omega * C \quad \Omega := \sum_x \omega_x T_x.$$  \hfill (14)

Note that the notion of performance associated to our test is completely characterized by the operator $\Omega$. We call $\Omega$ the performance operator.

For a given performance operator $\Omega$, the maximum expected score is given by

$$\omega_{\text{max}} := \max_{C \in \text{Comb}} \Omega * C = \max_{C \in \text{Comb}} \text{Tr}[\Omega C].$$  \hfill (15)

The second equality comes from the fact that the set of quantum combs is closed under transposition and, therefore, we can omit the transpose in Eq. (13). Using the notation $\langle A, B \rangle = \text{Tr}[AB]$, we can express the maximum score as

$$\omega_{\text{max}} := \max_{C \in \text{Comb}} \langle \Omega, C \rangle.$$  \hfill (16)

The above equation shows that the search for the maximum score is a semidefinite program. In the next section we will review the basic tools needed to address it.

3. Semidefinite programming and the max relative entropy

3.1. Basic facts about semidefinite programming

Here we review the background about semidefinite programming. For further details, we refer the reader to Watrous’ lecture notes [37].

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

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

A semidefinite program is associated to an optimization problem in the standard form

$$\text{maximize } \langle A, X \rangle$$
subject to $\phi(X) = B$
$X \geq 0$.

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

$$\text{minimize } \langle B, Y \rangle$$
subject to $\phi^\dagger(Y) \geq A$
$Y \in \text{Herm}(\mathcal{Y})$.

(17)
where $\phi^\dagger$ is the adjoint of $\phi$, 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 are related by duality. Let
\[
\omega_{\text{primal}} := \sup \langle A, X \rangle \quad \text{and} \quad \omega_{\text{dual}} := \inf \langle B, Y \rangle
\]
be the values of the primal and dual problem, respectively. For every semidefinite program, 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 [38]. In particular, we will use the following

**Proposition 3.** Let $(\phi, A, B)$ be a semidefinite program. 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. [37].

### 3.2. The max relative entropy

An important quantity in quantum information theory is the max relative entropy, introduced by Datta in Ref. [39]:

**Definition 2.** Let $A$ and $B$ be two positive operators on $\mathcal{X}$. The max entropy of $A$ relative to $B$ is given by
\[
D_{\text{max}}(A \| B) := -\log \max\{w \mid w A \leq B\}, \quad (19)
\]
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 $S \subset \text{Herm}(\mathcal{X})$ be a set of positive operators. The max entropy of $A$ relative the set $S$, denoted as $D_{\text{max}}(A \| S)$, is the quantity defined by
\[
D_{\text{max}}(A \| S) := \inf_{B \in S} D_{\text{max}}(A \| B). \quad (20)
\]

The max relative entropy between a quantum state and a set of quantum states plays a central role in entanglement theory [40] and thermodynamics [41, 42]. In this paper we will extend the application of the max relative entropy to dynamical scenarios, where $S$ represents a set of quantum networks. In fact, it turns out that the max relative entropy plays a central role in a large class of semidefinite programs.
3.3. From semidefinite programs to the max relative entropy

Here we provide a general bound on the primal value of an arbitrary semidefinite program. The bound can always be attained its value can be expressed in terms of a max relative entropy whenever the operator $A$ in Eq. (17) is positive. To state the result, we need some basic notation. 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 $\mathcal{S} \subseteq \mathcal{V}$, we define the dual affine space $\overline{\mathcal{S}}$ as

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

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

Given a semidefinite program $(A, B, \phi)$, we define the primal affine space as

$$\mathcal{S} := \{ X \in \text{Herm}(\mathcal{X}) | \phi(X) = B \}.$$ 

In words, $\mathcal{S}$ is the set of operators that satisfy the equality constraint of the primal problem. The dual affine space is given by

$$\overline{\mathcal{S}} = \{ \Gamma \in \text{Herm}(\mathcal{X}) | \langle \Gamma, X \rangle = 1, \forall X \in \mathcal{S} \},$$

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

**Theorem 1.** Let $(\phi, A, B)$ be a semidefinite program. The optimal solution of the primal problem is upper bounded as

$$\omega_{\text{primal}} \leq \inf_{\Gamma \in \overline{\mathcal{S}}} \min \{ \lambda \in \mathbb{R} | \lambda \Gamma \geq A \},$$

where $\overline{\mathcal{S}}$ is the dual affine space defined in Eq. (22). If $\mathcal{S}$ contains a positive operator and $\overline{\mathcal{S}}$ contains a strictly positive operator, then Eqs. (23) and (24) hold with the equality.

If, in addition, the operator $A$ is positive, then one has the expression

$$\omega_{\text{primal}} = 2^{D_{\text{max}}(A \parallel \mathcal{S}_+)}.$$ 

with $\mathcal{S}_+ := \{ \Gamma \in \mathcal{S} | \Gamma \geq 0 \}$.

The proof can be found in appendix Appendix A.

We call the quantity $D(A \parallel \mathcal{S}_+)$ the max divergence from normalization. This quantity measures how much the operator $A$ deviates from the set of functionals that are normalized on every element of the primal set. The connection between semidefinite programming and the max relative entropy has been previously known in the special case where the task is to optimize quantum channels [29, 43]. A special case of Theorem 1 was derived by one of us [44] in the context of parameter estimation. A related result was obtained by Jenčová in the framework of base norms [45].

In the next sections we will elaborate on the physical meaning of Theorem 1 and we will apply it to the optimization of quantum networks, both with definite and indefinite causal structure.
4. Optimizing quantum causal networks

Consider the scenario where a network of quantum devices is required to perform a desired task, such as implementing a distributed algorithm. To evaluate the performance of the network, one can set up a testing procedure where the network is presented with a set of possible inputs and its response is probed by a measurement on the outputs.

4.1. The dual networks

The mathematical description of the test is provided by a performance operator \( \Omega \), acting on the Hilbert spaces of the input and output systems of the network. The maximum performance achieved by an arbitrary causal network is determined by the following

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

\[
\begin{align*}
A_{1}^{\text{in}} & \xrightarrow{C_{1}} A_{1}^{\text{out}} \\
A_{2}^{\text{in}} & \xrightarrow{C_{2}} A_{2}^{\text{out}} \\
& \vdots
\end{align*}
\]

Then, \( \omega_{\text{max}} \) is given by

\[
\omega_{\text{max}} = \min_{\Gamma \in \text{DualComb}} \min \{ \lambda \in \mathbb{R} \mid \lambda \Gamma \geq \Omega \},
\]

where DualComb denotes the set of dual combs, that is, positive operators \( \Gamma \) representing networks of the form

\[
\begin{align*}
\sigma & \xrightarrow{A_{1}^{\text{in}}} \mathcal{E}_{1} \\
& \xrightarrow{A_{2}^{\text{in}}} \mathcal{E}_{2} \\
& \vdots
\end{align*}
\]

where \( \sigma \) is a quantum state, \( (\mathcal{E}_{1}, \mathcal{E}_{2}, \ldots, \mathcal{E}_{N-1}) \) 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 \( \Gamma \) satisfying the linear constraint

\[
\begin{align*}
\Gamma &= I_{A_{N}^{\text{out}}} \otimes \Gamma^{(N)} \\
\text{Tr}_{A_{N}^{\text{in}}} [\Gamma^{(N)}] &= I_{A_{N-1}^{\text{out}}} \otimes \Gamma^{(N-1)}, \quad n = 2, \ldots, N \\
\text{Tr}_{A_{1}^{\text{in}}} [\Gamma^{(1)}] &= 1,
\end{align*}
\]

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 \( \Omega \) is positive, the maximum performance can be expressed as

\[
\omega_{\text{max}} = 2^{D_{\text{max}}(\Omega || \text{DualComb})}
\]

The proof can be found in appendix Appendix B.

Theorem 2 has an intuitive interpretation. The dual networks (27) and the primal networks (25) “deterministically complement each other”: when two such networks are
Optimal quantum networks and one-shot entropies

Connected, one obtains the closed circuit

\[
\begin{array}{cccccc}
\sigma & A_{1i} & C_1 & A_{1o} & \cdots & A_{N-1i} & C_{N-1} & A_{N-1o} & \cdots & A_{Ni} & C_i & A_{Io} & \cdots & A_{N0} & C_N & A_{N0}^\text{out} & \text{Tr} \\
\end{array}
\]

which yields no information about the primal network and also makes it 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 \(\Omega\) 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. In this case, the performance operator is given by \(\Omega = T_{\text{yes}}\) and the probability to pass the test is given by

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

having used Eq. (26) with \(\lambda\) 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 \(T_{\text{yes}}\) under some dual comb \(\Gamma\).

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 \(\Gamma\). Then, define \(T'_{\text{yes}} := wT_{\text{yes}}\) and \(T'_{\text{no}} := \Gamma - T'_{\text{no}}\). By construction, the operators \(\{T'_{\text{yes}}, T'_{\text{no}}\}\) form a tester: they are positive and their sum satisfies Eq. (11).

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 \(T^*_{\text{yes}}\) is proportional to \(T_{\text{yes}}\). 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

\[
\langle T^*_{\text{yes}}, C^* \rangle = w^* \langle T_{\text{yes}}, C^* \rangle = 1.
\]
5. The conditional min-entropy of quantum causal networks

Theorem 2 allows us to extend the notion of conditional min-entropy [28] from quantum states to quantum causal networks. For a quantum state $\rho \in \text{St}(AB)$, the conditional min-entropy of system $A$, conditional on system $B$, is defined as [28]

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

(32)

König, Renner, and Schaffner [29] clarified the operational meaning of $H_{\text{min}}(A|B)_\sigma$ in terms of the following task: given the state $\rho_{AB}$, 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 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 $\rho$ 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

\[
\begin{array}{c}
\rho \\
\downarrow \\
| \begin{array}{c}
A \\
B \\
\end{array} \\
\downarrow \\
\begin{array}{c}
P_{\text{yes}} \\
A \\
\end{array}
\end{array}
\]

(33)

whose Choi operator is given by

$$T_{\text{yes}} := \rho \ast P_{\text{yes}} = \sigma / d_A.$$

Hence, the probability that the channel passes the test is

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

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

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

(34)

thus providing an operational interpretation to the conditional min-entropy $H_{\text{min}}(A|B)_\rho$. We now extend the notion of conditional min-entropy from states to networks with a definite causal structure. This can be done in two ways, illustrated in the following subsections.

5.1. The conditional min-entropy of a test

The first way to extend the notion of conditional min-entropy is to regard $H_{\text{min}}(A|B)_\rho$ as a quantity associated to a test—specifically, the test depicted in Eq. (33). From this point of view, it is natural to extend the definition to tests consisting of multiple time steps, as follows
Definition 4. Let $T_{\text{yes}}$ be a positive operator associated to a test of the form

$$\rho_{A_1^{\text{in}} A_1^{\text{out}}} D_1 \cdots D_{N-1} A_N^{\text{in}} A_N^{\text{out}} P_{\text{yes}}.$$ 

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

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

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

$$\begin{array}{cccccccccc}
\sigma & A_1^{\text{in}} & A_1^{\text{out}} & \mathcal{E}_1 & A_2^{\text{in}} & A_2^{\text{out}} & \cdots & \mathcal{E}_{N-1} & A_N^{\text{in}} & A_N^{\text{out}} \\
\end{array}.$$ (36)

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

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

$$\begin{array}{cccccccc}
A_1^{\text{in}} & C_1 & A_2^{\text{in}} & C_2 & \cdots & C_N & A_N^{\text{out}} \\
\end{array}$$

passes the test with operator $T_{\text{yes}}$ is

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

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

Another 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 $\rho_{AB}$ by acting on system $B$ alone. A natural generalization is to consider a quantum network of the form

$$\begin{array}{cccccccccc}
B_1^{\text{in}} & D_1^{\text{out}} & B_2^{\text{in}} & D_2^{\text{out}} & \cdots & D_N^{\text{out}} & B_N^{\text{out}} \\
\end{array}.$$ (37)

and to ask how much the first $N - 1$ time steps contribute to build up correlations with the output system at the last step. To generate the correlations, we can connect the network (37) with a second network that processes all input/output systems before $B_N^{\text{out}}$, such as the following

$$\begin{array}{cccccccccc}
\sigma & B_1^{\text{in}} & B_1^{\text{out}} & \mathcal{E}_1 & B_2^{\text{in}} & B_2^{\text{out}} & \cdots & \mathcal{E}_{N-1} & B_N^{\text{in}} & B_N^{\text{out}} \\
\end{array}.$$ (38)
When the two networks are connected, we obtain the bipartite state

$$\sigma_{B} \in \mathcal{D}_{1} \otimes \mathcal{D}_{2} \otimes \ldots \otimes \mathcal{D}_{N-1} \otimes \mathcal{D}_{N}$$

(39)

A measure of correlation is then provided by the fidelity between the state (39) and the maximally entangled state. Explicitly, the fidelity is given by

$$F = \text{Tr} [R C^T]$$

with

$$R := \mathcal{D}_{1} \ast \ldots \ast \mathcal{D}_{N} \quad \text{and} \quad C := \rho \ast \mathcal{E}_{1} \ast \ldots \ast \mathcal{E}_{N-1}.$$

The maximum of the fidelity over all networks of the form (38) can be computed via Theorem 2, which yields the expression

$$F_{\text{max}} = 2^{-H(t_{N} | t_{1} \ldots t_{N-1})_{R}} d_{B_{N}^{\text{out}}}$$

(40)

with $H(t_{N} | t_{1} \ldots t_{N-1})_{R}$ defined as follows

**Definition 5.** Let $R \in \text{Comb}(B_{1}^{\text{in}} \rightarrow B_{1}^{\text{out}}, \ldots, B_{N}^{\text{in}} \rightarrow B_{N}^{\text{out}})$ 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

$$H_{\text{min}}(t_{N} | t_{1} \ldots t_{N-1})_{R} := -\log \left[ \min_{\Gamma_{t_{1} \ldots t_{N-1}}} \min_{\lambda \in \mathbb{R}} \{ \lambda : \lambda \left( I_{N} \otimes \Gamma_{t_{1} \ldots t_{N-1}} \right) \geq R \} \right],$$

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

The above definition gives back the conditional min-entropy $H_{\text{min}}(A | B)_{\rho}$ in the special case $N = 2$, $B_{1}^{\text{in}} = B_{2}^{\text{in}} = I$, $B_{1}^{\text{out}} = B$, $B_{2}^{\text{out}} = A$, and $R = \rho$. Operationally, this choice of settings corresponds to a network of the form

$$\rho \xrightarrow{A} I \xrightarrow{A}$$

(42)

where $I$ is the identity transformation.

### 6. Non-causal networks

In the previous sections we restricted our focus to causal networks. We will now move to the general scenario, including networks that are not compatible with any pre-defined causal order [10, 11, 12, 13, 14, 15, 16, 18, 17]. Some of these networks arise when multiple quantum devices are connected in a way that is controlled by the state of a quantum system [11, 15]. Some other networks are not built from individual devices [12, 14] 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.
6.1. 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, Costa, and Brukner [12]. 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 the first laboratory has been prepared jointly with system $B_1$ in the second laboratory, and that, at a later time, system $A_1'$ and system $B_1'$ are discarded. Indulging into a bit of science fiction, one could imagine a scenario where systems $A_1$ and $B_1$ emerge from a wormhole at time $t_0$ and system $A_1'$ and $B_1'$ enter the same wormhole at time $t_1$. Between times $t_0$ and $t_1$ the systems $A_1$ and $B_1$ can interact with the other systems in Alice’s and Bob’s laboratories, here denoted as $A_2$ and $B_2$. The interaction is controlled locally by Alice and Bob, who implement the channels $\mathcal{A}$ and $\mathcal{B}$, as illustrated in Figure 3. The connection of Alice’s and Bob’s laboratories through the background spacetime structure can be described as a map

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

which transforms the quantum channels $\mathcal{A}$ and $\mathcal{B}$ into a new quantum channel $S(\mathcal{A} \otimes \mathcal{B})$. Maps that transform channels into channels are known as quantum supermaps [32, 31]. The basic requirements for quantum supermaps are linearity, complete positivity, and normalization. In this setting, linearity means that one has

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

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 $S$ acts non-trivially only on these systems. The requirement that channels are mapped into channels for every possible choice of systems $A_2$ and $B_2$ imposes that the map $S$ be of the form

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

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 $A_2$ ($B_2$) and output $A_2$ ($B_2$), and $\mathcal{C}$ is a supermap that annihilates channels with
input $A_1B_1$ and output $A_1B_1$. Physically, the map $C$ represents the piece of spacetime connecting $A_1$ and $B_1$ with $A_1'$ and $B_1'$.

6.2. Choi operator formulation

Since all the maps $A, B,$ and $C$ are completely positive, one can represent them with Choi operators $A, B,$ and $C$, respectively. In terms of Choi operators, Eq. (45) can be expressed as

$$C' = \text{Tr}_{A_1,A_1',B_1,B_1'} \left( (I_{A_2} \otimes I_{A_2} \otimes C \otimes I_{B_2} \otimes I_{B_2}) (A^T \otimes B^T) \right),$$

where $C'$ the Choi operator of the channel $S(A \otimes B)$. Oreshkov, Costa, and Brukner refer to the Choi operator $C$ as a process matrix [12].

Here the operator $C$ acts on the tensor product Hilbert space $\mathcal{H}_{A_1} \otimes \mathcal{H}_{A_1} \otimes \mathcal{H}_{B_1} \otimes \mathcal{H}_{B_1}$. 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

$$\text{Tr}_{A_1,A_1',B_1,B_1'} \left[C(\tilde{A} \otimes \tilde{B})\right] = 1$$

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

$$\text{Tr}_{A_1'}[\tilde{A}] = I_{A_1} \quad \text{and} \quad \text{Tr}_{B_1'}[\tilde{B}] = I_{B_1}.$$

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

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

**Proposition 4.** An operator $C$ is the Choi operator of a non-causal network as in Fig. 3 if and only if $C$ is positive and $\text{Tr}[CD] = 1$ for every operator $D$ satisfying the conditions

$$\text{Tr}_{A_1'}[D] = I_{A_1} \otimes \tilde{B}, \quad \text{Tr}_{B_1'}[\tilde{B}] = I_{B_1}.$$

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

Equivalently, the valid networks can be characterized as in the following:
Optimal quantum networks and one-shot entropies

and

$$\text{Tr}_{B_1}[D] = I_{B_1} \otimes \tilde{A}, \quad \text{Tr}_{A_1}[\tilde{A}] = I_{A_1},$$

(50)

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

For the proof, see Theorem 2 of [14]. Physically, the operator $D$ represents the Choi operator of a no-signalling channel [46, 47, 48], 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-signalling channel.

In the following we will denote by $\text{NoSig}(A_1 \rightarrow A_1' | B_1 \rightarrow B_1')$ is the set of positive operators satisfying the no-signalling conditions (49) and (50). With this notation, Proposition 4 can be reformulated as

**Corollary 4.** An operator $C$ is the Choi operator of a non-causal network as in Fig. 3 if and only if

$$C \geq 0 \quad \text{and} \quad C \in \overline{\text{NoSig}(A_1 \rightarrow A_1' | B_1 \rightarrow B_1')},$$

(51)

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

We will denote by $\text{DualNoSig}(A_1 \rightarrow A_1' | B_1 \rightarrow B_1')$ the set of operators satisfying conditions (51) and (52).

6.3. The max relative entropy of signalling

In some situations, such as the study of non-causal games [12], 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 Fig. 3. 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 $\rho$ and $\sigma$, 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.

The maximum score is achievable by quantum non-causal networks is

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

(53)

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 5. Let $\Omega \in \text{Herm}(\mathcal{H}_{A_1} \otimes \mathcal{H}_{A_1} \otimes \mathcal{H}_{B_1} \otimes \mathcal{H}_{B_1})$ be a generic performance operator a $\omega_{\text{max}}$ be the maximum score defined in Eq. (53). Then, one has

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

When $\Omega$ is positive, the maximum score is given by

$$\omega_{\text{max}} = 2^{D_{\text{max}}(\Omega \parallel \text{NoSig}(A_1 \rightarrow A_1', B_1 \rightarrow B_1'))}.$$  

(54)

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-signalling channels. We call $D_{\text{max}}(\Omega \parallel \text{NoSig}(A_1 \rightarrow A_1', B_1 \rightarrow B_1'))$ the max relative entropy of signalling, in analogy with the relative entropy of entanglement of a state $\rho$ [49, 50, 51].

6.4. Optimizing multipartite non-causal networks

The results presented in 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

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

for every set of Choi operators $(\tilde{A}_1, \tilde{A}_2, \cdots, \tilde{A}_k)$ representing local quantum devices.

Equivalently, the normalization condition can be expressed as

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

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

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

$$\text{Tr}_{A'_J} [D] = I_{A_J} \otimes D^c_J,$$

where $\text{Tr}_{A'_J}$ is the partial trace over the Hilbert space $\mathcal{H}_{A'_J} := \bigotimes_{i \in J} \mathcal{H}_{A_i}$, $I_{A_S}$ is the identity operator on the Hilbert space $\mathcal{H}_{A_S} := \bigotimes_{i \in J} \mathcal{H}_{A_i}$, and $D^c_J$ is the Choi operator of a quantum channel transforming density matrices on $\mathcal{H}_{A^c_J} := \bigotimes_{i \not\in J} \mathcal{H}_{A_i}$ into density matrices on $\mathcal{H}_{A'_J} := \bigotimes_{i \not\in J} \mathcal{H}_{A'_i}$.

We denote the set of $k$-partite no-signalling channels as $\text{NoSig}_k$, keeping implicit the specification of the Hilbert spaces.
Also in the multipartite 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 $\Omega$. The maximum score is then given by

$$\omega_{\text{max}} = \max \{ \text{Tr}[\Omega C] \mid C \in \overline{\text{NoSig}_k} \}.$$  \hfill (55)

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

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

or, when $\Omega$ is positive

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

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

7. 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 another set of target channels. Then, we will move the case of non-causal networks.

7.1. Transforming quantum channels

Consider the following scenario: A black box implements a quantum channel in the set $\{\mathcal{E}_x\}_{x \in \mathcal{X}}$, where $\mathcal{X}$ 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

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

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

$$\begin{array}{c}
A_0 \\
\xrightarrow{A_1} \mathcal{E}_x \\
\xrightarrow{A_2} \mathcal{C}_1 \\
\xrightarrow{A_2} \mathcal{C}_2
\end{array} = \begin{array}{c}
A_0 \\
\xrightarrow{A_1} \mathcal{E}'_x \\
\xrightarrow{A_3} \mathcal{C}_1
\end{array} ,$$

where $\mathcal{C}_1$ and $\mathcal{C}_2$ are suitable quantum channels. The Choi operator of the output channel $\mathcal{E}'_x$ is then given by

$$E'_x = C * E_x ,$$

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 [52, 53, 54], given by

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

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

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

In this case, the fidelity can be interpreted as the probability that the network passes a test, where the channel $E'_x$ 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

$$F(\mathcal{E}'_x, \mathcal{F}_x) := \frac{1}{d_0^2} \text{Tr} \left[ C \langle \langle V_x | (|V_x\rangle\langle V_x| \otimes E'^T_x) \rangle \rangle \right]$$

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

$$f = \sum_x p(x) F(\mathcal{E}'_x, \mathcal{F}_x) = \text{Tr}[\Omega C], \quad \Omega := \sum_x p(x) \langle \langle V_x | (|V_x\rangle\langle V_x| \otimes E'^T_x) \rangle \rangle.$$ 

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

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

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

$$\Gamma = I_3 \otimes T_{210}, \quad \text{Tr}_2[T_{210}] = I_1 \otimes T_0, \quad \text{Tr}[T_0] = 1.$$

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

7.2. 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 $U$ into its inverse? It turns that such a conversion is impossible. 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 inverse. Our result highlights an analogy between the optimal inversion of an unknown unitary dynamics and the optimal universal NOT (UNOT) gate, that is, the quantum channel that attempts to transform every pure quantum state into its orthogonal complement [55, 56]. In this case, 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 [55, 56].

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 Eq. (61) takes the form

$$\Omega = \frac{1}{d^2} \int dU \langle U| \langle U|_{30} \otimes |U\rangle \langle U|_{21},$$

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

$$F_{\text{max}} = \frac{2}{d^2}.$$ (65)

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 $P_{\hat{U}} = d |\hat{U}\rangle \langle \hat{U}|$ [57]. This strategy leads to the conditional probability distribution

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

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

$$F_{\text{est}}(U) = \int d\hat{U} \left| \text{Tr}[U^\dagger \hat{U}] \right|^4 /d^2$$

$$= 2/d^2$$

$$\equiv F_{\text{max}}, \quad \forall U \in SU(d).$$ (66)

In conclusion, no quantum network can invert a gate better than a classical network that generates the inverse by using gate estimation as an intermediate step.

### 7.3. 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 $\overline{U}$, 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 Eq. (58).
For the charge conjugation problem, the performance operator \( \Omega \) reads
\[
\Omega = \frac{1}{d^2} \int dU \left| \mathbf{U} \right\rangle \langle \mathbf{U}\rangle_{30} \otimes \left| \mathbf{U} \right\rangle \langle \mathbf{U}\rangle_{21} \left( \frac{P_{+,-32} \otimes P_{+,-10}}{d_+} + \frac{P_{-,-32} \otimes P_{-,-10}}{d_-} \right),
\]
where \( P_+ \) and \( P_- \) (\( d_+ \) and \( d_- \)) are the projectors on (the dimensions of) the symmetric and antisymmetric subspaces, respectively. In appendix Appendix E we evaluate Eq. (62), obtaining the maximum fidelity
\[
F_{\text{max}} = \frac{2}{d(d - 1)}.
\]
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 \( \mathbf{U} \) are unitarily equivalent—specifically, \( \mathbf{U} = YU\mathbf{Y} \), 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 unitary gates.

For systems of large dimension, the fidelity converges to \( 2/d^2 \), the value achieved by gate estimation [cf. Eq. (66) in the previous paragraph]. Remarkably, estimating the gate is not the optimal strategy for every finite dimension. 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 \( \text{Tr} [\Omega C] = 1/d_- = F_{\text{max}} \). Physically, \( C \) represents a network of the form
\[
A_0 \begin{pmatrix} \mathbf{K} \\ \mathbf{A}_1 \end{pmatrix} A_1, \quad A_2 \begin{pmatrix} \mathbf{K} \\ \mathbf{A}_3 \end{pmatrix} A_3,
\]
consisting of two subsequent uses of the channel \( \mathbf{K} \) with Choi operator \( \mathbf{K} := d P_-/d_- \). 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 \( \mathbf{F'} = \mathbf{K} \mathbf{U} \mathbf{K} \), which optimally simulates the charge conjugate evolution \( \mathbf{U} \).

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
\[
\rho \xrightarrow{\text{transpose}} \rho^T \xrightarrow{U} U\rho^T U^\dagger \xrightarrow{\text{transpose}} \mathbf{U}\rho\mathbf{U}^T.
\]
As the transpose is not a physical operation, one may try to use the optimal transpose channel [58, 59, 60, 61, 62], which has Choi operator \( T = dP_+/d_+ \). However, this choice would be suboptimal, leading to the fidelity
\[
F_{\text{transpose}} = 1/d_+ = 2/[d(d + 1)] < F_{\text{max}}.
\]
Instead, the optimal strategy is to approximate the transpose \textsc{not}, i.e. the impossible transformation that maps every projector into its orthogonal complement. In the Heisenberg picture, the transpose \textsc{not} maps every observable $A$ into the observable $I - A^T$, allowing us to reproduce the charge conjugate dynamics as

$$A \xrightarrow{\text{transpose \textsc{not}}} I - A^T \xrightarrow{U} I - U^\dagger A^T U \xrightarrow{\text{transpose \textsc{not}}} U^T A U.$$ 

It turns out that the optimal approximation of the transpose \textsc{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 \textsc{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 (68) would lead to a classical, estimation-based strategy. Instead, the transpose \textsc{not} cannot be achieved via state estimation. For example, the transpose \textsc{not} for qubits is a unitary transformation, corresponding to the Pauli matrix $Y$.

### 7.4. 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 by Araujo \textit{et al} [63]. Like the no-cloning Theorem, this “no-controlization” result establishes the impossibility of a perfect functionality. But what about approximate controlization? \textit{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.

For the controlization task, the performance operator $\Omega$ reads

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

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

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

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 $C_U$ defined by

$$C_U(\rho \otimes \sigma) := \langle 0|\sigma|0\rangle \rho + \langle 1|\sigma|1\rangle U\rho U^\dagger,$$

where $\rho$ is an arbitrary state of the system and $\sigma$ is an arbitrary state of the control qubit. It is immediate to check that the fidelity between the classically-controlled channel $C_U$ 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 measured.

7.5. Maximization of a non-causal Bell-type correlation

Here we consider the non-causal game introduced by Oreshkov, Costa, and Brukner in Ref. [12]. 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, and 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

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

(70)

In this game, Alice and Bob are not subject to the no-signalling 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.

Now, it is interesting to see how quantum resources can help Alice and Bob in this game. The most general quantum resource is described by a network that connects Alice’s operations to Bob’s operations. The network will provide inputs $A_1$ and $B_1$ to Alice and Bob, respectively. Alice and Bob then perform local operations, transforming systems $A_1$ and $B_1$ them into systems $A_2$ and $B_2$. Alice’s and Bob’s 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 Fig. 4.

Mathematically, the operations are described by two quantum instruments $\{\mathcal{M}_x^a\}_{x=0,1}$ and $\{\mathcal{N}_y^{b,b'}\}_{y=0,1}$, that is, two collections of completely positive, trace non-increasing maps with the property that the maps

$$\mathcal{M}^a := \sum_x \mathcal{M}_x^a \quad \text{and} \quad \mathcal{N}^{b,b'} := \sum_y \mathcal{N}_y^{b,b'}$$

are trace-preserving. With these settings, the probability distribution of the outputs is given by

$$p(x, y|a, b, b') = \text{Tr} \left[ \left( \mathcal{M}_x^a \otimes \mathcal{N}_y^{b,b'} \right) C \right]$$
where \( \{ M^a_x \}_{x=0,1} \) and \( \{ N^{b,b'}_y \}_{y=0,1} \) 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

\[
\omega = \frac{1}{8} \sum_{a,b,b',x,y} \omega(x,y|a,b,b') p(x,y|a,b,b')
\]

\[
= \text{Tr} \left[ \Omega C \right],
\]

where \( \Omega \) is the performance operator

\[
\Omega := \frac{1}{8} \sum_{a,b,b',x,y} \omega(x,y|a,b,b') \left( M^a_x \otimes N^{b,b'}_y \right).
\] (71)

The main result by Oreshkov, Costa, and Brukner 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 \( \{ M^a_x \}_{x=0,1} \) and \( \{ N^{b,b'}_y \}_{y=0,1} \) that achieve score

\[
\omega_* = \frac{1}{2} \left( 1 + \frac{1}{\sqrt{2}} \right).
\] (72)

Specifically, the score \( \omega_* \) is achieved by choosing systems \( A_1, B_1, A_2, B_2 \) to be qubits and by choosing the local operations with Choi operators

\[
M^a_x = \frac{1}{4} \left[ I + (-1)^x \sigma_z \right]_{A_1} \otimes \left[ I + (-1)^a \sigma_z \right]_{A_2}
\]

\[
N^{b,b'}_y = b' \left[ \frac{1}{2} \left[ I + (-1)^y \sigma_z \right]_{B_1} \otimes \rho_{B_2} \right]
\]

\[
+ (b' \oplus 1) \left\{ \frac{1}{4} \left[ I + (-1)^{b'} \sigma_z \right]_{B_1} \otimes \left[ I + (-1)^{b+y} \sigma_z \right]_{B_2} \right\},
\] (73)

where \( \oplus \) denotes the addition modulo 2 and \( \rho_{B_2} \) 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 \( \omega \) 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 \( \omega_* \)
is the maximum score attainable when Alice’s and Bob’s instruments (73) are connected by an arbitrary non-causal network. The question was answered in the affirmative by Brukner [20]. We now present an alternative derivation, which is comparatively shorter.

Inserting Eq. (73) into Eq. (71) we obtain the performance operator
\[
\Omega = \sum_{i,j,k} |i\rangle \langle i|_{A_1} \otimes |j\rangle \langle j|_{A_2} \otimes \Omega_{ijk} \otimes |k\rangle \langle k|_{B_2}
\]
where \(\Omega_{ijk}\) are operators acting on \(B_1\) and are defined as
\[
\begin{align*}
\Omega_{000} &= \frac{1}{8}(|+\rangle \langle +| + |0\rangle \langle 0|), \\
\Omega_{001} &= \frac{1}{8}(|-\rangle \langle -| + |0\rangle \langle 0|), \\
\Omega_{010} &= \frac{1}{8}(|+\rangle \langle +| + |1\rangle \langle 1|), \\
\Omega_{011} &= \frac{1}{8}(|-\rangle \langle -| + |1\rangle \langle 1|), \\
\Omega_{100} &= \frac{1}{8}(|-\rangle \langle -| + |0\rangle \langle 0|), \\
\Omega_{101} &= \frac{1}{8}(|+\rangle \langle +| + |0\rangle \langle 0|), \\
\Omega_{110} &= \frac{1}{8}(|-\rangle \langle -| + |1\rangle \langle 1|), \\
\Omega_{111} &= \frac{1}{8}(|+\rangle \langle +| + |1\rangle \langle 1|). 
\end{align*}
\]

Now, the dual optimization problem is to find the minimum \(\lambda\) such that \(\lambda \Gamma \geq \Omega\), for some Choi operator \(\Gamma\) representing a no-signalling channel. The key observation is that all the \(\Omega_{ijk}\) have the same maximum eigenvalue, equal to \(\epsilon_{\text{max}} = 1/8(1 + 1/\sqrt{2})\). As a result, we can satisfy the dual constraint by setting \(\lambda = 1/2(1 + 1/\sqrt{2})\) and \(\Gamma = I_{A_1A_2B_1B_2}/4\). Note that \(\Gamma\) is the Choi operator of a no-signalling channel, as it satisfies Eqs. (49) and (50). Hence, we obtain the bound
\[
\omega \leq \frac{1}{2} \left(1 + \frac{1}{\sqrt{2}}\right), \tag{74}
\]
valid for every non-causal network. The bound can be achieved, since r.h.s. matches the value in Eq. (72).

8. Conclusions

We developed a semidefinite programming 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, which measures the deviation of the figure of merit from the set of functions that are constant on the set of quantum networks under optimization. Using this fact, it is possible to extend the notion of conditional min-entropy from quantum states to quantum networks, providing a quantitative measure of the amount of quantum correlations generated over a sequence of time steps. We expect that this quantity will play a role in the study non-Markovian quantum evolutions, offering an opportunity to extend the results of Refs. [64, 65] to stochastic processes that can interact with external devices in a sequence of time steps.

As a general-purpose method, our semidefinite programming approach has applications to a number of new perspective scenarios, such as quantum sensor
networks, distributed algorithms and communication protocols, and fundamental tests of spacetime physics. A 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 resources. Exploring how resources like energy and coherence can be optimally allocated within a distributed system can potentially unveil new quantum advantages, leading to a further level of optimization in the design of quantum technologies.

Acknowledgments. This work is supported by the Foundational Questions Institute (FQXi-RFP3-1325 and FQXi-MGA-1502), by the National Natural Science Foundation of China through Grant 11450110096, and by the HKU Seed Funding for Basic Research. GC is grateful to F Buscemi for useful discussions and to the organizers of the workshop “Quantum Networks”, Barcelona, March 30-April 1 2016, which offered the occasion for a stimulating exchange of ideas that benefitted this paper.

[1] Xi-Lin Wang, Xin-Dong Cai, Zu-En Su, Ming-Cheng Chen, Dian Wu, Li Li, Naï-Le Liu, Chao-Yang Lu, and Jian-Wei Pan. Quantum teleportation of multiple degrees of freedom of a single photon. Nature, 518(7540):516–519, 2015.
[2] H. Haffner, C. Roos, and R. Blatt. Quantum computing with trapped ions. Physics Reports, 469(4):155–203, 2008.
[3] O. Morsch and M. Oberthaler. Dynamics of Bose-Einstein condensates in optical lattices. Reviews of Modern Physics, 78(1):179, 2006.
[4] J. Wrachtrup and F. Jelezko. Processing quantum information in diamond. Journal of Physics: Condensed Matter, 18(21):S807, 2006.
[5] Juan Yin, Yuan Cao, Shu-Bin Liu, Ge-Sheng Pan, Jin-Hong Wang, Tao Yang, Zhong-Ping Zhang, Fu-Min Yang, Yu-Ao Chen, Cheng-Zhi Peng, et al. Experimental quasi-single-photon transmission from satellite to earth. Optics express, 21(17):20032–20040, 2013.
[6] Anne Broadbent, Joseph Fitzsimons, and Elham Kashefi. Universal blind quantum computation. In Foundations of Computer Science, 2009. FOCS’09. 50th Annual IEEE Symposium on, pages 517–526. IEEE, 2009.
[7] Robert Beals, Stephen Brierley, Oliver Gray, Aram W Harrow, Samuel Kutin, Noah Linden, Dan Shepherd, and Mark Stather. Efficient distributed quantum computing. In Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences, volume 469, page 20120686. The Royal Society, 2013.
[8] Stefanie Barz, Elham Kashefi, Anne Broadbent, Joseph F Fitzsimons, Anton Zeilinger, and Philip Walther. Demonstration of blind quantum computing. Science, 335(6066):303–308, 2012.
[9] Jeff Kimble. The quantum internet. Nature, 453(7198):1023–1030, 2008.
[10] Lucien Hardy. Quantum gravity computers: On the theory of computation with indefinite causal structure. In Quantum Reality, Relativistic Causality, and Closing the Epistemic Circle, pages 379–401. Springer, 2009.
[11] Giulio Chiribella, GM D’Ariano, P Perinotti, and B Valiron. Beyond quantum computers. arXiv preprint arXiv:0912.0195, 2009.
[12] Ognyan Oreshkov, Fabio Costa, and Časlav Brukner. Quantum correlations with no causal order. Nature communications, 3:1092, 2012.
[13] Giulio Chiribella. Perfect discrimination of no-signalling channels via quantum superposition of causal structures. Physical Review A, 86(4):040301, 2012.
[14] Giulio Chiribella, Giacomo Mauro D’Ariano, Paolo Perinotti, and Benoît Valiron. Quantum computations without definite causal structure. Physical Review A, 88(2):022318, 2013.
[15] Timoteo Colnaghi, Giacomo Mauro D’Ariano, Stefano Facchini, and Paolo Perinotti. Quantum
computation with programmable connections between gates. *Physics Letters A*, 376(45):2940–2943, 2012.

[16] Å. Baumeler and S. Wolf. Perfect signaling among three parties violating predefined causal order. *ArXiv e-prints*, December 2013.

[17] Amin Baumeler and Stefan Wolf. Non-causal computation avoiding the grandfather and information antinomies. *arXiv preprint arXiv:1601.06522*, 2016.

[18] Mateus Araújo, Fabio Costa, and Časlav Brukner. Computational advantage from quantum-controlled ordering of gates. *Physical review letters*, 113(25):250402, 2014.

[19] Christopher Portmann, Christian Matt, Ueli Maurer, Renato Renner, and Björn Tackmann. Causal boxes: Quantum information-processing systems closed under composition. *arXiv preprint arXiv:1512.02240*, 2015.

[20] Časlav Brukner. Bounding quantum correlations with indefinite causal order. *New Journal of Physics*, 17(8):083034, 2015.

[21] Cyril Branciard, Mateus Araújo, Adrien Feix, Fabio Costa, and Časlav Brukner. The simplest causal inequalities and their violation. *New Journal of Physics*, 18(1):013008, 2015.

[22] Lorenzo M Procopio, Amir Moqanaki, Mateus Araújo, Fabio Costa, Irati Alonso Calafell, Emma G Dowd, Deny R Hamel, Lee A Rozema, Časlav Brukner, and Philip Walther. Experimental superposition of orders of quantum gates. *Nature Communications*, 6, 2015.

[23] Alexei Kitaev and John Watrous. Parallelization, amplification, and exponential time simulation of quantum interactive proof systems. In *Proceedings of the thirty-second annual ACM symposium on Theory of computing*, pages 608–617. ACM, 2000.

[24] Giulio Chiribella, Giacomo Mauro D’Ariano, and Paolo Perinotti. Quantum circuit architecture. *Physical review letters*, 101(6):060401, 2008.

[25] Gus Gutoski and John Watrous. Toward a general theory of quantum games. In *Proceedings of the thirty-ninth annual ACM symposium on Theory of computing*, pages 565–574. ACM, 2007.

[26] Renato Renner and Stefan Wolf. Smooth rényi entropy and applications. In *IEEE International Symposium on Information Theory*, pages 233–233, 2004.

[27] Renato Renner. Security of quantum key distribution. *International Journal of Quantum Information*, 6(01):1–127, 2008.

[28] Nilanjan Datta and Renato Renner. Smooth entropies and the quantum information spectrum. *Information Theory, IEEE Transactions on*, 55(6):2807–2815, 2009.

[29] Robert König, Renato Renner, and Christian Schaffner. The operational meaning of min-and max-entropy. *Information Theory, IEEE Transactions on*, 55(9):4337–4347, 2009.

[30] Marco Tomamichel. *Quantum Information Processing with Finite Resources: Mathematical Foundations*, volume 5. Springer, 2015.

[31] Giulio Chiribella, Giacomo Mauro D’Ariano, and Paolo Perinotti. Theoretical framework for quantum networks. *Physical Review A*, 80(2):022339, 2009.

[32] Giulio Chiribella, Giacomo Mauro D’Ariano, and Paolo Perinotti. Transforming quantum operations: Quantum supermaps. *EPL (Europhysics Letters)*, 83(3):30004, 2008.

[33] Karl Kraus, Arno Böhm, John D Dollard, and WH Wootters. States, effects, and operations fundamental notions of quantum theory. 1983.

[34] Man-Duen Choi. Completely positive linear maps on complex matrices. *Linear algebra and its applications*, 10(3):285–290, 1975.

[35] Thomas H. Cormen, Charles Eric Leiserson, Ronald L Rivest, and Clifford Stein. *Introduction to algorithms*, volume 6. MIT press Cambridge, 2001.

[36] Giulio Chiribella, Giacomo M D’Ariano, and Paolo Perinotti. Memory effects in quantum channel discrimination. *Physical review letters*, 101(18):180501, 2008.

[37] John Watrous. *Lecture notes on semidefinite programming*. University of Waterloo, 2011.

[38] Stephen Boyd and Lieven Vandenberghe. *Convex optimization*. Cambridge university press, 2004.

[39] Nilanjana Datta. Min-and max-relative entropies and a new entanglement monotone. *Information Theory, IEEE Transactions on*, 55(6):2816–2826, 2009.
[40] Fernando Brandao and Nilanjana Datta. One-shot rates for entanglement manipulation under non-entangling maps. *Information Theory, IEEE Transactions on*, 57(3):1754–1760, 2011.

[41] Michał Horodecki and Jonathan Oppenheim. Fundamental limitations for quantum and nanoscale thermodynamics. *Nature communications*, 4, 2013.

[42] Fernando Brandão, Michał Horodecki, Nelly Ng, Jonathan Oppenheim, and Stephanie Wehner. The second laws of quantum thermodynamics. *Proceedings of the National Academy of Sciences*, 112(11):3275–3279, 2015.

[43] Giulio Chiribella and Jinyu Xie. Optimal design and quantum benchmarks for coherent state amplifiers. *Physical review letters*, 110(21):213602, 2013.

[44] Giulio Chiribella. Optimal networks for quantum metrology: semidefinite programs and product rules. *New Journal of Physics*, 14(12):125008, 2012.

[45] A Jenčová. Base norms and discrimination of generalized quantum channels. *Journal of Mathematical Physics*, 55(2):022201, 2014.

[46] David Beckman, Daniel Gottesman, MA Nielsen, and John Preskill. Causal and localizable quantum operations. *Physical Review A*, 64(5):052309, 2001.

[47] Marco Piani, Michał Horodecki, Paweł Horodecki, and Ryszard Horodecki. Properties of quantum nonsignaling boxes. *Physical Review A*, 74(1):012305, 2006.

[48] Giacomo Mauro D’Ariano, Stefano Facchini, and Paolo Perinotti. No signaling, entanglement breaking, and localizability in bipartite channels. *Physical review letters*, 106(1):010501, 2011.

[49] Charles Bennett, Herbert Bernstein, Sandu Popescu, and Benjamin Schumacher. Concentrating partial entanglement by local operations. *Physical Review A*, 53:2046–2052, Apr 1996.

[50] Sandu Popescu and Daniel Rohrlich. Thermodynamics and the measure of entanglement. *Physical Review A*, 56(5):R3319, 1997.

[51] Vlatko Vedral, Martin B Plenio, Michael A Rippin, and Peter L Knight. Quantifying entanglement. *Physical Review Letters*, 78(12):2275, 1997.

[52] Maxim Raginsky. A fidelity measure for quantum channels. *Physics Letters A*, 290(1):11–18, 2001.

[53] Michael A Nielsen. A simple formula for the average gate fidelity of a quantum dynamical operation. *Physics Letters A*, 303(4):249–252, 2002.

[54] Michał Horodecki, Paweł Horodecki, and Ryszard Horodecki. General teleportation channel, singlet fraction, and quasidistillation. *Phys. Rev. A*, 60:1888–1898, Sep 1999.

[55] V Bužek, M Hillery, and RF Werner. Optimal manipulations with qubits: Universal-not gate. *Physical Review A*, 60(4):R2626, 1999.

[56] Pranaw Rungta, Vladimir Bužek, Carlton M Caves, Mark Hillery, and Gerard J Milburn. Universal state inversion and concurrence in arbitrary dimensions. *Physical Review A*, 64(4):042315, 2001.

[57] Giulio Chiribella, GM D’ariano, and MF Sacchi. Optimal estimation of group transformations using entanglement. *Physical Review A*, 72(4):042338, 2005.

[58] Paweł Horodecki and Artur Ekert. Method for direct detection of quantum entanglement. *Physical review letters*, 89(12):127902, 2002.

[59] Paweł Horodecki. From limits of quantum operations to multicopy entanglement witnesses and state-spectrum estimation. *Physical Review A*, 68(5):052101, 2003.

[60] F Buscemi, GM D’Ariano, P Perinotti, and MF Sacchi. Optimal realization of the transposition maps. *Physics Letters A*, 314(5):374–379, 2003.

[61] Amir Kalev and Joonwoo Bae. Optimal approximate transpose map via quantum designs and its applications to entanglement detection. *Physical Review A*, 87(6):062314, 2013.

[62] Hyang-Tag Lim, Yong-Su Kim, Young-Sik Ra, Joonwoo Bae, and Yoon-Ho Kim. Experimental realization of an approximate partial transpose for photonic two-qubit systems. *Phys. Rev. Lett.*, 107:160401, Oct 2011.

[63] Mateus Araújo, Adrien Feix, Fabio Costa, and Časlav Brukner. Quantum circuits cannot control unknown operations. *New Journal of Physics*, 16(9):093026, 2014.

[64] Francesco Buscemi and Nilanjana Datta. Equivalence between divisibility and monotonic decrease
Appendix A. Proof of Theorem 1

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

$$\omega_{\text{primal}} = \sup \{ \langle A, X \rangle \mid X \geq 0, X \in \mathcal{S} \}$$

$$= \sup \left\{ \langle A, X \rangle \mid X \geq 0, X \in \mathcal{S} \right\}$$

$$= \sup \left\{ \langle A, X \rangle \mid X \geq 0, \langle \Gamma, X \rangle = 1, \forall \Gamma \in \mathcal{S} \right\},$$

having used the relation $\mathcal{S} = \mathcal{S}$.

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

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

Weak duality then yields the relation

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

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

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

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

Now, suppose that $\mathcal{S}$ contains a positive operator $X_0$ and $\mathcal{S}$ contains a strictly positive operator $\Gamma_0$, then Slater’s Theorem implies the equality: indeed, one can choose the affine basis $(\Gamma_i)_{i=1}^K$ to contain the operator $\Gamma_0$. Since $\Gamma_0$ 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 r.h.s. of Eq. (A.1) admits a strictly positive solution. Hence, proposition 3 implies the equality in Eq. (A.1). The equality holds also in Eq. (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 $\lambda_1$ with $\lambda_1 + \epsilon$, $\epsilon > 0$. Since $\epsilon$ can be arbitrarily small, this substitution does not change the value of the infimum. If $A$ is positive, then one has the lower bound $\omega_{\text{primal}} \geq \langle A, X_0 \rangle \geq 0$. Eq. (A.3) then implies that every $\lambda$ satisfying $\lambda \Gamma \geq A$, $\Gamma \in \mathcal{S}$ must be non-negative. If $\lambda$ is strictly positive, the operator $\Gamma$ must be positive. If $\lambda = 0$, the operator $\Gamma$ can be chosen to be positive without loss of generality. In conclusion, the infimum in Eq. (A.3) can be restricted to $\mathcal{S}_+$. Setting $w := 1/\lambda$ one finally obtains the desired expression. 

$\square$
Appendix B. Proof of Theorem 2

Proof. The maximum performance is given by Eq. (16). The expression can be re-written as

\[ \omega_{\text{max}} := \max \{ \langle \Omega, C \rangle | C \in S, C \geq 0 \} , \quad (B.1) \]

where \( S \) is the affine space of all the operators on \( \bigotimes_{j=1}^{N} (H_{\text{out}}^j \otimes H_{\text{in}}^j) \) that are Hermitian and satisfy the linear constraint (9). Note that \( S \) contains the strictly positive operator

\[ C_0 = I_0 \otimes I_1 \otimes \cdots \otimes I_{2N-1}/(d_1 d_3 \ldots d_{2N-1}) \quad (B.2) \]

and the dual affine space \( \bar{S} \) contains the strictly positive operator

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

Since the sets \( S \) and \( \bar{S} \) contain strictly positive operators, the expression in Theorem 1 holds with the equality. Moreover, one can choose the performance operator \( \Omega \) to be positive without loss of generality: if \( \Omega \) is not positive, one can define \( \Omega' = \Omega + c \Gamma_0 \), where \( c \) is a positive constant and \( \Gamma_0 \) is the operator in Eq. (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 \( \bar{S}^+ \), namely

\[ \omega'_{\text{max}} = \inf_{\Gamma \in \bar{S}^+} \min \{ \lambda \in \mathbb{R} | \lambda \Gamma \geq \Omega' \} . \]

Now, the set \( \bar{S}^+ \) has been characterized in Ref. [31]: precisely, \( \bar{S}^+ \) is the set of all positive operators \( \Gamma \) satisfying the linear constraint

\[ \begin{align*}
\Gamma &= I_{A_2}^{\text{out}} \otimes \Gamma^{(N)} \\
\text{Tr}_{A_2^{\text{out}}}[\Gamma^{(N)}] &= I_{A_{n-1}}^{\text{out}} \otimes \Gamma^{(N-1)} , \quad n = 2, \ldots, N \\
\text{Tr}_{A_1^{\text{in}}}[\Gamma^{(1)}] &= 1 ,
\end{align*} \quad (B.4) \]

for suitable positive operators \( \Gamma^{(n)} \) acting on \( \mathcal{H}_{\text{in}}^n \otimes \left[ \bigotimes_{j=1}^{n-1} (\mathcal{H}_{\text{out}}^j \otimes \mathcal{H}_{\text{in}}^j) \right] \). Observing that \( I_{A_n}^{\text{out}} \) is the Choi operator of the trace channel \( \text{Tr}_{A_n^{\text{out}}} \) and comparing Eq. (B.4) with Eq. (9) we then obtain that every operator \( \Gamma \) in \( \bar{S}^+ \) is the Choi operator of a network of the form (27). Hence, \( \bar{S}^+ = \text{DualComb} \). Finally, note that the set \( \text{DualComb} \) is compact and therefore the infimum is a minimum. \qed

Appendix C. Normalization condition for supermaps on product channels

Equation (46) gives us the Choi operator \( C \). In order for \( C \) to be the Choi operator of a channel, we must have

\[ \text{Tr}_{A_2^{\text{out}},B_2^{\text{out}}}[C] = I_{A_2} \otimes I_{B_2}. \quad (C.1) \]

Inserting Eq. (46), we then obtain the condition

\[ \begin{align*}
\text{Tr}_{A_2^{\text{out}},B_2^{\text{out}},A_1^{\text{in}},B_1^{\text{in}}}[\left( (I_{A_2'} \otimes I_{A_2} \otimes N \otimes I_{B_2'} \otimes I_{B_2})(A \otimes B)\right)] &= I_{A_2} \otimes I_{B_2} \quad (C.2)
\end{align*} \]
which must be satisfied whenever $A$ and $B$ satisfy the conditions
\[ \text{Tr}_{A_1',A_2'}[A] = I_{A_1} \otimes I_{A_2} \quad \text{and} \quad \text{Tr}_{B_1',B_2'}[B] = I_{B_1} \otimes I_{B_2}. \] (C.3)

Now, we have the following

**Proposition 6.** For every operator $N$, the following conditions are equivalent:

(i) $N$ satisfies the condition (C.2) for every operators $A$ and $B$ satisfying the condition (C.3)

(ii) $N$ satisfies the condition
\[ \text{Tr}_{A_1',B_1,B_1'}[N (\bar{A} \otimes \bar{B})] = 1 \] (C.4)
for every operators $\bar{A} \in \text{Lin}(\mathcal{H}_{A_1'} \otimes \mathcal{H}_{A_1})$ and $\bar{B} \in \text{Lin}(\mathcal{H}_{B_1'} \otimes \mathcal{H}_{B_1})$ satisfying the conditions
\[ \text{Tr}_{A_1'}[\bar{A}] = I_{A_1} \quad \text{and} \quad \text{Tr}_{B_1'}[\bar{B}] = I_{B_1}. \] (C.5)

**Proof.** Suppose that the operators $\bar{A}$ and $\bar{B}$ satisfy the trace conditions (C.5). By defining the operators $A$ and $B$ as $A = \bar{A} \otimes I_{A_2}/d_{A_2}$ and $B = \bar{B} \otimes I_{B_2}/d_{B_2}$, we see that Eq. (C.3) is satisfied. Then, Eq. (C.2) becomes
\[ \text{Tr}_{A_2',B_2',A_1',B_1,B_1'} \left[ \left( I_{A_2'} \otimes I_{A_2} \otimes N \otimes I_{B_2'} \otimes I_{B_2} \right)(A \otimes B) \right] = \text{Tr}_{A_2',B_2',A_1',B_1,B_1'} \left\{ \frac{I_{A_2'}}{d_{A_2}} \otimes I_{A_2} \otimes \left[ N(\bar{A} \otimes \bar{B}) \right] \otimes \frac{I_{B_2'}}{d_{B_2}} \otimes I_{B_2} \right\} = I_{A_2} \otimes (C.6) \]
The above equation holds if and only if condition (C.4) is satisfied. Conversely, if the operator $N$ satisfies condition (C.4) and $\bar{A}$ and $\bar{B}$ the trace conditions (C.5), we obtain
\[ \text{Tr}_{A_2',B_2',A_1',B_1,B_1'} \left[ \left( I_{A_2'} \otimes I_{A_2} \otimes N \otimes I_{B_2'} \otimes I_{B_2} \right)(A \otimes B) \right] = \text{Tr}_{A_1,A_1',B_1',B_1'} \left[ \left( I_{A_2} \otimes N \otimes I_{B_2} \right) \left( \text{Tr}_{A_2',B_2'}[A \otimes B] \right) \right] = \text{Tr}_{A_1,A_1',B_1',B_1'} \left[ \left( I_{A_2} \otimes N \otimes I_{B_2} \right) \left( \bar{A} \otimes \bar{B} \right) \right] \] (C.7)
where we defined $\text{Tr}_{A_2'}[A] = \bar{A}$ and $\text{Tr}_{B_2'}[B] = \bar{B}$. Hence, Eq. (C.2) holds if and only if
\[ \text{Tr}_{A_1,A_1',B_1,B_1'} \left[ \left( I_{A_2} \otimes N \otimes I_{B_2} \right) \left( \bar{A} \otimes \bar{B} \right) \right] = I_{A_2} \otimes I_{B_2}. \] (C.8)
In turn, the above equation holds if and only if
\[ \text{Tr}_{A_1,A_1',B_1,B_1'} \left[ N \left( \bar{A}_\rho \otimes \bar{B}_\sigma \right) \right] = 1, \quad \forall \rho \in \text{St}(\mathcal{H}_{A_1}), \forall \sigma \in \text{St}(\mathcal{H}_{B_1}), \] (C.9)
where $\bar{A}_\rho$ and $\bar{B}_\sigma$ are defined as
\[ \bar{A}_\rho := \text{Tr}_{A_2}[(\rho \otimes I_{A_1'})\bar{A}] \quad \text{and} \quad \bar{B}_\sigma := \text{Tr}_{B_2}[(\rho \otimes I_{B_1'})\bar{B}] \] (C.10)
Now, the normalization condition (C.9) is nothing but Eq. (C.4). The condition is satisfied because the operators $\bar{A}_\rho$ and $\bar{B}_\sigma$ satisfy condition (C.5). \qed
Appendix D. Maximum fidelity for the inversion of an unknown dynamics

The performance operator $\Omega$ reads

$$\Omega = \frac{1}{d^2} \int dU \langle U | \langle U |_3 \otimes | U \rangle \langle U |_2 \rangle$$

$$= \frac{1}{d^2} \int dU \left( I_3 \otimes U_0 \otimes U_2 \otimes I_1 \right) \left( | I \rangle \langle I |_3 \otimes | I \rangle \langle I |_2 \right) (I_3 \otimes U_0 \otimes U_2 \otimes (I)\rangle \langle I) \rangle$$

 Explicit calculation using Schur's lemma yields the relations

$$[\Omega, I_3 \otimes U_2 \otimes I_1 \otimes U_0] = 0$$  \hspace{1cm} (D.2)

$$[\Omega, U_3 \otimes I_2 \otimes U_1 \otimes I_0] = 0,$$ \hspace{1cm} (D.3)

required to hold for every unitary $U$. Explicitly, the operator $\Omega$ is given by

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

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

The problem is to find the minimum $\lambda$ such that $\lambda \Gamma \geq \Omega$, for $\Gamma$ satisfying the conditions (63). The first condition requires $\Gamma$ to be of the form $\Gamma = I_3 \otimes T_{210}$. Now, Eq. (D.3) implies that, without loss of generality, the operator $T_{210}$ can be chosen to satisfy the condition

$$[T_{210}, I_2 \otimes U_1 \otimes I_0] = 0 \quad \forall \ U \in SU(d)$$ \hspace{1cm} (D.5)

which in turn implies

$$T_{210} = Q_{20} \otimes I_1$$ \hspace{1cm} (D.6)

where $Q_{20}$ is some positive operator on $H_{20}$. Similarly, Eq. (D.2) implies that we can choose $T_{210}$ to satisfy the condition

$$[T_{210}, U_2 \otimes I_1 \otimes U_0] = 0 \quad \forall \ U \in SU(d).$$ \hspace{1cm} (D.7)

Combined with Eq. (D.6), the above relation implies

$$[Q_{20}, U_2 \otimes U_0] = 0 \quad \forall \ U \in SU(d)$$ \hspace{1cm} (D.8)

and therefore

$$Q_{20} = \alpha P_+ + \beta P_-$$ \hspace{1cm} (D.9)

Finally, the last condition in Eq. (63) gives $\text{Tr}[Q_{20}] = 1$ and, therefore,

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

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

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

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

$$\lambda \geq \frac{1}{d_+ d_+^2} \text{ and } \lambda \geq \frac{1}{d^2 (d - \alpha d_+)}$$ \hspace{1cm} (D.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 $\lambda$ is $\lambda_{\text{min}} = 2/d^2$. 
Appendix E. 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 $\Omega$ is given by Eq. (67), namely
\[
\Omega = \frac{1}{d^2} \left( \frac{P_{+,32} \otimes P_{+,10}}{d_+} + \frac{P_{-,32} \otimes P_{-,10}}{d_-} \right).
\]
(E.1)

The form of $\Omega$ implies the relations
\[
[\Omega, U_3 \otimes U_2 \otimes I_{10}] = 0 \quad \text{(E.2)}
\]
\[
[\Omega, I_{32} \otimes U_1 \otimes U_0] = 0, \quad \text{(E.3)}
\]
valid for every $U$ in $\text{SU}(d)$. Now, one has to find the minimum $\lambda$ such that
\[
\lambda \left( I_{32} \otimes \Gamma \right) \geq \Omega,
\]
with some $\Gamma$ satisfying Eqs. (63). Eq. (E.2) implies that, without loss of generality, one has
\[
[T_{210}, U_2 \otimes I_{10}] = 0 \quad \forall U \in \text{SU}(d),
\]
(E.4)

and therefore $T_{210} = I_2 \otimes Q_{10}$. Moreover, the second condition in Eq. (63) 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, Eq. (E.3) implies that one can choose $\rho_0 = I / d$ without loss of generality. Summing everything up, $\Gamma$ 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
\[
\lambda \frac{I_{3210}}{d^2} \geq \frac{1}{d^2} \left( \frac{P_{+,32} \otimes P_{+,10}}{d_+} + \frac{P_{-,32} \otimes P_{-,10}}{d_-} \right),
\]
(E.5)
yielding the minimal value $\lambda_{\text{min}} = 1 / d_- = 2 / d(d - 1)$.

Appendix F. Maximum fidelity for unitary controlization

The performance operator for the controlization problem is
\[
\Omega = \frac{1}{4d^2} \int dg \left| \text{ctrl} - U \right> \left< \text{ctrl} - U \right|_{30Q'Q} \otimes \left| \overline{U} \right> \left< \overline{U} \right|_{21}
\]
\[
= \Omega^{(0)}_{3210} \otimes |0\rangle^Q \otimes |0\rangle^{Q'} + \Omega^{(1)}_{3210} \otimes |1\rangle^Q \otimes |1\rangle^{Q'},
\]
(F.1)

where $Q$ and $Q'$ denote the control qubit before and after the interaction, respectively, and
\[
\Omega^{(0)}_{3210} := \frac{1}{4d^2} \left( E_{30} \otimes I_{21} \right)
\]
\[
\Omega^{(1)}_{3210} := \frac{1}{4d^2} \left( E_{32} \otimes E_{10} + \frac{E_{32} \otimes E_{10}}{d_+} \right)
\]
(F.2)

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
Optimal quantum networks and one-shot entropies

\( \Omega_{3210}^{(0)} \) and \( \Omega_{3210}^{(1)} \) satisfy the conditions

\[
\begin{align*}
&\left[ \Omega_{3210}^{(1)}, U_3 \otimes I_{21} \otimes U_0 \right] = 0 \quad (F.4) \\
&\left[ \Omega_{3210}^{(1)}, U_3 \otimes U_2 \otimes I_{10} \right] = 0 \quad (F.5) \\
&\left[ \Omega_{3210}^{(1)}, I_{32} \otimes U_1 \otimes U_0 \right] = 0 , \quad (F.6)
\end{align*}
\]

for every group element \( U \in SU(d) \).

To solve the dual problem, we have to find the minimum \( \lambda \) satisfying the relation

\( \lambda \Gamma \geq \Omega \) for some dual comb \( \Gamma \). By Eq. (63), we have \( \Gamma = I_b \otimes I_3 \otimes T_{210a} \), for some suitable operator \( T_{210a} \) satisfying the conditions

\[
\begin{align*}
&\text{Tr}_2 \left[ T_{210a} \right] = I_1 \otimes \rho_{0a} \quad \text{and} \quad \text{Tr}_{0a} \left[ \rho_{0a} \right] = 1 .
\end{align*}
\]

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

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

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

\[
\begin{align*}
&\text{Tr}_2 \left[ T_{210}^{(0)} \right] = p_0 \left[ I_1 \otimes \rho_0^{(0)} \right] \quad \text{and} \quad \text{Tr}_2 \left[ T_{210}^{(1)} \right] = p_1 \left[ I_1 \otimes \rho_0^{(1)} \right] \quad (F.8)
\end{align*}
\]

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

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

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

\[
\left[ T_{210}^{(0)}, I_{21} \otimes U_0 \right] = 0 , \quad \forall U \in SU(d) ,
\]

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

\[
T_{210}^{(0)} = p_0 \frac{I_2 \otimes I_1 \otimes I_0}{d^2} . \quad (F.10)
\]

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

\[
\begin{align*}
&\left[ T_{210}^{(1)}, U_2 \otimes I_{10} \right] = 0 \quad (F.11) \\
&\left[ T_{210}^{(1)}, I_2 \otimes U_1 \otimes U_0 \right] = 0 , \quad (F.12)
\end{align*}
\]

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

\[
T_{210}^{(1)} = I_2 \otimes Q_{10}^{(1)} \quad (F.13)
\]
and Eq. (F.8) implies the condition
\[ dQ_{10}^{(1)} = \text{Tr}_2 \left[ T_{210}^{(1)} \right] \]
\[ = p_1 \left[ I_1 \otimes \rho_0^{(1)} \right] \]
for some probability \( p_1 \) and some quantum state \( \rho_0^{(1)} \). Combining Eqs. (F.13) and (F.12) one finally obtains \( Q_{10}^{(1)} = p_1 I_1 \otimes I_0 / d^2 \), and therefore
\[
T_{210}^{(1)} = p_1 \frac{I_2 \otimes I_1 \otimes I_0}{d^2}.
\]
Inserting the above relations into the dual constraint, we then obtain
\[
\lambda p_0 \frac{I_{3210}^{(1)}}{d^2} \geq \frac{1}{4d^2} (E_{30} \otimes I_{21})
\]
\[
\lambda p_1 \frac{I_{3210}^{(1)}}{d^2} \geq \frac{1}{4d^2} \left( E_{32} \otimes E_{10} + \frac{E_{32}^{\perp} \otimes E_{10}^{\perp}}{d_{\perp}} \right),
\]
having used Eqs. (F.10), and (F.14). To satisfy the constraint, the parameters \( \lambda, p_0, \) and \( p_1 \) must satisfy \( \lambda p_0 \geq 1/4 \) and \( \lambda p_1 \geq 1/4 \), leading to the minimum value \( \lambda_{\text{min}} = 1/2 \).