# Quantifying quantum causal networks

Sumeet Khatri

Hearne Institute for Theoretical Physics, Department of Physics and Astronomy, and Center for Computation and Technology, Louisiana State University, Baton Rouge, Louisiana, 70803, USA

November 13, 2020

**Abstract**

A quantum causal network can be thought of as a quantum channel with multiple inputs and outputs that, due to its causal structure, is constrained beyond the usual complete-positivity and trace-preservation requirements. Quantum causal networks can take adaptive inputs, making them a natural model for interactive protocols in quantum information processing, ranging from quantum communication, quantum metrology, quantum interactive proof systems, and even quantum generalizations of reinforcement learning. In these notes, we provide a review of quantum causal networks and we discuss information quantities and resource measures for them. Specifically, we look at the generalized divergence, fidelity, and Schatten $\alpha$-norms of quantum causal networks, and we study some of their basic properties. We then use these quantities to define resource measures for quantum causal networks. We end with a discussion of some problems for future work on quantum causal networks.

## Table of Contents

# 1 Introduction

A causal network is an object that is used to model the cause-and-effect relationships between physical systems. Classically, they model the causal relationships between random variables [1], and can be used to study the time-evolution of groups of systems that interact with each other. Similarly, quantum causal networks [2–5] model the relationships between quantum systems over time. The generality of quantum causal networks means that they arise in almost every area of quantum information processing, such as adaptive protocols for quantum metrology [6–8], quantum communication [9–14], and quantum error correction [15]. They are also used in the study of quantum interactive proof systems [16], and even in quantum generalizations of reinforcement learning [17, 18].

The formulation of a quantum causal network that we consider is illustrated in Fig. 1, and it is the one presented in [2] (also in [19], where it is referred to as a "quantum strategy"). The lines represent quantum systems, and the nodes, or boxes, represent quantum channels. As an example, the nodes can be thought of as devices, and an agent interacting with the device has access only to the input and output systems $A_i$ and $B_i$, while the systems $E_i$ model the device's internal memory and are inaccessible to the agent. Note that the nodes can also represent the same device at different points in time, so that different channels in each node corresponds to a device that changes its behavior with time.

Being a concatenation of quantum channels, a quantum causal network can simply be viewed as a quantum channel itself. However, a quantum causal network is more than just a quantum channel—its causal structure implies that the channels comprising a quantum causal network can be used *adaptively*, or interactively. In particular, continuing the example above, the agent need not provide the inputs $A_i$ to the device all at once. Instead, it can use the outputs $B_i$ of the device from previous time steps to determine the inputs to the device for future time steps. Typically, the agent's task in such a scenario is to use the device to achieve a desired goal. The agent must therefore determine an optimal sequence of inputs relative to a given figure of merit. From this perspective, it is natural to think of the device, and of the quantum causal network more generally, as a *resource* that the agent uses to accomplish its task.

Resource theories, in the form that they have been considered recently in the quantum information setting [20, 21], are based upon the idea that certain types of objects or physical entities are *free*, meaning that they are allowed to be used to accomplish a certain task. Any object that cannot be created using the free objects alone is called a resource. The idea is that the resources, while still physical, are somehow rare and should be used sparingly. Accordingly, the cost for a particular task is given by the number of resources needed to accomplish it. This viewpoint has been fruitful in the

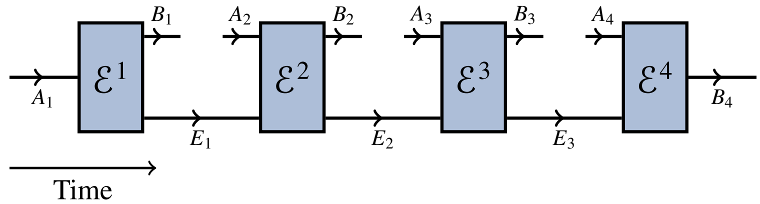



FIGURE 1: An example of a quantum causal network. Time proceeds from left to right. The lines represent quantum systems, and the nodes/boxes represent quantum channels—physical transformations of the input systems into the output systems. The network can be thought as representing a device to which an agent has access via the systems $A_i$ and from which it receives feedback via the systems $B_i$.

analysis of, e.g., quantum communication and quantum thermodynamics. Recently, there has been heightened interest in the development of a general resource theory of quantum channels [22–28] and in the understanding of the limits on resource interconversions, meaning how well one resource can be converted into another resource using free quantum channels only. Although current work on the resource theory of quantum channels covers quantum causal networks in principle, they ostensibly do not take into account explicitly the causal constraints, and thus the error estimates obtained in them are not necessarily best suited to quantum causal networks.

In this note, we set out on the task of understanding the resourcefulness of quantum causal networks. The first step in such a study is quantifying the information contained in a quantum causal network. To this end, we consider a definition of the generalized divergence between two quantum causal networks. Intuitively, the generalized divergence quantifies how "far" the states generated by the two networks are, and it is a natural extension of the definition of the generalized divergence between two quantum channels. In particular, the generalized divergence that we consider reduces to the generalized divergence for ordinary quantum channels if we consider a network with just one node. We also consider the fidelity between quantum causal networks, and we define a notion of Schatten $\alpha$-norm for quantum causal networks that extends prior work on the "strategy norm" [29] for quantum causal networks. An important property that quantities in information theory should satisfy is the data-processing inequality. The rich structure of quantum causal networks leads to a variety of different ways of transforming them, and thus different forms of the data-processing inequality, some of which we investigate. We also consider an application to hypothesis testing and discrimination of quantum causal networks.

The information quantities that we consider here are then used to define resource measures for quantum causal networks, in a similar manner to the study conducted in [22]. In particular, our measures reduce to the measures defined in [22] if we consider a quantum causal network with just one node. We prove some of the necessary properties for these measures, such as data processing and faithfulness, and we lay out resource interconversion problems. Again, the rich structure of quantum causal networks leads to many different kinds of possible transformations, and it also leads to different ways of distilling quantum states and quantum channels. Looking at quantum causal networks as resources also leads to an interesting, physically motivated question of how many rounds of interaction with a given device, modeled as a quantum causal network, are needed in order to accomplish a particular task (problems of this type are relevant in, e.g., reinforcement learning), and we provide the starting point for investigating these problems.

Because all of the measures that we define in this note reduce to the ones defined previously for quantum channels in the case that the quantum causal network contains only one node, one desirable outcome of the development here is that we obtain a general framework for the study of quantum dynamical processes. We also expect the information quantities defined in this note to be useful outside the realm of resource theories, such as investigations of the information-theoretic limits of quantum reinforcement learning [17, 18].

**Outline.** The rest of this note is structured as follows.

- In Sec. 2, we provide a review of quantum causal networks. Along with stating the constraints on the Choi representation of a network due to the causal constraints, an important part of the review is Proposition 1, which pertains to the combination of a causal network with a "pure" network that is central to definitions of our information quantities.
- In Sec. 3, we define information quantities for quantum causal networks, starting with quantities based on generalized divergences in Definition 2. We also define quantities based on fidelity (Definition 6), and we consider a definition of the Schatten $\alpha$-norm for quantum causal networks (Definition 7). In Sec. 3.4 we outline certain types of transformations of quantum causal networks and prove that our quantities satisfy a data-processing inequality under these transformations. As an application of the generalized divergence-based measures, we consider in Sec. 3.5 the tasks of (binary) hypothesis testing of quantum causal networks and the discrimination of multiple quantum causal networks.
- Finally, in Sec. 4, we use the quantities defined in Sec. 3 to define resource measures for quantum causal networks (Definition 11). These have a standard construction as the generalized divergence to a set of free networks. We prove that these resource measures satisfy a data-processing inequality under free networks and are faithful (Theorem 11). We also prove that the commonly-used log-robustness as a resource quantifier for state and channel resource theories does not take the adaptiveness of quantum causal networks into account, meaning that its value is equal to the value under non-adaptive uses of the network. We end Sec. 4 with some interesting directions for future work on transformations of quantum causal networks to quantum states and quantum channels. We provide concluding remarks in Sec. 5.

**Notation.** We let $\mathrm{Lin}(A)$ denote the set of linear operators acting on the finite-dimensional complex vector space (Hilbert space) $\mathcal{H}_A$ associated with the quantum system $A$, and we let $d_A$ denote the dimension of $\mathcal{H}_A$. For every pair $A$ and $B$ of quantum systems, we use $A \cong B$ to signify that the corresponding Hilbert spaces $\mathcal{H}_A$ and $\mathcal{H}_B$ are isomorphic (and thus $d_A = d_B$). For a collection $A_1, \dots, A_r$ of quantum systems, we let $A_1^r \equiv A_1 \cdots A_r$ denote the composite quantum system with Hilbert space $\mathcal{H}_{A_1} \otimes \cdots \otimes \mathcal{H}_{A_r}$. For brevity and simplicity of notation, throughout this note we suppress identity operators in tensor products.

# 2 Review of quantum causal networks

We start with a review of quantum causal networks. In particular, we review the definitions and results from [2, 19, 30, 31], in which the proofs of all of the results being presented can be found.

From now on, for brevity and due the visual look of a quantum causal network as in Fig. 1, we refer to a quantum causal network with $r$ nodes as an "$r$-comb".

## 2.1 Definition and basic properties

A quantum $r$-comb with input systems $A_1, \dots, A_r$, output systems $B_1, \dots, B_r$, and memory systems $E_1, \dots, E_{r-1}$, is depicted in Fig. 1, and it is defined as a sequence $(\mathcal{E}^1, \dots, \mathcal{E}^r)$ of quantum channels $\mathcal{E}^1_{A_1 \to B_1 E_1}$, $\mathcal{E}^j_{A_j E_{j-1} \to B_j E_j}$ for $2 \leq j \leq r-1$, and $\mathcal{E}_{A_r E_{r-1} \to B_r}$. We use the notation

$$\mathcal{E}^{[r]} \equiv (\mathcal{E}^1, \dots, \mathcal{E}^r) \tag{1}$$

to refer to a quantum comb. A quantum comb can be thought of as representing a sequence of actions over time, with the actions represented by quantum channels, so that the channel $\mathcal{E}^j$ is the channel applied at the $j^{\text{th}}$ time step.

The comb in Fig. 1 is often referred to as a *strategy* (see [19]). From a strategy comb, we can derive three other types of combs, as shown in Fig. 2.

- *Measuring strategy*: Illustrated in Fig. 2(a), a measuring strategy is defined by setting the map $\mathcal{E}^4$ to be the measurement channel $\rho \mapsto \sum_{x \in \mathcal{X}} \text{Tr}[\Lambda^x \rho] |x\rangle\langle x|$ corresponding to a positive operator-valued measure (POVM) $\{\Lambda^x\}_{x \in \mathcal{X}}$, where $\mathcal{X}$ is some finite set.
- *Co-strategy*: Illustrated in Fig. 2(b), a co-strategy is defined by setting $d_{A_1} = 1$ and setting the map $\mathcal{E}^1$ to be the preparation channel for a quantum state $\sigma_{B_1 E_1}$.
- *Measuring co-strategy*: Illustrated in Fig. 2(c), a measuring co-strategy is defined by setting $d_{A_1} = 1$, setting the map $\mathcal{E}^1$ to be the preparation channel for a quantum state $\sigma_{B_1 E_1}$, and setting $\mathcal{E}^4$ to be the measurement channel $\rho \mapsto \sum_{x \in \mathcal{X}} \text{Tr}[\Lambda^x \rho] |x\rangle\langle x|$ corresponding to a POVM $\{\Lambda^x\}_{x \in \mathcal{X}}$, where $\mathcal{X}$ is some finite set.

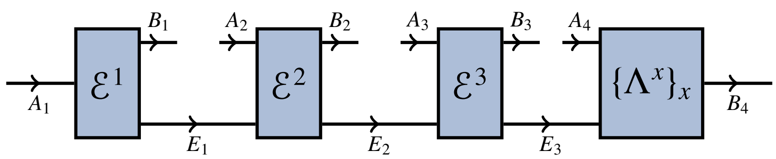
For emphasis, we often write $\mathcal{E}^{[r]}_{(\text{st})}$ to refer to a strategy $r$-comb and $\mathcal{E}^{[r]}_{(\text{co-st})}$ to refer to a co-strategy $r$-comb.

Given an $r$-comb $\mathcal{E}^{[r]} = (\mathcal{E}^1, \dots, \mathcal{E}^r)$ with input system $A_1^r$, output systems $B_1^r$, and memory systems $E_1^{r-1}$, we can define a set $\{\mathcal{E}^{[r_1;r_2]} : 1 \leq r_1 < r_2 \leq r\}$ of quantum combs as follows:

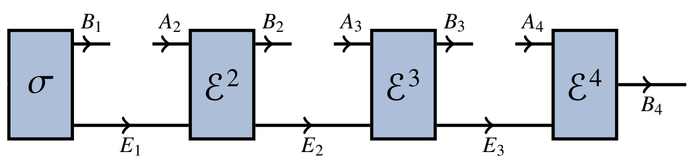
$$\mathcal{E}^{[r_1;r_2]} \coloneqq (\mathcal{E}^{r_1}, \mathcal{E}^{r_1+1}, \dots, \mathcal{E}^{r_2}). \tag{2}$$

In other words, $\mathcal{E}^{[r_1;r_2]}$ is a $(r_2 - r_1 + 1)$-comb with input systems $A_{r_1} E_{r_1 - 1}, \dots, A_{r_2}$, output systems $B_{r_1}, \dots, B_{r_2}, E_{r_2}$, and memory systems $E_{r_1}, \dots, E_{r_2 - 1}$. Another set of "truncated" combs is $\{\mathcal{E}^{[r];k}\}_{k=1}^{r-1}$, where

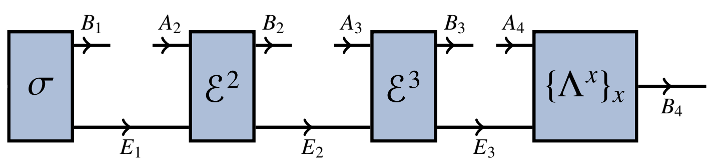
$$\mathcal{E}^{[r];k} \coloneqq (\mathcal{E}^1, \dots, \text{Tr}_{E_k} \circ \mathcal{E}^k). \tag{3}$$




(a) Measuring strategy




(b) Co-strategy




(c) Measuring co-strategy

FIGURE 2: The three types of quantum combs that derive from the quantum comb in Fig. 1. In (a), we set $\mathcal{E}^4$ to be the measurement channel corresponding to the POVM $\{\Lambda^x\}_{x\in\mathcal{X}}$. In (b), we set $d_{A_1} = 1$ and let $\mathcal{E}^1$ be the preparation channel for the state $\sigma_{B_1E_1}$. In (c), we set $d_{A_1} = 1$, we let $\mathcal{E}^1$ be the preparation channel for a quantum state $\sigma_{B_1E_1}$, and we set $\mathcal{E}^4$ to be the measurement channel corresponding to the positive operator-valued measure (POVM) $\{\Lambda^x\}_{x\in\mathcal{X}}$.

In other words, $\mathcal{E}^{[r];k}$ is a $k$-comb with input systems $A_1^k$, output systems $B_1^k$, and memory systems $E_1^{k-1}$, which we obtain by truncating $\mathcal{E}^{[r]}$ at the $k^{\text{th}}$ step and tracing out the memory system $E_k$. We also define the "time-shifted" comb $\mathcal{E}^{[r]+t}$, for $t \geq 0$, as follows:

$$\mathcal{E}^{[r]+t} \coloneqq (\underbrace{\mathrm{id}_{A_1}, \mathrm{id}_{A_1}, \dots, \mathrm{id}_{A_1}}_{t \text{ times}}, \mathcal{E}^1, \mathcal{E}^2, \dots, \mathcal{E}^r), \tag{4}$$

where $\mathrm{id}_{A_1}$ is the identity channel for the system $A_1$, which we recall is the input to the channel $\mathcal{E}^1$. In other words, $\mathcal{E}^{[r]+t}$ is a $(r+t)$-comb in which no action is performed for the first $t$ time steps, and the actions corresponding to the given comb are delayed until the $(t+1)^{\text{st}}$ step. Time-shifting a comb can be useful in the context of tensor products of combs (which we discuss below), because we might want to have a relative time difference between the application of different combs.

As mentioned in the Introduction, any quantum comb can be thought of as a multipartite channel (i.e., a quantum channel with multiple inputs and multiple outputs) with causal constraints. Specifically, to every quantum $r$-comb $\mathcal{E}^{[r]} = (\mathcal{E}^1, \dots, \mathcal{E}^r)$ of quantum channels $\mathcal{E}^j$, with input systems $A_1^r$, output systems $B_1^r$, and memory systems $E_1^{r-1}$, we can associate a quantum channel $\mathcal{N}^{\mathcal{E}^{[r]}} : \mathrm{Lin}(A_1^r) \to \mathrm{Lin}(B_1^r)$ defined as

$$\mathcal{N}^{\mathcal{E}^{[r]}}_{A_1^r \to B_1^r} \coloneqq \mathcal{E}^r_{A_r E_{r-1} \to B_r} \circ \cdots \mathcal{E}^1_{A_1 \to B_1 E_1}. \tag{5}$$

An important object associated with every quantum comb $\mathcal{E}^{[r]}$ is its *Choi representation*, which we denote by $\gamma(\mathcal{E}^{[r]})$, and we define it to be the Choi representation of the associated quantum channel $\mathcal{N}^{\mathcal{E}^{[r]}}$:

$$\gamma(\mathcal{E}^{[r]})_{A_1^r B_1^r} \coloneqq \mathcal{N}^{\mathcal{E}^{[r]}}_{(A')_1^r \to B_1^r} \left( \Gamma_{A_1 A_1'} \otimes \cdots \otimes \Gamma_{A_r A_r'} \right), \tag{6}$$

where $A_j \cong A_j'$ for all $1 \leq j \leq r$, $\Gamma_{A_j A_j'} \coloneqq |\Gamma\rangle\langle\Gamma|_{A_j A_j'}$ and

$$|\Gamma\rangle_{A_j A_j'} \coloneqq \sum_{i=0}^{d_{A_j}-1} |i, i\rangle_{A_j A_j'}. \tag{7}$$

The Choi representation of a quantum comb is positive semidefinite, as for all quantum channels. However, the causal structure of a quantum comb implies further constraints on its Choi representation. In particular, using the definition of the truncated combs $\mathcal{E}^{[r];k}$ in (3), it is straightforward to verify that

$$\mathrm{Tr}_{B_r}\left[\gamma(\mathcal{E}^{[r]})_{A_1^r B_1^r}\right] = \gamma(\mathcal{E}^{[r];r-1})_{A_1^{r-1} B_1^{r-1}} \otimes \mathbb{1}_{A_r}, \tag{8}$$

$$\mathrm{Tr}_{B_k}\left[\gamma(\mathcal{E}^{[r];k})_{A_1^k B_1^k}\right] = \gamma(\mathcal{E}^{[r];k-1})_{A_1^{k-1} B_1^{k-1}} \otimes \mathbb{1}_{A_k} \quad \forall\, r-1 \geq k \geq 2, \tag{9}$$

$$\mathrm{Tr}_{B_1}\left[\gamma(\mathcal{E}^{[r];1})_{A_1 B_1}\right] = \mathbb{1}_{A_1}. \tag{10}$$

An important fact is that the converse statement is also true: any set $\{C^{(k)}_{A_1^k B_1^k}\}_{k=1}^r$ of positive semi-definite operators satisfying the constraints in (8)–(10) can be associated with an $r$-comb of quantum channels.

**Theorem 1** (Characterization of quantum combs [2, 19]). *The set of quantum $r$-combs of quantum channels with input systems $A_1, \dots, A_r$ and output systems $B_1, \dots, B_r$ is in one-to-one correspondence with operators $C^{(k)}_{A_1^k B_1^k}$, $1 \leq k \leq r$, satisfying the following constraints:*

$$C^{(k)}_{A_1^k B_1^k} \geq 0 \quad \forall\ 1 \leq k \leq r, \tag{11}$$

$$\mathrm{Tr}_{B_k}\left[C^{(k)}_{A_1^k B_1^k}\right] = C^{(k-1)}_{A_1^{k-1} B_1^{k-1}} \otimes \mathbb{1}_{A_k} \quad \forall\ r \geq k \geq 2, \tag{12}$$

$$\mathrm{Tr}_{B_1}\left[C^{(1)}_{A_1 B_1}\right] = \mathbb{1}_{A_1}. \tag{13}$$

*Furthermore, given such a set of operators, the channels $\mathcal{E}^1_{A_1 \to B_1 E_1}$, $\mathcal{E}^j_{A_j E_{j-1} \to B_j E_j}$ for $2 \leq j \leq r-1$, and $\mathcal{E}^r_{A_r E_{r-1} \to B_r}$, can be constructed as isometric channels such that the dimensions of the memory systems $E_1, \dots, E_{r-1}$ satisfy $d_{E_k} \leq d_{A_1^k} d_{B_1^k}$ for all $1 \leq k \leq r-1$.*

For convenience, let us explicitly express the conditions in (11)–(13) for the three special types of quantum combs defined in Fig. 2.

**Constraints for a measuring strategy.** Using the fact that the Choi representation of the measurement channel $\rho \mapsto \sum_{x \in \mathcal{X}} \mathrm{Tr}[\Lambda^x \rho] |x\rangle\langle x|$ is $\sum_{x \in \mathcal{X}} (\Lambda^x)^{\mathsf{T}} \otimes |x\rangle\langle x|$, we have that

$$\gamma(\mathcal{E}^{[r]})_{A_1^r B_1^r} = \sum_{x \in \mathcal{X}} C^{(r);x}_{A_1^r B_1^{r-1}} \otimes |x\rangle\langle x|_{B_r}, \tag{14}$$

where

$$C^{(r);x}_{A_1^r B_1^{r-1}} \coloneqq \mathrm{Tr}_{E_{r-1}}\left[\left(\Lambda^x_{A_r E_{r-1}}\right)^{\mathsf{T}_{A_r}} \gamma((\mathcal{E}^1, \dots, \mathcal{E}^{r-1}))_{A_1^{r-1} B_1^{r-1} E_{r-1}}\right] \tag{15}$$

are positive semi-definite operators corresponding to each element of the POVM for the measurement channel. The constraint in (8) then translates to

$$\sum_{x \in \mathcal{X}} C^{(r);x}_{A_1^r B_1^{r-1}} = C^{(r-1)}_{A_1^{r-1} B_1^{r-1}} \otimes \mathbb{1}_{A_r}. \tag{16}$$

In other words, any measurement strategy $r$-comb is in one-to-one correspondence with operators $C^{(r);x}_{A_1^r B_1^{r-1}}$ for all $x \in \mathcal{X}$ and operators $C^{(k)}_{A_1^k B_1^k}$ for all $1 \leq k \leq r-1$ satisfying the following constraints:

$$C^{(r);x}_{A_1^r B_1^{r-1}} \geq 0 \quad \forall x \in \mathcal{X}, \quad C^{(k)}_{A_1^k B_1^k} \geq 0 \quad \forall\ 1 \leq k \leq r-1, \tag{17}$$

$$\sum_{x \in \mathcal{X}} C^{(r);x}_{A_1^r B_1^{r-1}} = C^{(r-1)}_{A_1^{r-1} B_1^{r-1}} \otimes \mathbb{1}_{A_r}, \tag{18}$$

$$\mathrm{Tr}_{B_k}\left[C^{(k)}_{A_1^k B_1^k}\right] = C^{(k-1)}_{A_1^{k-1} B_1^{k-1}} \otimes \mathbb{1}_{A_k} \quad \forall\ r-1 \geq k \geq 2, \tag{19}$$

$$\mathrm{Tr}_{B_1}\left[C^{(1)}_{A_1 B_1}\right] = \mathbb{1}_{A_1}. \tag{20}$$

**Constraints for a co-strategy.** A co-strategy is defined by setting $\mathcal{E}^1$ to be the preparation channel for a state $\sigma_{B_1E_1}$, so that $d_{A_1} = 1$. Then, because $\mathbb{1}_{A_1} = 1$, the conditions in (11)–(13) reduce to

$$\mathrm{Tr}_{B_k}\left[C^{(k)}_{A_1^k B_1^k}\right] = C^{(k-1)}_{A_1^{k-1}B_1^{k-1}} \otimes \mathbb{1}_{A_k} \quad \forall\, r \geq k \geq 2, \tag{21}$$

$$\mathrm{Tr}_{B_1}\left[C^{(1)}_{B_1}\right] = 1. \tag{22}$$

Intuitively, the operator $C^{(1)}_{B_1}$ corresponds to the partial trace of the state $\sigma_{B_1E_1}$ over $E_1$, i.e., $C^{(1)}_{B_1} \equiv \mathrm{Tr}_{E_1}[\sigma_{B_1E_1}]$.

**Constraints for a measuring co-strategy.** We can combine the conditions above for a measuring strategy and a co-strategy to get

$$C^{(r);x}_{A_1^r B_1^{r-1}} \geq 0 \quad \forall x \in \mathcal{X}, \quad C^{(k)}_{A_1^k B_1^k} \geq 0 \quad \forall\, 1 \leq k \leq r-1, \tag{23}$$

$$\sum_{x\in\mathcal{X}} C^{(r);x}_{A_1^r B_1^{r-1}} = C^{(r-1)}_{A_1^{r-1}B_1^{r-1}} \otimes \mathbb{1}_{A_r}, \tag{24}$$

$$\mathrm{Tr}_{B_k}\left[C^{(k)}_{A_1^k B_1^k}\right] = C^{(k-1)}_{A_1^{k-1}B_1^{k-1}} \otimes \mathbb{1}_{A_k} \quad \forall\, r-1 \geq k \geq 2, \tag{25}$$

$$\mathrm{Tr}_{B_1}\left[C^{(1)}_{B_1}\right] = 1. \tag{26}$$

In other words, a measuring co-strategy $r$-comb is defined by a set $\{C^{r;(x)}_{A_1^r B_1^{r-1}}\}_{x\in\mathcal{X}}$ of positive semi-definite operators such that

$$\sum_{x\in\mathcal{X}} C^{(r);x}_{A_1^r B_1^{r-1}} = \mathrm{Tr}_{B_r}\left[\gamma(\mathcal{D}^{[r]}_{(\text{co-st})})\right], \tag{27}$$

where $\mathcal{D}^{[r]}_{(\text{co-st})}$ is some co-strategy $r$-comb with input systems $A_1, \ldots, A_r$ and output systems $B_1, \ldots, B_r$.

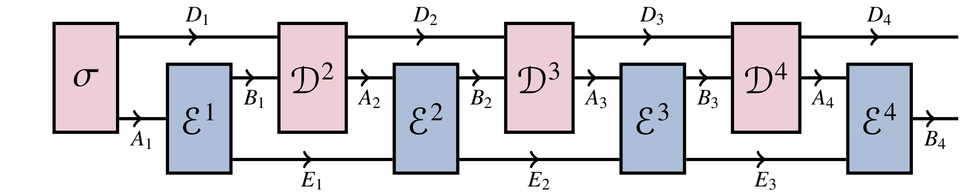
## 2.2 Combining combs

As described above, a quantum comb can be thought of simply as a multipartite quantum channel. However, a comb is more general, because it can take as input another comb and output another comb. Suppose we are given two combs $\mathcal{E}_1^{[r_1]}$ and $\mathcal{E}_2^{[r_2]}$, with some subset of equal input and output systems. Let us denote the collection of these systems that are common to the two combs by $S_{\text{in}}$, and let us refer to the remaining (uncommon) input and output systems of $\mathcal{E}_1^{[r_1]}$ and $\mathcal{E}_2^{[r_2]}$ by $S_1$ and $S_2$, respectively. To *combine*, or *link* the combs means to join the common systems of the two combs to create another comb. We denote the output of combining the two combs by

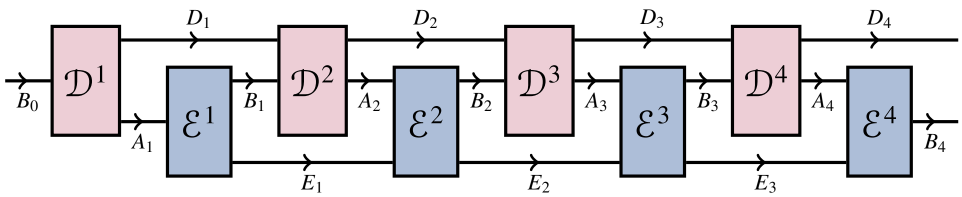
$$\mathcal{E}_1^{[r_1]} \circ \mathcal{E}_2^{[r_2]}, \tag{28}$$

and it is the comb whose Choi representation is given by

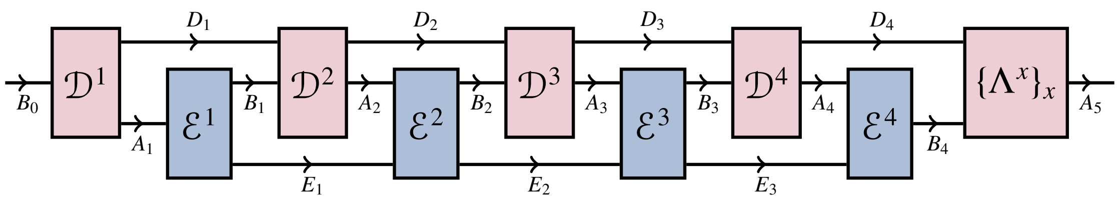
$$\gamma(\mathcal{E}_1^{[r_1]} \circ \mathcal{E}_2^{[r_2]})_{S_1S_2} = \gamma(\mathcal{E}_1^{[r_1]})_{S_1S_{\text{in}}} * \gamma(\mathcal{E}_2^{[r_2]})_{S_2S_{\text{in}}}, \tag{29}$$



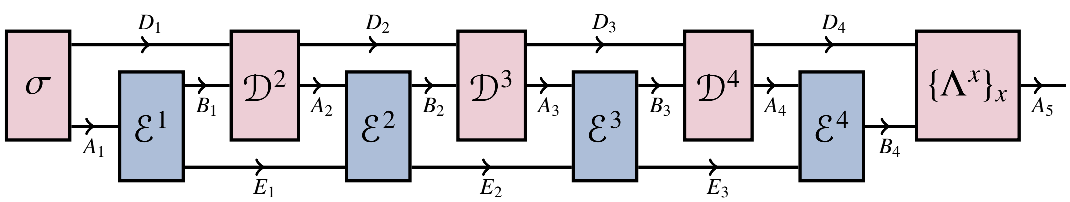

(a) Strategy + co-strategy




(b) Strategy + strategy




(c) Strategy + measuring strategy




(d) Strategy + measuring co-strategy

FIGURE 3: Four simple ways of combining strategy and co-strategy combs. In (a), we combine a strategy 4-comb with a co-strategy 4-comb to obtain a quantum state for the systems $D_4$ and $B_4$. In (b), we combine a strategy 4-comb with a strategy 4-comb to obtain a quantum channel $B_0 \to D_4B_4$. In (c), we combine a strategy 4-comb with a measuring strategy 5-comb to obtain a quantum-classical channel $B_0 \to A_5$. In (d), we combine a strategy 4-comb with a measuring co-strategy 5-comb to obtain a classical register $A_5$ containing the output of the measurement at the final step.

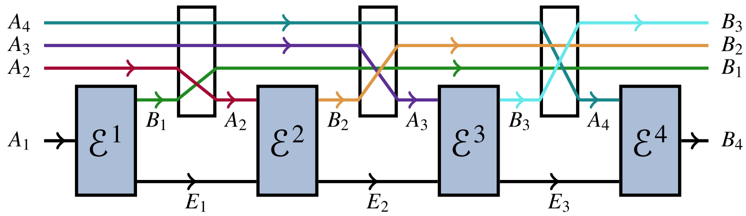



FIGURE 4: A co-strategy comb $\mathcal{D}^{[4]}_{(\text{co-st})}$ such that $\mathcal{E}^{[4]} \circ \mathcal{D}^{[4]}_{(\text{co-st})} = \mathcal{N}^{\mathcal{E}^{[4]}}_{A_1^4 \to B_1^4}(\sigma_{A_1^4})$, where $\sigma_{A_1^4}$ is the state prepared in the first step of the comb $\mathcal{D}^{[4]}_{(\text{co-st})}$. In other words, this co-strategy corresponds to using the comb $\mathcal{E}^{[4]}$ as an ordinary quantum channel.

where $\gamma(\mathcal{E}_1^{[r_1]})_{S_1 S_{\text{in}}} * \gamma(\mathcal{E}_2^{[r_2]})_{S_2 S_{\text{in}}}$ is the *link product* of $\gamma(\mathcal{E}_1^{[r_1]})$ and $\gamma(\mathcal{E}_2^{[r_2]})$, defined as

$$\gamma(\mathcal{E}_1^{[r_1]})_{S_1 S_{\text{in}}} * \gamma(\mathcal{E}_2^{[r_2]})_{S_2 S_{\text{in}}} := \operatorname{Tr}_{S_{\text{in}}}\left[\gamma(\mathcal{E}_1^{[r_1]})_{S_1 S_{\text{in}}}^{\mathsf{T}_{S_{\text{in}}}} \gamma(\mathcal{E}_2^{[r_2]})_{S_2 S_{\text{in}}}\right]. \tag{30}$$

The input and output systems of the comb $\mathcal{E}_1^{[r_1]} \circ \mathcal{E}_2^{[r_2]}$ depend on which of the uncommon systems between the constituent combs $\mathcal{E}_1^{[r_1]}$ and $\mathcal{E}_2^{[r_2]}$ are inputs and outputs. In Fig. 3, we illustrate four possible ways of combining strategy and co-strategy combs. In Sec. 3.4, we consider more general transformations of combs.

Let us elaborate on the combination of a strategy with a co-strategy. Let $\mathcal{E}^{[r]}_{(\text{st})}$ be a strategy $r$-comb with input systems $A_1, \dots, A_r$, output systems $B_1, \dots, B_r$, and memory systems $E_1, \dots, E_{r-1}$, and let $\mathcal{D}^{[r]}_{(\text{co-st})}$ be a co-strategy $r$-comb with input systems $B_1, \dots, B_{r-1}$, output systems $A_1, \dots, A_r, D_r$, and memory systems $D_1, \dots, D_{r-1}$; see Fig. 3(a) for the case $r = 4$. The combination of $\mathcal{E}^{[r]}_{(\text{st})}$ and $\mathcal{D}^{[r]}_{(\text{st})}$ is a quantum state for the systems $D_r$ and $B_r$, and we use the notation

$$\rho_{D_r B_r}^{(\mathcal{E}^{[r]}_{(\text{st})}, \mathcal{D}^{[r]}_{(\text{co-st})})} \equiv \mathcal{D}^{[r]}_{(\text{co-st})} \circ \mathcal{E}^{[r]}_{(\text{st})}. \tag{31}$$

Note that, because the output is a quantum state, we have that

$$\begin{aligned} \rho_{D_r B_r}^{(\mathcal{E}^{[r]}_{(\text{st})}, \mathcal{D}^{[r]}_{(\text{co-st})})} &= \gamma(\mathcal{D}^{[r]}_{(\text{co-st})}) * \gamma(\mathcal{E}^{[r]}_{(\text{st})}) &(32)\\ &= \mathcal{D}^{[r]}_{(\text{co-st})}) \circ \gamma(\mathcal{E}^{[r]}_{(\text{st})} &(33)\\ &= \operatorname{Tr}_{A_1^r B_1^{r-1}}\left[\gamma(\mathcal{D}^{[r]}_{(\text{co-st})})_{B_1^{r-1} A_1^r D_r} \gamma(\mathcal{E}^{[r]}_{(\text{st})})_{A_1^r B_1^r}^{\mathsf{T}_{A_1^r B_1^{r-1}}}\right]. &(34) \end{aligned}$$

Furthermore, if $\mathcal{E}^{[r]}_{(\text{st})} = (\mathcal{E}^1_{A_1 \to B_1 E_1}, \mathcal{E}^2_{A_2 E_1 \to B_2 E_2}, \dots, \mathcal{E}^r_{A_r E_{r-1} \to B_r})$ and $\mathcal{D}^{[r]}_{(\text{co-st})} = (\sigma_{A_1 D_1}, \mathcal{D}^2_{B_1 D_1 \to A_2 D_2}, \dots, \mathcal{D}^r_{D_{r-1} B_{r-1} \to A_r D_r})$, then

$$\rho_{D_r B_r}^{(\mathcal{E}^{[r]}_{(\text{st})}, \mathcal{D}^{[r]}_{(\text{co-st})})} = (\mathcal{E}^4_{A_r E_{r-1} \to B_r} \circ \mathcal{D}^r_{D_{r-1} B_{r-1} \to A_r D_r} \circ \dots \circ \mathcal{E}^1_{A_1 \to B_1 E_1})(\sigma_{A_1 E_1}). \tag{35}$$

A special case of a co-strategy is the one shown in Fig. 4 for the case $r = 4$. This co-strategy 4-comb is defined by letting the memory systems be $D_1 = A_2 A_3 A_4$, $D_2 = B_1 A_3 A_4$, $D_3 = B_1 B_2 A_4$, and $D_4 = B_1B_2B_3$. The channels $\mathcal{D}^j$, $2 \leq j \leq 4$, are defined to be swap channels as shown, and $\mathcal{D}^1$ prepares a quantum state $\sigma_{A_1^4}$. Combining this co-strategy with the strategy $\mathcal{E}^{[4]}$ leads to

$$\rho_{D_4B_4}^{(\mathcal{E}^{[4]},\mathcal{D}^{[4]}_{\text{(co-st)}})} = \mathcal{N}^{\mathcal{E}^{[4]}}_{A_1^4\to B_1^4}(\sigma_{A_1^4}) \tag{36}$$

We thus see that all strategy $r$-combs can be used as an ordinary quantum channel by combining them with the co-strategy comb of the form shown in Fig. 4, which has a straightforward generalization to an arbitrary number $r$ of rounds.

The following is an important fact that we use later. See [2, Theorem 12] for a closely related result.

**Proposition 1.** *Consider a strategy $r$-comb $\mathcal{E}^{[r]}$ with input systems $A_1,\dots,A_r$ and output systems $B_1,\dots,B_r$. Let $\mathcal{D}^{[r]}_{(co\text{-}st)} = (\sigma_{D_1A_1}, \mathcal{D}^2_{D_1B_1\to D_2A_2},\dots,\mathcal{D}^r_{D_{r-1}B_{r-1}\to D_rA_r})$ be a co-strategy consisting of an initial pure state and isometric channels $\mathcal{D}^k$, with input systems $B_1,\dots,B_{r-1}$, output systems $A_1,\dots,A_r,D_r$ such that $D_r \cong A_1^rB_1^{r-1}$, and memory systems $D_k$ such that $D_1 \cong A_1$ and $D_k \cong D_{k-1}A_kB_{k-1}$ for all $2 \leq k \leq r-1$; see Fig. 3(a)[1]. Then, there exists a linear operator $V_{A_1^rB_1^{r-1}}$ (as a function of $\mathcal{D}^{[r]}$) such that*

$$\gamma(\mathcal{E}^{[r]}) * \gamma(\mathcal{D}^{[r]}) = V\gamma(\mathcal{E}^{[r]})V^\dagger. \tag{37}$$

*In particular,*

$$V^{\mathsf{T}}\overline{V} = \gamma((\mathcal{D}^1,\dots,\mathrm{Tr}_{D_r}\circ\mathcal{D}^r)). \tag{38}$$

*Proof.* We start with the fact that the co-strategy comb $\mathcal{D}^{[r]}$ consists entirely of isometric channels, which implies that its Choi representation is a unit-rank operator: $\gamma(\mathcal{D}^{[r]}) = |\Psi\rangle\langle\Psi|_{B_1^{r-1}A_1^rD_r}$. Now, by assumption, it holds that $D_r = A_1^rB_1^{1-r}$. Letting $S \equiv B_1^{r-1}A_1^r$ and $S' \equiv D_r$, we have that $d_S = d_{S'}$, and it holds that there exists a linear operator $V_{S'}$ such that

$$|\Psi\rangle_{B_1^{r-1}A_1^rD_r} \equiv |\Psi\rangle_{SS'} = (\mathbb{1}_S \otimes V_{S'})|\Gamma\rangle_{SS'}. \tag{39}$$

Then, the link product $\gamma(\mathcal{E}^{[r]}) * \gamma(\mathcal{D}^{[r]})$ is

$$\gamma(\mathcal{E}^{[r]}) * \gamma(\mathcal{D}^{[r]}) = \mathrm{Tr}_{B_1^{r-1}A_1^r}\left[\gamma(\mathcal{E}^{[r]})_{A_1^rB_1^r}^{\mathsf{T}_{B_1^{r-1}A_1^r}}\gamma(\mathcal{D}^{[r]})_{B_1^{r-1}A_1^rD_r}\right] \tag{40}$$

$$= \mathrm{Tr}_S\left[\gamma(\mathcal{E}^{[r]})_{SB_r}^{\mathsf{T}_S}V_{S'}\Gamma_{SS'}V_{S'}^\dagger\right]. \tag{41}$$

Now, the operator $\gamma(\mathcal{E}^{[r]})_{SB_r}$ can in general be decomposed as follows:

$$\gamma(\mathcal{E}^{[r]})_{SB_r} = \sum_{i,j} X_S^i \otimes Y_{B_r}^j, \tag{42}$$

for some sets $\{X_S^i\}_i$ and $\{Y_{B_r}^j\}_j$ of linear operators. Using this, and making use of the fact that

$$Z_S|\Gamma\rangle_{SS'} = Z_{S'}^{\mathsf{T}}|\Gamma\rangle_{SS'} \tag{43}$$

[1] Note that we consider the system $D_r$ to be an output system of the comb $\mathcal{D}^{[r]}$ and not a memory system.

for every (square) linear operator $Z_S$, we obtain

$$\gamma(\mathcal{E}^{[r]}) * \gamma(\mathcal{D}^{[r]}) = \mathrm{Tr}_S\left[\gamma(\mathcal{E}^{[r]})^{\intercal_S} V_{S'}\Gamma_{SS'}V_{S'}^{\dagger}\right] \tag{44}$$

$$= \sum_{i,j} \mathrm{Tr}_S\left[(X_S^i)^{\intercal} V_{S'}\Gamma_{SS'}V_{S'}^{\dagger}\right] \otimes Y_{B_r}^j \tag{45}$$

$$= \sum_{i,j} \mathrm{Tr}_S\left[V_{S'}X_{S'}^i\Gamma_{SS'}V_{S'}^{\dagger}\right] \otimes Y_{B_r}^j \tag{46}$$

$$= \sum_{i,j} V_{S'}X_{S'}^i V_{S'}^{\dagger} \otimes Y_{B_r}^j \tag{47}$$

$$= V_{S'}\gamma(\mathcal{E}^{[r]})_{S'B_r}V_{S'}^{\dagger}, \tag{48}$$

which is precisely (37), as required. Now, to see (38), we use (43) in (39) to get

$$\mathrm{Tr}_{D_r}[\gamma(\mathcal{D}^{[r]})_{B_1^{r-1}A_1^rD_r}] = \mathrm{Tr}_{S'}[V_{S'}\Gamma_{SS'}V_{S'}^{\dagger}] \tag{49}$$

$$= \mathrm{Tr}_{S'}[V_S^{\intercal}\Gamma_{SS'}\overline{V}_S] \tag{50}$$

$$= V_S^{\intercal}\overline{V}_S. \tag{51}$$

Then, because

$$\mathrm{Tr}_{D_r}[\gamma(\mathcal{D}^{[r]})] = \gamma((\mathcal{D}^1, \dots, \mathrm{Tr}_{D_r} \circ \mathcal{D}^r)), \tag{52}$$

we obtain the desired result. □

**Remark 2.** *We refer to co-strategy $r$-combs $\mathcal{D}^{[r]}_{\text{(co-st)}}$ that consist of a pure state at the beginning followed by isometric channels as "pure" co-strategies, because their Choi representation is a unit-rank operator of the form $|\Psi\rangle\langle\Psi|$.*

*Observe that Proposition 1 holds more generally: for any operator $X \in \mathrm{Lin}(A_1^rB_1^r)$ and any co-strategy $r$-comb $\mathcal{D}^{[r]}_{\text{(co-st)}}$ with input systems $B_1\dots,B_{r-1}$, output systems $A_1,\dots,A_r,D_r$ with $D_r \cong A_1^rB_1^{r-1}$, and memory systems $D_k$ such that $D_1 \cong A_1$ and $D_k \cong D_{k-1}A_kB_{k-1}$ for all $2 \le k \le r-1$, it holds that*

$$X * \gamma(\mathcal{D}^{[r]}_{\text{(co-st)}}) = V_{A_1^rB_1^{r-1}}X_{A_1^rB_1^r}V_{A_1^rB_1^{r-1}}^{\dagger}, \tag{53}$$

*where the operator $V_{A_1^rB_1^{r-1}}$ satisfies*

$$V^{\intercal}\overline{V} = \gamma((\mathcal{D}^1, \dots, \mathrm{Tr}_{D_r} \circ \mathcal{D}^r)). \tag{54}$$

*We can further generalize Proposition 1 by assuming that the co-strategy comb $\mathcal{D}^{[r]}$ consists of general linear maps, not just quantum channels. In this case, with the same assumptions on the dimensions of the systems $D_k$, $1 \le k \le r$, we have that instead of isometric channels it suffices to take each map $\mathcal{D}^k$ to be of the form $\mathcal{D}^k(\cdot) = Y_k(\cdot)Z_k^{\dagger}$, where $\{(Y_k, Z_k)\}_{k=1}^r$ are pairs of linear operators. Then, it follows that the Choi representation of $\mathcal{D}^{[r]}$ is a unit-rank operator of the form $\gamma(\mathcal{D}^{[r]}) = |\Psi\rangle\langle\Phi|_{B_1^{r-1}A_1^rD_r}$, and there exist operators $V_{S'}$ and $W_{S'}$ such that*

$$|\Psi\rangle_{SS'} = (\mathbb{1}_S \otimes V_{S'})|\Gamma\rangle_{SS'}, \tag{55}$$

$$|\Phi\rangle_{SS'} = (\mathbb{1}_S \otimes W_{S'})|\Gamma\rangle_{SS'}. \tag{56}$$

*The steps in (44)–(48) and (49)–(51) follow through analogously to before, and we obtain*

$$X * \gamma(\mathcal{D}^{[r]}_{\textit{(co-st)}}) = V_{A_1^r B_1^{r-1}} X_{A_1^r B_1^r} W^\dagger_{A_1^r B_1^{r-1}}, \tag{57}$$

$$V^\intercal \overline{W} = \gamma((\mathcal{D}^1, \ldots, \mathrm{Tr}_{D_r} \circ \mathcal{D}^r)). \tag{58}$$

So far, we have discussed combining combs via composition. Let us now briefly discuss taking tensor products of combs. Roughly speaking, taking the tensor product of quantum channels corresponding to applying the channels in parallel, i.e., at the same time. This idea can be used to define what we mean by a tensor product of quantum combs. Specifically, let $\mathcal{E}^{[r]} = (\mathcal{E}^1, \ldots, \mathcal{E}^r)$ be an $r$-comb with input systems $A_1^r$, output systems $B_1^r$, and memory systems $E_1^{r-1}$. Let $\mathcal{M}^{[r]} = (\mathcal{M}^1, \ldots, \mathcal{M}^r)$ be an $r$-comb with input systems $C_1^r$, output systems $D_1^r$, and memory systems $M_1^{r-1}$. Then,

$$\mathcal{E}^{[r]} \otimes \mathcal{M}^{[r]} \coloneqq (\mathcal{E}^1 \otimes \mathcal{M}^1, \ldots, \mathcal{E}^r \otimes \mathcal{M}^r) \tag{59}$$

is an $r$-comb with input systems $A_1^r C_1^r$, output systems $B_1^r D_1^r$, and memory systems $E_1^{r-1} M_1^{r-1}$; see Fig. 5(a) for an illustration. In general,

$$\mathcal{E}^{[r_1]} \otimes \mathcal{M}^{[r_2]} = \mathcal{F}^{[r']}, \quad r' = \max\{r_1, r_2\}, \tag{60}$$

is the result of taking the tensor product of combs with different lengths $r_1$ and $r_2$. For example, if $r_1 < r_2$, then

$$\mathcal{F}^j = \begin{cases} \mathcal{E}^j \otimes \mathcal{M}^j & j \leq \min\{r_1, r_2\}, \\ \mathcal{M}^j & j > \min\{r_1, r_2\}, \end{cases} \tag{61}$$

so that

$$\mathcal{E}^{[r_1]} \otimes \mathcal{M}^{[r_2]} = (\mathcal{E}^1 \otimes \mathcal{M}^1, \mathcal{E}^2 \otimes \mathcal{M}^2, \ldots, \mathcal{E}^{r_1} \otimes \mathcal{M}^{r_1}, \mathcal{M}^{r_1+1}, \ldots, \mathcal{M}^{r_2}), \tag{62}$$

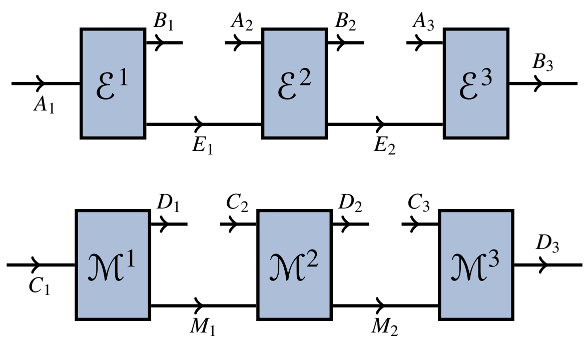
where for the channels after the $r_1^{\text{st}}$ time step we have suppressed the identity channel acting on the final output system of $\mathcal{E}^{[r_1]}$. We can also take the tensor product of time-shifted combs. For example, if we shift the comb $\mathcal{M}^{[r_2]}$ by one time step and then take the tensor product with $\mathcal{E}^{[r_1]}$, we obtain

$$\mathcal{E}^{[r_1]} \otimes \mathcal{M}^{[r_2]+1} = (\mathcal{E}^1, \mathcal{E}^2 \otimes \mathcal{M}^1, \ldots, \mathcal{E}^{r_1} \otimes \mathcal{M}^{r_1-1}, \mathcal{M}^{r_1}, \ldots, \mathcal{M}^{r_2}), \tag{63}$$

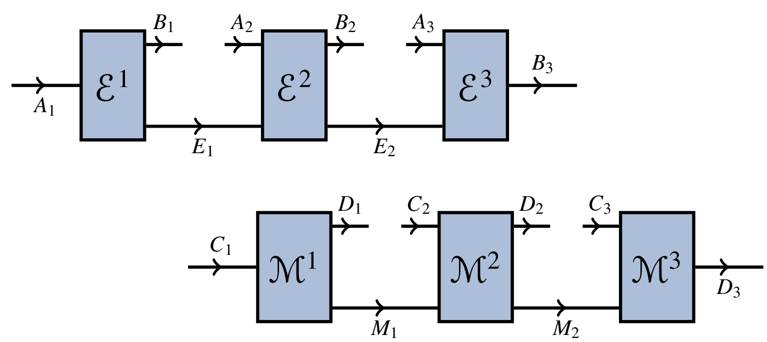
which is a $(r_2 + 1)$-comb; see Fig. 5(b) for an illustration. We refer to [2, Sec. IV.B] for a more detailed discussion on the tensor product of quantum combs.

## 3 Information quantities

We now proceed to defining information quantities for quantum combs. We start with quantities induced by a generalized divergence for quantum states. We then define fidelities for quantum combs using fidelities for states before proceeding to a definition of the Schatten $\alpha$-norm for quantum combs. We also discuss quantities defined using the concept of amortization, and we discuss combining strategies in the context of data processing.




(a)




(b)

FIGURE 5: Tensor products of the quantum 3-combs $\mathcal{E}^{[3]} = (\mathcal{E}^1, \mathcal{E}^2, \mathcal{E}^3)$ and $\mathcal{M}^{[3]} = (\mathcal{M}^1, \mathcal{M}^2, \mathcal{M}^3)$. (a) The tensor product $\mathcal{E}^{[3]} \otimes \mathcal{M}^{[3]}$ is another 3-comb. (b) The tensor product of $\mathcal{E}^{[3]}$ with the time-shifted comb $\mathcal{M}^{[3]+1}$ leads to a 4-comb.

### 3.1 Generalized divergence-based quantities

A generalized divergence $\boldsymbol{D}$ [32, 33] is a function $(\rho, \sigma) \mapsto \boldsymbol{D}(\rho\|\sigma)$, where $\rho$ and $\sigma$ are quantum states, that satisfies the data-processing inequality:

$$\boldsymbol{D}(\mathcal{N}(\rho)\|\mathcal{N}(\sigma)) \leq \boldsymbol{D}(\rho\|\sigma), \tag{64}$$

where $\mathcal{N}$ is any quantum channel. The well-known quantum relative entropy [34] is an example of a generalized divergence, as are quantum Rényi relative entropies; see [35] for a review. The trace distance $\frac{1}{2}\|\rho - \sigma\|_1$ is also a generalized divergence. Intuitively, a generalized divergence quantifies how far apart two quantum states are, although a generalized divergence is not required to be a distance/metric in the mathematical sense.

Given a generalized divergence $\boldsymbol{D}$ on quantum states, as defined above, we have the following definition of the generalized divergence for quantum channels $\mathcal{N}$ and $\mathcal{M}$ [36]:

$$\boldsymbol{D}(\mathcal{E}\|\mathcal{M}) = \sup_{\rho_{RA}} \boldsymbol{D}(\mathcal{E}_{A\to B}(\rho_{RA})\|\mathcal{M}_{A\to B}(\rho_{RA})), \tag{65}$$

where the optimization is over quantum states $\rho_{RA}$, with the dimension of $R$ unrestricted in general, although it is straightforward to show that it suffices to let $R \cong A$. In other words, in order to define a generalized divergence between two channels, we ultimately resort to the corresponding generalized divergence for states by acting with the channels on one-half of a bipartite state $\rho_{RA}$, over which we optimize. The diamond distance $\frac{1}{2}\|\mathcal{N} - \mathcal{M}\|_\diamond$ [37] is a well-known example of a generalized divergence for quantum channels.

By noting that a quantum channel is nothing but a strategy 1-comb, and by noting that $\mathcal{N}_{A\to B}(\rho_{RA})$ is the result of applying the co-strategy 1-comb defined by the preparation channel for $\rho_{RA}$, we see that the procedure to calculate $\boldsymbol{D}(\mathcal{N}\|\mathcal{M})$ is simply a special case of the combination of a strategy with a co-strategy depicted in Figure 3(a). With this insight, the generalization of (65) to quantum strategy $r$-combs, $r \geq 1$, is immediate; see [38, Definition 1]

**Definition 2** (Generalized divergence between quantum combs)**.** *Let $r \geq 1$. Given two strategy $r$-combs $\mathcal{E}^{[r]}$ and $\mathcal{M}^{[r]}$ with input systems $A_1^r$ and output systems $B_1^r$, we define the generalized divergence between them as follows:*

$$\boldsymbol{D}_r\left(\mathcal{E}^{[r]}\middle\|\mathcal{M}^{[r]}\right) \coloneqq \sup_{\mathcal{D}^{[r]}_{\text{(co-st)}}} \boldsymbol{D}\left(\rho_{D_rB_r}^{(\mathcal{E}^{[r]},\mathcal{D}^{[r]}_{\text{(co-st)}})}\middle\|\rho_{D_rB_r}^{(\mathcal{M}^{[r]},\mathcal{D}^{[r]}_{\text{(co-st)}})}\right) \tag{66}$$

*where the optimization is with respect to $r$-round co-strategy $r$-combs $\mathcal{D}^{[r]}_{\text{(co-st)}}$ with input systems $B_1^{r-1}$ and output systems $A_1^rD_r$; see Fig. 3(a). Using Theorem 1 along with (32), the generalized* *divergence is given by the solution to the following optimization problem:*

$$\begin{aligned}
\textit{maximize}\quad & \boldsymbol{D}\left(\gamma(\mathcal{E}^{[r]}) * C^{(r)} \middle\| \gamma(\mathcal{M}^{[r]}) * C^{(r)}\right) \\
\textit{subject to}\quad & C^{(r)} \geq 0, \\
& \mathrm{Tr}_{A_r D_r}[C^{(r)}_{B_1^{r-1} A_1^r D_r}] = C^{(r-1)}_{B_1^{r-2} A_1^{r-1}} \otimes \mathbb{1}_{B_{r-1}}, \\
& C^{(r-1)} \geq 0, \\
& \mathrm{Tr}_{A_k}[C^{(k)}_{B_1^{k-1} A_1^k}] = C^{(k-1)}_{B_1^{k-2} A_1^{k-1}} \otimes \mathbb{1}_{B_{k-1}}, \\
& C^{(k-1)} \geq 0, \quad r-1 \geq k \geq 2, \\
& \mathrm{Tr}_{A_1}[C^{(1)}_{A_1}] = 1, \\
& C^{(1)} \geq 0.
\end{aligned} \tag{67}$$

Note that for $r = 1$, the definition above reduces to the definition in (65) for the generalized divergence between channels, because channels are 1-combs.

Intuitively, the generalized divergence between two strategy $r$-combs is given by taking the divergence of the output states obtained by combining the strategies with a compatible co-strategy. This method is natural, and it has been used in prior work [39] to define the max-relative entropy and the conditional min-entropy of quantum combs. This definition is also natural when considering the problems of discrimination and hypothesis testing of quantum combs, which we discuss in Sec. 3.5.

**Proposition 3.** *Let $r \geq 1$, and consider two strategy $r$-combs $\mathcal{E}^{[r]}$ and $\mathcal{M}^{[r]}$ with input systems $A_1^r$ and output systems $B_1^r$. When calculating the generalized divergence $\boldsymbol{D}_r\left(\mathcal{E}^{[r]} \middle\| \mathcal{M}^{[r]}\right)$, it suffices to optimize with respect to isometric co-strategies, i.e., co-strategies $\mathcal{D}^{[r]}_{\textit{(co-st)}} = (\sigma_{D_1 A_1}, \mathcal{D}^2_{D_1 B_1 \to D_2 A_2}, \ldots, \mathcal{D}^r_{D_{r-1} B_{r-1} \to D_r A_r})$ in which the starting state $\sigma_{D_1 A_1}$ is pure and the channels $\mathcal{D}^k$, $2 \leq k \leq r$, are isometric channels, with the dimensions of the systems $D_k$, $1 \leq k \leq r$, given by $d_{D_1} = d_{A_1}$ and $d_{D_k} = d_{D_{k-1}} d_{B_{k-1}} d_{A_k}$ for all $2 \leq k \leq r$.*

*Proof.* Consider an arbitrary co-strategy as depicted in Fig. 3(a), which consists of a state $\sigma_{D_1 A_1}$ and quantum channels $\mathcal{D}^k_{D_{k-1} B_{k-1} \to D_k A_k}$ for $2 \leq k \leq r$. Now, it holds that the state $\sigma_{D_1 A_1}$ has a purification $\psi_{R_1 D_1 A_1}$, so that $\sigma_{D_1 A_1} = \mathrm{Tr}_{R_1}[\psi_{R_1 D_1 A_1}]$. Furthermore, by the Stinespring dilation theorem (see, e.g., [40, Corollary 2.27]) there exist isometries $V^k_{D_{k-1} B_{k-1} \to R_k D_k A_k}$ such that

$$\mathcal{D}^k_{D_{k-1} B_{k-1} \to D_k A_k}(\cdot) = \mathrm{Tr}_{R_k}[V^k_{D_{k-1} B_{k-1} \to R_k D_k A_k}(\cdot)(V^k_{D_{k-1} B_{k-1} \to R_k D_k A_k})^\dagger], \tag{68}$$

The pure state $\psi_{R_1 D_1 A_1}$ and the isometries $V^k_{D_{k-1} B_{k-1} \to R_k D_k A_k}$ define another co-strategy $r$-comb $\widetilde{\mathcal{D}}^{[r]}_{\text{(co-st)}}$ with input systems $B_1^{r-1}$, output systems $A_1^r R_1^r D_r$ and memory systems $D_1^{r-1}$ such that

$$\begin{aligned}
&\boldsymbol{D}\left(\rho^{(\mathcal{E}^{[r]}, \mathcal{D}^{[r]}_{\text{(co-st)}})}_{D_r B_r} \middle\| \rho^{(\mathcal{M}^{[r]}, \mathcal{D}^{[r]}_{\text{(co-st)}})}_{D_r B_r}\right) \\
&\quad = \boldsymbol{D}\left(\mathrm{Tr}_{R_1^r}\left[\rho^{(\mathcal{E}^{[r]}, \widetilde{\mathcal{D}}^{[r]}_{\text{(co-st)}})}_{R_1^r D_r B_r}\right] \middle\| \mathrm{Tr}_{R_1^r}\left[\rho^{(\mathcal{M}^{[r]}, \widetilde{\mathcal{D}}^{[r]}_{\text{(co-st)}})}_{R_1^r D_r B_r}\right]\right)
\end{aligned} \tag{69}$$

$$\leq \boldsymbol{D}\left(\rho^{(\mathcal{E}^{[r]},\widetilde{\mathcal{D}}^{[r]}_{(\text{co-st})})}_{R_1^r D_r B_r}\middle\|\rho^{(\mathcal{M}^{[r]},\widetilde{\mathcal{D}}^{[r]}_{(\text{co-st})})}_{R_1^r D_r B_r}\right), \tag{70}$$

where the inequality is due to the data-processing inequality for $\boldsymbol{D}$. Now, because the co-strategy $\widetilde{\mathcal{D}}^{[r]}_{(\text{co-st})}$ consists of a pure state and isometric channels, its Choi representation is a rank-one operator of the form $|\Psi\rangle\langle\Psi|_{B_1^{r-1}A_1^r R_1^r D_r}$. By the Schmidt decomposition theorem (see, e.g., [40, Theorem 2.10]), the dimension of $R_1^r D_r$ need not exceed the dimension of $B_1^{r-1}A_1^r$. Let $D_r' \equiv R_1^r D_r$. Then, the vector $|\Psi\rangle_{B_1^{r-1}A_1^r R_1^r D_r} \equiv |\Psi\rangle_{B_1^{r-1}A_1^r D_r'}$ corresponds to the Choi representation of a pure co-strategy $(\psi'_{A_1 D_1'}, \mathcal{U}^1_{D_1' B_1 \to D_2' A_2}, \dots, \mathcal{U}^r_{D_{r-1}' B_{r-1} \to D_r' B_r})$, where the $\mathcal{U}^k$ are isometric channels such that $D_1' \cong A_1$ and $D_k' \cong D_{k-1}' B_{k-1} A_k$. Then, by (70), the generalized divergence between the states arising from this co-strategy is never less than the generalized divergence arising from the original co-strategy $\mathcal{D}^{[r]}_{(\text{co-st})}$. It therefore suffices to restrict the optimization to such pure co-strategies, and because the systems $D_k'$ in this co-strategy have the dimensions as specified in the statement of the proposition, the proof is complete. □

As mentioned above, quantum channels are strategy 1-combs, which means that, for $r = 1$, the definition in (66) coincides with the definition in (65). Also, because every quantum strategy comb can be used as an ordinary quantum channel via the co-strategy illustrated in Fig. 4, we immediately have that

$$\boldsymbol{D}_r\left(\mathcal{E}^{[r]}\middle\|\mathcal{M}^{[r]}\right) \geq \boldsymbol{D}_1\left(\mathcal{N}^{\mathcal{E}^{[r]}}\middle\|\mathcal{N}^{\mathcal{M}^{[r]}}\right), \tag{71}$$

where the divergence $\boldsymbol{D}_1$ is the generalized divergence for quantum channels defined in (65). Another simple consequence of definitions is the following fact.

**Proposition 4.** *Given any two strategy $r$-combs $\mathcal{E}^{[r]}$ and $\mathcal{M}^{[r]}$, with $r \geq 1$, it holds that the generalized divergence between the truncated combs $\mathcal{E}^{[r];k}$ and $\mathcal{M}^{[r];k}$ (recall (3)), as well as the combs $\mathcal{E}^{[1;k]}$ and $\mathcal{M}^{[1;k]}$ (recall (2)) satisfy*

$$\boldsymbol{D}_r\left(\mathcal{E}^{[r]}\middle\|\mathcal{M}^{[r]}\right) \leq \boldsymbol{D}_k\left(\mathcal{E}^{[1;k]}\middle\|\mathcal{M}^{[1;k]}\right), \tag{72}$$

$$\boldsymbol{D}_k\left(\mathcal{E}^{[r];k}\middle\|\mathcal{M}^{[r];k}\right) \leq \boldsymbol{D}_k\left(\mathcal{E}^{[1;k]}\middle\|\mathcal{M}^{[1;k]}\right) \tag{73}$$

*for all $1 \leq k \leq r$.*

*Proof.* For any co-strategy $r$-comb $\mathcal{D}^{[r]}_{(\text{co-st})} = (\sigma, \mathcal{D}^1, \dots, \mathcal{D}^r)$ in the optimization for $\boldsymbol{D}_r\left(\mathcal{E}^{[r]}\middle\|\mathcal{M}^{[r]}\right)$, we use (35) to see that

$$\begin{aligned}\rho^{(\mathcal{E}^{[r]},\mathcal{D}^{[r]}_{(\text{co-st})})}_{D_r B_r} &= (\mathcal{E}^r \circ \mathcal{D}^r \circ \cdots \circ \mathcal{E}^{k+1} \circ \mathcal{D}^{k+1} \circ \\ &\qquad \mathcal{E}^k \circ \mathcal{D}^k \circ \cdots \circ \mathcal{D}^2 \circ \mathcal{E}^1)(\sigma) \qquad (74)\\ &= (\mathcal{E}^r \circ \mathcal{D}^r \circ \cdots \circ \mathcal{E}^{k+1} \circ \mathcal{D}^{k+1})\left(\rho^{(\mathcal{E}^{[1;k]},\mathcal{D}^{[1;k]}_{(\text{co-st})})}_{D_k B_k E_k}\right). \qquad (75)\end{aligned}$$

An analogous expression holds for $\rho^{(\mathcal{M}^{[r]},\mathcal{D}^{[r]}_{(\text{co-st})})}_{D_r B_r}$. Then, by the data-processing inequality for $\boldsymbol{D}$,

$$\boldsymbol{D}\left(\rho^{(\mathcal{E}^{[r]},\mathcal{D}^{[r]}_{(\text{co-st})})}_{D_r B_r}\middle\|\rho^{(\mathcal{M}^{[r]},\mathcal{D}^{[r]}_{(\text{co-st})})}_{D_r B_r}\right)$$

$$\leq \boldsymbol{D}\left(\rho_{D_k B_k E_k}^{(\mathcal{E}^{[1;k]},\mathcal{D}^{[1;k]}_{(\text{co-st})})} \middle\| \rho_{D_k B_k E_k}^{(\mathcal{E}^{1;k]},\mathcal{D}^{[1;k]}_{(\text{co-st})})}\right) \tag{76}$$

$$\leq \sup_{\mathcal{F}^{[k]}_{(\text{co-st})}} \boldsymbol{D}\left(\rho_{D_k B_k E_k}^{(\mathcal{E}^{[1;k]},\mathcal{F}^{[k]}_{(\text{co-st})})} \middle\| \rho_{D_k B_k E_k}^{(\mathcal{E}^{1;k]},\mathcal{F}^{[k]}_{(\text{co-st})})}\right) \tag{77}$$

$$= \boldsymbol{D}_k\left(\mathcal{E}^{[1;k]} \middle\| \mathcal{M}^{[1;k]}\right). \tag{78}$$

Since the co-strategy $r$-comb $\mathcal{D}^{[r]}_{(\text{co-st})} = (\sigma, \mathcal{D}^1, \ldots, \mathcal{D}^r)$ was arbitrary, we obtain the inequality in (72). To see the inequality in (73), consider any co-strategy $k$-comb $\widetilde{\mathcal{D}}^{[k]}_{(\text{co-st})}$ in the optimization for $\boldsymbol{D}_k\left(\mathcal{E}^{[r];k} \middle\| \mathcal{M}^{[r];k}\right)$. By observing that

$$\gamma(\mathcal{E}^{[r];k}) = \operatorname{Tr}_{E_k}\left[\gamma(\mathcal{E}^{[1;k]})\right], \tag{79}$$

the data-processing inequality for $\boldsymbol{D}$ yields

$$\begin{aligned}
&\boldsymbol{D}\left(\rho_{D_k B_k}^{(\mathcal{E}^{[r];k},\widetilde{\mathcal{D}}^{[k]}_{(\text{co-st})})} \middle\| \rho_{D_k B_k}^{(\mathcal{M}^{[r];k},\widetilde{\mathcal{D}}^{[k]}_{(\text{co-st})})}\right) \\
&\quad = \boldsymbol{D}\left(\gamma(\mathcal{E}^{[r];k}) * \gamma(\widetilde{\mathcal{D}}^{[k]}_{(\text{co-st})}) \middle\| \gamma(\mathcal{M}^{[r];k}) * \gamma(\widetilde{\mathcal{D}}^{[k]}_{(\text{co-st})})\right) &(80)\\
&\quad = \boldsymbol{D}\left(\operatorname{Tr}_{E_k}\left[\gamma(\mathcal{E}^{[1;k]})\right] * \gamma(\widetilde{\mathcal{D}}^{[k]}_{(\text{co-st})}) \middle\| \right. \\
&\qquad\qquad \left. \operatorname{Tr}_{E_k}\left[\gamma(\mathcal{M}^{[1;k]})\right] * \gamma(\widetilde{\mathcal{D}}^{[k]}_{(\text{co-st})})\right) &(81)\\
&\quad \leq \boldsymbol{D}\left(\gamma(\mathcal{E}^{[1;k]}) * \gamma(\widetilde{\mathcal{D}}^{[k]}_{(\text{co-st})}) \middle\| \gamma(\mathcal{M}^{[1;k]}) * \gamma(\widetilde{\mathcal{D}}^{[k]}_{(\text{co-st})})\right). &(82)
\end{aligned}$$

Then, because $\widetilde{\mathcal{D}}^{[k]}_{(\text{co-st})}$ is an example of a co-strategy in the optimization for $\boldsymbol{D}_k\left(\mathcal{E}^{[1;k]} \middle\| \mathcal{M}^{[1;k]}\right)$, we obtain the inequality in (73), which completes the proof. □

Note that the defining data-processing property of a generalized divergence, shown in (64), is enough to conclude that all generalized divergences are isometrically invariant: for all isometries $V$ and for all states $\rho$ and $\sigma$,

$$\boldsymbol{D}(V\rho V^\dagger \| V\sigma V^\dagger) = \boldsymbol{D}(\rho\|\sigma). \tag{83}$$

Using the isometric invariance of generalized divergences leads to an alternate expression for the generalized divergence between $r$-combs.

**Proposition 5.** *For all $r \geq 1$, and for all pairs $\mathcal{E}^{[r]}$, $\mathcal{M}^{[r]}$ of strategy $r$-combs, the generalized* *divergence* $\boldsymbol{D}_r(\mathcal{E}^{[r]}\|\mathcal{M}^{[r]})$ *is the solution to the following optimization problem:*

$$\begin{aligned}
\textit{maximize} \quad & \boldsymbol{D}\left(\sqrt{P}\gamma(\mathcal{E}^{[r]})\sqrt{P}\middle\|\sqrt{P}\gamma(\mathcal{M}^{[r]})\sqrt{P}\right) \\
\textit{subject to} \quad & P_{A_1^r B_1^{r-1}} \geq 0, \\
& \operatorname{Tr}_{A_r}[P_{A_1^r B_1^{r-1}}] = C^{(r-1)}_{B_1^{r-2}A_1^{r-1}} \otimes \mathbb{1}_{B_{r-1}}, \\
& C^{(r-1)} \geq 0, \\
& \operatorname{Tr}_{A_k}[C^{(k)}_{B_1^{k-1}A_1^k}] = C^{(k-1)}_{B_1^{k-2}A_1^{k-1}} \otimes \mathbb{1}_{B_{k-1}}, \\
& C^{(k-1)} \geq 0, \quad r-1 \geq k \geq 2, \\
& \operatorname{Tr}_{A_1}[C^{(1)}_{A_1}] = 1, \\
& C^{(1)} \geq 0.
\end{aligned} \tag{84}$$

*Proof.* First, we use the fact that in the optimization with respect to co-strategies it suffices to take co-strategies consisting entirely of isometric channels such that the final output system satisfies $D_r \cong A_1^r B_1^{r-1}$ and the memory systems satisfy $D_1 \cong A_1$ and $D_k \cong D_{k-1}A_kB_{k-1}$ for all $2 \leq k \leq r-1$; see Proposition 3. Then, by Proposition 1, for any such co-strategy $\mathcal{D}^{[r]}_{(\text{co-st})}$, we have that

$$\gamma(\mathcal{E}^{[r]}) * \gamma(\mathcal{D}^{[r]}_{(\text{co-st})}) = V\gamma(\mathcal{E}^{[r]})V^\dagger, \tag{85}$$

$$\gamma(\mathcal{M}^{[r]}) * \gamma(\mathcal{D}^{[r]}_{(\text{co-st})}) = V\gamma(\mathcal{M}^{[r]})V^\dagger \tag{86}$$

for a linear operator $V_{A_1^r B_1^{r-1}}$ that satisfies (38). Consider now a polar decomposition of $V$ as $V = U\sqrt{P}$, where $U$ is unitary and $P$ is positive semi-definite. Then, $V^{\intercal}\overline{V} = \overline{P}$, where we have used the fact that $P$ is Hermitian, which means that

$$\overline{P} = \gamma((\mathcal{D}^1, \ldots, \operatorname{Tr}_{D_r} \circ \mathcal{D}^r)). \tag{87}$$

Finally, by isometric invariance of the generalized divergence $\boldsymbol{D}$, we obtain

$$\begin{aligned}
&\boldsymbol{D}(\gamma(\mathcal{E}^{[r]}) * \gamma(\mathcal{D}^{[r]}_{(\text{co-st})})\|\gamma(\mathcal{M}^{[r]}) * \gamma(\mathcal{D}_{(\text{co-st})})) && (88)\\
&\quad = \boldsymbol{D}(V\gamma(\mathcal{E}^{[r]})V^\dagger\|V\gamma(\mathcal{M}^{[r]})V^\dagger) && (89)\\
&\quad = \boldsymbol{D}\left(U\sqrt{P}\gamma(\mathcal{E}^{[r]}\sqrt{P}U^\dagger\middle\|U\sqrt{P}\gamma(\mathcal{M}^{[r]})\sqrt{P}U^\dagger\right) && (90)\\
&\quad = \boldsymbol{D}\left(\sqrt{P}\gamma(\mathcal{E}^{[r]})\sqrt{P}\middle\|\sqrt{P}\gamma(\mathcal{M}^{[r]})\sqrt{P}\right). && (91)
\end{aligned}$$

By (87), and because $P$ is Hermitian, the first constraint in (67) can be written as

$$\begin{aligned}
\operatorname{Tr}_{A_rD_r}[C^{(r)}_{B_1^{r-1}A_1^rD_r}] &= \operatorname{Tr}_{A_r}[\overline{P}_{A_1^rB_1^{r-1}}] && (92)\\
&= \operatorname{Tr}_{A_r}[P_{A_1^rB_1^{r-1}}] && (93)\\
&= C^{(r-1)}_{B_1^{r-2}A_1^{r-1}} \otimes \mathbb{1}_{B_{r-1}}. && (94)
\end{aligned}$$

The result then follows. □

Before moving on, we mention in passing that to every generalized divergence for channels, as defined in (65), there is an associated so-called *amortized channel divergence* [41] defined as

$$\boldsymbol{D}^{\mathcal{A}}(\mathcal{N}\|\mathcal{M}) \coloneqq \sup_{\rho_{RA},\sigma_{RA}} \left[\boldsymbol{D}(\mathcal{N}_{A\to B}(\rho_{RA})\|\mathcal{M}_{A\to B}(\sigma_{RA})) - \boldsymbol{D}(\rho_{RA}\|\sigma_{RA})\right], \tag{95}$$

where the optimization is with respect to states $\rho_{RA}, \sigma_{RA}$, with the dimension of $R$ unrestricted in general. Unlike in the generalized divergence for quantum channels, the optimization here cannot in general be restricted to pure states with $R \cong A$. If $\boldsymbol{D}$ satisfies $\boldsymbol{D}(\rho\|\rho) \leq 0$ for all states $\rho$, then

$$\boldsymbol{D}^{\mathcal{A}}(\mathcal{N}\|\mathcal{M}) \geq \boldsymbol{D}(\mathcal{N}\|\mathcal{M}). \tag{96}$$

The generalized divergence for $r$-combs can be extended to an amortized quantity as follows: for all strategy $r$-combs $\mathcal{E}^{[r]}$ and $\mathcal{M}^{[r]}$ with input systems $A_1^r$ and output systems $B_1^r$,

$$\boldsymbol{D}_r^{\mathcal{A}}\left(\mathcal{E}^{[r]}\middle\|\mathcal{M}^{[r]}\right) \coloneqq \sup_{\substack{\mathcal{D}_{\text{(co-st)}}^{[r]},\\ \mathcal{F}_{\text{(co-st)}}^{[r]}}} \left[\boldsymbol{D}\left(\rho_{D_rB_r}^{(\mathcal{E}^{[r]},\mathcal{D}_{\text{(co-st)}}^{[r]})}\middle\|\rho_{D_rB_r}^{(\mathcal{M}^{[r]},\mathcal{F}_{\text{(co-st)}}^{[r]})}\right) - \boldsymbol{D}(\sigma_{D_1A_1}\|\tau_{D_1A_1})\right], \tag{97}$$

where the optimization is with respect to co-strategy $r$-combs $\mathcal{D}_{\text{(co-st)}}^{[r]} = (\sigma_{D_1A_1}, \mathcal{D}^2, \ldots, \mathcal{D}^r)$ and $\mathcal{F}_{\text{(co-st)}}^{[r]} = (\tau_{D_1A_1}, \mathcal{F}^2, \ldots, \mathcal{F}^r)$ with input systems $B_1^{r-1}$, output systems $A_1^r D_r$, and memory systems $D_1^{r-1}$; see Fig. 3(a). Because the optimization in (95) is in general unbounded, the optimization in (97) is also unbounded in general, meaning that a result analogous to Proposition 3 does not hold in general.

### 3.2 Fidelity-based quantities

The fidelity between two quantum states $\rho$ and $\sigma$ is defined to be [42]

$$F(\rho,\sigma) \coloneqq \|\sqrt{\rho}\sqrt{\sigma}\|_1^2. \tag{98}$$

We also let

$$\sqrt{F}(\rho,\sigma) \coloneqq \sqrt{F(\rho,\sigma)} = \|\sqrt{\rho}\sqrt{\sigma}\|_1 \tag{99}$$

denote the "root-fidelity".

As with generalized divergences, the fidelity obeys a data-processing inequality: for all pairs $\rho, \sigma$ of states, and for all quantum channels $\mathcal{N}$,

$$F(\mathcal{N}(\rho), \mathcal{N}(\sigma)) \geq F(\rho,\sigma). \tag{100}$$

The fidelity between two quantum channels $\mathcal{N}$ and $\mathcal{M}$ is defined to be [43]

$$F(\mathcal{N},\mathcal{M}) \coloneqq \inf_{\rho_{RA}} F(\mathcal{N}_{A\to B}(\rho_{RA}), \mathcal{M}_{A\to B}(\rho_{RA})), \tag{101}$$

where the optimization is with respect to states $\rho_{RA}$, with the dimension of $R$ unrestricted in general, although it is straightforward to show that it suffices to let $R \cong A$. Note that this defintion is entirely analgous to the definition of the generalized divergence between two channels. Therefore, following the same line of reasoning as before, we arrive at the following definition of the fidelity between two strategy $r$-combs.

**Definition 6** (Fidelity between quantum combs [44]). *Let $r \geq 1$. Given two strategy $r$-combs $\mathcal{E}^{[r]}$ and $\mathcal{M}^{[r]}$ with input systems $A_1^r$ and output systems $B_1^r$, we define the fidelity and root-fidelity between them as follows:*

$$F_r(\mathcal{E}^{[r]}, \mathcal{M}^{[r]}) \coloneqq \inf_{\mathcal{D}^{[r]}_{(co\text{-}st)}} F\left(\rho_{D_rB_r}^{(\mathcal{E}^{[r]}, \mathcal{D}^{[r]}_{(co\text{-}st)})}, \rho_{D_rB_r}^{(\mathcal{M}^{[r]}, \mathcal{D}^{[r]}_{(co\text{-}st)})}\right), \tag{102}$$

$$\sqrt{F_r}(\mathcal{E}^{[r]}, \mathcal{M}^{[r]}) \coloneqq \inf_{\mathcal{D}^{[r]}_{(co\text{-}st)}} \sqrt{F}\left(\rho_{D_rB_r}^{(\mathcal{E}^{[r]}, \mathcal{D}^{[r]}_{(co\text{-}st)})}, \rho_{D_rB_r}^{(\mathcal{M}^{[r]}, \mathcal{D}^{[r]}_{(co\text{-}st)})}\right) \tag{103}$$

*where the optimization is with respect to co-strategy $r$-combs $\mathcal{D}^{[r]}_{(co\text{-}st)}$ with input systems $B_1^{r-1}$ and output systems $A_1^rD_r$; see Fig. 3(a). Using Theorem 1 along with (32), the fidelity is given by the solution to the following optimization problem:*

$$\begin{aligned}
\textit{minimize} \quad & F\left(\gamma(\mathcal{E}^{[r]}) * C^{(r)}, \gamma(\mathcal{M}^{[r]}) * C^{(r)}\right) \\
\textit{subject to} \quad & C^{(r)} \geq 0 \\
& \mathrm{Tr}_{A_rD_r}[C^{(r)}_{B_1^{r-1}A_1^rD_r}] = C^{(r-1)}_{B_1^{r-2}A_1^{r-1}} \otimes \mathbb{1}_{B_{r-1}}, \\
& C^{(r-1)} \geq 0, \\
& \mathrm{Tr}_{A_k}[C^{(k)}_{B_1^{k-1}A_1^k}] = C^{(k-1)}_{B_1^{k-2}A_1^{k-1}} \otimes \mathbb{1}_{B_{k-1}}, \\
& C^{(k-1)} \geq 0, \quad r-1 \geq k \geq 2, \\
& \mathrm{Tr}_{A_1}[C^{(1)}_{A_1}] = 1, \\
& C^{(1)} \geq 0,
\end{aligned} \tag{104}$$

*with an analogous optimization problem for the root-fidelity.*

**Remark 3.** *Due to the data-processing inequality for the fidelity between states, by arguments analogous to those in the proof of Proposition 3, it suffices in (102) and (103) to optimize over co-strategy $r$-combs $\mathcal{D}^{[r]}_{(co\text{-}st)} = (\sigma_{D_1A_1}, \mathcal{D}^2_{D_1B_1\to D_2A_2}, \dots, \mathcal{D}^r_{D_{r-1}B_{r-1}\to D_rA_r})$ in which the starting state $\sigma_{D_1A_1}$ is pure and the channels $\mathcal{D}^k$, $2 \leq k \leq r$, are isometric channels, with the dimensions of the systems $D_k$ given by $d_{D_1} = d_{A_1}$ and $d_{D_k} = d_{D_{k-1}}d_{B_{k-1}}d_{A_k}$ for all $2 \leq k \leq r$. Then, because of the fact that the root-fidelity can be calculated using an SDP, as shown in [45, Theorem 7.1.5] and [46], when we replace fidelity with root-fidelity in (104), we obtain an SDP.*

*Also, because the fidelity is isometrically invariant, we get that the fidelity between quantum combs* *can be calculated using as the solution to an optimization problem analogous to the one in (84):*

$$\begin{aligned}
\textit{minimize} \quad & F\left(\sqrt{P}\gamma(\mathcal{E}^{[r]})\sqrt{P}, \sqrt{P}\gamma(\mathcal{M}^{[r]})\sqrt{P}\right) \\
\textit{subject to} \quad & P_{A_1^r B_1^{r-1}} \geq 0, \\
& \mathrm{Tr}_{A_r}[P_{A_1^r B_1^{r-1}}] = C^{(r-1)}_{B_1^{r-2} A_1^{r-1}} \otimes \mathbb{1}_{B_{r-1}}, \\
& C^{(r-1)} \geq 0, \\
& \mathrm{Tr}_{A_k}[C^{(k)}_{B_1^{k-1} A_1^k}] = C^{(k-1)}_{B_1^{k-2} A_1^{k-1}} \otimes \mathbb{1}_{B_{k-1}}, \\
& C^{(k-1)} \geq 0, \quad r-1 \geq k \geq 2, \\
& \mathrm{Tr}_{A_1}[C^{(1)}_{A_1}] = 1, \\
& C^{(1)} \geq 0.
\end{aligned} \tag{105}$$

### 3.3 Schatten $\alpha$-norms

The Schatten $\alpha$-norm of a linear operator $Y$, for $1 \leq \alpha \leq \infty$, is defined as

$$\|Y\|_\alpha := \left(\mathrm{Tr}\left[(Y^\dagger Y)^{\frac{\alpha}{2}}\right]\right)^{\frac{1}{\alpha}}, \quad 1 \leq \alpha < \infty, \tag{106}$$

$$\|Y\|_\infty := \lim_{\alpha \to \infty} \|Y\|_\alpha. \tag{107}$$

In order to generalize the notion of Schatten $\alpha$-norm to $r$-combs, we use the same reasoning that we used to define the generalized divergence and the fidelity between $r$-combs. Specifically, we take a given strategy $r$-comb $\mathcal{E}^{[r]}$, apply a co-strategy $r$-comb $\mathcal{D}^{[r]}_{\text{(co-st)}}$ to it as in Fig. 3(a), then take the Schatten-norm of the resulting quantum state.

**Definition 7** (Schatten norm of a quantum comb)**.** *Given any operator $X \in \mathrm{Lin}(A_1^r B_1^r)$, for $1 \leq \alpha \leq \infty$ we define its strategy Schatten $\alpha$-norm as*

$$\|X\|_{r,\alpha} := \sup_{\mathcal{D}^{[r]}_{\text{(co-st)}}} \|X * \gamma(\mathcal{D}^{[r]}_{\text{(co-st)}})\|_\alpha. \tag{108}$$

*where the optimization is with respect to co-strategy $r$-combs $\mathcal{D}^{[r]}_{\textit{(co-st)}} = (\sigma_{D_1 A_1}, \mathcal{D}^2_{D_1 B_1 \to D_2 A_2}, \ldots, \mathcal{D}^r_{D_{r-1} B_{r-1} \to D_r A_r})$ (see Fig. 3(a)) in which the starting state $\sigma_{D_1 A_1}$ is pure and the channels $\mathcal{D}^k$, $2 \leq k \leq r$, are isometric channels, with the dimensions of the systems $D_k$ given by $d_{D_1} = d_{A_1}$ and $d_{D_k} = d_{D_{k-1}} d_{B_{k-1}} d_{A_k}$ for all $2 \leq k \leq r$. Then, given any strategy $r$-comb $\mathcal{E}^{[r]}$ consisting of arbitrary linear maps with input systems $A_1^r$ and output systems $B_1^r$, for $1 \leq \alpha \leq \infty$ we define its strategy Schatten $\alpha$-norm as*

$$\|\mathcal{E}^{[r]}\|_{r,\alpha} := \|\gamma(\mathcal{E}^{[r]})\|_{r,\alpha}. \tag{109}$$

*By Theorem 1, $\|X\|_{r,\alpha}$ is the solution to the following optimization problem:*

$$\begin{aligned}
\textit{maximize} \quad & \|X * C^{(r)}\|_\alpha \\
\textit{subject to} \quad & C^{(r)} \geq 0, \\
& \mathrm{Tr}_{A_r D_r}[C^{(r)}_{B_1^{r-1} A_1^r D_r}] = C^{(r-1)}_{B_1^{r-2} A_1^{r-1}} \otimes \mathbb{1}_{B_{r-1}}, \\
& C^{(r-1)} \geq 0, \\
& \mathrm{Tr}_{A_k}[C^{(k)}_{B_1^{k-1} A_1^k}] = C^{(k-1)}_{B_1^{k-2} A_1^{k-1}} \otimes \mathbb{1}_{B_{k-1}}, \\
& C^{(k-1)} \geq 0, \quad r-1 \geq k \geq 2, \\
& \mathrm{Tr}_{A_1}[C^{(1)}_{A_1}] = 1, \\
& C^{(1)} \geq 0,
\end{aligned} \tag{110}$$

**Remark 4.** *Note that in (108), $X$ can be any linear operator, not just a Hermitian or a positive semi-definite operator. In particular, the definition of the Schatten $\alpha$-norm of an $r$-comb applies to quantum combs consisting of general linear maps, not just quantum channels or completely positive maps. In this more general setting, however, one should in principle define the norm using an optimization over co-strategies consisting of arbitrary linear maps, not just quantum channels, but by doing so we lose the all of the constraints in (110). We thus leave the optimization in (108) restricted to co-strategies consisting entirely of quantum channels.*

*In this context, let us also mention that the definition in (108) in the case $\alpha = 1$ has been defined already in [2, 29, 47] and denoted by $\|\cdot\|_{\diamond r}$ in [29]. The definitions therein were given based on the operational task of state discrimination, and hold only for Hermitian operators. If the operator $X$ in (108) is Hermitian, then $\|X\|_{1,\alpha} = \|X\|_{\diamond r}$.*

Because the Schatten $\alpha$-norms are isometrically invariant for all $1 \leq \alpha \leq \infty$, we obtain a result analogous to Proposition 5.

**Proposition 8.** *For all Hermitian operators $X \in \mathrm{Lin}(A_1^r B_1^r)$, the strategy Schatten $\alpha$-norm $\|X\|_{r,\alpha}$ can be calculated as the solution of the following optimization problem:*

$$\begin{aligned}
\textit{maxmize} \quad & \|\sqrt{P} X \sqrt{P}\|_\alpha \\
\textit{subject to} \quad & P_{A_1^r B_1^{r-1}} \geq 0, \\
& \mathrm{Tr}_{A_r}[P_{A_1^r B_1^{r-1}}] = C^{(r-1)}_{B_1^{r-2} A_1^{r-1}} \otimes \mathbb{1}_{B_{r-1}}, \\
& C^{(r-1)} \geq 0, \\
& \mathrm{Tr}_{A_k}[C^{(k)}_{B_1^{k-1} A_1^k}] = C^{(k-1)}_{B_1^{k-2} A_1^{k-1}} \otimes \mathbb{1}_{B_{k-1}}, \\
& C^{(k-1)} \geq 0, \quad r-1 \geq k \geq 2, \\
& \mathrm{Tr}_{A_1}[C^{(1)}_{A_1}] = 1, \\
& C^{(1)} \geq 0,
\end{aligned} \tag{111}$$

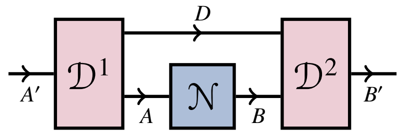



FIGURE 6: The most general transformation of the quantum channel $\mathcal{N}_{A\to B}$ to a quantum channel $A' \to B'$ via a "superchannel", which is the strategy 2-comb $(\mathcal{D}^1, \mathcal{D}^2)$.

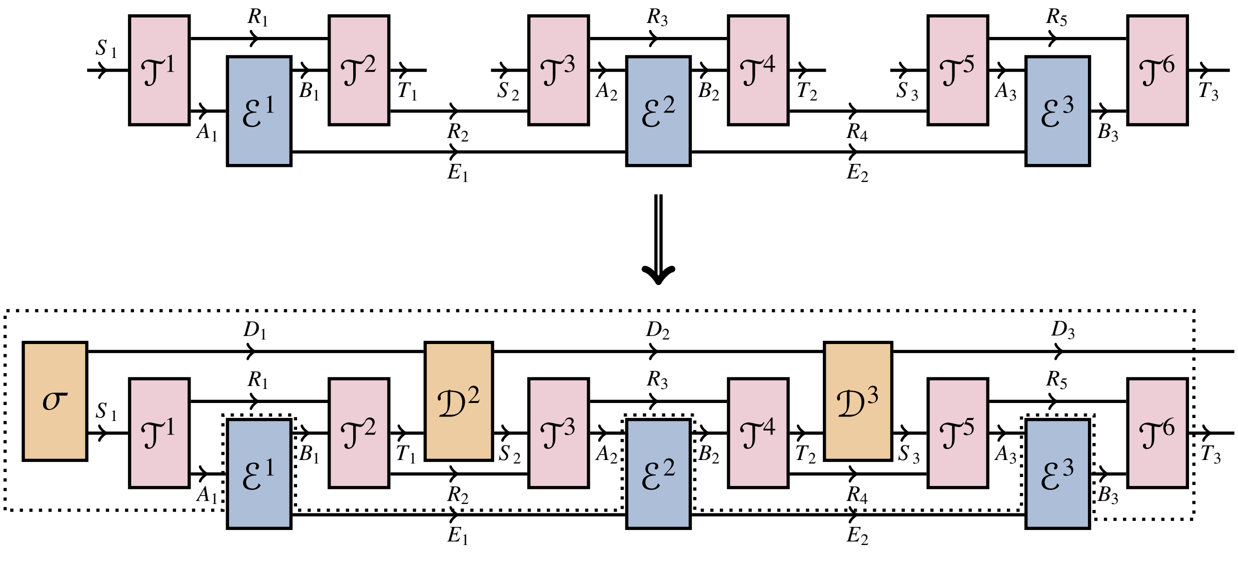


FIGURE 7: (Top) Transformation of the strategy 3-comb $\mathcal{E}^{[3]}$ into another strategy 3-comb via combination with the strategy 6-comb $\mathcal{T}^{[6]}$, which consists of pre- and post-processing of each of the channels $\mathcal{E}^j$ in the comb $\mathcal{E}^{[3]}$. (Bottom) Any co-strategy $\mathcal{D}^{[3]}_{\text{(co-st)}}$ for the transformed comb, when combined with $\mathcal{T}^{[6]}$, is a co-strategy for the original 3-comb $\mathcal{E}^{[3]}$.

*Proof.* The proof is analogous to the proof of Proposition 5, except that we make use of the generalization of Proposition 1 explained in Remark 2, specifically, (53) and (53). □

## 3.4 Comb transformations and data processing

The defining property of any generalized divergence $\boldsymbol{D}$ for states is the data-processing inequality. Having defined the generalized divergence $\boldsymbol{D}_r$ for $r$-combs, it is important to understand what analogues of the data-processing inequality hold for $\boldsymbol{D}_r$.

Let us first note that, given a quantum channel, it is known [48] that the unique way of transforming it into another channel is via a "superchannel", as depicted in Fig. 6, which in this context is simply a strategy 2-comb that consists of a pre- and post-processing of the given channel with two other channels. However, the transformation from an $r$-comb to another $r$-comb (with $r > 1$) can be much more varied in general. We now discuss two possible ways of transforming quantum combs and show how these transformations lead to a data-processing inequality for the generalized divergence $\boldsymbol{D}_r$.

One method for transforming an $r$-comb to another $r$-comb is directly analgous to the transformation of quantum channel shown in Fig. 6. As shown in Fig. 7 for $r = 3$, we can transform a given $r$-comb $\mathcal{E}^{[r]}$ into another $r$-comb by combining $\mathcal{E}^{[r]}$ with a comb $\mathcal{T}^{[2r]}$ that consist of a pre- and post-processing of each channel in $\mathcal{E}^{[r]}$. For this type of transformation of a quantum comb, we obtain the following data-processing inequality.

**Theorem 5** (Data-processing inequalities I)**.** *Given any two strategy $r$-combs $\mathcal{E}^{[r]}$ and $\mathcal{M}^{[r]}$, consider the $r$-combs $\mathcal{T}^{[2r]} \circ \mathcal{E}^{[r]}$ and $\mathcal{T}^{[2r]} \circ \mathcal{M}^{[r]}$ resulting from the combination of each comb with an arbitrary $2r$-comb $\mathcal{T}^{[2r]}$ of the form shown in Fig. 7, in which each channel of the combs $\mathcal{E}^{[r]}$ and $\mathcal{M}^{[r]}$ undergoes a pre- and post-processing. Then,*

$$\boldsymbol{D}_r\left(\mathcal{T}^{[2r]} \circ \mathcal{E}^{[r]} \middle\| \mathcal{T}^{[2r]} \circ \mathcal{M}^{[r]}\right) \leq \boldsymbol{D}_r\left(\mathcal{E}^{[r]} \middle\| \mathcal{M}^{[r]}\right), \tag{112}$$

$$F_r\left(\mathcal{T}^{[2r]} \circ \mathcal{E}^{[r]}, \mathcal{T}^{[2r]} \circ \mathcal{M}^{[r]}\right) \geq F_r\left(\mathcal{E}^{[r]}, \mathcal{M}^{[r]}\right). \tag{113}$$

*Proof.* We base our reasoning on the case $r = 3$ by referring to Fig. 7, with the understanding that this reasoning extends to arbitrary $r$. Consider any co-strategy $r$-comb $\mathcal{D}^{[r]}_{(\text{co-st})}$ for the transformed $r$-combs $\mathcal{T}^{[2r]} \circ \mathcal{E}^{[r]}$ and $\mathcal{T}^{[2r]} \circ \mathcal{M}^{[r]}$. From the bottom part of Fig. 7, it is clear that $\widetilde{\mathcal{D}}^{[r]}_{(\text{co-st})} \coloneqq \mathcal{D}^{[r]}_{(\text{co-st})} \circ \mathcal{T}^{[2r]}$ is a valid co-strategy for $\mathcal{E}^{[r]}$ and $\mathcal{M}^{[r]}$. Therefore,

$$\begin{aligned}
&\boldsymbol{D}\left(\mathcal{D}^{[r]}_{(\text{co-st})} \circ \mathcal{T}^{[2r]} \circ \mathcal{E}^{[r]} \middle\| \mathcal{D}^{[r]}_{(\text{co-st})} \circ \mathcal{T}^{[2r]} \circ \mathcal{M}^{[r]}\right) \\
&\quad = \boldsymbol{D}\left(\widetilde{\mathcal{D}}^{[r]}_{(\text{co-st})} \circ \mathcal{E}^{[r]} \middle\| \widetilde{\mathcal{D}}^{[r]}_{(\text{co-st})} \circ \mathcal{M}^{[r]}\right) && (114) \\
&\quad \leq \sup_{\widetilde{\mathcal{D}}^{[r]}_{(\text{co-st})}} \boldsymbol{D}\left(\widetilde{\mathcal{D}}^{[r]}_{(\text{co-st})} \circ \mathcal{E}^{[r]} \middle\| \widetilde{\mathcal{D}}^{[r]}_{(\text{co-st})} \circ \mathcal{M}^{[r]}\right) && (115) \\
&\quad = \boldsymbol{D}_r\left(\mathcal{E}^{[r]} \middle\| \mathcal{M}^{[r]}\right). && (116)
\end{aligned}$$

Because the co-strategy $\mathcal{D}^{[r]}_{(\text{co-st})}$ was arbitrary, we obtain

$$\begin{aligned}
&\boldsymbol{D}_r\left(\mathcal{T}^{[2r]} \circ \mathcal{E}^{[r]} \middle\| \mathcal{T}^{[2r]} \circ \mathcal{M}^{[r]}\right) \\
&\quad = \sup_{\mathcal{D}^{[r]}_{(\text{co-st})}} \boldsymbol{D}\left(\mathcal{D}^{[r]}_{(\text{co-st})} \circ \mathcal{T}^{[2r]} \circ \mathcal{E}^{[r]} \middle\| \mathcal{D}^{[r]}_{(\text{co-st})} \circ \mathcal{T}^{[2r]} \circ \mathcal{M}^{[r]}\right) && (117) \\
&\quad \leq \boldsymbol{D}_r\left(\mathcal{E}^{[r]} \middle\| \mathcal{M}^{[r]}\right), && (118)
\end{aligned}$$

as required. Analogous reasoning applies to the fidelity. □

**Remark 6.** *By applying similar reasoning as in the proof above, we can also prove the following data-processing inequality for the Schatten $\alpha$-norms for all $\alpha \in [1, \infty]$:*

$$\|\gamma(\mathcal{T}^{[2r]}) * X\|_{r,\alpha} \leq \|X\|_{r,\alpha}, \tag{119}$$

*for all linear operators $X \in \mathrm{Lin}(A_1^r B_1^r)$, where $\mathcal{T}^{[2r]}$ is the $(2r)$-comb shown in Fig. 7.*

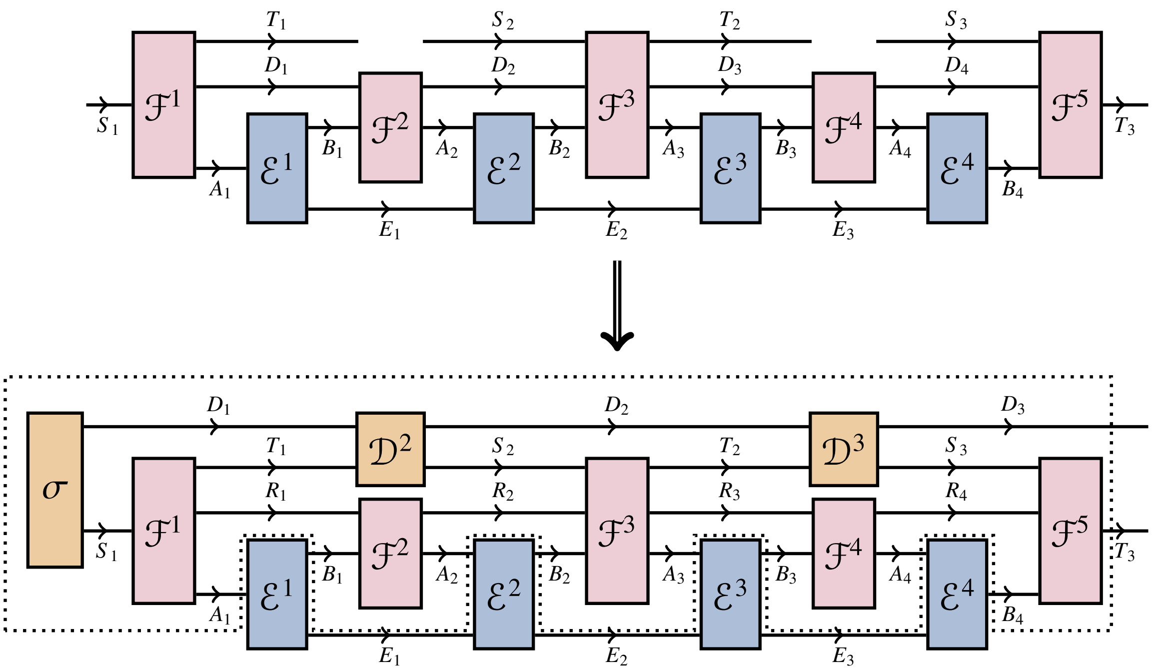



FIGURE 8: (Top) Transformation of the strategy 4-comb $\mathcal{E}^{[4]}$ into a strategy 3-comb via combination with the strategy 5-comb $\mathcal{F}^{[5]}$. (Bottom) Any co-strategy $\mathcal{D}^{[3]}_{\text{(co-st)}}$ for the transformed comb, when combined with $\mathcal{F}^{[5]}$, is a co-strategy for the original 4-comb $\mathcal{E}^{[4]}$.

The transformation shown in the top part of Fig. 7 takes an $r$-comb and transforms it into another $r$-comb. It is also possible in general to take an $r$-comb and transform it into an $r'$-comb with $r'$ different from $r$. One example of such a transformation is shown in Fig. 8, in which a 4-comb is transformed into a 3-comb. This type of transformation can be generalized in order to transform any $r$-comb into an $(r-1)$-comb. We then obtain the following data-processing inequality.

**Theorem 7** (Data-processing inequalities II)**.** *Given any two strategy $r$-combs $\mathcal{E}^{[r]}$ and $\mathcal{M}^{[r]}$, consider the $(r-1)$-combs $\mathcal{F}^{[r+1]} \circ \mathcal{E}^{[r]}$ and $\mathcal{F}^{[r+1]} \circ \mathcal{M}^{[r]}$ resulting from the combination of each comb with an arbitrary $(r+1)$-comb $\mathcal{F}^{[r+1]}$ of the form shown in Fig. 8. Then,*

$$\boldsymbol{D}_{r-1}\left(\mathcal{F}^{[r+1]} \circ \mathcal{E}^{[r]} \middle\| \mathcal{F}^{[r+1]} \circ \mathcal{M}^{[r]}\right) \leq \boldsymbol{D}_r\left(\mathcal{E}^{[r]} \middle\| \mathcal{M}^{[r]}\right), \tag{120}$$

$$F_{r-1}\left(\mathcal{F}^{[r+1]} \circ \mathcal{E}^{[r]}, \mathcal{F}^{[r+1]} \circ \mathcal{M}^{[r]}\right) \geq F_r\left(\mathcal{E}^{[r]}, \mathcal{M}^{[r]}\right). \tag{121}$$

*Proof.* We base our reasoning on the case $r = 4$ by referring to Fig. 8, with the understanding that this reasoning extends to arbitrary $r$. Consider any co-strategy $(r-1)$-comb $\mathcal{D}^{[r-1]}_{\text{(co-st)}}$ for the transformed $r$-combs $\mathcal{T}^{[r+1]} \circ \mathcal{E}^{[r]}$ and $\mathcal{T}^{[r+1]} \circ \mathcal{M}^{[r]}$. From the bottom part of Fig. 8, it is clear that $\widetilde{\mathcal{D}}^{[r]}_{\text{(co-st)}} := \mathcal{D}^{[r-1]}_{\text{(co-st)}} \circ \mathcal{T}^{[r+1]}$ is a valid co-strategy for $\mathcal{E}^{[r]}$ and $\mathcal{M}^{[r]}$. Therefore,

$$\begin{aligned}&\boldsymbol{D}\left(\mathcal{D}^{[r-1]}_{\text{(co-st)}} \circ \mathcal{F}^{[r+1]} \circ \mathcal{E}^{[r]} \middle\| \mathcal{D}^{[r-1]}_{\text{(co-st)}} \circ \mathcal{F}^{[r+1]} \circ \mathcal{M}^{[r]}\right) \\ &\quad = \boldsymbol{D}\left(\widetilde{\mathcal{D}}^{[r]}_{\text{(co-st)}} \circ \mathcal{E}^{[r]} \middle\| \widetilde{\mathcal{D}}^{[r]}_{\text{(co-st)}} \circ \mathcal{M}^{[r]}\right)\end{aligned} \tag{122}$$

$$\leq \sup_{\widetilde{\mathcal{D}}^{[r]}_{\text{(co-st)}}} \boldsymbol{D}\left(\widetilde{\mathcal{D}}^{[r]}_{\text{(co-st)}} \circ \mathcal{E}^{[r]} \middle\| \widetilde{\mathcal{D}}^{[r]}_{\text{(co-st)}} \circ \mathcal{M}^{[r]}\right) \tag{123}$$

$$= \boldsymbol{D}_r\left(\mathcal{E}^{[r]} \middle\| \mathcal{M}^{[r]}\right). \tag{124}$$

Because the co-strategy $\mathcal{D}^{[r-1]}_{\text{(co-st)}}$ was arbitrary, we obtain

$$\boldsymbol{D}_{r-1}\left(\mathcal{F}^{[r+1]} \circ \mathcal{E}^{[r]} \middle\| \mathcal{F}^{[r+1]} \circ \mathcal{M}^{[r]}\right)$$

$$= \sup_{\mathcal{D}^{[r-1]}_{\text{(co-st)}}} \boldsymbol{D}\left(\mathcal{D}^{[r-1]}_{\text{(co-st)}} \circ \mathcal{F}^{[r+1]} \circ \mathcal{E}^{[r]} \middle\| \mathcal{D}^{[r-1]}_{\text{(co-st)}} \circ \mathcal{F}^{[r+1]} \circ \mathcal{M}^{[r]}\right) \tag{125}$$

$$\leq \boldsymbol{D}_r\left(\mathcal{E}^{[r]} \middle\| \mathcal{M}^{[r]}\right), \tag{126}$$

as required. Analogous reasoning applies to the fidelity. □

**Remark 8.** *By applying similar reasoning as in the proof above, we can also prove the following data-processing inequality for the Schatten $\alpha$-norms for all $\alpha \in [1, \infty]$:*

$$\|\gamma(\mathcal{F}^{[r+1]}) * X\|_{r-1,\alpha} \leq \|X\|_{r,\alpha}, \tag{127}$$

*for all linear operators $X \in \text{Lin}(A_1^r B_1^r)$, where $\mathcal{F}^{[r+1]}$ is the $(r+1)$-comb shown in Fig. 8.*

Another type of comb transformation, which we discussed in Sec. 2, is taking the tensor product of combs.

**Theorem 9.** *Given any two strategy $r$-combs $\mathcal{E}^{[r]}$ and $\mathcal{M}^{[r]}$, for all $r$-combs $\mathcal{G}^{[r]}$ it holds that*

$$\boldsymbol{D}_r\left(\mathcal{G}^{[r]} \otimes \mathcal{E}^{[r]} \middle\| \mathcal{G}^{[r]} \otimes \mathcal{M}^{[r]}\right) = \boldsymbol{D}_r\left(\mathcal{E}^{[r]} \middle\| \mathcal{M}^{[r]}\right), \tag{128}$$

$$F_r\left(\mathcal{G}^{[r]} \otimes \mathcal{E}^{[r]}, \mathcal{G}^{[r]} \otimes \mathcal{M}^{[r]}\right) = F_r\left(\mathcal{E}^{[r]}, \mathcal{M}^{[r]}\right). \tag{129}$$

*Proof.* Suppose that $\mathcal{E}^{[r]}$ and $\mathcal{M}^{[r]}$ have input systems $A_1^r$, output systems $B_1^r$, and memory systems $E_1^{r-1}$, and suppose that $\mathcal{G}^{[r]}$ has input systems $C_1^r$, output systems $D_1^r$, and memory systems $M_1^{r-1}$. Then,

$$\boldsymbol{D}_r\left(\mathcal{G}^{[[r]} \otimes \mathcal{E}^{[r]} \middle\| \mathcal{G}^{[r]} \otimes \mathcal{M}^{[r]}\right)$$

$$= \sup_{\mathcal{D}^{[r]}_{\text{(co-st)}}} \boldsymbol{D}\left(\mathcal{D}^{[r]}_{\text{(co-st)}} \circ \left(\mathcal{G}^{[r]} \otimes \mathcal{E}^{[r]}\right) \middle\| \mathcal{D}^{[r]}_{\text{(co-st)}} \circ \left(\mathcal{G}^{[r]} \otimes \mathcal{M}^{[r]}\right)\right), \tag{130}$$

where the optimization is with respect to co-strategy $r$-combs $\mathcal{D}^{[r]}_{\text{(co-st)}}$ with input systems $B_1^{r-1} D_1^{r-1}$ and output systems $A_1^r C_1^r D_r'$. A particular choice for a costrategy is a tensor product co-strategy $\mathcal{D}^{[r]}_{1,\text{(co-st)}} \otimes \mathcal{D}^{[r]}_{2,\text{(co-st)}}$, where $\mathcal{D}^{[r]}_{1,\text{(co-st)}}$ has input systems $B_1^{r-1}$ and output systems $A_1^r K_r$ and $\mathcal{D}^{[r]}_{2,\text{(co-st)}}$ has input systems $D_1^{r-1}$ and output systems $C_1^r L_r$. Then,

$$\left(\mathcal{D}^{[r]}_{1,\text{(co-st)}} \otimes \mathcal{D}^{[r]}_{2,\text{(co-st)}}\right) \circ \left(\mathcal{G}^{[r]} \otimes \mathcal{E}^{[r]}\right) = \left(\mathcal{D}^{[r]}_{1,\text{(co-st)}} \circ \mathcal{G}^{[r]}\right) \otimes \left(\mathcal{D}^{[r]}_{2,\text{(co-st)}} \circ \mathcal{E}^{[r]}\right), \tag{131}$$

with an analogous expression involving $\mathcal{M}^{[r]}$. Then, because the generalized divergence $\boldsymbol{D}$ satisfies $\boldsymbol{D}(\tau\otimes\rho\|\tau\otimes\sigma)=\boldsymbol{D}(\rho\|\sigma)$ for all states $\rho,\sigma,\tau$ by virtue of the data-processing inequality, we obtain

$$\boldsymbol{D}\left(\left(\mathcal{D}^{[r]}_{1,(\text{co-st})}\otimes\mathcal{D}^{[r]}_{2,(\text{co-st})}\right)\circ\left(\mathcal{G}^{[r]}\otimes\mathcal{E}^{[r]}\right)\Big\|\left(\mathcal{D}^{[r]}_{1,(\text{co-st})}\otimes\mathcal{D}^{[r]}_{2,(\text{co-st})}\right)\circ\left(\mathcal{G}^{[r]}\otimes\mathcal{M}^{[r]}\right)\right) \tag{132}$$

$$=\boldsymbol{D}\left(\left(\mathcal{D}^{[r]}_{1,(\text{co-st})}\circ\mathcal{G}^{[r]}\right)\otimes\left(\mathcal{D}^{[r]}_{2,(\text{co-st})}\circ\mathcal{E}^{[r]}\right)\Big\|\left(\mathcal{D}^{[r]}_{1,(\text{co-st})}\circ\mathcal{G}^{[r]}\right)\otimes\left(\mathcal{D}^{[r]}_{2,(\text{co-st})}\circ\mathcal{M}^{[r]}\right)\right) \tag{133}$$

$$=\boldsymbol{D}\left(\mathcal{D}^{[r]}_{2,(\text{co-st})}\circ\mathcal{E}^{[r]}\Big\|\mathcal{D}^{[r]}_{2,(\text{co-st})}\circ\mathcal{M}^{[r]}\right) \tag{134}$$

Because $\mathcal{D}^{[r]}_{2,(\text{co-st})}$ is arbitrary, we can optimize with respect to $\mathcal{D}^{[r]}_{2,(\text{co-st})}$ to obtain

$$\boldsymbol{D}_r\left(\mathcal{G}^{[r]}\otimes\mathcal{E}^{[r]}\Big\|\mathcal{G}^{[r]}\otimes\mathcal{M}^{[r]}\right)\geq\boldsymbol{D}_r\left(\mathcal{E}^{[r]}\Big\|\mathcal{M}^{[r]}\right), \tag{135}$$

where the inequality is due to the fact that we restricted the ourselves to tensor-product co-strategies.

Now, for the reverse inequality, consider for any choice of co-strategy $\mathcal{D}^{[r]}_{(\text{co-st})}$ in the optimization in (130) the combination

$$\mathcal{D}^{[r]}_{(\text{co-st})}\circ\mathcal{G}^{[r]}. \tag{136}$$

Observe that this combination is a co-strategy $r$-comb with input systems $B_1^{r-1}$ and output systems $A_1^rD_r'$. It is therefore a valid co-strategy $\widetilde{\mathcal{D}}^{[r]}_{(\text{co-st})}$ in the optimization for $\boldsymbol{D}_r\left(\mathcal{E}^{[r]}\|\mathcal{M}^{[r]}\right)$. Therefore,

$$\begin{aligned}&\boldsymbol{D}\left(\mathcal{D}^{[r]}_{(\text{co-st})}\circ\left(\mathcal{G}^{[r]}\otimes\mathcal{E}^{[r]}\right)\Big\|\mathcal{D}^{[r]}_{(\text{co-st})}\circ\left(\mathcal{G}^{[r]}\otimes\mathcal{M}^{[r]}\right)\right)\\&\quad=\boldsymbol{D}\left(\widetilde{\mathcal{D}}^{[r]}_{(\text{co-st})}\circ\mathcal{E}^{[r]}\Big\|\widetilde{\mathcal{D}}^{[r]}_{(\text{co-st})}\circ\mathcal{M}^{[r]}\right)\end{aligned} \tag{137}$$

$$\leq\sup_{\widetilde{\mathcal{D}}^{[r]}_{(\text{co-st})}}\boldsymbol{D}\left(\widetilde{\mathcal{D}}^{[r]}_{(\text{co-st})}\circ\mathcal{E}^{[r]}\Big\|\widetilde{\mathcal{D}}^{[r]}_{(\text{co-st})}\circ\mathcal{M}^{[r]}\right) \tag{138}$$

$$=\boldsymbol{D}_r\left(\mathcal{E}^{[r]}\Big\|\mathcal{M}^{[r]}\right). \tag{139}$$

Because the costrategy $\mathcal{D}^{[r]}_{(\text{co-st})}$ was arbitrary, we obtain

$$\boldsymbol{D}_r\left(\mathcal{G}^{[r]}\otimes\mathcal{E}^{[r]}\Big\|\mathcal{G}^{[r]}\otimes\mathcal{M}^{[r]}\right)\leq\boldsymbol{D}_r\left(\mathcal{E}^{[r]}\Big\|\mathcal{M}^{[r]}\right), \tag{140}$$

which completes the proof. □

**Remark 10.** *The reasoning in the proof of Theorem 9 can be applied to conclude the more general statement that*

$$\boldsymbol{D}_{r''}\left(\mathcal{G}^{[r']}\otimes\mathcal{E}^{[r]}\Big\|\mathcal{G}^{[r']}\otimes\mathcal{M}^{[r]}\right)=\boldsymbol{D}_r\left(\mathcal{E}^{[r]}\Big\|\mathcal{M}^{[r]}\right), \tag{141}$$

*for all $r'$-combs $\mathcal{G}^{[r']}$, where $r''=\max\{r,r'\}$.*

### 3.5 Application to hypothesis testing

We now discuss hypothesis testing in the context of quantum combs/quantum causal networks. The task of ordinary (asymmetric) hypothesis testing (see, e.g., [49]) is to distinguish between two "hypotheses", respresented as quantum states $\rho$ and $\sigma$, via a binary measurement given by a POVM $\{\Lambda^0, \Lambda^1\}$. The goal is to minimize the so-called type-II error $\mathrm{Tr}[\Lambda^0\sigma]$ while maintaining the type-I error $\mathrm{Tr}[\Lambda^1\rho]$ below a specified threshold $\varepsilon \in [0,1]$, i.e., $\mathrm{Tr}[\Lambda^1\rho] \leq \varepsilon$. The minimum type-II error, obtained by optimizing over all binary POVMs, is given by the hypothesis testing relative entropy [50–52]:

$$D_H^\varepsilon(\rho\|\sigma) \coloneqq -\log_2 \inf_{\Lambda^0,\Lambda^1\geq 0} \left\{\mathrm{Tr}[\Lambda^0\sigma] : \mathrm{Tr}[\Lambda^1\rho] \leq \varepsilon, \Lambda^0 + \Lambda^1 = \mathbb{1}\right\}. \tag{142}$$

Observe that this optimization problem (without the logarithm) is an SDP.

The hypothesis testing relative entropy is a generalized divergence, meaning that we can extend its definition to quantum combs using Definition 2. Doing so, we obtain the following.

**Proposition 9.** *Let $r \geq 1$. Given any two $r$-combs $\mathcal{E}^{[r]}$ and $\mathcal{M}^{[r]}$, for all $\varepsilon \in [0,1]$ their hypothesis testing relative entropy $D_{H,r}^\varepsilon\left(\mathcal{E}^{[r]}\|\mathcal{M}^{[r]}\right) = -\log_2 \alpha_{opt}$, where $\alpha_{opt}$ is the solution to the following optimization problem, which is an SDP:*

$$\begin{aligned}
\textit{minimize} \quad & \mathrm{Tr}[P^0\gamma(\mathcal{M}^{[r]})] \\
\textit{subject to} \quad & \mathrm{Tr}[P^1\gamma(\mathcal{E}^{[r]})] \leq \varepsilon, \\
& P^0, P^1 \geq 0, \\
& P^0_{A_1^rB_1^r} + P^1_{A_1^rB_1^r} = C^{(r)}_{B_1^{r-1}A_1^r} \otimes \mathbb{1}_{B_r}, \\
& C^{(r)} \geq 0, \\
& \mathrm{Tr}_{A_k}[C^{(k)}_{B_1^{k-1}A_1^k}] = C^{(k-1)}_{B_1^{k-2}A_1^{k-1}} \otimes \mathbb{1}_{B_{k-1}}, \\
& C^{(k-1)} \geq 0, \quad r \geq k \geq 2, \\
& \mathrm{Tr}_{A_1}[C^{(1)}_{A_1}] = 1, \\
& C^{(1)} \geq 0.
\end{aligned} \tag{143}$$

*Proof.* Let the $r$-combs $\mathcal{E}^{[r]}$ and $\mathcal{M}^{[r]}$ have input systems $A_1^r$ and output systems $B_1^r$. By definition, we have

$$D_{H,r}^\varepsilon\left(\mathcal{E}^{[r]}\|\mathcal{M}^{[r]}\right) = \sup_{\mathcal{D}^{[r]}_{(\text{co-st})}} D_H^\varepsilon\left(\rho_{D_rB_r}^{(\mathcal{E}^{[r]},\mathcal{D}^{[r]})}\|\rho_{D_rB_r}^{(\mathcal{M}^{[r]},\mathcal{D}^{[r]})}\right) \tag{144}$$

$$= \sup_{\mathcal{D}^{[r]}_{(\text{co-st})}} D_H^\varepsilon\left(\gamma(\mathcal{D}^{[r]}_{(\text{co-st})}) * \gamma(\mathcal{E}^{[r]})\|\gamma(\mathcal{D}^{[r]}_{(\text{co-st})}) * \gamma(\mathcal{M}^{[r]})\right) \tag{145}$$

$$= \sup_{\mathcal{D}^{[r]}_{(\text{co-st})}} \left(-\log_2 \inf_{\Lambda^0,\Lambda^1\geq 0} \left\{\mathrm{Tr}\left[\Lambda^0\left(\gamma(\mathcal{D}^{[r]}_{(\text{co-st})}) * \gamma(\mathcal{M}^{[r]})\right)\right] : \mathrm{Tr}\left[\Lambda^1\left(\gamma(\mathcal{D}^{[r]}_{(\text{co-st})}) * \gamma(\mathcal{E}^{[r]})\right)\right] \leq \varepsilon,\ \Lambda^0 + \Lambda^1 = 1\right\}\right) \tag{146}$$

$$= -\log_2 \inf_{\substack{\mathcal{D}^{[r]}_{(\text{co-st})},\\ \Lambda^0,\Lambda^1\geq 0}} \left\{\mathrm{Tr}\left[\Lambda^0\left(\gamma(\mathcal{D}^{[r]}_{(\text{co-st})}) * \gamma(\mathcal{M}^{[r]})\right)\right] : \mathrm{Tr}\left[\Lambda^1\left(\gamma(\mathcal{D}^{[r]}_{(\text{co-st})}) * \gamma(\mathcal{E}^{[r]})\right)\right] \leq \varepsilon,\ \Lambda^0 + \Lambda^1 = 1\right\} \tag{147}$$

where the optimization is with respect to co-strategy $r$-combs $\mathcal{D}^{[r]}_{(\text{co-st})}$ with input systems $B_1^{r-1}$, output systems $A_1^r D_r$, and memory systems $D_1^{r-1}$. Now, for any such co-strategy,

$$\begin{aligned}
&\operatorname{Tr}\left[\Lambda^0_{D_r B_r}\left(\gamma(\mathcal{D}^{[r]}_{(\text{co-st})})_{B_1^{r-1}A_1^r D_r} * \gamma(\mathcal{M}^{[r]})_{A_1^r B_1^r}\right)\right] \\
&\quad = \operatorname{Tr}_{D_r B_r}\left[\Lambda^0_{D_r B_r}\operatorname{Tr}_{A_1^r B_1^{r-1}}\left[\gamma(\mathcal{D}^{[r]}_{(\text{co-st})})_{B_1^{r-1}A_1^r D_r}\gamma(\mathcal{M}^{[r]})_{A_1^r B_1^r}^{\mathsf{T}_{A_1^r B_1^{r-1}}}\right]\right] &(148)\\
&\quad = \operatorname{Tr}\left[P^0_{A_1^r B_1^r}\gamma(\mathcal{M}^{[r]})_{A_1^r B_1^r}^{\mathsf{T}_{A_1^r B_1^{r-1}}}\right], &(149)
\end{aligned}$$

where

$$P^0_{A_1^r B_1^r} \coloneqq \operatorname{Tr}_{D_r}\left[\Lambda^0_{D_r B_r}\gamma(\mathcal{D}^{[r]}_{(\text{co-st})})_{B_1^{r-1}A_1^r D_r}\right]. \tag{150}$$

Similarly, we have

$$\begin{aligned}
&\operatorname{Tr}\left[\Lambda^1_{D_r B_r}\left(\gamma(\mathcal{D}^{[r]}_{(\text{co-st})})_{B_1^{r-1}A_1^r D_r} * \gamma(\mathcal{E}^{[r]})_{A_1^r B_1^r}\right)\right] \\
&\quad = \operatorname{Tr}\left[P^1_{A_1^r B_1^r}\gamma(\mathcal{E}^{[r]})_{A_1^r B_1^r}^{\mathsf{T}_{A_1^r B_1^{r-1}}}\right], &(151)
\end{aligned}$$

where

$$P^1_{A_1^r B_1^r} \coloneqq \operatorname{Tr}_{D_r}\left[\Lambda^1_{D_r B_r}\gamma(\mathcal{D}^{[r]}_{(\text{co-st})})_{B_1^{r-1}A_1^r D_r}\right]. \tag{152}$$

Note that both $P^0$ and $P^1$ are positive semi-definite. Also, using the fact that $\Lambda^0 + \Lambda^1 = \mathbb{1}$, we find that

$$P^0_{A_1^r B_1^r} + P^1_{A_1^r B_1^r} = \operatorname{Tr}_{D_r}\left[\gamma(\mathcal{D}^{[r]}_{(\text{co-st})})\right] \otimes \mathbb{1}_{B_r}. \tag{153}$$

Now, because $\mathcal{D}^{[r]}_{(\text{co-st})}$ is a co-strategy $r$-comb with input systems $B_1^{r-1}$ and output systems $A_1^r D_r$, it holds that $\operatorname{Tr}_{D_r}\left[\gamma(\mathcal{D}^{[r]}_{(\text{co-st})})\right]$ is the Choi representation of a co-strategy $r$-comb with input systems $B_1^{r-1}$ and output systems $A_1^r$. Therefore, letting

$$\operatorname{Tr}_{D_r}\left[\gamma(\mathcal{D}^{[r]}_{(\text{co-st})})\right] \equiv C^{(r)}_{B_1^{r-1}A_1^r}, \tag{154}$$

and using the constraints for a co-strategy in (21)–(22), we obtain the desired optimization problem. Because of the linear objective function and the linear and semi-definite constraints, it is manifestly an SDP. □

Let us now briefly discuss the case of hypothesis testing of multiple hypotheses in the "symmetric" setting, i.e., in which we consider the average error/success probability instead of individual errors. We mention that discrimination between two combs has been considered already in [2, 29, 47].

Suppose that an agent has multiple-round access to a quantum device, but the agent does not know with which device they are interacting. All the agent knows is that the device is picked randomly from a finite set $\mathcal{X}$ of devices and that the devices can be described mathematically by quantum $r$-combs $\mathcal{E}_x^{[r]}$, where $r$ is the number of allowed rounds. Furthermore, each device is picked with probability $p(x)$, where $p : \mathcal{X} \to [0, 1]$ is a probability distribution. The agent's task is to guess which device they are using. How should the agent proceed? The most general thing that the agent could do is interact with the device via a co-strategy $r$-comb followed by a measurement of the resulting quantum state. They can then use the measurement outcome to make a guess about the device. For any co-strategy $\mathcal{D}^{[r]}_{(\text{co-st})}$, the problem reduces to the multiple-state discrimination problem for the following ensemble of states:

$$\left\{\left(p(x), \rho_{D_r B_r}^{(\mathcal{E}_x^{[r]}, \mathcal{D}^{[r]}_{(\text{co-st})})}\right)\right\}_{x\in\mathcal{X}}, \tag{155}$$

for which the optimal guessing probability is [53, 54]:

$$\sup_{\Lambda^x \geq 0\ \forall x\in\mathcal{X}} \sum_{x\in\mathcal{X}} p(x) \operatorname{Tr}\left[\Lambda^x_{D_r B_r} \rho_{D_r B_r}^{(\mathcal{E}_x^{[r]}, \mathcal{D}^{[r]}_{(\text{co-st})})}\right]. \tag{156}$$

By optimizing this quantity with respect to all possible co-strategies, we obtain

$$p_{\text{guess}}(\{(p(x), \mathcal{E}_x^{[r]})\}) := \sup_{\substack{\mathcal{D}^{[r]}_{(\text{co-st})},\\ \Lambda^x \geq 0\ \forall x\in\mathcal{X}}} \sum_{x\in\mathcal{X}} p(x) \operatorname{Tr}\left[\Lambda^x_{D_r B_r} \rho_{D_r B_r}^{(\mathcal{E}^{[r]}, \mathcal{D}^{[r]}_{(\text{co-st})})}\right]. \tag{157}$$

By following the reasoning in the proof of Proposition 9, it is straightforward to show that this optimal guessing probability is equal to the solution of the following optimization problem, which is an SDP:

$$\begin{aligned}
\text{maximize} \quad & \sum_{x\in\mathcal{X}} \operatorname{Tr}[P^x \gamma(\mathcal{E}_x^{[r]})] \\
\text{subject to} \quad & P^x \geq 0 \quad \forall\, x \in \mathcal{X}, \\
& \sum_{x\in\mathcal{X}} P^x_{A_1^r B_1^r} = C^{(r)}_{B_1^{r-1} A_1^r} \otimes \mathbb{1}_{B_r}, \\
& C^{(r)} \geq 0, \\
& \operatorname{Tr}_{A_k}[C^{(k)}_{B_1^{k-1} A_1^k}] = C^{(k-1)}_{B_1^{k-2} A_1^{k-1}} \otimes \mathbb{1}_{B_{k-1}}, \\
& C^{(k-1)} \geq 0, \quad r \geq k \geq 2, \\
& \operatorname{Tr}_{A_1}[C^{(1)}_{A_1}] = 1, \\
& C^{(1)} \geq 0.
\end{aligned} \tag{158}$$

The optimal guessing probability, as given by the optimization problem above, leads to a generalization to quantum combs of the conditional min-entropy for classical-quantum states, based on the fact that the conditional min-entropy of a classical-quantum state is related to the optimal guessing probability for multiple-state discrimination [55]. The conditional min-entropy has already been defined for quantum combs in previous work [39].

## 4 Resource measures

In the previous section, we considered various general quantifiers for quantum causal networks. Let us now consider quantum causal networks as resources, and use the measures defined in the previous section to quantify their resourcefulness.

A resource theory is defined by a particular set $\mathfrak{F}$ of *free objects* that represents constrains on what can be achieved in a particular physical setting. Here, the free objects consist of a subset of quantum channels, i.e., $\mathfrak{F} \equiv \mathfrak{F}(A \to B)$ is the subset of quantum channels from a given system $A$ to a given system $B$. This free set should satisfy the following axioms [20, 22, 24]:

- *Identity*: The identity channel $\mathrm{id}_A$ for a system $A$ is contained in the set $\mathfrak{F}(A \to A)$.
- *Closure under composition*: For systems $A$, $B$, $C$, if $\mathcal{M} \in \mathfrak{F}(A \to B)$ and $\mathcal{N} \in \mathfrak{F}(B \to C)$, then $\mathcal{N} \circ \mathcal{M} \in \mathfrak{F}(A \to C)$.
- *Stability under tensor product*: For systems $A$, $B$, $C$, if $\mathcal{M} \in \mathfrak{F}(A \to B)$, then $\mathrm{id}_C \otimes \mathcal{M} \in \mathfrak{F}(CA \to CB)$.
- *Topological closure*: For systems $A$, $B$, $\mathfrak{F}(A \to B)$ is a topologically closed set.

Note that closure under composition and stability under tensor product imply that if $\mathcal{M}_1$ and $\mathcal{M}_2$ are free channels, then so is $\mathcal{M}_1 \otimes \mathcal{M}_2$.

Since quantum causal networks consist of quantum channels with a causal ordering, one way to construct a resource theory of quantum causal networks is to take a particular resource theory for quantum channels and consider a causal network to be free if its all of its constituent channels are free.

**Definition 10** (Set of free quantum combs)**.** *For $r \geq 1$, given any collection $A_1, \dots, A_r$ and $B_1, \dots, B_r$ of quantum systems, we define $\mathfrak{F}_r(A \to B)$ to be the set of quantum $r$-combs with input systems $A_1, \dots, A_r$ and output systems $B_1, \dots, B_r$ obtained by composing channels from given free sets in the manner depicted in Fig. 1. Specifically, an $r$-comb $\mathcal{E}^{[r]} = (\mathcal{E}^1, \dots, \mathcal{E}^r) \in \mathfrak{F}_r$ if and only if $\mathcal{E}^1 \in \mathfrak{F}(A_1 \to E_1 B_1)$, $\mathcal{E}^k \in \mathfrak{F}(A_k E_{k-1} \to B_k E_k)$ for all $2 \leq k \leq r-1$, and $\mathcal{E}^r \in \mathfrak{F}(A_r E_{r-1} \to B_r)$, where $E_1, \dots, E_{r-1}$ are memory systems. For brevity, we simply write $\mathfrak{F}_r$ when the input and output systems are understood from context or are unimportant.*

Note that, due to the causal constraints on quantum combs, $\mathfrak{F}_r(A \to B)$ is a subset of all free channels from $A_1^r \to B_1^r$, i.e.,

$$\mathfrak{F}_r(A \to B) \subseteq \mathfrak{F}(A_1^r \to B_1^r). \tag{159}$$

Free-channel sets commonly used in the analysis of quantum communication protocols are the sets of LOCC and PPT channels between the sending and receiving ends of a given quantum channel. In quantum thermodynamics, one considers the so-called "Gibbs preserving" channels; see, e.g., [21] for a review.

We now define the following resource measures analogously to the measures defined in [22, 24] for quantum channels.

**Definition 11** (Divergence-based resource measures)**.** *Let $\boldsymbol{D}$ be a generalized divergence for quantum states. Then, for an $r$-comb $\mathcal{E}^{[r]}$ with input systems $A_1^r$ and output systems $B_1^r$, we define two resource measures:*

$$\boldsymbol{D}_r^{\mathfrak{F}}(\mathcal{E}^{[r]}) := \inf_{\mathcal{M}^{[r]} \in \mathfrak{F}_r} \boldsymbol{D}_r\left(\mathcal{E}^{[r]} \middle\| \mathcal{M}^{[r]}\right) \tag{160}$$

$$= \inf_{\mathcal{M}^{[r]}\in\mathfrak{F}_r} \sup_{\mathcal{D}^{[r]}_{(co\text{-}st)}} \boldsymbol{D}\left(\rho^{(\mathcal{E}^{[r]},\mathcal{D}^{[r]}_{(co\text{-}st)})}_{D_rB_r}\middle\|\rho^{(\mathcal{M}^{[r]},\mathcal{D}^{[r]}_{(co\text{-}st)})}_{D_rB_r}\right), \tag{161}$$

*where the optimization is with respect to all co-strategies $\mathcal{D}^{[r]}_{(co\text{-}st)}$ with input systems $B_1^{r-1}$ and output systems $A_1^r D_r$. We also define*

$$\widetilde{\boldsymbol{D}}^{\mathfrak{F}}_r(\mathcal{E}^{[r]}) := \inf_{\mathcal{M}^{[r]}\in\mathfrak{F}_r} \sup_{\mathcal{D}^{[r]}_{(co\text{-}st)}\in\mathfrak{F}_r} \boldsymbol{D}\left(\rho^{(\mathcal{E}^{[r]},\mathcal{D}^{[r]}_{(co\text{-}st)})}_{D_rB_r}\middle\|\rho^{(\mathcal{M}^{[r]},\mathcal{D}^{[r]}_{(co\text{-}st)})}_{D_rB_r}\right) \tag{162}$$

$$= \inf_{\mathcal{M}^{[r]}\in\mathfrak{F}_r} \sup_{\mathcal{D}^{[r]}_{(co\text{-}st)}\in\mathfrak{F}_r} \boldsymbol{D}\left(\mathcal{E}^{[r]}\circ\mathcal{D}^{[r]}_{(co\text{-}st)}\middle\|\mathcal{M}^{[r]}\circ\mathcal{D}^{[r]}_{(co\text{-}st)}\right), \tag{163}$$

*where the second line is due to* (31).

Intuitively, the measures defined above quantify the resourcefulness of a quantum causal network by calculating its distance to the set of free networks. The larger the value of the quantity, the more resourceful the network is.

We note that for $r = 1$, the resource measures defined above reduce to the ones defined in [22] for quantum channels (which are 1-combs). This fact leads to another way of quantifying the resourcefulness of a quantum causal network, which is by considering its corresponding multipartite quantum channel as defined in (5). Then, for an $r$-comb $\mathcal{E}^{[r]}$ with input systems $A_1^r$ and output systems $B_1^r$, we can consider the quantity

$$\boldsymbol{D}^{\mathfrak{F}}_1(\mathcal{N}^{\mathcal{E}^{[r]}}) = \inf_{\mathcal{L}\in\mathfrak{F}(A_1^r\to B_1^r)} \boldsymbol{D}_1\left(\mathcal{N}^{\mathcal{E}^{[r]}}\middle\|\mathcal{L}\right), \tag{164}$$

where now the optimization is with respect to all free channels $\mathcal{L} : \mathrm{Lin}(A_1^r) \to \mathrm{Lin}(B_1^r)$. Now, as a result of (71), we have that

$$\boldsymbol{D}^{\mathfrak{F}}_r(\mathcal{E}^{[r]}) = \inf_{\mathcal{M}^{[r]}\in\mathfrak{F}_r(A\to B)} \boldsymbol{D}_r\left(\mathcal{E}^{[r]}\middle\|\mathcal{M}^{[r]}\right) \tag{165}$$

$$\geq \inf_{\mathcal{M}^{[r]}\in\mathfrak{F}_r(A\to B)} \boldsymbol{D}_1\left(\mathcal{N}^{\mathcal{E}^{[r]}}\middle\|\mathcal{N}^{\mathcal{M}^{[r]}}\right) \tag{166}$$

$$\geq \inf_{\mathcal{L}\in\mathfrak{F}(A_1^r\to B_1^r)} \boldsymbol{D}_1\left(\mathcal{N}^{\mathcal{E}^{[r]}}\middle\|\mathcal{L}\right) \tag{167}$$

$$= \boldsymbol{D}^{\mathfrak{F}}_1(\mathcal{N}^{\mathcal{E}^{[r]}}), \tag{168}$$

where the last inequality holds due to (159), i.e., because for all free $r$-combs $\mathcal{M}^{[r]}$, $\mathcal{N}^{\mathcal{M}^{[r]}}$ is an element of the set of free channels from $A_1^r$ to $B_1^r$. This tells us something that is intuitively clear: we can in general get more out of a quantum causal network by making use of its causal structure (and therefore using adaptive inputs) than by simply using it as an ordinary quantum channel.

The quantities defined in Definition 11 satisfy two properties that are necessary in order for them to be considered resource measures.

**Theorem 11.** *For all $r \geq 1$, the measures $\boldsymbol{D}^{\mathfrak{F}}_r$ and $\widetilde{\boldsymbol{D}}^{\mathfrak{F}}_r$ satisfy the following properties:*

- *Data processing (monotonicity): If $\mathcal{T}^{[2r]} \in \mathfrak{F}_{2r}$ is a free $(2r)$-comb of the type shown in Fig. 7, then for all $r$-combs $\mathcal{E}^{[r]}$,*

$$\boldsymbol{D}_r^{\mathfrak{F}}(\mathcal{T}^{[2r]} \circ \mathcal{E}^{[r]}) \leq \boldsymbol{D}_r^{\mathfrak{F}}(\mathcal{E}^{[r]}), \tag{169}$$

$$\widetilde{\boldsymbol{D}}_r^{\mathfrak{F}}(\mathcal{T}^{[2r]} \circ \mathcal{E}^{[r]}) \leq \widetilde{\boldsymbol{D}}_r^{\mathfrak{F}}(\mathcal{E}^{[r]}). \tag{170}$$

*Similarly, if $\mathcal{F}^{[r+1]} \in \mathfrak{F}_{r+1}$ is a free $(r+1)$-comb of the type shown in Fig. 8, then for all $r$-combs $\mathcal{E}^{[r]}$,*

$$\boldsymbol{D}_{r-1}^{\mathfrak{F}}(\mathcal{F}^{[r+1]} \circ \mathcal{E}^{[r]}) \leq \boldsymbol{D}_r^{\mathfrak{F}}(\mathcal{E}^{[r]}), \tag{171}$$

$$\widetilde{\boldsymbol{D}}_{r-1}^{\mathfrak{F}}(\mathcal{F}^{[r+1]} \circ \mathcal{E}^{[r]}) \leq \widetilde{\boldsymbol{D}}_r^{\mathfrak{F}}(\mathcal{E}^{[r]}). \tag{172}$$

- *Faithfulness: If the generalized divergence $\boldsymbol{D}$ (for states) is faithful, meaning that $\boldsymbol{D}(\rho\|\sigma) = 0$ for all states $\rho$, $\sigma$ if and only if $\rho = \sigma$, then*

$$\boldsymbol{D}_r^{\mathfrak{F}}(\mathcal{E}^{[r]}) = 0 \quad \Longleftrightarrow \quad \mathcal{E}^{[r]} \in \mathfrak{F}_r, \tag{173}$$

$$\widetilde{\boldsymbol{D}}_r^{\mathfrak{F}}(\mathcal{E}^{[r]}) = 0 \quad \Longleftarrow \quad \mathcal{E}^{[r]} \in \mathfrak{F}_r. \tag{174}$$

*Proof.* To see (169), note that because $\mathcal{T}^{[2r]}$ is free, and the optimization in $\boldsymbol{D}_r^{\mathfrak{F}}(\mathcal{T}^{[2r]} \circ \mathcal{E}^{[r]})$ is with respect to free $r$-combs, we have that $\{\mathcal{T}^{[2r]} \circ \widetilde{\mathcal{M}}^{[r]} : \widetilde{\mathcal{M}}^{[r]} \in \mathfrak{F}_r\} \subseteq \mathfrak{F}_r$. Restricting the optimization to this set, and using the data-processing inequality in Theorem 5, gives us

$$\begin{aligned}
\boldsymbol{D}_r^{\mathfrak{F}}(\mathcal{T}^{[2r]} \circ \mathcal{E}^{[r]}) &= \inf_{\mathcal{M}^{[r]} \in \mathfrak{F}_r} \boldsymbol{D}_r\left(\mathcal{T}^{[2r]} \circ \mathcal{E}^{[r]} \middle\| \mathcal{M}^{[r]}\right) &(175)\\
&\leq \inf_{\widetilde{\mathcal{M}}^{[r]} \in \mathfrak{F}_r} \boldsymbol{D}_r\left(\mathcal{T}^{[2r]} \circ \mathcal{E}^{[r]} \middle\| \mathcal{T}^{[2r]} \circ \widetilde{\mathcal{M}}^{[r]}\right) &(176)\\
&\leq \inf_{\widetilde{\mathcal{M}}^{[r]} \in \mathfrak{F}_r} \boldsymbol{D}_r\left(\mathcal{E}^{[r]} \middle\| \widetilde{\mathcal{M}}^{[r]}\right) &(177)\\
&= \boldsymbol{D}_r^{\mathfrak{F}}(\mathcal{E}^{[r]}), &(178)
\end{aligned}$$

as required. The same reasoning, except using the data-processing inequality in Theorem 7, leads to a proof of (171).

To see (170), we again restrict optimization to the set $\{\mathcal{T}^{[2r]} \circ \widetilde{\mathcal{M}}^{[r]} : \widetilde{\mathcal{M}}^{[r]} \in \mathfrak{F}_r\}$ and use Theorem 5 to obtain

$$\begin{aligned}
&\widetilde{\boldsymbol{D}}_r^{\mathfrak{F}}(\mathcal{T}^{[2r]} \circ \mathcal{E}^{[r]})\\
&= \inf_{\mathcal{M}^{[r]} \in \mathfrak{F}_r} \sup_{\mathcal{D}^{[r]}_{(\text{co-st})} \in \mathfrak{F}_r} \boldsymbol{D}\left(\mathcal{T}^{[2r]} \circ \mathcal{E}^{[r]} \circ \mathcal{D}^{[r]}_{(\text{co-st})} \middle\| \mathcal{M}^{[r]} \circ \mathcal{D}^{[r]}_{(\text{co-st})}\right) &(179)\\
&\leq \inf_{\widetilde{\mathcal{M}} \in \mathfrak{F}_r} \sup_{\mathcal{D}^{[r]}_{(\text{co-st})} \in \mathfrak{F}_r} \boldsymbol{D}\left(\mathcal{T}^{[2r]} \circ \mathcal{E}^{[r]} \circ \mathcal{D}^{[r]}_{(\text{co-st})} \middle\| \mathcal{T}^{[2r]} \circ \widetilde{\mathcal{M}}^{[r]} \circ \mathcal{D}^{[r]}_{(\text{co-st})}\right) &(180)\\
&\leq \inf_{\widetilde{\mathcal{M}} \in \mathfrak{F}_r} \sup_{\mathcal{D}^{[r]}_{(\text{co-st})} \in \mathfrak{F}_r} \boldsymbol{D}\left(\mathcal{E}^{[r]} \circ \mathcal{D}^{[r]}_{(\text{co-st})} \middle\| \widetilde{\mathcal{M}}^{[r]} \circ \mathcal{D}^{[r]}_{(\text{co-st})}\right) &(181)
\end{aligned}$$

$$= \widetilde{\boldsymbol{D}}_r^{\mathfrak{F}}(\mathcal{E}^{[r]}), \tag{182}$$

as required. The same reasoning, except using using the data-processing inequality in Theorem 7, leads to a proof of (172).

For faithfulness, we first note that the property $\boldsymbol{D}(\rho\|\sigma) = 0$ if and only if $\rho = \sigma$ implies that $\boldsymbol{D}(\rho\|\sigma) \geq 0$ for all $\rho$ and $\sigma$. This implies that, if $\mathcal{E}^{[r]} \in \mathfrak{F}_r$, then $\boldsymbol{D}_r^{\mathfrak{F}}(\mathcal{E}^{[r]}) = 0$ simply because the optimal $\mathcal{M}^{[r]} \in \mathfrak{F}_r$ in (160) can be taken to be $\mathcal{E}^{[r]}$ itself. The same reasoning can be applied to $\widetilde{\boldsymbol{D}}_r^{\mathfrak{F}}$ to conclude (174). Finally, if $\boldsymbol{D}_r^{\mathfrak{F}}(\mathcal{E}^{[r]}) = 0$, then

$$\mathcal{E}^{[r]} \circ \mathcal{D}^{[r]}_{\text{(co-st)}} = \mathcal{M}^{[r]} \circ \mathcal{D}^{[r]}_{\text{(co-st)}} \tag{183}$$

for all choices of $\mathcal{D}^{[r]}_{\text{(co-st)}}$ and $\mathcal{M}^{[r]} \in \mathfrak{F}_r$. In particular, for any choice of $\mathcal{M}^{[r]} \in \mathfrak{F}_r$, the equality above holds for all $\mathcal{D}^{[r]}_{\text{(co-st)}}$, which implies that $\mathcal{E}^{[r]} = \mathcal{M}^{[r]}$, i.e., $\mathcal{E}^{[r]} \in \mathfrak{F}_r$. □

A quantity that is used very commonly in resource theories is the so-called *robustness*, which can be written in terms of the max-relative entropy [56], which is defined for two positive semi-definite operators $X$ and $Y$ as

$$D_{\max}(X\|Y) \coloneqq \log_2 \inf\{\lambda : X \leq \lambda Y\}. \tag{184}$$

The logarithm of the robustness, called *log-robustness*, is then defined for $r$-combs using the max-relative entropy for $r$-combs via the construction in Definition 2 and Definition 11:

$$\mathrm{LR}_r^{\mathfrak{F}}(\mathcal{E}^{[r]}) \equiv D^{\mathfrak{F}}_{\max,r}(\mathcal{E}^{[r]}) \tag{185}$$

$$= \inf_{\mathcal{M}^{[r]} \in \mathfrak{F}_r} D_{\max,r}\left(\mathcal{E}^{[r]} \middle\| \mathcal{M}^{[r]}\right), \tag{186}$$

which is a direct generalization of the definition of log-robustness for quantum channels (see, e.g., [24]). In particular, for $r = 1$, the definition above coincides with the definition in [24].

Since the max-relative entropy is isometrically invariant, we can use Proposition 5 to conclude a very simple expression for the $r$-comb max-relative entropy $D_{\max,r}$. (See also [39, Proposition 7] for a proof of this fact using a different method.)

**Corollary 12** (Max-relative entropy between quantum combs)**.** *For any two $r$-combs $\mathcal{E}^{[r]}$ and $\mathcal{M}^{[r]}$,*

$$D_{\max,r}\left(\mathcal{E}^{[r]} \middle\| \mathcal{M}^{[r]}\right) = D_{\max}\left(\gamma(\mathcal{E}^{[r]}) \middle\| \gamma(\mathcal{M}^{[r]})\right). \tag{187}$$

*Proof.* By Proposition 5, we have that

$$D_{\max,r}\left(\mathcal{E}^{[r]} \middle\| \mathcal{M}^{[r]}\right)$$

$$= \sup_{\substack{\mathcal{D}^{[r]}_{\text{(co-st)}},\\ P}} D_{\max}\left(\sqrt{P}\gamma(\mathcal{E}^{[r]})\sqrt{P} \middle\| \sqrt{P}\gamma(\mathcal{M}^{[r]})\sqrt{P}\right) \tag{188}$$

$$= \sup_{\substack{\mathcal{D}^{[r]}_{\text{(co-st)}},\\ P}} \inf_{\lambda} \left\{\lambda : \sqrt{P}\gamma(\mathcal{E}^{[r]})\sqrt{P} \leq \lambda \sqrt{P}\gamma(\mathcal{M}^{[r]})\sqrt{P}\right\}. \tag{189}$$

Now, for any choice of $\mathcal{D}^{[r]}_{(\text{co-st})}$ and its associated positive semi-definite operator $P$, it holds that

$$\sqrt{P}\gamma(\mathcal{E}^{[r]})\sqrt{P} \leq \lambda\sqrt{P}\gamma(\mathcal{M}^{[r]})\sqrt{P} \Leftrightarrow \gamma(\mathcal{E}^{[r]}) \leq \lambda\gamma(\mathcal{M}^{[r]}). \tag{190}$$

In other words, the optimal choice of $\lambda$ is independent of the co-strategy. The optimization over co-strategies is therefore unnecessary, and we obtain

$$D_{\max,r}\left(\mathcal{E}^{[r]}\middle\|\mathcal{M}^{[r]}\right) = \inf_{\lambda}\{\lambda : \gamma(\mathcal{E}^{[r]}) \leq \lambda\gamma(\mathcal{M}^{[r]})\} \tag{191}$$

$$= D_{\max}\left(\gamma(\mathcal{E}^{[r]})\middle\|\gamma(\mathcal{M}^{[r]})\right), \tag{192}$$

as required. □

Corollary 12 tells us that the max-relative entropy between quantum combs is given simply by the usual max-relative entropy in (184) between the Choi representations. Recalling the fact that the Choi representation of any comb $\mathcal{E}^{[r]}$ is simply the Choi representation of the corresponding multipartite channel $\mathcal{N}^{\mathcal{E}^{[r]}}$ (see (6)), we see that equality holds in (168) for the max-relative entropy, i.e.,

$$\text{LR}^{\mathfrak{F}}_r(\mathcal{E}^{[r]}) = \text{LR}^{\mathfrak{F}}_1(\mathcal{N}^{\mathcal{E}^{[r]}}), \tag{193}$$

so that the log-robustness of a quantum causal network is equal to the log-robustness of its corresponding multipartite quantum channel. This fact has an important consequence: using the log-robustness to quantify the resourcefulness of a quantum causal network does not take adaptiveness into account. Moreover, the error bounds in [27, 28] based on log-robustness do not lead to new results when applied to quantum causal networks.

So far, we have considered resource measures that essentially quantify how far the given causal network is from the set of free causal networks. Let us now consider measures that are based on how well a given network can be transformed into another given network using free networks. In Sec. 3.4, we considered two types of transformations of $r$-combs, as shown in Fig. 7 and Fig. 8. The first type of transformation takes an $r$-comb and transforms it into another $r$-comb by applying a $(2r)$-comb that performs a pre- and post-processing of the channels in the given $r$-comb. Let us denote these "transformations of the first kind" by $\mathcal{T}^{[2r]}_{\text{I}}$. The second type of transformation takes an $r$-comb and transforms it into an $(r-1)$-comb by applying a $(r+1)$-comb with the structure shown in Fig. 8. We denote these "transformations of the second kind" by $\mathcal{F}^{[r+1]}_{\text{II}}$. Given an $r$-comb $\mathcal{E}^{[r]}$ and a target $r$-comb $\mathcal{M}^{[r]}$, we can quantify the error in the transformation of $\mathcal{E}^{[r]}$ into $\mathcal{M}^{[r]}$ via a transformation of the first kind by

$$d^{\text{I}}_{\boldsymbol{D}}\left(\mathcal{E}^{[r]} \xrightarrow{\mathcal{T}^{[2r]}_{\text{I}}} \mathcal{M}^{[r]}\right) \coloneqq \boldsymbol{D}_r\left(\mathcal{E}^{[r]} \circ \mathcal{T}^{[2r]}_{\text{I}}\middle\|\mathcal{M}^{[r]}\right). \tag{194}$$

For a target $(r-1)$-comb $\mathcal{N}^{[r-1]}$, the error in the transformation of $\mathcal{E}^{[r]}$ into $\mathcal{N}^{[r-1]}$ via a transformation of the second kind is

$$d^{\text{II}}_{\boldsymbol{D}}\left(\mathcal{E}^{[r]} \xrightarrow{\mathcal{F}^{[r+1]}_{\text{II}}} \mathcal{N}^{[r-1]}\right) \coloneqq \boldsymbol{D}_{r-1}\left(\mathcal{E}^{[r]} \circ \mathcal{F}^{[r+1]}_{\text{II}}\middle\|\mathcal{N}^{[r-1]}\right). \tag{195}$$

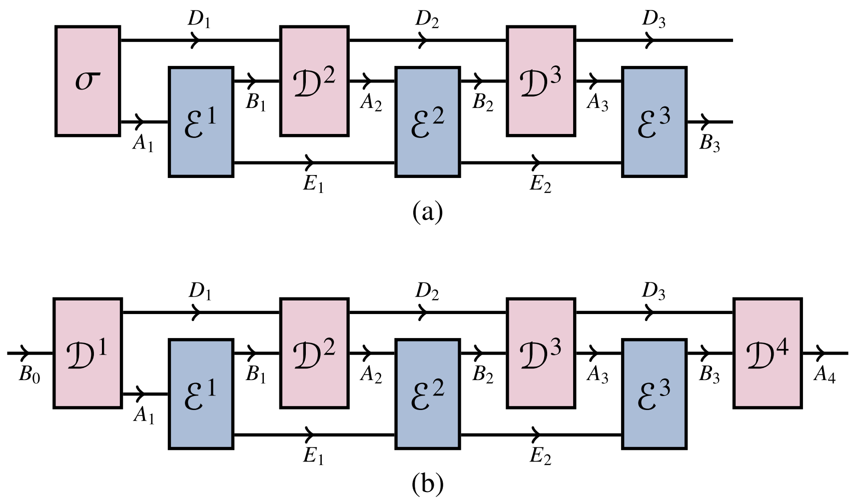



FIGURE 9: (a) Transformation of a strategy 3-comb $\mathcal{E}^{[3]}$ into a quantum state for systems $D_3$ and $B_3$ via combination with a co-strategy 3-comb $\mathcal{D}^{[3]}_{(\text{co-st})}$. (b) Transformation of a strategy 3-comb $\mathcal{E}^{[3]}$ into a quantum channel from $B_0$ to $A_4$ via combination with a strategy 4-comb $\mathcal{D}^{[4]}_{(\text{st})}$.

We can define analogous error quantities using the fidelity and root-fidelity of quantum combs, as defined in Definition 6.

Now, in a resource theory, we are allowed only free transformations. In other words, the channels in the combs $\mathcal{T}_{\text{I}}^{[2r]}$ and $\mathcal{F}_{\text{II}}^{[r+1]}$ must be free channels. Under this restriction, the optimal errors are

$$d^{\text{I}}_{\boldsymbol{D}}\left(\mathcal{E}^{[r]} \xrightarrow{\mathfrak{F}_{2r}} \mathcal{M}^{[r]}\right) := \inf_{\mathcal{T}_{\text{I}}^{[2r]} \in \mathfrak{F}_{2r}} d^{\text{I}}_{\boldsymbol{D}}\left(\mathcal{E}^{[r]} \xrightarrow{\mathcal{T}_{\text{I}}^{[r]}} \mathcal{M}^{[r]}\right) \tag{196}$$

$$= \inf_{\mathcal{T}_{\text{I}}^{[2r]} \in \mathfrak{F}_{2r}} \boldsymbol{D}_r\left(\mathcal{E}^{[r]} \circ \mathcal{T}_{\text{I}}^{[r]} \middle\| \mathcal{M}^{[r]}\right), \tag{197}$$

$$d^{\text{II}}_{\boldsymbol{D}}\left(\mathcal{E}^{[r]} \xrightarrow{\mathfrak{F}_{r+1}} \mathcal{M}^{[r]}\right) := \inf_{\mathcal{F}_{\text{II}}^{[r+1]} \in \mathfrak{F}_{r+1}} d^{\text{II}}_{\boldsymbol{D}}\left(\mathcal{E}^{[r]} \xrightarrow{\mathcal{F}_{\text{II}}^{[r+1]}} \mathcal{M}^{[r]}\right) \tag{198}$$

$$= \inf_{\mathcal{F}_{\text{II}}^{[r+1]} \in \mathfrak{F}_{r+1}} \boldsymbol{D}_{r-1}\left(\mathcal{E}^{[r]} \circ \mathcal{F}_{\text{II}}^{[r+1]} \middle\| \mathcal{M}^{[r]}\right). \tag{199}$$

Let us now consider transformation tasks that are based on taking a given resource $\mathcal{E}^{[r]}$ and transforming it to either a quantum state or a quantum channel using free combs.

## 4.1 Transformations to states

Consider transforming a given (resourceful) strategy $r$-comb $\mathcal{E}^{[r]}$ with input systems $A_1^r$ and output systems $B_1^r$ into a given (resourceful) quantum state $\sigma_{D_r B_r}$ via a free co-strategy $r$-comb $\mathcal{D}^{[r]}_{(\text{co-st})}$ with input systems $B_1^{r-1}$ and output systems $A_1^r D_r$, as shown in Fig. 9(a). For this task, the figure of merit (i.e., the transformation error) is

$$d_{\boldsymbol{D}}\left(\mathcal{E}^{[r]} \overset{\mathcal{D}^{[r]}_{(\text{co-st})}}{\to} \sigma\right) := \boldsymbol{D}\left(\rho^{(\mathcal{E}^{[r]},\mathcal{D}^{[r]}_{(\text{co-st})})}_{D_r B_r} \middle\| \sigma_{D_r B_r}\right). \tag{200}$$

The optimal error is then

$$d_{\boldsymbol{D}}\left(\mathcal{E}^{[r]} \overset{\mathfrak{F}_r}{\to} \sigma\right) := \inf_{\mathcal{D}^{[r]}_{(\text{co-st})} \in \mathfrak{F}_r} d_{\boldsymbol{D}}\left(\mathcal{E}^{[r]} \overset{\mathcal{D}^{[r]}_{(\text{co-st})}}{\to} \sigma\right) \tag{201}$$

$$= \inf_{\mathcal{D}^{[r]}_{(\text{co-st})} \in \mathfrak{F}_r} \boldsymbol{D}\left(\rho^{(\mathcal{E}^{[r]},\mathcal{D}^{[r]}_{(\text{co-st})})}_{D_r B_r} \middle\| \sigma\right), \tag{202}$$

where the optimization is with respect to free co-strategy $r$-combs $\mathcal{D}^{[r]}_{(\text{co-st})}$ with input systems $B_1^r$ and output systems $A_1^r D_r$. We typically use the trace distance as the divergence in (202) due to its operational interpretation in the context of binary state discrimination, so that

$$d_{\text{tr}}\left(\mathcal{E}^{[r]} \to \sigma\right) := \inf_{\mathcal{D}^{[r]}_{(\text{co-st})} \in \mathfrak{F}_r} \frac{1}{2}\left\| \rho^{(\mathcal{E}^{[r]},\mathcal{D}^{[r]}_{(\text{co-st})})}_{D_r B_r} - \sigma_{D_r B_r} \right\|_1. \tag{203}$$

The error quantity in (202) can be thought of as a resource measure for the pair $(\mathcal{E}^{[r]}, \sigma)$ of resources.

## 4.2 Transformations to channels

We can consider all of the aforementioned questions in the channel scenario as well. Specifically, consider transformation a given strategy $r$-comb $\mathcal{E}^{[r]}$ with input systems $A_1^r$ and output systems $B_1^r$ into a given quantum channel $\mathcal{N}_{B_0 \to A_{r+1}}$ via a free strategy $(r+1)$-comb $\mathcal{D}^{[r+1]}_{(st)}$ with input systems $B_0^r$ and output systems $A_1^{r+1}$, as shown in Fig. 9(b). For this task, the figure of merit (i.e., the transformation error) is

$$d_{\boldsymbol{D}}\left(\mathcal{E}^{[r]} \overset{\mathcal{D}^{[r+1]}_{(\text{st})}}{\to} \mathcal{N}\right) = \boldsymbol{D}_1\left(\mathcal{E}^{[r]} \circ \mathcal{D}^{[r+1]}_{(\text{st})} \middle\| \mathcal{N}\right), \tag{204}$$

The optimal error is then

$$d_{\boldsymbol{D}}\left(\mathcal{E}^{[r]} \overset{\mathfrak{F}_{r+1}}{\to} \mathcal{N}\right) := \inf_{\mathcal{D}^{[r+1]}_{(\text{st})} \in \mathfrak{F}_{r+1}} d_{\boldsymbol{D}}\left(\mathcal{E}^{[r]} \overset{\mathcal{D}^{[r+1]}_{(\text{st})}}{\to} \mathcal{N}\right) \tag{205}$$

$$= \inf_{\mathcal{D}^{[r+1]}_{(\text{st})} \in \mathfrak{F}_{r+1}} \boldsymbol{D}_1\left(\mathcal{E}^{[r]} \circ \mathcal{D}^{[r+1]}_{(\text{st})} \middle\| \mathcal{N}\right), \tag{206}$$

where the optimization is with respect to free strategy $(r+1)$-combs $\mathcal{D}^{[r+1]}_{(\text{st})}$ with input systems $B_0^r$ and output systems $A_1^{r+1}$. We typically use the diamond distance as the divergence in (206) due to its operational interpretation in terms of binary channel discrimination, so that

$$d_{\diamond}\left(\mathcal{E}^{[r]} \overset{\mathfrak{F}_{r+1}}{\to} \mathcal{N}\right) := \inf_{\mathcal{D}^{[r+1]}_{(\text{st})} \in \mathfrak{F}_{r+1}} \frac{1}{2}\left\| \mathcal{E}^{[r]} \circ \mathcal{D}^{[r+1]}_{(\text{st})} - \mathcal{N} \right\|_{\diamond}. \tag{207}$$

This definition of the transformation error in terms of the diamond distance is analogous what has been defined before [57] (see also [23]). As before, the error quantity in (206) can be thought of as a resource measure for the pair $(\mathcal{E}^{[r]}, \mathcal{N})$ of resources.

### 4.3 Distillation tasks

The task of distillation is about using multiple (parallel) instances of a given resource $\mathcal{E}^{[r]}$ and using them to obtain another resource. Let us consider distilling to states and channels. We define the following quantities:

$$R_{\boldsymbol{D}}^{n,\varepsilon}\left(\mathcal{E}^{[r]} \overset{\mathfrak{F}_{nr}}{\to} \sigma\right) \coloneqq \sup_{m\in\mathbb{N}} \left\{\frac{m}{n} : d_{\boldsymbol{D}}\left((\mathcal{E}^{[r]})^{\times n} \to \sigma^{\otimes m}\right) \leq \varepsilon\right\}, \tag{208}$$

$$R_{\boldsymbol{D}}^{n,\varepsilon}\left(\mathcal{E}^{[r]} \overset{\mathfrak{F}_{nr+1}}{\to} \mathcal{N}\right) \coloneqq \sup_{m\in\mathbb{N}} \left\{\frac{m}{n} : d_{\boldsymbol{D}}\left((\mathcal{E}^{[r]})^{\times n} \to \mathcal{N}^{\otimes m}\right) \leq \varepsilon\right\}, \tag{209}$$

which are the maximum rates at which copies of $\sigma$ and $\mathcal{N}$, respectively, can be distilled from $n$ uses of the comb $\mathcal{E}^{[r]}$. Note that

$$(\mathcal{E}^{[r]})^{\times n} = (\underbrace{\mathcal{E}^{[r]}, \mathcal{E}^{[r]}, \dots, \mathcal{E}^{[r]}}_{n \text{ times}}) \tag{210}$$

is an $(nr)$-comb, which corresponds to the fact that we can use comb $\mathcal{E}^{[r]}$ itself adaptively as opposed to simply in parallel, which would be described by the $r$-comb $(\mathcal{E}^{[r]})^{\otimes n}$. This is also why the free operations for the case of states are $(nr)$-combs, because each use of the comb involves $r$ rounds. Similarly, for the transformation to a channel, the free operations are $(nr+1)$-combs.

In the asymptotic setting, we are interested in the values

$$\inf_{\varepsilon\in(0,1)} \liminf_{n\to\infty} R_{\boldsymbol{D}}^{n,\varepsilon}\left(\mathcal{E}^{[r]} \overset{\mathfrak{F}_{nr}}{\to} \sigma\right), \tag{211}$$

$$\inf_{\varepsilon\in(0,1)} \liminf_{n\to\infty} R_{\boldsymbol{D}}^{n,\varepsilon}\left(\mathcal{E}^{[r]} \overset{\mathfrak{F}_{nr+1}}{\to} \mathcal{N}\right), \tag{212}$$

which are the highest rates at which the comb $\mathcal{E}^{[r]}$ can be distilled to a $\sigma$ or $\mathcal{N}$, respectively, with asymptotically vanishing error.

## 5 Summary & outlook

In this note, we have looked at various ways of quantifying quantum causal networks. We considered a family of measures based on generalized divergences for states that quantify how far apart two quantum causal networks are. We also considered a notion of fidelity between quantum causal networks and a notion of the norm of a causal network via the Schatten $\alpha$-norms. The main idea behind all of these definitions is to take a quantum causal network, apply a corresponding co-strategy to it, and then evaluate the usual state measure on the state obtained at the end of the interaction. We looked at data-processing inequalities for these quantities, and we looked at an application to hypothesis testing and discrimination of quantum causal networks.

Using the generalized divergence between quantum causal networks leads to definitions of resource measures for them. Because quantum causal networks can allow for adaptive inputs, we show that these resource measures are in principle different from the usual quantum channel resource measures that been previously defined. However, for the log-robustness, we showed that the resource measure for quantum causal networks coincides with the measure for ordinary quantum channels. This means that prior results on resource interconversions for quantum channels, which make use of the log-robustness, can be directly applied to quantum causal networks, leading to estimates of their resourcefulness. It also means, however, that there is potential for improvement in the analysis of the resourcefulness of quantum causal networks using different techniques, because the log-robustness does not take adaptiveness into account. Specifically, for future work, it would be interesting to investigate the resource interconversion problems defined in Sec. 4.1, 4.2, and 4.3.

Other directions for future work include: extending the generalized divergence and fidelity measures considered here to continuous-variable systems, in manner similar to recent work [58]. It is also worthwhile to prove continuity results for the information quantities considered here, in a manner similar to [22], and to prove analogues of Stein's lemma for (asymmetric) hypothesis testing in the asymptotic setting.

## References


[1] J. Pearl, *Causality*, 2nd ed. (Cambridge University Press, 2009).

[2] G. Chiribella, G. M. D'Ariano, and P. Perinotti, "Theoretical framework for quantum networks", Physical Review A **80**, 022339 (2009).

[3] M. S. Leifer and R. W. Spekkens, "Towards a formulation of quantum theory as a causally neutral theory of bayesian inference", Physical Review A **88**, 052130 (2013).

[4] J.-M. A. Allen, J. Barrett, D. C. Horsman, C. M. Lee, and R. W. Spekkens, "Quantum common causes and quantum causal models", Physical Review X **7**, 031021 (2017).

[5] D. Schmid, K. Ried, and R. W. Spekkens, "Why initial system-environment correlations do not imply the failure of complete positivity: a causal perspective", Physical Review A **100**, 022112 (2019).

[6] V. Giovannetti, S. Lloyd, and L. Maccone, "Quantum metrology", Physical Review Letters **96**, 010401 (2006).

[7] A. Hentschel and B. C. Sanders, "Efficient algorithm for optimizing adaptive quantum metrology processes", Physical Review Letters **107**, 233601 (2011).

[8] R. Demkowicz-Dobrza ński, J. Czajkowski, and P. Sekatski, "Adaptive quantum metrology under general markovian noise", Physical Review X **7**, 041009 (2017).

[9] M. Takeoka, S. Guha, and M. M. Wilde, "Fundamental rate-loss tradeoff for optical quantum key distribution", Nature Communications **5**, 5235 (2014).

[10] K. Azuma, A. Mizutani, and H.-K. Lo, "Fundamental rate-loss trade-off for the quantum internet", Nature Communications **7**, 13523 (2016).

[11] S. Pirandola, R. Laurenza, C. Ottaviani, and L. Banchi, "Fundamental limits of repeaterless quantum communications", Nature Communications **8**, 15043 (2017).

[12] S. Pirandola, "Bounds for multi-end communication over quantum networks", Quantum Science and Technology **4**, 045006 (2019).

[13] S. Pirandola, "End-to-end capacities of a quantum communication network", Communications Physics **2**, 51 (2019).

[14] S. Khatri, "Policies for elementary link generation in quantum networks", arXiv:2007.03193 (2020).

[15] B. M. Terhal, "Quantum error correction for quantum memories", Reviews of Modern Physics **87**, 307 (2015).

[16] T. Vidick and J. Watrous, "Quantum proofs", Foundations and Trends® in Theoretical Computer Science **11**, 1 (2016).

[17] V. Dunjko, J. M. Taylor, and H. M. Briegel, "Framework for learning agents in quantum environments", arXiv:1507.08482 (2015).

[18] V. Dunjko, J. M. Taylor, and H. J. Briegel, "Quantum-enhanced machine learning", Physical Review Letters **117**, 130501 (2016).

[19] G. Gutoski and J. Watrous, "Toward a general theory of quantum games", in Proceedings of the Thirty-Ninth Annual ACM Symposium on Theory of Computing, STOC '07, arXiv:quant-ph/0611234 (2007), pp. 565–574.

[20] F. G. S. L. Brandão and G. Gour, "Reversible framework for quantum resource theories", Physical Review Letters **115**, 070503 (2015).

[21] E. Chitambar and G. Gour, "Quantum resource theories", Reviews of Modern Physics **91**, 025001 (2019).

[22] G. Gour and A. Winter, "How to quantify a dynamical quantum resource", Physical Review Letters **123**, 150401 (2019).

[23] G. Gour and C. M. Scandolo, "The entanglement of a bipartite channel", arXiv:1907.02552 (2019).

[24] Z.-W. Liu and A. Winter, "Resource theories of quantum channels and the universal role of resource erasure", arXiv:1904.04201 (2019).

[25] K. Fang and Z.-W. Liu, "No-go theorems for quantum resource purification", Physical Review Letters **125**, 060405 (2020).

[26] Y. Liu and X. Yuan, "Operational resource theory of quantum channels", Physical Review Research **2**, 012035 (2020).

[27] K. Fang and Z.-W. Liu, "No-go theorems for quantum resource purification: new approach and channel theory", arXiv:2010.11822 (2020).

[28] B. Regula and R. Takagi, "Fundamental limitations on quantum channel manipulation", arXiv:2010.11942 (2020).

[29] G. Gutoski, "On a measure of distance for quantum strategies", Journal of Mathematical Physics **53**, 032202 (2012).

[30] G. Chiribella, G. M. D'Ariano, and P. Perinotti, "Quantum circuit architecture", Physical Review Letters **101**, 060401 (2008).

[31] A. Bisio, G. Chiribella, G. M. D'Ariano, and P. Perinotti, "Quantum Networks: General Theory and Applications", Acta Physica Slovaca **61**, arXiv:1601.04864, 273 (2011).

[32] Y. Polyanskiy and S. Verdú, "Arimoto channel coding converse and Rényi divergence", in 2010 48th Annual Allerton Conference on Communication, Control, and Computing (Allerton) (2010), pp. 1327–1333.

[33] N. Sharma and N. A. Warsi, "On the strong converses for the quantum channel capacity theorems", arXiv:1205.1712 (2012).

[34] H. Umegaki, "Conditional expectation in an operator algebra. IV. Entropy and information", Kodai Mathematical Seminar Reports **14**, 59 (1962).

[35] M. Tomamichel, *Quantum Information Processing with Finite Resources* (Springer International Publishing, 2016).

[36] F. Leditzky, E. Kaur, N. Datta, and M. M. Wilde, "Approaches for approximate additivity of the holevo information of quantum channels", Physical Review A **97**, 012332 (2018).

[37] A. Kitaev, "Quantum computations: algorithms and error correction", Russian Mathematical Surveys **52**, 1191 (1997).

[38] X. Wang and M. M. Wilde, "Resource theory of asymmetric distinguishability for quantum channels", Physical Review Research **1**, 033169 (2019).

[39] G. Chiribella and D. Ebler, "Optimal quantum networks and one-shot entropies", New Journal of Physics **18**, 093053 (2016).

[40] J. Watrous, *The Theory of Quantum Information* (Cambridge University Press, 2018).

[41] M. M. Wilde, M. Berta, C. Hirche, and E. Kaur, "Amortized channel divergence for asymptotic quantum channel discrimination", Letters in Mathematical Physics **110** (2020).

[42] A. Uhlmann, "The "transition probability" in the state space of a *-algebra", Reports on Mathematical Physics **9**, 273 (1976).

[43] A. Gilchrist, N. K. Langford, and M. A. Nielsen, "Distance measures to compare real and ideal quantum processes", Physical Review A **71**, 062310 (2005).

[44] G. Gutoski, A. Rosmanis, and J. Sikora, "Fidelity of quantum strategies with applications to cryptography", Quantum **2**, 89 (2018).

[45] Killoran, Nathan, "Entanglement quantification and quantum benchmarking of optical communication devices", PhD thesis (University of Waterloo, 2012).

[46] J. Watrous, "Simpler semidefinite programs for completely bounded norms", Chicago Journal of Theoretical Computer Science **2013** (2013).

[47] G. Chiribella, G. M. D'Ariano, and P. Perinotti, "Memory effects in quantum channel discrimination", Physical Review Letters **101**, 180501 (2008).

[48] G. Chiribella, G. M. D'Ariano, and P. Perinotti, "Transforming quantum operations: Quantum supermaps", Europhysics Letters **83**, 30004 (2008).

[49] "Quantum Hypothesis Testing and Discrimination of Quantum States", in *Quantum Information: An Introduction* (Springer Berlin Heidelberg, Berlin, Heidelberg, 2006), pp. 69–91.

[50] F. Buscemi and N. Datta, "The Quantum Capacity of Channels With Arbitrarily Correlated Noise", IEEE Transactions on Information Theory **56**, 1447 (2010).

[51] F. G. S. L. Brandao and N. Datta, "One-Shot Rates for Entanglement Manipulation Under Non-entangling Maps", IEEE Transactions on Information Theory **57**, 1754 (2011).

[52] L. Wang and R. Renner, "One-Shot Classical-Quantum Capacity and Hypothesis Testing", Physical Review Letters **108**, 200501 (2012).

[53] A. S. Holevo, "Statistical decision theory for quantum systems", Journal of Multivariate Analysis **3**, 337 (1973).

[54] H. Yuen, R. Kennedy, and M. Lax, "Optimum testing of multiple hypotheses in quantum detection theory", IEEE Transactions on Information Theory **21**, 125 (1975).

[55] R. König, R. Renner, and C. Schaffner, "The Operational Meaning of Min- and Max-Entropy", IEEE Transactions on Information Theory **55**, 4337 (2009).

[56] N. Datta, "Min- and max-relative entropies and a new entanglement monotone", IEEE Transactions on Information Theory **55**, 2816 (2009).

[57] K. Fang, X. Wang, M. Tomamichel, and M. Berta, "Quantum channel simulation and the channel's smooth max-information", IEEE Transactions on Information Theory **66**, 2129 (2020).

[58] B. Regula, L. Lami, G. Ferrari, and R. Takagi, "Operational quantification of continuous-variable quantum resources", arXiv:2009.11302 (2020).