

# Unitary channel discrimination beyond group structures: Advantages of sequential and indefinite-causal-order strategies

Cite as: J. Math. Phys. 63, 042203 (2022); <https://doi.org/10.1063/5.0075919>

Submitted: 19 October 2021 • Accepted: 13 March 2022 • Published Online: 18 April 2022

 Jessica Bavaresco,  Mio Murao and  Marco Túlio Quintino



View Online



Export Citation



CrossMark



Journal of  
Mathematical Physics

YOUR RESEARCH Belongs Here

LEARN MORE



# Unitary channel discrimination beyond group structures: Advantages of sequential and indefinite-causal-order strategies

Cite as: *J. Math. Phys.* **63**, 042203 (2022); doi: [10.1063/5.0075919](https://doi.org/10.1063/5.0075919)

Submitted: 19 October 2021 • Accepted: 13 March 2022 •

Published Online: 18 April 2022






View Online



Export Citation



CrossMark

Jessica Bavaresco,<sup>1,a)</sup>  Mio Murao,<sup>2,3</sup>  and Marco Túlio Quintino<sup>1,4</sup> 

## AFFILIATIONS

<sup>1</sup>Institute for Quantum Optics and Quantum Information (IQOQI), Austrian Academy of Sciences, Boltzmanngasse 3, A-1090 Vienna, Austria

<sup>2</sup>Department of Physics, Graduate School of Science, The University of Tokyo, Hongo 7-3-1, Bunkyo-ku, Tokyo 113-0033, Japan

<sup>3</sup>Trans-scale Quantum Science Institute, The University of Tokyo, Hongo 7-3-1, Bunkyo-ku, Tokyo 113-0033, Japan

<sup>4</sup>Vienna Center for Quantum Science and Technology (VCQ), Faculty of Physics, University of Vienna, Boltzmanngasse 5, A-1090 Vienna, Austria

<sup>a)</sup> Author to whom correspondence should be addressed: [jessica.bavaresco@oeaw.ac.at](mailto:jessica.bavaresco@oeaw.ac.at)

## ABSTRACT

For minimum-error channel discrimination tasks that involve only unitary channels, we show that sequential strategies may outperform the parallel ones. Additionally, we show that general strategies that involve indefinite causal order are also advantageous for this task. However, for the task of discriminating a uniformly distributed set of unitary channels that forms a group, we show that parallel strategies are, indeed, optimal, even when compared to general strategies. We also show that strategies based on the quantum switch cannot outperform sequential strategies in the discrimination of unitary channels. Finally, we derive an absolute upper bound for the maximal probability of successfully discriminating any set of unitary channels with any number of copies for the most general strategies that are suitable for channel discrimination. Our bound is tight since it is saturated by sets of unitary channels forming a group  $k$ -design.

Published under an exclusive license by AIP Publishing. <https://doi.org/10.1063/5.0075919>

## I. INTRODUCTION

The discrimination of different hypotheses is a fundamental part of the scientific method that finds application in the most distinct areas, such as information theory,<sup>1</sup> bioinformatics,<sup>2</sup> machine learning,<sup>3</sup> and behavioral and social sciences.<sup>4</sup> In a discrimination task, one seeks for the best manner to decide whether a particular hypothesis is the most likely to be the best description of some scenario or experiment. An important, albeit general, instance of a discrimination task consists in identifying between different input–output relations or different causal-effect dynamics a physical system may undergo. For instance, in an x-ray examination that identifies whether a person has a broken bone, the examiner prepares an initial state, which is subjected to certain dynamics when passing through the person’s body, before being measured by some physical apparatus. In this case, the goal is to distinguish between the dynamics related to a broken and a healthy bone by implementing the most appropriate input state and measurement apparatus.

In its most fundamental level, closed-system dynamics in quantum theory are described by unitary operations. Hence, being able to discriminate between different unitary operations is a ubiquitous task within quantum theory and quantum technologies. Examples of tasks directly related to our ability to discriminate unitary operations are unitary equivalence determination,<sup>5,6</sup> quantum metrology,<sup>7,8</sup> quantum hypothesis testing,<sup>9</sup> quantum parameter estimation,<sup>10</sup> alignment and transmission of reference frames,<sup>11,12</sup> and discrimination and tomography of quantum circuit elements.<sup>13</sup>

Discrimination tasks are also relevant to the field of computer science. An oracle, which is an abstract machine used to study decision problems, may be understood as a black box that solves certain problems with a single operation. From a quantum computational perspective,

a quantum oracle is a unitary operation whose internal mechanisms are unknown and are employed in seminal quantum algorithms, such as the Deutsch–Jozsa algorithm,<sup>14</sup> Grover’s algorithm,<sup>15</sup> and Simon’s algorithm.<sup>16</sup> These oracle-based quantum algorithms may be recast as unitary discrimination tasks.<sup>17</sup>

Such practical and fundamental interest has motivated an extensive study of the discrimination of unitary channels within the context of quantum information theory, leading to a plethora of interesting results.

Contrarily to the problem of quantum state discrimination,<sup>18</sup> in which two states cannot be perfectly distinguished with a finite number of uses, or copies, unless they are orthogonal, it has been remarkably shown that any pair of unitary channels can indeed always be perfectly distinguished with a finite number of copies.<sup>19,20</sup> Moreover, perfect discrimination of a pair of unitary channels can always be achieved by a parallel scheme<sup>19,20</sup> (see also Ref. 21). Even when perfect discrimination is not possible, sequential strategies can never outperform parallel strategies in a task of discrimination between a pair of unitary channels.<sup>22</sup> Concerning the discrimination of sets of more than two unitary channels, when considering unitaries, which are a representation of a group and uniformly distributed, Ref. 22 showed once more that, for any number of copies, sequential strategies are not advantageous when compared to parallel strategies. For related tasks such as error-free and unambiguous unitary channel discrimination,<sup>23</sup> unitary estimation,<sup>22</sup> unitary learning,<sup>24</sup> and unitary store-and-retrieve,<sup>25</sup> parallel strategies were also proven to be optimal. Up to this point, no unitary channel minimum-error discrimination tasks in which sequential strategies outperform parallel strategies are known, to the extent of our knowledge.

In this work, we focus on the discrimination of sets of more than two unitary channels with multiple copies to study the potential advantages that different classes of strategies can bring to this task. The first contribution of our work is precisely to show examples of discrimination tasks of unitary channels in which sequential strategies are advantageous, when compared to parallel strategies that allow the same number of copies. In fact, contrarily to the tasks of error-free and unambiguous unitary channel discrimination,<sup>23</sup> we show that sequential strategies can achieve perfect discrimination in tasks that parallel strategies cannot.

Then, motivated by the recent advances in channel discrimination theory that have established the advantage of general discrimination strategies that involve indefinite causal order for general channels,<sup>26</sup> we study the potential advantages of these general strategies for the specific case of unitary channel discrimination.

Extending the framework developed in Ref. 26 to discrimination tasks that allow for the use of multiple copies, we achieve the following results. We prove the optimality of parallel strategies, even when compared against general strategies, in tasks of discrimination of uniformly distributed unitary channels that form a unitary representation of some group for any number of copies. However, the power of general strategies is revealed when applied to discrimination tasks that fail to satisfy at least one of these requirements. In these cases, we show that general indefinite-causal-order strategies can outperform sequential strategies and again that sequential strategies can outperform the parallel ones. Then, we show that a particular case of general strategies, one that applies processes related to the quantum switch<sup>27</sup> and its generalizations,<sup>28–30</sup> can never outperform sequential strategies in the discrimination of unitary channels.

The final contribution of our work is to derive an ultimate upper bound for the maximal probability of successful discrimination of any ensembles of unitary channels with a uniform probability distribution. Our result represents an upper bound for the most general strategy that can possibly be employed in a task of channel discrimination. We show that this bound is saturated by parallel strategies for the discrimination of unitary groups that form a  $k$ -design, where  $k$  is the number of allowed copies.

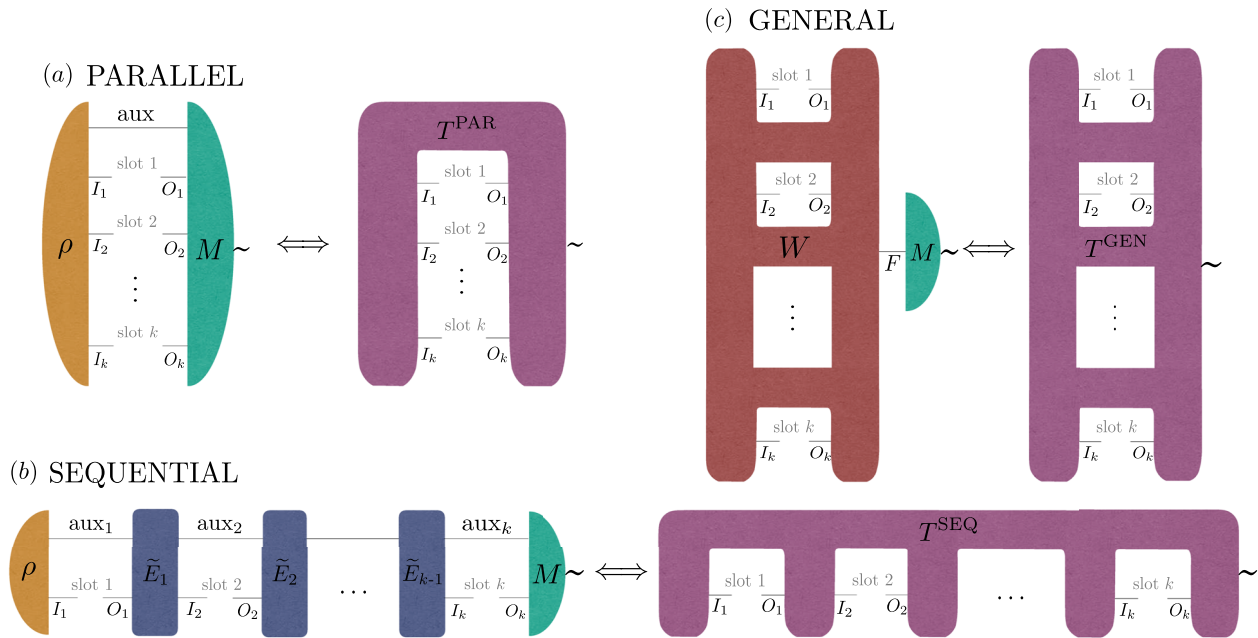
## II. MINIMUM-ERROR CHANNEL DISCRIMINATION

In a task of minimum-error channel discrimination, one is given access to an unknown quantum channel  $\tilde{C}_i : \mathcal{L}(\mathcal{H}^I) \rightarrow \mathcal{L}(\mathcal{H}^O)$ , which maps quantum states from an input linear space  $\mathcal{H}^I$  to an output linear space  $\mathcal{H}^O$ . This quantum channel is known to have been drawn with probability  $p_i$  from a known ensemble of channels  $\mathcal{E} = \{p_j, \tilde{C}_j\}_{j=1}^N$ . The task is to determine which channel from the ensemble was received using a limited amount of uses/queries of it, which is essentially to determine the classical label  $i$  of channel  $\tilde{C}_i$ . In order to accomplish this task in the case where only a single use of the unknown channel is allowed, one may send part of a potentially entangled state  $\rho \in \mathcal{L}(\mathcal{H}^I \otimes \mathcal{H}^{aux})$  through the channel  $\tilde{C}_i$  and subsequently jointly measure the output state with a positive-operator valued measure (POVM)  $M = \{M_a\}_{a=1}^N$ ,  $M_a \in \mathcal{L}(\mathcal{H}^O \otimes \mathcal{H}^{aux})$ . When both the state and measurement are optimized according to the knowledge of the ensemble, the outcome of the measurement will correspond to the most likely value of the label  $i$  of the unknown channel. Then, the maximal probability of successfully determining which channel is at hand is given by

$$P := \max_{\rho, \{M_i\}} \sum_{i=1}^N p_i \text{Tr}[(\tilde{C}_i \otimes \tilde{\mathbb{I}})(\rho) M_i], \quad (1)$$

where  $\tilde{\mathbb{I}} : \mathcal{L}(\mathcal{H}^{aux}) \rightarrow \mathcal{L}(\mathcal{H}^{aux})$  is the identity map.

When more than one use, or copy, as we will refer to from now on, is allowed, different strategies come into play, each exploring a different order in which the copies of the unknown channel are applied. In Figs. 1(a) and 1(b), we illustrate two such possibilities, a parallel and a sequential strategy, respectively. However, a more general strategy can be defined by considering the most general higher-order transformation that can map  $k$  quantum channels to a valid probability distribution [see Fig. 1(c)]. It has been shown that some of these general strategies may employ processes with an indefinite causal order and that these strategies may outperform parallel and sequential ones in tasks of channel discrimination.<sup>26</sup>



**FIG. 1.** Schematic representation of the realization of every  $k$ -copy (a) parallel tester  $T^{\text{PAR}}$  with a state  $\rho$  and a POVM  $M$ ; (b) sequential tester  $T^{\text{SEQ}}$  with a state  $\rho$ , channels  $\tilde{E}_i, i \in \{1, k - 1\}$ , and a POVM  $M$ ; and (c) general tester  $T^{\text{GEN}}$  with a process matrix  $W$  and a POVM  $M$ .

### III. TESTER FORMALISM

To facilitate the approach to this problem, a concise and unified formalism of testers,<sup>31</sup> also referred to as process POVMs,<sup>32,33</sup> was developed in Ref. 26, providing practical tools for both the comparison between different strategies and for the efficient computation of the maximal probability of successful discrimination of a channel ensemble under different classes of strategies. We now revise the tester formalism while extending its definitions to strategies that involve a finite number of copies  $k$  of the unknown channel.

In order to apply the tester formalism, we will make use of the Choi–Jamiołkowski (CJ) representation of quantum maps. The CJ isomorphism is a one-to-one correspondence between completely positive maps and positive semidefinite operators, which allows one to represent any linear map  $\tilde{L} : \mathcal{L}(\mathcal{H}^I) \mapsto \mathcal{L}(\mathcal{H}^O)$  by a linear operator  $L \in \mathcal{L}(\mathcal{H}^I \otimes \mathcal{H}^O)$  defined by

$$L := (\tilde{\mathbb{1}} \otimes \tilde{L})(\Phi^+), \tag{2}$$

where  $\Phi^+ = \sum_{ij} |ii\rangle\langle jj| \in \mathcal{L}(\mathcal{H}^I \otimes \mathcal{H}^I)$ , with  $\{|i\rangle\}$  being an orthonormal basis, is an unnormalized maximally entangled state. In this representation, a quantum channel, i.e., a completely positive trace-preserving (CPTP) map  $\tilde{C} : \mathcal{L}(\mathcal{H}^I) \mapsto \mathcal{L}(\mathcal{H}^O)$ , is represented by a linear operator  $C \in \mathcal{L}(\mathcal{H}^I \otimes \mathcal{H}^O)$ , often called the “Choi operator” of channel  $\tilde{C}$ , which satisfies

$$C \geq 0, \quad \text{Tr}_O C = \mathbb{1}^I, \tag{3}$$

where  $\text{Tr}_O$  denotes the partial trace over  $\mathcal{H}^O$  and  $\mathbb{1}^I$  denotes the identity operator on  $\mathcal{H}^I$ . In particular, the Choi operator of a unitary channel is proportional to a maximally entangled state. Using Choi operators of quantum channels, we can equivalently represent the channel ensemble  $\{p_i, \tilde{C}_i\}_{i=1}^N$  as  $\{p_i, C_i\}_{i=1}^N$ , where  $C_i$  is the Choi operator of channel  $\tilde{C}_i$ .

A tester is a set of positive semidefinite operators  $T = \{T_i\}_{i=1}^N$ ,  $T_i \in \mathcal{L}(\mathcal{H}^I \otimes \mathcal{H}^O)$ , which obey certain normalization constraints and which, when taking the trace with the Choi operator of a quantum channel  $C$ , lead to a valid probability distribution according to  $p(i|C) = \text{Tr}(T_i C)$ . In this sense, testers act on quantum channels similarly to how POVMs act on quantum states and can, therefore, be interpreted as a “measurement” of a quantum channel. Testers allow us to rewrite the maximal probability of successful discrimination of the channel ensemble  $\mathcal{E} = \{p_i, C_i\}_{i=1}^N$  as

$$P = \max_{\{T_i\}} \sum_{i=1}^N p_i \text{Tr}(T_i C_i). \tag{4}$$

The advantage of this representation is the simplification of the optimization problem that defines the maximal probability of success: now, optimization over different discrimination strategies may be achieved by maximizing  $P$  over the set of valid testers, as opposed to optimizing over both states and measurements [as in Eq. (1)].

For the case of  $k$  copies, different normalization constraints define testers that represent different classes of strategies. We now define  $k$ -copy testers that represent parallel, sequential, and general strategies.

Let  $\mathcal{H}^I := \otimes_{i=1}^k \mathcal{H}^{I_i}$  and  $\mathcal{H}^O := \otimes_{i=1}^k \mathcal{H}^{O_i}$  be the joint input and output spaces, respectively, of  $k$  copies of a quantum channel, let  $d_I := \dim(\mathcal{H}^I)$  and  $d_O := \dim(\mathcal{H}^O)$  be their dimension, and let

$$x(\cdot) := \text{Tr}_X(\cdot) \otimes \frac{\mathbb{1}^X}{d_X} \quad (5)$$

denote a trace-and-replace operation in  $\mathcal{H}^X$ .

### A. Parallel strategies

Parallel strategies are the ones that consist of sending each system that composes a multipartite state through one of the copies of the unknown channel, in such a way that the output of each copy does not interact with the input of the others, and jointly measuring the output state at the end [see Fig. 1(a), left]. These strategies are characterized by parallel testers [see Fig. 1(a), right]  $T^{\text{PAR}} = \{T_i^{\text{PAR}}\}_{i=1}^N$ ,  $T_i^{\text{PAR}} \in \mathcal{L}(\mathcal{H}^I \otimes \mathcal{H}^O)$ . Let  $W^{\text{PAR}} := \sum_i T_i^{\text{PAR}}$ . Then, parallel testers are defined as

$$T_i^{\text{PAR}} \geq 0 \quad \forall i, \quad (6)$$

$$\text{Tr} W^{\text{PAR}} = d_O, \quad (7)$$

$$W^{\text{PAR}} = {}_O W^{\text{PAR}}. \quad (8)$$

This is equivalent to defining  $W^{\text{PAR}}$  to be a parallel process, satisfying  $W^{\text{PAR}} = \sigma^I \otimes \mathbb{1}^O$ , where  $\text{Tr}(\sigma^I) = 1$ .

### B. Sequential strategies

In a sequential strategy, a quantum system is sent through the first copy of the channel, and its output system is allowed to be sent as input of the next copy, while general CPTP maps may act on the systems in between copies. The final output is measured by a POVM [see Fig. 1(b), left]. These strategies are represented by sequential testers [see Fig. 1(b), right]  $T^{\text{SEQ}} = \{T_i^{\text{SEQ}}\}_{i=1}^N$ ,  $T_i^{\text{SEQ}} \in \mathcal{L}(\mathcal{H}^I \otimes \mathcal{H}^O)$ . Let  $W^{\text{SEQ}} := \sum_i T_i^{\text{SEQ}}$ . Then, sequential testers are defined as

$$T_i^{\text{SEQ}} \geq 0 \quad \forall i, \quad (9)$$

$$\text{Tr} W^{\text{SEQ}} = d_O, \quad (10)$$

$$W^{\text{SEQ}} = {}_{O_k} W^{\text{SEQ}}, \quad (11)$$

$$I_k O_k W^{\text{SEQ}} = {}_{O_{(k-1)}} I_k O_k W^{\text{SEQ}}, \quad (12)$$

⋮

$$I_2 O_2 \dots I_k O_k W^{\text{SEQ}} = {}_{O_1 I_2 O_2 \dots I_k O_k} W^{\text{SEQ}}. \quad (13)$$

This is equivalent to defining  $W^{\text{SEQ}}$  to be a  $k$ -slot comb<sup>31</sup> or a  $k$ -partite ordered process matrix<sup>34</sup> (see also Refs. 35 and 36).

### C. General strategies

Finally, general strategies are defined by the most general higher-order operations that can transform  $k$  quantum channels into a joint probability distribution. They can be regarded as the most general “measurement” that acts jointly on  $k$  quantum channels, yielding a classical output. Crucially, and contrarily to parallel and sequential strategies, general strategies do not physically impose any particular order in which the copies of the channel must be applied. The testers that characterize these strategies, the general testers, are the most general sets of positive semidefinite operators  $T^{\text{GEN}} = \{T_i^{\text{GEN}}\}_{i=1}^N$ ,  $T_i^{\text{GEN}} \in \mathcal{L}(\mathcal{H}^I \otimes \mathcal{H}^O)$  that satisfy  $p(i|C_1, \dots, C_k)_{\text{break}} = \text{Tr}[T_i^{\text{GEN}}(C_1 \otimes \dots \otimes C_k)]$ . Let  $W^{\text{GEN}} := \sum_i T_i^{\text{GEN}}$ . Then, general testers can be equivalently defined as

$$T_i^{\text{GEN}} \geq 0 \quad \forall i, \quad (14)$$

$$\text{Tr}[W^{\text{GEN}}(C_1 \otimes \dots \otimes C_k)] = 1 \quad (15)$$

for all Choi states of quantum channels  $C_i \in \mathcal{L}(\mathcal{H}^I \otimes \mathcal{H}^O)$ . This is equivalent to defining  $W^{\text{GEN}}$  to be a general  $k$ -partite process matrix.<sup>34,37</sup> For the cases of  $k = 2$  and  $k = 3$ , we provide a characterization of  $W^{\text{GEN}}$  in terms of linear constraints in [Appendix A](#).

Ordered processes, such as  $W^{\text{PAR}}$  and  $W^{\text{SEQ}}$ , form a subset of the general processes  $W^{\text{GEN}}$ . However, some general processes are known not to respect the causal constraints of ordered processes, that is, they are neither parallel nor sequential.<sup>37</sup> Moreover, they cannot be described as convex mixtures of ordered processes, in the bipartite case, or other more appropriate notions of mixtures of causal order, in the multipartite case.<sup>38</sup> Such general processes have been termed to exhibit an *indefinite causal order*<sup>37</sup> and have been shown to bring advantages to several quantum information tasks, such as quantum computation,<sup>28</sup> communication complexity,<sup>39,40</sup> discrimination of bipartite non-signaling channels,<sup>41</sup> violation of fully and semi-device-independent causal inequalities,<sup>37,42,43</sup> inversion of unknown unitary operations,<sup>44</sup> and, more recently, discrimination of general quantum channels.<sup>26</sup>

Contrarily to the parallel and sequential cases, a realization of general testers in terms of quantum operations is an open problem. More specifically, a general tester can always be constructed from a general process matrix and a POVM,<sup>26</sup> as illustrated in [Fig. 1\(c\)](#). However, only a subset of process matrices are currently known to be realizable with quantum operations.<sup>45</sup> This subset, known as “coherent quantum control of causal orders,” has been shown to bring advantage to the discrimination of general channels.<sup>26</sup> See [Ref. 26](#) for a more detailed discussion about the realization of testers.

For any chosen strategy, the maximal probability of successful discrimination of an ensemble of  $N$  channels  $\mathcal{E} = \{p_i, C_i\}_{i=1}^N$  using  $k$  copies is given by

$$P^{\mathcal{S}} := \max_{\{T_i^{\mathcal{S}}\}_{i=1}^N} \sum_{i=1}^N p_i \text{Tr}(T_i^{\mathcal{S}} C_i^{\otimes k}), \quad (16)$$

where  $\mathcal{S} \in \{\text{PAR}, \text{SEQ}, \text{GEN}\}$ .

For all three classes of strategies,  $P^{\mathcal{S}}$  can be computed via semidefinite programming (SDP). A short review of dual and primal SDP problems associated with the maximal probability of successful discrimination is given in [Appendix A](#).

#### IV. DISCRIMINATION OF UNITARY CHANNELS

In this section, we present our results concerning the discrimination of unitary channels. Here, unitary channels, also called unitary operations, will be simply denoted by unitary operators  $U$  that satisfy  $UU^\dagger = \mathbb{1}$ , and these terms will be used interchangeably. The Choi operators of unitary channels will be denoted as  $|U\rangle\rangle \langle\langle U| \in \mathcal{L}(\mathcal{H}^I \otimes \mathcal{H}^O)$ , where

$$|U\rangle\rangle := \sum_i (\mathbb{1} \otimes U)|ii\rangle. \quad (17)$$

We start considering a discrimination task that involves unitary channels that form a group (see [Appendix B](#)). We show that, for an ensemble composed of a set of unitary channels that forms a group and a uniform distribution, parallel strategies do not only perform as well as the sequential ones<sup>22</sup> but are, indeed, the optimal strategies for discrimination—even considering general strategies that may involve indefinite causal order.

**Theorem 1.** *For ensembles composed of a uniform probability distribution and a set of unitary channels that forms a group up to a global phase, in discrimination tasks that allow for  $k$  copies, parallel strategies are optimal, even when considering general strategies.*

*More specifically, let  $\mathcal{E} = \{p_i, U_i\}_i$  be an ensemble with  $N$  unitary channels, where  $p_i = \frac{1}{N} \forall i$  and the set  $\{U_i\}_i$  forms a group up to a global phase. Then, for any number of copies  $k$  and for every general tester  $\{T_i^{\text{GEN}}\}$ , there exists a parallel tester  $\{T_i^{\text{PAR}}\}_i$  such that*

$$\frac{1}{N} \sum_{i=1}^N \text{Tr}(T_i^{\text{PAR}} |U_i\rangle\rangle \langle\langle U_i|^{\otimes k}) = \frac{1}{N} \sum_{i=1}^N \text{Tr}(T_i^{\text{GEN}} |U_i\rangle\rangle \langle\langle U_i|^{\otimes k}). \quad (18)$$

The proof of this theorem can be found in [Appendix B](#).

Theorem 1 has two crucial hypotheses: (1) the set of unitary operators  $\{U_i\}$  forms a group and (2) the distribution  $\{p_i\}$  is uniform. If at least one of these hypotheses is not satisfied, then Theorem 1, in fact, does not hold, as we show in the following.

**Theorem 2.** *There exist ensembles of unitary channels for which sequential strategies of discrimination outperform parallel strategies. Moreover, sequential strategies can achieve perfect discrimination in some scenarios where the maximal probability of success of parallel strategies is strictly less than one.*

Let us start with the case where the set of unitary channels does not form a group, but the probability distribution of the ensemble is uniform. In the following,  $\sigma_x$ ,  $\sigma_y$ , and  $\sigma_z$  are the Pauli operators and  $H := |+\rangle\langle 0| + |-\rangle\langle 1|$ , where  $|\pm\rangle := \frac{1}{\sqrt{2}}(|0\rangle \pm |1\rangle)$ , is the Hadamard gate.

*Example 1.* The ensemble composed of a uniform probability distribution and  $N = 4$  qubit unitary channels given by  $\{U_i\} = \{1, \sqrt{\sigma_x}, \sqrt{\sigma_y}, \sqrt{\sigma_z}\}$ , in a discrimination task that allows for  $k = 2$  copies, can be discriminated under a sequential strategy with the probability of success  $P^{\text{SEQ}} = 1$ , while any parallel strategy yields  $P^{\text{PAR}} < 1$ .

A straightforward sequential strategy that attains perfect discrimination of this ensemble can be constructed by first noting that  $\sqrt{\sigma_i}\sqrt{\sigma_i} = \sigma_i$ ; hence, a simple composition of the unitary operators  $U_i$  leads to the Pauli operators, which are perfectly discriminated with a bipartite maximally entangled state and a joint measurement in the Bell basis. The proof that any parallel strategy that applies two copies can never attain perfect discrimination is provided in [Appendix C](#) and applies the method of computer-assisted proofs developed in [Ref. 26](#).

Another example of this phenomenon is shown by the ensemble  $\{U_i\} = \{1, \sigma_x, \sigma_y, \sqrt{\sigma_z}\}$  with uniform probability distribution, which also satisfies  $P^{\text{PAR}} < P^{\text{SEQ}} = 1$  for a discrimination task with  $k = 2$  copies. In [Appendix C](#), we show that such an ensemble can actually be discriminated perfectly by a sequential strategy that uses, on average, 1.5 copies.

The next example concerns a set of unitary channels that forms a group, but the probability distribution of the ensemble is not uniform.

*Example 2.* Let  $\{1, \sigma_x, \sigma_y, \sigma_z, H, \sigma_x H, \sigma_y H, \sigma_z H\} = \{U_i\}$  be a tuple of  $N = 8$  unitary channels that forms a group up to a global phase, and let  $\{p_i\}$  be a probability distribution in which each element  $p_i$  is proportional to the  $i$ th digit of the number  $\pi \approx 3.141\,592\,6$ , that is,  $\{p_i\} = \{\frac{3}{31}, \frac{1}{31}, \frac{4}{31}, \dots, \frac{6}{31}\}$ . For the ensemble  $\{p_i, U_i\}$ , in a discrimination task that allows for  $k = 2$  copies, sequential strategies outperform parallel strategies, i.e.,  $P^{\text{PAR}} < P^{\text{SEQ}}$ .

The proof of this example is also in [Appendix C](#) and applies the same method of computer-assisted proofs. In [Example 2](#), we have set the distribution  $\{p_i\}$  to be proportional to the  $i$ th digit of the constant  $\pi$  to emphasize that the phenomenon of sequential strategies outperforming the parallel ones when the set of unitary channels forms a group does not require a particularly well chosen non-uniform distribution. In practice, we have observed that with randomly generated distributions, optimal strategies often respect  $P^{\text{PAR}} < P^{\text{SEQ}}$ .

In both of the aforementioned examples, general strategies do not outperform sequential strategies. However, for the case of discrimination of unitary channels using  $k = 3$  copies, we show that general strategies are, indeed, advantageous.

**Theorem 3.** *There exist ensembles of unitary channels for which general strategies of discrimination outperform sequential strategies.*

Let us start again with the case where the set of unitary channels does not form a group, but the probability distribution of the ensemble is uniform. For the following, we define  $H_y := |+\rangle\langle 0| + |-\rangle\langle 1|$ , where  $|\pm\rangle := \frac{1}{\sqrt{2}}(|0\rangle \pm |1\rangle)$ , and  $H_p := |+\rangle\langle 0| + |-\rangle\langle 1|$ , where  $|\pm\rangle := \frac{1}{5}(3|0\rangle + 4|1\rangle)$  and  $|-\rangle := \frac{1}{5}(4|0\rangle - 3|1\rangle)$ .

*Example 3.* For the ensemble composed of a uniform probability distribution and  $N = 4$  unitary channels given by  $\{U_i\} = \{\sqrt{\sigma_x}, \sqrt{\sigma_z}, \sqrt{H_y}, \sqrt{H_p}\}$ , in a discrimination task that allows for  $k = 3$  copies, general strategies outperform sequential strategies and sequential strategies outperform parallel strategies. Therefore, the maximal probabilities of success satisfy the strict hierarchy  $P^{\text{PAR}} < P^{\text{SEQ}} < P^{\text{GEN}}$ .

The proof of this example can be found in [Appendix D](#).

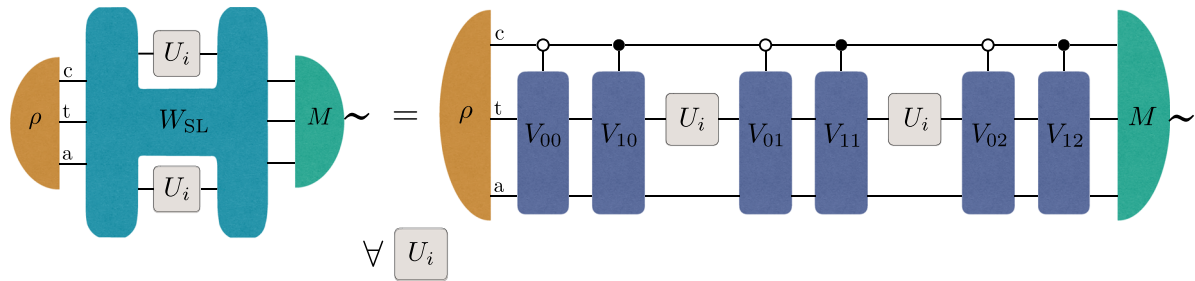
General strategies can also be advantageous for the discrimination of an ensemble composed of a non-uniform probability distribution and a set of unitary channels that forms a group. Let the set of unitary operators in [Example 3](#) be the set of generators of a group (potentially with an infinite number of elements). Now, consider the ensemble composed of such a group and a probability distribution given by  $p_i = \frac{1}{4}$  for the four values of  $i$  corresponding to the four unitary operators, which are the generators of the group, and  $p_i = 0$  otherwise. It is straightforward to see that the maximal probabilities of successfully discriminating this ensemble would be the same as the ones in [Example 3](#), hence satisfying  $P^{\text{PAR}} < P^{\text{SEQ}} < P^{\text{GEN}}$ . Although somewhat artificial, this example shows that advantages of general strategies are, indeed, possible for this kind of unitary channel ensemble.

Although general indefinite-causal-order strategies can be advantageous for the discrimination of unitary channels, this is not the case for one particular sub-class of general strategies: those which can be constructed from the quantum switch.<sup>27</sup>

Let  $V_{mn}$ , with  $m \in \{0, 1\}$ ,  $n \in \{0, 1, 2\}$ , be unitary operators that act on a target and an auxiliary system and  $U_1, U_2$  be unitary operators that act only on the target system. Finally, let  $\{|m\rangle\langle m|^c\}_m$  be projectors that act on a control system. Then, we define the *switch-like* superchannel, which transforms a pair of unitary channels into one unitary channel, according to

$$\mathcal{W}_{\text{SL}}(U_1, U_2) := |0\rangle\langle 0|^c \otimes V_{02} (U_2 \otimes \mathbb{1}) V_{01} (U_1 \otimes \mathbb{1}) V_{00} + |1\rangle\langle 1|^c \otimes V_{12} (U_1 \otimes \mathbb{1}) V_{11} (U_2 \otimes \mathbb{1}) V_{10}, \quad (19)$$

where  $\mathbb{1}$  is the identity operator acting on the auxiliary system. In the case where  $V_{mn} = \mathbb{1} \forall m, n$ , one recovers the standard quantum switch.<sup>27</sup> The switch-like superchannel has been previously considered in [Refs. 29 and 30](#) in the context of reversibility-preserving transformations. Generalizations of the switch-like superchannel that transform  $k$  instead of 2 unitaries are presented in detail in [Appendix E](#), applying unitaries



**FIG. 2.** Illustration of a two-copy sequential strategy, which attains the same probability of successful discrimination as any two-copy switch-like strategy, for all sets of unitary channels  $\{U_i\}_{i=1}^N$ . Line “c” represents a control system, “t” represents a target system, and “a” represents an auxiliary system. Both strategies can be straightforwardly extended to  $k$  copies (see Appendix E).

$\{V_{mn}\}_{m,n}$ , with  $m \in \{0, \dots, k! - 1\}$  and  $n \in \{0, \dots, k\}$ , and considering all permutations of the target unitaries  $\{U_i\}_{i=1}^k$ . Such  $k$ -slots switch-like superchannels have been shown to be implementable via coherent quantum control of causal orders.<sup>45</sup>

Now, let  $W^{\text{SL}} \in \mathcal{L}(\mathcal{H}^p \otimes \mathcal{H}^t \otimes \mathcal{H}^a)$  be the  $k$ -slot switch-like process associated with the  $k$ -slot generalization of the switch-like superchannel in Eq. (19). A general discrimination strategy, given by the  $k$ -copy switch-like tester  $T^{\text{SL}} = \{T_i^{\text{SL}}\}$ ,  $T_i^{\text{SL}} \in \mathcal{L}(\mathcal{H}^t \otimes \mathcal{H}^a)$ , can be constructed using the  $k$ -slot switch-like process  $W^{\text{SL}}$ , a quantum state  $\rho \in \mathcal{L}(\mathcal{H}^p)$  that acts on the “past” space of the  $k$ -slots of  $W^{\text{SL}}$ , and a POVM  $\{M_i\}$ ,  $M_i \in \mathcal{L}(\mathcal{H}^t \otimes \mathcal{H}^a)$  that acts on the “future” space according to

$$T_i^{\text{SL}} := \text{Tr}_{PF}[(\rho \otimes \mathbb{1})W^{\text{SL}}(\mathbb{1} \otimes M_i)], \quad (20)$$

where the identity operators  $\mathbb{1}$  act on the correspondent complementary spaces.

We show that such switch-like strategies exhibit no advantage over sequential strategies for the discrimination of  $N$  unitary channels using  $k$  copies.

**Theorem 4.** *The action of the switch-like process on  $k$  copies of a unitary channel can be equivalently described by a sequential process that acts on  $k$  copies of the same unitary channel.*

Consequently, in a discrimination task involving the ensemble  $\mathcal{E} = \{p_i, U_i\}_i$  composed of  $N$  unitary channels and some probability distribution and that allows for  $k$  copies, for every switch-like tester  $\{T_i^{\text{SL}}\}_i$ , there exists a sequential tester  $\{T_i^{\text{SEQ}}\}_i$  that attains the same probability of success according to

$$\sum_{i=1}^N p_i \text{Tr}(T_i^{\text{SL}} |U_i\rangle\langle\langle U_i|^{\otimes k}) = \sum_{i=1}^N p_i \text{Tr}(T_i^{\text{SEQ}} |U_i\rangle\langle\langle U_i|^{\otimes k}). \quad (21)$$

The proof can be found in Appendix E, where we provide a simple construction of a sequential strategy that performs as well as any switch-like strategy using the same number of copies for unitary channel discrimination. A graphical representation of such a strategy in the case of  $k = 2$  copies is represented in Fig. 2.

## V. UNIFORMLY SAMPLED UNITARY CHANNELS

The advantage of sequential and general strategies in the discrimination of unitary channels is not restricted to the main examples given for Theorems 2 and 3. In fact, by sampling sets of unitary operators uniformly distributed according to the Haar measure and using these sets to construct ensembles with uniform probability distribution, one can find several other examples of the advantage of sequential and general strategies. A summary of our numerical findings is presented in Table I.

**Qubits,  $k = 2$ .** In the scenario of qubit unitary channels and  $k = 2$  copies, we have observed gaps between the performance of parallel and sequential strategies for ensembles of  $N \in \{4, \dots, 25\}$  unitary channels. For the case of  $N = 2$ , it is known that such gaps do not exist,<sup>22</sup> and for the case of  $N = 3$ , no gap was discovered. By calculating the averages of the maximal probabilities of success, we observed that the minimum ratio  $\langle P^{\text{PAR}} \rangle / \langle P^{\text{SEQ}} \rangle$  occurred at  $N = 6$ . At  $N = 25$ , gaps were hardly detected. This behavior can be visualized in the plot of Fig. 3 (top).

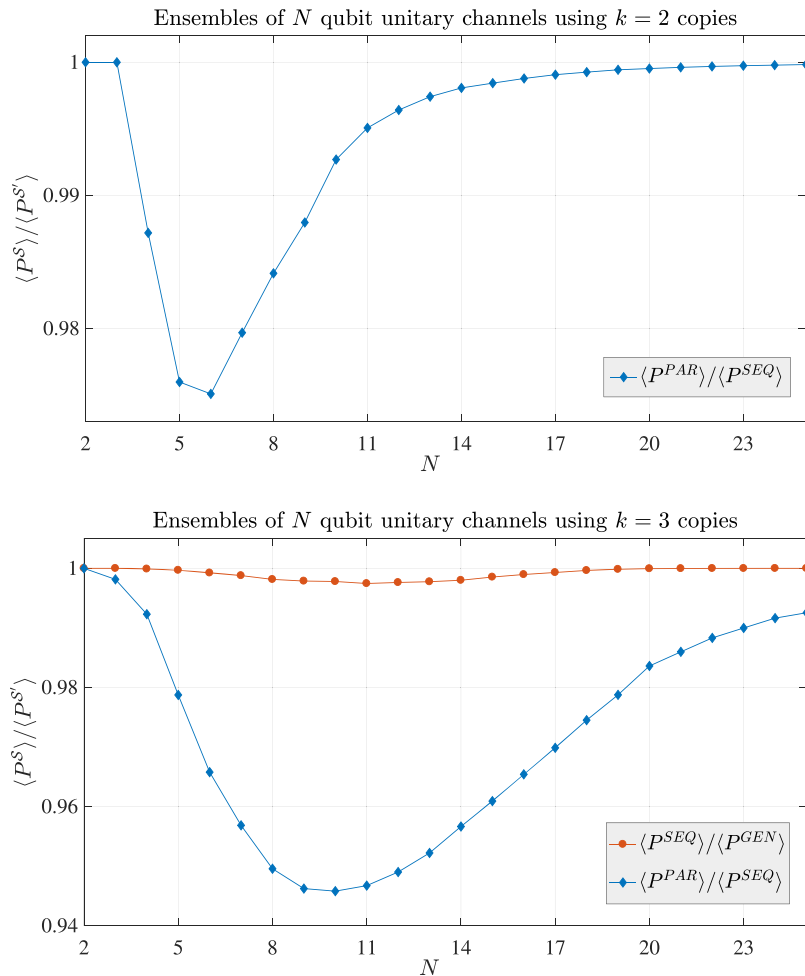
**Qutrits,  $k = 2$ .** For the case of qutrit unitary channels and  $k = 2$  copies, we discovered a gap between the performance of parallel and sequential strategies already for a discrimination task of only  $N = 3$  unitary channels, while in the qubit case, the first example of this phenomenon was found only for  $N = 4$ . In this scenario as well, no advantages of general strategies were found. These results are not plotted.

**Qubits,  $k = 3$ .** In the scenario of qubit unitary channels and  $k = 3$  copies, the advantage of general strategies over causally ordered ones (parallel and sequential) is common. Still considering uniformly sampled qubit unitary channels, in the three-copy case, we have found a strict hierarchy of discrimination strategies in scenarios of  $N \in \{4, \dots, 19\}$ . For  $N \in \{20, \dots, 25\}$ , the advantage of sequential strategies was clear



**TABLE I.** Summary of numerical findings: Gaps between different strategies of discrimination using  $k$  copies of ensembles of  $N$  uniformly distributed qubit unitary channels sampled according to the Haar measure. The bold equalities on row  $N = 2$  denote known analytical results (see Ref. 22). A strict inequality indicates that examples of ensembles that exhibit such gaps were encountered. An equality indicates that, for all sampled ensembles, no gap was encountered, up to numerical precision. The number of sampled sets ranged from 500 to 50 000.

Uniformly sampling qubit unitary channels		
$N$	$k = 2$	$k = 3$
2	$\mathbf{P}^{\text{PAR}} = \mathbf{P}^{\text{SEQ}} = \mathbf{P}^{\text{GEN}}$	$\mathbf{P}^{\text{PAR}} = \mathbf{P}^{\text{SEQ}} = \mathbf{P}^{\text{GEN}}$
3	$P^{\text{PAR}} = P^{\text{SEQ}} = P^{\text{GEN}}$	$P^{\text{PAR}} < P^{\text{SEQ}} = P^{\text{GEN}}$
4	$P^{\text{PAR}} < P^{\text{SEQ}} = P^{\text{GEN}}$	$P^{\text{PAR}} < P^{\text{SEQ}} < P^{\text{GEN}}$
$\vdots$	$\vdots$	$\vdots$
10	$P^{\text{PAR}} < P^{\text{SEQ}} = P^{\text{GEN}}$	$P^{\text{PAR}} < P^{\text{SEQ}} < P^{\text{GEN}}$
$\vdots$	$\vdots$	$\vdots$
25	$P^{\text{PAR}} \approx P^{\text{SEQ}} = P^{\text{GEN}}$	$P^{\text{PAR}} < P^{\text{SEQ}} \approx P^{\text{GEN}}$



**FIG. 3.** Ratios of the averages of the maximal probability of successful discrimination of ensembles of  $N \in \{2, \dots, 25\}$  uniformly distributed qubit unitary channels sampled according to the Haar measure. Top: For  $k = 2$  copies, the ratio between parallel and sequential strategies  $\langle P^{\text{PAR}} \rangle / \langle P^{\text{SEQ}} \rangle$ . To compute the averages, 1000 sets were sampled for each  $N$ . Bottom: For  $k = 3$  copies, the ratio between parallel and sequential strategies  $\langle P^{\text{PAR}} \rangle / \langle P^{\text{SEQ}} \rangle$  and between sequential and general strategies  $\langle P^{\text{SEQ}} \rangle / \langle P^{\text{GEN}} \rangle$ . To compute the averages, 500 sets were sampled for each  $N$ .

but that of general strategies was hardly detected. In addition, in the case of  $N = 3$ , only an advantage of sequential over parallel strategies was found. The minimum ratios of the averages  $\langle P^{\text{PAR}} \rangle / \langle P^{\text{SEQ}} \rangle$  and  $\langle P^{\text{SEQ}} \rangle / \langle P^{\text{GEN}} \rangle$  were found at  $N = 10$  and  $N = 11$ , respectively. These results are plotted in Fig. 3 (bottom).

The number of unitary channel ensembles sampled to calculate the average values of the maximal probabilities of success ranged from 50 000 for  $d = 2$ ,  $k = 2$ , and  $N = 2$  to 100 for  $d = 3$ ,  $k = 2$ , and  $N = 3$ .

In Fig. 3, one can see how the ratios between the maximal probabilities of success of different classes of strategies decrease as the number  $N$  of uniformly sampled unitary channels being discriminated increases. This observation is in line with the idea that, in the limit where the ensemble is composed of all qubit unitary channels, therefore forming the group  $SU(2)$ , it is expected that parallel strategies would be optimal. In Sec. VI, we formally analyze the asymptotic behavior of  $\langle P^S \rangle$ , while providing an absolute upper bound for the maximal probability of success under any strategy.

## VI. ULTIMATE UPPER BOUND FOR $P^S$ FOR ANY SET OF UNITARY CHANNELS

We now present an upper bound for the maximal probability of success for discriminating a set of  $N$   $d$ -dimensional unitary channels with general strategies when  $k$  copies are available. Our result applies to any ensemble of unitary channels  $\mathcal{E} = \{p_i, U_i\}_{i=1}^N$ , where  $p_i = \frac{1}{N}$  is a uniform probability distribution. Since general testers are the most general strategies, which are consistent with a channel discrimination task, our result constitutes an ultimate upper bound for discriminating uniformly distributed unitary channels. In addition, as we show later, our bound can be saturated by particular choices of unitary channels.

**Theorem 5** (upper bound for general strategies). *Let  $\mathcal{E} = \{p_i, U_i\}_{i=1}^N$  be an ensemble composed of  $N$   $d$ -dimensional unitary channels and a uniform probability distribution. The maximal probability of successful discrimination of a general strategy with  $k$  copies is upper bounded by*

$$P^{\text{GEN}} \leq \frac{1}{N} \gamma(d, k), \quad (22)$$

where  $\gamma(d, k)$  is given by

$$\gamma(d, k) := \binom{k + d^2 - 1}{k} = \frac{(k + d^2 - 1)!}{k!(d^2 - 1)!}. \quad (23)$$

The proof of this theorem is in Appendix F.

The upper bound given by Theorem 5 can be attained by some particular choices of unitary channels that form a group (and hence, by Theorem 3, attainable by a parallel strategy). In particular, it follows from Refs. 46–48 that when sets of unitary operators  $\{U_i\}_i$  form group  $k$ -designs, there exists a parallel strategy such that

$$P^{\text{PAR}} = \frac{1}{N} \gamma(d, k). \quad (24)$$

A set of unitary operators  $\{U_i\}_{i=1}^N$  forms a group  $k$ -design when the set forms a group and respects the relation

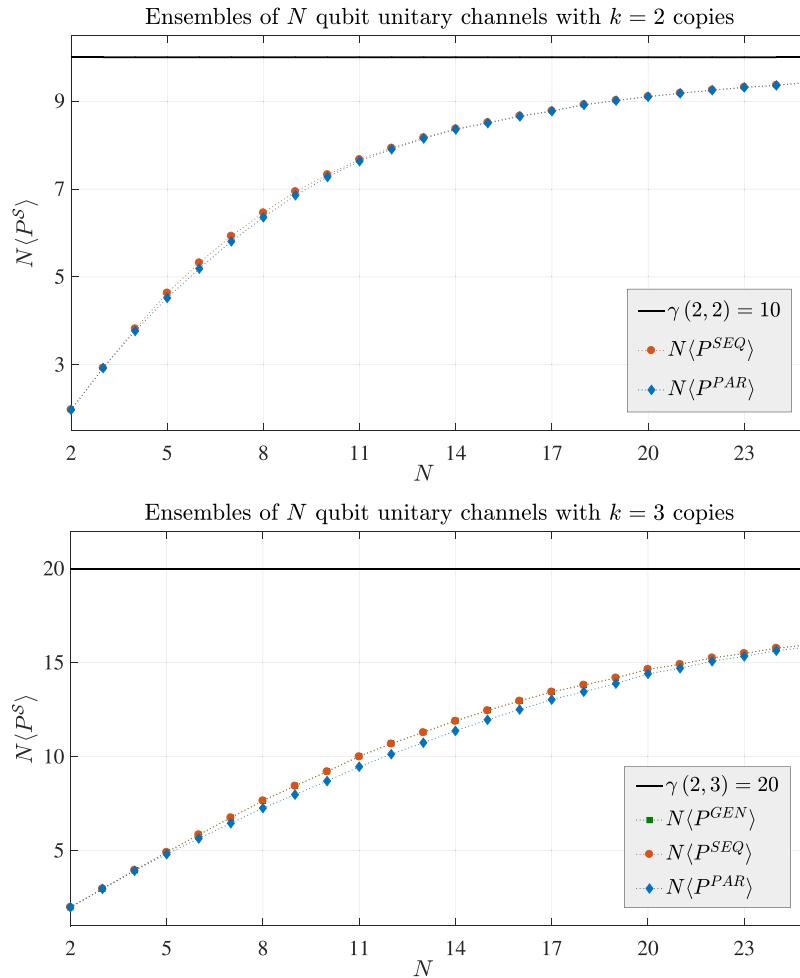
$$\int_{\text{Haar}} U^{\otimes k} \rho U^{\otimes k \dagger} dU = \frac{1}{N} \sum_{i=1}^N U_i^{\otimes k} \rho U_i^{\otimes k \dagger} \quad (25)$$

for every linear operator  $\rho$ .

In Sec. V, we analyzed the problem of discriminating  $N$  unitary operators uniformly sampled according to the Haar measure. For any fixed dimension  $d$  and number of copies  $k$ , if we sample a very large number  $N$  of unitary operators ( $N \rightarrow \infty$ ), when the distribution  $\{p_i\}_i$  is uniform, the maximal probability of success will be very small ( $P^S \rightarrow 0$ ). At the same time, due to the uniform properties of the Haar measure, if we uniformly sample a large number  $N$  of unitary operators, we obtain a set that approaches the group of all unitary operators of the given dimension, which is equivalent to the group  $SU(d)$ , approximately satisfying the group  $k$ -design condition in Eq. (25). Therefore, for large  $N$ , we have the asymptotic behavior

$$N \langle P^{\text{PAR}} \rangle \approx N \langle P^{\text{SEQ}} \rangle \approx N \langle P^{\text{GEN}} \rangle \approx \gamma(d, k). \quad (26)$$

This argument can be made more rigorous by recognizing that  $SU(d)$  is a compact Lie group, treating probabilities as probability densities, and maximizing likelihood.<sup>46,48</sup>



**FIG. 4.** Upper bound and average of the maximal probability of successful discrimination of sets of  $N \in \{2, \dots, 25\}$  qubit unitary operators sampled according to the Haar measure. The plotted curves correspond to the average  $\langle P^S \rangle$  multiplied by  $N$ , and the constant line is the analytical upper bound of  $\gamma(2, k)$  derived in Theorem 5. Note how all  $N\langle P^S \rangle$  asymptotically approach the upper bound  $\gamma(2, k)$  as  $N$  increases, although somewhat slowly. Top: For  $k = 2$  copies,  $N\langle P^{\text{PAR}} \rangle$  and  $N\langle P^{\text{SEQ}} \rangle$  are plotted against the upper bound of  $\gamma(2, 2) = 10$ . To compute the averages, 1000 sets were sampled for each  $N$ . Bottom: For  $k = 3$  copies,  $N\langle P^{\text{PAR}} \rangle$ ,  $N\langle P^{\text{SEQ}} \rangle$ , and  $N\langle P^{\text{GEN}} \rangle$  are plotted against the upper bound of  $\gamma(2, 3) = 20$ . The difference between plotted points of  $N\langle P^{\text{SEQ}} \rangle$  and  $N\langle P^{\text{GEN}} \rangle$  is not visible. To compute the averages, 500 sets were sampled for each  $N$ .

As an example from our numerical calculations, in Fig. 4, for the case of  $d = 2$  and  $k = 2, 3$ , we can visualize the behavior of  $N\langle P^S \rangle$  as a function of  $N$  and how it asymptotically approaches the constant  $\gamma(d, k)$ , which takes the values of  $\gamma(2, 2) = 10$  and  $\gamma(2, 3) = 20$ , with increasing  $N$ .

## VII. CONCLUSION

We extended the unified tester formalism of Ref. 26 to the case of  $k$  copies and applied it particularly to the study of unitary channel discrimination. Our first contribution was to prove that, in a discrimination task of an ensemble of a set of unitary channels that forms a group and a uniform probability distribution, parallel strategies are always optimal, even when compared against the performance of general strategies.

Subsequently, we showed an example of a unitary channel discrimination task in which a sequential strategy outperforms any parallel strategy. Our result consists of the first demonstration of such a phenomenon, to the extent of our knowledge. Our task involves four unitary

channels and two copies, which can be perfectly discriminated with a sequential strategy but not with a parallel one. We explicitly provided the optimal discrimination strategy for this task.

We also showed that general strategies that involve indefinite causal order are advantageous for the discrimination of unitary channels. Our simplest example of this phenomenon is a task of discriminating among four unitary channels using three copies. While our optimal parallel and sequential strategies can be straightforwardly implemented with (ordered) quantum circuits, a potential quantum realization of the optimal general strategies that are advantageous in this scenario remains an open problem. We then demonstrated that general strategies that are created from switch-like transformations, which are known to be, in principle, implementable with quantum operations,<sup>45</sup> can never perform better than sequential strategies for unitary channel discrimination.

The final result of our work was an upper bound for the maximal probability of success of any set of unitary channels using any number of copies under any strategy. This ultimate bound applies to any possible discrimination strategy and was shown to be tight, attained by discrimination tasks of unitary groups that form a  $k$ -design.

The concrete examples of the advantages of sequential and general strategies in this work focused on discrimination tasks that use  $k = 2$  or 3 copies. An open question of our work is how these strategy gaps would scale with larger values of  $k$ . The preliminary results presented here indicate that the advantage of sequential over parallel strategies and of general over sequential strategies should be even more accentuated as a higher number of copies are allowed. This idea is supported by the intuition that the number of different ways in which one can construct sequential strategies, as compared to parallel strategies, increases with the number of slots  $k$ . Similarly, we expect such a phenomenon to exist for the general case. It would then be interesting to find out exactly the rate with which these gaps increase with  $k$ .

No advantage of general strategies was found in scenarios involving discrimination of unitary channels using only  $k = 2$  copies. We conjecture that, when considering  $k = 2$  copies, such an advantage is, indeed, not possible for any number  $N$  of unitary channels. We also remark that, when considering  $k = 2$  copies, Refs. 29 and 30 proved that superchannels that preserve reversibility (i.e., transform unitary channels into unitary channels) are necessarily of the switch-like form. Intuitively, it seems plausible that the optimal general strategy for discriminating unitary channels would be one that transforms unitary channels into unitary channels. This argument of reversibility preservation combined with our Theorem 4 might lead to a proof of our conjecture.

Furthermore, we also conjecture that, when considering  $N = 2$  unitary channels, general strategies are not advantageous for any number of copies  $k$ . In this scenario, it has been proven that sequential strategies cannot outperform parallel ones,<sup>22</sup> and we believe this to also be the case for general strategies. The task of discriminating between two unitary channels can always be recast as the problem of discriminating a unitary operator from the identity operator. In the parallel case, the probability of successful discrimination has been shown to be related to the spread of the eigenvalues of this unitary operator.<sup>19,20</sup> The proof presented in Ref. 22 explores how sequential strategies affect the spread of the eigenvalues of unitary operators, to conclude that they cannot outperform the parallel ones. A better understanding of how general strategies affect the spread of the eigenvalues of unitary operators could lead to a conclusive answer for this conjecture.

Finally, an interesting open question that could follow our work is how quantum-specific are the phenomena presented here. For instance, Ref. 49 showed that adaptive strategies may outperform the parallel ones in classical channel discrimination; however, their examples did not concern a classical analog of unitary channels. Assuming the analog of a unitary channel in classical information theory to be a classical channel that maps deterministic probability distributions into deterministic probability distributions, i.e., channels that can be described by a permutation matrix, it would be interesting to investigate whether their discrimination could also be enhanced by sequential strategies. Additionally, Refs. 50 and 51 showed that indefinite causal order also manifests itself in classical processes. Hence, an interesting question would be whether indefinite causality is also a useful resource for classical channel discrimination.

## ACKNOWLEDGMENTS

The authors thank Cyril Branciard for comments on the first version of the manuscript. The authors are also grateful for their computer, “Princess Bubblegum,” for performing very demanding calculations. J.B. acknowledges funding from the Austrian Science Fund (FWF) through Zukunftscolleg ZK03 and START Project No. Y879-N27. M.M. was supported by the MEXT Quantum Leap Flagship Program (MEXT Q-LEAP) under Grant No. JPMXS0118069605 and by the Japan Society for the Promotion of Science (JSPS) through KAKENHI under Grant Nos. 17H01694, 18H04286, and 21H03394. M.T.Q. acknowledges the Austrian Science Fund (FWF) through the SFB project BeyondC (sub-project F7103), a grant from the Foundational Questions Institute (FQXi) as part of the Quantum Information Structure of Spacetime (QISS) Project (qiss.fr). The opinions expressed in this publication are those of the authors and do not necessarily reflect the views of the John Templeton Foundation. This project received funding from the European Union’s Horizon 2020 Research and Innovation Programme under Marie Skłodowska-Curie Grant Agreement No. 801110 and the Austrian Federal Ministry of Education, Science and Research (BMBWF). It reflects only the authors’ view, and the EU Agency is not responsible for any use that may be made of the information it contains.

## AUTHOR DECLARATIONS

### Conflict of Interest

The authors have no conflicts to disclose.

DATA AVAILABILITY

Data sharing is not applicable to this article as no new data were created or analyzed in this study. The code that supports the findings of this study are openly available at [https://github.com/mtcq/unitary\\_channel\\_discrimination](https://github.com/mtcq/unitary_channel_discrimination).<sup>52</sup>

The Appendix is composed of the following sections: Appendix A presents the definition of two- and three-slot process matrices in terms of positivity and linear normalization constraints and a small review of the formulation of the maximal probability of successful discrimination in terms of primal and dual SDP problems; Appendix B presents the Proof of Theorem 1; Appendix C presents the proof of Examples 1 and 2; Appendix D presents the proof of Example 3, Appendix E presents the Proof of Theorem 4; and, finally, Appendix F contains the Proof of Theorem 5.

Some of the sections in the Appendix will make use of the *link product*<sup>13</sup> between two linear operators, which is a useful mathematical tool to compose linear maps that are represented by their Choi operators. If  $\tilde{C} := \tilde{B} \circ \tilde{A}$  is the composition of the linear maps  $\tilde{A} : \mathcal{L}(\mathcal{H}^1) \rightarrow \mathcal{L}(\mathcal{H}^2)$  and  $\tilde{B} : \mathcal{L}(\mathcal{H}^2) \rightarrow \mathcal{L}(\mathcal{H}^3)$ , the Choi operator of  $\tilde{C}$  is given by  $C = A * B$ , where  $A$  and  $B$  are the Choi operators of  $\tilde{A}$  and  $\tilde{B}$ , respectively, and  $*$  stands for the link product, which we now define. Let  $A \in \mathcal{L}(\mathcal{H}^1 \otimes \mathcal{H}^2)$  and  $B \in \mathcal{L}(\mathcal{H}^2 \otimes \mathcal{H}^3)$  be linear operators. The link product  $A * B \in \mathcal{L}(\mathcal{H}^1 \otimes \mathcal{H}^3)$  is defined as

$$A^{12} * B^{23} := \text{Tr}_2 \left[ \left( (A^{12})^{T_2} \otimes \mathbb{1}^3 \right) \left( \mathbb{1}^1 \otimes B^{23} \right) \right], \tag{27}$$

where  $(\cdot)^{T_2}$  stands for the partial transposition on the linear space  $\mathcal{H}^2$ .

We remark that identifying the linear spaces where the operators act is an important part of the link product, also, if we keep track on these linear spaces, the link product is commutative and associative.

APPENDIX A: SEMIDEFINITE PROGRAMMING FORMULATION AND LINEAR CONSTRAINTS FOR GENERAL PROCESSES OF  $k = 2$  AND  $k = 3$  SLOTS

For sake of completeness, in the following, we present a characterization of general processes with  $k = 2$  and  $k = 3$  slots, which are bipartite and tripartite process matrices, in terms of positivity constraints and linear normalization constraints. While positivity is a consequence of physical considerations, the linear normalization constraints are a direct implication of Eq. (15). We refer to Ref. 34 for a detailed explanation of the method and Ref. 53 for the explicit characterization of general processes with  $k = 3$  slots.

We recall that we represent the trace-and-replace operation on  $\mathcal{H}^X$  as  $X(\cdot) = \text{Tr}_X(\cdot) \otimes \frac{\mathbb{1}^X}{d_X}$ , where  $d_X = \dim(\mathcal{H}^X)$ .

Let  $W \in \mathcal{L}(\mathcal{H}^{A_1 A_0} \otimes \mathcal{H}^{B_1 B_0})$  be a two-slot (bipartite) process matrix. Then,

$$W \geq 0, \tag{A1}$$

$$\text{Tr}(W) = d_{A_0} d_{B_0}, \tag{A2}$$

$$_{A_1 A_0} W = _{A_1 A_0 B_0} W, \tag{A3}$$

$$_{B_1 B_0} W = _{A_0 B_1 B_0} W, \tag{A4}$$

$$W + _{A_0 B_0} W = _{A_0} W + _{B_0} W. \tag{A5}$$

Now, let  $W \in \mathcal{L}(\mathcal{H}^{A_1 A_0} \otimes \mathcal{H}^{B_1 B_0} \otimes \mathcal{H}^{C_1 C_0})$  be a three-slot (tripartite) process matrix. Then,

$$W \geq 0, \tag{A6}$$

$$\text{Tr}(W) = d_{A_0} d_{B_0} d_{C_0}, \tag{A7}$$

$$_{A_1 A_0 B_1 B_0} W = _{A_1 A_0 B_1 B_0 C_0} W, \tag{A8}$$

$$_{B_1 B_0 C_1 C_0} W = _{A_0 B_1 B_0 C_1 C_0} W, \tag{A9}$$

$$_{A_1 A_0 C_1 C_0} W = _{A_1 A_0 B_0 C_1 C_0} W, \tag{A10}$$

$$_{A_1 A_0} W + _{A_1 A_0 B_0 C_0} W = _{A_1 A_0 B_0} W + _{A_1 A_0 C_0} W, \tag{A11}$$

$${}_{B_i B_0} W + {}_{A_0 B_i B_0 C_0} W = {}_{A_0 B_i B_0} W + {}_{B_i B_0 C_0} W, \tag{A12}$$

$${}_{C_i C_0} W + {}_{A_0 B_0 C_i C_0} W = {}_{A_0 C_i C_0} W + {}_{B_0 C_i C_0} W, \tag{A13}$$

$$W + {}_{A_0 B_0} W + {}_{A_0 C_0} W + {}_{B_0 C_0} W = {}_{A_0} W + {}_{B_0} W + {}_{C_0} W + {}_{A_0 B_0 C_0} W. \tag{A14}$$

We also recall that the maximal probability of successful discrimination in Eq. (16) can be equivalently expressed by the primal SDP problem

$$\begin{aligned} & \text{given } \{p_i, C_i\} \\ & \text{maximize } \sum_i p_i \text{Tr}(T_i^S C_i^{\otimes k}) \\ & \text{subject to } T_i^S \geq 0 \ \forall i, \quad \sum_i T_i^S = W^S, \end{aligned} \tag{A15}$$

and the dual problem

$$\begin{aligned} & \text{given } \{p_i, C_i\} \\ & \text{minimize } \lambda \\ & \text{subject to } p_i C_i^{\otimes k} \leq \lambda \overline{W}^S \ \forall i, \end{aligned} \tag{A16}$$

the latter of which can be straightforwardly phrased as an SDP by absorbing the coefficient  $\lambda$  into  $\overline{W}^S$ , which is the dual affine of process matrix  $W^S$ .<sup>26</sup> For parallel processes  $W^{\text{PAR}}$ , their dual affine  $\overline{W}^{\text{PAR}}$  is a general  $k$ -partite quantum channel, for sequential processes  $W^{\text{SEQ}}$ , their dual affine  $\overline{W}^{\text{SEQ}}$  is a  $k$ -partite channel with memory, and for general processes  $W^{\text{GEN}}$ , their dual affine  $\overline{W}^{\text{GEN}}$  is a  $k$ -partite non-signaling channel. We refer to Ref. 26 for more details. In principle, any solution of the primal problem gives an upper bound to  $P^S$  and any solution of the dual problem, a lower bound. From strong duality, it is guaranteed that the optimal solution of both problems coincides. Both problems are used in the computer-assisted-proof method applied in this paper.

## APPENDIX B: PROOF OF THEOREM 1

We start this section with Lemma 1, which plays a main role in the Proof of Theorem 1 and may be of independent interest. The theorems presented in this section employ methods that are similar to the ones in Refs. 20 and 54, which exploit the covariance of processes to parallelize strategies.

*Lemma 1.* Let  $\{T_U\}_U$ ,  $T_U \in \mathcal{L}(\mathcal{H}^I \otimes \mathcal{H}^O)$ , be a general  $k$ -slot tester associated with the general process  $W := \sum_U T_U$ , which respects the commutation relation

$$W^{IO} \left( \mathbb{1} \otimes U^{\otimes k} \right)^{IO} = \left( \mathbb{1} \otimes U^{\otimes k} \right)^{IO} W^{IO} \tag{B1}$$

for every unitary operator  $U \in \mathcal{L}(\mathbb{C}^d)$  from a set  $\{U\}_U$ .

Then, there exists a parallel  $k$ -slot tester  $\{T_U^{\text{PAR}}\}_i$  such that

$$\text{Tr}(T_U^{\text{PAR}} |U\rangle\langle U|^{\otimes k}) = \text{Tr}(T_U |U\rangle\langle U|^{\otimes k}) \quad \forall U \in \{U\}_U. \tag{B2}$$

Moreover, this parallel tester can be written as  $T_U^{\text{PAR}} = \rho^{I'I} * M_U^{I'O}$ , where  $\mathcal{H}^{I'}$  is an auxiliary space, which is isomorphic to  $\mathcal{H}^I$ ,  $\rho \in \mathcal{L}(\mathcal{H}^I \otimes \mathcal{H}^{I'})$  is a quantum state defined by

$$\rho^{I'I} := \sqrt{W}^{T'I'I} |\mathbb{1}\rangle\langle \mathbb{1}|^{I'I} \sqrt{W}^{T'I'I}, \tag{B3}$$

and  $\{M_U\}_U$  is a POVM defined by<sup>55</sup>

$$M_U^{I'O} := \sqrt{W}^{-1'I'O} T_U^{I'O} \sqrt{W}^{-1'I'O}. \quad (\text{B4})$$

*Proof.* We start our proof by verifying that  $\rho \in \mathcal{L}(\mathcal{H}^{I'} \otimes \mathcal{H}^I)$  is a valid quantum state. The operator  $\rho$  is positive semidefinite because it is a composition of positive semidefinite operators and the normalization condition follows from

$$\text{Tr}(\rho) = \text{Tr}\left(\sqrt{W}^{T'I'I} |\mathbb{1}\rangle\langle\mathbb{1}|^{I'I} \sqrt{W}^{T'I'I}\right) \quad (\text{B5})$$

$$= \text{Tr}\left(W^{T'I'I} |\mathbb{1}\rangle\langle\mathbb{1}|^{I'I}\right) \quad (\text{B6})$$

$$= \text{Tr}\left(W^{I'I} |\mathbb{1}\rangle\langle\mathbb{1}|^{I'I}\right) \quad (\text{B7})$$

$$= \text{Tr}\left(W^{I'I} |\mathbb{1}\rangle\langle\mathbb{1}|^{I'I}\right) \quad (\text{B8})$$

$$= 1, \quad (\text{B9})$$

where the last equation holds because since  $W$  is a general process, it satisfies  $\text{Tr}(WC) = 1$  for any  $C$ , that is, the Choi operator of a channel.

Let us now verify that the set of operators  $\{M_U\}_U$  forms a valid POVM. For that, it is enough to recognize that all operators  $M_U$  are compositions of positive semidefinite operators that add up to the identity according to

$$\sum_U M_U = \sqrt{W}^{-1'I'O} \sum_U T_U^{I'O} \sqrt{W}^{-1'I'O} \quad (\text{B10})$$

$$= \sqrt{W}^{-1'I'O} W^{I'O} \sqrt{W}^{-1'I'O} \quad (\text{B11})$$

$$= \mathbb{1}^{I'O}. \quad (\text{B12})$$

The relation  $\sqrt{W}^{-1} W \sqrt{W}^{-1} = \mathbb{1}$  can be shown by writing  $W$  in an orthonormal basis as  $W = \sum_i \alpha_i |i\rangle\langle i|$  and  $\sqrt{W}^{-1} = \sum_i \alpha_i^{-1/2} |i\rangle\langle i|$ .

Recall that for any unitary operator  $U$ , we have the identity  $|U\rangle\langle U|^T = |U^*\rangle\langle U^*|$ , and if  $C^{IO}$  is the Choi operator of a linear map  $\tilde{C} : \mathcal{L}(\mathcal{H}^I) \rightarrow \mathcal{L}(\mathcal{H}^O)$ ,  $\rho^{I'I} \in \mathcal{L}(\mathcal{H}^{I'} \otimes \mathcal{H}^I)$ , it holds that  $\rho^{I'I} * C^{IO} = \left(\mathbb{1} \otimes \tilde{C}(\rho^{I'I})\right)^{I'O}$ . In addition, if a diagonalizable operator  $W^{IO}$  commutes with  $(\mathbb{1} \otimes U^{\otimes k})^{IO}$ , its positive semidefinite square root  $\sqrt{W}$  also commutes with<sup>56</sup>  $(\mathbb{1} \otimes U^{\otimes k})^{IO}$ ; hence, we have

$$\sqrt{W} (\mathbb{1} \otimes U^{\otimes k}) = (\mathbb{1} \otimes U^{\otimes k}) \sqrt{W}. \quad (\text{B13})$$

By taking the complex conjugation on both sides of Eq. (B13) and exploiting the fact that  $\sqrt{W} = \sqrt{W}^\dagger$  implies  $\sqrt{W}^T = \sqrt{W}^*$ , it holds that

$$\sqrt{W}^T (\mathbb{1} \otimes U^{*\otimes k}) = (\mathbb{1} \otimes U^{*\otimes k}) \sqrt{W}^T. \quad (\text{B14})$$

With these identities in hand, we can evaluate the link product  $\rho^{I'I} * (|U^{\otimes k}\rangle\langle U^{\otimes k}|^T)^{IO}$ , which will be used in the next step of the proof, to obtain

$$\rho^{I'I} * (|U\rangle\langle U|^{\otimes k T})^{IO} = \rho^{I'I} * (|U^*\rangle\langle U^*|^{\otimes k})^{IO} \quad (\text{B15})$$

$$= \left[ (\mathbb{1} \otimes U^{*\otimes k}) \rho (\mathbb{1} \otimes U^{T\otimes k}) \right]^{I'O} \quad (\text{B16})$$

$$= \left[ (\mathbb{1} \otimes U^{*\otimes k}) \sqrt{W}^T |\mathbb{1}\rangle\langle\mathbb{1}| \sqrt{W}^T (\mathbb{1} \otimes U^{T\otimes k}) \right]^{I'O} \quad (\text{B17})$$

$$= \left[ \sqrt{W}^T (\mathbb{1} \otimes U^{*\otimes k}) |\mathbb{1}\rangle\langle\mathbb{1}| (\mathbb{1} \otimes U^{T\otimes k}) \sqrt{W}^T \right]^{I'O} \quad (\text{B18})$$

$$= \left( \sqrt{W}^T |U^*\rangle\langle U^*|^{\otimes k} \sqrt{W}^T \right)^{I'O}. \quad (\text{B19})$$

We now finish the proof by verifying that

$$\text{Tr}\left(T_U^{\text{PAR}} |U\rangle\langle U|^{\otimes k}\right) = \text{Tr}\left[(\rho^{I'I} * M_U^{I'O}) |U\rangle\langle U|^{\otimes k IO}\right] \quad (\text{B20})$$

$$= (\rho^{I'I} * M_U^{I'O}) * (|U\rangle\langle U|^{\otimes k})^{T IO} \quad (\text{B21})$$

$$= M_U^{I'O} * \left(\rho^{I'I} * |U^*\rangle\langle U^*|^{\otimes k IO}\right) \quad (\text{B22})$$

$$= M_U^{I'O} * \left(\sqrt{W}^T |U^*\rangle\langle U^*|^{\otimes k} \sqrt{W}^T\right)^{I'O} \quad [\text{applying Eq. (B19)}] \quad (\text{B23})$$

$$= \text{Tr}\left[M_U^{I'O} \left(\sqrt{W}^T |U^*\rangle\langle U^*|^{\otimes k} \sqrt{W}^T\right)^{T I'O}\right] \quad (\text{B24})$$

$$= \text{Tr}\left[M_U^{I'O} \left(\sqrt{W} |U\rangle\langle U|^{\otimes k} \sqrt{W}\right)^{I'O}\right] \quad (\text{B25})$$

$$= \text{Tr}\left[\left(\sqrt{W}^{-1} T_U \sqrt{W}^{-1}\right)^{I'O} \left(\sqrt{W} |U\rangle\langle U|^{\otimes k} \sqrt{W}\right)^{I'O}\right] \quad (\text{B26})$$

$$= \text{Tr}\left(T_U |U\rangle\langle U|^{\otimes k}\right). \quad (\text{B27})$$

□

Now, we prove Theorem 1.

**Theorem 1.** Let  $\mathcal{E} = \{p_U, |U\rangle\langle U|\}_U$  be an ensemble of unitary channels where the set of unitary operators  $U \in \mathcal{L}(\mathbb{C}^d)$ ,  $\{U\}_U$  forms a group up to a global phase—that is, there exist real numbers  $\phi_i$  such that the following holds:

- $e^{i\phi_1} \mathbb{1} \in \{U\}_U$ .
- If  $A \in \{U\}_U$ , then  $e^{i\phi_A} A^{-1} \in \{U\}_U$ .
- If  $A, B \in \{U\}_U$ , then  $e^{i\phi_{AB}} AB \in \{U\}_U$

and the distribution  $\{p_U\}_U$  is uniform—that is, if the set has  $N$  elements,  $p_U = \frac{1}{N}$ .

Then, for any number of uses  $k$  and every general tester  $\{T_U^{\text{GEN}}\}_U$ ,  $T_U^{\text{GEN}} \in \mathcal{L}(\mathcal{H}^I \otimes \mathcal{H}^O)$ , there exists a parallel tester  $\{T_U^{\text{PAR}}\}_U$ ,  $T_U^{\text{PAR}} \in \mathcal{L}(\mathcal{H}^I \otimes \mathcal{H}^O)$ , such that

$$\frac{1}{N} \sum_{U \in \{U\}_U} \text{Tr}\left(T_U^{\text{GEN}} |U\rangle\langle U|^{\otimes k}\right) = \frac{1}{N} \sum_{U \in \{U\}_U} \text{Tr}\left(T_U^{\text{PAR}} |U\rangle\langle U|^{\otimes k}\right). \quad (\text{B28})$$

Before presenting the proof, we recall that unitary operators that are equivalent up to a global phase represent equivalent unitary channels. That is, if  $U' = e^{i\phi} U : \mathcal{H}^I \rightarrow \mathcal{H}^O$  is a linear operator, its associated map is given by

$$\tilde{U}'(\rho) = U' \rho U'^{\dagger} \quad (\text{B29})$$

$$= e^{i\phi} e^{-i\phi} U \rho U^{\dagger} \quad (\text{B30})$$

$$= U \rho U^{\dagger} \quad (\text{B31})$$

$$= \tilde{U}(\rho), \quad (\text{B32})$$

and its Choi operator  $|U\rangle\langle U|$  respects

$$|U\rangle\langle U| = |U'\rangle\langle U'|. \quad (\text{B33})$$

Due to this fact, the two sets of operators  $\{U_i\}_i$  and  $\{e^{i\phi_i} U_i\}_i$  represent the same set of quantum channels.



*Proof.* The proof goes as follows: we start by using the general tester  $\{T_U^{\text{GEN}}\}_U$  to construct another general tester  $\{T_U\}_U$ , which obeys

$$\frac{1}{N} \sum_{U \in \{U\}_U} \text{Tr}(T_U^{\text{GEN}} |U\rangle\langle U|^{\otimes k}) = \frac{1}{N} \sum_{U \in \{U\}_U} \text{Tr}(T_U |U\rangle\langle U|^{\otimes k}). \quad (\text{B34})$$

Then, we prove that the general tester  $\{T_U\}_U$  we defined respects the hypothesis of Lemma 1. Hence, there exists a parallel tester  $\{T_U^{\text{PAR}}\}_U$ , which is equivalent to  $\{T_U\}_U$  when acting on the set of unitary operators  $\{U\}_U$ .

Let us start by defining the general tester  $\{T_U\}_U$  as

$$T_U := \frac{1}{N} \sum_{V \in \{U\}_U} \left( \mathbb{1}^I \otimes V^{\dagger \otimes k} \right)^{IO} T_{VU}^{\text{GEN}} \left( \mathbb{1}^I \otimes V^{\otimes k} \right)^{IO}, \quad (\text{B35})$$

where  $VU$  stands for the standard operator composition up to a global phase, that is, if  $VU$  is not in the set  $\{U\}_U$ , we pick  $e^{i\phi_{VU}} VU$ , which is ensured to be an element of  $\{U\}_U$ . Before proceeding, we should verify that the set of operators  $\{T_U\}_U$  is, indeed, a valid general tester. Note that since  $T_U$  is a composition of positive semidefinite operators, it holds that  $T_U \geq 0$  for every  $U$ . We now show that  $W := \sum_U T_U$  is a valid general process. First, note that

$$W := \sum_U T_U \quad (\text{B36})$$

$$= \sum_U \frac{1}{N} \sum_V (\mathbb{1} \otimes V^{\dagger \otimes k}) T_{VU}^{\text{GEN}} (\mathbb{1} \otimes V^{\otimes k}) \quad (\text{B37})$$

$$= \frac{1}{N} \sum_U \sum_V (\mathbb{1} \otimes V^{\dagger \otimes k}) T_{V(V^{-1}U)}^{\text{GEN}} (\mathbb{1} \otimes V^{\otimes k}) \quad (\text{B38})$$

$$= \frac{1}{N} \sum_U \sum_V (\mathbb{1} \otimes V^{\dagger \otimes k}) T_U^{\text{GEN}} (\mathbb{1} \otimes V^{\otimes k}) \quad (\text{B39})$$

$$= \frac{1}{N} \sum_V (\mathbb{1} \otimes V^{\dagger \otimes k}) \sum_U T_U^{\text{GEN}} (\mathbb{1} \otimes V^{\otimes k}) \quad (\text{B40})$$

$$= \frac{1}{N} \sum_V (\mathbb{1} \otimes V^{\dagger \otimes k}) W^{\text{GEN}} (\mathbb{1} \otimes V^{\otimes k}), \quad (\text{B41})$$

where  $W^{\text{GEN}} := \sum_U T_U^{\text{GEN}}$ , and in Eq. (B38), we have used the change of variable  $U \mapsto V^{-1}U$ , which does not affect the sum because the set  $\{U\}_U$  is a group.

Note also that, if  $C^{IO}$  is the Choi operator of a quantum channel, the operator defined by

$$C^{IO} := \frac{1}{N} \sum_V (\mathbb{1} \otimes V^{\otimes k})^{IO} C^{IO} (\mathbb{1} \otimes V^{\dagger \otimes k})^{IO} \quad (\text{B42})$$

is a valid channel since it is positive semidefinite and  $\text{Tr}_O(C^{IO}) = \text{Tr}_O(C^{IO}) = \mathbb{1}^I$ . It then follows that, for every quantum channel of the form  $C = \otimes_{i=1}^k C_i^{I_i O_i}$ , we have

$$\text{Tr}(W^{IO} C^{IO}) = \frac{1}{N} \text{Tr} \left[ \sum_V (\mathbb{1} \otimes V^{\dagger \otimes k}) W^{\text{GEN}} (\mathbb{1} \otimes V^{\otimes k}) C \right] \quad (\text{B43})$$

$$= \frac{1}{N} \text{Tr} \left[ W^{\text{GEN}} \sum_V (\mathbb{1} \otimes V^{\otimes k}) C (\mathbb{1} \otimes V^{\dagger \otimes k}) \right] \quad (\text{B44})$$

$$= \text{Tr}(W^{IO} C^{IO}) \quad (\text{B45})$$

$$= 1, \quad (\text{B46})$$

ensuring that  $\{T_U\}_U$  is a valid general tester.

The next step is to show that the tester  $\{T_U\}_U$  attains the same success probability for discriminating the ensemble  $\mathcal{E} = \{p_U, |U\rangle\langle U|\}_U$  as the tester  $\{T_U^{\text{GEN}}\}_U$ . This claim follows from direct calculation, that is,

$$\frac{1}{N} \sum_U \text{Tr}(T_U |U\rangle\langle U|^{\otimes k}) = \frac{1}{N^2} \sum_U \sum_V \text{Tr} \left[ (1 \otimes V^{\dagger \otimes k}) T_{VU}^{\text{GEN}} (1 \otimes V^{\otimes k}) |U\rangle\langle U|^{\otimes k} \right] \quad (\text{B47})$$

$$= \frac{1}{N^2} \sum_U \sum_V \text{Tr} \left[ T_{VU}^{\text{GEN}} (1 \otimes V^{\otimes k}) |U\rangle\langle U|^{\otimes k} (1 \otimes V^{\dagger \otimes k}) \right] \quad (\text{B48})$$

$$= \frac{1}{N^2} \sum_U \sum_V \text{Tr} \left( T_{VU}^{\text{GEN}} |VU\rangle\langle VU|^{\otimes k} \right) \quad (\text{B49})$$

$$= \frac{1}{N^2} \sum_U \sum_V \text{Tr} \left( T_{V(V^{-1}U)}^{\text{GEN}} |V(V^{-1}U)\rangle\langle V(V^{-1}U)|^{\otimes k} \right) \quad (\text{B50})$$

$$= \frac{1}{N^2} \sum_U \sum_V \text{Tr} \left( T_U^{\text{GEN}} |U\rangle\langle U|^{\otimes k} \right) \quad (\text{B51})$$

$$= \frac{1}{N} \sum_U \text{Tr} \left( T_U^{\text{GEN}} |U\rangle\langle U|^{\otimes k} \right). \quad (\text{B52})$$

The final step is to verify that the process  $W := \sum_U T_U$  commutes with  $1 \otimes U^{\otimes k}$  for every unitary operator  $U \in \{U\}_U$  to ensure that the tester  $\{T_U\}_U$  fulfills the hypothesis of Lemma 1. Direct calculation shows that

$$(1 \otimes U^{\otimes k}) W (1 \otimes U^{\dagger \otimes k}) = (1 \otimes U^{\otimes k}) \frac{1}{N} \sum_V (1 \otimes V^{\dagger \otimes k}) W^{\text{GEN}} (1 \otimes V^{\otimes k}) (1 \otimes U^{\dagger \otimes k}) \quad (\text{B53})$$

$$= \frac{1}{N} \sum_V (1 \otimes (UV^{\dagger})^{\otimes k}) W^{\text{GEN}} (1 \otimes (VU^{\dagger})^{\otimes k}) \quad (\text{B54})$$

$$= \frac{1}{N} \sum_V [1 \otimes (U(VU)^{\dagger})^{\otimes k}] W^{\text{GEN}} [1 \otimes ((VU)U^{\dagger})^{\otimes k}] \quad (\text{B55})$$

$$= \frac{1}{N} \sum_V (1 \otimes V^{\dagger \otimes k}) W^{\text{GEN}} (1 \otimes V^{\otimes k}) \quad (\text{B56})$$

$$= W. \quad (\text{B57})$$

Hence, we have that

$$W^{IO} (1 \otimes U^{\otimes k})^{IO} = (1 \otimes U^{\otimes k})^{IO} W^{IO}, \quad (\text{B58})$$

and by Lemma 1, one can construct a parallel tester  $\{T_U^{\text{PAR}}\}_U$ , which respects

$$\text{Tr} \left( T_U^{\text{PAR}} |U\rangle\langle U|^{\otimes k} \right) = \text{Tr} \left( T_U |U\rangle\langle U|^{\otimes k} \right) \quad \forall U \in \{U\}_U, \quad (\text{B59})$$

and therefore, by applying Eq. (B52), we have

$$\frac{1}{N} \sum_U \text{Tr} \left( T_U^{\text{PAR}} |U\rangle\langle U|^{\otimes k} \right) = \frac{1}{N} \sum_U \text{Tr} \left( T_U^{\text{GEN}} |U\rangle\langle U|^{\otimes k} \right), \quad (\text{B60})$$

concluding our proof.  $\square$

### APPENDIX C: PROOF OF EXAMPLES 1 AND 2

The examples in this section show the advantage of sequential strategies over parallel strategies in channel discrimination tasks that involve only unitary channels and using  $k = 2$  copies. In the examples of this section, general strategies cannot outperform sequential ones. We recall that in the following,  $\sigma_x$ ,  $\sigma_y$ , and  $\sigma_z$  denote the Pauli operators and  $H := |+\rangle\langle 0| + |-\rangle\langle 1|$ , where  $|\pm\rangle := \frac{1}{\sqrt{2}}(|0\rangle \pm |1\rangle)$ , denotes the

Hadamard gate. In addition, if  $A \in \mathcal{L}(\mathbb{C}^d)$  is a linear operator with spectral decomposition  $A = \sum_i \alpha_i |\psi_i\rangle\langle\psi_i|$ , its square root is defined as  $\sqrt{A} := \sum_i \sqrt{\alpha_i} |\psi_i\rangle\langle\psi_i|$ .

We start by proving Example 1 from the main text. It concerns the discrimination of an ensemble composed of a uniform probability distribution and a set of unitaries that does not form a group.

*Example 1.* The ensemble composed of a uniform probability distribution and  $N = 4$  qubit unitary channels given by  $\{U_i\} = \{\mathbb{1}, \sqrt{\sigma_x}, \sqrt{\sigma_y}, \sqrt{\sigma_z}\}$ , in a discrimination task that allows for  $k = 2$  copies, can be discriminated under a sequential strategy with success probability  $P^{\text{SEQ}} = 1$ , while any parallel strategy yields  $P^{\text{PAR}} < 1$ .

*Proof.* The sequential strategy that attains perfect discrimination is easily understood by realizing that when the  $k = 2$  copies of the unitaries  $\{U_i\}$  are applied in sequence, one recovers  $U_i U_i = \sqrt{\sigma_i} \sqrt{\sigma_i} = \sigma_i$ , where  $\sigma_1 = \mathbb{1}$ ,  $\sigma_2 = \sigma_x$ ,  $\sigma_3 = \sigma_y$ ,  $\sigma_4 = \sigma_z$ . Therefore, the task reduces to the discrimination of the four Pauli operators with  $k = 1$  copy, which can be perfectly realized with a two-qubit maximally entangled state and a Bell measurement.

In order to show that the probability  $P^{\text{PAR}}$  of discriminating these unitary channels with  $k = 2$  copies in a parallel strategy is strictly less than one, we make use of the dual problem associated with the SDP, which computes the maximal probability of successful discrimination, given in Eq. (A16). Hence, in order to obtain an upper bound for the maximal success probability  $P^{\text{PAR}}$ , it is enough to find a value  $\lambda < 1$  and the Choi state of a quantum channel  $\bar{W}$ , that is,  $\bar{W} \geq 0$  and  $\text{Tr}_O(\bar{W}) = \mathbb{1}^I$ , which respect

$$\frac{1}{4} |U_i\rangle\langle\langle U_i|^{\otimes 2} \leq \lambda \bar{W} \quad \text{for } i \in \{1, 2, 3, 4\}. \tag{C1}$$

Using the computer-assisted-proof method presented in Ref. 26, we obtain an operator  $\bar{W}$  that satisfies all the quantum channel conditions exactly and, for  $\lambda = \frac{9571}{1000}$ , satisfies the inequality (C1). Hence,

$$P^{\text{PAR}} \leq \frac{9571}{1000}. \tag{C2}$$

In our online repository,<sup>52</sup> we present a Mathematica™ notebook that can be used to verify that  $\bar{W}$  is a valid Choi state of a quantum channel. □

Another similar example with an interesting property is given by the unitary channel ensemble composed of a uniform probability distribution and  $\{U_i\} = \{\mathbb{1}, \sigma_x, \sigma_y, \sqrt{\sigma_z}\}$ .

To prove this example as well, let us start by constructing a perfect sequential strategy. We start by noting that the four Bell states can be written as

$$\begin{aligned} |\phi^+\rangle &:= \frac{|00\rangle + |11\rangle}{\sqrt{2}} = (\mathbb{1} \otimes \mathbb{1})|\phi^+\rangle, & |\phi^-\rangle &:= \frac{|00\rangle - |11\rangle}{\sqrt{2}} = (\mathbb{1} \otimes \sigma_z)|\phi^+\rangle, \\ |\psi^+\rangle &:= \frac{|01\rangle + |10\rangle}{\sqrt{2}} = (\mathbb{1} \otimes \sigma_x)|\phi^+\rangle, & |\psi^-\rangle &:= \frac{|01\rangle - |10\rangle}{\sqrt{2}} = (-i)(\mathbb{1} \otimes \sigma_y)|\phi^+\rangle. \end{aligned}$$

In addition, since  $\sqrt{\sigma_z} = |0\rangle\langle 0| + i|1\rangle\langle 1|$ , we can check that the state

$$(\mathbb{1} \otimes \sqrt{\sigma_z})|\phi^+\rangle = \frac{|00\rangle + i|11\rangle}{\sqrt{2}} \tag{C3}$$

is orthogonal to  $|\psi^+\rangle$  and  $|\psi^-\rangle$ . We will now exploit these identities to construct a sequential strategy that attains  $P^{\text{SEQ}} = 1$  with  $k = 2$  uses.

The strategy goes as follows: Define the auxiliary space  $\mathcal{H}^{\text{aux}_1}$  to be isomorphic to  $\mathcal{H}^I$  and prepare the initial state  $\rho \in \mathcal{L}(\mathcal{H}^I \otimes \mathcal{H}^{\text{aux}_1})$  as

$$\rho^{I_1 \text{ aux}_1} := |\phi^+\rangle\langle\phi^+|^{\otimes I_1 \text{ aux}_1}. \tag{C4}$$

The state  $\rho^{I_1 aux_1}$  will then be subjected to the first copy of a unitary channel  $U_i$ , leading to the output state  $\rho^{O_1 aux_1}$ ,

$$\rho^{O_1 aux_1} = (U_i \otimes \mathbb{1}^{aux_1}) \rho^{I_1 aux_1} (U_i \otimes \mathbb{1}^{aux_1})^\dagger = (U_i \otimes \mathbb{1}^{aux_1}) |\phi^+\rangle \langle \phi^+|^{I_1 aux_1} (U_i \otimes \mathbb{1}^{aux_1})^\dagger. \quad (C5)$$

Then, we apply the following transformation on the state  $\rho^{O_1 aux_1}$ . We perform a projective measurement with POVM elements given by

$$M_{\psi^+} := |\psi^+\rangle \langle \psi^+|, \quad (C6)$$

$$M_{\psi^-} := |\psi^-\rangle \langle \psi^-|, \quad (C7)$$

$$M_\phi := |\phi^+\rangle \langle \phi^+| + |\phi^-\rangle \langle \phi^-|, \quad (C8)$$

and after the measurement, re-prepare the quantum system in the state  $\rho^{I_2 aux_2} = \widetilde{M}_i(\rho) = \sqrt{\widetilde{M}_i} \rho \sqrt{\widetilde{M}_i}^\dagger$  with probability  $\text{Tr}(\widetilde{M}_i(\rho)) = \text{Tr}(\rho M_i)$ . This transformation can be described by a Lüders instrument. It can be checked that, if  $U_i = \sigma_x$ , one obtains the outcome associated with  $M_{\psi^+}$  with probability one. Similarly, if  $U_i = \sigma_y$ , one obtains the outcome associated with  $M_{\psi^-}$  with probability one. Hence, in these two cases, we have perfect channel discrimination. Now, if we obtain the outcome associated with  $M_\phi$ , the unitary  $U_i$  can be either  $\mathbb{1}$  or  $\sqrt{\sigma_z}$ .

After performing the projective measurement with elements  $\{M_{\psi^+}, M_{\psi^-}, M_\phi\}$  and a Lüders instrument, the state  $\rho^{I_2 aux_2}$  is subjected to a second copy of  $U_i$ . Direct calculation shows that if  $U_i = \mathbb{1}$ , then after the use of the second copy of unitary  $U_i$ , the state  $\rho^{O_2 aux_2}$  of the system is

$$(\mathbb{1} \otimes \mathbb{1})^2 |\phi^+\rangle = |\phi^+\rangle. \quad (C9)$$

If  $U_i = \sqrt{\sigma_z}$  after the second use of the unitary  $U_i$ , the state  $\rho^{O_2 aux_2}$  of the system is

$$(\mathbb{1} \otimes \sqrt{\sigma_z})^2 |\phi^+\rangle = |\phi^-\rangle. \quad (C10)$$

Since  $|\phi^+\rangle$  and  $|\phi^-\rangle$  are orthogonal, they can be discriminated with probability one. Hence, the set of unitary operators  $\{U_i\}_{i=1}^4$  can be perfectly discriminated in a sequential strategy with  $k = 2$  copies.

Using the tester formalism, this sequential strategy would be presented in terms of a sequential tester  $T^{\text{SEQ}} = \{T_i^{\text{SEQ}}\}$ , which can be implemented by an input quantum state  $\rho$ , a quantum encoder channel  $\widetilde{E}$ , and a quantum measurement  $\{N_i\}_i$ . For completeness, we now present an explicit sequential tester that attains  $P^{\text{SEQ}} = 1$ . As in the strategy described earlier, we set the initial state as  $\rho^{I_1 aux_1} := |\phi^+\rangle \langle \phi^+|^{I_1 aux_1}$ . Now, instead of using an instrument, we define a quantum encoder channel  $\widetilde{E} : \mathcal{L}(\mathcal{H}^{O_1} \otimes \mathcal{H}^{aux_1}) \rightarrow \mathcal{L}(\mathcal{H}^{I_2} \otimes \mathcal{H}^{aux_2} \otimes \mathcal{H}^{aux'_2})$  as

$$\widetilde{E}(\rho) = \left(\sqrt{M_\phi} \rho \sqrt{M_\phi}^\dagger\right) \otimes M_\phi^{aux'_2} + \left(\sqrt{M_{\psi^+}} \rho \sqrt{M_{\psi^+}}^\dagger\right) \otimes M_{\psi^+}^{aux'_2} + \left(\sqrt{M_{\psi^-}} \rho \sqrt{M_{\psi^-}}^\dagger\right) \otimes M_{\psi^-}^{aux'_2} \quad (C11)$$

so that  $\rho^{I_2 aux_2 aux'_2} = \widetilde{E}(\rho^{O_1 aux_1})$ . We finish our sequential tester construction by presenting quantum measurement given by operators  $N_i \in \mathcal{L}(\mathcal{H}^{O_2} \otimes \mathcal{H}^{aux} \otimes \mathcal{H}^{aux'_2})$ ,

$$N_1 := |\phi^+\rangle \langle \phi^+| \otimes M_\phi, \quad (C12)$$

$$N_2 := |\phi^-\rangle \langle \phi^-| \otimes M_\phi, \quad (C13)$$

$$N_3 := \mathbb{1} \otimes M_{\psi^+}, \quad (C14)$$

$$N_4 := \mathbb{1} \otimes M_{\psi^-}. \quad (C15)$$

In this way, if  $E$  is the Choi operator of the channel  $\widetilde{E}$ , the sequential tester with elements  $T_i^{\text{SEQ}} := \rho * E * N_i^T$  respects  $\sum_i \text{Tr}(T_i^{\text{SEQ}} |U_i\rangle \langle U_i| \otimes \mathbb{1}^{\otimes 2}) = 1$ .

In order to show that the probability  $P^{\text{PAR}}$  of discriminating these unitary channels with  $k = 2$  copies in a parallel strategy is strictly less than one, we apply the method of computer-assisted proof again to obtain the upper bound of

$$P^{\text{PAR}} \leq \frac{9741}{1000}. \quad (C16)$$

An interesting property of this example is that, with the above described sequential strategy, for the cases in which the unknown channel is either  $U_2 = \sigma_x$  or  $U_3 = \sigma_y$ , a conclusive answer is achieved after only one use of the unknown channel. From the uniform probability of the ensemble, we know that this scenario would occur with probability  $\frac{1}{2}$ . Only when the unknown channel is either  $U_1 = \mathbb{1}$  or  $U_4 = \sqrt{\sigma_z}$  is

it necessary to use the second copy of the unknown channel to arrive at a conclusive answer. This scenario would also occur with probability  $\frac{1}{2}$ . If one considers this discrimination task as being performed repeatedly, always drawing the unknown channel with uniform probability from the ensemble  $\{U_i\} = \{\mathbb{1}, \sigma_x, \sigma_y, \sqrt{\sigma_z}\}$ , then one can see that, on average over the multiples runs of the task, perfect discrimination will be achieved using only 1.5 copies of the unknown channel under this sequential strategy.

We now prove Example 2 from the main text. It concerns the discrimination of an ensemble composed of a non-uniform probability distribution and a set of unitaries that forms a group.

*Example 2.* Let  $\{U_i\} = \{\mathbb{1}, \sigma_x, \sigma_y, \sigma_z, H, \sigma_x H, \sigma_y H, \sigma_z H\}$  be a tuple of  $N = 8$  unitary channels that forms a group up to a global phase, and let  $\{p_i\}$  be a probability distribution in which each element  $p_i$  is proportional to the  $i$ th digit of the number  $\pi \approx 3.141\,592\,6$ , that is,  $\{p_i\} = \{\frac{3}{31}, \frac{1}{31}, \frac{4}{31}, \frac{1}{31}, \frac{5}{31}, \frac{9}{31}, \frac{2}{31}, \frac{6}{31}\}$ . For the ensemble  $\{p_i, U_i\}$ , in a discrimination task that allows for  $k = 2$  copies, sequential strategies outperform parallel strategies, i.e.,  $P^{\text{PAR}} < P^{\text{SEQ}}$ .

*Proof.* The first step of the proof is to ensure that the tuple  $\{\mathbb{1}, \sigma_x, \sigma_y, \sigma_z, H, \sigma_x H, \sigma_y H, \sigma_z H\}$  forms a group up to a global phase. This is done by direct inspection. The second step of the proof is to ensure that there is a sequential strategy, which outperforms any parallel one. We accomplish this step with the aid of the computer-assisted-proof methods presented in Ref. 26. These methods allow us to compute rigorous and explicit upper and lower bounds for the maximal probability of success under parallel and sequential strategies. We obtain

$$\frac{8196}{10\,000} < P^{\text{PAR}} < \frac{8197}{10\,000} < P^{\text{SEQ}} < \frac{8198}{10\,000}, \tag{C17}$$

ensuring that  $P^{\text{PAR}} < P^{\text{SEQ}}$ .

The code used in the computer-assisted proof of the this example is publicly available at our online repository,<sup>52</sup> along with a Mathematica notebook file, which shows that this set of unitaries forms a group.  $\square$

#### APPENDIX D: PROOF OF EXAMPLE 3

The example in this section shows the advantage of general strategies over sequential strategies and of sequential strategies over parallel strategies in channel discrimination tasks that only involve unitary channels and using  $k = 3$  copies.

We start by proving Example 3 from the main text. It concerns the discrimination of an ensemble composed of a uniform probability distribution and a set of unitaries that does not form a group. We recall that for the following, we define  $H_y := |+_y\rangle\langle 0| + |-_y\rangle\langle 1|$ , where  $|\pm_y\rangle := \frac{1}{\sqrt{2}}(|0\rangle \pm i|1\rangle)$ , and  $H_p := |+_p\rangle\langle 0| + |-_p\rangle\langle 1|$ , where  $|\pm_p\rangle := \frac{1}{5}(3|0\rangle + 4|1\rangle)$  and  $|-_p\rangle := \frac{1}{5}(4|0\rangle - 3|1\rangle)$ .

*Example 3.* For the ensemble composed of a uniform probability distribution and  $N = 4$  unitary channels given by  $\{U_i\} = \{\sqrt{\sigma_x}, \sqrt{\sigma_z}, \sqrt{H_y}, \sqrt{H_p}\}$ , in a discrimination task that allows for  $k = 3$  copies, general strategies outperform sequential strategies and sequential strategies outperform parallel strategies. Therefore, the maximal probabilities of success satisfy the strict hierarchy  $P^{\text{PAR}} < P^{\text{SEQ}} < P^{\text{GEN}}$ .

*Proof.* The proof follows from the direct application of the computer-assisted methods presented in Ref. 26. These methods allow us to find explicit and exact parallel/sequential/general testers, which attain a given success probability, ensuring, then, a lower bound for the maximal success probability for its class. In addition, we can obtain an exact parallel/sequential/general upper bound, given the SDP dual formulation. The code used to obtain the computer-assisted proof of the present example is publicly available in our online repository.<sup>52</sup>

The computed bounds for the maximal probability of successful discrimination are

$$\begin{aligned} \frac{9570}{1000} < P^{\text{PAR}} < \frac{9571}{1000} \\ &< \frac{9876}{1000} < P^{\text{SEQ}} < \frac{9877}{1000} \\ &< \frac{9881}{1000} < P^{\text{GEN}} < \frac{9882}{1000}, \end{aligned} \tag{D1}$$

showing the advantage of strategies that apply indefinite causal order over ordered ones and proving a strict hierarchy between strategies for the discrimination of a set of unitary channels.  $\square$

### APPENDIX E: PROOF OF THEOREM 4

In this section, we prove Theorem 4 from the main text, which concerns the inability of switch-like strategies to outperform sequential strategies on channel discrimination tasks that involve only unitary channels.

**Theorem 4.** *The action of the switch-like superchannel on  $k$  copies of a unitary channel can be equivalently described by a sequential circuit that acts on  $k$  copies of the same unitary channel.*

Consequently, in a discrimination task involving the ensemble  $\mathcal{E} = \{p_i, U_i\}_i$  composed of  $N$  unitary channels and some probability distribution and that allows for  $k$  copies, for every switch-like tester  $\{T_i^{\text{SL}}\}_i$ , there exists a sequential tester  $\{T_i^{\text{SEQ}}\}_i$  that attains the same probability of success according to

$$\sum_{i=1}^N p_i \text{Tr} \left( T_i^{\text{SL}} |U_i\rangle\langle U_i|^{\otimes k} \right) = \sum_{i=1}^N p_i \text{Tr} \left( T_i^{\text{SEQ}} |U_i\rangle\langle U_i|^{\otimes k} \right). \quad (\text{E1})$$

In order to provide a better intuition on this result, before presenting the formal definition of the switch-like process with  $k$ -slots and proving Theorem 4 in full generality, we present a proof for the  $k = 2$  case, which is illustrated in Fig. 2, in the main text.

For the case  $k = 2$ , the switch-like superchannel transforms a pair of unitary channels  $\{U_1, U_2\}$  into one unitary channel according to

$$\mathcal{W}_{\text{SL}}(U_1, U_2) := |0\rangle\langle 0|^c \otimes V_{02} (U_2 \otimes \mathbb{1}) V_{01} (U_1 \otimes \mathbb{1}) V_{00} + |1\rangle\langle 1|^c \otimes V_{12} (U_1 \otimes \mathbb{1}) V_{11} (U_2 \otimes \mathbb{1}) V_{10}, \quad (\text{E2})$$

where  $\mathbb{1}$  is the identity operator acting on the auxiliary system and  $V_{\pi i}$  are fixed unitary operators. Note that if  $U_1 = U_2 = U$ , we have

$$\mathcal{W}_{\text{SL}}(U, U) = |0\rangle\langle 0|^c \otimes V_{02} (U \otimes \mathbb{1}) V_{01} (U \otimes \mathbb{1}) V_{00} + |1\rangle\langle 1|^c \otimes V_{12} (U \otimes \mathbb{1}) V_{11} (U \otimes \mathbb{1}) V_{10}. \quad (\text{E3})$$

We now define a controlled version of the unitary operators  $V_{0i}$  as

$$V_{0i}^{\text{ctrl}} := |0\rangle\langle 0|^c \otimes V_{0i} + |1\rangle\langle 1|^c \otimes \mathbb{1} \quad (\text{E4})$$

and a controlled version of  $V_{1i}$  as

$$V_{1i}^{\text{ctrl}} := |0\rangle\langle 0|^c \otimes \mathbb{1} + |1\rangle\langle 1|^c \otimes V_{1i}. \quad (\text{E5})$$

We first note that due to the orthogonality of  $|0\rangle$  and  $|1\rangle$ , we have  $V_{1i}^{\text{ctrl}} V_{0i}^{\text{ctrl}} = |0\rangle\langle 0|^c \otimes V_{0i} + |1\rangle\langle 1|^c \otimes V_{1i}$ . Hence, a direct calculation shows that

$$\begin{aligned} V_{12}^{\text{ctrl}} V_{02}^{\text{ctrl}} (U \otimes \mathbb{1}) \cdot V_{11}^{\text{ctrl}} V_{01}^{\text{ctrl}} (U \otimes \mathbb{1}) \cdot V_{10}^{\text{ctrl}} V_{00}^{\text{ctrl}} &= (|0\rangle\langle 0|^c \otimes V_{02} + |1\rangle\langle 1|^c \otimes V_{12}) (U \otimes \mathbb{1}) \\ &\cdot (|0\rangle\langle 0|^c \otimes V_{01} + |1\rangle\langle 1|^c \otimes V_{11}) (U \otimes \mathbb{1}) \\ &\cdot (|0\rangle\langle 0|^c \otimes V_{00} + |1\rangle\langle 1|^c \otimes V_{10}) (U \otimes \mathbb{1}) \end{aligned} \quad (\text{E6})$$

$$= \mathcal{W}_{\text{SL}}(U, U). \quad (\text{E7})$$

This shows that, when  $U_1 = U_2 = U$ , a two-slot sequential circuit, which performs the operations  $V_{12}^{\text{ctrl}} V_{02}^{\text{ctrl}}$ ,  $V_{11}^{\text{ctrl}} V_{01}^{\text{ctrl}}$ , and  $V_{10}^{\text{ctrl}} V_{00}^{\text{ctrl}}$ , can perfectly simulate the two-slot switch-like superchannel. See Fig. 2 in the main text for an illustration.

Now, we extend this result to an arbitrary finite number of copies  $k$ .

**Definition 1** (switch-like superchannel). *Let  $\{\pi\}_\pi$ ,  $\pi \in \{0, \dots, k! - 1\}$ , be a set in which each integer  $\pi$  represents a permutation of the set  $\{1, \dots, k\}$  and  $\sigma_\pi : \{1, \dots, k\} \rightarrow \{1, \dots, k\}$  be the permutation function such that, after permutation  $\pi$ , the element  $i \in \{1, \dots, k\}$  is mapped to  $\sigma_\pi(i)$ . The  $k$ -slot switch-like superchannel acts on a set of  $k$  unitary operators  $\{U_i\}_{i=1}^k$ ,  $U_i : \mathcal{H}^{i_i} \rightarrow \mathcal{H}^{O_i}$  according to*

$$\mathcal{W}_{\text{SL}}(U_1, \dots, U_k) := \sum_{\pi=0}^{k!-1} |\pi\rangle\langle \pi| \otimes [V_{\pi k} (U_{\sigma_\pi(k)} \otimes \mathbb{1}^{\text{aux}}) V_{\pi(k-1)} (U_{\sigma_\pi(k-1)} \otimes \mathbb{1}^{\text{aux}}) V_{\pi(k-2)} \dots (U_{\sigma_\pi(1)} \otimes \mathbb{1}^{\text{aux}}) V_{\pi 0}], \quad (\text{E8})$$

where  $\{V_{\pi n}\}_{\pi n}$  is a set of unitary operators defined as

$$V_{\pi 0} : \mathcal{H}^{P_i} \otimes \mathcal{H}^{aux} \rightarrow \mathcal{H}^{I_{\sigma_n(i)}} \otimes \mathcal{H}^{aux}, \tag{E9}$$

$$V_{\pi n} : \mathcal{H}^{I_{\sigma_n(i)}} \otimes \mathcal{H}^{aux} \rightarrow \mathcal{H}^{O_{\sigma_n(i+1)}} \otimes \mathcal{H}^{aux} \quad \text{for } n \in \{1, \dots, k-1\}, \tag{E10}$$

$$V_{\pi k} : \mathcal{H}^{I_{\sigma_n(k)}} \otimes \mathcal{H}^{aux} \rightarrow \mathcal{H}^{F_i} \otimes \mathcal{H}^{aux}. \tag{E11}$$

Here, we have defined the switch-like superchannel only by its action on unitary channels, without explicitly stating how the switch-like superchannel acts on general quantum operations<sup>57,61</sup> or its process  $W^{SL} \in \mathcal{L}(\mathcal{H}^P \otimes \mathcal{H}^I \otimes \mathcal{H}^O \otimes \mathcal{H}^F)$ . In order to prove Theorem 4 and for the main purpose of this paper, knowing the action of switch-like superchannels only on unitary channels will be enough, but for the sake of concreteness, we also present an explicit process, which implements the switch-like superchannel. For that, we define the process  $W^{SL} := |U_{SL}\rangle\rangle\langle\langle U_{SL}|$ , where

$$U_{SL} := \bigoplus_{\pi} V_{\pi k} V_{\pi k-1} \dots V_{\pi 1} V_{\pi 0}. \tag{E12}$$

Following Lemma 1 in Ref. 29 (see also Theorem 2 of Ref. 51), one can verify that the process  $W^{SL}$  acts on unitary operators according to the switch-like superchannel, as presented in Definition 1.

*Proof.* We start our proof by defining the generalized controlled operation

$$V_n^{ctrl} := \sum_{\pi=0}^{k-1} |\pi\rangle\langle\pi| \otimes V_{\pi n} \quad \forall n \in \{0, \dots, k\}, \tag{E13}$$

which is a valid unitary operator since  $V_n^{ctrl} (V_n^{ctrl})^\dagger = \mathbb{1}$ . Now, note that, due to orthogonality of the vectors  $|\pi\rangle$ , we have

$$V_k^{ctrl} (U \otimes \mathbb{1}) V_{(k-1)}^{ctrl} (U \otimes \mathbb{1}) V_{(k-2)}^{ctrl} \dots (U \otimes \mathbb{1}) V_0^{ctrl} = W_{SL}(U, \dots, U). \tag{E14}$$

Hence, similarly to the  $k = 2$  case, a simple concatenation of the operators  $V_i^{ctrl}$  provides a  $k$ -slot sequential quantum circuit, which perfectly simulates the switch-like  $k$ -slot superchannel when all input unitary channels are equal.

Since every sequential quantum circuit can be written as an ordered process  $W^{SEQ} \in \mathcal{L}(\mathcal{H}^P \otimes \mathcal{H}^I \otimes \mathcal{H}^O \otimes \mathcal{H}^F)$ ,<sup>13</sup> when  $k$  identical unitary operators  $U$  are plugged into the process  $W^{SEQ}$ , the output operation is described by

$$W^{SEQ} * |U\rangle\rangle\langle\langle U|^{\otimes k} = W^{SL} * |U\rangle\rangle\langle\langle U|^{\otimes k}, \tag{E15}$$

where  $*$  is the link product and  $W^{SL}$  is a process associated with the switch-like superchannel. Hence, if

$$T_i^{SL} := \text{Tr}_{PF}[(\rho \otimes \mathbb{1}) W^{SL}(\mathbb{1} \otimes M_i)] \tag{E16}$$

is the tester associated with the switch-like strategy, then one can construct a sequential tester

$$T_i^{SEQ} = \text{Tr}_{PF}[(\rho \otimes \mathbb{1}) W^{SEQ}(\mathbb{1} \otimes M_i)] \tag{E17}$$

such that, for any unitary operator  $U$ , one has

$$\text{Tr}(T_i^{SL} |U\rangle\rangle\langle\langle U|^{\otimes k}) = \text{Tr}(T_i^{SEQ} |U\rangle\rangle\langle\langle U|^{\otimes k}), \tag{E18}$$

ensuring that there is always a sequential tester, which performs as well as any switch-like one. □

## APPENDIX F: UPPER BOUND

We start this section by stating a lemma from Ref. 58, which will be very useful for proving Theorem 5. We would also like to mention that step 2 of Ref. 59 and Theorem 3 of Ref. 48 are essentially equivalent to the lemma that we now state.

*Lemma 2* (Refs. 48, 58, and 59). Let  $E \in \mathcal{L}(\mathbb{C}^d)$  be a positive semidefinite operator,  $U_g \in \mathcal{L}(\mathbb{C}^d)$  be unitary operators, and  $G = \{U_g\}_g$  be a (compact Lie) group of unitary operators up to a global phase. It holds true that

$$E \leq \gamma_G \int_{\text{Haar}} U_g E U_g^\dagger dg, \tag{F1}$$

with

$$\gamma_G := \sum_{\mu \in \text{irrep}\{G\}} d_\mu \min(m_\mu, d_\mu), \tag{F2}$$

where  $\text{irrep}\{G\}$  is the set of all inequivalent irreducible representations (irreps) of  $G$  in  $\mathcal{L}(\mathbb{C}^d)$ ,  $d_\mu$  is the dimension of the linear space corresponding to the irrep  $\mu$ , and  $m_\mu$  is its associated multiplicity.

We are now in conditions to prove Theorem 5.

**Theorem 5** (upper bound for general strategies). Let  $\mathcal{E} = \{p_i, U_i\}_{i=1}^N$  be an ensemble composed of  $N$   $d$ -dimensional unitary channels and a uniform probability distribution. The maximal probability of successful discrimination of a general strategy with  $k$  copies is upper bounded by

$$P^{\text{GEN}} \leq \frac{1}{N} \gamma(d, k), \tag{F3}$$

where  $\gamma(d, k)$  is given by

$$\gamma(d, k) := \binom{k + d^2 - 1}{k} = \frac{(k + d^2 - 1)!}{k!(d^2 - 1)!}. \tag{F4}$$

*Proof.* The dual formulation of the channel discrimination problem [see Eq. (A16)] guarantees that if  $\overline{W}$  is the Choi operator of a non-signaling channel and the constraints

$$p_i |U_i\rangle\langle U_i|^{\otimes k} \leq \lambda \overline{W} \quad \forall i \in \{1, \dots, N\} \tag{F5}$$

are respected, the coefficient  $\lambda$  is an upper bound for the maximal probability of successfully discriminating the ensemble  $\mathcal{E} = \{p_i, U_i\}_{i=1}^N$  with  $k$  copies. Our proof consists in explicitly presenting the Choi operator of a non-signaling channel  $\overline{W}$  and a real number  $\lambda$  that respects the constraints of Eq. (F5) for any ensemble  $\mathcal{E} = \{p_i, U_i\}_{i=1}^N$  with  $p_i = \frac{1}{N}$ .

Consider the following ansatz:

$$\overline{W} := \int_{\text{Haar}} |U\rangle\langle U|^{\otimes k} dU, \tag{F6}$$

$$\lambda := \frac{1}{N} \sum_{\mu \in \text{irrep}\{SU(d)^{\otimes k}\}} d_\mu^2, \tag{F7}$$

where  $\text{irrep}\{SU(d)^{\otimes k}\}$  is the set of all inequivalent irreducible representations (irreps) of  $SU(d)^{\otimes k}$  and  $d_\mu$  is the dimension of the linear space corresponding to the irrep  $\mu$ .

Our first step is to show that  $\overline{W}$  is the Choi operator of a non-signaling channel. The operator  $\overline{W}$  is positive semidefinite because it is a convex mixture of positive semidefinite operators. In addition, for any  $d$ -dimensional unitary operator  $U$ , we have that  $\text{Tr}(|U\rangle\langle U|) = d$ . Hence, from the normalization of the Haar measure, we have that  $\text{Tr}(\overline{W}) = d^k$ .

The last step to certify that  $\overline{W}$  is indeed a non-signaling channel is then to guarantee that if  $j \in \{1, \dots, k\}$  stands for a slot of our process, we have that

$${}_j \overline{W} = {}_j 0, \overline{W} \quad \forall j \in \{1, \dots, k\}, \tag{F8}$$



where  $\mathcal{H}_{I_j}$  and  $\mathcal{H}_{O_j}$  correspond to the input and output space associated with the slot  $j$ , respectively. Since for any  $|U\rangle\langle U| \in \mathcal{L}(\mathcal{H}_{I_j} \otimes \mathcal{H}_{O_j})$ , we have

$$\text{Tr}_{O_j}(|U\rangle\langle U|) = \text{Tr}_{O_j}((1 \otimes U)|1\rangle\langle 1|(1 \otimes U^\dagger)) \tag{F9}$$

$$= \text{Tr}_{O_j}((1 \otimes U^\dagger U)|1\rangle\langle 1|) \tag{F10}$$

$$= \text{Tr}_{O_j}(|1\rangle\langle 1|) \tag{F11}$$

$$= 1_{I_j}, \tag{F12}$$

it holds that

$${}_j \overline{W} = \int_{\text{Haar } O_j} (|U\rangle\langle U|^{\otimes k}) dU \tag{F13}$$

$$= \int_{\text{Haar } O_i} (|U\rangle\langle U|_{I_1 O_1} \otimes |U\rangle\langle U|_{I_2 O_2} \otimes \cdots \otimes |U\rangle\langle U|_{I_k O_k}) dU \tag{F14}$$

$$= {}_{I_i O_i} \overline{W}. \tag{F15}$$

The last step of the proof is then to certify that for  $p_i = \frac{1}{N}$ , we indeed have

$$p_i |U_i\rangle\langle U_i|^{\otimes k} \leq \lambda \overline{W} \tag{F16}$$

for any set of unitary operators  $\{U\}_{i=1}^N$ . First, we observe that due to the left and right invariance of the Haar measure, for any unitary operator  $V \in SU(d)$ , the operator  $\overline{W}$  can be written as

$$\overline{W} := \int_{\text{Haar}} |U\rangle\langle U|^{\otimes k} dU \tag{F17}$$

$$= \int_{\text{Haar}} [(1 \otimes U)|1\rangle\langle 1|(1 \otimes U^\dagger)]^{\otimes k} dU \tag{F18}$$

$$= \int_{\text{Haar}} [(1 \otimes (UVU^\dagger)U)|1\rangle\langle 1|(1 \otimes U^\dagger(UVU^\dagger))]^{\otimes k} dU \tag{F19}$$

$$= \int_{\text{Haar}} [(1 \otimes U)|V\rangle\langle V|(1 \otimes U^\dagger)]^{\otimes k} dU. \tag{F20}$$

Additionally, the set  $\{1^{\otimes k} \otimes U^{\otimes k}\}_{U \in SU(d)}$  is a compact Lie group. Moreover, the dimensions of the linear spaces associated with the irreducible representation of  $\{1^{\otimes k} \otimes U^{\otimes k}\}_{U \in SU(d)}$  coincide with the dimension of the irreducible representations of  $\{U^{\otimes k}\}_{U \in SU(d)}$ . Since  $\min(d_\mu, m_\mu) \leq d_\mu$ , Lemma 2 ensures that

$$\frac{1}{N} |U_i\rangle\langle U_i|^{\otimes k} \leq \frac{1}{N} \left( \sum_{\mu \in \text{irrep}\{SU(d)^{\otimes k}\}} d_\mu^2 \right) \int_{\text{Haar}} [(1 \otimes U)|U_i\rangle\langle U_i|(1 \otimes U^\dagger)]^{\otimes k} dU \tag{F21}$$

$$= \frac{1}{N} \left( \sum_{\mu \in \text{irrep}\{SU(d)^{\otimes k}\}} d_\mu^2 \right) \int_{\text{Haar}} [(1 \otimes U)|1\rangle\langle 1|(1 \otimes U^\dagger)]^{\otimes k} dU \tag{F22}$$

$$= \lambda \overline{W}. \tag{F23}$$

Hence,  $\lambda := \frac{1}{N} (\sum_{\mu \in \text{irrep}\{SU(d)^{\otimes k}\}} d_\mu^2)$  is indeed an upper bound for the maximum probability of discriminating any set of  $N$   $d$ -dimensional unitary channels with  $k$  copies with general strategies.

We finish the proof by recognizing that, as proven in Ref. 60 [p. 57, Eq. (57)], we have the following identity:

$$\sum_{\mu \in \text{irrep}\{SU(d)^{\otimes k}\}} d_\mu^2 = \binom{k + d^2 - 1}{k}. \tag{F24}$$

□

## REFERENCES

- <sup>1</sup>R. Blahut, "Hypothesis testing and information theory," *IEEE Trans. Inf. Theory* **20**, 405–417 (1974).
- <sup>2</sup>S. L. K. Pond, S. D. W. Frost, and S. V. Muse, "HyPhy: Hypothesis testing using phylogenies," *Bioinformatics* **21**, 676–679 (2004).
- <sup>3</sup>J. Pearl, *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference* (Elsevier, 2014).
- <sup>4</sup>R. S. Nickerson, "Null hypothesis significance testing: A review of an old and continuing controversy," *Psychol. Methods* **5**, 241 (2000).
- <sup>5</sup>A. Shimbo, A. Soeda, and M. Muraio, "Equivalence determination of unitary operations," [arXiv:1803.11414](https://arxiv.org/abs/1803.11414) [quant-ph] (2018).
- <sup>6</sup>A. Soeda, A. Shimbo, and M. Muraio, "Optimal quantum discrimination of single-qubit unitary gates between two candidates," *Phys. Rev. A* **104**, 022422 (2021); [arXiv:2103.08208](https://arxiv.org/abs/2103.08208) [quant-ph].
- <sup>7</sup>V. Giovannetti, S. Lloyd, and L. Maccone, "Quantum metrology," *Phys. Rev. Lett.* **96**, 010401 (2006); [arXiv:quant-ph/0509179](https://arxiv.org/abs/quant-ph/0509179) [quant-ph].
- <sup>8</sup>L. Pezzè, A. Smerzi, M. K. Oberthaler, R. Schmied, and P. Treutlein, "Quantum metrology with nonclassical states of atomic ensembles," *Rev. Mod. Phys.* **90**, 035005 (2018); [arXiv:1609.01609](https://arxiv.org/abs/1609.01609) [quant-ph].
- <sup>9</sup>M. Hayashi, "Quantum hypothesis testing and discrimination of quantum states," in *Quantum Information* (Springer, Berlin, Heidelberg, 2006), pp. 69–91.
- <sup>10</sup>M. G. A. Paris, "Quantum estimation for quantum technology," *Int. J. Quantum Inf.* **7**, 125–137 (2009); [arXiv:0804.2981](https://arxiv.org/abs/0804.2981) [quant-ph].
- <sup>11</sup>G. Chiribella, G. M. D'Ariano, P. Perinotti, and M. F. Sacchi, "Efficient use of quantum resources for the transmission of a reference frame," *Phys. Rev. Lett.* **93**, 180503 (2004); [arXiv:quant-ph/0405095](https://arxiv.org/abs/quant-ph/0405095).
- <sup>12</sup>S. D. Bartlett, T. Rudolph, and R. W. Spekkens, "Reference frames, superselection rules, and quantum information," *Rev. Mod. Phys.* **79**, 555–609 (2007); [arXiv:quant-ph/0610030](https://arxiv.org/abs/quant-ph/0610030).
- <sup>13</sup>G. Chiribella, G. M. D'Ariano, and P. Perinotti, "Quantum circuit architecture," *Phys. Rev. Lett.* **101**, 060401 (2008); [arXiv:0712.1325](https://arxiv.org/abs/0712.1325) [quant-ph].
- <sup>14</sup>D. Deutsch and R. Jozsa, "Rapid solution of problems by quantum computation," *Proc. R. Soc. London, Ser. A* **439**, 553–558 (1992).
- <sup>15</sup>L. K. Grover, "A framework for fast quantum mechanical algorithms," in *Proceedings of the Thirtieth Annual ACM Symposium on Theory of Computing* (Association for Computing Machinery, New York, NY, 1998), pp. 53–62; [arXiv:quant-ph/9711043](https://arxiv.org/abs/quant-ph/9711043).
- <sup>16</sup>D. R. Simon, "On the power of quantum computation," in *Proceedings of the 35th Annual Symposium on Foundations of Computer Science* (IEEE, 1994), pp. 116–123.
- <sup>17</sup>A. Chefles, A. Kitagawa, M. Takeoka, M. Sasaki, and J. Twamley, "Unambiguous discrimination among oracle operators," *J. Phys. A: Math. Gen.* **40**, 10183–10213 (2007); [arXiv:quant-ph/0702245](https://arxiv.org/abs/quant-ph/0702245).
- <sup>18</sup>C. W. Helstrom, "Quantum detection and estimation theory," *J. Stat. Phys.* **1**, 231–252 (1969).
- <sup>19</sup>A. Acín, "Statistical distinguishability between unitary operations," *Phys. Rev. Lett.* **87**, 177901 (2001); [arXiv:quant-ph/0102064](https://arxiv.org/abs/quant-ph/0102064).
- <sup>20</sup>G. M. D'Ariano, P. Lo Presti, and M. G. A. Paris, "Using entanglement improves the precision of quantum measurements," *Phys. Rev. Lett.* **87**, 270404 (2001); [arXiv:quant-ph/0109040](https://arxiv.org/abs/quant-ph/0109040).
- <sup>21</sup>R. Duan, Y. Feng, and M. Ying, "Entanglement is not necessary for perfect discrimination between unitary operations," *Phys. Rev. Lett.* **98**, 100503 (2007); [arXiv:quant-ph/0601150](https://arxiv.org/abs/quant-ph/0601150).
- <sup>22</sup>G. Chiribella, G. M. D'Ariano, and P. Perinotti, "Memory effects in quantum channel discrimination," *Phys. Rev. Lett.* **101**, 180501 (2008); [arXiv:0803.3237](https://arxiv.org/abs/0803.3237) [quant-ph].
- <sup>23</sup>G. Chiribella, G. M. D'Ariano, and M. Roetteler, "Identification of a reversible quantum gate: Assessing the resources," *New J. Phys.* **15**, 103019 (2013); [arXiv:1306.0719](https://arxiv.org/abs/1306.0719) [quant-ph].
- <sup>24</sup>A. Bisio, G. Chiribella, G. M. D'Ariano, S. Facchini, and P. Perinotti, "Optimal quantum learning of a unitary transformation," *Phys. Rev. A* **81**, 032324 (2010); [arXiv:0903.0543](https://arxiv.org/abs/0903.0543) [quant-ph].
- <sup>25</sup>M. Sedlák, A. Bisio, and M. Ziman, "Optimal probabilistic storage and retrieval of unitary channels," *Phys. Rev. Lett.* **122**, 170502 (2019); [arXiv:1809.04552](https://arxiv.org/abs/1809.04552) [quant-ph].
- <sup>26</sup>J. Bavaresco, M. Muraio, and M. T. Quintino, "Strict hierarchy between parallel, sequential, and indefinite-causal-order strategies for channel discrimination," *Phys. Rev. Lett.* **127**, 200504 (2021); [arXiv:2011.08300](https://arxiv.org/abs/2011.08300) [quant-ph].
- <sup>27</sup>G. Chiribella, G. M. D'Ariano, P. Perinotti, and B. Valiron, "Quantum computations without definite causal structure," *Phys. Rev. A* **88**, 022318 (2013); [arXiv:0912.0195](https://arxiv.org/abs/0912.0195) [quant-ph].
- <sup>28</sup>M. Araújo, F. Costa, and Č. Brukner, "Computational advantage from quantum-controlled ordering of gates," *Phys. Rev. Lett.* **113**, 250402 (2014); [arXiv:1401.8127](https://arxiv.org/abs/1401.8127) [quant-ph].
- <sup>29</sup>W. Yokojima, M. T. Quintino, A. Soeda, and M. Muraio, "Consequences of preserving reversibility in quantum superchannels," *Quantum* **5**, 441 (2021); [arXiv:2003.05682](https://arxiv.org/abs/2003.05682) [quant-ph].
- <sup>30</sup>J. Barrett, R. Lorenz, and O. Oreshkov, "Cyclic quantum causal models," *Nat. Commun.* **12**, 885 (2021); [arXiv:2002.12157](https://arxiv.org/abs/2002.12157) [quant-ph].
- <sup>31</sup>G. Chiribella, G. M. D'Ariano, and P. Perinotti, "Theoretical framework for quantum networks," *Phys. Rev. A* **80**, 022339 (2009); [arXiv:0904.4483](https://arxiv.org/abs/0904.4483) [quant-ph].
- <sup>32</sup>M. Ziman, "Process positive-operator-valued measure: A mathematical framework for the description of process tomography experiments," *Phys. Rev. A* **77**, 062112 (2008); [arXiv:0802.3862](https://arxiv.org/abs/0802.3862) [quant-ph].
- <sup>33</sup>M. Ziman and M. Sedlák, "Single-shot discrimination of quantum unitary processes," *J. Mod. Opt.* **57**, 253–259 (2010); [arXiv:1003.1488](https://arxiv.org/abs/1003.1488) [quant-ph].
- <sup>34</sup>M. Araújo, C. Branciard, F. Costa, A. Feix, C. Giarmatzis, and Č. Brukner, "Witnessing causal nonseparability," *New J. Phys.* **17**, 102001 (2015); [arXiv:1506.03776](https://arxiv.org/abs/1506.03776) [quant-ph].
- <sup>35</sup>G. Gutoski and J. Watrous, "Toward a general theory of quantum games," in *STOC'07: Proceedings of the Thirty-Ninth Annual ACM Symposium on Theory of Computing* (Association for Computing Machinery, New York, NY, 2007), pp. 565–574; [arXiv:quant-ph/0611234](https://arxiv.org/abs/quant-ph/0611234).
- <sup>36</sup>D. Kretschmann and R. F. Werner, "Quantum channels with memory," *Phys. Rev. A* **72**, 062323 (2005); [arXiv:quant-ph/0502106](https://arxiv.org/abs/quant-ph/0502106).
- <sup>37</sup>O. Oreshkov, F. Costa, and Č. Brukner, "Quantum correlations with no causal order," *Nat. Commun.* **3**, 1092 (2012); [arXiv:1105.4464](https://arxiv.org/abs/1105.4464) [quant-ph].
- <sup>38</sup>J. Wechs, A. A. Abbott, and C. Branciard, "On the definition and characterisation of multipartite causal (non)separability," *New J. Phys.* **21**, 013027 (2019); [arXiv:1807.10557](https://arxiv.org/abs/1807.10557) [quant-ph].
- <sup>39</sup>A. Feix, M. Araújo, and Č. Brukner, "Quantum superposition of the order of parties as a communication resource," *Phys. Rev. A* **92**, 052326 (2015); [arXiv:1508.07840](https://arxiv.org/abs/1508.07840) [quant-ph].
- <sup>40</sup>P. A. Guérin, A. Feix, M. Araújo, and Č. Brukner, "Exponential communication complexity advantage from quantum superposition of the direction of communication," *Phys. Rev. Lett.* **117**, 100502 (2016); [arXiv:1605.07372](https://arxiv.org/abs/1605.07372) [quant-ph].

- <sup>41</sup>G. Chiribella, “Perfect discrimination of no-signalling channels via quantum superposition of causal structures,” *Phys. Rev. A* **86**, 040301(R) (2012); [arXiv:1109.5154](#) [quant-ph].
- <sup>42</sup>C. Branciard, M. Araújo, A. Feix, F. Costa, and Č. Brukner, “The simplest causal inequalities and their violation,” *New J. Phys.* **18**, 013008 (2016); [arXiv:1508.01704](#) [quant-ph].
- <sup>43</sup>J. Bavaresco, M. Araújo, Č. Brukner, and M. T. Quintino, “Semi-device-independent certification of indefinite causal order,” *Quantum* **3**, 176 (2019); [arXiv:1903.10526](#) [quant-ph].
- <sup>44</sup>M. T. Quintino, Q. Dong, A. Shimbo, A. Soeda, and M. Muraio, “Reversing unknown quantum transformations: Universal quantum circuit for inverting general unitary operations,” *Phys. Rev. Lett.* **123**, 210502 (2019); [arXiv:1810.06944](#) [quant-ph].
- <sup>45</sup>J. Wechs, H. Dourdent, A. A. Abbott, and C. Branciard, “Quantum circuits with classical versus quantum control of causal order,” *PRX Quantum* **2**, 030335 (2021); [arXiv:2101.08796](#) [quant-ph].
- <sup>46</sup>G. Chiribella, G. M. D’Ariano, P. Perinotti, and M. F. Sacchi, “Covariant quantum measurements that maximize the likelihood,” *Phys. Rev. A* **70**, 062105 (2004); [arXiv:quant-ph/0403083](#).
- <sup>47</sup>A. Hayashi, T. Hashimoto, and M. Horibe, “Extended quantum color coding,” *Phys. Rev. A* **71**, 012326 (2005); [arXiv:quant-ph/0409173](#).
- <sup>48</sup>G. Chiribella, G. M. D’Ariano, P. Perinotti, and M. F. Sacchi, “Maximum likelihood estimation for a group of physical transformations,” *Int. J. Quantum Inf.* **4**, 453–472 (2006); [arXiv:quant-ph/0507007](#).
- <sup>49</sup>A. W. Harrow, A. Hassidim, D. W. Leung, and J. Watrous, “Adaptive versus nonadaptive strategies for quantum channel discrimination,” *Phys. Rev. A* **81**, 032339 (2010); [arXiv:0909.0256](#) [quant-ph].
- <sup>50</sup>Ä. Baumeler and S. Wolf, “The space of logically consistent classical processes without causal order,” *New J. Phys.* **18**, 013036 (2016); [arXiv:1507.01714](#) [quant-ph].
- <sup>51</sup>M. Araújo, A. Feix, M. Navascués, and Č. Brukner, “A purification postulate for quantum mechanics with indefinite causal order,” *Quantum* **1**, 10 (2017); [arXiv:1611.08535](#) [quant-ph].
- <sup>52</sup>M. Túlio Quintino and J. Bavaresco, Code to accompany: “Unitary channel discrimination beyond group structures: Advantages of sequential and indefinite-causal-order-strategies” (2021). [https://github.com/mtcq/unitary\\_channel\\_discrimination](https://github.com/mtcq/unitary_channel_discrimination).
- <sup>53</sup>M. T. Quintino, Q. Dong, A. Shimbo, A. Soeda, and M. Muraio, “Probabilistic exact universal quantum circuits for transforming unitary operations,” *Phys. Rev. A* **100**, 062339 (2019); [arXiv:1909.01366](#) [quant-ph].
- <sup>54</sup>A. Bisio, G. M. D’Ariano, P. Perinotti, and M. Sedláč, “Optimal processing of reversible quantum channels,” *Phys. Lett. A* **378**, 1797–1808 (2014); [arXiv:1308.3254](#) [quant-ph].
- <sup>55</sup>Here,  $\sqrt{W}^{-1}$  stands for the inverse of  $\sqrt{W}$  on its range. If the operator  $W$  is not full rank, the composition  $W W^{-1} =: \Pi_W$  is not the identity  $\mathbb{1}$  but the projector onto the subspace spanned by the range of  $\sqrt{W}$ . Due to this technicality, when the operator  $W$  is not full rank, we should define the measurements as  $M_U^{I'O} := \sqrt{W}^{-1'I'O} T_U^{I'O} \sqrt{W}^{-1'I'O} + \frac{1}{N} (\mathbb{1} - W W^{-1})$ . With that, the proof written here also applies to the case where the operator  $W$  is not full rank.
- <sup>56</sup>Indeed, two diagonalisable operators  $A$  and  $B$  commute if and only if they are diagonal in the same basis. Now, if  $A =: \sum_i \alpha_i |i\rangle\langle i|$ , its positive semidefinite square root is also diagonal in the same basis, in particular,  $\sqrt{A} = \sum_i \sqrt{\alpha_i} |i\rangle\langle i|$ . Hence, if  $A$  commutes with  $B$ ,  $\sqrt{A}$  also commutes with  $B$ .
- <sup>57</sup>Interestingly, it can be proved that for the  $k = 2$ -slot case, the action of the standard quantum switch superchannel, i.e.,  $V_{\text{ms}} = \mathbb{1}$ , on unitary channels uniquely defines its action on general operations.<sup>61</sup> The possibility of extending this result for general switch-like superchannels is still open, but the existence of general switch-like superchannels is ensured by the construction presented via Eq. (E12).
- <sup>58</sup>T. Hashimoto, A. Hayashi, M. Hayashi, and M. Horibe, “Unitary-process discrimination with error margin,” *Phys. Rev. A* **81**, 062327 (2010); [arXiv:0912.2610](#) [quant-ph].
- <sup>59</sup>J. von Korff and J. Kempe, “Quantum advantage in transmitting a permutation,” *Phys. Rev. Lett.* **93**, 260502 (2004).
- <sup>60</sup>I. Schur, “Ueber eine klasse von matrizen, die sich einer gegebenen matrix zuordnen lassen,” Ph.D. thesis, Friedrich-Wilhelms-Universität zu Berlin, 1901.
- <sup>61</sup>Q. Dong, M. T. Quintino, A. Soeda, and M. Muraio, “The quantum switch is uniquely defined by its action on unitary operations,” [arXiv:2106.00034](#) [quant-ph] (2021).