Strict Hierarchy between Parallel, Sequential, and Indefinite-Causal-Order Strategies for Channel Discrimination

Jessica Bavaresco,[1,*] Mio Murao,[2,4] and Marco Túlio Quintino[1,3,2]

[1]Institute for Quantum Optics and Quantum Information (IQOQI), Austrian Academy of Sciences, Boltzmanngasse 3, A-1090 Vienna, Austria

[2]Department of Physics, Graduate School of Science, The University of Tokyo, Hongo 7-3-1, Bunkyo-ku, Tokyo 113-0033, Japan

[3]Vienna Center for Quantum Science and Technology (VCQ), Faculty of Physics, University of Vienna, Boltzmanngasse 5, A-1090 Vienna, Austria

[4]Trans-scale Quantum Science Institute, The University of Tokyo, Hongo 7-3-1, Bunkyo-ku, Tokyo 113-0033, Japan

(Received 27 November 2020; accepted 7 October 2021; published 12 November 2021)

We present an instance of a task of minimum-error discrimination of two qubit-qubit quantum channels for which a sequential strategy outperforms any parallel strategy. We then establish two new classes of strategies for channel discrimination that involve indefinite causal order and show that there exists a strict hierarchy among the performance of all four strategies. Our proof technique employs a general method of computer-assisted proofs. We also provide a systematic method for finding pairs of channels that showcase this phenomenon, demonstrating that the hierarchy between strategies is not exclusive to our main example.

DOI: 10.1103/PhysRevLett.127.200504

The discrimination of physical operations is a task related to the elementary ability to experimentally distinguish among different dynamics, or time evolutions, to which physical systems are subjected. From a fundamental perspective, the capacity to test and discriminate between hypotheses lies at the core of statistical analysis and constitutes one of the pillars of the scientific method. From a more practical standpoint, the discrimination of physical operations comes into play in problems such as the identification of cause-effect relations [1], computational complexity analysis of oracle-based algorithms [2–5], certification of circuit elements [6–8], and in tasks related to metrology [9].

Within the context of quantum physics, pioneering work connecting hypothesis testing with discrimination of quantum objects dates back to Holevo and Helstrom [10,11]. Most of these initial results concern the discrimination of quantum states, while developing concepts and methods relevant to the fundamental problem of discriminating between quantum operations—a task also referred to as *quantum channel discrimination*. A plethora of interesting results on this topic has since been demonstrated [12–20]. In channel discrimination tasks that allow multiple interactions, different discrimination strategies become relevant, the most common being parallel and sequential (i.e., adaptive). For pairs of unitary channels, optimal minimum-error discrimination is achieved by parallel schemes [21]. The advantage of sequential strategies first became apparent for a task regarding two qubit-ququart entanglement-breaking channels [22], which cannot transmit quantum information [23]. Recent results also indicate this advantage [8,24–26], including its numerical observation in a discrimination task of two qubit-qubit channels [27].

In the related task of discriminating two nonsignaling bipartite channels, a more general strategy constructed from the quantum switch [28], involving indefinite causal order, provides an advantage over causal (sequential and parallel) strategies, even allowing for perfect discrimination [29]. This phenomenon hints that indefinite causal order could be useful for tasks of channel discrimination, similarly to how it is advantageous for quantum computation [30], communication complexity [31,32], and inversion of unknown unitary operations [33].

In this Letter, we have two main contributions to the study of channel discrimination. The first is the rigorous demonstration of the advantage of sequential over parallel strategies. Our example concerns the simplest scenario of channel discrimination—between two qubit-qubit channels using two copies—and channels with nonzero quantum capacity—an amplitude-damping and a bit-flip channel. The second is the demonstration that strategies involving indefinite causal order can outperform parallel and sequential ones for the same task of channel discrimination. In order to do so, we define two new classes of discrimination strategies that make use of indefinite causal order—which we call separable and general. Together, these results constitute a strict hierarchy between four strategies of channel discrimination. To prove our results, we develop a method of computer-assisted proofs.

The task of minimum-error channel discrimination works as follows: With probability p_i, Alice is given an unknown quantum channel $\tilde{C}_i : \mathcal{L}(\mathcal{H}^I) \to \mathcal{L}(\mathcal{H}^O)$, drawn from an ensemble $\mathcal{E} = \{p_j, \tilde{C}_j\}_{j=1}^N$ that is known to her. Being allowed to use a finite number of copies of $\tilde{C}_i$, her task is to determine which channel she received by performing operations on it and guessing the value of $i \in \{1, \ldots, N\}$. If Alice is allowed to use one copy of the channel she received, the most general quantum operations she could apply are to send part of a potentially entangled state $\rho \in \mathcal{L}(\mathcal{H}^I \otimes \mathcal{H}^{\text{aux}})$ through the channel $\tilde{C}_i$, and jointly measure the output with a positive operator-valued measure (POVM) $M = \{M_a\}, M_a \in \mathcal{L}(\mathcal{H}^O \otimes \mathcal{H}^{\text{aux}})$, announcing the outcome of her measurement as her guess. Then, her probability of correctly guessing the value i is given by $p_{\text{succ}} \coloneqq \sum_{i=1}^N p_i \operatorname{Tr}[(\tilde{C}_i \otimes \tilde{\mathbb{1}})(\rho)M_i]$, where $\tilde{\mathbb{1}}$ is the identity map on $\mathcal{L}(\mathcal{H}^{\text{aux}})$. Alice can improve her chances by optimizing her operations based on her knowledge of the ensemble. Her maximal probability of success is then given by $p^*_{\text{succ}} \coloneqq \max_{\{\rho,M\}} p_{\text{succ}}$.

By means of the Choi-Jamiołkowski isomorphism (see Ref. [34]) [35–37] one can represent a channel $\tilde{C}$ (a completely positive, trace-preserving map) as a positive semidefinite operator $C \in \mathcal{L}(\mathcal{H}^I \otimes \mathcal{H}^O)$, called "Choi operator," that satisfies $C \geq 0$ and $\operatorname{Tr}_O C = \mathbb{1}^I$, where Tr_X denotes the partial trace over $\mathcal{H}^X$ and $\mathbb{1}^X$ the identity operator on $\mathcal{H}^X$. Using Choi operators and the link product (see Ref. [38]) [39] to represent a concatenation of operations, we can rewrite the maximal probability of success as $p^*_{\text{succ}} = \max_{\{\rho,M\}} \sum_{i=1}^N p_i C_i * \rho * M_i^T$.

In principle, Alice could apply a more general strategy by constructing the most general map that takes a channel to a set of probability distributions. This map is defined by the most general set of operators $T = \{T_i\}_{i=1}^N, T_i \in \mathcal{L}(\mathcal{H}^I \otimes \mathcal{H}^O)$ that respect the relation $p(i|C) = \operatorname{Tr}(CT_i)$ for all Choi operators of channels C, where $\{p(i|C)\}$ is a probability distribution. This set of operators has been characterized as a *general tester*, a set $T = \{T_i\}$ that satisfies $T_i \geq 0\, \forall i$ and $\sum_i T_i = \sigma \otimes \mathbb{1}^O$, where $\sigma \in \mathcal{L}(\mathcal{H}^I)$ is a quantum state [17,39] (see also Supplemental Material [40]). Remarkably, it has been shown that every general tester has a *quantum realization*. Namely, from any general tester a state and measurement can always be constructed, in such a way that each tester element is recovered as $T_i = \rho * M_i^T$. This mathematical equivalence allows for a simpler characterization of Alice's strategies, who can now optimize over general testers T to achieve a maximal probability of success that is equivalently given by $p^*_{\text{succ}} = \max_{\{T\}} \sum_{i=1}^N p_i \operatorname{Tr}(T_i C_i)$ (see Ref. [50]).

Now let us analyze the more interesting case where Alice has access to two copies of the channel C_i. With two copies, Alice has the freedom of choosing how to concatenate these channels in order to gain more information about them.

One option is to apply the two copies of the unknown channel in parallel, by sending a joint state $\rho \in \mathcal{L}(\mathcal{H}^{I_1} \otimes \mathcal{H}^{I_2} \otimes \mathcal{H}^{\text{aux}})$ through both copies of C_i and then measuring

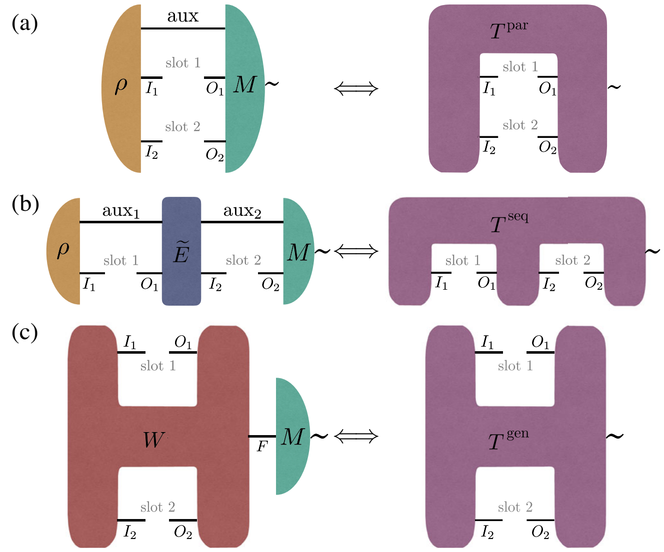

FIG. 1. Schematic representation of the realization of every two-copy (a) parallel tester T^{par} with a state ρ and a POVM M; (b) sequential tester T^{seq} with a state ρ, a channel $\tilde{E}$, and a POVM M; and (c) general tester T^{gen} with a process matrix W and a POVM M.

the output with a POVM $M = \{M_i\}, M_i \in \mathcal{L}(\mathcal{H}^{O_1} \otimes \mathcal{H}^{O_2} \otimes \mathcal{H}^{\text{aux}})$, where $\mathcal{H}^{I_1}$ ($\mathcal{H}^{I_2}$) represents the input space of the first (second) copy of C_i, and equivalently for the output spaces. Just like in the one-copy case, this strategy can be expressed by a two-copy parallel tester, a set of operators $T^{\text{par}} = \{T_i^{\text{par}}\}$ that satisfy linear constraints defined below, and that always accepts a quantum realization according to $T_i^{\text{par}} = \rho * M_i^T$ [39] [see Fig. 1(a)]. In the following we use the notation ${}_X A \coloneqq \operatorname{Tr}_X A \otimes (\mathbb{1}^X / d_X)$ and $d_X = \dim(\mathcal{H}^X)$.

Definition 1 (two-copy parallel tester).—A parallel tester is a set of linear operators $T^{\text{par}} = \{T_i^{\text{par}}\}_{i=1}^N, T_i^{\text{par}} \in \mathcal{L}(\mathcal{H}^{I_1 O_1 I_2 O_2})$ such that $T_i^{\text{par}} \geq 0$, $\forall i$ and $W^{\text{par}} \coloneqq \sum_i T_i^{\text{par}}$ satisfies $\operatorname{Tr}(W^{\text{par}}) = d_{O_1} d_{O_2}$, and

$$W^{\text{par}} = {}_{O_1 O_2} W^{\text{par}}. \tag{1}$$

W^{par} is called a parallel process.

More generally, Alice could use her two copies of C_i in a sequential manner, first sending a state $\rho \in \mathcal{L}(\mathcal{H}^{I_1} \otimes \mathcal{H}^{\text{aux}_1})$ through the first copy of C_i, next applying to the output a general channel $\tilde{E}: \mathcal{L}(\mathcal{H}^{O_1} \otimes \mathcal{H}^{\text{aux}_1}) \to \mathcal{L}(\mathcal{H}^{I_2} \otimes \mathcal{H}^{\text{aux}_2})$, then sending part of the output of channel $\tilde{E}$ through the second copy of C_i, and finally measuring the output with a POVM $M = \{M_i\}, M_i \in \mathcal{L}(\mathcal{H}^{O_2} \otimes \mathcal{H}^{\text{aux}_2})$. Analogously to the parallel case, the tester associated to this strategy—a sequential tester $T^{\text{seq}} = \{T_i^{\text{seq}}\}$ expressed as $T_i^{\text{seq}} = \rho * E * M_i^T$, where $E \in \mathcal{L}(\mathcal{H}^{O_1} \otimes \mathcal{H}^{\text{aux}_1} \otimes \mathcal{H}^{I_2} \otimes \mathcal{H}^{\text{aux}_2})$ is the Choi operator of map $\tilde{E}$, meaning it can always be realized by quantum circuits [39] [see Fig. 1(b)]—has been characterized as

Definition 2 (two-copy sequential tester).—A sequential tester is a set of linear operators $T^{\mathrm{seq}} = \{T_i^{\mathrm{seq}}\}_{i=1}^N, T_i^{\mathrm{seq}} \in \mathcal{L}(\mathcal{H}^{I_1O_1I_2O_2})$ such that $T_i^{\mathrm{seq}} \geq 0,\ \forall i$ and $W^{\mathrm{seq}} \coloneqq \sum_i T_i^{\mathrm{seq}}$ satisfies $\mathrm{Tr}(W^{\mathrm{seq}}) = d_{O_1}d_{O_2}$, and

$$W^{\mathrm{seq}} = {}_{O_2}W^{\mathrm{seq}} \tag{2}$$

$${}_{I_2O_2}W^{\mathrm{seq}} = {}_{O_1I_2O_2}W^{\mathrm{seq}}. \tag{3}$$

W^{seq} is called a sequential process.

Parallel and sequential strategies have long been regarded as the most general strategies for channel discrimination. We now propose more general strategies than sequential ones, that arise from the following reasoning: In the same fashion of the definition of the general one-copy tester, one may define a *general two-copy tester* as the most general set of operators $T^{\mathrm{gen}} = \{T_i^{\mathrm{gen}}\}$ that map a pair of channels, represented by their Choi operators $C_A \in \mathcal{L}(\mathcal{H}^{I_1} \otimes \mathcal{H}^{O_1})$ and $C_B \in \mathcal{L}(\mathcal{H}^{I_2} \otimes \mathcal{H}^{O_2})$, to a valid probability distribution according to $p(i|C_A, C_B) = \mathrm{Tr}[(C_A \otimes C_B)T_i^{\mathrm{gen}}]$. It is shown in the Supplemental Material [40] that this definition is equivalent to

Definition 3 (two-copy general tester).—A general tester is a set of linear operators $T^{\mathrm{gen}} = \{T_i^{\mathrm{gen}}\}_{i=1}^N, T_i^{\mathrm{gen}} \in \mathcal{L}(\mathcal{H}^{I_1O_1I_2O_2})$ such that $T_i^{\mathrm{gen}} \geq 0,\ \forall i$ and $W^{\mathrm{gen}} \coloneqq \sum_i T_i^{\mathrm{gen}}$ satisfies $\mathrm{Tr}(W^{\mathrm{gen}}) = d_{O_1}d_{O_2}$, and

$${}_{I_1O_1}W^{\mathrm{gen}} = {}_{I_1O_1O_2}W^{\mathrm{gen}} \tag{4}$$

$${}_{I_2O_2}W^{\mathrm{gen}} = {}_{O_1I_2O_2}W^{\mathrm{gen}} \tag{5}$$

$$W^{\mathrm{gen}} = {}_{O_1}W^{\mathrm{gen}} + {}_{O_2}W^{\mathrm{gen}} - {}_{O_1O_2}W^{\mathrm{gen}}. \tag{6}$$

W^{gen} is called a general process.

A general tester can be seen as the most general transformation that acts globally on a pair of independent channels, extracting probability distributions. In this sense, a general tester is to a pair of channels as a POVM is to a pair of states, and it can be analogously interpreted as a "global measurement" of a pair of channels. Different from parallel and sequential testers, the definition of general testers does not take into account the order in which the channels may be acted upon.

Both parallel and sequential processes are particular cases of general processes [see Fig. 2(b)]. Nevertheless, the formalism of process matrices shows that there are general processes that do not respect a definite causal order [51,52]—which is defined as a process being parallel, sequential, or a classical mixture of sequential processes, called "causally separable" processes, motivating the definition of our final class of testers:

Definition 4 (two-copy separable tester).—A separable tester is a set of linear operators $T^{\mathrm{sep}} = \{T_i^{\mathrm{sep}}\}_{i=1}^N, T_i^{\mathrm{sep}} \in \mathcal{L}(\mathcal{H}^{I_1O_1I_2O_2})$ such that $T_i^{\mathrm{sep}} \geq 0,\ \forall i$ and $W^{\mathrm{sep}} \coloneqq \sum_i T_i^{\mathrm{sep}}$ satisfies $\mathrm{Tr}(W^{\mathrm{sep}}) = d_{O_1}d_{O_2}$ and

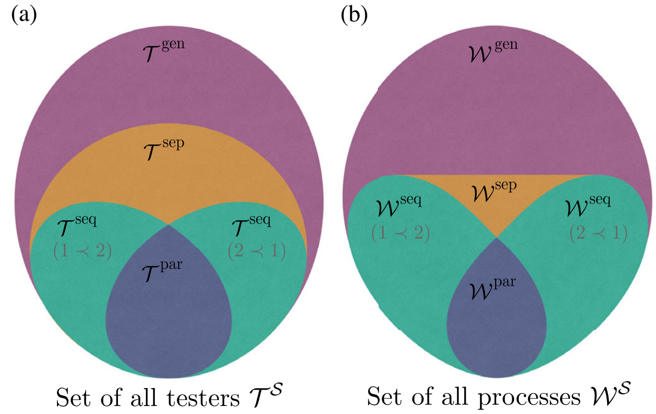

FIG. 2. Graphical representation of the nesting relations between (a) the sets of all testers $\mathcal{T}^{\mathcal{S}} \coloneqq [T^{\mathcal{S}}; T^{\mathcal{S}} = \{T_i^{\mathcal{S}}\}]$ and (b) the sets of all processes $\mathcal{W}^{\mathcal{S}} \coloneqq [W^{\mathcal{S}}; W^{\mathcal{S}} = \sum_i T_i^{\mathcal{S}}]$, where $\mathcal{S}$ represents parallel, sequential, separable, or general strategies.

$$W^{\mathrm{sep}} = qW^{1\prec 2} + (1-q)W^{2\prec 1}, \tag{7}$$

where $0 \leq q \leq 1$ and $W^{1\prec 2(2\prec 1)}$ is a sequential process with slot 1(2) coming before slot 2(1). W^{sep} is called a separable process.

In our terminology, the set of separable processes is the convex hull of the set of sequential processes whose slots follow orders $1 \prec 2$ and $2 \prec 1$, while parallel processes are those at the intersection of these sets [see Fig. 2(b)].

The definition of separable processes came from the idea that one could plug two different channels C_A and C_B in the two slots of process W^{sep} implementing a mixture of a process that applies channel C_A before C_B with one that applies C_B before C_A. One could then expect that this classical mixture of orders should not be relevant for a problem where the two channels being plugged into the separable tester are identical: two copies of C_i. Nonetheless, we show that separable testers indeed provide an advantage over sequential testers, which hints at a more complicated structure of separable *testers* than of separable *processes* themselves. This advantage implies that separable testers cannot be simply realized by ordered circuits and classical randomness, and that the set of separable testers is strictly larger than the convex hull of the set of sequential testers with different orders [see Fig. 2(a)].

With our unified framework for channel discrimination at hand, we can now define the maximal probability of success under each of our four strategies by allowing Alice to optimize over different classes of testers. The maximal probability of successful discrimination of a channel ensemble $\mathcal{E} = \{p_i, C_i\}$ using two copies under strategy $\mathcal{S}$ then reads as

$$P^{\mathcal{S}} \coloneqq \max_{\{T^{\mathcal{S}}\}} \sum_{i=1}^{N} p_i \mathrm{Tr}\,(T_i^{\mathcal{S}} C_i^{\otimes 2}), \tag{8}$$

where $\mathcal{S}$ represents parallel, sequential, separable, or general strategies. It is clear that these four strategies—parallel, sequential, separable, and general—form a hierarchy since each set of testers is a superset of the previous one, implying the relation $P^{\text{par}} \leq P^{\text{seq}} \leq P^{\text{sep}} \leq P^{\text{gen}}$ for any fixed ensemble. We show that, in fact, all these three inequalities can be simultaneously strictly satisfied.

To compute the values of $P^{\mathcal{S}}$, we phrase the optimization problems that define it in terms of semidefinite programming (SDP). Essentially,

$$\begin{aligned} &\text{given } \{p_i, C_i\} \\ &\text{maximize } \sum_i p_i \mathrm{Tr}(T_i^{\mathcal{S}} C_i^{\otimes 2}) \\ &\text{subject to } \{T_i^{\mathcal{S}}\} \text{ is a tester with strategy } \mathcal{S}. \end{aligned} \tag{9}$$

This optimization can be equivalently solved by its dual problem:

$$\begin{aligned} &\text{Given } \{p_i, C_i\} \\ &\text{minimize } \lambda \\ &\text{subject to } p_i C_i^{\otimes 2} \leq \lambda \bar{W}^{\mathcal{S}} \quad \forall i, \end{aligned} \tag{10}$$

where $\bar{W}^{\mathcal{S}}$ lies in the dual affine of the set of processes $\mathcal{W}^{\mathcal{S}}$ (see Ref. [53]) (see the Supplemental Material [40]).

SDPs can be solved by efficient numerical packages which, despite being in practice accurate, suffer from imprecision that arises from the use of floating-point variables [57,58]. In order to overcome this issue, we provide in the Supplemental Material [40] an algorithm for computer-assisted proofs (see Refs. [59,60] for other examples). Using our method, we obtain rigorous upper and lower bounds for $P^{\mathcal{S}}$, arriving at a result that has the same mathematical rigor of an analytical proof.

Theorem 1.—In the simplest instance of a channel discrimination task using $k = 2$ copies, i.e., discrimination between $N = 2$ qubit-qubit channels, there exist ensembles for which the maximal probability of successful discrimination of parallel, sequential, separable, and general strategies obey the strict hierarchy

$$P^{\text{par}} < P^{\text{seq}} < P^{\text{sep}} < P^{\text{gen}}. \tag{11}$$

Sketch of the proof.—The proof is constructive and considers the channel ensemble composed by $p_1 = p_2 = 1/2$, an amplitude-damping channel $\tilde{C}_{\text{AD}}$ (see Ref. [61]) with parameter $\gamma = 67/100$, and a bit-flip channel $\tilde{C}_{\text{BF}}$ (see Ref. [62]) with parameter $\eta = 87/100$. We start by applying standard numerical packages to solve the primal SDP [Eq. (9)] and obtain an ansatz for the optimal tester of each strategy. From the numerically imperfect ansatz, we construct a valid tester, following Algorithm 2 in the Suppl. Material. We then compute probabilities of success with these valid testers, which provide rigorous lower bounds for the maximal probabilities of success. To calculate rigorous upper bounds, we repeat this procedure, now taking as ansatz the numerical solutions of the dual problems [Eq. (10)] for dual affine processes, and following Algorithm 1 in the Supplemental Material [40]. Applying this method, we computed the following bounds: $(8346/10000) < P^{\text{par}} < (8347/10000)$, $(8446/10000) < P^{\text{seq}} < (8447/10000)$, $(8486/10000) < P^{\text{sep}} < (8487/10000)$, and $(8514/10000) < P^{\text{gen}} < (8515/10000)$. The clear gap between strategies concludes the proof. ■

Similar gaps can also be found for different ensembles of amplitude-damping and bit-flip channels, and also for two amplitude-damping channels, a problem which has been previously studied [8,24–27]. Moreover, this phenomenon is not particular to these channels. We have constructed a simple method of sampling pairs of channels that present a gap between all four strategies, for the case of qubit-qubit channels, in approximately 94% of the rounds (see Supplemental Material [40]).

Having demonstrated the theoretical advantage of these strategies, we would now like to discuss their potential implementation. As already mentioned, for parallel and sequential strategies, it is known that from every tester one can construct, in an algorithmic manner, a state, channel, and measurement that constitute its quantum realization [39]. Therefore, these testers can be physically implemented with quantum circuits, as depicted in Figs. 1(a) and 1(b).

For general testers, given a tester $T^{\text{gen}} = \{T_i^{\text{gen}}\}$, it can be realized by a process $W \coloneqq \sum_{i=1}^{N} T_i^{\text{gen}} \otimes |i\rangle\langle i|^F \in \mathcal{L}(\mathcal{H}^{I_1 O_1 I_2 O_2} \otimes \mathcal{H}^F)$, where $\mathcal{H}^F$ represents the space of a system in the common future of slots 1 and 2 of T^{gen}, and a POVM $M = \{M_i\}_i$, $M_i = |i\rangle\langle i| \in \mathcal{H}^F$. Each general tester element is recovered by $T_i^{\text{gen}} = W * M_i^T$. However, a *quantum* realization of T^{gen} would then depend on the ability of physically implementing any process W, as depicted in Fig. 1(c). Unfortunately, at this point, the physical implementation of general processes remains an open question.

For separable testers, however, a physical implementation is known. Similar to the general case, every separable tester can be recovered by a process $W \coloneqq \sum_{i=1}^{N} T_i^{\text{sep}} \otimes |i\rangle\langle i|^F$ and a POVM $\{M_i\}$, $M_i \coloneqq |i\rangle\langle i|^F$. However, when constructed from separable testers, the process W always satisfies the condition that $\mathrm{Tr}_F W = \sum_{i=1}^{N} T_i^{\text{sep}} = W^{\text{sep}}$, that is, they are (potentially nonseparable) processes that become separable when the future space is traced out. Such processes always lead to separable strategies, and can be used to realize every separable tester. Remarkably, these processes have recently been shown to be realized by circuits that employ a coherent quantum control of causal orders [63], implying that all separable strategies, including the ones that we have shown to be advantageous, can be physically implemented. One example of such a process that only leads to separable testers is the well-studied quantum switch [28]. Nevertheless, we have not been able to construct a discrimination task for which the testers generated by the quantum switch specifically were advantageous.

Conclusions.—We have demonstrated a new example of the advantage of sequential over parallel strategies for a task of minimum-error discrimination between two qubit-qubit channels. We also established two new classes of strategies that involve indefinite causal order and showed that they can outperform causal ones. Moreover, we proved a strict hierarchy between these four classes of strategies. Our main example concerns the discrimination of an amplitude-damping and a bit-flip channel; however, we showed that this phenomenon is not unique, by presenting a simple method of constructing pairs of channels that, with very high probability, respect this strict hierarchy. The main technique we developed is a method of computer-assisted proofs that finds immediate application in a plethora of physics problems that currently rely on numerical optimization. We hope this method can contribute to paving the way to more rigorous numerical proofs in quantum information science. It is furthermore our hope that our demonstration of the theoretical advantage of indefinite causal order for channel discrimination will further motivate the investigation of potential implementations of general processes.

The supporting code for this article are openly available from [64].

We are thankful to Alastair Abbott and Simon Milz for interesting discussions and to Mateus Araújo, Nicolai Friis, and Cyril Branciard for comments on the manuscript. J. B. acknowledges the Austrian Science Fund (FWF) through the START Project No. Y879-N27 and the Zukunftskolleg Project No. ZK03. M. M. is supported by the MEXT Quantum Leap Flagship Program (MEXT Q-LEAP) Grants No. JPMXS0118069605 and No. JPMXS0120351339 and by the Japan Society for the Promotion of Science (JSPS) through KAKENHI Grants No. 17H01694, No. 18H04286, and No. 21H03394. M. T. Q. acknowledges the Austrian Science Fund (FWF) through the SFB project BeyondC (Subproject No. 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 has received funding from the European Unions Horizon 2020 research and innovation program under the Marie Skłodowska-Curie Grant Agreement No. 801110. It reflects only the authors' view; the EU Agency is not responsible for any use that may be made of the information it contains. The Erwin Schrödinger Center for Quantum Science & Technology (ESQ) has received funding from the Austrian Federal Ministry of Education, Science and Research (BMBWF).

*jessica.bavaresco@oeaw.ac.at

[1] G. Chiribella and D. Ebler, Quantum speedup in the identification of causeeffect relations, Nat. Commun. **10**, 1472 (2019).

[2] D. Deutsch and R. Jozsa, Rapid solution of problems by quantum computation, Proc. R. Soc. A **439**, 553 (1992).

[3] L. K. Grover, A framework for fast quantum mechanical algorithms, in *Proceedings of the 30th annual ACM symposium on Theory of computing* (Association for Computing Machinery, New York, 1998), p. 53, https://doi.org/10.1145/276698.276712.

[4] A. Chefles, A. Kitagawa, M. Takeoka, M. Sasaki, and J. Twamley, Unambiguous discrimination among oracle operators, J. Phys. A Math. Gen. **40**, 10183 (2007).

[5] D. Reitzner and M. Ziman, Two notes on Grover's search: Programming and discriminating, Eur. Phys. J. Plus **129**, 128 (2014).

[6] G. Chiribella, G. M. D'Ariano, and P. Perinotti, Quantum Circuit Architecture, Phys. Rev. Lett. **101**, 060401 (2008).

[7] M. Skotiniotis, R. Hotz, J. Calsamiglia, and R. Munoz-Tapia, Identification of malfunctioning quantum devices, arXiv:1808.02729.

[8] J. L. Pereira and S. Pirandola, Bounds on amplitude-damping-channel discrimination, Phys. Rev. A **103**, 022610 (2021).

[9] V. Giovannetti, S. Lloyd, and L. Maccone, Advances in quantum metrology, Nat. Photonics **5**, 222 (2011).

[10] A. S. Holevo, Statistical decision theory for quantum systems, J. Multivariate Anal. **3**, 337 (1973).

[11] C. W. Helstrom, Quantum detection and estimation theory, J. Stat. Phys. **1**, 231 (1969).

[12] A. Y. Kitaev, Quantum computations: algorithms and error correction, Russ. Math. Surv. **52**, 1191 (1997).

[13] D. Aharonov, A. Kitaev, and N. Nisan, Quantum circuits with mixed states, in *Proceedings of the Thirtieth Annual ACM Symposium on Theory of Computing STOC '98* (Association for Computing Machinery, New York, 1998), p. 20, https://doi.org/10.1145/276698.276708.

[14] A. Acín, Statistical Distinguishability Between Unitary Operations, Phys. Rev. Lett. **87**, 177901 (2001).

[15] 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).

[16] R. Duan, Y. Feng, and M. Ying, Entanglement is Not Necessary for Perfect Discrimination Between Unitary Operations, Phys. Rev. Lett. **98**, 100503 (2007).

[17] M. Ziman, Process positive-operator-valued measure: A mathematical framework for the description of process tomography experiments, Phys. Rev. A **77**, 062112 (2008).

[18] R. Duan, Y. Feng, and M. Ying, Perfect Distinguishability of Quantum Operations, Phys. Rev. Lett. **103**, 210501 (2009).

[19] R. Duan, C. Guo, C. Li, and Y. Li, Parallel distinguishability of quantum operations, in *Proceedings of the 2016 IEEE International Symposium on Information Theory (ISIT)* (IEEE, Barcelona, 2016), p. 2259, https://doi.org/10.1109/ISIT.2016.7541701.

[20] S. Becker, N. Datta, L. Lami, and C. Rouzé, Energy-Constrained Discrimination of Unitaries, Quantum Speed Limits, and a Gaussian Solovay-Kitaev Theorem, Phys. Rev. Lett. **126**, 190504 (2021).

[21] G. Chiribella, G. M. D'Ariano, and P. Perinotti, Memory Effects in Quantum Channel Discrimination, Phys. Rev. Lett. **101**, 180501 (2008).

[22] 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).

[23] M. Horodecki, P. W. Shor, and M. B. Ruskai, Entanglement breaking channels, Rev. Math. Phys. **15**, 629 (2003).

[24] S. Pirandola, R. Laurenza, C. Lupo, and J. L. Pereira, Fundamental limits to quantum channel discrimination, npj Quantum Inf. **5**, 50 (2019).

[25] Q. Zhuang and S. Pirandola, Ultimate Limits for Multiple Quantum Channel Discrimination, Phys. Rev. Lett. **125**, 080505 (2020).

[26] M. Rexiti and S. Mancini, Discriminating qubit amplitude damping channels, J. Phys. A **54**, 165303 (2021).

[27] V. Katariya and M. M. Wilde, Evaluating the advantage of adaptive strategies for quantum channel distinguishability, arXiv:2001.05376.

[28] G. Chiribella, G. M. D'Ariano, P. Perinotti, and B. Valiron, Quantum computations without definite causal structure, Phys. Rev. A **88**, 022318 (2013).

[29] G. Chiribella, Perfect discrimination of no-signalling channels via quantum superposition of causal structures, Phys. Rev. A **86**, 040301(R) (2012).

[30] M. Araújo, F. Costa, and Č. Brukner, Computational Advantage from Quantum-Controlled Ordering of Gates, Phys. Rev. Lett. **113**, 250402 (2014).

[31] A. Feix, M. Araújo, and Č. Brukner, Quantum superposition of the order of parties as a communication resource, Phys. Rev. A **92**, 052326 (2015).

[32] 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).

[33] M. T. Quintino, Q. Dong, A. Shimbo, A. Soeda, and M. Murao, Reversing Unknown Quantum Transformations: Universal Quantum Circuit for Inverting General Unitary Operations, Phys. Rev. Lett. **123**, 210502 (2019).

[34] The Choi-Jamiołkowski isomorphism is a one-to-one correspondence between linear maps $\tilde{\Lambda}\colon \mathcal{L}(\mathcal{H}^I) \to \mathcal{L}(\mathcal{H}^O)$ and linear operators $L \in \mathcal{L}(\mathcal{H}^I \otimes \mathcal{H}^O)$ defined by $L \coloneqq (\tilde{\mathbb{1}} \otimes \tilde{\Lambda})(\Phi^+)$, 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.

[35] J. de Pillis, Linear transformations which preserve hermitian and positive semidefinite operators, Pac. J. Math. **23**, 129 (1967).

[36] A. Jamiołkowski, Linear transformations which preserve trace and positive semidefiniteness of operators, Rep. Math. Phys. **3**, 275 (1972).

[37] M.-D. Choi, Completely positive linear maps on complex matrices, Linear Algebra Appl. **10**, 285 (1975).

[38] Let $F \in \mathcal{L}(\mathcal{H}^A \otimes \mathcal{H}^B)$ and $G \in \mathcal{L}(\mathcal{H}^B \otimes \mathcal{H}^C)$ be two linear operators. We define the link product between them as $F * G \coloneqq \mathrm{Tr}_B[(F^{AB} \otimes \mathbb{1}^C)(\mathbb{1}^A \otimes G^{BC^{T_B}})]$, where $\cdot^{T_B}$ is the transposition in the computational basis of $\mathcal{H}^B$. If $C_i \in \mathcal{L}(\mathcal{H}^I \otimes \mathcal{H}^O)$ is the Choi operator of a quantum channel $\tilde{C}_i\colon \mathcal{L}(\mathcal{H}^I) \to \mathcal{L}(\mathcal{H}^O)$, then for every state $\rho \in \mathcal{L}(\mathcal{H}^I)$ we have that $\tilde{C}_i(\rho) = C_i * \rho = \mathrm{Tr}_I[C_i(\rho^T \otimes \mathbb{1}^O)] \in \mathcal{L}(\mathcal{H}^O)$.

[39] G. Chiribella, G. M. D'Ariano, and P. Perinotti, Theoretical framework for quantum networks, Phys. Rev. A **80**, 022339 (2009).

[40] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevLett.127.200504 for a characterization theorem for general testers in Sec. I, semidefinite programming and dual affine spaces in Sec. II, a detailed description of our computer-assisted proofs methods in Sec. III, and an analysis of the sampling of general channels and the typicality of the hierarchy between discrimination strategies in Sec. IV. The Supplemental Material includes also Refs. [41–49].

[41] D. Kretschmann and R. F. Werner, Quantum channels with memory, Phys. Rev. A **72**, 062323 (2005).

[42] D. Beckman, D. Gottesman, M. A. Nielsen, and J. Preskill, Causal and localizable quantum operations, Phys. Rev. A **64**, 052309 (2001).

[43] T. Eggeling, D. Schlingemann, and R. F. Werner, Semicausal operations are semilocalizable, Europhys. Lett. **57**, 782 (2002).

[44] J. Wechs, A. A. Abbott, and C. Branciard, On the definition and characterisation of multipartite causal (non)separability, New J. Phys. **21**, 013027 (2019).

[45] CVX: Matlab Software for Disciplined Convex Programming, http://cvxr.com/cvx/.

[46] The MOSEK optimization toolbox for MATLAB, https://www.mosek.com.

[47] SeDuMi: Optimization over symmetric cones, http://sedumi.ie.lehigh.edu/.

[48] SDPT3 — a MATLAB software for semidefinite-quadratic-linear programming, https://blog.nus.edu.sg/mattohkc/softwares/sdpt3/.

[49] N. Johnston, QETLAB: A MATLAB Toolbox for Quantum Entanglement, http://www.qetlab.com.

[50] Notice that we can drop the transpose over T_i in the expression of the maximal probability of success since the set of testers is equal to the set of testers with a transposition, not affecting the optimization.

[51] O. Oreshkov, F. Costa, and Č. Brukner, Quantum correlations with no causal order, Nat. Commun. **3**, 1092 (2012).

[52] M. Araújo, C. Branciard, F. Costa, A. Feix, C. Giarmatzi, and Č. Brukner, Witnessing causal nonseparability, New J. Phys. **17**, 102001 (2015).

[53] For the case of separable testers, instead of imposing that $\bar{W}^{\mathcal{S}}$ lies in the dual affine set of separable processes, we should impose that $p_i C_i^{\otimes 2} \le H$ with $H \le \bar{W}^{1\prec 2}$ and $H \le \bar{W}^{2\prec 1}$ where $\bar{W}^{i\prec j}$ lies in the dual affine set of sequential processes $\mathcal{W}^{i\prec j}$. See the Suppl. Material for details. Notice that our formulation of the dual problem is analogous to the one presented in Ref. [54,55] and may be seen as a generalization of the Yuen-Kennedy-Lax bound for state discrimination [56].

[54] G. Chiribella, Optimal networks for quantum metrology: Semidefinite programs and product rules, New J. Phys. **14**, 125008 (2012).

[55] G. Chiribella and D. Ebler, Optimal quantum networks and one-shot entropies, New J. Phys. **18**, 093053 (2016).

[56] H. Yuen, R. Kennedy, and M. Lax, Optimum testing of multiple hypotheses in quantum detection theory, IEEE Trans. Inf. Theory **21**, 125 (1975).

[57] The Floating-Point Guide, https://floating-point-gui.de.

[58] Floating-point arithmetic, https://wikipedia.org/wiki/Floating-point_arithmetic.

[59] H. Peyrl and P. A. Parrilo, Computing sum of squares decompositions with rational coefficients, Theor. Comput. Sci. **409**, 269 (2008).

[60] S. M. Rump, Verification methods: Rigorous results using floating-point arithmetic, Acta Numer. **19**, 287 (2010).

[61] The action of an amplitude damping channel on a qubit state is given by $\tilde{C}_{\mathrm{AD}}(\rho) = K_0 \rho K_0^\dagger + K_1 \rho K_1^\dagger$, where $K_0 = |0\rangle\langle 0| + \sqrt{1-\gamma}|1\rangle\langle 1|$ and $K_1 = \sqrt{\gamma}|0\rangle\langle 1|$.

[62] The action of a bit-flip channel on a qubit state is given by $\tilde{C}_{\mathrm{BF}}(\rho) = \eta\rho + (1-\eta)X\rho X$, where X is the Pauli X operator.

[63] 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).

[64] M. T. Quintino and J. Bavaresco, Code to accompany: Strict hierarchy between parallel, sequential, and indefinite-causal-order strategies for channel discrimination, https://github.com/mtcq/channel_discrimination.