



universität  
wien

# DISSERTATION / DOCTORAL THESIS

Titel der Dissertation /Title of the Doctoral Thesis

“Certifying complex quantum properties:  
High-dimensional entanglement and indefinite causal order”

verfasst von / submitted by

Jessica Bavaresco

angestrebter akademischer Grad / in partial fulfilment of the requirements for the degree of  
Doktorin der Naturwissenschaften (Dr. rer. nat.)

Wien, 2021 / Vienna, 2021

Studienkennzahl lt. Studienblatt /  
degree programme code as it appears on the  
student record sheet:

A 796 605 411

Dissertationsgebiet lt. Studienblatt /  
field of study as it appears on the student  
record sheet:

Physik

Betreut von / Supervisor:

Priv. Doz. Ass. Prof. Mag. Dr. Marcus Huber



*“Eu quase que nada não sei. Mas desconfio de muita coisa.”*

— João Guimarães Rosa



# Abstract

The technological application of quantum features relies upon the certification of the corresponding desirable properties of quantum states and processes. As our technical capabilities advance, so does our need for understanding complex properties that go beyond simple qubit entangled states and causal processes. This cumulative thesis addresses two of the current challenges in quantum state and process characterisation. First, we study the certification of high-dimensional entanglement that arises when systems are entangled in more than two degrees of freedom. Second, we study the certification of processes with indefinite causal order and the problem of channel discrimination.

High-dimensional encoding of quantum information provides a promising method of transcending current limitations in quantum communication. The main goal of this part of this thesis was to develop practically implementable tools to certify the dimensionality of the entanglement taking into consideration the least possible amount of assumptions about the state, and when viable, no assumptions at all. In the first work, we developed an adaptive method that certifies the Schmidt number of a state using only two global product measurements. We put our method to test in an experiment that certified entanglement in 9 dimensions on a state encoded in the orbital angular momentum of two photons.

When studying the most general transformations that act on a pair of quantum operations, an intriguing phenomenon emerges: some higher-order transformations may act on their input operations in an indefinite causal order. Such non-causal properties have found several theoretical advantages, from communication complexity to quantum computing. In the second part of the thesis, a formalism of certification of non-causal properties under different levels of assumptions was developed. This formalism can witness indefinite causal order in device-dependent, independent, and semi-device-independent experiments, while showing that a prominent higher-order operation, the quantum switch, can demonstrate stronger non-causal properties than what was previously known. For the other works in this part of the thesis, the focus was on the famous quantum information problem of channel discrimination, proving several novel instances of tasks for which sequential strategies outperform parallel ones, both for sets of unitary and general channels. We defined new classes of strategies for channel discrimination that employ indefinite causal order and proved them to be advantageous when compared to causally ordered ones. These strategies may themselves be interpreted as methods for certification of indefinite causal order.



# Zusammenfassung

Die technische Anwendung quantenmechanischer Systeme erfordert die Zertifizierung der gewünschten Eigenschaften der Quantenzustände und Prozesse. Mit dem Fortschreiten der technologischen Möglichkeiten braucht es das Verständnis von komplexen Eigenschaften die über simple qubit-verschränkte Zustände und kausale Prozesse hinausgehen. Diese kumulative Arbeit beschäftigt sich mit zwei der kontemporären Herausforderungen in Quantenzustands und-Prozess Charakterisierung. Zuerst behandelt die Arbeit die Zertifizierung von hochdimensionaler Verschränkung die Zustände kommt wenn Systeme in mehr als zwei Freiheitsgraden verschränkt sind. Danach geht es weiter mit der Zertifizierung von Prozessen mit indefiniter Kausalordnung und dem Problem der Kanalunterscheidung.

Hochdimensionale Kodierung von Quanteninformation bietet eine vielversprechende Methode um gegenwärtige Beschränkungen in der Quantenkommunikation zu überwinden. Das Hauptziel diesen Teils der Arbeit war die Entwicklung praktisch implementierbarer Theoriewerkzeuge zur Zertifizierung der Dimensionalität der Verschränkung unter Berücksichtigung der mindestmöglichen Annahmen über den Zustand, oder, wenn möglich, gar keiner Annahmen. In der ersten Arbeit haben wir eine adaptive Methode entwickelt, die den Schmidt-Zahl eines Zustandes anhand von nur zwei globalen Produktmessungen nachweist. Wir haben unsere Methode in einem Experiment getestet, bei dem die Verschränkung in 9 Dimensionen an einem Zustand bestätigt wurde, die im Drehimpuls zweier Photonen kodiert ist.

Bei der Untersuchung der allgemeinsten Transformationen, die auf ein Paar von Quantenoperationen einwirken können, tritt ein faszinierendes Phänomen auf: Einige Transformationen höherer Ordnung können in einer unbestimmten kausalen Reihenfolge auf die Eingabeoperationen einwirken. Solche nicht-kausalen Eigenschaften haben mehrere theoretische Vorteile von der Kommunikationskomplexität bis zum Quantencomputer gezeigt. Im zweiten Teil der Arbeit wurde ein Formalismus für die Zertifizierung einer solchen nicht-kausalen Ordnung unter verschiedenen Annahmeebenen entwickelt. In der ersten Arbeit haben wir einen Zertifizierungsformalismus entwickelt, der dieses Phänomen in geräteabhängigen, geräteunabhängigen und semi-geräteunabhängigen Experimenten zeigen kann, und gleichzeitig gezeigt dass eine herausragende Operation höherer Ordnung, der Quantenschalter, stärkere nicht-kausale Eigenschaften aufweisen kann als bisher bekannt. Bei den anderen Arbeiten im zweiten Teil der Arbeit lag der Schwerpunkt auf dem bekannten

Quanteninformationsproblem der Kanalunterscheidung, das mehrere neuartige Instanzen von Aufgaben aufzeigt, bei denen sequentielle Strategien parallele Strategien sowohl für einheitliche als auch für allgemeine Kanäle übertreffen. Wir haben dann neue Klassen von Strategien zur Kanaldiskriminierung definiert, die eine unbestimmte kausale Ordnung verwenden, und sie als vorteilhaft im Vergleich zu kausal geordneten erwiesen. Diese Strategien können selbst als Methoden zur Zertifizierung einer unbestimmten kausalen Ordnung interpretiert werden.

# Acknowledgements

The last four and a half years have brought me amazing experiences that go well beyond the work contained in this thesis, and have enriched me both as a scientist and as a person. None of it would have been possible without the support, collaboration, and guidance of the several brilliant people that I met along the way.

Firstly, I would like to express my deep gratitude to my supervisor, Marcus Huber. You not only gave me the opportunity to do a PhD in Vienna but you went out of your way to make sure I found my way around here. This support, particularly in the beginning but also on the toughest times, made all the difference. I am grateful to you for giving me the freedom to explore as many topics and paths as I wanted, from foundations to experiments, while always giving me the opportunity to be a part of your projects. As much as I felt free, I never felt lost or discouraged under your supervision or as a part of the awesome group you have created. I am indebted to you not only for the science you taught me but also for advising and guiding me in every aspect of my academic career. Although the techno-German-rap you listen to is a bit too loud and your non-compilable tex files drive me absolutely insane, I consider myself extremely fortunate to have had you as my supervisor.

I am also very grateful to all of my collaborators, without whom the work in this thesis could not have been accomplished. In particular, I would like to sincerely thank Nicolai Friis, for being a beacon of sanity and order amidst the Quitter madness and for always being there for me, personally and professionally, throughout my entire PhD; Simon Milz, for our chaotically productive discussions and for sharing your exquisite taste in rom-coms; and Mateus Araújo, for always being down for discussions and getting your hands in the calculations, and of course, for the most fantastic picanhas.

I am truly thankful to Mio Murao, for welcoming me into her group in Tokyo for several visits, for our interesting projects and discussions, and for being a role model to me. I absolutely loved the time I got to spend in Japan.

I want to thank my PhD-brothers, Fabien Clivaz, Manu Schwarzhans, Simon Morelli, and Phil Taranto, for giving me the feeling of a real PhD family and for all the pleasant beers we shared. A special thanks to Fabien, for being such a great friend, and to Phil, for the extremely useful feedback about this thesis and for giving me the energy to write it, with your life-saving, mouth-watering #thesismisu.

To all the Quitters, for creating such an open, relaxed and stimulating research environment, and for all the good times we spent together. To the Quitmas Master, Yelena Guryanova, for the incredible holiday joy you have never failed to deliver and for your amazing company all these years. And to Tiago Debarba, for sharing with me such a special time in Vienna.

I would like to particularly thank Yelena Guryanova, Natalia H. Valencia, Chrysoula Vlachou, Xiaoqin Gao, Ayaka Usui, Eliza Agudelo, and Cornelia Spee, for making our group a better one.

I am also indebted to Orkideh Badamchian, for showing me which battles are worth picking and for fighting them with me. Many thanks as well to Sonja Redl, Lukas Edengruber, and Benjamin Skarabela from the IQOQI admin-team for your support and your kindness.

To Princess Bonnibel Bubblegum, for performing countless calculations around the clock, and specially to Marco Túlio Quintino and Mateus Araújo, for frequently providing her with much appreciated maintenance. She would not be alive if it were not for you. Or maybe she would, but she would be running Windows, and what kind of life would that be?

Most specially, a heartfelt thank you to Marco Túlio, for all that you have taught me, all the support you have given me, and for being the most amazing person I have ever met.

Finally, to my parents and my fabulous sister, for holding my education as their highest priority and for constantly supporting me in pursuing my dream of becoming a scientist. This journey would have never even started if it were not for your love and dedication, and for that I am forever grateful.

I would like to thank the Austrian Science Fund (FWF), for funding my PhD through the START project Y879-N27.

# List of Publications

The indicated publications are featured in this thesis.

## Peer-reviewed

- [P1] **Verification of high-dimensional entanglement generated in quantum interference.**

Yuanyuan Chen, Sebastian Ecker, Jessica Bavaresco, Thomas Scheidl, Lixiang Chen, Fabian Steinlechner, Marcus Huber, and Rupert Ursin.

*Physical Review A* **101**, 032302 (2020).

[arXiv:1910.07684 \[quant-ph\]](#).

[DOI: 10.1103/PhysRevA.101.032302](#).

- [P2] **Semi-device-independent certification of indefinite causal order.**

Jessica Bavaresco, Mateus Araújo, Časlav Brukner, and Marco Túlio Quintino.

*Quantum* **3**, 176 (2019).

[arXiv:1903.10526 \[quant-ph\]](#).

[DOI: 10.22331/q-2019-08-19-176](#).

(Chapter 2)

- [P3] **Measurements in two bases are sufficient for certifying high-dimensional entanglement.**

Jessica Bavaresco, Natalia H. Valencia, Claude Klöckl, Matej Pivoluska, Paul Erker, Nicolai Friis, Mehul Malik, and Marcus Huber.

*Nature Physics* **14**, 1032 (2018).

[arXiv:1709.07344 \[quant-ph\]](#).

[DOI: 10.1038/s41567-018-0203-z](#).

(Chapter 1)

- [P4] **Simulations of submonolayer Xe on Pt(111): The case for a chaotic low temperature phase.**

Anthony D. Novaco and Jessica Bavaresco.

*Journal of Chemical Physics* **148**, 144704 (2018).

arXiv:1708.01493 [cond-mat.mtrl-sci].

DOI: 10.1063/1.5024027.

[P5] **Operational framework for quantum measurement simulability.**

Leonardo Guerini, Jessica Bavaresco, Marcelo Terra Cunha, and Antonio Acín.

*Journal of Mathematical Physics* **58**, 092102 (2017).

arXiv:1705.06343 [quant-ph].

DOI: 10.1063/1.4994303.

[P6] **Most incompatible measurements for robust steering tests.**

Jessica Bavaresco, Marco Túlio Quintino, Leonardo Guerini, Thiago O. Maciel, Daniel Cavalcanti, and Marcelo Terra Cunha.

*Physical Review A* **96**, 022110 (2017).

arXiv:1704.02994 [quant-ph].

DOI: 10.1103/PhysRevA.96.022110.

[P7] **Distribution of high-dimensional entanglement via an intra-city free-space link.**

Fabian Steinlechner, Sebastian Ecker, Matthias Fink, Bo Liu, Jessica Bavaresco, Marcus Huber, Thomas Scheidl, and Rupert Ursin.

*Nature Communications* **8**, 15971 (2017).

arXiv:1612.00751 [quant-ph].

DOI: 10.1038/ncomms15971.

## Under review

[P8] **Strict hierarchy between parallel, sequential, and indefinite-causal-order strategies for channel discrimination.**

Jessica Bavaresco, Mio Murao, and Marco Túlio Quintino.

arXiv:2011.08300 [quant-ph] (2020).

(Chapter 3)

## In preparation

[P9] **Resource theory of causal connection.**

Jessica Bavaresco, Giulio Chiribella, Simon Milz\*.

\*Authors listed in alphabetical order.

[P10] **Unitary operation discrimination beyond group structures: advantages of adaptive and indefinite-causal-order strategies.**

Jessica Bavaresco, Mio Murao, and Marco Túlio Quintino.

(Chapter 4)

# Contents

<b>Preamble</b>	<b>1</b>
<b>I High-dimensional entanglement</b>	<b>11</b>
<b>1 Measurements in two bases are sufficient for certifying high-dimensional entanglement</b>	<b>13</b>
1.1 Introduction . . . . .	14
1.2 Entanglement dimensionality . . . . .	16
1.3 Target state identification . . . . .	17
1.4 Dimensionality witnesses . . . . .	18
1.5 Fidelity bounds . . . . .	18
1.6 Experimental certification of high-dimensional entanglement . . . . .	20
1.7 Discussion and outlook . . . . .	24
1.8 Methods . . . . .	25
1.8.1 Derivation of the fidelity lower bound . . . . .	25
1.8.2 Tightness of the fidelity bound . . . . .	29
1.8.3 Role of the target state . . . . .	30
1.8.4 Experimental details . . . . .	31
<b>II Indefinite causal order</b>	<b>35</b>
<b>2 Semi-device-independent certification of indefinite causal order</b>	<b>37</b>
2.1 Introduction . . . . .	38
2.2 Preliminaries . . . . .	40
2.3 Certification . . . . .	42
2.3.1 Device-dependent . . . . .	44
2.3.2 Device-independent . . . . .	46
2.3.3 Semi-device-independent . . . . .	49
2.4 The Quantum Switch . . . . .	54
2.5 Conclusions . . . . .	60

<b>3 Strict hierarchy between parallel, sequential and indefinite-causal-order strategies for channel discrimination</b>	<b>63</b>
3.1 Introduction . . . . .	64
3.2 Minimum-error channel discrimination . . . . .	65
3.3 Tester formalism for two copies . . . . .	66
3.4 Discrimination of general channels . . . . .	69
3.5 Quantum realization of testers . . . . .	72
3.6 Conclusions . . . . .	73
<b>4 Unitary operation discrimination beyond group structures: Advantages of adaptive and indefinite-causal-order strategies</b>	<b>75</b>
4.1 Introduction . . . . .	76
4.2 Minimum-error channel discrimination . . . . .	77
4.3 Tester formalism for $k$ copies . . . . .	78
4.4 Discrimination of unitary channels . . . . .	81
4.5 Numerical investigation . . . . .	84
4.6 Conclusions . . . . .	85
<b>Concluding Discussion</b>	<b>87</b>
<b>Appendix</b>	<b>90</b>
<b>Appendix A Supplemental Information of Chapter 1</b>	<b>91</b>
A.1 Normalization for measurements in the tilted bases . . . . .	93
A.2 Noise robustness . . . . .	94
A.3 Improved bounds using multiple bases . . . . .	95
A.4 Bounds on the entanglement of formation . . . . .	99
A.5 Entanglement quantification using mutually unbiased bases . . . . .	103
A.6 Multipartite entanglement certification . . . . .	105
A.7 Effects of a wrong choice of Schmidt basis on the fidelity lower bounds . . . . .	107
A.8 Classical prepare-and-measure experiment: LG basis . . . . .	108
A.8.1 Examples of MUBs in other experimental degrees-of-freedom/platforms . . . . .	111
A.8.2 Systematic errors . . . . .	113
<b>Appendix B Supplemental Information of Chapter 2</b>	<b>117</b>
B.1 Behaviours and assemblages unattainable by process matrices . . . . .	118
B.2 Behaviours and assemblages attainable by causally separable process matrices	121
B.3 Characterization theorems for general and causal assemblages . . . . .	125
B.4 Causally nonseparable process matrices that cannot be certified in a semi-device-independent way . . . . .	128
B.5 Certification of tripartite process matrices . . . . .	129

B.5.1	Device-dependent – TTT . . . . .	131
B.5.2	Device-independent – UUU . . . . .	131
B.5.3	Semi-device-independent – TTU . . . . .	134
B.5.4	Semi-device-independent – TUU . . . . .	136
B.5.5	Semi-device-independent – UTT . . . . .	138
B.5.6	Semi-device-independent – UUT . . . . .	139
B.6	Proof that the quantum switch processes are causal in the UUT scenario . .	141
B.7	Analysis of experimental implementations of the quantum switch . . . . .	143
<b>Appendix C Supplemental Information of Chapter 3</b>		<b>147</b>
C.1	Characterization theorem for general testers . . . . .	147
C.2	Semidefinite programming formulation and dual affine spaces . . . . .	150
C.3	Computer-assisted proofs . . . . .	156
C.4	Sampling general channels and the typicality of the hierarchy between discrimination strategies . . . . .	162
<b>Appendix D Supplemental Information of Chapter 4</b>		<b>165</b>
D.1	Proof of Theorem 4.1 . . . . .	165
D.2	Proof of Examples 4.1 and 4.2 . . . . .	172
D.3	Proof of Example 4.3 . . . . .	175
D.4	Proof of Theorem 4.4 . . . . .	176
D.5	Numerical Findings . . . . .	180
<b>Bibliography</b>		<b>183</b>



# Preamble

A certification procedure is one that can infer, in an unambiguous manner, that a physical system or mathematical object exhibits a certain property. Typically, certification procedures require only partial information about the object whose properties it intends to certify. In other words, they do not require its complete description. A concrete example is the certification of the entanglement of a quantum state by an entanglement witness, which is a procedure that only requires the knowledge of the expectation value of the given state for a certain observable. The full characterization of the state, which could be achieved by state tomography, is not required for the evaluation of an entanglement witness, making it more practical and efficient.

Such procedures come in handy particularly when considering complex quantum properties, for which complete characterization can be far too costly. Although somewhat arbitrary, the term complex quantum properties is hereby used to refer to properties of quantum systems which go beyond their simplest manifestation and have been shown to bring advantages to information-theoretic tasks. It means high-dimensional, as opposed to qubit entanglement. Multipartite, as opposed to bipartite states. Non-causal, as opposed to causally ordered processes.

The main challenge addressed in this thesis is the development of methods to certify two remarkable complex quantum properties: high-dimensional entanglement and indefinite causal order. The certification of each of these two properties are driven by different motivations and currently face contrasting stages of technological implementation.

In this Preamble, we outline the context and the contribution of this thesis to both of these topics. The thesis is then composed of two parts: Part I, containing Chapter 1, addresses the topic of certification of high-dimensional entanglement, while Part II, containing Chapters 2–4, addresses the topic of certification of indefinite causal order and channel discrimination. In the Concluding Discussion, an outlook on future research directions is presented.

Fast-developing quantum technologies have been proving their merit in the last couple of decades. Quantum advantages for, e.g., computation [1], communication [2], security for key distribution [3, 4], and simulations of complex systems [5, 6] have motivated remarkable experimental achievements. Some examples are the distribution of pairs of entangled

states across long distances [7, 8], over free-space [9, 10] and satellite links [11, 12], and the fine-tuned control over various physical platforms, as for instance, photonics [13, 14], ultracold atoms [15], integrated-optics [16], and superconducting circuits [17].

While the generation of quantum states that are entangled across two degrees of freedom is readily available, the canonical way in photonics of generating two-dimensional polarization entanglement in down-conversion processes already offers the potential for exploring more complex entanglement properties in higher dimensions, by exploiting spatial degrees of freedom [18, 19], orbital angular momentum [20–22], energy-time based encodings [23–26], or combinations thereof, to create hyper-entangled quantum systems [27, 28].

High-dimensional entanglement has long been known to overcome some limitations of qubit entanglement [29, 30], by offering better key rates [31], higher noise resistance [32, 33] and improved security against different attacks [34]. Hence, the necessity of developing solid theoretical tools that can allow us to take full advantage of the available technology emerges.

One main concern is on certification methods that do not rely on assumptions about the state being certified. Realistic scenarios of entanglement distribution for, e.g., communication purposes, cannot rely on the assumption that the state that arrives at the detectors after traveling potentially long distances is the same as that which was generated by its source. Moreover, assumptions about the source itself may put quantum protocols at risk of suffering security breaches [35].

Nevertheless, due to the complexity of realizing measurements in high-dimensional spaces, previous methods that aimed to certify dimensionality of entanglement often had to resort to assumptions about the underlying quantum state, including, amongst others, conservation of angular momentum [36], subtraction of accidentals [37], perfect correlations in a desired basis [38], or that the experimentally-generated state is pure [39]. Although often plausible, such assumptions decrease the strength of the certification result, which will only hold as long as the assumptions can be guaranteed to do so.

Part I of this thesis focuses on these challenges: the development and testing of high-dimensional entanglement certification methods that can be applied across multiple physical platforms in a practical and efficient manner without relying on any assumptions about the quantum state.

In Chapter 1, we devise and implement an experimentally-accessible method for certifying the dimensionality of the entanglement of an unknown quantum state. In contrast to state tomography, which requires  $(d + 1)^2$  global product measurements, where  $d$  is the local dimension of the subsystems, and measurement of fidelity with respect to a pure state, which requires  $(d + 1)$  global product measurements, our method requires only two measurements, regardless of the dimension of the underlying state.

Consider a bipartite quantum state, prepared in a laboratory, about which no information is assumed — not even the dimension of its local systems. The goal is to certify whether or not this state is entangled and, in case it is, quantify the dimensionality of its entanglement.

Our method goes as follows. The first step is to choose a basis onto which the first measurement, a local projective measurement, will be performed, along with the dimension of the space that will be measured. Ideally, the chosen basis should be as close as possible to the Schmidt basis of the state. Since no assumptions about the state are made, one cannot know what the actual Schmidt basis is. Nevertheless, an educated guess is often possible, and can be made based on the physical platform and degree of freedom used to encode the state. For example, if the state of interest is encoded in the orbital-angular-momentum degree of freedom of a pair of photons that were prepared through a process of spontaneous parametric down conversion (SPDC), one may explore the natural symmetry that should arise from the conservation of angular momentum to guess that the Schmidt basis of this state should be close to the Laguerre-Gauss basis. The experiment reported in Chapter 1 follows this reasoning. Another example would be an experiment involving two-mode squeezed states, for which the Fock basis could be the most appropriate choice for the first measurement. While this educated guess will guide the choice of measurement, its correspondence to the actual Schmidt basis of the measured state is not a requirement of the method. This guess will not be part of the analysis of the data, therefore not constituting an assumption.

The result of this first measurement will yield partial information about the state, and this partial knowledge will be exploited in the next step to serve two purposes. The first is the determination of a target state with respect to which the fidelity of the measured state will be estimated. This target state is pure and has maximal entanglement dimensionality for its subspace dimensions. The second purpose is the determination of the second basis of measurement. This basis – which we have dubbed *tilted* basis – is a set of non-orthogonal, normalized, and linearly independent vectors which are constructed by transforming a basis that is mutually unbiased with respect to the first basis of measurement, according to coefficients that are calculated from the outcome statistics of the first measurement.

Subsequently, a measurement onto the projectors of the second basis is performed locally on each subsystem. With the outcome statistics of these two measurements, one now has enough information about the measured state to compute a lower bound of its fidelity with respect to the target state, and consequently infer a lower bound on the Schmidt number of the measured state. This procedure guarantees a minimal dimensionality for the entanglement of the measured state without any assumptions about its preparation.

Of course, in practice, *some* information about the state preparation is often available. For example, if a state is prepared in the polarization degree of freedom of two photons, and polarization is the only degree of freedom that will be measured in the experiment, one could be confident that they are dealing with a two-qubit state. If, moreover, this state was prepared through a process of SPDC, one could expect it to be close to a maximally entangled state. The important question is then whether or not this knowledge will be used/assumed in the analysis of the experimental data. By choosing *not* to use such information in order to make assumptions about the state, but rather to simply guide

the choice of the first measurement basis, we guarantee an entanglement certification that stands regardless of the validity of any state preparation assumption.

It is true that a successful certification – in the sense of a certification of a value for the entanglement dimensionality that is as close as possible to the actual entanglement dimensionality of the measured state – relies on a good choice of the initial measurement basis. Indeed, if this choice were made at random, most likely no entanglement would be certified, even when applying the method to an entangled state. However, the point of the method is to take advantage of the commonly available information about the potential symmetries of the physical systems and degrees of freedom where the state is prepared to make the best possible choice of measurement. Since the choice of initial measurement basis is not taken as an assumption in the certification process, its result holds regardless of whether the chosen basis was the actual Schmidt basis. Consequently, and crucially, a poor choice of the first measurement basis will never lead to an over-estimation of entanglement, although it could lead to an under-estimation. Another advantage of this adaptive method is that the target state does not have to be guessed or established *a priori*, but it is rather learned from the outcomes of the first measurement.

Although the theoretical merit of this method was well established by our proofs of tightness of our fidelity bounds for several classes of states, described in Chapter 1, and its independence of physical platform, its practical accessibility still remained to be put to test in an experiment carried out under realistic circumstances. We therefore designed and implemented an experiment using the orbital angular momentum of a pair of photons, and certified 9-dimensional entanglement, a record for assumption-free high-dimensional entanglement certification at the time of publication [40].

Another complex quantum property whose certification this thesis addresses is that of the causal order of quantum processes. While in hindsight it is easy to see that both of the topics of high-dimensional entanglement and of indefinite-causal-order approached in this thesis fall under the same umbrella of certification of complex quantum properties, the original motivation for the work of Part I of the thesis was, in many ways, the opposite of that of Part II. While the development of methods for the certification of high-dimensional entanglement was motivated by the current advances in quantum technologies, our study of the certification of indefinite causal order instead had the original intention of motivating such technological advances in its corresponding field.

The study of causal order in quantum information arises from the analysis of signaling properties of probability distributions. Joint probability distributions are signalling when the marginal probabilities of one party do not depend on the inputs of another. If, let us say, a set of joint probability distributions are signalling from party A to B, it means that, by their choice of inputs, party A can affect the marginal probabilities of party B, and exploit this property to, for example, signal messages. Although signalling is not enough to

establish a causal *influence* between the two parties, it is enough to establish the causal *order* between them. A set of joint probability distributions can determine whether A is in the past of B, B is in the past of A, or whether the events observed by the parties are not directly causally related, which is the case when the marginal probabilities of each party do not depend on the other.

To relate causal order relations to quantum physics, one can analyse what kind of probability distributions can be generated by the operations allowed by quantum mechanics. It is then convenient to analyse this question from the point of view of transformations on quantum operations – the so-called higher-order operations [41–43]. Here, by quantum operations, we are referring to operations that transform quantum states. The same way that one can see, e.g., a unitary operation to be a transformation that maps quantum states into quantum states, or a measurement to be a transformation that maps a quantum state into a probability distribution, one can also think of higher-order operations to be transformations of quantum operations. One example would be a transformation that maps quantum channels into quantum channels, measurements into measurements, or any quantum operation into a probability distribution. While transformations of quantum states are reasonably well-understood, complex properties of transformations of quantum operations arise when one considers their underlying causal structures. A useful formalism to approach causal properties of higher-order operations is the one of *process matrices* [44].

Let us first consider a set of joint probability distributions whose marginals are independent of the inputs of the other party and hence establish no causal ordering between the events they describe. Then, one can ask what is the most general transformation that can take a pair of quantum operations to a set of probability distributions of this kind. The answer to this question is what we call a *parallel* process [42, 45]. Such transformations, the parallel processes, can be realized by sending a bipartite quantum state through the pair of operations that will be transformed and then measuring it, resulting in probability distributions.

One could also ask what is the most general transformation that can take a pair of quantum operations to a set of joint probability distributions that is signalling in one direction, and hence establishes a fixed causal order between the events it describes. The answer this time is an *ordered* or *sequential* process [42, 44, 46]. Ordered processes can be physically implemented by sending a quantum state that is first acted upon by the operations of the first party, then sent through a quantum channel to the second party, who acts upon it with their quantum operations. The interaction of the process with the operations of each party yields probability distributions. Auxiliary systems may also be involved in the implementation of both parallel and ordered processes. The aforescribed process effectively takes a pair of quantum operations, each realized by one party, to probability distributions that allow for signalling in one direction.

Another consideration is that of joint probability distributions which can be expressed as a mixture (i.e. convex combination) of one-way signalling distributions. One can check

that these distributions can always be achieved by the action of a process that can also be described as a mixture of processes that are ordered in different directions. A mixture of processes can be realized by flipping a coin and implementing a process that is order in one direction when the coin lands on ‘heads’ and in another direction when it lands on ‘tails’. Such processes, in turn, are called *causally separable* processes [44, 46].

Finally, following this reasoning, one question is still left to be asked: *what is the most general transformation that can take a pair of quantum operations to any general set of probability distributions, without restrictions on their signalling relations?* The answer to this question is a *general process* or *general process matrix* [44, 46].

A general process is a valid transformation of pairs of quantum operations because it generates valid probability distributions. However, it is known that there exist general processes that are neither parallel, sequential, nor causally separable [44]. Moreover, some of the probability distributions that these processes can generate when acting on a pair of quantum operations also cannot themselves be described as mixtures of one-way signalling distributions. Although transmission of information is allowed between the parties in this scenario, it can be then concluded that neither the process itself nor the probability distributions that it can generate can define a *definite causal order* between the observed events. This interesting phenomenon has been called *indefinite causal order* [44].

An important follow-up question would then be: Can one describe a general indefinite-causal-order process *in terms of quantum states and quantum operations*, similarly to how we previously described parallel, sequential and causally separable process to be physically realized by quantum states, channels, measurements, and coin-throwing? For some of these processes, the answer is yes. A subclass of them, to which a much-studied process, the quantum switch, belongs, has been shown to be implementable by coherent quantum control of quantum operations [47, 48]. These results are nevertheless very recent and still remain to be experimentally tested. But for the remaining indefinite-causal-order processes that do not fit in this class, the answer to this question is currently an open avenue of research.

Currently, a general process matrix may most comfortably be seen as a mathematical object of interesting properties about which we still lack physical intuition. However, the limitations of our natural human intuition has, in the past, shown itself not to be a good enough reason to discard the study of puzzling mathematical objects<sup>1</sup>. Some physical and informational principles have been suggested as potential divisors of the quantumly realizable and non-realizable processes [54–56]. Importantly, so far no such principle has been able to conclusively rule out any process matrix, nor has any physically or mathematically absurd consequence (such as the ability to send information to the past or the ability to provide implausible computational advantages) been shown to arise from

---

<sup>1</sup>Looking back at the history of physics, one can easily find examples of theoretical findings which were initially seen as potential mathematical pathologies only to later be confirmed as accurate theoretical models for experimentally-observed phenomena. A few examples are the Poisson spot [49], the Dirac sea [50], the gravitational lens [51], the gravitational wave [52], and the Higgs boson [53].

the analysis of such objects.

Theoretically, the application of indefinite-causal-order has been shown to be advantageous to numerous tasks. Some examples are the outperformance of indefinite-causal-order processes over causally ordered ones in tasks such the discrimination of non-signaling channels [57], communication complexity [58, 59], quantum computation [60], and inverting unknown unitary operations [61].

Here enters the main motivation of the Part II of this thesis. By developing theoretical frameworks for the experimental certification of indefinite causal order and demonstrating the theoretical advantage of the application of these processes to quantum information tasks, we aim to provide tools to assist the further technological development of protocols that involve indefinite causal order and the continued investigation of their physical implementation.

In Chapter 2, we develop a framework for the certification of indefinite causal order in process matrices under different sets of assumptions. Similarly to the certification of high-dimensional entanglement in Part I, no assumptions are ever made about the object whose properties the framework proposes to certify: the quantum state or the process matrix. In contrast to our method of high-dimensional entanglement certification, which assumes that the measurements implemented experimentally are exactly the ones described in theory, for the certification of process matrices we consider scenarios in which the level of assumptions about the operations being applied also varies.

Processes, just like quantum states, may be certified by full characterization – achieved through process or state tomography [62–64] – or by the correlations they are able to produce. An analysis of the correlations may or may not take into consideration information about what operations were applied. A device-dependent approach is one that assumes full knowledge of the involved quantum operations in the analysis of experimental data. It is in this scenario that entanglement witnesses [65] and the analogous causal witnesses [46, 66] may be tested. A device-independent analysis, on the other hand, is the one that dispenses any potential knowledge about the quantum operations and takes conclusions solely based on the analysis of probability distributions. In this scenario, entanglement of quantum states can be certified through the violation of a Bell inequality [67, 68] and indefinite causal order of processes can be certified through the violation of causal inequalities [44, 69].

Even though the process matrix formalism predicts general process matrices that can generate correlations that allow for device-independent certification [44], a method for the experimental implementation of such processes is currently unknown. The only process matrix that has so far been the subject of experimental investigation is the quantum switch [43, 57], which is known not to be device-independently certifiable [46, 70], and all of its experiments to date have relied on a device-dependent analysis [71–80].

Another potential scenario for certification is the semi-device-independent scenario, which makes assumptions about the operations of some, but not all, of the involved parties, providing an intermediate test bed for the analysis of statistical data. These semi-device-independent scenarios have been extensively used for the certification of entanglement,

related to a phenomenon called quantum steering [81, 82], while the semi-device-independent certification of process matrices was first proposed in the work of Chapter 2. This work has unified certification for process matrices under different levels of assumptions, providing a plethora of mathematical tools for the analysis of potential future experiments. We have also shown that the indefinite-causal-order properties of the quantum switch can not only be certified under device-dependent assumptions, but also in semi-device-independent scenarios. This proves that the quantum switch can exhibit stronger non-causal properties than what was previously known and hopefully contributes to paving the way to stronger experimental demonstrations of indefinite causal order.

The last two chapters of the thesis, Chapter 3 and 4, study the classic<sup>2</sup> quantum information problem of channel discrimination and the application of indefinite-causal-order strategies to this problem.

Analogously to the task of quantum state discrimination [83], in a task of channel discrimination [83], one is given access to an unknown quantum channel, known to have been drawn from a channel ensemble with a certain probability. The task is then to perform transformations on this channel to determine which of the channels in the ensemble is the one at hand. Typically, only a finite number of uses of the unknown channel are allowed, which can also be seen as one being allowed to operate on a finite number of copies of the unknown channel. Depending on the causal structure in which the copies of the channel are applied, different discrimination strategies can be defined. Related to higher-order transformations as well, these strategies may be described as parallel, sequential, or indefinite-causal-order strategies [45].

The task of channel discrimination can also be interpreted as a certification of causal properties of processes. In principle, if the maximal probability of successful discrimination of an ensemble of channels under, for example, a parallel strategy is known, this information can be used to certify that an unknown process is not parallel by performing a discrimination task with said ensemble of channels and achieving a higher probability of success than would be achievable by a parallel strategy. The same reasoning applies to the certification of sequential and indefinite-causal-order processes.

Chapter 3 sets a tester formalism that unifies mathematical tools for the treatment of parallel, sequential, and indefinite-causal-order strategies using two copies and applies it to the discrimination of general channels. Chapter 4 extends this formalism to the task of channel discrimination using  $k$  copies and applies it to the particular case of unitary channels.

For general channels, we show a new example of the simplest instance of a channel discrimination task, that of discriminating between two qubit-qubit channels using two copies, for which a sequential strategy outperforms any parallel strategy. Our main example involves an amplitude-damping channel and a bit-flip channel. For this same discrimination

---

<sup>2</sup>classic, not classical.

task, we show that two kinds of strategies that involve indefinite causal order outperform causal (i.e. parallel and sequential) strategies, composing a strict hierarchy of performance.

This same phenomenon can be observed in the discrimination of unitary channels. We show both the first example of the advantage of sequential strategies, as compared to parallel strategies, in the discrimination of a set of unitary channels and the first example of the advantage of general indefinite-causal-order strategies as compared to sequential strategies. We also prove two cases in which indefinite causal order is not advantageous. The first is for the discrimination of a uniformly distributed set of unitary channels defined by unitary operators that form a group, for which we prove that parallel strategies are optimal even when considering general strategies for any finite number of copies. The second is for a particular case of indefinite-causal order strategies constructed from switch-like processes. We show that this class of strategies can never outperform sequential strategies in the discrimination of sets of unitary channels for any finite number of copies.



## Part I

# High-dimensional entanglement



# Measurements in two bases are sufficient for certifying high-dimensional entanglement

Jessica Bavaresco, Natalia H. Valencia, Claude Klöckl, Matej Pivoluska, Paul Erker, Nicolai Friis, Mehul Malik, and Marcus Huber

**Abstract.** High-dimensional encoding of quantum information provides a promising method of transcending current limitations in quantum communication. One of the central challenges in the pursuit of such an approach is the certification of high-dimensional entanglement. In particular, it is desirable to do so without resorting to inefficient full state tomography. Here, we show how carefully constructed measurements in two bases (one of which is not orthonormal) can be used to faithfully and efficiently certify bipartite high-dimensional states and their entanglement for any physical platform. To showcase the practicality of this approach under realistic conditions, we put it to the test for photons entangled in their orbital angular momentum. In our experimental setup, we are able to verify 9-dimensional entanglement for a pair of photons on a 11-dimensional subspace each, at present the highest amount certified without any assumptions on the state.

Published in:

*Nature Physics* **14**, 1032 (2018)

arXiv:1709.07344 [quant-ph]

## Author Contribution

In this work, the doctoral candidate contributed to the formulation of the theoretical methods, to the construction of the initial experimental set up, the analysis of the experimental data, and to the writing and revision of the manuscript.

### 1.1 Introduction

Quantum communication offers advantages such as enhanced security in quantum key distribution (QKD) protocols [84] and increased channel capacities [85] with respect to classical means of communication. All of these improvements, ranging from early proposals [86] to recent exciting developments such as fully device-independent QKD [87, 88], rely on one fundamental phenomenon: quantum entanglement. Currently, the workhorse of most implementations is entanglement between qubits, i.e., between two-dimensional quantum systems (e.g. photon polarization). However, it has long been known that higher-dimensional entanglement can be useful in overcoming the limitations of qubit entanglement [29, 30], offering better key rates [31], higher noise resistance [32, 33] and improved security against different attacks [34].

Attempting to capitalize on this insight, recent experiments have successfully generated and certified high-dimensional entanglement in different degrees of freedom. In particular, the canonical way of generating two-dimensional polarization entanglement in down-conversion processes already offers the potential for exploring entanglement in higher dimensions. This can be achieved by exploiting spatial degrees of freedom [18, 19], orbital angular momentum (OAM) [20–22], energy-time based encodings [23–26], or combinations thereof in hyper-entangled quantum systems [27, 28]. High-dimensional quantum systems have recently also been explored in matter-based systems such as Cesium atoms [89] and superconducting circuits [90]. Thus, high-dimensional quantum systems are not only of fundamental interest but are also becoming more readily available.

In this context, the certification and quantification of entanglement in many dimensions is a crucial challenge since full state tomography (FST) for bipartite systems of local dimension  $d$  requires measurements in  $(d + 1)^2$  global product bases (i.e., tensor product bases for the global state) [91], which quickly becomes impractical in high dimensions. Due to the complexity of realizing measurements in high-dimensional spaces, previous experiments that aimed to certify entanglement dimensionality (also known as Schmidt number) often had to resort to assumptions about the underlying quantum state  $\rho$ , including, amongst others, conservation of angular momentum [36], subtraction of accidentals [37], perfect correlations in a

desired basis [38], or the assumption that the experimentally generated state is pure [39]. Although relying on such assumptions allows for a plausible quantification of entanglement dimensionality, it is not enough for unambiguous certification, which is desirable for secure quantum communication based on high-dimensional entanglement. The certification of the Schmidt number of a state is crucial for this task since a high-dimensional entangled state with low Schmidt number is LOCC equivalent to a low-dimensional entangled state. Hence, unwieldy or inefficient entanglement estimation would strongly mitigate possible advantages of high-dimensional encoding. It is therefore of high significance to determine efficient and practical strategies for certifying high-dimensional states and quantifying their entanglement.

Here, we present an efficient adaptive method that is tailored to better harvest the information about entangled states generated in a given experiment, without the need for any assumptions about the (generally mixed) underlying state and requiring measurements in only two global product bases, regardless of the dimension of the state. Our certification method can be implemented in any physical platform that is suitable for high-dimensional quantum information processing. For the purpose of assumption-free state certification, we certify the fidelity  $F(\rho, \Phi)$  of the experimental state  $\rho$  to a previously identified suitable target state  $|\Phi\rangle$ . We show that measurements in only two global product bases,  $\{|mn\rangle\}_{m,n}$  and  $\{|\tilde{i}\tilde{j}^*\rangle\}_{i,j}$ , are sufficient to select  $|\Phi\rangle$  and to bound the fidelity from below by a quantity  $\tilde{F}(\rho, \Phi) \leq F(\rho, \Phi)$ . For the purpose of assumption-free entanglement certification and quantification, we use our fidelity bound  $\tilde{F}(\rho, \Phi)$  to certify the Schmidt number of the state.

One of the most surprising consequences of our results is the fact that *all pure bipartite quantum states in any dimension can be faithfully certified by measurements in only two global product bases*. We prove this statement by deriving a tight lower bound to the fidelity with an appropriately chosen target state. All that is required for this certification is an educated guess of the corresponding Schmidt bases, which can be inferred from the physical setup at hand for all typical quantum optical platforms. The more accurate the identification of these bases, the higher the confidence in the certified state.

For any identified target state  $|\Phi\rangle$ , the fidelity bound becomes exact when the setup indeed generates the pure state  $|\Phi\rangle$  or the mixed state obtained by dephasing  $|\Phi\rangle$ . We demonstrate that this method can be generalized to measurements in multiple global product bases, yielding  $\tilde{F}^{(M)}(\rho, \Phi)$ , in which  $M + 1$  is the total number of measurements bases, and in prime dimensions the fidelity bounds using measurements in  $d+1$  bases ( $M = d$ ) become exact for all states, i.e.,  $\tilde{F}^{(d)}(\rho, \Phi) = F(\rho, \Phi)$ . Moreover, deriving general decompositions for dephased maximally entangled states further

allows us to prove that unbiased measurement bases indeed provide a necessary and sufficient condition for tight *Schmidt number bounds* for all pure states  $\rho = |\Phi\rangle\langle\Phi|$  and for maximally entangled states subject to pure dephasing. Our method can also be used for *entanglement quantification* by providing lower bounds on the entanglement of formation [92, 93]. Here, our bounds outperform previous methods in terms of their noise robustness and the number of certified e-bits [94]. Finally, our bounds are also shown to be applicable for the certification of certain types of multipartite quantum states.

To put these theoretical predictions to the test in realistic circumstances with actual noise, we devise and carry out an experiment based on photons entangled in their orbital angular momentum, allowing our approach to prove its mettle. In our experimental implementation, measurements are realised using computer programmable holograms implemented on spatial light modulators (SLMs). Employing the theoretical methods developed here, we are able to certify high target-state fidelities and verify record entanglement dimensionality: 9-dimensional entanglement in 11-dimensional subspaces, without any assumptions on the state itself. We use our experimental setup to fully explore the performance of our criteria for non-maximally entangled states, showcasing the flexibility of the derived results.

## 1.2 Entanglement dimensionality

Consider a typical laboratory situation for preparing a high-dimensional quantum system in a bipartite state  $\rho$  that is to be employed for quantum information processing between two parties. In order to be useful, this state should be close to some highly entangled target state that is normally required to have a high purity. Let us therefore consider a pure target state  $|\Phi\rangle$  with a desired Schmidt rank  $k = k_{\max}$ . The Schmidt rank is a measure of the entanglement dimensionality of the state and represents the minimum number of levels one needs to faithfully represent the state and its correlations in any global product basis. Ideally, the target state's Schmidt rank is equal (or close) to the (accessible) local dimension,  $k_{\max} = d$ , where we take the local Hilbert spaces to have the same dimension,  $\dim(\mathcal{H}^A) = \dim(\mathcal{H}^B) = d$ . For mixed states  $\rho$  the Schmidt rank generalizes to the Schmidt number

$$k(\rho) = \inf_{\mathcal{D}(\rho)} \left\{ \max_{|\psi_i\rangle \in \{p_i, |\psi_i\rangle\}_i} \left\{ \text{rank}(\text{Tr}_B|\psi_i\rangle\langle\psi_i|) \right\} \right\}, \quad (1.1)$$

where the infimum is taken over all pure state decompositions, i.e.,  $\mathcal{D}(\rho)$  is the set of all sets  $\{(p_i, |\psi_i\rangle)\}_i$  for which  $\rho = \sum_i p_i |\psi_i\rangle\langle\psi_i|$ ,  $\sum_i p_i = 1$ , and  $0 \leq p_i \leq 1$ .

The Schmidt number hence quantifies the maximal local dimension in which any of the pure state contributions to  $\rho$  can be considered to be entangled and we

hence call  $k$  the *entanglement dimensionality* of  $\rho$ . Note that this implies  $k \leq d$ . For example, any two-qubit entangled state (for which  $d = 2$ ) has an entanglement dimensionality  $k = 2$ . A higher-dimensional entangled state, like a two-qutrit state ( $d = 3$ ), could have Schmidt number  $k = 3$ , in which case it would indeed carry qutrit entanglement, or it could have only  $k = 2$ , in which case the state would be LOCC equivalent to a two-qubit entangled state. In the latter example, even though the state has a higher local dimension, the entanglement dimensionality, which is our quantity of interest, is not higher. Trivially, all separable states have  $k = 1$ .

### 1.3 Target state identification

The task at hand is to certify that the state  $\rho$  generated in the lab is indeed close to the intended target state  $|\Phi\rangle$  and thus provides the desired high-dimensional entanglement. One immediate first approach is to start with local projective measurements in the local Schmidt bases, i.e., the global product basis  $\{|mn\rangle\}_{m,n=0}^{d-1}$ , which we designate as our *standard basis*. These bases can typically be identified from conserved quantities or the setup design, but depending on the physical setup, the corresponding measurements are realised in different ways. In essence, a good choice for the standard basis provides a good target state. For instance, in an optical setting using OAM (as we employ in the experiment reported in this article) the chosen standard basis is the Laguerre-Gauss (LG) basis. In this case, these measurements are performed by coincidence post-selection after local projective filtering. That is, SLMs programmed with the phase pattern of a specific state  $|mn\rangle$  act as local unitary operations, which are followed by single mode fibers (SMF) as local filters, and the number  $N_{mn}$  of coincidences between local photon detectors is counted for each setting corresponding to fixed values of  $m$  and  $n$ . In this way one can obtain the matrix elements

$$\langle mn|\rho|mn\rangle = \frac{N_{mn}}{\sum_{k,l} N_{kl}}. \quad (1.2)$$

A measurement in one global product basis can be realized by one  $d$ -outcome local measurement or equivalently replaced by  $d$  single-outcome local measurements. The latter case employs the use of  $d$  local filter settings ( $d^2$  filter settings globally) to obtain the values  $\langle mm|\rho|mm\rangle$ . These are used to nominate a *target state*  $|\Phi\rangle = \sum_{m=0}^{d-1} \lambda_m |mm\rangle$  by identifying

$$\lambda_m = \sqrt{\frac{\langle mm|\rho|mm\rangle}{\sum_n \langle nn|\rho|nn\rangle}}. \quad (1.3)$$

This association alone by no means guarantees that the state  $\rho$  really is equivalent to the target state  $|\Phi\rangle$ . Although the information about the diagonal elements of  $\rho$

provides an informed guess, it is not enough to infer entanglement properties. In order to access this information, one could in principle perform costly FST. This requires measurements in  $(d+1)^2$  global product bases [91], which is equivalent to  $d^2(d+1)^2$  global filter settings. Here, we propose a much more efficient alternative method to obtain a lower bound on the Schmidt rank of  $\rho$  and on its fidelity to the target state.

## 1.4 Dimensionality witnesses

For the certification of the Schmidt rank of  $\rho$  we consider the fidelity  $F(\rho, \Phi)$  to the target state  $|\Phi\rangle$ , given by

$$F(\rho, \Phi) = \text{Tr}(|\Phi\rangle\langle\Phi|\rho) = \sum_{m,n=0}^{d-1} \lambda_m \lambda_n \langle mm|\rho|nn\rangle. \quad (1.4)$$

For any state  $\rho$  of Schmidt rank  $k \leq d$  the fidelity of Eq. (1.4) is bounded by [95, 96]

$$F(\rho, \Phi) \leq B_k(\Phi) := \sum_{m=0}^{k-1} \lambda_{i_m}^2, \quad (1.5)$$

where the sum runs over the  $k$  largest Schmidt coefficients, i.e.,  $i_m, m \in \{0, \dots, d-1\}$  such that  $\lambda_{i_m} \geq \lambda_{i_{m'}} \forall m' \leq m$ . Consequently, any state for which  $F(\rho, \Phi) > B_k(\Phi)$  is incompatible with a Schmidt rank of  $k$  or less, implying an entanglement dimensionality of at least  $k+1$ .

## 1.5 Fidelity bounds

The next step is hence to experimentally estimate the value of the fidelity  $F(\rho, \Phi)$ . To see how this can be done, we split the fidelity into two contributions, one that depends on the terms of Eq. (1.4) that are diagonal in the basis  $\{|mn\rangle\}_{m,n}$ , which will be called  $F_1(\rho, \Phi)$ , and the other that depends on the off-diagonal terms, called  $F_2(\rho, \Phi)$  (see Sec. 1.8).

The contribution  $F_1(\rho, \Phi)$  can be calculated directly from the already performed measurements in the basis  $\{|mn\rangle\}_{m,n}$ . However, exactly determining the term  $F_2(\rho, \Phi)$  would require a number of measurements that scales with the dimension. To avoid such a high overhead, we employ bounds for  $F_2(\rho, \Phi)$  that can be calculated from measurements in only one additional basis  $\{|\tilde{j}\rangle\}_j$  (see Sec. 1.8).

Number of measurements			
Method	FST	$F(\rho, \Phi)$	$\tilde{F}(\rho, \Phi)$
Global product bases	$(d+1)^2$	$d+1$	2
Local filter settings	$(d+1)^2 d^2$	$(d+1)d^2$	$2d^2$

Table 1.1: The table shows the number of required measurements for optimal full state tomography (FST), optimal fidelity measurement ( $F(\rho, \Phi)$ ), and to calculate the fidelity bounds presented in this work ( $\tilde{F}(\rho, \Phi)$ ). The first line corresponds to the necessary number of measured global product bases (which can be realised with at most  $d+1$ -outcome local measurements), and the second line, the necessary number of local filter settings (which can be realised with single-outcome local measurements).

Using the previously obtained values  $\{\lambda_m\}_m$ , we define the basis  $\{|\tilde{j}\rangle\}_{j=0,\dots,d-1}$  according to

$$|\tilde{j}\rangle = \frac{1}{\sqrt{\sum_n \lambda_n}} \sum_{m=0}^{d-1} \omega^{jm} \sqrt{\lambda_m} |m\rangle, \quad (1.6)$$

where  $\omega = e^{2\pi i/d}$  and  $\{|m\rangle\}_m$  is the standard basis. Notice that, although the basis vectors  $|\tilde{j}\rangle$  are normalized by construction, they are not necessarily orthogonal, but become orthogonal and even mutually unbiased w.r.t. to  $\{|m\rangle\}_m$  when all  $\lambda_m$  are the same. We hence refer to  $\{|\tilde{j}\rangle\}_j$  as the *tilted basis*.

Due to this general non-orthogonality, the relation of Eq. (1.2) between the diagonal matrix elements  $\langle \tilde{i}\tilde{j}^* | \rho | \tilde{i}\tilde{j}^* \rangle$  and the coincidence counts  $\tilde{N}_{ij}$  for the local filter setting  $|\tilde{i}\tilde{j}^*\rangle$  requires a small modification in terms of an additional normalization factor  $c_\lambda := \frac{d^2}{(\sum_k \lambda_k)^2} \sum_{m,n} \lambda_m \lambda_n \langle mn | \rho | mn \rangle$ , i.e.,

$$\langle \tilde{i}\tilde{j}^* | \rho | \tilde{i}\tilde{j}^* \rangle = \frac{\tilde{N}_{ij}}{\sum_{k,l} \tilde{N}_{kl}} c_\lambda. \quad (1.7)$$

Apart from the inclusion of  $c_\lambda$  (see detailed derivation in Appendix A), measurements in the tilted basis are in principle not different from measurements in any orthonormal basis.

The terms of Eq. (1.7), along with the measurement results in the standard basis, allow us to bound the fidelity term  $F_2(\rho, \Phi)$ , which in turn provides a lower bound  $\tilde{F}(\rho, \Phi)$  for the fidelity  $F(\rho, \Phi)$  that is experimentally easily accessible.

We thus immediately obtain the dimensionality witness inequality

$$\tilde{F}(\rho, \Phi) \leq F(\rho, \Phi) \leq B_k(\Phi), \quad (1.8)$$

which is satisfied by any state  $\rho$  with Schmidt rank  $k$  or less. Conversely, the entanglement dimensionality  $d_{\text{ent}}$  that is certifiable with our method is the maximal  $k$  such that  $\tilde{F}(\rho, \Phi) > B_{k-1}(\Phi)$ .

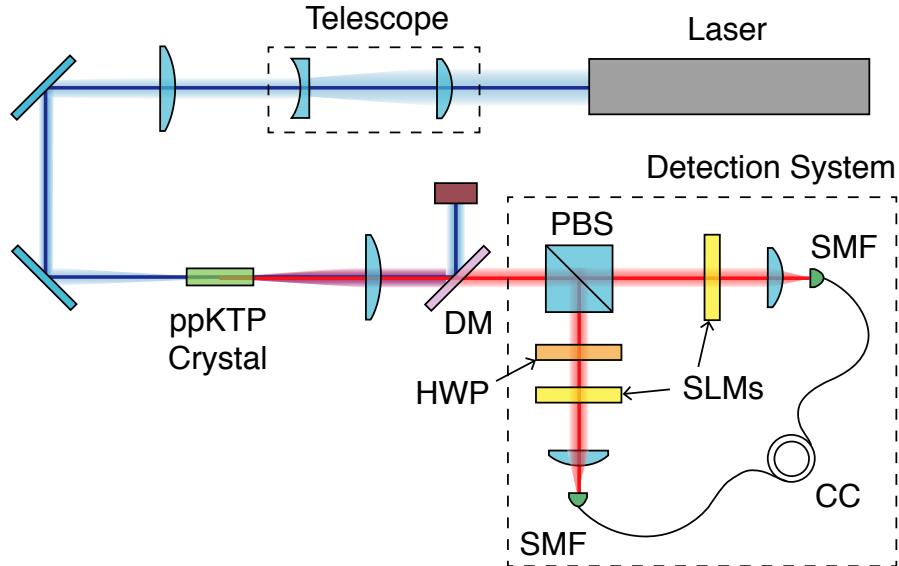
A detailed derivation of this bound along with the proofs of its tightness can be found in the Sec. 1.8. In Appendix A we further present a generalization of the fidelity bound to multiple measurement bases, the derivation of bounds for entanglement of formation that arise from our method, and an extension of our fidelity bound to a family of multipartite states.

Crucially, our witness requires only 2 global product bases to be evaluated, and is hence significantly more efficient than the  $d+1$  and  $(d+1)^2$  bases required for the exact evaluation of the fidelity or even a FST, respectively. For projective filtering the overall number of filter settings is obtained by multiplying the number of required bases by  $d^2$ . A comprehensive comparison of required number of measurement settings is given in Table 1.1.

## 1.6 Experimental certification of high-dimensional entanglement

We now apply our witness to certify high-dimensional orbital-angular-momentum (OAM) entanglement between two photons generated by Type-II SPDC in a non-linear ppKTP crystal (see Fig. 1.1 for details). To this end, we display computer-programmed holograms (Fig. 1.2(a) and (b)) on spatial light modulators (SLMs) designed to manipulate the phase and amplitude of incident photons [97]. In this manner, we are able to projectively measure the photons in any spatial mode basis, e.g., the Laguerre-Gaussian (LG) basis, any mutually unbiased (MUB) [98] or any tilted basis (TILT) composed of superpositions of elements of the standard basis (Eq. (1.6)). Additional details of the experimental implementation, including information on the holograms, can be found in Sec. 1.8 Methods and in Appendix A.

For local dimensions up to  $d = 11$  (i.e., for azimuthal quantum numbers  $\ell \in \{-5, \dots, 5\}$ ) we then proceed in the following way. First, we measure the two-photon state in the LG basis  $\{|m\rangle\}_m$  to obtain a cross-talk matrix of coincidence counts  $N_{mn}$  (Fig. 1.3 (a)), taking into account the effects of mode-dependent loss (see Appendix A). This allows us to calculate the density matrix elements  $\langle mn|\rho|mn\rangle$ , estimate the  $\lambda_m$ , and nominate the target state  $|\Phi\rangle$ . We then use the set  $\{\lambda_m\}_m$  to construct the *tilted* basis  $\{|\tilde{j}\rangle\}_j$  according to Eq. (1.6) and perform correlation measurements (Fig. 1.3 (b)) that allow us to calculate  $\langle \tilde{j}\tilde{j}^*|\rho|\tilde{j}\tilde{j}^*\rangle$ . From these measurements, we calculate the lower bound of the fidelity to the target state, for which we find high



**Figure 1.1: Experimental setup.** A 405nm CW laser pumps a 5mm ppKTP crystal to generate a pair of infrared (810nm) photons via the process of Type-II spontaneous parametric down-conversion (SPDC), which are entangled in their orbital angular momentum (OAM). The pump is removed by a dichroic mirror (DM) and the two photons are separated by a polarizing beam splitter (PBS) and incident on two phase-only spatial light modulators (SLMs). A half-wave plate (HWP) is used to rotate the polarization of the reflected photon from vertical to horizontal, allowing it to be manipulated by the SLM. In combination with single-mode fibers (SMFs), the SLMs act as spatial mode filters. The filtered photons are detected by single-photon avalanche photodiodes (not shown) and time-coincident events are registered by a coincidence counting logic (CC).

values, e.g.,  $\tilde{F}(\rho, \Phi) = 76.2 \pm 0.6\%$  for  $d = 11$  (data for other dimensions is presented in Table 1.2). However, in our setup, the certification thresholds  $B_k$  for the tilted basis are higher than for the MUB (e.g.,  $B_7 = 0.72$  vs  $B_7 = 0.64$  for  $d = 11$  in tilted versus MUB respectively). We therefore also measure the correlations in the first MUB  $\{|j\rangle\}_j$  (Fig. 1.3 (c)) following the standard MUB construction by Wootters et al. [98], corresponding to  $\lambda_m = 1/\sqrt{d}$  for all  $m$  in Eq. (1.6). Using these measurements, we calculate lower bounds of the fidelity to the maximally entangled state, and find  $\tilde{F}(\rho, \Phi^+) = 74.8 \pm 0.4\%$  for  $d = 11$ , which is significantly above the bound of  $B_8(\Phi^+) = \frac{8}{11} \approx 0.727$ , but below  $B_9(\Phi^+) = \frac{9}{11} \approx 0.818$ . We hence certify 9-dimensional entanglement in this way. Note that the asymmetry in the counts just below and above the diagonal in Figs. 1.3 (b) and (c) corresponds to a slight misalignment in the experiment. Errors in the fidelity are calculated by propagating statistical Poissonian errors in photon-count rates via Monte-Carlo simulation of the experiment. This demonstrates that our witness indeed works for efficiently certifying high-dimensional entanglement. Moreover, this shows that although the

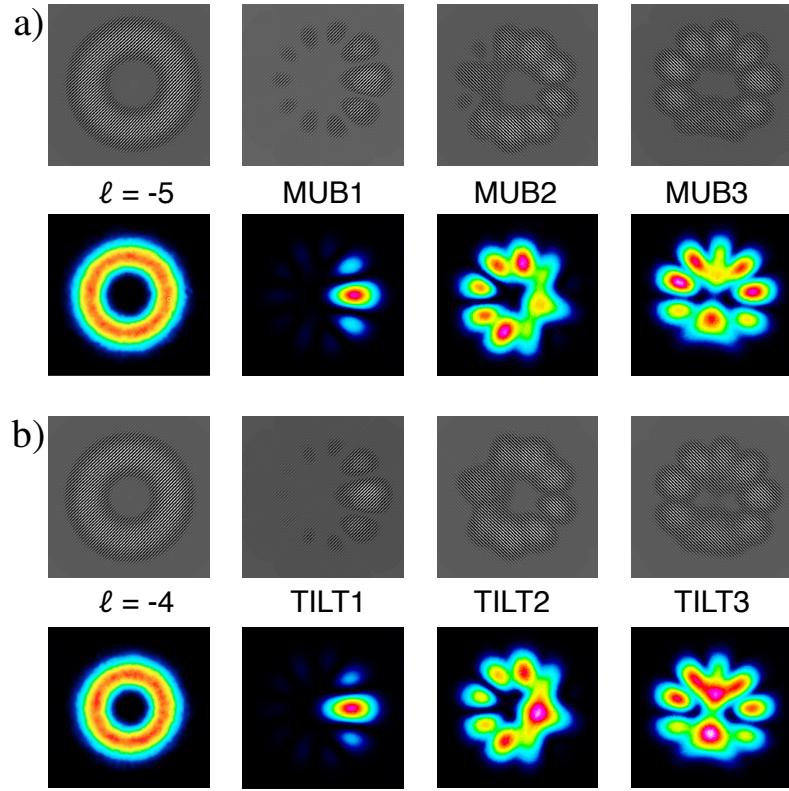


Figure 1.2: **Holograms.** (a) and (b) upper rows: examples of computer-generated holograms displayed on the SLMs for measuring the photons in a  $d = 11$  dimensional space; (a) and (b) left panels: standard LG basis modes with azimuthal quantum number  $\ell = -5$  and  $-4$ ; right panels of (a): 3 basis states from a MUB (denoted MUB1, MUB2, MUB3); right panel of (b): 3 basis states from a tilted basis (Eq. (1.6)) (denoted TILT1, TILT2, TILT3); (a) and (b) lower rows: intensity images of the modes filtered by these holograms (see Appendix A for details on how these intensity images were obtained).

tilted basis measurements can achieve higher fidelities, one pays a price in terms of increased certification thresholds, and thus an increased sensitivity to noise.

Our approach hence provides a lower bound for  $F(\rho, \Phi)$  and  $k(\rho)$  using measurements in as little as 2 global product bases. Each of these are realized by  $d$  local filter settings on each side, totalling to  $2d^2$  global filter settings instead of  $d^2(d + 1)^2$  for FST. For our state in a  $11 \times 11$ -dimensional Hilbert space this corresponds to 242 filter settings, versus the 17,424 filter settings required for FST, which is a reduction by two orders of magnitude.

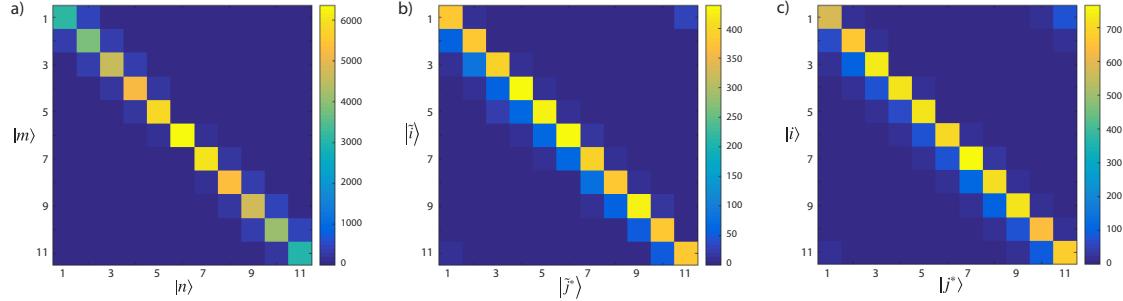


Figure 1.3: **Experimental data certifying 9-dimensional entanglement.** Two-photon coincidence counts showing orbital angular momentum correlations in: (a) the standard LG basis  $\{|m\rangle, |n\rangle\}_{m,n}$ , (b) the tilted basis  $\{|\tilde{i}\rangle, |\tilde{j}^*\rangle\}_{i,j}$ , and (c) the first mutually unbiased basis  $\{|i\rangle, |j^*\rangle\}_{i,j}$ . As seen in (a), our generated state is not maximally entangled (measured Schmidt coefficients  $\lambda_m$  can be found in Appendix A). For each set of two-basis measurements, we calculate a fidelity to the  $d = 11$  target state of  $\tilde{F}(\rho, \Phi) = 76.2 \pm 0.6\%$  (LG and tilted bases), and  $\tilde{F}(\rho, \Phi^+) = 74.8 \pm 0.4\%$  (LG and MUB). Even though the fidelity bound in the tilted case (b) is higher, the Schmidt number bounds are also higher and more difficult to overcome, yielding a certified entanglement dimensionality of  $d_{\text{ent}} = 8$ , slightly lower than the bound of  $d_{\text{ent}} = 9$  obtained in the MUB case (c).

### Experimental results

$d$	$d_{\text{ent}}$	$\tilde{F}(\rho, \Phi^+)$	$\tilde{F}(\rho, \Phi)$
3	3	$91.5 \pm 0.4\%$	$92.5 \pm 0.4\%$
5	5	$89.9 \pm 0.4\%$	$90.0 \pm 0.5\%$
7	6	$84.2 \pm 0.5\%$	$86.9 \pm 0.6\%$
11	9	$74.8 \pm 0.4\%$	$76.2 \pm 0.6\%$

Table 1.2: Fidelities  $\tilde{F}(\rho, \Phi^+)$  and  $\tilde{F}(\rho, \Phi)$  to the maximally entangled state and to the target state, obtained via measurements in two MUBs and two ( $M = 1$ ) tilted bases in dimension  $d$ , respectively. The second column lists the entanglement dimensionality  $d_{\text{ent}}$  certified using  $\tilde{F}_2(\rho, \Phi^+)$ .

## 1.7 Discussion and outlook

A remarkable trait of high-dimensional entanglement is that measurements in two bases are enough to certify any entangled pure state for arbitrarily large Hilbert space dimension. We make use of this insight to establish fidelity bounds for states produced under realistic laboratory conditions. Using two (or, if desirable more, see Appendix A) local basis choices, these bounds can be employed to certify the Schmidt rank and entanglement of formation in a much more efficient way than is possible via full state tomography or even complete measurements of the fidelity. It is also interesting to note that the two measurement bases required for optimal fidelity certification become unbiased whenever the target state is maximally entangled. This procedure could be viewed as a trusted device analogue to self-testing [99], requiring significantly fewer measurements and exhibiting a much greater noise resistance.

The strength of our method has its origin in the fact that we use readily available knowledge about the quantum system under investigation in terms of an educated guess for the Schmidt bases. This is close in spirit to assumptions commonly used in many experiments where preserved quantities in non-linear processes are harnessed to create entanglement. For the case of our experimental setup, this amounted to the conservation of transverse momenta [21]. Using holograms and couplings to single-mode fibers essentially implements single-outcome measurements (projective filtering), leading to  $2d^2$  filter settings globally. This could be further improved by means of a mode sorter [100, 101], reducing the global measurement settings to merely 2 (see Table 1.1 for a comprehensive overview) at the cost of using  $d$ -coincidence detectors. But our proposed method is not limited to transverse momenta and OAM. In energy-time based setups [23], conservation of energy leads to the frequency or time-bin basis to be the natural Schmidt basis. Canonically these systems even feature  $d$ -outcome measurements, making them ideal candidates for the application of our method. Indeed, the states generated in the time-bin basis are generically close to being maximally entangled [25] and thus the tilted measurement would ideally be close to MUBs. There are various proposals as to how mutually unbiased measurements could also be directly implemented as  $d$ -outcome measurements in such systems [102, 103]. Finally, our method can be directly implemented using multi-path interferometers [18] where the natural Schmidt basis is the path degree of freedom. Let us stress again, however, that even deviations from the assumed situation do not invalidate the bounds employed in our approach, but lead (at most) to suboptimal performance, and an unambiguous certification is still ensured.

To demonstrate the practical utility of our method, we have performed an

experiment using two photons entangled in their orbital angular momenta. We were able to certify 9-dimensional entanglement in a  $11 \times 11$ -dimensional Hilbert space, which is the highest number achieved so far without further assumptions on the underlying quantum state. This is achieved using only two local, unbiased measurement bases (11-outcomes each), which are realized by 242 local filters and coincidence counting. Using similar measurements in the tilted bases we are able to achieve target state fidelities of 92.5% in 3 dimensions and 76.2% in 11 dimensions. As we have shown, the certification method proposed here is thus surprisingly robust to noise and enables straightforward and assumption-free entanglement characterization in realistic quantum optics experiments. This further illustrates the usefulness of MUBs for the detection of entanglement [94, 104–109] and correlations [110].

Our certification method can also be generalized to operate with more than two bases, enabling an adaptable increase in noise resistance when required, as discussed in Appendix A. There we also show how our bounds can be extended to certify entanglement of formation. Remarkably, this approach can also be generalized to Greenberger-Horne-Zeilinger (GHZ)-like multipartite states recently created using OAM [37, 111], making large multipartite states generated by the methods of Ref. [112] certifiable in a scalable manner. We give a brief exposition of this result in Appendix A.

## 1.8 Methods

### 1.8.1 Derivation of the fidelity lower bound

In this section, we provide a proof for the fidelity bound

$$F(\rho, \Phi) \geq \tilde{F}(\rho, \Phi), \quad (1.9)$$

i.e., the right-hand side of Eq. (1.8) of the main text, where  $F(\rho, \Phi) = F_1(\rho, \Phi) + F_2(\rho, \Phi)$  and  $\tilde{F}(\rho, \Phi) = F_1(\rho, \Phi) + \tilde{F}_2(\rho, \Phi)$ , each split into two contributions. Since the first of these, given by

$$F_1(\rho, \Phi) := \sum_m \lambda_m^2 \langle mm | \rho | mm \rangle, \quad (1.10)$$

is the same for both  $F$  and  $\tilde{F}$ , we hence want to concentrate on showing that  $F_2 \geq \tilde{F}_2$ , where

$$F_2(\rho, \Phi) := \sum_{m \neq n} \lambda_m \lambda_n \langle mm | \rho | nn \rangle, \quad (1.11)$$

whereas the lower bound to  $F_2(\rho, \Phi)$  is

$$\begin{aligned} \tilde{F}_2 := & \frac{(\sum_m \lambda_m)^2}{d} \sum_{j=0}^{d-1} \langle \tilde{j} \tilde{j}^* | \rho | \tilde{j} \tilde{j}^* \rangle - \sum_{m,n=0}^{d-1} \lambda_m \lambda_n \langle m n | \rho | m n \rangle \\ & - \sum_{\substack{m \neq m', m \neq n \\ n \neq n', n' \neq m'}} \tilde{\gamma}_{mm'nn'} \sqrt{\langle m' n' | \rho | m' n' \rangle \langle m n | \rho | m n \rangle}, \end{aligned} \quad (1.12)$$

where the asterisk denotes complex conjugation of the vector components w.r.t.  $\{|m\rangle\}_m$  and the prefactor  $\tilde{\gamma}_{mm'nn'}$  is given by

$$\tilde{\gamma}_{mm'nn'} = \begin{cases} 0 & \text{if } (m - m' - n + n') \bmod d \neq 0 \\ \sqrt{\lambda_m \lambda_n \lambda_{m'} \lambda_{n'}} & \text{otherwise,} \end{cases} \quad (1.13)$$

as we will show in the following. Here, the quantity  $F_1(\rho, \Phi)$ , as well as the second and third terms of  $\tilde{F}_2$  in Eq. (1.12) can be obtained directly from measurements in the standard basis  $\{|mn\rangle\}_{m,n}$ , whereas the first term of  $\tilde{F}_2$  is constructed from diagonal density matrix elements w.r.t. to the tilted bases with elements

$$|\tilde{j}\rangle = \frac{1}{\sqrt{\sum_n \lambda_n}} \sum_{m=0}^{d-1} \omega^{jm} \sqrt{\lambda_m} |m\rangle, \quad (1.14)$$

where  $\omega = e^{2\pi i/d}$ . This non-orthogonal construction is motivated by the observations that  $|\Phi\rangle$  is in general non-maximally entangled and that the tilted basis interpolates between the measurement bases required to obtain unit fidelities for pure product states  $|\Phi\rangle = |mn\rangle$  (where the standard basis suffices) and for maximally entangled states  $|\Phi\rangle = |\Phi^+\rangle$  (where the tilted basis becomes unbiased w.r.t. to the standard basis). The tilted basis  $\{|\tilde{j}\rangle\}_j$  can be seen as a particular construction of a basis that satisfies the condition  $|\langle m | \tilde{j} \rangle|^2 = \lambda_m \lambda_j \forall m, j$  with the standard basis  $\{|m\rangle\}_m$ . Notice that the standard definition of mutually unbiased bases (MUBs) is recovered when  $\lambda_i = \frac{1}{\sqrt{d}} \forall i$ .

For the proof, we then focus on the matrix elements obtained from measurements w.r.t. the tilted basis. That is, we define the quantity

$$\Sigma := \sum_{j=0}^{d-1} \langle \tilde{j} \tilde{j}^* | \rho | \tilde{j} \tilde{j}^* \rangle \quad (1.15)$$

$$= \frac{1}{(\sum_k \lambda_k)^2} \sum_{\substack{m,m' \\ n,n'}} \sqrt{\lambda_m \lambda_n \lambda_{m'} \lambda_{n'}} \times \sum_{j=0}^{d-1} \omega^{j(m-m'-n+n')} \langle m' n' | \rho | m n \rangle. \quad (1.16)$$

The sums over the standard basis components can then be split into several contributions. When  $m = m'$  and  $n = n'$ , the phases all cancel, the sum over the tilted

basis elements has  $d$  equal contributions, and we hence have

$$\Sigma_1 := \frac{d}{(\sum_k \lambda_k)^2} \sum_{m,n} \lambda_m \lambda_n \langle mn | \rho | mn \rangle. \quad (1.17)$$

When  $m = m'$  but  $n \neq n'$  (or vice versa) one finds terms containing the sum

$$\sum_{j=0}^{d-1} \omega^{j(n'-n)} = \delta_{nn'}. \quad (1.18)$$

Since  $n \neq n'$ , these terms vanish. For all remaining contributions to  $\Sigma$  one has  $m \neq m'$  and  $n \neq n'$ . These terms then again split into three sets. First, for  $m = n$  and  $m' = n'$  we recover the desired terms of the form

$$\Sigma_2 := \frac{d}{(\sum_k \lambda_k)^2} \sum_{m \neq n} \lambda_m \lambda_n \langle mm | \rho | nn \rangle, \quad (1.19)$$

which also appear in  $F_2(\rho, \Phi)$  in Eq. (1.11). The terms where  $m = n$  but  $m' \neq n'$  (or vice versa) again vanish due to Eq. (1.18). Finally, this leaves the term

$$\Sigma_3 := \frac{1}{(\sum_k \lambda_k)^2} \sum_{\substack{m \neq m' \\ m \neq n \\ n \neq n' \\ n' \neq m'}} \sqrt{\lambda_m \lambda_n \lambda_{m'} \lambda_{n'}} \times \sum_{j=0}^{d-1} \omega^{j(m-m'-n+n')} \langle m'n' | \rho | mn \rangle, \quad (1.20)$$

$$= \frac{1}{(\sum_k \lambda_k)^2} \sum_{\substack{m \neq m' \\ m \neq n \\ n \neq n' \\ n' \neq m'}} \sqrt{\lambda_m \lambda_n \lambda_{m'} \lambda_{n'}} \times \text{Re} \left( c_{mn m' n'} \langle m'n' | \rho | mn \rangle \right), \quad (1.21)$$

where we have used the abbreviation  $c_{mn m' n'} := \sum_j \omega^{j(m-m'-n+n')}$ . In the last step we have replaced  $c_{mn m' n'}$  by its real part, since for each combination of values for  $m, n, m', n'$  the sum contains a term where the pairs  $(m, n)$  and  $(m', n')$  are exchanged. Each term in the sum is hence paired with another term that is its complex conjugate, and the total sum is hence real.

While  $\Sigma_1$  and  $\Sigma_2$  are accessible via measurements in the standard basis, the off-diagonal matrix elements in  $\Sigma_3$  cannot be obtained from measurements w.r.t.  $\{|mn\rangle\}_{m,n}$ . In order to provide a useful lower bound for  $\Sigma$  we therefore have to provide a bound for  $\Sigma_3$ . To this end, we can bound the real part by the modulus, i.e.,

$$\text{Re} \left( c_{mn m' n'} \langle m'n' | \rho | mn \rangle \right) \leq |c_{mn m' n'} \langle m'n' | \rho | mn \rangle| \quad (1.22)$$

$$= |c_{mn m' n'}| \cdot |\langle m'n' | \rho | mn \rangle|. \quad (1.23)$$

We then use the Cauchy-Schwarz inequality to bound the second factor on the right-hand side of (1.23) by writing  $\rho = \sum_i p_i |\psi_i\rangle\langle\psi_i|$  such that

$$|\langle m'n'|\rho|mn\rangle| = \left| \sum_i \sqrt{p_i} \langle m'n'|\psi_i\rangle \sqrt{p_i} \langle\psi_i|mn\rangle \right| \quad (1.24)$$

$$\leq \sqrt{\sum_i p_i \langle m'n'|\psi_i\rangle\langle\psi_i|m'n'\rangle} \times \sqrt{\sum_i p_i \langle mn|\psi_i\rangle\langle\psi_i|mn\rangle} \quad (1.25)$$

$$= \sqrt{\langle m'n'|\rho|m'n'\rangle\langle mn|\rho|mn\rangle}. \quad (1.26)$$

In Eq. (1.23), note that in the first factor,  $|c_{mnm'n'}|$ , the sum  $\sum_j \omega^{j(m-m'-n+n')}$  vanishes whenever  $(m-m'-n+n') \bmod d \neq 0$ , and equals to  $d$  otherwise. Collecting  $c_{mnm'n'}/d$  with  $\sqrt{\lambda_m \lambda_n \lambda_{m'} \lambda_{n'}}$  into  $\tilde{\gamma}_{mm'nn'}$  as defined in Eq. (1.13), this allows us to bound the quantity  $\Sigma_3$  according to

$$\Sigma_3 \leq \frac{d}{(\sum_k \lambda_k)^2} \sum_{\substack{m \neq m' \\ m \neq n \\ n \neq n' \\ n' \neq m' \\ m - m' - n + n' = 0}} \tilde{\gamma}_{mm'nn'} \sqrt{\langle m'n'|\rho|m'n'\rangle\langle mn|\rho|mn\rangle}. \quad (1.27)$$

Collecting the different contributions to  $\Sigma$  we thus have

$$\Sigma = \Sigma_1 + \Sigma_2 + \Sigma_3 = \sum_{j=0}^{d-1} \langle \tilde{j}\tilde{j}^* | \rho | \tilde{j}\tilde{j}^* \rangle \quad (1.28)$$

$$\begin{aligned} &\leq \frac{d}{(\sum_k \lambda_k)^2} \left( \sum_{m,n} \lambda_m \lambda_n \langle mn|\rho|mn\rangle + \sum_{m \neq n} \lambda_m \lambda_n \langle mm|\rho|nn\rangle \right. \\ &\quad \left. + \sum_{\substack{m \neq m' \\ m \neq n \\ n \neq n' \\ n' \neq m' \\ m - m' - n + n' = 0}} \tilde{\gamma}_{mm'nn'} \sqrt{\langle m'n'|\rho|m'n'\rangle\langle mn|\rho|mn\rangle} \right). \end{aligned} \quad (1.29)$$

Conversely, this means that the term  $F_2$  can be bounded by

$$F_2 = \sum_{m \neq n} \lambda_m \lambda_n \langle mm|\rho|nn\rangle \quad (1.30)$$

$$\begin{aligned} &\geq \frac{(\sum_k \lambda_k)^2}{d} \sum_{j=0}^{d-1} \langle \tilde{j}\tilde{j}^* | \rho | \tilde{j}\tilde{j}^* \rangle - \sum_{m,n} \lambda_m \lambda_n \langle mn|\rho|mn\rangle \\ &\quad - \sum_{\substack{m \neq m' \\ m \neq n \\ n \neq n' \\ n' \neq m' \\ m - m' - n + n' = 0}} \tilde{\gamma}_{mm'nn'} \sqrt{\langle m'n'|\rho|m'n'\rangle\langle mn|\rho|mn\rangle}, \end{aligned} \quad (1.31)$$

as claimed for the quantity  $\tilde{F}_2$  in Eq. (1.12). The fidelity  $F(\rho, \Phi)$  can hence be bounded by measurements in only two local bases,  $\{|m\rangle\}_m$  and  $\{|\tilde{j}\rangle\}_j$ , for each party, i.e., two global product bases  $\{|mn\rangle\}_{m,n}$  and  $\{|\tilde{i}\tilde{j}^*\rangle\}_{i,j}$ .

### 1.8.2 Tightness of the fidelity bound

In this section, we show that whenever the system state  $\rho$  is either equal to the (pure) target state  $\rho = |\Phi\rangle\langle\Phi|$  or is a dephased maximally entangled state  $\rho_{\text{deph}}(p) = p|\Phi^+\rangle\langle\Phi^+| + \frac{1-p}{d} \sum_m |mm\rangle\langle mm|$ , the Schmidt number witness  $\tilde{F}(\rho, \Phi) > B_{k-1}(\Phi)$  is not only a sufficient, but also a necessary condition for  $|\Phi\rangle$  or  $\rho_{\text{deph}}$  to have a Schmidt rank larger or equal than  $k$ . For the state  $|\Phi\rangle$  this is obvious. Since the coefficients  $\lambda_m$  are determined by measurements in the Schmidt basis of  $\rho = |\Phi\rangle\langle\Phi|$ , the fidelity bound is tight, and we have  $\tilde{F}(\rho, \Phi) = F(\rho, \Phi) = 1$  and  $B_k(\Phi)$  is equal to 1 if and only if  $k = d$ .

For dephased maximally entangled states we proceed by showing that there exists a Schmidt-rank  $k$  state  $\rho_{\text{deph}}(p = p_k)$  such that  $F(\rho_{\text{deph}}(p_k), \Phi) = B_k(\Phi)$  for every  $k$ . To this end, first note that  $\rho_{\text{deph}}$  can be written as

$$\rho_{\text{deph}} = p|\Phi^+\rangle\langle\Phi^+| + \frac{1-p}{d} \sum_m |mm\rangle\langle mm| \quad (1.32)$$

$$= \frac{1}{d} \sum_m |mm\rangle\langle mm| + \frac{p}{d} \sum_{m \neq n} |mm\rangle\langle nn|, \quad (1.33)$$

which implies that  $\lambda_m = \frac{1}{\sqrt{d}} \forall m$ . That is, the corresponding target state is  $|\Phi\rangle = |\Phi^+\rangle$  and  $B_k = \frac{k}{d}$ . The relevant fidelity then evaluates to

$$F(\rho, \Phi) = F(\rho_{\text{deph}}, \Phi^+) = \frac{1 + p(d-1)}{d}, \quad (1.34)$$

and  $F(\rho_{\text{deph}}, \Phi^+) = B_k$  for  $p = p_k = \frac{k-1}{d-1}$ . All we need to do now is to show that  $\rho_{\text{deph}}(p_k)$  has a Schmidt rank no larger than  $k$ . To see this, consider the family of maximally entangled states in dimension  $k$ , i.e.,

$$|\Phi_\alpha^+\rangle := \frac{1}{\sqrt{|\alpha|}} \sum_{m \in \alpha} |mm\rangle, \quad (1.35)$$

where  $\alpha \subset \{0, 1, \dots, d-1\}$  with cardinality  $|\alpha| = k$ . In dimension  $d$ , we can find  $\binom{d}{k}$  such states and consider their incoherent mixture, i.e.,

$$\rho_k = \frac{1}{\binom{d}{k}} \sum_{\alpha \text{ s.t. } |\alpha|=k} |\Phi_\alpha^+\rangle\langle\Phi_\alpha^+|. \quad (1.36)$$

Since each of the  $\Phi_\alpha^+$  has Schmidt rank  $k$ , the convex sum  $\rho_k$  cannot have a Schmidt rank larger than  $k$ . Since there are  $\binom{d-1}{k-1}$  terms contributing to every nonzero diagonal matrix element, we have  $\langle mn|\rho_k|mn\rangle = \frac{1}{d}\delta_{mn}$ . Similarly, every nonvanishing off-diagonal matrix element has  $\binom{d-2}{k-2}$  contributions, and we hence have  $\langle mn|\rho_k|ij\rangle = \frac{k-1}{d(d-1)}\delta_{mn}\delta_{ij}$  for  $m \neq i$ . It is then easy to see that the fidelity with the maximally

entangled state (in dimension  $d$ ) is  $F(\rho_k, \Phi^+) = \frac{k}{d}$ . More specifically, comparison with Eq. (1.33) reveals that  $\rho_{\text{deph}} = \rho_k$  for  $p = p_k = \frac{k-1}{d-1}$ . Since the Schmidt rank of  $\rho_k$  is smaller or equal than  $k$ , we have hence shown that the Schmidt rank of the dephased maximally entangled state  $\rho_{\text{deph}}(p_k)$  with  $F(\rho_{\text{deph}}(p_k), \Phi) = B_k$  is  $k$  or less. Consequently,  $F(\rho_{\text{deph}}, \Phi^+) > B_{k-1}$  is a necessary and sufficient condition for  $\rho_{\text{deph}}$  to have Schmidt rank  $k$ .

Moreover, since the fidelity bound  $\tilde{F} \leq F$  is tight for  $\rho_{\text{deph}}$  already for  $M = 1$  and the tilted basis is unbiased w.r.t. the standard basis for dephased maximally entangled states, we can conclude that measurements in two unbiased bases provide the necessary and sufficient condition  $\tilde{F}(\rho_{\text{deph}}, \Phi^+) > B_{k-1}$  for Schmidt rank  $k$  for these states.

### 1.8.3 Role of the target state

The initial designation of the target state  $|\Phi\rangle$ , or rather its Schmidt basis, helps to suitably adapt the dimensionality witness to the experimental situation. Although identifying the Schmidt basis from the setup could in principle be seen as an assumption about the underlying state, choosing a basis that is far from the Schmidt basis doesn't invalidate our certification method. Since the latter is based on lower-bounding the fidelity to the target state, such a misidentification would simply result in a reduced performance by using lower bounds on the fidelity to a state that is far from the actual state. An analysis of how our fidelity bounds are affected by a “wrong” choice of basis is provided in Appendix A. In other words, a non-optimal guess can lead to what is called a type-II-error (i.e., a “false negative”), but never to a type-I-error (i.e., a “false positive”). This means that a suboptimal guess of the target state may lead to a less than optimal value for the certified fidelity and/or Schmidt number. The entanglement dimensionality (Schmidt number) certified by a wrong choice of basis may hence be lower than the actual entanglement dimensionality (Schmidt number) of the underlying state  $\rho$ , but never higher. In summary, it can be concluded that the performance of our method may depend on the expected target state, but the method does not require any assumptions about the true system state  $\rho$ .

While this certification method is thus independent of the specific circumstances in the laboratory, it can be noted that it works particularly well for certain types of states. For instance, whenever the target state matches the underlying state up to pure dephasing, i.e., when  $\rho = p|\Phi\rangle\langle\Phi| + \frac{1-p}{d}\sum_m |mm\rangle\langle mm|$ , the fidelity bound  $\tilde{F}(\rho, \Phi) \leq F(\rho, \Phi)$  is tight, since the last term in Eq. (1.12) vanishes in this case. Moreover, whenever these states are pure ( $p = 1$ ) or dephased maximally entangled

states (arbitrary  $p$  but  $|\Phi\rangle = |\Phi^+\rangle$ ) one can further show that the Schmidt number bound  $F(\rho, \Phi) \leq B_k(\Phi)$  is also tight (see Appendix A for derivation), in which case we have  $\tilde{F}(\rho, \Phi) = F(\rho, \Phi) = B_{d_{\text{ent}}}(\Phi)$ .

In addition, it can sometimes be helpful to select a “wrong” target state on purpose. For example, the maximally entangled state  $|\Phi^+\rangle = \frac{1}{\sqrt{d}} \sum_m |mm\rangle$ , i.e., a target state whose coefficients were chosen to be  $\lambda_m = \frac{1}{\sqrt{d}} \forall m$ , may at times offer a higher Schmidt number lower bound than a target state with coefficients  $\lambda_m$  taken from the measurement results in the standard basis, even though the fidelity bound would be lower. In the case of the maximally entangled target state, the tilted basis becomes an orthonormal basis that is mutually unbiased w.r.t. to the standard basis and we have  $B_k(\Phi^+) = \frac{k}{d}$ . Since this bound is lower than for general values of  $\lambda_m$ , it may be easier to overcome, particularly in the presence of noise, and hence yield a higher certified Schmidt number. Indeed, this is the case in our experimental realization (see Table 1.2 of the main text), where higher fidelity bounds are attained with the tilted basis but higher Schmidt number is obtained using MUBs. It is important to point out again, however, that regardless of the choice of target state, the certified fidelity and Schmidt number will always be correct and never over-estimated. In practice this means that a bad choice of basis may lead to a worse noise resistance and it may be harder to certify any entanglement, but when one manages to certify it, this result can be trusted.

#### 1.8.4 Experimental details

Finally, let us discuss the experimental implementation of our entanglement certification method in more detail. As shown in Fig. 1.1 of the main text, our source consists of a single-spatial mode, continuous wave 405nm diode laser (Toptica iBeam Smart 405 HP) with  $\sim 140$ mW of power. The laser is demagnified with a 3:1 telescope system of lenses and focused by a 500mm lens to a spot size of  $330\mu\text{m}$  ( $1/e^2$  beam diameter) at the ppKTP crystal. The 5mm long ppKTP crystal is designed for degenerate Type-II spontaneous parametric downconversion (SPDC) from 405nm to 810nm at  $25^\circ\text{C}$ , and is housed in a custom-built oven for this purpose. The SPDC process generates orthogonally polarized pairs of photons entangled in the Laguerre-Gaussian (LG) basis. The photon pairs are recollimated by a 200mm lens, separated by a polarizing beamsplitter (PBS), and incident on phase-only spatial light modulators (SLMs).

The SLMs (Holoeye PLUTO) have a parallel-aligned LCOS design with a dimension of  $15.36\text{mm} \times 8.64\text{mm}$ , resolution of  $1920 \times 1080$  pixels, reflectivity of approximately 60%, and a diffraction efficiency of 80% at 810nm. The photons are

transformed and reflected by these SLMs (shown in transmission for simplicity) and coupled into single-mode fibers (SMFs) with a coupling efficiency of approx. 50%. The SMFs carry the photons to single-photon avalanche detectors (not shown, Excelitas SPCM-AQRH-14-FC) with a detection efficiency of 60% at 810nm. The detectors are connected to a custom-built coincidence counting logic (CC) with a coincidence-time window of 5ns.

The SLMs and SMFs together act as projective filters for the photon spatial modes. The SLMs are used to display a computer-generated hologram (CGH) that multiplies the incident photon amplitude by an arbitrary amplitude and phase. In this manner, photons in a particular spatial mode (Laguerre-Gaussian or superpositions thereof) are converted to a fundamental Gaussian mode, which then effectively couples to the SMF. The manipulation of both the phase and amplitude of a photon by means of a phase-only device such as an SLM requires the design of a class of phase-only CGHs that allow one to encode arbitrary scalar complex fields. Following the Type 3 method in Ref. [97], our CGH encodes the modulation of a complex field given by  $s(x, y) = A(x, y) \exp[i\phi(x, y)]$  into a phase-only function whose functional form depends explicitly on the amplitude and phase of the field  $s(x, y)$ . This allows arbitrary complex amplitudes to be generated/measured by a phase-only device, albeit at the expense of additional loss. Additionally, we divide the measurement amplitude  $s(x, y)$  by an offset fundamental Gaussian amplitude in order to maximize its overlap with the SMF mode.

A two-photon count rate of approximately 23,000 pairs/sec (Gaussian modes) is measured at the detectors (with blazed gratings displayed on the SLMs), and singles rates of 160,000 and 173,000 counts/sec in the reflected and transmitted PBS arms respectively. The resulting coincidence-to-singles ratios are consistent with the losses described above in each arm. The lossy complex amplitude hologram described above further reduces the two-photon Gaussian-mode count rate to 668 pairs/sec. These holograms have a mode-dependent loss that varies for different incident modes. In Appendix A, we discuss how the coincidence and singles rates allow us to account for this mode-dependent loss. As shown in Fig. 1.3 (a) of the main text, the resultant state measured by these holograms in the standard Laguerre-Gaussian basis is close to  $|\Phi\rangle = \sum_{m=0}^{10} \lambda_m |mm\rangle$ , with 89% counts on the diagonal. The individual  $\lambda_m$  values are:  $\lambda_0 = 0.255$ ,  $\lambda_1 = 0.259$ ,  $\lambda_2 = 0.292$ ,  $\lambda_3 = 0.315$ ,  $\lambda_4 = 0.335$ ,  $\lambda_5 = 0.349$ ,  $\lambda_6 = 0.339$ ,  $\lambda_7 = 0.316$ ,  $\lambda_8 = 0.305$ ,  $\lambda_9 = 0.272$ , and  $\lambda_{10} = 0.260$ . Note that  $m \in \{0, \dots, 10\}$  corresponds to Laguerre-Gaussian modes with an OAM of  $\ell \in \{-5, \dots, 5\}$ . The measured state is correlated in OAM, as the reflection at the PBS flips the sign of one photon from the initially OAM-anti-correlated state.

The probability that one CW pump photon downconverts into a pair of photons in our ppKTP crystal is  $10^{-9}$ . While this is two orders of magnitude higher than  $\beta$ -BBO, it is still quite low. The corresponding probability of two pairs being produced simultaneously is then significantly lower at  $10^{-18}$  and can be neglected. The rate of accidental counts becomes a factor when the singles rates are high and the measurement integration time is long. For example, in the Gaussian (brightest) modes, there are 6675 pairs measured in 10 seconds. The total singles are 230438 and 249617, an accidental rate of  $\approx 2.9/\text{sec}$ . Correcting for accidental coincidences in this manner increases the measured fidelities of our states slightly.





## Part II

# Indefinite causal order



# CHAPTER 2

## Semi-device-independent certification of indefinite causal order

Jessica Bavaresco, Mateus Araújo, Časlav Brukner, Marco Túlio Quintino

**Abstract.** When transforming pairs of independent quantum operations according to the fundamental rules of quantum theory, an intriguing phenomenon emerges: some such higher-order operations may act on the input operations in an indefinite causal order. Recently, the formalism of process matrices has been developed to investigate these noncausal properties of higher-order operations. This formalism predicts, in principle, statistics that ensure indefinite causal order even in a device-independent scenario, where the involved operations are not characterised. Nevertheless, all physical implementations of process matrices proposed so far require full characterisation of the involved operations in order to certify such phenomena. Here we consider a semi-device-independent scenario, which does not require all operations to be characterised. We introduce a framework for certifying noncausal properties of process matrices in this intermediate regime and use it to analyse the quantum switch, a well-known higher-order operation, to show that, although it can only lead to causal statistics in a device-independent scenario, it can exhibit noncausal properties in semi-device-independent scenarios. This proves that the quantum switch generates stronger noncausal correlations than it was previously known.

Published in:

*Quantum* **3**, 176 (2019)  
arXiv:1903.10526 [quant-ph]

## Author Contribution

The doctoral candidate envisioned this project, formulated the main questions that were investigated, contributed to the theoretical results, wrote the majority of the code applied in the work, and wrote the manuscript.

### 2.1 Introduction

A common quantum information task consists in certifying that some uncharacterised source is preparing a system with some features. By making the assumption that the measurement devices are completely characterised, that is, that they are known exactly, it is possible to infer properties of the system. In this *device-dependent* scenario, fidelity of a quantum state with respect to a target state can be estimated, entanglement witnesses can be evaluated [65], and even complete characterisation of the source via state tomography is possible [64].

Remarkably, it is possible to certify properties of systems even without fully characterizing the measurement devices [67, 68]. In such a *device-independent* scenario it is only assumed that the measurements are done by separated parties and compose under a tensor product, which is justified by implementing them with a space-like separation. Under these circumstances, Bell scenarios can be used to certify properties like entanglement of quantum states [68], incompatibility of quantum measurements [113], or to perform device-independent state estimation via self-testing [114, 115].

Since the assumptions are weaker, demonstrations of device-independent certification are usually experimentally challenging. For instance, although experimental device-independent certification of entanglement has been reported [116–119], its experimental difficulty has so far prevented its use in practical applications such as device-independent quantum key distribution [120] and randomness certification [121].

An interesting middle ground is the *semi-device-independent* scenario, where assumptions are made about some parties but not others. Semi-device-independent schemes have been developed and extensively studied for the certification of entanglement [81] and measurement incompatibility [122, 123], known as EPR-steering, and applied to quantum key distribution protocols where some but not all parties can be trusted [124].

A close analogy can be developed with regard to the certification of indefinite causal order, as encoded in a *process matrix* [44]. A process matrix is a higher-order operation [41, 42, 54] – i.e. a transformation of quantum operations – that acts on

independent sets of operations. Fundamental laws of quantum theory predict the existence of process matrices that act on these operations in a such a way that a well-defined causal order cannot be established among them. Process matrices with indefinite causal order were proven to be a powerful resource, outperforming causally ordered ones in tasks such as quantum channel discrimination [57], communication complexity [58, 59], quantum computation [60], and inverting unknown unitary operations [61].

To certify that a process matrix in fact does not act in a causally ordered way, there are two standard methods available in the literature. The first is to evaluate a causal witness [46, 66]. Analogous to the evaluation of an entanglement witness, this method relies on detailed knowledge of the quantum operations being implemented, and as such it allows for a device-dependent certification. All experimental certifications of indefinite causal order to date either measure a causal witness [72, 74] or rely on similar device-dependent assumptions [71, 73, 75]. The second method is the violation of a causal inequality, phenomenon which is also predicted by quantum mechanics [44, 69]. Analogous to the violation of a Bell inequality, this method does not rely on detailed knowledge of the quantum operations implemented by the parties, but rather only that they compose under a tensor product. As such, it allows for a device-independent certification. When a causal inequality is violated, it is verified that least one of the principles used to derived it is not respected. Moreover, this claim holds true independently of the physical theory that supports the experiment that led to the violation. Hence, although we focus on quantum theory, indefinite causal order could in principle be certified even without relying on the laws of quantum mechanics. Although it would be highly desirable to perform such device-independent certification of indefinite causal order, no physical implementation of process matrices that would violate a causal inequality is currently known.

In this work, we introduce a semi-device-independent framework for certifying noncausal properties of process matrices that allows for an experimental certification of indefinite causal order that relies on fewer assumptions than previous ones. In our semi-device-independent scenario, the operations of some parties are fully characterized while no assumptions are made about the others.

We begin by considering the bipartite case, for which we construct a general framework for certifying indefinite causal order in a semi-device-independent scenario, while contextualizing previously developed device-dependent and -independent ones. We then extend our framework to a tripartite case in which the third party is always in the future of the other two, and provide an extensive machinery that may be generalized to other multipartite scenarios. We apply our methods to the notorious

quantum switch [43, 57], a process matrix that, despite having an indefinite causal order, only leads to causal correlations in a device-independent scenario. We show that the noncausal properties of the quantum switch can be certified in a semi-device-independent way, proving that it generates stronger noncausal correlations than it was previously known.

## 2.2 Preliminaries

In our certification scheme, we will deal with statistical data in the form of behaviours.

A *general bipartite behaviour*  $\{p(ab|xy)\}$  is a set of joint probability distributions, that is, a set in which each element  $p(ab|xy)$  is a real non-negative number such that  $\sum_{a,b} p(ab|xy) = 1$  for all  $x, y$ , where  $a \in \{1, \dots, O_A\}$  and  $b \in \{1, \dots, O_B\}$  are labels for outcomes and  $x \in \{1, \dots, I_A\}$  and  $y \in \{1, \dots, I_B\}$  are labels for inputs, for parties Alice and Bob, respectively.

The most general operation allowed by quantum theory is modelled by a quantum instrument, which is a set of completely-positive (CP) maps that sum to a completely-positive trace-preserving (CPTP) map [125].

The Choi-Jamiołkowski (CJ) isomorphism [126–128] allows us to represent every linear map<sup>1</sup>  $\mathcal{M} : \mathcal{L}(\mathcal{H}^I) \rightarrow \mathcal{L}(\mathcal{H}^O)$  by a linear operator  $M \in \mathcal{L}(\mathcal{H}^I \otimes \mathcal{H}^O)$  acting on the joint input and output Hilbert spaces. In this representation, a *set of instruments* is a set of operators  $\{I_{a|x}\}$ ,  $I_{a|x} \in \mathcal{L}(\mathcal{H}^I \otimes \mathcal{H}^O)$ , that satisfies

$$I_{a|x} \geq 0, \quad \forall a, x \tag{2.1}$$

$$\text{Tr}_O \sum_a I_{a|x} = \mathbb{1}^I, \quad \forall x, \tag{2.2}$$

where  $x \in \{1, \dots, I\}$  labels the instrument in the set and  $a \in \{1, \dots, O\}$  its outcomes, and  $\mathbb{1}^I$  is the identity operator on  $\mathcal{H}^I$ .

Now let us consider the most general set of behaviours which respects quantum theory. For that we follow the steps of ref. [44] to analyse behaviours that can be extracted by pairs of independent quantum instruments. Let  $\{A_{a|x}\}$ ,  $A_{a|x} \in \mathcal{L}(\mathcal{H}^{A_I} \otimes \mathcal{H}^{A_O})$  and  $\{B_{b|y}\}$ ,  $B_{b|y} \in \mathcal{L}(\mathcal{H}^{B_I} \otimes \mathcal{H}^{B_O})$  be the Choi operators of Alice's and Bob's local instruments. We then seek a function which assigns probabilities to a pair of instrument elements  $A_{a|x}$  and  $B_{b|y}$ . In order to preserve the structure of quantum mechanics, we assume that this function is linear in both arguments. It follows from the Riesz representation lemma [129] that this general linear function

---

<sup>1</sup>In this paper we only consider finite dimensional complex linear spaces. That is, all linear spaces are isomorphic to  $\mathbb{C}^d$  for some natural number  $d$ .

necessarily has the form of  $p(ab|xy) = \text{Tr} \left[ (A_{a|x}^{A_I A_O} \otimes B_{b|y}^{B_I B_O}) W \right]$  for some linear operator  $W \in \mathcal{L}(\mathcal{H}^{A_I} \otimes \mathcal{H}^{A_O} \otimes \mathcal{H}^{B_I} \otimes \mathcal{H}^{B_O})$ . In order to be consistent with extended quantum scenarios, we also consider the case where Alice and Bob may share a (potentially entangled) auxiliary quantum state  $\rho \in \mathcal{L}(\mathcal{H}^{A_{I'}} \otimes \mathcal{H}^{B_{I'}})$ , and have instruments  $\{A'_{a|x}\}$ ,  $A'_{a|x} \in \mathcal{L}(\mathcal{H}^{A_{I'} A_I A_O})$ ,  $\{B'_{b|y}\}$ ,  $B'_{b|y} \in \mathcal{L}(\mathcal{H}^{B_{I'} B_I B_O})$  which acts on the space of the operator  $W$  and the auxiliary state  $\rho$ . A *process matrix* is then defined as the most general linear operator  $W$  such that

$$p(ab|xy) = \text{Tr} \left[ (A'_{a|x}^{A_{I'} A_I A_O} \otimes B'_{b|y}^{B_{I'} B_I B_O}) (W^{A_I A_O B_I B_O} \otimes \rho^{A_{I'} B_{I'}}) \right] \quad (2.3)$$

represents<sup>2</sup> elements of valid probability distributions for every state  $\rho$  and sets of instruments  $\{A'_{a|x}\}$  and  $\{B'_{b|y}\}$ .

It was shown in ref. [46] that a linear operator  $W$  is a process matrix if and only if it respects

$$W \geq 0 \quad (2.4)$$

$$\text{Tr } W = d_{A_O} d_{B_O} \quad (2.5)$$

$${}_{A_I A_O} W = {}_{A_I A_O B_O} W \quad (2.6)$$

$${}_{B_I B_O} W = {}_{A_O B_I B_O} W \quad (2.7)$$

$$W = {}_{A_O} W + {}_{B_O} W - {}_{A_O B_O} W, \quad (2.8)$$

where  ${}_X W := \text{Tr}_X W \otimes \frac{1^X}{d_X}$  is the trace-and-replace operation and  $d_X = \dim(\mathcal{H}^X)$ .

We then define *process behaviours*  $\{p^Q(ab|xy)\}$  as behaviours which can be obtained by process matrices according to

$$p^Q(ab|xy) = \text{Tr} \left[ (A_{a|x}^{A_I A_O} \otimes B_{b|y}^{B_I B_O}) W^{A_I A_O B_I B_O} \right], \quad (2.9)$$

for all  $a, b, x, y$ .

Now we begin to discuss the causal properties of behaviours and process matrices.

A behaviour is considered causally ordered when it can be established that one party acted before the other because the marginal probability distributions of one party do not depend on the inputs of the other. Formally, we have that a behaviour  $\{p^{A \prec B}(ab|xy)\}$  is causally ordered from Alice to Bob if it satisfies

$$\sum_b p^{A \prec B}(ab|xy) = \sum_b p^{A \prec B}(ab|xy'), \quad (2.10)$$

for all  $a, x, y, y'$ , and equivalently from Bob to Alice.

---

<sup>2</sup>In eq. (2.3), as well as in some other equations, we add superscripts on the operators to indicate the Hilbert spaces in which they act, for sake of clarity.

Behaviours that are within the convex hull of causally ordered behaviours are also considered causal, as they can be interpreted as a classical mixture of causally ordered behaviours. Hence, a *causal behaviour*  $\{p^{\text{causal}}(ab|xy)\}$  is a behaviour that can be expressed as a convex combination of causally ordered behaviours, i.e.,

$$p^{\text{causal}}(ab|xy) := qp^{A \prec B}(ab|xy) + (1 - q)p^{B \prec A}(ab|xy), \quad (2.11)$$

for all  $a, b, x, y$ , where  $0 \leq q \leq 1$  is a real number. Behaviours that do not satisfy eq. (2.11) are called noncausal behaviours.

Following the same reasoning as the one in the definition of a general process matrix, in order to associate causal properties to process matrices we now define causally ordered process matrices as the most general operator that takes pairs of local instruments to causally ordered behaviours, that is,

$$p^{A \prec B}(ab|xy) = \text{Tr} [(A_{a|x} \otimes B_{b|y}) W^{A \prec B}], \quad (2.12)$$

for all  $a, b, x, y$ . This definition is equivalent to the one in refs. [42, 44], which states that a bipartite process matrix  $W^{A \prec B} \in \mathcal{L}(\mathcal{H}^{A_I} \otimes \mathcal{H}^{A_O} \otimes \mathcal{H}^{B_I} \otimes \mathcal{H}^{B_O})$  is causally ordered from Alice to Bob if it satisfies

$$W^{A \prec B} =_{B_O} W^{A \prec B}, \quad (2.13)$$

and equivalently from Bob to Alice.

In line with the definition of a causal behaviour, a *causally separable process matrix*  $W^{\text{sep}} \in \mathcal{L}(\mathcal{H}^{A_I} \otimes \mathcal{H}^{A_O} \otimes \mathcal{H}^{B_I} \otimes \mathcal{H}^{B_O})$  is a process matrix that can be expressed as a convex combination of causally ordered process matrices, i.e.,

$$W^{\text{sep}} := qW^{A \prec B} + (1 - q)W^{B \prec A}, \quad (2.14)$$

where  $0 \leq q \leq 1$  is a real number. Process matrices that do not satisfy eq. (2.14) are called causally nonseparable process matrices.

## 2.3 Certification

Let us consider the following task: we are given a behaviour that describes the statistics of a quantum experiment. We analyse this behaviour in the process matrix formalism, that is, we assume that there exists a process matrix  $W$  and sets of local instruments that give rise to this behaviour according to the rules of quantum theory. Without any information about  $W$  – i.e., without direct assumptions about the process matrix – the goal is to verify whether it is causally nonseparable. Additionally, information about the instruments which were performed may or may not be given.

The assumptions about the instruments can be split in three: device-dependent, -independent, and semi-device-independent. A device-dependent certification scenario is one in which the operations of all parties are fully characterised, i.e., the whole matrix description of the elements of all applied instruments is known. A device-independent certification scenario is the opposite, no knowledge or assumption is made regarding the operations performed by any parties, not even the dimension of the linear spaces used to describe them. Finally, a semi-device-independent certification scenario is one in which at least one party is device-dependent, which is often called trusted, and at least one is device-independent, often called untrusted.

In the following we formalise our notions of certification for bipartite process matrices. We refer to appendix B.5 and to section 2.4 for a discussion of more general scenarios.

**Definition 2.1** (Device-dependent certification). *Given a process behaviour  $\{p^Q(ab|xy)\}$ , that arises from known instruments  $\{\overline{A}_{a|x}\}$  and  $\{\overline{B}_{b|y}\}$  and an unknown bipartite process matrix, one certifies that this process matrix is causally nonseparable in a device-dependent way if, for some  $a, b, x, y$ ,*

$$p^Q(ab|\overline{A}_{a|x}, \overline{B}_{b|y}) \neq \text{Tr} [(\overline{A}_{a|x} \otimes \overline{B}_{b|y}) W^{sep}], \quad (2.15)$$

for all causally separable process matrices  $W^{sep}$ .

**Definition 2.2** (Device-independent certification). *Given a process behaviour  $\{p^Q(ab|xy)\}$ , that arises from unknown instruments and an unknown bipartite process matrix, one certifies that this process matrix is causally nonseparable in a device-independent way if, for some  $a, b, x, y$ ,*

$$p^Q(ab|xy) \neq \text{Tr} [(A_{a|x} \otimes B_{b|y}) W^{sep}] \quad (2.16)$$

for all causally separable process matrices  $W^{sep}$  and all general instruments  $\{A_{a|x}\}$  and  $\{B_{b|y}\}$ .

**Definition 2.3** (Semi-device-independent certification). *Given a process behaviour  $\{p^Q(ab|xy)\}$ , that arises from unknown instruments on Alice's side, known instruments  $\{\overline{B}_{b|y}\}$  on Bob's side, and an unknown bipartite process matrix, one certifies that this process matrix is causally nonseparable in a semi-device-independent way if, for some  $a, b, x, y$ ,*

$$p^Q(ab|x, \overline{B}_{b|y}) \neq \text{Tr} [(A_{a|x} \otimes \overline{B}_{b|y}) W^{sep}] \quad (2.17)$$

for all causally separable process matrices  $W^{sep}$  and all general instruments  $\{A_{a|x}\}$ .

On eqs. (2.15), (2.16), and (2.17) one can identify the quantities that are given – the behaviour and the trusted instruments that belong to the device-dependent parties – and the variables in the certification problem – the unknown instruments that belong to the device-independent parties and any causally separable process matrix. If one can guarantee that, for any sets of instruments for the device-independent parties and any causally separable process matrix, the given behaviour cannot be described by the left-hand side of eqs. (2.15), (2.16), or (2.17), then the fact that the process matrix that generated this behaviour is causally nonseparable is certified. A summary of the given quantities and variables in each certification scenario is provided by table 2.1.

Before proceeding we remark an analogy with the entanglement certification problem in which behaviours are assumed to arise from quantum measurements performed on a quantum state. In the entanglement certification case, device-dependent scenarios are related to entanglement witnesses [65], device-independent scenarios to Bell nonlocality [68], and the semi-device-independent ones to EPR-steering [81].

### 2.3.1 Device-dependent

We start be analysing the device-dependent scenario and definition 2.1. This scenario has been thoroughly studied before using the concept of causal witnesses [46], analogous to entanglement witnesses [65], and now we formulate it under our certification paradigm.

In this scenario, the behaviour and sets of instruments of both parties are given and we aim to check whether the given process behaviour is consistent with performing these exact instruments on a causally separable process matrix. This problem can be solved by semidefinite programming (SDP) with the following formulation:

$$\begin{aligned} \text{given } & \{p^Q(ab|xy)\}, \{\bar{A}_{a|x}\}, \{\bar{B}_{b|y}\} \\ \text{find } & W \\ \text{subject to } & p^Q(ab|xy) = \text{Tr}[(\bar{A}_{a|x} \otimes \bar{B}_{b|y})W] \quad \forall a, b, x, y \\ & W \in \text{SEP}, \end{aligned} \tag{2.18}$$

where SEP denotes the set of causally separable matrices (i.e.  $W$  is constrained to eq. (2.14)) which can be characterized by SDP.

If the problem is infeasible, that is, if there does not exist a process matrix  $W$  that satisfies the constraints of eq. (2.18), then the process matrix that generated  $\{p^Q(ab|xy)\}$  is certainly causally nonseparable. Consequently, there exists a causal witness that can certify it, without the need of performing full tomography of the

**Device-dependent**

Given quantities	Variables
$\{p^Q(ab xy)\}$	$W$
$\{\bar{A}_{a x}\}, \{\bar{B}_{b y}\}$	

**Device-independent**

Given quantities	Variables
$\{p^Q(ab xy)\}$	$d_{A_I}, d_{A_O}, d_{B_I}, d_{B_O}$
	$\{A_{a x}\}, \{B_{b y}\}$
	$W$

**Semi-device-independent**

Given quantities	Variables
$\{p^Q(ab xy)\}$	$d_{A_I}, d_{A_O}$
$\{\bar{B}_{b y}\}$	$\{A_{a x}\}$
	$W$

Table 2.1: Comparison between the given (known) quantities and the variables (unknown quantities) in each scenario with different levels of assumptions about the operations of each party involved in the task of certifying noncausal properties of a process matrix.

process matrix [46]. On the contrary, if the problem is feasible, then the solution provides a causally separable process matrix that could have generated the given behaviour.

We now show that all causally nonseparable process matrices can be certified to be so in a device-dependent way.

**Theorem 2.1.** *All noncasual process matrices can be certified in a device-dependent way for some choice of instruments.*

The proof of the above theorem follows from the fact that one can always consider a scenario where Alice and Bob have access to tomographically complete instruments, which allows for the complete characterization of the process matrix [62, 63].

### 2.3.2 Device-independent

A device-independent approach for the process matrix formalism has been previously studied in terms of causal inequalities [69], analogous to Bell inequalities [68], and now we formulate it under our certification paradigm, exploring definition 2.2.

In a device-independent scenario, besides assuming that the given behaviour is a process behaviour following eq. (2.9), no extra assumptions are made. It is then necessary to check whether the given behaviour is consistent with probability distributions that come from any tensor product of pair of sets of instruments, with fixed number of inputs and outputs, performed on any causally separable process matrix of any dimension.

Before characterizing the problem of whether a certain behaviour can be obtained by a causally separable process matrix, we ask the more fundamental question of whether any behaviour can be obtained by a general process matrix on which instruments are performed locally. That is, whether all general (valid) behaviours are process behaviours.

By definition, every process matrix leads to valid general behaviours. However, it is shown in ref. [130] that in the scenario where all parties have dichotomic inputs and outputs, the deterministic two-way signalling behaviour defined by  $p^{2WS}(ab|xy) := \delta_{a,y}\delta_{b,x}$ , where  $\delta_{i,j} = 1$  if  $i = j$  and  $\delta_{i,j} = 0$  otherwise, cannot be obtained exactly by any process matrix. Here we show that one cannot obtain this two-way signalling behaviour even approximately for finite-dimensional process matrices. The proof can be found in appendix B.1.

**Theorem 2.2.** *All process behaviours are valid behaviours, however, not all valid behaviours are process behaviours.*

*In particular, in the scenario where all parties have dichotomic inputs and outputs, any behaviour  $\{p(ab|xy)\}$  such that*

$$\frac{1}{4} \sum_{a,b,x,y} \delta_{a,y}\delta_{b,x} p(ab|xy) > 1 - \frac{1}{d+1}$$

*is not a process behaviour for process matrices with total dimension  $d_{A_I}d_{A_O}d_{B_I}d_{B_O} = d$ .*

Here we can make a parallel with Bell nonlocality, where the Popescu-Rohrlich behaviour is known to respect the non-signalling conditions which arise naturally in Bell scenarios but cannot be obtained by performing local measurements on entangled states [131–133].

As for causal behaviours, the analogous question is also pertinent. Can all causal behaviours be obtained by pair of sets of instruments and causally separable process

matrices? We answer this question positively, which allows us to relate the properties of a behaviour directly to the properties of the process matrices that could have given rise to it.

**Lemma 2.1.** *A general behaviour is causal if and only if it is a process behaviour that can be obtained by a causally separable process matrix.*

The proof is made by explicitly constructing the instruments and causally separable process matrix that can recover any causal behaviour. It can be found in appendix B.2.

This result allows us to identify which causally nonseparable process matrices can be certified in a device-independent way:

**Theorem 2.3.** *A process matrix can be certified to be causally nonseparable in a device-independent way if and only if it can generate a noncausal behaviour for some choice of instruments for Alice and Bob.*

*Proof.* If a process matrix is causally separable then its behaviours will be causal. If a behaviour is causal, even though it could in principle have been generated by a causally nonseparable process matrix, according to lemma 2.1 it can always be reproduced by a causally separable process matrix. Hence, no causal properties of the process matrix can be inferred.  $\square$

From the formulation of the device-independent certification problem in definition 2.2, it is not clear whether one could obtain a simple characterisation to solve it. In particular, because there are no constraints on the dimension of the linear spaces and there is a product of variables (that represent the unknown instruments and process matrix). Interestingly, we can explore the above theorem to present simple necessary and sufficient conditions for a general behaviour to allow for device-independent certification of indefinite causal order. This follows from the fact that a behaviour can be checked to be noncausal by linear programming [69]. More explicitly, the certification problem can be formulated as follows:

$$\begin{aligned} & \text{given } \{p^Q(ab|xy)\} \\ & \text{find } q_1(\lambda), q_2(\lambda) \\ & \text{s.t. } p^Q(ab|xy) = \sum_{\lambda} \left[ q_1(\lambda) D_{\lambda}^{A \prec B}(ab|xy) + q_2(\lambda) D_{\lambda}^{B \prec A}(ab|xy) \right], \forall a, b, x, y \quad (2.19) \\ & q_1(\lambda) \geq 0, q_2(\lambda) \geq 0, \forall \lambda, \end{aligned}$$

where  $\{D_{\lambda}^{A \prec B}(ab|xy)\}$  and  $\{D_{\lambda}^{B \prec A}(ab|xy)\}$  are the finite set of deterministic causal distributions described in ref. [69].

If the problem is infeasible, then the process matrix that was used to generate the process behaviour  $\{p^Q(ab|xy)\}$  is certainly causally nonseparable and there exists a causal inequality that can witness it [69]. If the problem is feasible, then one can use the results presented in appendix B.2 to explicitly find a causally separable process matrix  $W^{\text{sep}}$  and sets of instruments  $\{A_{a|x}\}$  and  $\{B_{b|y}\}$  such that  $p^Q(ab|xy) = \text{Tr}[(A_{a|x} \otimes B_{b|y}) W^{\text{sep}}]$ .

Differently from the device-dependent scenario, it is known that some causally nonseparable process matrices cannot be certified in a device-independent way [46, 70, 134]. In particular, there exist causally nonseparable bipartite process matrices that, for any choice of instruments of Alice and Bob, will always lead to causal behaviours. This result was first presented in ref. [134] and we rephrase it here:

**Proposition 2.1** (Device-dependent certifiable, device-independent noncertifiable process matrix). *There exist causally nonseparable process matrices that, for any sets of instruments, always give rise to causal behaviours. That is, a causally nonseparable process matrix that cannot be certified in a device-independent way.*

In particular, let  $W \in \mathcal{L}(\mathcal{H}^{A_I A_O B_I B_O})$  be a process matrix and  $W^{T_B}$  be the partial transposition of  $W$  with respect to some basis in  $\mathcal{L}(\mathcal{H}^{B_I B_O})$  for Bob. If  $W^{T_B}$  is causally separable, the behaviour generated by  $p^Q(ab|xy) = \text{Tr}[(A_{a|x} \otimes B_{b|y}) W]$  is causal for every sets of instruments  $\{A_{a|x}\}$  and  $\{B_{b|y}\}$ .

We would like to remark that this phenomenon can be seen as a consequence of the choice of definition of causally separable process matrices. Recalling section 2.2, a causally ordered process matrix is defined as the most general operator  $W^{A \prec B}$  that takes any pairs of sets of instruments to causally ordered behaviours according to  $p^{A \prec B}(ab|xy) = \text{Tr}[(A_{a|x} \otimes B_{b|y}) W^{A \prec B}]$ . On the other hand, the definition of a causally separable process matrix  $W^{\text{sep}}$  as a convex combination of process matrices with definite causal orders, instead of focusing on the behaviours, has an arguably more physical motivation of a classical mixture of causal orders. If the definition were to, alternatively, focus on the behaviours, then a natural choice would be to define a ‘causally separable’ process matrix  $\widetilde{W}^{\text{sep}}$  as the most general operator that takes any pairs of sets of instruments to causal behaviours according to  $p^{\text{causal}}(ab|xy) = \text{Tr}[(A_{a|x} \otimes B_{b|y}) \widetilde{W}^{\text{sep}}]$ . With this alternative, inequivalent definition, ‘causally nonseparable’ process matrices would always lead to noncausal behaviours, for some choice of instruments, by definition. We observe the phenomenon of causally nonseparable process matrices leading exclusively to causal behaviours, presented in proposition 2.1, when we take the physically motivated definition, and it exposes an

intrinsic difference between these two kinds of reasoning. In the next sections, we show how this phenomenon manifests itself in the semi-device-independent scenario.

### 2.3.3 Semi-device-independent

In this final scenario, which has not been explored for process matrices before, we have the information of the behaviour and the instruments of one party, in this case, Bob. According to definition 2.3, one needs to check whether the given behaviour can be reproduced by performing these exact given instruments for Bob, and any set of instruments for Alice with fixed number of inputs and outputs, on any causally separable process matrix that has a fixed dimension on Bob's side. Since both the process matrix and Alice's instruments are variables in eq. (2.17), it is not clear whether this problem can be solved by SDP.

Our approach contrasts a previous one which exploits communication complexity tasks to certify indefinite causal order in process matrices assuming an upper bound for communication capacity between parties and the dimension of their local systems [58, 59].

However, consider the following expression for a process behaviour,

$$p(ab|xy) = \text{Tr} [(A_{a|x} \otimes B_{b|y}) W] \quad (2.20)$$

$$= \text{Tr} [B_{b|y} \text{Tr}_A (A_{a|x} \otimes \mathbb{1}^B W)] \quad (2.21)$$

$$= \text{Tr} (B_{b|y} w_{a|x}^Q), \quad \forall a, b, x, y, \quad (2.22)$$

which motivates us to define  $w_{a|x}^Q$ .

**Definition 2.4** (Process assemblage). *A process assemblage  $\{w_{a|x}^Q\}$  is a set of operators  $w_{a|x}^Q \in \mathcal{L}(\mathcal{H}^{B_I} \otimes \mathcal{H}^{B_O})$  for which there exist a process matrix  $W^{A_I A_O B_I B_O}$  and a set of instruments  $\{A_{a|x}^{A_I A_O}\}$  such that*

$$w_{a|x}^Q = \text{Tr}_{A_I A_O} \left[ (A_{a|x}^{A_I A_O} \otimes \mathbb{1}^{B_I B_O}) W^{A_I A_O B_I B_O} \right], \quad (2.23)$$

for all  $a, x$ .

By defining the process assemblage, we gather all the variables in the certification problem in one object and can start to relate properties of this object to properties of the process matrix. We remark that the process assemblage generalizes the notion of assemblage in EPR-steering [81, 135], which is recovered when both Alice's and Bob's output spaces have  $d_{A_O} = d_{B_O} = 1$ . Consequently,  $\{A_{a|x}\}$  becomes a set of POVMs,  $W$  becomes a bipartite quantum state, and the process assemblage recovers the steering assemblage  $\sigma_{a|x} = \text{Tr}_A (A_{a|x} \otimes \mathbb{1}^B \rho^{AB})$  [135].

Let us first examine the equation below more closely:

$$p(ab|xy) = \text{Tr} (B_{b|y} w_{a|x}) . \quad (2.24)$$

In the same way that a process matrix was defined as the most general operator that takes sets of local instruments to a behaviour, we can define a *general assemblage* to be the most general object that takes a set of instruments to a valid behaviour and respects linearity. In the appendix B.3, we prove that this definition is equivalent to:

**Definition 2.5** (General assemblage). *A general assemblage  $\{w_{a|x}\}$  is a set of operators  $w_{a|x} \in \mathcal{L}(\mathcal{H}^{B_I} \otimes \mathcal{H}^{B_O})$  that satisfies*

$$w_{a|x} \geq 0 \quad \forall a, x \quad (2.25)$$

$$\text{Tr} \sum_a w_{a|x} = d_{B_O} \quad \forall x \quad (2.26)$$

$$\sum_a w_{a|x} =_{B_O} \sum_a w_{a|x} \quad \forall x. \quad (2.27)$$

By defining the general assemblage as the most general set of operators that takes a set of instruments to a behaviour and respects linearity, we are no longer considering its relation with a process matrix or requiring that it is a process assemblage.

If one compares the set of all general assemblages to the set of all process assemblages, it is clear that the set of general assemblages contains the set of process assemblages, since one can see from eq. (2.21) that all process assemblages lead to valid behaviours. But the former set is in principle larger, an outer approximation with a simpler characterisation. We show that, indeed, the set of general assemblages is larger than the set of process assemblages, because just like general behaviours, not all general assemblages can be realised by process matrices.

**Theorem 2.4.** *All process assemblages are valid assemblages, however, not all valid assemblages are process assemblages.*

*In particular, in the scenario where Alice has dichotomic inputs and outputs, the general assemblage  $\{w_{a|x}\}$  given by  $w_{a|x} = |x\rangle\langle x| \otimes |a\rangle\langle a|$  is not a process assemblage.*

The proof presented in appendix B.1 is based on the fact that this assemblage can lead to a deterministic two-way signalling behaviour, which we know not to be attainable by process matrices from theorem 2.2.

Although the process assemblage can be regarded as a generalisation of the steering assemblage that arises in EPR-steering scenarios, here we point out an important difference between semi-device-independent certification of indefinite causal order

and entanglement. A fundamental result on EPR-steering theory is that all steering assemblages admit a quantum realisation by performing POVMs on a quantum state [136–138]. On the other hand, theorem 2.4 shows that some process assemblages do not admit a quantum realisation by performing a set of instruments on a process matrix.

The next step is to assign causal properties to assemblages. A natural approach is to define that an assemblage  $\{w_{a|x}^{Q,A \prec B}\}$  is causally ordered from Alice to Bob if it is a process assemblage that can be obtained from a process matrix that is causally ordered from Alice to Bob, namely,

$$w_{a|x}^{Q,A \prec B} = \text{Tr}_A[(A_{a|x} \otimes \mathbb{1}^B) W^{A \prec B}] \quad \forall a, x, \quad (2.28)$$

for some set of instruments  $\{A_{a|x}\}$  and some causally ordered process matrix  $W^{A \prec B}$ , and equivalently from Bob to Alice.

Since the above definition depends on an unknown set of instruments  $\{A_{a|x}\}$ , it is not easy to check whether a given general assemblage  $\{w_{a|x}\}$  is causally ordered. We therefore derive a simpler characterisation of causally ordered assemblages that is equivalent to the one above.

Analogously to how we defined a general assemblage, we characterise the most general set of operators  $\{w_{a|x}^{A \prec B}\}$  that give rise to a causally ordered behaviour  $\{p^{A \prec B}(ab|xy)\}$  according to the equation

$$p^{A \prec B}(ab|xy) = \text{Tr}(B_{b|y} w_{a|x}^{A \prec B}), \quad (2.29)$$

for any set of instruments  $\{B_{b|y}\}$  for Bob. We then prove its equivalence to the definition below in appendix B.3.

**Definition 2.6** (Causally ordered assemblages). *An assemblage  $\{w_{a|x}^{A \prec B}\}$  is causally ordered from Alice to Bob if it satisfies*

$$w_{a|x}^{A \prec B} =_{BO} w_{a|x}^{A \prec B} \quad \forall a, x, \quad (2.30)$$

while an assemblage  $\{w_{a|x}^{B \prec A}\}$  is causally ordered from Bob to Alice if it satisfies

$$\sum_a w_{a|x}^{B \prec A} = \sum_a w_{a|x'}^{B \prec A} \quad \forall x, x'. \quad (2.31)$$

In appendix B.2, we show that all causally ordered assemblages  $\{w_{a|x}^{A \prec B}\}$  and  $\{w_{a|x}^{B \prec A}\}$  can be realized by some set of instruments  $\{A_{a|x}\}$  and some causally ordered process matrix  $W^{A \prec B}$  and  $W^{B \prec A}$ , respectively. That is, we show that all  $\{w_{a|x}^{A \prec B}\}$  satisfy eq. (2.28), and analogously for the causal order  $B \prec A$ .

We now contrast the statement made in the previous paragraph with general assemblages. As stated before, the technique of defining the general assemblage as the most general set of linear operators that takes instruments to general behaviours results in an object that cannot always be described by process matrices. On the other hand, in the case of causally ordered assemblages, this technique yielded an object that can always be described by (causal) process matrices. The main point to be taken here is that to characterize the most general set of linear operators that takes instruments to some kind of behaviour is a mathematical artifice to find an outer approximation to the set of assemblages that are described by process matrices. The goal is to find an approximation of this set with a potentially simpler characterization. This approximation may be tight, as in the case of causal assemblages, or may not be tight, as in the case of general assemblages. We explore this further in appendix B.5 for assemblages in tripartite scenarios.

We now define a causal assemblage by taking the elements of the convex hull of causally ordered assemblages.

**Definition 2.7** (Causal assemblage). *An assemblage  $\{w_{a|x}^{\text{causal}}\}$  is causal if it can be expressed as a convex combination of causally ordered assemblages, i.e.,*

$$w_{a|x}^{\text{causal}} := q w_{a|x}^{A \prec B} + (1 - q) w_{a|x}^{B \prec A}, \quad (2.32)$$

for all  $a, x$ , where  $0 \leq q \leq 1$  is a real number. An assemblage that does not satisfy eq. (2.32) is called a noncausal assemblage.

We can now express our result in terms of the following lemma, proved in appendix B.2.

**Lemma 2.2.** *A general assemblage is causal if and only if it is a process assemblage that can be obtained from a causally separable process matrix.*

This result allows us to identify which causally nonseparable process matrices can be certified in a semi-device-independent way:

**Theorem 2.5.** *A process matrix is certified to be causally nonseparable in a semi-device-independent way if and only if it can generate a noncausal assemblage for some choice of instruments for Alice.*

*Proof.* If a process matrix is causally separable then its assemblages will be causal. If the assemblage is causal, even though it could in principle have been generated by a causally nonseparable process matrix, according to lemma 2.2 it can always be reproduced by a causally separable process matrix and hence this property cannot be certified.  $\square$

Note that all requirements for an assemblage to be causal are linear and positive semidefinite constraints, hence one can check whether an assemblage is causal via SDP.

However, in our semi-device-independent certification scenario, the only information available is the process behaviour and Bob's instruments, not the assemblage itself. If it were the case that Bob's instruments are tomographically complete, he could obtain full information about the assemblage, and check whether it is causal via SDP. Nevertheless, we show that it is possible to check whether a given behaviour can certify indefinite causal order in a semi-device-independent scenario using SDP even without the knowledge of the assemblage. We do this by rephrasing our certification task in terms of an unknown assemblage.

**Definition 3'** (Semi-device-independent certification, with assemblages). *Given a behaviour  $\{p^Q(ab|xy)\}$ , that arises from unknown instruments on Alice's side, known instruments  $\{\overline{B}_{b|y}\}$  on Bob's side, and an unknown bipartite process matrix, one certifies that this process matrix is causally nonseparable in a semi-device-independent way if, for some  $a, b, x, y$ ,*

$$p^Q(ab|x, \overline{B}_{b|y}) \neq \text{Tr}(\overline{B}_{b|y} w_{a|x}^{causal}) \quad (2.33)$$

for all causal assemblages  $\{w_{a|x}^{causal}\}$ .

Now we are able to formulate the semi-device-independent certification problem in terms of SDP:

$$\begin{aligned} & \text{given } \{p^Q(ab|xy)\}, \{\overline{B}_{b|y}\} \\ & \text{find } \{w_{a|x}\} \\ & \text{s.t. } p^Q(ab|xy) = \text{Tr}(\overline{B}_{b|y} w_{a|x}) \quad \forall a, b, x, y \\ & \quad \{w_{a|x}\} \in \text{CAUSAL}, \end{aligned} \quad (2.34)$$

where CAUSAL denotes the set of causal assemblages, that is,  $\{w_{a|x}\}$  is constrained to eq. (2.32).

As in the previous cases, if the problem is infeasible, then the process matrix that was used to generate the process behaviour  $\{p^Q(ab|xy)\}$  is certainly causally nonseparable. If the problem is feasible, then one can use the results presented in appendix B.2 to explicitly find a causally separable process matrix  $W^{\text{sep}}$  and sets of instruments  $\{A_{a|x}\}$  such that  $w_{a|x} = \text{Tr}_A[(A_{a|x} \otimes \mathbb{1}^B) W^{\text{sep}}]$ .

All three SDP formulations we presented in eqs. (2.18), (2.19) and (2.34) are feasibility problems which can be turned into optimisation problems that allow for a robust certification of indefinite causal order. We discuss this further in section 2.4.

We now show that not all process matrices can be certified in a semi-device-independent way, as some process matrices cannot lead to noncausal assemblages. The proof is in appendix B.4.

**Theorem 2.6** (Device-dependent certifiable, semi-device-independent noncertifiable process matrix). *There exist causally nonseparable process matrices that, for any sets of instruments on Alice’s side, always give rise to causal assemblages. That is, causally nonseparable process matrices that cannot be certified in a semi-device-independent way.*

In particular, let  $W \in \mathcal{L}(\mathcal{H}^{A_I A_O B_I B_O})$  be a process matrix and  $W^{T_A}$  be the partial transposition of  $W$  with respect to some basis in  $\mathcal{L}(\mathcal{H}^{A_I A_O})$  for Alice. If  $W^{T_A}$  is causally separable, the assemblages generated by  $w_{a|x} = \text{Tr}_A[(A_{a|x} \otimes \mathbb{1}^B) W]$  are causal for every set of instruments  $\{A_{a|x}\}$ .

We remark that the above theorem strictly extends proposition 2.1 which was first proved in ref. [134]. That is, with the same hypothesis – that  $W^{T_A}$  is causally separable – we can make a stronger claim – that  $W$  cannot be certified as causally nonseparable even if Bob is treated in a device-dependent way (is trusted).

In ref. [134], the authors show that, when extended with an entangled state, the resulting process matrix can violate a causal inequality and can therefore be certified in a device-independent way. This implies that this extended process matrix can also be certified in a semi-device-independent way. Ref. [134] leaves as an open question the existence of a causally nonseparable bipartite process matrix that cannot be certified in a device-independent way even when extended by entanglement. We remark that this open question is also relevant in the context of semi-device-independent certification.

Another natural question also emerges: is there a bipartite process matrix that can be certified to be causally nonseparable in a semi-device-independent scenario but that cannot be certified to be causally nonseparable in any device-independent scenario? Although we believe such process matrix exists, no example is currently known.

## 2.4 The Quantum Switch

The concepts of certification presented in the previous section have a natural generalisation to different multipartite scenarios. We are now going to illustrate one particular tripartite case by discussing and presenting some results involving the quantum switch [43, 57]. In appendix B.5, we present a detailed extension of the concepts and results from the bipartite case, introduced in section 2.3, to the

tripartite case in which the quantum switch is defined, and pave the way to future more general tripartite and multipartite extensions.

On its first appearance, the *quantum switch* was defined as a higher-order transformation that maps quantum channels into quantum channels and it can be defined as the following. Let  $U_A$  and  $U_B$  be two unitary operators that act on the same space of a target state  $|\psi\rangle^t$ . Let  $|c\rangle^c := \alpha|0\rangle + \beta|1\rangle$ ,  $|\alpha|^2 + |\beta|^2 = 1$ , be a ‘control’ state that is able to coherently control the order in which the operations  $U_A$  and  $U_B$  are applied. The quantum switch acts as following:

$$\text{switch}(U_A, U_B) = |0\rangle\langle 0|^c \otimes U_A U_B + |1\rangle\langle 1|^c \otimes U_B U_A. \quad (2.35)$$

When applied to the state  $|c\rangle^c \otimes |\psi\rangle^t$  we have

$$\text{switch}(U_A, U_B)|c\rangle \otimes |\psi\rangle = \alpha|0\rangle \otimes U_A U_B |\psi\rangle + \beta|1\rangle \otimes U_B U_A |\psi\rangle. \quad (2.36)$$

Physically, the equation above can be understood as the control qubit determining which unitary is going to be applied first on the target state  $|\psi\rangle$ . If the control qubit is in the state  $|0\rangle$  ( $\alpha = 1, \beta = 0$ ), the unitary  $U_B$  is performed before the unitary  $U_A$ . If the control qubit is in the state  $|1\rangle$  ( $\alpha = 0, \beta = 1$ ), the unitary  $U_B$  is performed before the unitary  $U_A$ . In general, if the control qubit is in the state  $|c\rangle = \alpha|0\rangle + \beta|1\rangle$ ,  $\alpha \neq 0, \beta \neq 0$ , the output state will be in a coherent superposition of two different causal orders.

In the process matrix formalism, ref. [54] has analysed the quantum switch as a four-partite process matrix of which the first party has only an output space, which defines the input target and control states, one party inputs  $U_A$ , another party inputs  $U_B$ , and a final party obtains the output control and target states. Here we follow instead the steps of ref. [46] to associate *tripartite process matrices* to the quantum switch. This can be done by absorbing the input target and control states into the process matrix and setting the third party with output space of dimension equal to one. Hence, the quantum switch is described by a family of tripartite process matrices that is shared among three parties, Alice, Bob, and Charlie, for which Charlie is always in the future of Alice and Bob, and the causal order between Alice and Bob may or may not be well defined. A consequence of the fact that Charlie is last and his output space  $\mathcal{H}^{Co}$  has  $d = 1$  is that the most general instrument Charlie can perform is a POVM.

Formally, we define a family of tripartite process matrices associated to the quantum switch according to:

**Definition 2.8** (Quantum switch processes). Let  $|w(\psi, \alpha, \beta)\rangle \in \mathcal{H}^{A_I A_O B_I B_O C_I^t C_I^c}$  be

$$|w(\psi, \alpha, \beta)\rangle = \alpha |\psi\rangle^{A_I} |\Phi^+\rangle^{A_O B_I} |\Phi^+\rangle^{B_O C_I^t} |0\rangle^{C_I^c} + \beta |\psi\rangle^{B_I} |\Phi^+\rangle^{B_O A_I} |\Phi^+\rangle^{A_O C_I^t} |1\rangle^{C_I^c}, \quad (2.37)$$

where  $|\psi\rangle$  is a  $d$ -dimensional pure state,  $\alpha, \beta$  are complex numbers such that  $|\alpha|^2 + |\beta|^2 = 1$ , and  $|\Phi^+\rangle = \sum_{i=1}^d |ii\rangle$  is a maximally entangled unnormalised bipartite qudit state, the Choi representation of the identity channel.

Then, the pure quantum switch processes are a family of tripartite process matrices given by

$$W_{\text{switch}}(\psi, \alpha, \beta) = |w(\psi, \alpha, \beta)\rangle \langle w(\psi, \alpha, \beta)|. \quad (2.38)$$

When the control state is in a nontrivial superposition of  $|0\rangle$  and  $|1\rangle$ , the quantum switch processes have been shown to have some interesting properties [46, 70]. They are causally nonseparable process matrices, meaning they cannot be expressed as a convex combination of tripartite process matrices with definite causal ordered between Alice and Bob, with Charlie in their common future. Namely, when  $\alpha \neq 0$  and  $\beta \neq 0$ ,

$$W_{\text{switch}}(\psi, \alpha, \beta) \neq q W^{A \prec B \prec C} + (1 - q) W^{B \prec A \prec C}, \quad (2.39)$$

for all  $|\psi\rangle$ , and all real numbers  $0 \leq q \leq 1$ . The exact definitions of causally ordered and general tripartite process matrices can be found in appendix B.5.

However, when Charlie is traced out, the resulting bipartite process matrices shared by Alice and Bob are causally separable, namely,

$$\text{Tr}_{C_I^c C_I^t} [W_{\text{switch}}(\psi, \alpha, \beta)] = q W^{A \prec B} + (1 - q) W^{B \prec A}, \quad (2.40)$$

for all  $|\psi\rangle$ ,  $\alpha$ , and  $\beta$ , where  $q = |\alpha|^2$ .

These causally nonseparable tripartite process matrices can be certified in a device-dependent way, since theorem 2.1 also holds for tripartite process matrices. Yet, it has been shown in refs. [46, 70] that the quantum switch processes cannot be certified in a device-independent way, as they always lead to causal behaviours for any choice of instruments of Alice, Bob, and Charlie. It remains to find out whether these processes can be certified in semi-device-independent scenarios.

For this purpose, we extend all concepts and methods from bipartite semi-device-independent certification. Much like in the bipartite case, we make different assumptions about the knowledge of the operations performed by each party. We call *untrusted* (U) a party that is treated in a device-independent way and *trusted* (T) a party that is treated in a device-dependent way, and we use the convention Alice Bob Charlie for denoting the parties. For example, a scenario TTU means

Alice = T (device-dependent), Bob = T (device-dependent), and Charlie = U (device-independent). The four inequivalent semi-device-independent tripartite scenarios are TTU, TUU, UTT, and UUT.

The core idea of the certification task remains the same. For a given process behaviour and given sets of instruments for the trusted parties, one needs to check whether it is possible that this behaviour comes from performing this instruments on a causally separable tripartite process matrix. We also derive the concepts of general, process, causally ordered, and causal assemblages for each scenario. In appendix B.5, we provide all details and calculations, including for the tripartite device-dependent TTT and -independent UUU scenarios.

Our next theorem strengthens the previous result [46] that showed that the quantum switch processes cannot be certified in a full device-independent scenario, i.e., in the UUU scenario. We show that when the instruments of Alice and Bob are unknown, even if the measurements performed by Charlie are known, the quantum switch processes can never be proven to be causally nonseparable, for any pairs of sets of instruments for Alice and Bob. In other words, we prove that the quantum switch processes cannot be certified to be causally nonseparable in the UUT scenario. The previous result of the full device-independent scenario can now be seen as a particular case of the theorem we now present, whose proof is in appendix B.6.

**Theorem 2.7.** *The quantum switch processes cannot be certified to be causally nonseparable on a semi-device-independent scenario where Alice and Bob are untrusted and Charlie is trusted (UUT).*

Moreover, any tripartite process matrix  $W \in \mathcal{L}(\mathcal{H}^{A_I A_O B_I B_O C_I})$ , with Charlie in the future of Alice and Bob, that satisfies

$$\text{Tr}[(A_{a|x}^{A_I A_O} \otimes B_{b|y}^{B_I B_O} \otimes \mathbb{1}^{C_I}) W^{A_I A_O B_I B_O C_I}] = qp^{A \prec B}(ab|xy) + (1-q)p^{B \prec A}(ab|xy), \quad (2.41)$$

for all  $a, b, x, y$ , where  $0 \leq q \leq 1$  is a real number, cannot be certified to be causally nonseparable in a UUT scenario.

We now show that in the three remaining semi-device-independent scenarios, TTU, TUU, and UTT, the quantum switch processes *can* be certified to be causally nonseparable, proving that they can demonstrate stronger noncausal properties than it was previously known.

For our remaining calculations we use the *reduced quantum switch process*

$$W_{\text{red}} := \text{Tr}_{C_I^t} \left[ W_{\text{switch}} \left( |0\rangle, \frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}} \right) \right]. \quad (2.42)$$

By choosing a reduced, mixed switch process in these scenarios, in which the space of the target state is traced out, we guarantee that the pure switch process version can also be certified without performing any measurements on the output target space.

We show that the reduced quantum switch process  $W_{\text{red}}$  can be certified in the TTU, TUU, and UTT scenarios by providing sets of instruments for the device-independent (untrusted) parties that, when applied to  $W_{\text{red}}$ , generate TTU-, TUU-, and UTT-assemblages that are noncausal. We prove these assemblages to be noncausal by means of SDP.

To calculate the assemblages, we choose the following instruments for each untrusted party:

$$\begin{aligned} A_{0|0}^{A_I A_O} &= B_{0|0}^{B_I B_O} = |0\rangle\langle 0| \otimes |0\rangle\langle 0|, & M_{0|0}^{C_I^c} &= |+\rangle\langle +|, \\ A_{1|0}^{A_I A_O} &= B_{1|0}^{B_I B_O} = |1\rangle\langle 1| \otimes |1\rangle\langle 1|, & M_{1|0}^{C_I^c} &= |-\rangle\langle -|, \\ A_{0|1}^{A_I A_O} &= B_{0|1}^{B_I B_O} = |+\rangle\langle +| \otimes |+\rangle\langle +|, \\ A_{1|1}^{A_I A_O} &= B_{1|1}^{B_I B_O} = |-\rangle\langle -| \otimes |-\rangle\langle -|, \end{aligned} \quad (2.43)$$

where  $|\pm\rangle = \frac{1}{\sqrt{2}}(|0\rangle \pm |1\rangle)$ .

Let us illustrate with the TUU case. To construct the TUU-assemblage, we use the instruments from eq. (2.43) for the untrusted parties, Bob and Charlie, according to

$$w_{bc|yz}^{\text{switch}} = \text{Tr}_{BC}[(\mathbb{1}^A \otimes B_{b|y} \otimes M_{c|z}) W_{\text{red}}] \quad \forall b, c, y, z. \quad (2.44)$$

Then, we show via SDP that

$$w_{bc|yz}^{\text{switch}} \neq q w_{bc|yz}^{A \prec B \prec C} + (1 - q) w_{bc|yz}^{B \prec A \prec C}, \quad \forall b, c, y, z, \quad (2.45)$$

proving that the switch process can be certified in this scenario. This is possible due to the SDP characterization of causal assemblages presented in appendix B.5 (for instance, see definition B.20 for the TUU scenario). It follows analogously for the scenarios TTU and UTT.

To be able to compare and quantify the causal properties of the quantum switch across different certification scenarios, from full device-dependent to full device-independent, we introduce a robustness measure. We start by defining the noisy version of the switch, a mixture of the reduced quantum switch process with a ‘trivial’ process matrix (the normalised identity):

$$W_{\text{red}}^\eta := \eta \frac{\mathbb{1}}{d_I} + (1 - \eta) W_{\text{red}}, \quad (2.46)$$

where  $d_I = d_{A_I} d_{B_I} d_{C_I^c}$  is the dimension of the joint input spaces.

We then estimate the minimum value of  $\eta$  for which  $W_{\text{red}}^\eta$  has only causal properties in a given scenario. For example, in the device-dependent scenario, this is the

minimum value of  $\eta$  for which  $W_{\text{red}}^\eta$  is causally separable. In a semi-device-dependent scenario, this is the minimum value of  $\eta$  for which  $W_{\text{red}}^\eta$  only generates causal assemblages. Finally, for the device-independent scenario, this is the minimum value of  $\eta$  for which  $W_{\text{red}}^\eta$  only generates causal behaviours.

It is immediate to see that in the UUU (device-independent) scenario,  $\eta^* = 0$ , since the switch process always generates causal behaviours [46]. Equivalently for the UUT scenario, as a direct consequence of our theorem 2.7. For the TTT scenario, we evaluate via SDP a value of  $\eta^* = 0.6118$  for which  $W_{\text{red}}^{\eta^*}$  is causally separable and below which it is causally nonseparable.

In the remaining scenarios, TTU, TUU, and UTT, in order to calculate the exact value of  $\eta^*$ , one should consider every possible assemblage that could be generated from  $W_{\text{red}}^\eta$ , by optimizing over the set of instruments of the device-independent (untrusted) parties. Since we only consider the fixed instruments of eq. (2.43), we calculate lower bounds for  $\eta^*$ . Additionally, as detailed in appendix B.5, in some of these scenarios our SDP characterization of the set of causal assemblages only constitutes an outer approximation of the set of assemblages that can be described by causal process matrices. Since with SDP we calculate the minimum  $\eta$  for the assemblages to be inside this outer approximation, this again gives only a lower bound for  $\eta^*$ .

Let us illustrate again with the TUU case. We construct a noisy TUU-assemblage using the instruments from eq. (2.43) and  $W_{\text{red}}^\eta$  according to

$$w_{bc|yz}^{\eta,\text{switch}} = \text{Tr}_{BC}[(\mathbb{1}^A \otimes B_{b|y} \otimes M_{c|z}) W_{\text{red}}^\eta] \quad \forall b, c, y, z. \quad (2.47)$$

We then calculate via SDP the minimum value of  $\eta$  for which  $w_{bc|yz}^{\eta,\text{switch}}$  is causal and below which it is noncausal. This value constitutes a lower bound for  $\eta^*$ . In the TUU scenario, we evaluate  $\eta^* \geq 0.1621$  for the reduced switch process. Analogously, we evaluate  $\eta^* \geq 0.1802$  in the UTT scenario. Finally, in the TTU scenario,  $\eta^* \geq 0.5687$ . All these values are summarized in table 2.2. All code used to obtain these results is freely available in an online repository [139]. We remark that the set of instruments required to obtain a robust certification of noncausal separability of the quantum switch is relatively simple and that Charlie can be restricted to perform a single POVM.

The quantum switch has motivated several experiments that explore optical interferometers to certify indefinite causal order of process matrices [71, 72, 74, 75]. Up to now, all experimental results rely on, among other assumptions, complete knowledge of the instruments to certify of causal nonseparability, i.e., they are

TTT	
$\eta^* = 0.6118$	
NONCAUSAL	
<b>UTT</b>	<b>TTU</b>
$\eta^* \geq 0.1802$	$\eta^* \geq 0.5687$
NONCAUSAL	NONCAUSAL
<b>UUT</b>	<b>TUU</b>
$\eta^* = 0$	$\eta^* \geq 0.1621$
CAUSAL	NONCAUSAL
<b>UUU</b>	
$\eta^* = 0$	
CAUSAL	

Table 2.2: The quantum switch can be certified to be causally nonseparable in scenarios TTT, UTT, TTU, and TUU and cannot be certified in scenarios UUT and UUU, where T stands for trusted (device-dependent), U for untrusted (device-independent) and we have chosen the order Alice, Bob, and Charlie (for instance, TTU represents the scenario where Alice and Bob are treated in a device-dependent and Charlie in a device-independent way). The values and bounds for  $\eta^*$  concern the critical value of the mixing parameter  $0 \leq \eta \leq 1$  in eq. (2.46) for which the quantum switch cannot be certified to causally nonseparable on each scenario, and below which, it can be certified. All non-zero values were obtained via SDP. All code is publicly available in an online repository [139].

fully device-dependent. As shown in this section, one can also certify the causal nonseparability of the quantum switch without trusting some of the instruments and measurement apparatuses, in a semi-device-independent way. We have used the machinery developed in this paper to analyse the experiments of refs. [72] and [74] and concluded that, the instruments used in these experiments could allow us to make a stronger claim than what was reported. More precisely, the instruments used to certify that the quantum switch is causally nonseparable on refs. [72] and [74] can lead to a semi-device-independent certification of the noncausal properties of the quantum switch in the TTU scenario. We discuss this results further in appendix B.7.

## 2.5 Conclusions

We developed a framework for certifying indefinite causal order in the process matrix formalism under different sets of assumptions about the operations of the involved parties. In particular, we constructed a semi-device-independent approach

to certification of causally nonseparable process matrices, and unified previously explored device-dependent and -independent approaches. We showed that the sets of causally nonseparable process matrices that can be certified in each scenario are different. More specifically, we proved that some bipartite process matrices can be certified to be causally nonseparable in a device-dependent way but not in a semi-device-independent way, and that some tripartite process matrices can be certified to be causally nonseparable in a semi-device-independent way but not in a device-independent way.

In our framework, we formulated the problem of certifying causally nonseparable process matrices in the device-dependent, semi-device-independent, and device-independent scenarios in terms of semidefinite programming (SDP), implying they can be efficiently solved.

We also showed that some noncausal behaviours and some noncausal assemblages cannot be obtained by process matrices according to the rules of quantum mechanics. For the device-independent case, we presented non-trivial bounds that relate the dimension of a process matrix with its maximal attainable violation of a causal game inequality. Concerning bipartite causal behaviours and causal assemblages, we explicitly showed how to obtain them from causally separable process matrices.

Concerning the quantum switch, we proved that it can produce noncausal correlations, that is, can be certified to be causally nonseparable, in three out of four semi-device-independent scenarios, and proved that its noncausal properties cannot be certified in the remaining one. Finally, we showed that previous experiments that claim to have certified causal nonseparability with the quantum switch under device-dependent assumptions [72, 74] could have, in principle, dropped some assumptions to achieve a stronger form of certification. Our results provide the theoretical basis for a future experimental demonstration of stronger noncausal phenomena that will rely on weaker assumptions than previous ones.

All our code is available in an online repository [139] and can be freely used and edited.





# CHAPTER 3

## Strict hierarchy between parallel, sequential and indefinite-causal-order strategies for channel discrimination

Jessica Bavaresco, Mio Murao, Marco Túlio Quintino

**Abstract.** 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 the strategies is not exclusive to our main example.

Under review. Manuscript submitted on 27 Nov 2020.

arXiv:2011.08300 [quant-ph]

## Author Contribution

In this work, the doctoral candidate contributed to the formulation of the theoretical formalism, to the construction of the main example, wrote the code applied on the initial investigation of the main questions, conducted the typicality analysis, and wrote the main text of the manuscript.

### 3.1 Introduction

The discrimination of quantum operations is one of the most fundamental tasks in quantum information science. It relates to the elementary ability to experimentally distinguish among different quantum dynamics, which comes into play, for example, in tasks associated to certification of quantum circuits.

A plethora of interesting results on this topic has been demonstrated over the course of the years. For the scenario in which the task consists of the discrimination of a pair of channels using only one query, or copy, of an unknown channel, the problem of finding the maximal probability of successful discrimination has been related to the Helstrom measurement [83] and the diamond norm [140, 141]. In striking contrast with the problem of state discrimination – in which any two states can only be perfectly discriminated with a finite number of copies if they are orthogonal – it has been shown that any pair of unitary channels can always be perfectly discriminated for some finite number of copies [142]. Still concerning pairs of unitary channels, it has been shown that there is no advantage of sequential strategies over parallel strategies for discrimination with any finite number of copies [143]. However, for general channels, there can be an advantage of sequential strategies over parallel ones, as demonstrated in Ref. [144] with an example of two qubit-ququart entanglement breaking channels, and in Ref. [145] with an example of two qubit-qubit generalized amplitude damping channels.

In a related task that consists of the discrimination of two ‘no-signalling bipartite channels’, a more general strategy was constructed from the quantum switch [43]. This strategy involved indefinite causal order, and it was shown to not only provide an advantage over causal, i.e. sequential and parallel, strategies, but also to allow for perfect discrimination, which would otherwise not be achievable [57]. This phenomenon already hints that indefinite causal order could be useful for the task of channel discrimination, similarly to how it has proven to be advantageous for other tasks, such as the inversion of unknown unitary operations [61], communication complexity [58, 59], and quantum computation [60].

In this paper, we have two main contributions to the study of channel discrimina-

tion. The first is the demonstration of a new example of an advantage of a sequential strategy over any parallel strategies for the simplest task of channel discrimination – the one between a pair of qubit-qubit channels, an amplitude damping channel and a bit-flip channel. The second is the demonstration that strategies that involve indefinite causal order can outperform parallel and sequential strategies for the same task of channel discrimination. In order to do so, we define two new classes of strategies – which we call separable and general – that may be applied to this problem and make use of indefinite causal order. Together, these results constitute a strict hierarchy between four different strategies of channel discrimination. To demonstrate our results, we develop and apply a general method of computer-assisted proofs.

### 3.2 Minimum-error channel discrimination

The task of minimum-error channel discrimination works as follows: With probability  $p_i$ , Alice is given an unknown quantum channel  $\tilde{C}_i : \mathcal{L}(\mathcal{H}^I) \rightarrow \mathcal{L}(\mathcal{H}^O)$ , 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 the channel  $\tilde{C}_i$ , her task is to determine which channel she received, by performing operations on this channel and guessing the value of  $i \in \{1, \dots, N\}$ . This problem is equivalent to Alice extracting the ‘classical information’  $i$  which is encoded in the channel  $\tilde{C}_i$ . In the simplest case of this task, when Alice is allowed to use one copy of the channel she received, the most general quantum operations that Alice could apply in her laboratory 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 of  $i$  is given by  $p_{\text{succ}} := \sum_{i=1}^N p_i \text{Tr} \left[ (\tilde{C}_i \otimes \tilde{\mathbb{1}})(\rho) M_i \right]$ , where  $\tilde{\mathbb{1}}$  is the identity map on  $\mathcal{L}(\mathcal{H}^{\text{aux}})$ .

Alice can improve her chances by optimizing over the operations she applies on the unknown channel based on her knowledge of the ensemble. Her maximal probability of success in this case is then given by  $p_{\text{succ}}^* := \max_{\{\rho, M\}} p_{\text{succ}}$ , where the optimization occurs over all possible strategies  $\{\rho, M\}$ .

By means of the Choi-Jamiołkowski isomorphism<sup>1</sup> [126–128], we can represent a quantum channel  $\tilde{C}$  (i.e. a completely positive, trace-preserving map) as a positive semidefinite operator  $C \in \mathcal{L}(\mathcal{H}^I \otimes \mathcal{H}^O)$ , called its ‘Choi operator’, that satisfies

---

<sup>1</sup>The Choi-Jamiołkowski isomorphism is a one-to-one correspondence between linear maps  $\tilde{\Lambda} : \mathcal{L}(\mathcal{H}^I) \rightarrow \mathcal{L}(\mathcal{H}^O)$  and linear operators  $L \in \mathcal{L}(\mathcal{H}^I \otimes \mathcal{H}^O)$  defined by  $L := (\tilde{\mathbb{1}} \otimes \tilde{\Lambda})(\Phi^+)$ , 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.

$C \geq 0$  and  $\text{Tr}_O C = \mathbb{1}^I$ , where  $\text{Tr}_O$  denotes the partial trace over  $\mathcal{H}^O$  and  $\mathbb{1}^I$  the identity operator on  $\mathcal{H}^I$ . Using the link product notation<sup>2</sup> [42], we can rewrite the probability of successful discrimination as  $p_{\text{succ}} = \sum_{i=1}^N p_i C_i * \rho * M_i^T$ .

### 3.3 Tester formalism for two copies

In principle, Alice could apply a more general strategy by constructing the most general map that takes a quantum 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  $p(i|C) = \text{Tr}(C T_i)$  for all Choi states 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 [42, 146] (see also App. C.1). Remarkably, it has been shown that every general tester has a *quantum realization* in terms of states and measurements. Namely, for any strategy given by a general tester, a state and measurement that are able to implement it can always be constructed, in such a way that each tester element can be 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 successful discrimination that is equivalently<sup>3</sup> given by  $p_{\text{succ}}^* = \max_{\{T\}} \sum_{i=1}^N p_i \text{Tr}(T_i C_i)$ .

Now let us analyse the more interesting case in which Alice receives 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.

The first and simplest 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 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 a number of linear constraints defined below, and that always accept a quantum realization in terms of states and measurements, according to  $T_i^{\text{PAR}} = \rho * M_i^T$  [42] (see Fig. 3.1(a)). In the following we use the notation  $_X A := \text{Tr}_X A \otimes \frac{\mathbb{1}^X}{d_X}$  and  $d_X = \dim(\mathcal{H}^X)$ .

---

<sup>2</sup>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  $F * G := \text{Tr}_B[(F^{AB} \otimes \mathbb{1}^C)(\mathbb{1}^A \otimes G^{BC} T_B)]$ , where  $.^T_B$  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  $\tilde{C}_i : \mathcal{L}(\mathcal{H}^I) \rightarrow \mathcal{L}(\mathcal{H}^O)$ , then for every state  $\rho \in \mathcal{L}(\mathcal{H}^I)$  we have that  $\tilde{C}_i(\rho) = C_i * \rho = \text{Tr}_I[C_i(\rho^T \otimes \mathbb{1}^O)] \in \mathcal{L}(\mathcal{H}^O)$ .

<sup>3</sup>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.

**Definition 3.1** (Two-copy Parallel Tester). A parallel tester is a set of linear operators  $T^{PAR} = \{T_i^{PAR}\}_{i=1}^N, T_i^{PAR} \in \mathcal{L}(\mathcal{H}^{I_1 O_1 I_2 O_2})$  such that  $T_i^{PAR} \geq 0, \forall i$  and  $W^{PAR} := \sum_i T_i^{PAR}$  satisfies  $\text{Tr}(W^{PAR}) = d_{O_1} d_{O_2}$  and

$$W^{PAR} =_{O_1 O_2} W^{PAR}. \quad (3.1)$$

$W^{PAR}$  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  $\tilde{E} : \mathcal{L}(\mathcal{H}^{O_1} \otimes \mathcal{H}^{\text{aux}_1}) \rightarrow \mathcal{L}(\mathcal{H}^{I_2} \otimes \mathcal{H}^{\text{aux}_2})$ , 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  $T^{SEQ} = \{T_i^{SEQ}\}$  which can be expressed as  $T_i^{SEQ} = \rho * E * M_i^T$ , 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 circuit [42] (see Fig. 3.1(b)) – has been characterized as the following:

**Definition 3.2** (Two-copy Sequential Tester). A sequential tester is a set of linear operators  $T^{SEQ} = \{T_i^{SEQ}\}_{i=1}^N, T_i^{SEQ} \in \mathcal{L}(\mathcal{H}^{I_1 O_1 I_2 O_2})$  such that  $T_i^{SEQ} \geq 0, \forall i$  and  $W^{SEQ} := \sum_i T_i^{SEQ}$  satisfies  $\text{Tr}(W^{SEQ}) = d_{O_1} d_{O_2}$  and

$$W^{SEQ} =_{O_2} W^{SEQ} \quad (3.2)$$

$${}_{I_2 O_2} W^{SEQ} =_{O_1 I_2 O_2} W^{SEQ}. \quad (3.3)$$

$W^{SEQ}$  is called a sequential process.

Parallel and sequential strategies have long been regarded as the most general strategies for channel discrimination. We now propose a more general strategy for channel discrimination than the sequential one, that arises from the following reasoning: In the same fashion of the characterization of the general one-copy tester, we may characterize a *general two-copy tester* as the most general set of operators  $T^{GEN} = \{T_i^{GEN}\}$  that map a pair of quantum 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  $p(i|C_A, C_B) = \text{Tr}[(C_A \otimes C_B)T_i^{GEN}]$ . It is shown in the App. C.1 that this definition is equivalent to:

**Definition 3.3** (Two-copy General Tester). A general tester is a set of linear operators  $T^{GEN} = \{T_i^{GEN}\}_{i=1}^N, T_i^{GEN} \in \mathcal{L}(\mathcal{H}^{I_1 O_1 I_2 O_2})$  such that  $T_i^{GEN} \geq 0, \forall i$  and

$W^{GEN} := \sum_i T_i^{GEN}$  satisfies  $\text{Tr}(W^{GEN}) = d_{O_1}d_{O_2}$  and

$${}_{I_1 O_1} W^{GEN} = {}_{I_1 O_1 O_2} W^{GEN} \quad (3.4)$$

$${}_{I_2 O_2} W^{GEN} = {}_{O_1 I_2 O_2} W^{GEN} \quad (3.5)$$

$$W^{GEN} = {}_{O_1} W^{GEN} + {}_{O_2} W^{GEN} - {}_{O_1 O_2} W^{GEN}. \quad (3.6)$$

$W^{GEN}$  is called a general process.

Both parallel and sequential processes are particular cases of general processes (see Fig. 3.2(b)). Nevertheless, the formalism of process matrices has shown that there are general processes that do not respect a definite causal order [44, 46] – which is defined as the ability of a process to be described as a parallel, ordered, or as a classical mixture of ordered processes, called ‘causally separable’ process matrices, motivating the definition of our final class of testers:

**Definition 3.4** (Two-copy Separable Tester). A separable tester is a set of linear operators  $T^{SEP} = \{T_i^{SEP}\}_{i=1}^N$ ,  $T_i^{SEP} \in \mathcal{L}(\mathcal{H}^{I_1 O_1 I_2 O_2})$  such that  $T_i^{SEP} \geq 0$ ,  $\forall i$  and  $W^{SEP} := \sum_i T_i^{SEP}$  satisfies  $\text{Tr}(W^{SEP}) = d_{O_1}d_{O_2}$  and

$$W^{SEP} = q W^{1 \prec 2} + (1 - q) W^{2 \prec 1}, \quad (3.7)$$

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.

Notice that our characterizations are equivalent to imposing that  $W^{SEQ}$ ,  $W^{GEN}$ , and  $W^{SEP}$  are ordered, general, and causally separable process matrices, respectively [44, 46].

The set of separable processes is then the convex hull of the set of sequential processes whose slots follow the order  $1 \prec 2$  and  $2 \prec 1$ . Parallel processes are the ones at the intersection of these two sets, satisfying both conditions of the ordering of the slots. Finally, separable processes are also a particular case of general processes [44, 46] (see Fig. 3.2(b)).

The definition of separable processes was conceived from the idea that one could plug two different channels  $C_A$  and  $C_B$  in the two slots of process  $W^{SEP}$ , which would then represent a mixture of a process that applies channel  $C_A$  before channel  $C_B$  with one that applies channel  $C_B$  before  $C_A$ . One could then expect that this classical mixture of causal orders should not be relevant for the problem in which the two channels being plugged into the separable tester are identical: two copies of  $C_i$ . Nonetheless, we show that separable testers indeed provide advantage over sequential testers in the task of channel discrimination, which hints at a more complicated

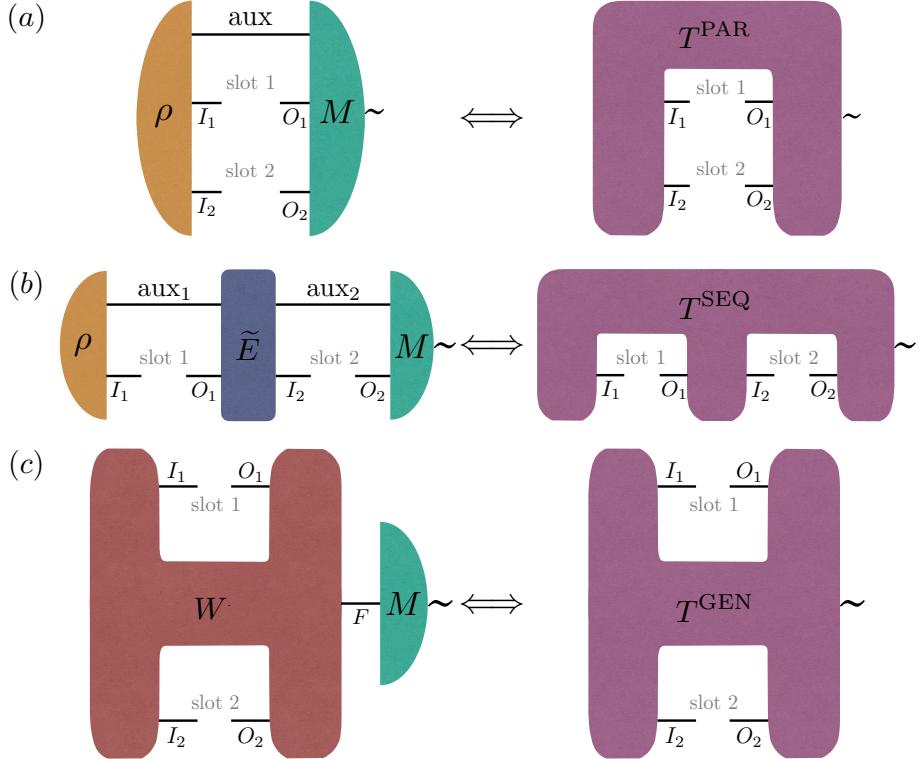


Figure 3.1: Schematic representation of the realization of every two-copy (a) parallel tester  $T^{\text{PAR}}$  with a state  $\rho$  and a POVM  $M$ , (b) sequential tester  $T^{\text{SEQ}}$  with a state  $\rho$ , a channel  $\tilde{E}$ , and a POVM  $M$ , and (c) general tester  $T^{\text{GEN}}$  with a process matrix  $W$  and a POVM  $M$ .

structure of separable *testers* than of separable *processes* themselves. Notice that, if Alice had access to two sequential testers  $T^{1 \leftarrow 2}$  and  $T^{2 \leftarrow 1}$ , and in each round of her experiment she would throw a coin to decide in which tester to plug her two copies of  $C_i$ , the tester that she would be implementing would in fact be such that  $T_i = q T_i^{1 \leftarrow 2} + (1 - q) T_i^{2 \leftarrow 1}$ . This subclass of separable testers, in which the convex combination condition is satisfied by each tester element individually, can never achieve a better performance than a sequential tester in a channel discrimination task. The fact that we indeed encounter advantages by using separable testers shows that they cannot be simply realized by ordered circuits and classical randomness, and implies that the set of separable testers is strictly larger than the convex hull of the set of sequential testers that are ordered in different directions (see Fig. 3.2(a)).

### 3.4 Discrimination of general channels

With our constructed unified framework for channel discrimination at hand, we can now define the maximal probability of successful discrimination under each of

the four described 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  $\mathcal{S} \in \{\text{PAR}, \text{SEQ}, \text{SEP}, \text{GEN}\}$  then reads

$$P^{\mathcal{S}} := \max_{\{T_i^{\mathcal{S}}\}} \sum_{i=1}^N p_i \text{Tr} (T_i^{\mathcal{S}} C_i^{\otimes 2}). \quad (3.8)$$

It is clear that these four strategies – parallel, sequential, separable, and general – form a hierarchy since the set of testers that they define is a superset of the previous one, in this exact order, 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 strictly satisfied by explicitly calculating all  $P^{\mathcal{S}}$  for a specific ensemble.

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

$$\begin{aligned} & \text{given } \{p_i, C_i\} \\ & \text{maximize } \sum_i p_i \text{Tr} (T_i^{\mathcal{S}} C_i^{\otimes 2}) \\ & \text{subject to } \{T_i^{\mathcal{S}}\} \text{ is a tester with strategy } \mathcal{S}. \end{aligned} \quad (3.9)$$

This problem can be equivalently solved by its dual problem

$$\begin{aligned} & \text{given } \{p_i, C_i\} \\ & \text{minimize } \lambda \\ & \text{subject to } p_i C_i^{\otimes 2} \leq \lambda \bar{W}^{\mathcal{S}} \quad \forall i, \end{aligned} \quad (3.10)$$

where  $\bar{W}^{\mathcal{S}}$  lies in the dual affine<sup>4</sup> of the set of processes  $\mathcal{W}^{\mathcal{S}}$ , as demonstrated in App. C.2.

SDPs can be solved by efficient numerical packages which, despite being in practice accurate, suffer from imprecisions that come from the use of floating-point variables. In order to overcome this issue, we provide in the App. C.3 an algorithm for computer-assisted proofs (see [150, 151] for other examples). Using our computer-assisted proof method, which does not make use of floating-point variables, we obtain exact upper and lower bounds for  $P^{\mathcal{S}}$ , arriving at a result that has the same mathematical rigour as an analytical proof.

---

<sup>4</sup>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  $p_i C_i^{\otimes 2} \leq H$  with  $H \leq \bar{W}^{1 \leftarrow 2}$  and  $H \leq \bar{W}^{2 \leftarrow 1}$  where  $\bar{W}^{i \leftarrow j}$  lies in the dual affine set of sequential processes  $\mathcal{W}^{i \leftarrow j}$ . See App. C.2 for details. Notice that our formulation of the dual problem is analogous to the one presented in Ref. [147, 148] and may be seen as a generalization of the Yuen-Kennedy-Lax bound for state discrimination [149].

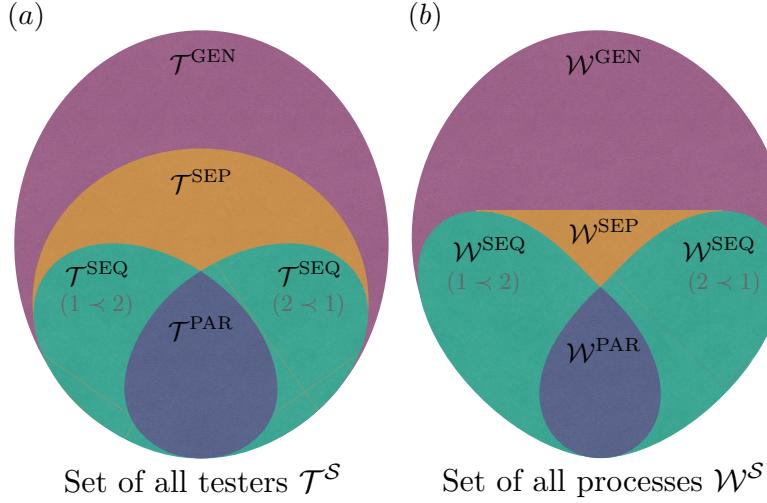


Figure 3.2: Graphical representation of the nesting relations between (a) the sets of all testers  $\mathcal{T}^{\mathcal{S}} := \{T^{\mathcal{S}}; T^{\mathcal{S}} = \{T_i^{\mathcal{S}}\}\}$  and (b) the sets of all processes  $\mathcal{W}^{\mathcal{S}} := \{W^{\mathcal{S}}; W^{\mathcal{S}} = \sum_i T_i^{\mathcal{S}}\}$ , where  $\mathcal{S} \in \{\text{PAR}, \text{SEQ}, \text{SEP}, \text{GEN}\}$  represents parallel, sequential, separable, or general strategies.

**Theorem 3.1.** *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*

$$P^{\text{PAR}} < P^{\text{SEQ}} < P^{\text{SEP}} < P^{\text{GEN}}. \quad (3.11)$$

*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<sup>5</sup>  $\tilde{C}_{\text{AD}}$  with damping parameter  $\gamma = 67/100$ , and a bit-flip channel<sup>6</sup>  $\tilde{C}_{\text{BF}}$  with flipping parameter  $\eta = 87/100$ . We start by applying standard numerical packages to solve the primal SDP (3.9) and obtain an ansatz for the optimal tester of each discrimination strategy. From the numerically imperfect ansatz, we construct a valid tester, following the steps of Algorithm 2 in App. C.3. We then calculate the probability of successful discrimination with this valid tester, which provides a rigorous lower bound for the maximal probability of success. To calculate a rigorous upper bound, we repeat this procedure, now taking as ansatz the numerical solution of the dual SDP (3.10) for a dual affine process, and following the steps of Algorithm 1 in the App. C.3. Applying this method, we calculated the following bounds:  $\frac{8346}{10000} < P^{\text{PAR}} < \frac{8347}{10000}$ ,  $\frac{8446}{10000} < P^{\text{SEQ}} < \frac{8447}{10000}$ ,  $\frac{8486}{10000} < P^{\text{SEP}} < \frac{8487}{10000}$ , and  $\frac{8514}{10000} < P^{\text{GEN}} < \frac{8515}{10000}$ . The clear gap

<sup>5</sup>The action of an amplitude damping channel on a qubit state is given by  $\tilde{C}_{\text{AD}}(\rho) = K_0\rho K_0^\dagger + K_1\rho K_1^\dagger$ , where  $K_0 = |0\rangle\langle 0| + \sqrt{1-\gamma}|1\rangle\langle 1|$  and  $K_1 = \sqrt{\gamma}|0\rangle\langle 1|$ .

<sup>6</sup>The action of a bit-flip channel on a qubit state is given by  $\tilde{C}_{\text{BF}}(\rho) = \eta\rho + (1-\eta)X\rho X$ , where  $X$  is the Pauli  $X$  operator.

between the upper bound of one strategy and the lower bound of the next concludes the proof.  $\square$

Similar gaps can also be found for different ensembles of amplitude damping and bit-flip channels, and also for ensembles of two amplitude damping channels, a problem which has already been previously studied [152–155] and for which we provide more details in App. C.4. Moreover, this phenomenon is not particular to these channels. In App. C.4, we present a simple method of sampling pairs of quantum channels that present a gap between all four strategies, for the case of qubit-qubit channels, in approximately 94% of the rounds.

### 3.5 Quantum realization of testers

Having demonstrated the theoretical advantage of these strategies, we would now like to discuss their potential implementation. As already mentioned, for the case of parallel and sequential strategies, it is known that from every tester that can be obtained, for example, as the optimal solution of an SDP, one can always construct in an algorithmic manner a state, a channel, and a POVM that constitute a quantum realization for each tester element [42]. Therefore, they can be used to physically implement these testers by means of a quantum circuit, as depicted on Fig. 3.1(a) and (b).

We now address the class of separable testers. Given a separable tester  $\{T_i^{\text{SEP}}\}$ , one can always construct a process matrix  $W^F := \sum_{i=1}^N T_i^{\text{SEP}} \otimes |i\rangle\langle i|^F \in \mathcal{L}(\mathcal{H}^{I_1 O_1 I_2 O_2} \otimes \mathcal{H}^F)$  and a POVM  $M = \{M_i\}_i$ ,  $M_i := |i\rangle\langle i| \in \mathcal{H}^F$  such that each tester element can be recovered as  $T_i^{\text{SEP}} = W * M_i^T$ . Notice that the process matrix  $W^F$  satisfies the condition that  $\text{Tr}_F W^F = \sum_{i=1}^N T_i^{\text{SEP}} = W^{\text{SEP}}$ . Remarkably, process matrices that satisfy this condition have recently been shown in Ref. [47] to be realizable with circuits that employ a coherent quantum control of causal orders, implying that all separable strategies of channel discrimination, including the ones that we have shown to be advantageous over causal strategies, can be physically implemented. This class of process matrices was also shown not to be device-independently certifiable [46] nor semi-device-independently certifiable when only the future space is characterized [156]. One notable example of this class is the quantum switch [43]. Even though we have shown that separable strategies may outperform causal strategies of channel discrimination, we have not been able to construct an example of an advantageous strategy that employs the quantum switch specifically.

Finally, for the case of general testers, a ‘realization’ analogous to the case of separable testers but that involves a general process matrix and a POVM is always

possible, as depicted on Fig. 3.1(c). A *quantum* realization of general testers would then depend on the ability to physically implement any process matrix. Unfortunately, at this point, the physical implementation of general process matrices remains an open question. It is our hope that our demonstration of the theoretical advantage of general strategies for channel discrimination will further motive the investigation of potential implementation of general processes.

### 3.6 Conclusions

We have demonstrated a new example of the advantage of a sequential strategy over any parallel strategy for a task of minimum-error discrimination, one 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 discrimination strategies. Our main example concerns the discrimination of an amplitude-damping channel 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, exhibit a strict hierarchy between all four strategies of discrimination. The main technique used in this paper was a method of computer assisted proofs, that finds immediate application in a plethora of quantum information problems that involve semidefinite programming. We hope that this method can contribute to paving the way to more rigorous numerical proofs in quantum information science.

All our code is available in an online repository [157] and can be freely used and edited.





# CHAPTER 4

## Unitary operation discrimination beyond group structures: Advantages of adaptive and indefinite-causal-order strategies

Jessica Bavaresco, Mio Murao, Marco Túlio Quintino

**Abstract.** For minimum-error channel discrimination tasks that involve more than two unitary channels, we show that sequential (i.e. adaptive) strategies may outperform parallel ones. Additionally, we show that general strategies that involve indefinite causal order may outperform sequential ones. For the case in which the set of unitaries being discriminated form a group, we show that parallel strategies are indeed optimal, even when considering general strategies. Finally, we show that strategies based on the quantum switch cannot outperform sequential strategies for the discrimination of unitary channels.

Manuscript in preparation.

## Author Contribution

The role of the doctoral candidate in this work was to contribute to the theoretical results, to write the code applied on the initial investigation of the main questions, to conduct the typicality analysis, and to write the main text of the manuscript.

### 4.1 Introduction

The ability to discriminate between different quantum objects plays a fundamental role in quantum information processing [83], quantum hypothesis testing [158], and quantum parameter estimation [159]. A particular instance of main relevance is the discrimination between unitary quantum operations, problem which physically corresponds to discriminating between different closed quantum dynamics. From a quantum computational perspective, algorithms based on quantum oracles such as Deustch-Josza [160], Grover's [161], and Simon's [162] algorithm may be recast as a discrimination task between quantum oracles, which are unitary operations [163].

The discrimination of unitary channels is a topic that has been extensively studied: Contrarily to the problem of quantum state discrimination, in which two states cannot be perfectly distinguished with a finite number of copies unless their are orthogonal, in Refs. [142, 164] it was shown that any pair of unitary channels can indeed be always perfectly distinguished with a finite number of copies. These references have also shown that the maximal probability of successful discrimination of a pair of unitary channels in a parallel scheme can always be achieved without the need for an auxiliary system, which implies without an entangled input state in the single-copy case, and with a  $k$ -partite entangled state in the  $k$ -copy case. Such  $k$ -copy parallel strategies were also shown to achieve perfect discrimination for a pair of unitary channels [142, 164]. Subsequently, a sequential scheme that can also achieve perfect discrimination of a pair of unitary channels using a finite number of copies, without the need for an auxiliary system – and hence without entanglement – was provided in Ref. [165]. Furthermore, Ref. [143] demonstrated the fact that a sequential scheme can never outperform a parallel scheme in a task of discrimination between two unitary channels using a finite number of copies. Concerning the discrimination of sets of more than two unitary channels, several necessary and sufficient conditions for perfect discrimination were developed in Ref. [166]. When considering unitaries which are a representation of a group and uniformly sampled, Ref. [143] showed that sequential strategies can never outperform parallel strategies, for any finite number of copies.

Motivated by the recent advances in channel discrimination theory [45] that have

established the advantage of discrimination strategies that involve indefinite causal order for general channels, we now turn our attention to the particular case of unitary channels.

Extending the framework developed in Ref. [45] to discrimination tasks that allow for the use of multiple copies of an unknown channel, we achieve the following main results. The first is the proof of the optimality of parallel strategies, even when compared against general strategies, in tasks of discrimination of an ensemble with a uniform probability distribution over a set of unitary channels that form a group. However, the optimality of parallel strategies ends there. For ensembles of unitary channels that either do not form a group or that are distributed according to a non-uniform probability distribution, we show that sequential (a.k.a. adaptive) strategies can be advantageous over parallel strategies. Moreover, we show that general strategies that apply indefinite causal order can outperform sequential strategies.

Finally, we show that for a particular case of general strategies, one that applies processes related to the quantum switch [43], cannot outperform sequential strategies in the discrimination of unitary channels.

## 4.2 Minimum-error channel discrimination

In a task of minimum-error channel discrimination, one is given access to an unknown quantum channel  $\tilde{C}_i : \mathcal{L}(\mathcal{H}^I) \rightarrow \mathcal{L}(\mathcal{H}^O)$ , that was 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. 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  $\rho \in \mathcal{L}(\mathcal{H}^I \otimes \mathcal{H}^{\text{aux}})$  through the channel  $\tilde{C}_i$  and subsequently jointly measure the output state with positive-operator valued measure (POVM)  $M = \{M_a\}_{a=1}^N, M_a \in \mathcal{L}(\mathcal{H}^O \otimes \mathcal{H}^{\text{aux}})$ . When both 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  $P := \max_{\rho, \{M_i\}} \sum_{i=1}^N p_i \text{Tr} \left[ (\tilde{C}_i \otimes \tilde{\mathbb{1}})(\rho) M_i \right]$ , where  $\tilde{\mathbb{1}}$  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. 4.1(a) and (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. It has been shown that some 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 [45].

### 4.3 Tester formalism for $k$ copies

To facilitate the approach to this problem, a concise formalism of testers [42], also referred to as process POVMs [146, 167], was developed in Ref. [45], 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 a given strategy.

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 that, when taken the trace with the Choi state of a quantum channel<sup>1</sup>, lead to a valid probability distribution. In this sense, testers are to channels what POVMs are to states. The simplest example is the general single-copy tester  $T = \{T_i\}_{i=1}^N$ , which satisfies  $T_i \geq 0 \forall i$ ,  $\sum_i T_i = \sigma^I \otimes \mathbb{1}^O$ , where  $\text{Tr}(\sigma^I) = 1$ . These constraints guarantee that every tester  $T$  can be realized by a quantum state  $\rho$  and a POVM  $M = \{M_i\}_i$  according to  $T_i = \text{Tr}_{\text{aux}}[(\rho^{I,\text{aux}} \otimes \mathbb{1}^O)(\mathbb{1}^I \otimes M_i^{\text{aux},O})]$ , and, conversely, that every state and measurement can construct a valid tester [42, 146]. Thus, the maximal probability of success can be expressed equivalently in terms of testers and Choi states of channels, according to  $P = \max_{\{T_i\}} \sum_{i=1}^N p_i \text{Tr}(T_i C_i)$ . 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, while guaranteeing that the optimal tester can be implemented by quantum states and measurements.

For the case of  $k$  copies, different normalization constraints define testers that represent different classes of strategies. Parallel strategies are represented by parallel testers  $T^{\text{PAR}} = \{T_i^{\text{PAR}}\}_{i=1}^N, T_i^{\text{PAR}} \in \mathcal{L}(\mathcal{H}^I \otimes \mathcal{H}^O)$ , where  $\mathcal{H}^I := \bigotimes_{i=1}^k \mathcal{H}^{I_i}$  and  $\mathcal{H}^O :=$

---

<sup>1</sup>In order to apply the tester formalism, we will make use of the Choi-Jamiolkowski (CJ) representation of quantum maps. The CJ isomorphism is a one-to-one correspondence between completely-positive maps and positive semidefinite operators, that 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  $L := (\tilde{\mathbb{1}} \otimes \tilde{L})(\Phi^+)$ , where  $\tilde{\mathbb{1}} : \mathcal{L}(\mathcal{H}^I) \mapsto \mathcal{L}(\mathcal{H}^I)$  is the identity map and  $\Phi^+ = \sum_{ij} |ii\rangle\langle jj| \in \mathcal{L}(\mathcal{H}^I \otimes \mathcal{H}^I)$ , where  $\{|i\rangle\}$  is an orthonormal basis, is an unnormalized maximally entangled state. In this representation, a quantum channel, i.e., a 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 state’ of channel  $\tilde{C}$ , that satisfies  $C \geq 0$  and  $\text{Tr}_O C = \mathbb{1}^I$ , where  $\text{Tr}_O$  denotes the partial trace over  $\mathcal{H}^O$  and  $\mathbb{1}^I$  the identity operator on  $\mathcal{H}^I$ . In particular, the Choi state of a unitary channel is proportional to a maximally entangled state.

$\bigotimes_{i=1}^k \mathcal{H}^{O_i}$ . Parallel testers, similarly to the single-copy case, can always be expressed in terms of states and measurements. Defining  $W^{\text{PAR}} := \sum_i T_i^{\text{PAR}}$ , parallel testers are the ones that satisfy

$$T_i^{\text{PAR}} \geq 0 \quad \forall i \quad (4.1)$$

$$\text{Tr } W^{\text{PAR}} = d_O \quad (4.2)$$

$$W^{\text{PAR}} =_O W^{\text{PAR}}, \quad (4.3)$$

where  $d_O = \dim(\mathcal{H}^O) = d_{O_1} \dots d_{O_k}$  is the dimension of the collective output space and  $_X(\cdot) = \text{Tr}_X(\cdot) \otimes \mathbb{1}^X / d_X$  denotes a trace-and-replace operation.

Sequential strategies are represented by sequential testers [42],  $T^{\text{SEQ}} = \{T_i^{\text{SEQ}}\}_{i=1}^N$ ,  $T_i^{\text{SEQ}} \in \mathcal{L}(\mathcal{H}^I \otimes \mathcal{H}^O)$ . Sequential testers are testers which can always be constructed from an input state, a sequence of CPTP maps, and a final POVM. Defining  $W^{\text{SEQ}} := \sum_i T_i^{\text{SEQ}}$ , sequential testers are the ones that satisfy

$$T_i^{\text{SEQ}} \geq 0 \quad \forall i \quad (4.4)$$

$$\text{Tr } W^{\text{SEQ}} = d_O \quad (4.5)$$

$$W^{\text{SEQ}} =_{O_k} W^{\text{SEQ}} \quad (4.6)$$

$${}_{I_k O_k} W^{\text{SEQ}} =_{O_{(k-1)} I_k O_k} W^{\text{SEQ}} \quad (4.7)$$

...

$${}_{I_2 O_2 \dots I_k O_k} W^{\text{SEQ}} =_{O_1 I_2 O_2 \dots I_k O_k} W^{\text{SEQ}}. \quad (4.8)$$

This is equivalent to defining  $W^{\text{SEQ}}$  to be a  $k$ -slot comb [42], or a  $k$ -partite ordered process matrix [46] (see also Refs. [168, 169]).

Finally, general strategies are represented by general testers  $T^{\text{GEN}} = \{T_i^{\text{GEN}}\}_{i=1}^N$ ,  $T_i^{\text{GEN}} \in \mathcal{L}(\mathcal{H}^I \otimes \mathcal{H}^O)$ . General testers are defined as the most general set of positive semidefinite operators  $\{T_i^{\text{GEN}}\}$  that takes  $k$  different channels to an element of a probability distribution, according to  $p(i|C_1 \otimes \dots \otimes C_k) = \text{Tr}[T_i^{\text{GEN}}(C_1 \otimes \dots \otimes C_k)]$ . Defining  $W^{\text{GEN}} := \sum_i T_i^{\text{GEN}}$ , general testers must satisfy

$$T_i^{\text{GEN}} \geq 0 \quad \forall i \quad (4.9)$$

$$\text{Tr}[W^{\text{GEN}}(C_1 \otimes \dots \otimes C_k)] = 1, \quad (4.10)$$

for all Choi states of quantum channels  $C_i \in \mathcal{L}(\mathcal{H}^{I_i} \otimes \mathcal{H}^{O_i})$ . This is equivalent to defining  $W^{\text{GEN}}$  to be a general  $k$ -partite process matrix [44, 46]. We refer to Ref. [45] for a detailed derivation of the two-copy general tester and to Ref. [46] for a derivation of multipartite process matrices. Ordered processes, such as  $W^{\text{PAR}}$  and  $W^{\text{SEQ}}$ , form a subset of general processes. However, some general processes exhibit

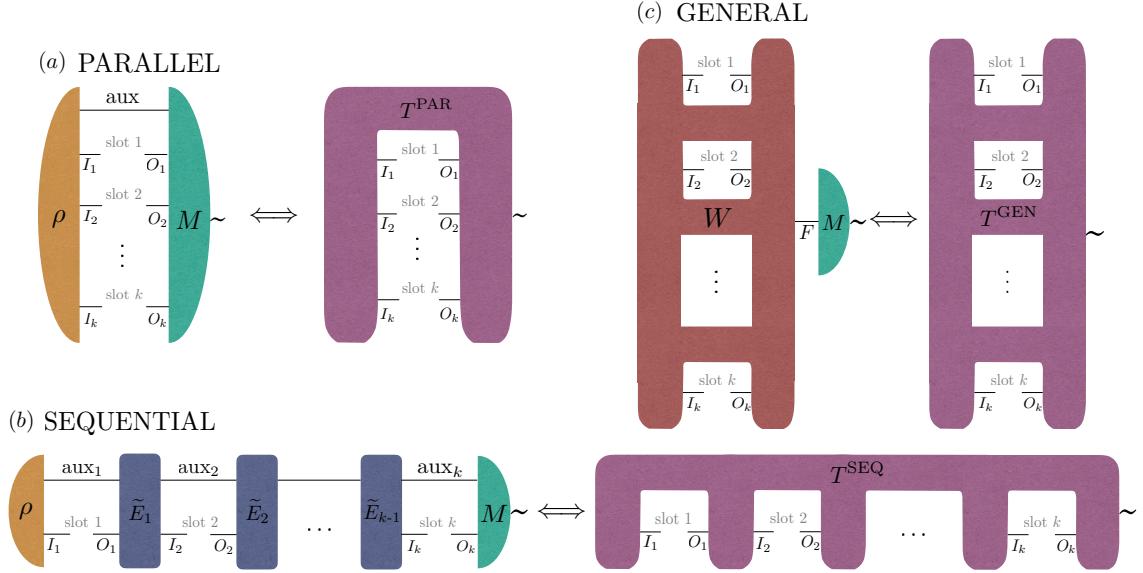


Figure 4.1: Schematic representation of the realization of every  $k$ -copy (a) parallel tester  $T^{\text{PAR}}$  with a state  $\rho$  and a POVM  $M$ , (b) sequential tester  $T^{\text{SEQ}}$  with a state  $\rho$ , channels  $\tilde{E}_i$ ,  $i \in \{1, k - 1\}$ , and a POVM  $M$ , and (c) general tester  $T^{\text{GEN}}$  with a process matrix  $W$  and a POVM  $M$ .

an indefinite causal order, that is, they are valid processes that are neither ordered, nor convex combinations of ordered processes.

Contrarily to the parallel and sequential cases, a realization of general testers in terms of quantum operations (states, channels, and measurements) is an open problem. More specifically, a general tester can always be constructed from a general process matrix and a POVM [45], as illustrated in Fig. 4.1(c), however, only a subset of process matrices are currently known to be realizable with quantum operations [47] (see also [48]). This subset, known as ‘quantum control of causal order’, has been shown to bring advantage to the discrimination of general channels [45].

Regardless of the 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

$$P^{\mathcal{S}} = \max_{\{T_i^{\mathcal{S}}\}} \sum_{i=1}^N p_i \text{Tr} (T_i^{\mathcal{S}} C_i^{\otimes k}), \quad (4.11)$$

where  $\mathcal{S} \in \{\text{PAR}, \text{SEQ}, \text{GEN}\}$ .  $P^{\mathcal{S}}$  can be computed via semidefinite programming (SDP).

## 4.4 Discrimination of unitary channels

We start considering a discrimination task that involves unitary operations that form a group. We show that, for an ensemble composed of a set of unitaries that form a group and a uniform distribution, parallel strategies do not only perform as well as sequential ones [143] but are indeed the optimal strategies for discrimination – even considering general strategies that may involve indefinite causal order. In the following, unitary channels will be simply denoted by unitary operators  $U$  that satisfy  $UU^\dagger = \mathbb{1}$ , and the Choi state of a unitary channel will be denoted as  $|U\rangle\langle U| \in \mathcal{L}(\mathcal{H}^I \otimes \mathcal{H}^O)$ , with  $|U\rangle := \sum_i (\mathbb{1} \otimes U)|ii\rangle$ .

**Theorem 4.1.** *For ensembles composed of a uniform probability distribution and a set of unitary channels that form 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  $\{T_i^{GEN}\}$ , there exists a parallel tester  $\{T_i^{PAR}\}_i$ , such that

$$\sum_{i=1}^N \text{Tr}(T_i^{PAR}|U_i\rangle\langle U_i|^{\otimes k}) = \sum_{i=1}^N \text{Tr}(T_i^{GEN}|U_i\rangle\langle U_i|^{\otimes k}). \quad (4.12)$$

The proof of this theorem can be found in Appendix D.1.

Theorem 4.1 has two crucial hypotheses: (1) the set of unitary operators is a representation of some group and (2) the distribution  $\{p_i\}$  is uniform. If at least one of these hypotheses is not satisfied, then Theorem 4.1 in fact does not hold, as we show in the following.

**Theorem 4.2.** *There exist ensembles of unitary channels for which sequential strategies of discrimination outperform parallel strategies.*

Let us start with the case where the set of unitaries is not a unitary representation of a group but the probability distribution of the ensemble is uniform. In the following,  $\sigma_x, \sigma_y, \sigma_z$  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 4.1.** *The ensemble composed by a uniform probability distribution and  $N = 4$  qubit-qubit unitary channels given by  $U_1 = \mathbb{1}$ ,  $U_2 = \sigma_x$ ,  $U_3 = \sigma_y$ , and  $U_4 = \sqrt{\sigma_z}$ , in a discrimination task that allows for  $k = 2$  copies, can be discriminated under a sequential strategy success probability  $P^{SEQ} = 1$  while any parallel strategy copies yields  $P^{PAR} < 1$ .*

The proof of this example, in which we explicitly show a sequential strategy that allows for perfect discrimination, is in Appendix D.2.

The next example concerns a set of unitaries that form a group up to a global phase but the probability distribution of the ensemble is not uniform.

**Example 4.2.** Let  $\{U_i\}_i := \{\mathbb{1}, \sigma_x, \sigma_y, \sigma_z, H, \sigma_x H, \sigma_y H, \sigma_z H\}$  be a tuple of eight unitary channels that forms a group up to a global phase, and let  $\{p_i\}_i$  be a tuple of probabilities in which each element  $p_i$  is proportional the  $i$ -th digit of the number  $\pi \approx 3.1415926$ , that is,  $\{p_i\}_i := \{\frac{3}{31}, \frac{1}{31}, \frac{4}{31}, \frac{1}{31}, \frac{5}{31}, \frac{9}{31}, \frac{2}{31}, \frac{6}{31}\}$ . For the ensemble  $\{p_i, U_i\}_i$ , 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 D.2, and applies the method of computer-assisted proofs developed in Ref. [45]. In Example 4.2, we have set the distribution  $\{p_i\}_i$  to be proportional to the digits of the constant  $\pi$  to emphasise that the phenomenon of sequential strategies outperforming parallel ones when the set of unitary channels form a group does not require a particularly well chosen non-uniform distribution. In practice, we have observed that even with randomly generated distributions, optimal strategies often respect  $P^{\text{PAR}} < P^{\text{SEQ}}$ .

In the particular 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 4.3.** 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 unitaries is not a group but the probability distribution of the ensemble is uniform. 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  $| +_P \rangle := \frac{1}{5}(3| 0 \rangle + 4| 1 \rangle)$  and  $| -_P \rangle := \frac{1}{5}(4| 0 \rangle - 3| 1 \rangle)$ .

**Example 4.3.** For the ensemble composed by a uniform probability distribution and  $N = 4$  qubit-qubit unitary channels given by  $U_1 = \sqrt{\sigma_x}$ ,  $U_2 = \sqrt{\sigma_z}$ ,  $U_3 = \sqrt{H_P}$ , and  $U_4 = \sqrt{H_y}$ , 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 form a strict hierarchy  $P^{\text{PAR}} < P^{\text{SEQ}} < P^{\text{GEN}}$ .

The proof of this example can be found in Appendix D.3.

General strategies can also be advantageous for the discrimination of an ensemble composed by a non-uniform probability distribution and a set of unitaries that forms

a group. Let the set of unitaries in Example 4.3 be the set of generators of a group of unitary operators. Now consider the ensemble composed by such group and a probability distribution given by  $p_i = \frac{1}{4}$  for the four values of  $i$  corresponding to the four unitaries 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 4.3, 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 [43].

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  $\{|m\rangle\langle m|^c\}_i$  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

$$\begin{aligned} \mathcal{W}_{\text{SL}}(U_1, U_2) := & |0\rangle\langle 0|^c \otimes V_{02}(U_2 \otimes \mathbb{1}) V_{01}(U_1 \otimes \mathbb{1}) V_{00} \\ & + |1\rangle\langle 1|^c \otimes V_{12}(U_1 \otimes \mathbb{1}) V_{11}(U_2 \otimes \mathbb{1}) V_{10}, \end{aligned} \quad (4.13)$$

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 [43].

The switch-like superchannel has been previously considered in Refs. [56, 170], in the context of reversability-preserving transformations. Generalizations of the switch-like superchannel that transform  $k$  instead of 2 unitaries are presented in detail in Appendix D.4, applying unitaries  $\{V_{mn}\}_{m,n}$ , with  $m \in \{0, \dots, k! - 1\}$  and  $n \in \{0, \dots, k\}$ , and considering all permutations of the target unitaries  $\{U_l\}_{l=1}^k$ .

Now, let  $W^{\text{SL}} \in \mathcal{L}(\mathcal{H}^P \otimes \mathcal{H}^I \otimes \mathcal{H}^O \otimes \mathcal{H}^F)$  be the  $k$ -slot switch-like process associated with the  $k$ -slot generalization of the switch-like superchannel in Eq. (4.13). A general discrimination strategy, given by the  $k$ -copy switch-like tester  $T^{\text{SL}} = \{T_i^{\text{SL}}\}$ ,  $T_i^{\text{SL}} \in \mathcal{L}(\mathcal{H}^I \otimes \mathcal{H}^O)$  can be constructed using the a  $k$ -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  $\{M_i\}$ ,  $M_i \in \mathcal{L}(\mathcal{H}^F)$ , that acts on the ‘future’ space, according to

$$T_i^{\text{SL}} := \text{Tr}_{PF}[(\rho \otimes \mathbb{1})W^{\text{SL}}(\mathbb{1} \otimes M_i)] \quad (4.14)$$

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

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

$$\sum_{i=1}^N p_i \text{Tr} (T_i^{SL} |U_i\rangle\langle U_i|^{\otimes k}) = \sum_{i=1}^N p_i \text{Tr} (T_i^{SEQ} |U_i\rangle\langle U_i|^{\otimes k}). \quad (4.15)$$

The proof can be found in Appendix D.4, 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.

## 4.5 Numerical investigation

The advantage of sequential and general strategies in the discrimination of unitary channels is not restricted to main examples given in Theorems 4.2 and 4.3. In fact, by sampling sets of unitary channels uniformly distributed according to the Haar measure, and using these sets to construct ensembles with probability  $p_i = 1/N \forall i$ , one can find several other examples of the advantage of sequential and general strategies.

For the case of qubit-qubit unitaries and  $k = 2$  copies, we have observed gaps between parallel and sequential strategies for ensembles of  $N \in \{4, \dots, 25\}$  unitary channels. By calculating the averages of the maximal probabilities of success  $\langle P^{\text{PAR}} \rangle$  and  $\langle P^{\text{SEQ}} \rangle$ , we observed that for  $N \in \{4, \dots, 6\}$  the ratio  $\langle P^{\text{PAR}} \rangle / \langle P^{\text{SEQ}} \rangle$  decreases with  $N$ , the minimum ratio occurring at  $N = 6$ , and then increasing for  $N \in \{7, \dots, 25\}$ . At  $N = 25$ , gaps are hardly detected. This observation is in line with the idea that, in the limit where the ensemble is composed of all qubit-qubit unitary channels, therefore forming the group  $U(2)$ , it is expected that parallel strategies would be optimal.

We also remark that, for the case of qutrit-qutrit unitaries, we discovered a gap between the performance of parallel and sequential strategies already for a discrimination task of only  $N = 3$  unitaries using  $k = 2$  copies, while in the qubit case, the first example of this phenomenon was found only for  $N = 4$ . For the case of  $k = 2$  copies and uniformly sampled unitary channels, we have not found any advantage of general strategies over sequential ones.

Uniformly sampling qubit-qubit unitary channels		
$N$	$k = 2$	$k = 3$
2	$\mathbf{P}^{\text{PAR}} = \mathbf{P}^{\text{SEQ}} = P^{\text{GEN}}$	$\mathbf{P}^{\text{PAR}} = \mathbf{P}^{\text{SEQ}} = P^{\text{GEN}}$
3	$P^{\text{PAR}} = P^{\text{SEQ}} = P^{\text{GEN}}$	$P^{\text{PAR}} < P^{\text{SEQ}} = P^{\text{GEN}}$
4	$P^{\text{PAR}} < P^{\text{SEQ}} = P^{\text{GEN}}$	$P^{\text{PAR}} < P^{\text{SEQ}} < P^{\text{GEN}}$
$\vdots$	$\vdots$	$\vdots$
9	$P^{\text{PAR}} < P^{\text{SEQ}} = P^{\text{GEN}}$	$P^{\text{PAR}} < P^{\text{SEQ}} < P^{\text{GEN}}$
$\vdots$	$\vdots$	
25	$P^{\text{PAR}} \approx P^{\text{SEQ}} = P^{\text{GEN}}$	

Table 4.1: Summary of numerical findings.  $N$  denotes the number of unitary channels in the ensemble and  $k$  denotes the number of copies. The bold equalities on row  $N = 2$  mark analytical results [143]. A strict inequality between the maximal probabilities of success of different strategies in a certain scenario indicates that examples of ensembles that exhibit such gap were encountered. An equality indicates that, for all sampled ensembles, the maximal probabilities of success of different strategies were equal, up to numerical precision. More details on Appendix D.5.

However, the advantage of general strategies over causally ordered ones (parallel and sequential) is common for the task of discriminating unitaries with  $k = 3$  copies. Still considering uniformly sampled qubit-qubit unitaries, in the 3-copy case, we have found a strict hierarchy of discrimination strategies in scenarios of  $N \in \{4, \dots, 9\}$ , and an advantage of sequential over parallel strategies for the case of  $N = 3$ . Both the ratios of the averages  $\langle P^{\text{PAR}} \rangle / \langle P^{\text{SEQ}} \rangle$  and  $\langle P^{\text{SEQ}} \rangle / \langle P^{\text{GEN}} \rangle$  increase with  $N$  in the range in question.

A summary of these findings is presented in Table 4.1. More details can be found in Appendix D.5, including a plot of the ratios of the averages of the probabilities of success under different strategies.

## 4.6 Conclusions

We extended the unified tester formalism of Ref. [45] to the case of  $k$  copies and applied it particularly to the studied of discrimination tasks among only unitary channels. Our first contribution was to prove that, in a discrimination task among a set of unitaries that form a group, parallel strategy are always optimal, even when comparing against the performance of general strategies.

Subsequently, we showed the first example of a unitary discrimination task in which a sequential strategy outperforms any parallel strategy. Our task involves

the discrimination among 4 unitary channels using 2 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 4 unitary channels using 3 copies. A potential quantum realization of the optimal general strategies that are advantageous in this scenario is, unfortunately, still an open problem. We then demonstrated that, general strategies that are created from switch-like transformations can never perform better than sequential strategies for unitary channel discrimination.

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 advantage is indeed not possible, for any number of unitaries  $N$ . We also remark that, when considering  $k = 2$  copies, Refs. [56, 170] prove that superchannels which 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 a unitary channels, hence the argument of reversibility preservation combined with our Theorem 4.4 might lead to a proof for our conjecture.

All our code is available in an online repository [171] and can be freely used and edited.



# Concluding Discussion

As bipartite and high-dimensional entanglement become theoretically and experimentally tamed, the next step in entanglement research seems to be in the direction of quantum networks.

A quantum network is composed of multiple distant parties who share entangled states with each other and are equipped with the tools to execute multipartite quantum protocols of communication or cryptography, for example. As a current technological challenge is the construction of a single source that can prepare and distribute quantum states that carry genuine multipartite correlations across several parties, we can focus on the already well-developed bipartite entanglement sources and develop the methods to apply them as building blocks of a fully-connected quantum network.

To approach this task, recent works have been presented [172, 173]. Some of the topics addressed so far are how bipartite entanglement can be distributed pairwise through a network that is able to apply joint operations to the subsystems of each node in order to transform several bipartite states into a genuinely connected multipartite network and what is the most appropriate notion of entanglement in this scenario.

This line of research is showing itself to be very promising, with also previous developments having been achieved on the study of Bell nonlocality in networks [174–177]. It goes very much in the line of the work developed in the first part of this thesis, and takes the goal of developing practically implementable theoretical tools for the certification of complex quantum properties that can make maximal use of currently available technologies from the topic of high-dimensional entanglement to that of network entanglement.

On a different front, as more advantages for information-theoretic and computational tasks due to indefinite causal order are uncovered, the motivation for understanding precisely which process matrices can be physically implemented becomes stronger.

An interesting parallel<sup>1</sup> to the problem of the implementability of process matrices can be drawn for the (no-longer-a) problem of the implementability of transformations of quantum states.

When asking what are the most general operations that can transform quantum states into quantum states, one can quickly arrive at the answer: positive and trace-preserving maps. Since these mathematical objects, when applied to a density matrix will yield a density matrix, one could expect them to plausibly correspond to physical operations. Except, as we know, if one applies a positive map to *part* of a larger density matrix, the resulting object might no longer correspond to a quantum state (as the resulting operator may have negative eigenvalues). Even though the answer is correct, one might then realize that the question initially asked might not have been the most relevant one. What is the most general operation that transforms quantum states into quantum states even when applied to only a part of a quantum state? The answer then is a *completely-positive* and trace-preserving map and it is the correct answer to this question. But how can we know whether the question was sufficiently broad to guarantee that we have now arrived at an answer that indeed corresponds to physical operations? One could argue, because of Stinespring's dilation theorem [178]. It guarantees that every completely-positive trace-preserving map can be implemented with the use of an auxiliary system and a unitary map acting on a larger-dimensional space. It is generally agreed that unitaries can be realized in the lab, so case closed.

When it comes to higher-order operations, we are not quite there yet. We do not know<sup>2</sup> how to go to the lab and implement general transformations that map quantum channels to quantum channels, or quantum instruments to probability distributions. But a fair amount of care has been put into posing the right question. The definition of a process matrix does not satisfy only the requirement of transforming quantum operations accordingly but also the requirement of properly transforming parts of larger quantum operations, and when allowing the involved parties to share additional non-signaling resources like entangled states. These requirements forbid, for example, objects who could achieve full two-way signalling or feed information to one's own past, creating closed time-like curves. The question now is whether these requirements, imposed from within the perspective of quantum information theory, are sufficient to single out the processes that can be implemented. Only further investigation will be able to tell us.

---

<sup>1</sup>Brought to my attention in a discussion with Marcelo Terra Cunha.

<sup>2</sup>And not just because we are theorists. The experimentalists also do not know.

# **Appendix**



# APPENDIX A

## Supplemental Information of Chapter 1

In this supplemental material, we provide detailed proofs and additional calculations illustrating the versatility of the results presented in the main text, as well as more information on the experimental implementation. To provide some context, let us compactly summarize the main results:

**Fidelity bound:**  $\tilde{F}^{(M)}(\rho, \Phi) \leq F(\rho, \Phi)$

- Obtained from measurements in  $M + 1$  *global product* bases;
- *Exact* for dephased pure states with only *two* bases ( $M = 1$ );
- Free of assumptions about the state  $\rho$ ;
- *Exact* in prime dimensions for  $M = d$ ;
- Also works for certain classes of multipartite entangled states;

**Schmidt number witness:**  $\tilde{F}(\rho, \Phi) \Rightarrow d_{\text{ent}}$

- *Exact* for all pure states;
- *Exact* for dephased max. entangled states;

**Entanglement bound:**  $\tilde{F}(\rho, \Phi^+) \Rightarrow \mathcal{E}_{\text{oF}}(\rho)$

- *Improvement* w.r.t. previous bounds [94].

The basis for these results are measurements in two ( $M = 1$ ) (or more ( $M > 1$ )) global product bases, one of which – the standard basis  $\{|mn\rangle\}_{m,n}$  – provides initial data (a set of values  $\{\lambda_m\}$ ) that is used to construct the other (“tilted”) basis. To summarize this method:

**Adaptive strategy for certifying entanglement dimensionality:**

- (1) Identify standard basis  $\{|mn\rangle\}$  and measure coincidences  $\{N_{mn}\}$  to obtain  $\{\langle mn|\rho|mn\rangle\}$ .
- (2) Calculate  $\{\lambda_m\}$  and nominate target state  $|\Phi\rangle$ .
- (3) Construct tilted basis  $\{|\tilde{j}\rangle\}$  and measure coincidences  $\{\tilde{N}_{ij}\}$  to obtain  $\{\langle \tilde{j}\tilde{j}^*|\rho|\tilde{j}\tilde{j}^*\rangle\}$ .
- (4) Evaluate  $\tilde{F}(\rho, \Phi)$  and  $B_{k=1}(\Phi), \dots, B_{k=d-1}(\Phi)$ . The certified entanglement dimensionality is  $d_{\text{ent}} = \max\{k \mid \tilde{F}(\rho, \Phi) > B_{k-1}(\Phi)\}$ .

To be more precise, the (first) local tilted basis  $\{|\tilde{j}\rangle\}_{j=0,\dots,d-1}$  is constructed from the local standard basis  $\{|m\rangle\}_{m=0,\dots,d-1}$  according to

$$|\tilde{j}\rangle = \frac{1}{\sqrt{\sum_n \lambda_n}} \sum_{m=0}^{d-1} \omega^{jm} \sqrt{\lambda_m} |m\rangle. \quad (\text{A.1})$$

To obtain the values  $\{\lambda_m\}$ , we use the following method. As explained in the main text, local filters [e.g., an appropriately programmed spatial light modulator (SLM)] are employed to allow only systems in particular states to be detected. For a particular setting with fixed  $m$  and  $n$  corresponding to the global orthonormal basis  $\{|mn\rangle\}_{m,n}$  one then counts the coincidences  $N_{mn}$ , which give an estimate of the diagonal density matrix elements of the underlying system state  $\rho$  via

$$\langle mn|\rho|mn\rangle = \frac{N_{mn}}{\sum_{i,j} N_{ij}}. \quad (\text{A.2})$$

These matrix elements in turn determine the values

$$\lambda_m = \sqrt{\frac{\langle mm|\rho|mm\rangle}{\sum_n \langle nn|\rho|nn\rangle}}, \quad (\text{A.3})$$

which can be interpreted as nominating a *target state*  $|\Phi\rangle = \sum_{m=0}^{d-1} \lambda_m |mm\rangle$ . Measurements in the second (tilted) basis (and potential additional tilted bases) then allow to evaluate a lower bound  $\tilde{F}(\rho, \Phi)$  for the fidelity  $F(\rho, \Phi) \geq \tilde{F}(\rho, \Phi)$  to the target state, as well of a number of threshold values  $B_{k=1}(\Phi), \dots, B_{k=d-1}(\Phi)$ . A

Schmidt-rank of  $k$  is then certified if the fidelity bound  $\tilde{F}(\rho, \Phi)$  surpasses the value  $B_{k-1}(\Phi)$ , given by

$$B_k(\Phi) := \sum_{m=0}^{k-1} \lambda_{i_m}^2. \quad (\text{A.4})$$

Additional information on various aspects of this method and its implementation are given in the following. Section A.1 details how measurements in the tilted basis can be performed. In Sec. A.2, the noise robustness of our approach is discussed for the important special case of maximally entangled target states subject to white noise. In Sec. A.3, we discuss the generalization of the fidelity bounds to measurements in more than two bases. We continue by discussing some simple bounds for the entanglement of formation in Sec. A.4, before showing the connection to the fidelity bounds to the maximally entangled state and discussing the robustness of these quantification techniques in comparison to previous methods in Sec. A.5. We show how the method can naturally be extended to the multipartite case in Sec. A.6. In Sec. A.7 we analyse the effects of a non-ideal choice of the standard basis, while Sec. A.8 shows evidence for the mutual unbiasedness of the implemented measurement bases. In Sec. A.8.1, we show an experimental example of a second spatial mode basis and discuss how mutually unbiased measurements can be readily implemented in a wide range of high-dimensional quantum systems using current technology. Finally, in Sec. A.8.2, we discuss two sources of systematic error introduced by our specific measurement devices – mode-dependent loss and imperfect hologram measurements.

## A.1 Normalization for measurements in the tilted bases

We now discuss in more detail how the measurements in the bases  $\{|mn\rangle\}_{m,n}$  and  $\{|\tilde{i}\tilde{j}^*\rangle\}_{i,j}$  can be performed by means of a post-selection procedure that we refer to as projective filtering. As explained above, estimates of the diagonal matrix elements  $\langle mn|\rho|mn\rangle$  of  $\rho$  w.r.t. the standard basis can be obtained from coincidence counting. For the standard basis, one finds  $\sum_{m,n} \langle mn|\rho|mn\rangle = 1$  by construction, which is sensible, since this expression corresponds to  $\text{Tr}(\rho)$  for an orthonormal basis. In other words,  $\sum_{m,n} |mn\rangle\langle mn| = \mathbb{1}$  is a resolution of the identity.

The same cannot be said for the (generally non-orthogonal) basis  $\{|\tilde{j}\rangle\}_j$ . However, the projectors  $\{|\tilde{j}\rangle\langle\tilde{j}|\}_j$  can be used to construct a valid non-projective  $(d+1)$ -outcome positive operator-valued measure (POVM). The first  $d$  elements of this POVM correspond to projectors in the tilted basis divided by a factor of  $d$ , while the last POVM element is obtained by subtracting the sum of the aforementioned elements from the identity, which results in a positive semi-definite operator, that is, the set of POVM

elements for a measurement in a tilted basis is  $\left\{ \frac{1}{d} \{ |\tilde{j}\rangle \langle \tilde{j}| \}_{j=0,\dots,d-1}, 1 - \frac{1}{d} \sum_{j=0}^{d-1} |\tilde{j}\rangle \langle \tilde{j}| \right\}$ . By construction this is a  $(d+1)$ -outcome measurement. However, when measurements are performed using projective filtering, only  $d$  filter settings, corresponding to the  $d$  projectors  $|\tilde{j}\rangle \langle \tilde{j}|$ , need to be performed if the measurement results of the standard basis are already available. To see this, note that projective filtering implies that instead of the probabilities  $p_j = \langle \tilde{j} | \rho | \tilde{j} \rangle$  and  $\bar{p} = \text{Tr}((1 - \sum_{j=0}^{d-1} |\tilde{j}\rangle \langle \tilde{j}|) \rho) = 1 - \sum_{j=0}^{d-1} p_j$ , one obtains only the count rates  $N_j = N p_j$  and  $\bar{N} = N \bar{p}$ , where  $N$  is the overall number of photons such that  $N = \bar{N} + \sum_{j=0}^{d-1} N_j$ . The  $d$  values  $N_j$  alone hence do not fully determine the desired values  $p_j = N_j/N$ , but the normalization factor  $N$  can be determined from  $\sum_{j=0}^{d-1} N_j$  together with the measurements already performed in the standard basis  $\{|m\rangle\}_m$ , which yield  $\sum_{j=0}^{d-1} p_j = \frac{1}{N} \sum_{j=0}^{d-1} N_j$ .

For the two-party scenario with measurements w.r.t. the global product basis  $\{|\tilde{i}\tilde{j}^*\rangle\}_{i,j}$ , this sum of density matrix elements in the tilted basis is calculated as

$$\sum_{i,j} \langle \tilde{i}\tilde{j}^* | \rho | \tilde{i}\tilde{j}^* \rangle = \frac{1}{(\sum_k \lambda_k)^2} \sum_{\substack{m,m' \\ n,n'}} \sqrt{\lambda_m \lambda_n \lambda_{m'} \lambda_{n'}} \times \langle m'n' | \rho | mn \rangle \sum_i \omega^{i(m-m')} \sum_j \omega^{j(n-n')} \quad (\text{A.5})$$

$$= \frac{d^2}{(\sum_k \lambda_k)^2} \sum_{m,n} \lambda_m \lambda_n \langle mn | \rho | mn \rangle =: c_\lambda, \quad (\text{A.6})$$

where we have defined the normalization factor  $c_\lambda$  as the inverse of the overall photon number and added the subscript  $\lambda$  to emphasize the dependence on the initial measurements in the standard basis. If we had naively considered the coincidence counts  $\tilde{N}_{ij}$  in the tilted basis, and the quantity analogous to the right-hand side of Eq. (A.2), we would have found  $\sum_{i,j} \frac{\tilde{N}_{ij}}{\sum_{k,l} \tilde{N}_{k,l}} = 1$ , by construction. To relate the coincidences to the matrix elements w.r.t. to the tilted basis, we hence include the additional normalization factor  $c_\lambda$  of Eq. (A.6), i.e.,

$$\langle \tilde{i}\tilde{j}^* | \rho | \tilde{i}\tilde{j}^* \rangle = c_\lambda \frac{\tilde{N}_{ij}}{\sum_{k,l} \tilde{N}_{k,l}}, \quad (\text{A.7})$$

as stated in the main text.

## A.2 Noise robustness

In this section, we discuss the special case of a maximally entangled target state, which is particularly interesting for several reasons. First, it provides a simple theoretical testing ground to evaluate the performance of our method in the presence of noise, as illustrated in Fig. A.1. There, we assume  $\rho$  to be a mixture of  $|\Phi^+\rangle$  with

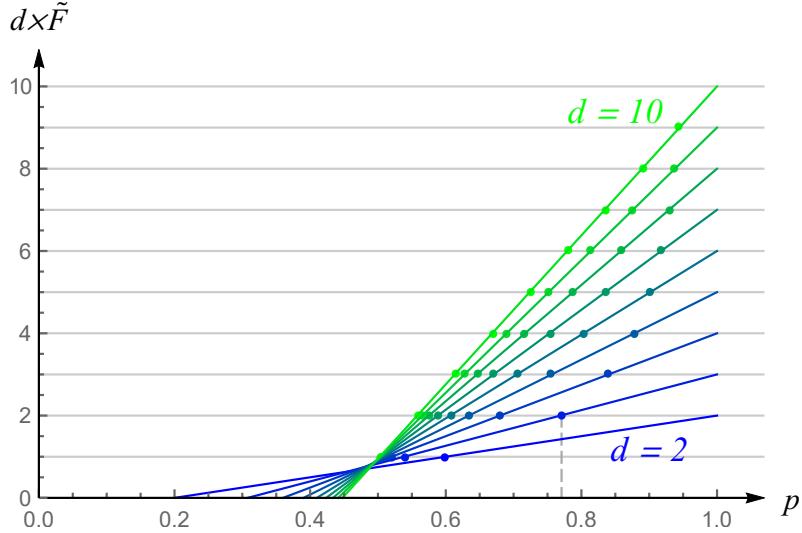


Figure A.1: **Noise-resistance of the fidelity bound for high-dimensional isotropic states.** The curves show the fidelity bound  $\tilde{F}(\rho_{\text{iso}}(p), \Phi^+)$  (weighted by the dimension  $d$ ) for isotropic states  $\rho_{\text{iso}}(p) = p|\Phi^+\rangle\langle\Phi^+| + \frac{1-p}{d^2}\mathbb{1}$  in  $d \times d$  dimensions as functions of the visibility  $p$  for  $d = 2$  (blue) to  $d = 10$  (green) in steps of 1. The intersections of the curves with the horizontal lines at the points  $(p_k(d), d \times \tilde{F}(\rho_{\text{iso}}(p_k), \Phi^+))$  (colored dots), where the intersection coordinates on the vertical axis are  $d \times \tilde{F}(\rho_{\text{iso}}(p_k), \Phi^+) = d \times B_k(\Phi^+) = k \in \{1, \dots, 9\}$ , indicate that visibilities  $p > p_k$  certify an entanglement dimensionality of at least  $d_{\text{ent}} = k + 1$ . In other words, for any  $p$  the certified dimension is  $d_{\text{ent}} = \lceil d \times \tilde{F}(\rho_{\text{iso}}, \Phi^+) \rceil$ . For instance, for isotropic states in local dimension  $d = 3$  with visibility  $p > p_{k=2}(d = 3) = \frac{10}{13}$  (vertical dashed line), our fidelity bound certifies Schmidt rank  $d_{\text{ent}} = 3$ .

a maximally mixed state, i.e., an isotropic state  $\rho_{\text{iso}} = p|\Phi^+\rangle\langle\Phi^+| + \frac{1-p}{d^2}\mathbb{1}$ , where the visibility  $p$  satisfies  $0 \leq p \leq 1$  and  $\mathbb{1}$  is the identity operator in dimension  $d^2$ . This allows us to identify the visibility thresholds for the certification of the Schmidt ranks of maximally entangled states subject to white noise. Second, the fidelity bounds for the target state  $|\Phi^+\rangle$  can be used to construct bounds on the entanglement of formation, as explained in the Supplementary Information. Although the selection of  $|\Phi^+\rangle$  as a target state may not be optimally suited for a given experimental situation, it thus nonetheless provides an efficient method for the direct certification of the number of e-bits in the system. In Sec. A.5, we show that this entanglement quantification method outperforms previous approaches [94] in terms of detected e-bits and noise robustness.

### A.3 Improved bounds using multiple bases

Next, we will show how measurements in more than one tilted basis can be included to improve the fidelity bounds. To this end, first note that the choice of

tilted basis is not unique. For instance, all of the statements made so far about the properties of the tilted basis would remain unaffected if additional phase factors independent of  $j$  were to be included in the definition of  $|\tilde{j}\rangle$ . That is, we have only relied on using identities such as  $\sum_j \omega^{j(m-n)} = d \delta_{mn}$ . For example, let us consider a family of tilted bases  $\{|\tilde{j}_k\rangle\}_{j,k}$  parameterized by an integer  $k \geq 0$ , such that

$$|\tilde{j}_k\rangle = \frac{1}{\sqrt{\sum_n \lambda_n}} \sum_{m=0}^{d-1} \omega^{jm+km^2} \sqrt{\lambda_m} |m\rangle. \quad (\text{A.8})$$

For  $k = 0$  we hence recover the original tilted basis. When the target state is a product state (and hence separable), all vectors within any tilted basis collapse to the same standard basis vector (up to a global phase factor), and are hence fully contained within the standard basis. In this case, and indeed, whenever any of the Schmidt coefficients vanish identically, tilted bases are no longer complete, and hence cannot technically even be considered to be bases anymore. However, when the target state is maximally entangled,  $|\Phi\rangle = |\Phi^+\rangle$ , we have  $\lambda_n = \frac{1}{\sqrt{d}} \forall n$ , in which case all of the tilted bases become orthonormal. Moreover, in this case one can recognize this construction as that of Ref. [98], i.e., for prime dimensions the choices  $k \in \{0, 1, \dots, d-1\}$  provide a maximal set of  $d$  mutually unbiased bases (MUBs),  $d+1$  if one includes the standard basis  $\{|m\rangle\}_m$ . For non-prime dimensions, the construction still provides an MUB w.r.t. to the standard basis for every choice of  $k$ , but the bases for different  $k$  are in general not unbiased w.r.t to each other. We will return to these interesting special cases in Sec. A.5.

In the more realistic scenario where  $|\Phi\rangle$  is not separable but also not maximally entangled and all Schmidt coefficients  $\lambda_m$  (as estimated from initial measurements in the standard basis) have arbitrary nonzero values, we may construct nonorthogonal but complete tilted bases  $\{|\tilde{j}_k\rangle\}_{j,k}$  according to Eq. (A.8). As for the MUBs, this construction provides  $d$  inequivalent tilted bases for odd prime dimensions, measurements w.r.t. which are sufficient for the fidelity bound to become tight, as we shall discuss in the following. To see this, first note that the only contribution of the additional phases  $\omega^{km^2}$  appears in the complex coefficient  $c_{mnm'n'} = \sum_j \omega^{j(m-m'-n+n')}$ , which we can then replace by

$$c_{mnm'n'}^{(k)} := \sum_j \omega^{j(m-m'-n+n')} \omega^{k(m^2-m'^2-n^2+n'^2)}. \quad (\text{A.9})$$

Clearly, when using any single one of the bases  $\{|\tilde{j}_k\rangle\}_{j,k}$ , the modification of the constant  $c_{mnm'n'}$  becomes irrelevant again due to the modulus, i.e.,  $|c_{mnm'n'}^{(k)}| = |c_{mnm'n'}^{(0)}|$  for all  $k$ .

However, we may use several of the tilted bases simultaneously to obtain an advantage. Replacing the term  $\Sigma = \sum_{j=0}^{d-1} \langle \tilde{j}\tilde{j}^* | \rho | \tilde{j}\tilde{j}^* \rangle$  by an average over  $M$  different tilted bases as defined by Eq. (A.8), i.e.,

$$\Sigma \rightarrow \Sigma^{(M)} = \frac{1}{M} \sum_{k=0}^{M-1} \sum_{j=0}^{d-1} \langle \tilde{j}_k \tilde{j}_k^* | \rho | \tilde{j}_k \tilde{j}_k^* \rangle, \quad (\text{A.10})$$

one finds that the only affected term in the bound  $\tilde{F}_2$  for  $F_2$  is  $\Sigma_3$ . That is, we may replace the coefficient  $\tilde{\gamma}_{mm'nn'}$  by the modified coefficient

$$\tilde{\gamma}_{mm'nn'}^{(M)} = \tilde{\gamma}_{mm'nn'} \frac{1}{M} \left| \sum_{k=0}^{M-1} \omega^{k(m^2 - m'^2 - n^2 + n'^2)} \right|, \quad (\text{A.11})$$

and define the quantity  $\tilde{F}^{(M)} := F_1 + \tilde{F}_2^{(M)} \leq F$ , where

$$\tilde{F}_2^{(M)} := \frac{(\sum_m \lambda_m)^2}{d} \Sigma^{(M)} - \sum_{m,n=0}^{d-1} \lambda_m \lambda_n \langle mn | \rho | mn \rangle - \sum_{\substack{m \neq m', m \neq n \\ n \neq n', n' \neq m'}} \tilde{\gamma}_{mm'nn'}^{(M)} \sqrt{\langle m'n' | \rho | m'n' \rangle \langle mn | \rho | mn \rangle}. \quad (\text{A.12})$$

In the least favourable possible case all phases in the sum over  $k$  are aligned and  $\tilde{\gamma}_{mm'nn'}^{(M)} = \tilde{\gamma}_{mm'nn'}$ , but in general  $\tilde{\gamma}_{mm'nn'}^{(M)} \leq \tilde{\gamma}_{mm'nn'}$ . Consequently, the fidelity bounds can only be improved by including measurements in more than one tilted basis.

In fact, when the dimension  $d$  is a (non-even) prime, we have  $\tilde{F}^{(M')} \geq \tilde{F}^{(M)}$  for  $M' \geq M$ , and for  $M = d$  the prefactor  $\tilde{\gamma}_{mm'nn'}^{(M=d)}$  vanishes exactly and the fidelity bound becomes tight, i.e.,  $F = \tilde{F}^{(M=d)}$ . In order to show this, we need to examine the sum in Eq. (A.11). At first it is important to realize that since the value of  $\tilde{\gamma}_{mm'nn'}$  does not depend on  $k$ , only cases for which  $(m - m' - n + n') \bmod d = 0$  need to be examined, otherwise  $\tilde{\gamma}_{mm'nn'} = 0$  leads to  $\tilde{\gamma}_{mm'nn'}^{(M)} = 0$ . Let us therefore prove the following claim. For parameter choices fulfilling the conditions

$$\begin{aligned} m &\neq m', m \neq n, \\ n &\neq n', n' \neq m', \\ (m - m' - n + n') &\bmod d = 0 \end{aligned} \quad (\text{A.13})$$

it holds that  $(m^2 - m'^2 - n^2 + n'^2) \neq 0$ . We will prove this claim by contradiction. In order to do so, suppose that both of the following equalities hold

$$m + n' = m' + n \bmod d \quad (\text{A.14})$$

$$m^2 + (n')^2 = (m')^2 + n^2 \bmod d. \quad (\text{A.15})$$

Without loss of generality suppose  $m > n$ , which also implies  $m' > n'$ . Let us define  $c := m - n = m' - n'$ , which allows us to rewrite Eq. (A.15) as

$$\begin{aligned} m^2 + n'^2 &= (n' + c)^2 + (m - c)^2 \pmod{d} \\ m^2 + n'^2 &= (n')^2 + 2cn' + c^2 + m^2 - 2cm + c^2 \pmod{d} \\ 0 &= 2c^2 + 2cn' - 2cm \pmod{d} \\ 0 &= 2c(c + n' - m) \pmod{d} \\ 0 &= 2c(m' - m) \pmod{d}. \end{aligned} \tag{A.16}$$

The last equality holds, if and only if  $2c(m' - m)$  is a multiple of  $d$ . Since  $d$  is an odd prime, the only possibility is that either  $c$  or  $(m' - m)$  are multiples of  $d$ . Clearly, since  $c = m - n$ ,  $m > n$  and  $m, n \in \{0, \dots, d-1\}$ ,  $0 < c < d$ , and  $c$  is therefore not a multiple of  $d$ . Similarly, since  $m \neq m'$  and  $m, m' \in \{0, \dots, d-1\}$ ,  $-d < (m' - m) < d$ , therefore  $(m' - m)$  is not a multiple of  $d$ . We hence arrive at a contradiction with Eq. (A.16) and conclude that under the conditions of (A.13) we have  $(m^2 - m'^2 - n^2 + n'^2) \neq 0$ .

Therefore, when working with  $M$  different tilted bases,  $\sum_{k=0}^{M-1} \omega^{k(m^2 - m'^2 - n^2 + n'^2)}$  is a sum of  $M$  *different*<sup>1</sup> powers of  $\omega$ . We subsequently have to show that the absolute value of this sum can be bounded to be strictly lower than  $M$ . Moreover, the bound improves with increasing  $M$ , and whenever  $M = d$ , the sum in Eq. (A.11) and hence also the sum in the last line of Eq. (A.12) vanishes. Before we turn to the more general statement for arbitrary  $M$ , let us briefly focus on the case  $M = d$ , where it can be easily seen that for non-zero  $(m^2 - m'^2 - n^2 + n'^2) \sum_{k=0}^{d-1} \omega^{k(m^2 - m'^2 - n^2 + n'^2)} = 0$ .

For general values  $M < d$  let us now analytically bound  $|\sum_{k=0}^{M-1} \omega^{kc}|$ , where  $c$  is a non-zero integer. Naturally, the exact value of this sum depends on the particular value of  $c$ , but here we give a general bound. To this end, we first argue that the worst case (the highest possible sum) corresponds to the situation, where  $kc$  ranges over subsequent powers of  $\omega$  (i.e.  $c = 1$ ). This can be seen from the fact that powers of  $\omega$  can be represented in the complex plane as vectors lying on the unit circle with the centre at the origin. The absolute value of the sum of several different powers of  $\omega$  can therefore be seen as the size of the sum of their corresponding vectors. Recall that for odd-prime dimension  $d$ , the exponent  $kc$  ranges over  $M$  different numbers between 0 and  $d-1$ . Now it is not hard to see that by fixing the number of vectors  $M$ , the worst case sum (i.e., the largest absolute value) corresponds to the sum of the  $M$  vectors next to each other on the complex plane, which in turn corresponds to

---

<sup>1</sup>The difference of the powers results from the fact that in the mod prime multiplicative group, every non-zero element is a generator of the whole group. This means that since  $(m^2 - m'^2 - n^2 + n'^2)$  is non-zero, iterating over different values of  $k$  results in different values of the whole exponent.

the subsequent powers of  $\omega$ . With this knowledge, we have to bound one particular worst case sum, given by

$$\sum_{k=0}^{M-1} \omega^k = \sum_{k=0}^{M-1} e^{\frac{2\pi i k}{d}}. \quad (\text{A.17})$$

Using a variant of the Dirichlet kernel [179], i.e.,

$$\sum_{k=0}^{M-1} e^{iMx} = e^{\frac{i(M-1)x}{2}} \frac{\sin\left(\frac{Mx}{2}\right)}{\sin\left(\frac{x}{2}\right)} \quad (\text{A.18})$$

with  $x = \frac{2\pi}{d}$ , we have

$$\sum_{k=0}^{M-1} \omega^k = e^{\frac{i(M-1)\pi}{d}} \frac{\sin\left(\frac{M\pi}{d}\right)}{\sin\left(\frac{\pi}{d}\right)}. \quad (\text{A.19})$$

Taking the absolute value reveals that for any choice of non-zero integer  $c$  we have

$$\left| \sum_{k=0}^{M-1} \omega^{kc} \right| \leq \frac{|\sin\left(\frac{M\pi}{d}\right)|}{|\sin\left(\frac{\pi}{d}\right)|}. \quad (\text{A.20})$$

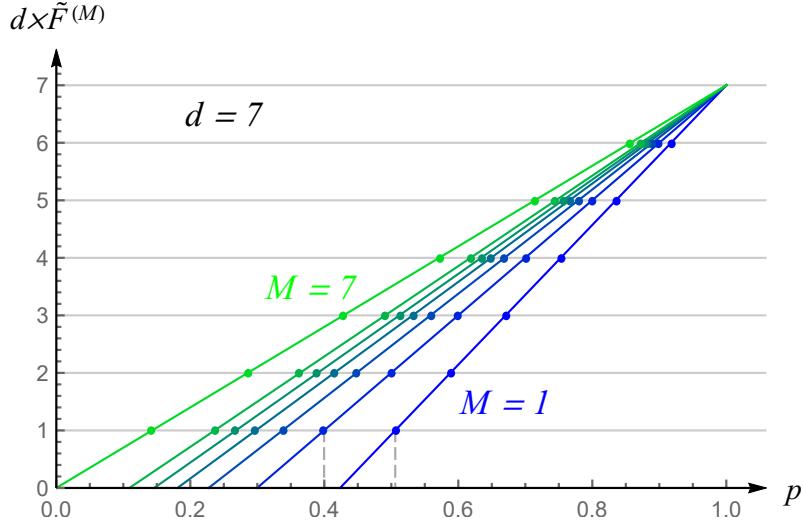
After plugging this lower bound into Eq. (A.11), all (non-zero) prefactors  $\tilde{\gamma}_{mm'nn'}^{(M)}$  become decreasing functions of  $M$ , on the interval  $1 \leq M \leq d$ , which concludes the proof that  $\tilde{F}^{(M')} \geq \tilde{F}^{(M)}$  for  $M' \geq M$  in odd prime dimensions.

For general dimension  $d$ , however, it is not the case that  $\tilde{F}^{(M')} \geq \tilde{F}^{(M)}$  for  $M' \geq M$ , except for the case when  $M = 1$  (for any dimension).

An illustration of the improvement obtained by including multiple tilted bases is given in Fig. A.2 for an isotropic state  $\rho_{\text{iso}} = p|\Phi^+\rangle\langle\Phi^+| + \frac{1-p}{d^2}\mathbb{1}$  in dimension  $d = 7$ . Such a state highlights the influence of white noise on the certification method, since the isotropic state is a mixture of a maximally entangled and a maximally mixed state. We have hence shown that an improvement of the bounds by using more than two global product bases is possible in principle. In Sec. A.5 we will further illustrate this improvement for quantifying entanglement.

## A.4 Bounds on the entanglement of formation

In this section, we discuss a method for bounding the entanglement of formation in bipartite systems of arbitrary dimension. To provide a self-contained approach, let us first give a pedagogical review of the entanglement of formation and useful bounds for it also discussed in Ref. [94], before we make use of the fidelity bounds established thus far in Sec. A.5. To begin, recall that the subsystems  $A$  and  $B$



**Figure A.2: Improved fidelity bound & dimensionality witness for isotropic state.** The curves show the fidelity bound  $\tilde{F}^{(M)}(\rho_{\text{iso}}(p), \Phi^+)$  (weighted by the local dimension  $d = 7$ ) for isotropic states  $\rho_{\text{iso}}(p) = p|\Phi^+\rangle\langle\Phi^+| + \frac{1-p}{d^2}\mathbb{1}$  in  $d \times d$  dimensions as functions of the visibility  $p$  for local dimension  $d = 7$  for different numbers of global product bases, i.e.,  $M = 1$  (blue) to  $M = 7$  (green) in steps of 1. The intersections of the curves with the horizontal lines at the points  $(p_k^{(M)}(d), d \times \tilde{F}^{(M)}(\rho_{\text{iso}}(p_k^{(M)}), \Phi^+))$  (colored dots), where the intersection coordinates on the vertical axis are  $d \times \tilde{F}^{(M)}(\rho_{\text{iso}}(p_k^{(M)}), \Phi^+) = d \times B_k(\Phi^+) = k \in \{1, \dots, 6\}$ , indicate that visibilities  $p > p_k^{(M)}$  certify an entanglement dimensionality of at least  $d_{\text{ent}} = k + 1$  when  $M$  tilted bases are used. In other words, for any  $p$  the certified dimension is  $d_{\text{ent}} = \lceil d \times \tilde{F}^{(M)}(\rho_{\text{iso}}, \Phi^+) \rceil$ . For instance, for isotropic states in local dimension  $d = 7$  our fidelity bound with one tilted basis ( $M = 1$ ) certifies Schmidt rank  $d_{\text{ent}} = 2$  for a visibility  $p > p_{k=1}^{(M=1)}(d = 7) = \frac{43}{85}$  (right vertical dashed line), whereas for two tilted bases ( $M = 2$ ) a visibility  $p > p_{k=1}^{(M=2)}(d = 7) \approx 0.3997$  (left vertical dashed line) is enough to certify  $d_{\text{ent}} = 2$ .

of a pure bipartite state  $|\psi\rangle_{AB}$  are entangled if and only if their reduced states  $\rho_A = \text{Tr}_B(|\psi\rangle\langle\psi|)$  and  $\rho_B = \text{Tr}_A(|\psi\rangle\langle\psi|)$  are mixed. This fact can easily be seen from the Schmidt decomposition, i.e., that any pure state  $|\psi\rangle_{AB} \in \mathcal{H}_{AB} = \mathcal{H}_A \otimes \mathcal{H}_B$  may be written as

$$|\psi\rangle_{AB} = \sum_{m=0}^{k-1} \lambda_m |\phi_m\rangle_A |\chi_m\rangle_B \quad (\text{A.21})$$

with respect to the Schmidt bases  $\{|\phi_m\rangle_A\}_m$  and  $\{|\chi_m\rangle_B\}_m$ , and where  $k \leq \min\{\dim(\mathcal{H}_A), \dim(\mathcal{H}_B)\}$ . The entanglement of the state  $|\psi\rangle_{AB}$  may therefore be quantified by the mixedness  $1 - \text{Tr}(\rho_A^2)$  of the reduced states. More specifically, we can define the *entropy of entanglement*  $\mathcal{E}_L$  via the linear entropy  $S_L$  as

$$\mathcal{E}_L(|\psi\rangle) = S_L(\rho_A) = \sqrt{2(1 - \text{Tr}(\rho_A^2))}. \quad (\text{A.22})$$

This method for entanglement quantification can be extended to mixed states via a convex-roof construction, i.e., we define

$$\mathcal{E}_L(\rho) := \inf_{\mathcal{D}(\rho)} \sum_i p_i S_L(\rho_A^{(i)}) , \quad (\text{A.23})$$

where the infimum is taken over the set of all pure state decompositions of  $\rho$ , i.e.,

$$\mathcal{D}(\rho) = \left\{ \left\{ (p_i, \psi_i) \right\}_i \mid \rho = \sum_i p_i |\psi_i\rangle\langle\psi_i|, 0 \leq p_i \leq 1, \sum_i p_i = 1 \right\}, \quad (\text{A.24})$$

where  $\rho_A^{(i)} = \text{Tr}_B(|\psi_i\rangle\langle\psi_i|)$ .

A simple bound on this convex roof of the linear entropy was derived in Refs. [180, 181]. Defining the quantity

$$I(\rho) = \sqrt{\frac{2}{d(d-1)}} \sum_{m \neq n} \left( |\langle mm|\rho|nn\rangle| - \sqrt{\langle mn|\rho|mn\rangle\langle nm|\rho|nm\rangle} \right), \quad (\text{A.25})$$

for bipartite systems of equal local dimension  $d$ , i.e.,  $\dim(\mathcal{H}_A) = \dim(\mathcal{H}_B) = d$ , with bases  $\{|\phi_n\rangle_A \equiv |n\rangle_A\}$  and  $\{|\chi_n\rangle_B \equiv |n\rangle_B\}$ , it was shown in [180, 181] that

$$I(\rho) \leq \mathcal{E}_L(\rho). \quad (\text{A.26})$$

Now, we want to see how  $I(\rho)$  can be used to bound also the entanglement of formation (EoF) [92, 93], defined as the convex roof extension of the entropy of entanglement when the von Neumann entropy  $S(\rho) = -\text{Tr}(\rho \log(\rho))$  is used instead of the linear entropy, i.e.,

$$\mathcal{E}_{\text{oF}}(\rho) := \inf_{\mathcal{D}(\rho)} \sum_i p_i S(\rho_A^{(i)}). \quad (\text{A.27})$$

To understand this connection, let us briefly expand upon the derivation given in Ref. [94]. First, note that for pure states  $|\psi\rangle$  we have

$$I(|\psi\rangle) \leq \mathcal{E}_L(|\psi\rangle) = \sqrt{2(1 - \text{Tr}(\rho_A^2))}. \quad (\text{A.28})$$

Therefore, if  $I(|\psi\rangle) \geq 0$  we can write

$$\text{Tr}(\rho_A^2) \leq 1 - \frac{1}{2} I^2(|\psi\rangle), \quad (\text{A.29})$$

which implies that

$$-\log(\text{Tr}(\rho_A^2)) \geq -\log\left(1 - \frac{1}{2} I^2(|\psi\rangle)\right) \quad (\text{A.30})$$

since  $\log x$  is a monotonically increasing function. With the additional negative sign we can recognize the left-hand side as the Rényi 2-entropy, defined as

$$S_\alpha(\rho) := \frac{1}{1-\alpha} \log \text{Tr}(\rho^\alpha) \quad (\text{A.31})$$

for  $\alpha = 2$ . For all  $\alpha, \beta \in \mathbb{N}$  and for all  $\rho$ , the Rényi entropies satisfy  $S_\alpha(\rho) \geq S_\beta(\rho)$  for  $\alpha \leq \beta$ . In particular, this means that

$$S_1(\rho) = \lim_{\alpha \rightarrow 1} S_\alpha(\rho) \geq S_2(\rho) = -\log(\text{Tr}(\rho^2)) \quad (\text{A.32})$$

and consequently one has

$$S_1(\rho_A) \geq -\log\left(1 - \frac{1}{2}I^2(|\psi\rangle)\right). \quad (\text{A.33})$$

For pure states, the (von Neumann) entropy of the subsystem is equal to the EoF and we have hence obtained the desired bound. To see that the bound also holds for mixed states, simply note that  $-\log(1 - x^2/2)$  is a convex function. Similarly, the function  $I(\rho)$  is convex, since

$$I_1 := \sum_{m \neq n} |\langle mm|\rho|nn\rangle| \quad (\text{A.34})$$

is convex, while

$$I_2 := \sum_{m \neq n} \sqrt{\langle mn|\rho|mn\rangle\langle nm|\rho|nm\rangle} \quad (\text{A.35})$$

is concave, i.e., by Jensen's inequality [182]

$$I_1\left(\sum_i p_i \rho_i\right) \leq \sum_i p_i I_1(\rho_i), \quad (\text{A.36})$$

$$I_2\left(\sum_i p_i \rho_i\right) \geq \sum_i p_i I_2(\rho_i), \quad (\text{A.37})$$

for  $0 \leq p_i \leq 1$  and  $\sum_i p_i = 1$ . This allows us to conclude that for all states  $\rho$ , for which  $I(\rho) \geq 0$  one has

$$\mathcal{E}_{\text{oF}}(\rho) \geq -\log\left(1 - \frac{1}{2}I^2(\rho)\right). \quad (\text{A.38})$$

Here, it is first useful to note here that the value of  $I(\rho)$  (in particular, whether or not  $I$  is non-negative) for a given state depends on the bases  $\{|m\rangle_A\}_m$  and  $\{|n\rangle_B\}_n$  that are chosen. For instance, if both bases are chosen to be the same single-qubit bases and the quantum state in question is the singlet state  $|\psi^-\rangle = \frac{1}{\sqrt{2}}(|01\rangle - |10\rangle)$ , where  $|0\rangle$  and  $|1\rangle$  are assumed to be the eigenstates of the third Pauli matrix  $Z = \text{diag}\{1, -1\}$ , then  $I(|\psi^-\rangle) = -1$ . In other words, the bases  $\{|m\rangle_A\}_m$  and

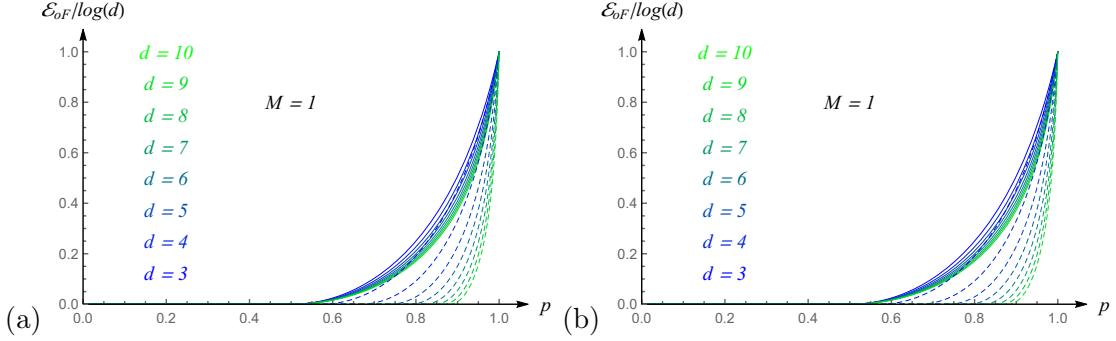


Figure A.3: **Entanglement bounds for isotropic state.** (a) The dashed and solid curves show the lower bounds for  $\mathcal{E}_{\text{OF}}$  obtained for  $M = 1$  and  $\rho_{\text{iso}}(p)$  using the bounds from Ref. [94] (dashed) and using the bound presented here in (A.42) (solid curves), respectively, for dimensions  $d = 3$  (blue) to  $d = 10$  (green) in steps of 1 and in units of  $\log d$ . It can be seen that the newly improved bounds can certify higher entanglement for given visibilities  $p$ . (b) The bound of Ref. [94] (orange, dashed) is compared with the bound of (A.42) (solid curves) for fixed dimension  $d = 7$  and varying numbers of bases,  $M = 1$  (blue) to  $M = 7$  (green) in steps of 1.

$\{|n\rangle_B\}_n$  should be chosen with a specific family of states in mind. For pure states, it is most useful to choose the Schmidt bases of the two subsystems.

Second, observe that, on the one hand, the term  $I_2$  contains only diagonal matrix elements and hence can practically easily be estimated using measurements in one pair of global product bases only. That is, counting the coincidences  $N_{mn}$  in the basis setting  $|m\rangle_A|n\rangle_B$ , we can reconstruct the desired matrix elements as  $\langle mn|\rho|mn\rangle = N_{mn}/(\sum_{i,j} N_{ij})$ . On the other hand, to estimate the off-diagonal matrix elements of the term  $I_1$  precisely, one would be required to reconstruct the entire density matrix by way of state tomography. However, this costly procedure can be avoided by supplementing the measurements in the basis  $\{|mn\rangle\}_{m,n}$  by measurements in one (or more) MUBs w.r.t.  $\{|mn\rangle\}_{m,n}$  to provide a lower bound on  $I_2(\rho)$ .

## A.5 Entanglement quantification using mutually unbiased bases

Having established the usefulness of the quantity  $I(\rho)$  for bounding the entanglement of formation, let us now relate it to the fidelity bounds we have discussed before. Inspection of the fidelity to the maximally entangled state, i.e.,

$$F(\rho, \Phi^+) = \frac{1}{d} \sum_m \langle mm|\rho|mm\rangle + \frac{1}{d} \sum_{m \neq n} \langle mm|\rho|nn\rangle, \quad (\text{A.39})$$

immediately lets us obtain the bound

$$\sum_{m \neq n} |\langle mm|\rho|nn\rangle| \geq \sum_{m \neq n} \langle mm|\rho|nn\rangle \quad (\text{A.40})$$

$$= d F(\rho, \Phi^+) - \sum_m \langle mm|\rho|mm\rangle. \quad (\text{A.41})$$

Since  $F(\rho, \Phi^+) \geq \tilde{F}^{(M)}$ , this, in turn, implies that

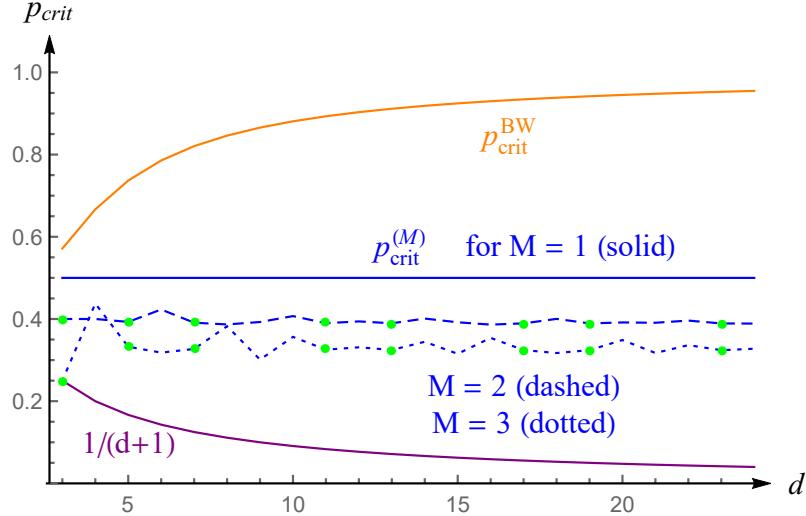
$$I(\rho) \geq \sqrt{\frac{2}{d(d-1)}} \left( d \tilde{F}^{(M)}(\rho, \Phi^+) - \sum_m \langle mm|\rho|mm\rangle - \sum_{m \neq n} \sqrt{\langle mn|\rho|mn\rangle \langle nm|\rho|nm\rangle} \right) \quad (\text{A.42})$$

$$\begin{aligned} &\geq \sqrt{\frac{2}{d(d-1)}} \left( d \Sigma^{(M)} - 1 - \sum_{m \neq n} \sqrt{\langle mn|\rho|mn\rangle \langle nm|\rho|nm\rangle} \right. \\ &\quad \left. - \sum_{\substack{m \neq m', m \neq n \\ n \neq n', n' \neq m'}} \tilde{\gamma}_{mm'nn'}^{(M)} \sqrt{\langle m'n'|\rho|m'n'\rangle \langle mn|\rho|mn\rangle} \right), \end{aligned} \quad (\text{A.43})$$

where we have inserted the fidelity bound  $\tilde{F}^{(M)}$  for multiple MUBs derived in Sec. A.3. The measurements performed to lower-bound the entanglement dimensionality of  $\rho$  may hence directly be used to also obtain a lower bound on the entanglement of formation.

We further note that the bound of (A.42) can also be considered to be a generalization of the bounds discussed in Ref. [94], where a similar, but strictly weaker bound for  $I(\rho)$  is provided, corresponding to setting  $M = 1$  and  $\tilde{\gamma}_{mm'nn'}^{(M)} \rightarrow 1$ . To provide direct comparisons of our bounds with the methods of Ref. [94], we again turn to the example of the isotropic state  $\rho_{\text{iso}} = p|\Phi^+\rangle\langle\Phi^+| + \frac{1-p}{d^2}\mathbb{1}$ , where  $0 \leq p \leq 1$ ,  $|\Phi^+\rangle = \frac{1}{\sqrt{d}} \sum_n |nn\rangle$ , and  $\mathbb{1}$  is the identity in dimension  $d^2$ . A comparison of the performance of these bounds for entanglement quantification for the assumed state  $\rho_{\text{iso}}$  is shown in Fig. A.3.

The isotropic state also provides an ideal theoretical testing ground for the noise robustness of these bounds, since it corresponds to mixing a maximally entangled state with white noise and hence allows to characterize the robustness of the entanglement bounds against decoherence. To this end, we compare the critical visibilities  $p_{\text{crit}}$ , that is, the parameters appearing in  $\rho_{\text{iso}}(p)$  for which the different methods stop detecting entanglement. Ideally, this could be the case for the value  $p_{\text{crit}} = \frac{1}{d+1}$ , below which the isotropic state is separable [32]. For the bound of Ref. [94] we find  $p_{\text{crit}}^{\text{BW}} = \frac{d^2-3d+4}{d^2-2d+4}$ , whereas our bound from (A.42) provides  $p_{\text{crit}}^{(M)} = \frac{d(d-1)+f(M)}{d(d^2-1)+f(M)}$ , where  $f(M) = \sum_{\substack{m \neq m', m \neq n \\ n \neq n', n' \neq m'}} \tilde{\gamma}_{mm'nn'}^{(M)}$ . As illustrated in Fig. A.4, the improved bounds presented here significantly improve on the noise resistance of the bounds.



**Figure A.4: Critical visibilities.** The curves show the parameters  $p$  for which the entanglement of the isotropic states in  $d \times d$  dimensions become undetectable using the bound of Ref. [94] (upper orange curve) and the bound of (A.42) for  $M = 1, 2, 3$  (blue solid, dashed, dotted curves), respectively. The bottom purple curves indicates the value below which  $\rho_{\text{iso}}$  is separable. The irregular behaviour of the curves for  $M > 1$  originates from the fact that the bases we use are all unbiased w.r.t. each other only in prime dimensions (green dots).

## A.6 Multipartite entanglement certification

In this appendix, we give a brief outlook on the multipartite case. For this purpose we define a family of generalized GHZ states for arbitrary local dimension and arbitrary weights  $\{\lambda_i\}_i$  as

$$|\text{GHZ}_{\lambda,n,d}\rangle := \sum_{i=0}^{d-1} \lambda_i |i\rangle^{\otimes n}, \quad (\text{A.44})$$

with  $\sum_i \lambda_i^2 = 1$ . The GHZ-weights  $\lambda_i$  can be interpreted as generalized Schmidt coefficients for this particular family of states and our fidelity method can be applied in full analogy to the bipartite cases discussed previously. As, before, we can introduce local tilted bases for the  $n$ -partite case as

$$|\tilde{j}^{(n)}\rangle := \frac{1}{\sqrt{\sum_k \lambda_k^{2/n}}} \sum_{m=0}^{d-1} \omega^{jm} \lambda_m^{1/n} |m\rangle, \quad (\text{A.45})$$

such that  $|\tilde{j}^{(n=2)}\rangle \equiv |\tilde{j}\rangle$  coincides with our previous definition for bipartite systems. We are then interested in determining a fidelity bound  $\tilde{F}(\rho, \text{GHZ}_{\lambda,n,d})$  such that

$$F(\rho, \text{GHZ}_{\lambda,n,d}) := \text{Tr}(\rho |\text{GHZ}_{\lambda,n,d}\rangle \langle \text{GHZ}_{\lambda,n,d}|) \quad (\text{A.46})$$

$$\geq \tilde{F}(\rho, \text{GHZ}_{\lambda,n,d}). \quad (\text{A.47})$$

Such a bound can indeed be found and, as we shall see, it takes the form

$$\tilde{F}(\rho, \text{GHZ}_{\lambda,n,d}) := \left( \sum_k \lambda_k^{2/n} \right)^n \langle \tilde{0}^{(n)} |^{\otimes n} \rho | \tilde{0}^{(n)} \rangle^{\otimes n} - \sum_{(\alpha,\beta) \in \gamma} \lambda_\alpha \lambda_\beta \sqrt{\langle \alpha | \rho | \alpha \rangle \langle \beta | \rho | \beta \rangle}. \quad (\text{A.48})$$

where  $\alpha = (i_1, \dots, i_n)$  and  $\beta = (j_1, \dots, j_n)$  are multi-indices with  $i_k, j_l \in \{0, 1, \dots, d-1\}$ , and we have used the notation  $|\alpha\rangle = |i_1, \dots, i_n\rangle$  and  $\lambda_\alpha := \prod_{i_k \in \alpha} \lambda_{i_k}^{1/n}$ . The sum in the second line of Eq. (A.48) runs over pairs of multi-indices in the set  $\gamma$ , which is given by

$$\gamma := \{(\alpha, \beta) | \alpha \notin \gamma_\alpha \vee \beta \notin \gamma_\beta\}, \quad (\text{A.49})$$

and  $\gamma_\alpha := \{\alpha = (i, i, \dots, i) | i = 0, 1, \dots, d-1\}$  are the sets of multi-indices where all sub-indices  $i_k$  are the same.

To prove the relation of Eq. (A.48), we expand the all-zero diagonal element in the tilted basis w.r.t. the standard basis, that is, inserting from Eq. (A.45) we write

$$\langle \tilde{0}^{(n)} |^{\otimes n} \rho | \tilde{0}^{(n)} \rangle^{\otimes n} = \left( \sum_k \lambda_k^{2/n} \right)^{-n} \sum_{\alpha, \beta} \lambda_\alpha \lambda_\beta \langle \alpha | \rho | \beta \rangle \quad (\text{A.50})$$

and observe that, just as in the bipartite case, all density matrix elements appear. We can then use this to replace terms in the fidelity on the left-hand side of Eq. (A.47) i.e.,

$$F(\rho, \text{GHZ}_{\lambda,n,d}) = \sum_{i,j} \lambda_i \lambda_j \langle i |^{\otimes n} \rho | j \rangle^{\otimes n} \quad (\text{A.51})$$

$$= \left( \sum_k \lambda_k^{2/n} \right)^n \langle \tilde{0}^{(n)} |^{\otimes n} \rho | \tilde{0}^{(n)} \rangle^{\otimes n} - \sum_{(\alpha,\beta) \in \gamma} \lambda_\alpha \lambda_\beta \langle \alpha | \rho | \beta \rangle. \quad (\text{A.52})$$

Now, we invoke the Cauchy-Schwarz inequality  $|\langle \alpha | \rho | \beta \rangle| \leq \sqrt{\langle \alpha | \rho | \alpha \rangle \langle \beta | \rho | \beta \rangle}$  to bound the last term in Eq. (A.51) as we have done in the case of bipartite states, such that we get

$$F(\rho, \text{GHZ}_{\lambda,n,d}) \geq \left( \sum_k \lambda_k^{2/n} \right)^n \langle \tilde{0}^{(n)} |^{\otimes n} \rho | \tilde{0}^{(n)} \rangle^{\otimes n} - \sum_{(\alpha,\beta) \in \gamma} \lambda_\alpha \lambda_\beta \sqrt{\langle \alpha | \rho | \alpha \rangle \langle \beta | \rho | \beta \rangle} \quad (\text{A.53})$$

$$= \tilde{F}(\rho, \text{GHZ}_{\lambda,n,d}). \quad (\text{A.54})$$

Note that in the case that  $\rho = |\text{GHZ}_{\lambda,n,d}\rangle \langle \text{GHZ}_{\lambda,n,d}|$  all the elements in the sum over  $(\alpha, \beta) \in \gamma$  vanish, as only terms  $\langle i |^{\otimes n} \rho | j \rangle^{\otimes n}$  appear and (A.47) becomes an equality. This shows that it is in principle possible to certify a unit fidelity with a multipartite and multi-dimensional target state for any  $n$  and  $d$ . However, using only a single

tilted basis element  $|\tilde{0}^{(n)}\rangle$  comes at the expense of reduced noise resistance, as we have seen in the bipartite case. Although this leaves room for improving the bound by the inclusion of additional tilted basis elements, the practical optimization over all potential combinations of phases is beyond the scope of this brief outlook.

## A.7 Effects of a wrong choice of Schmidt basis on the fidelity lower bounds

In this section we provide an example of how a choice of standard basis that does not correspond exactly to the Schmidt basis of the generated state affects the value of our fidelity lower bound  $\tilde{F}(\rho, \Phi)$ . Our example is based on the physically motivated situation in which there is a misalignment between the local reference frames of each party.

For the two-qutrit maximally entangled state  $|\Phi_3^+\rangle = \frac{1}{\sqrt{3}}(|00\rangle + |11\rangle + |22\rangle)$  we can assume without loss of generality that one side, Alice, performs the first measurement in the correct Schmidt basis while the other side, Bob, measures in a basis that is rotated w.r.t to Alice's measurement basis. This is due to the  $U \otimes U^*$  invariance of the isotropic states. Hence, let Alice measure in the standard basis  $\{|0\rangle, |1\rangle, |2\rangle\}$  and let Bob measure in a one-parameter rotation of a two-dimensional subspace of Alice's basis, namely,

$$|\bar{0}\rangle = \cos \theta |0\rangle + \sin \theta |1\rangle \quad (\text{A.55})$$

$$|\bar{1}\rangle = \sin \theta |0\rangle - \cos \theta |1\rangle \quad (\text{A.56})$$

$$|\bar{2}\rangle = |2\rangle. \quad (\text{A.57})$$

From the results of the measurements in the global product basis  $\{|m\bar{n}\rangle\}_{m,n}$ , one can compute the target state and the tilted basis for each party and complete the procedure outlined in the main text to obtain a fidelity lower bound and a certified Schmidt number. The results for this case are plotted in Fig. A.5 for this example.

This result illustrates how a sub-optimal choice of Schmidt basis can lead to suboptimal fidelity bounds and certified entanglement dimensionality. Crucially, however, it does not invalidate our method as the certified fidelity and entanglement are nonetheless valid. Moreover, one can see that, in our example, small deviations do not cause our fidelity bound to drop drastically, on the contrary, one can still certify maximal entanglement dimensionality up to at least 20% rotation.

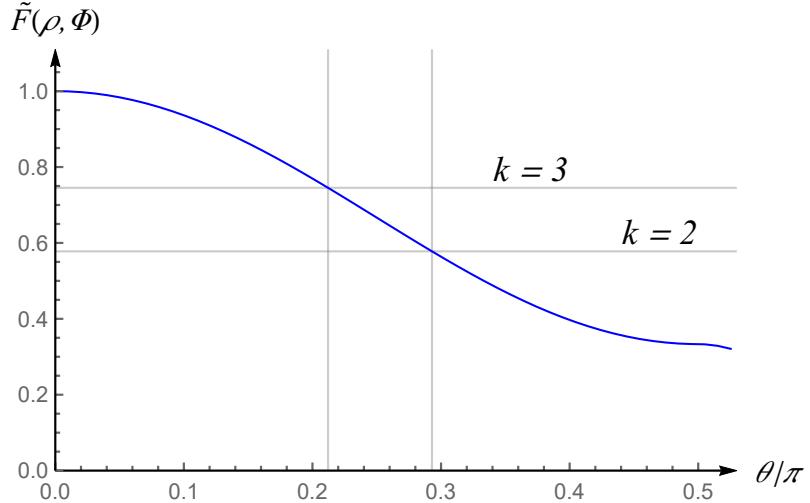


Figure A.5: Plot of fidelity lower bound  $\tilde{F}(\Phi_3^+, \Phi)$  for the maximally entangled two-qutrit state as a function of the rotation angle  $\theta$  when one of the sides measures in a standard basis that is rotated in a two-dimensional subspace w.r.t. the measurement basis on the other side. The horizontal lines show the threshold of the fidelity bound in which Schmidt numbers  $k = 3$  and  $k = 2$  can be certified.

## A.8 Classical prepare-and-measure experiment: LG basis

Here we demonstrate a classical experiment in which we prepare modes in the standard Laguerre-Gaussian (LG) basis and the first mutually unbiased basis (MUB), and then perform measurements in both bases using the technique discussed in the Methods section. The purpose of this experiment is to perform an unfolded, classical version of our two-photon entanglement setup. Also referred to as the Klyshko advanced-wave picture [183], this is equivalent to replacing the crystal with a mirror, propagating light back through one of the detectors, reflecting it at the crystal plane, and then propagating it back to the other detector (compare Fig. A.6 (b) with the setup figure in the main text). In this manner, we can show that we are able to generate and measure arbitrary complex amplitudes, and that our measured bases are indeed mutually unbiased with respect to each other.

As shown in Fig. A.6 (b), modes in seven-dimensional LG and MUB bases are generated using computer generated holograms (CGH) implemented on the SLM labelled (g). Intensity images of these modes obtained on a CCD camera are shown in Fig. A.6 (a). The modes generated by SLM (g) are imaged onto SLM (m) by a 4f system of lenses (l3, 400mm) through a pinhole to pick off the first diffraction order of the SLM and remove zero-order diffraction noise. The pinhole is also where the crystal plane would be in the Klyshko picture (dotted rectangle). A measurement of a particular mode is performed by the spatial-mode filter implemented by SLM (m),

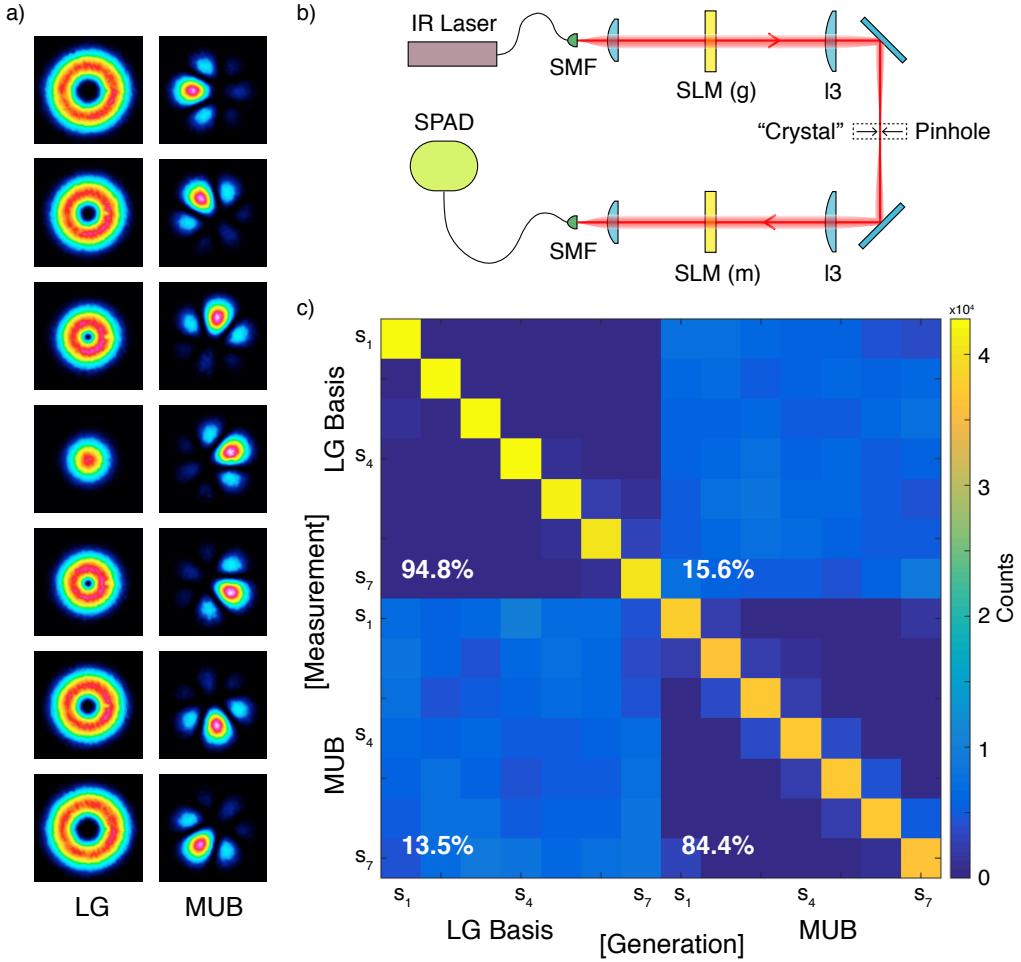


Figure A.6: **Classical prepare-and-measure experiment: LG basis.** a) CCD images of the 7-dimensional Laguerre-Gaussian (LG) basis and first mutually unbiased basis (MUB) modes. b) The experiment consists of a strongly attenuated IR laser incident on a spatial light modulator (SLM (g)) used for generating arbitrary spatial modes. SLM (g) is imaged onto SLM (m), which displays measurement holograms for arbitrary spatial modes. A pinhole is used to remove zero-order diffraction noise from SLM (g), and is also located at the ‘‘crystal’’ plane in the unfolded Klyshko picture [183]. The light from SLM (m) is coupled into a single-mode fiber (SMF), which is connected to a single-photon avalanche diode (SPAD). c) Experimental data showing measured counts when states are prepared and measured in both bases. The data are strongly correlated when the preparation and measurement bases are the same, and completely uncorrelated when they are not.

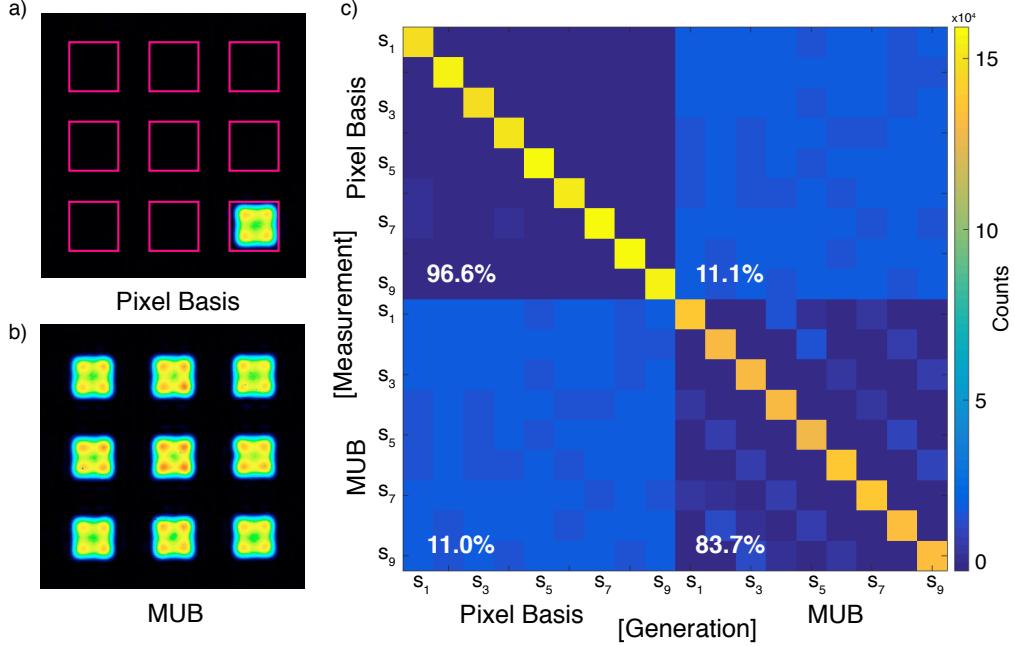


Figure A.7: **Classical prepare-and-measure experiment: 9-dimensional pixel basis.** CCD images of a) the first Pixel basis mode, and b) first mutually unbiased basis (MUB) mode. c) Experimental data showing measured counts when states are prepared and measured in both bases. The data are strongly correlated when the preparation and measurement bases are the same, and completely uncorrelated when they are not.

a single-mode fiber (SMF), and a single-photon avalanche photodiode (SPAD). The measurement holograms on SLM (m) are scanned through modes in both LG and MUB bases to obtain a  $14 \times 14$  element matrix of counts shown in Fig. A.6 (c). The counts are normalised such that the total counts measured across one basis are equal for each generated mode. As can be clearly seen, when modes in the same basis are generated and measured, a strong diagonal matrix is obtained, with a visibility of 94.8% (LG) and 84.4% (MUB) — defined as the sum of diagonal counts divided by total counts. The visibility in the LG basis is lower than the near-unity theoretical value due to imperfect alignment. The MUB visibility is further lowered due to errors introduced by the CGH method for approximating a more complex scalar field with a phase-only hologram, which is confirmed by simulation. When the generation and measurement bases are different, the data sets are seen to be mutually unbiased (flat), with a visibility approaching  $1/7=14.3\%$  in both cases (15.6% and 13.5%).

### A.8.1 Examples of MUBs in other experimental degrees-of-freedom/platforms

The purpose of this section is to show that our entanglement certification technique can be readily applied to other photonic degrees-of-freedom (DOFs), as well as to other physical platforms such as atoms. We do this by first demonstrating a second set of measurement bases with our classical prepare-and-measure experiment: the Pixel basis. Then, we discuss recent experimental examples of mutually unbiased measurements in the time-frequency and path degrees-of-freedom. Finally, building on recent experiment results, we show how such measurements are also feasible in high-dimensional atomic systems consisting of Cesium atoms.

First, we use the classical prepare-and-measure experiment discussed in Sec. A.8 to demonstrate a second set of mutually unbiased bases for the photonic position-momentum DOF. As shown in Fig. A.7 (a), the Pixel basis is composed of nine position states, defined by nine discrete macro-pixels. The figure shows the intensity profile of the first Pixel basis state recorded on a CCD, with the eight empty boxes indicating the positions of the remaining Pixel basis states. Fig. A.7 (b) shows the intensity profile of the first state from the first mutually unbiased basis (MUB) to the Pixel basis, constructed according to the standard method discussed in Ref. [98]. Using the setup from Fig. A.6 (b), every state in the Pixel and MUB bases is generated using SLM (g) and imaged onto the measurement SLM (m). The measurement SLM (m) is used to display measurement holograms in both bases, resulting in an  $18 \times 18$  element matrix of counts shown in Fig. A.7 (c). As can be clearly seen, when modes in the same basis are generated and measured, a strong diagonal matrix is obtained, with a visibility of 96.6% (Pixel) and 83.7% (MUB). As for the LG basis, the MUB visibility is slightly lower than the Pixel basis due to errors introduced by the CGH. When the generation and measurement bases are different, the data are again seen to be mutually unbiased (flat), with a visibility approaching  $1/9=11.1\%$  in both cases (11.1% and 11.0%).

Despite the significant difficulties in the implementation of arbitrary measurements on high-dimensional quantum systems, measurements in specific bases (such as MUBs) are quite common, with recent advances allowing for this in several experimental platforms. Here we briefly discuss how mutually unbiased and tilted basis measurements can be implemented in these platforms, allowing our entanglement certification technique to be directly applied in a wide range of future experiments. While we have demonstrated precise control and measurement over photonic spatial modes, recent experiments have been performed that show similar capabilities for other high-dimensional DOFs such as time-frequency and path.

For example, the experiment of Kues et al. [39] demonstrates on-chip, high-dimensional frequency-mode entanglement via spontaneous four-wave mixing in a micro-ring resonator. In order to measure their entangled state, the authors use a combination of two programmable phase filters and an electro-optic phase modulator to perform projective measurements corresponding to state vectors of the form  $|\psi_{\text{proj}}\rangle = \sum_{k=0}^{d-1} \alpha_k e^{i\phi_k} |\bar{k} + k\rangle$  where the projection amplitudes  $\alpha_k$  and the phases  $\phi_k$  can be chosen arbitrarily for a given frequency mode  $\bar{k}$ . This is precisely the type of transformation that would be required for a measurement in an arbitrary tilted or mutually unbiased basis of frequency modes, allowing the direct application of our method to this platform.

Another recent experiment by Karpiński et al. [184] used an electro-optic modulator to carry out a temporal Fourier transform of heralded single-photon pulses, while preserving their quantum coherence. This “time lens” applies the exact transformation required to measure temporal pulse-mode-entangled states in the mutually unbiased frequency basis. In the recent experiment by Carolan et al. [185], the authors demonstrate exquisite control over a rapidly reprogrammable 6-mode integrated photonic circuit, implementing Haar random unitaries with an extremely high fidelity. Combined with multi-outcome measurements at the end of the linear circuit, their system can readily be used to perform measurements in a six-dimensional mutually unbiased basis of path modes.

In the recent experiment by Anderson et al. [89], the electron and nuclear spins of individual  $^{133}\text{Cs}$  atoms were used as a test bed for implementing high-dimensional unitary transformations on an atomic system. Radio frequency and microwave magnetic fields were used to generate control Hamiltonians with excellent performance even in the presence of static and dynamic perturbations, allowing the implementation of unitary maps in a 16-dimensional Hilbert space with fidelities greater than 0.98. This was followed by a Stern-Gerlach measurement apparatus that measured the population in the 16-dimensional Hilbert space. Together, these unitary operations and multi-outcome measurements are precisely what is required to measure in a mutually unbiased basis of electron and nuclear  $^{133}\text{Cs}$  atoms spins.

Finally, one may note that multi-qubit systems, such as have been realized in photonic systems [186–189], superconducting qubits [190], or trapped ions [191], can also yield subsystems with high local dimension for suitable bipartitions of groups of multiple qubits. Such platforms are often composed of individually controllable qubits, e.g., for quantum computation or simulation [191], and usually permit arbitrary (projective) single-qubit measurements, and hence allow measurements w.r.t. mutually unbiased or tilted bases for any bipartition. For instance, measurements

### Systematic error due to imperfect measurements

$d$	$\tilde{F}(\rho, \Phi^+)$	$\tilde{F}_{s1}(\rho, \Phi^+)$	$\tilde{F}_{s2}(\rho, \Phi^+)$
3	$91.5 \pm 0.4\%$	98.0%	95.6%
5	$89.9 \pm 0.4\%$	96.4%	92.3%
7	$84.2 \pm 0.5\%$	94.6%	87.6%
11	$74.8 \pm 0.4\%$	89.7%	80.6%

Table A.1:  $\tilde{F}(\rho, \Phi^+)$  and  $\tilde{F}_{s1/2}(\rho, \Phi^+)$  are experimental and simulated fidelities to the maximally entangled state obtained via measurements in two MUBs in dimension  $d$ , respectively.  $\tilde{F}_{s1}(\rho, \Phi^+)$  is obtained by incorporating the effects of imperfect hologram measurements on an ideal state estimated from measurements in the LG basis.  $\tilde{F}_{s2}(\rho, \Phi^+)$  is obtained by additionally taking into account the misalignment-induced crosstalk measured in the LG basis.

w.r.t. the local Pauli  $Z$  and  $X$  operators for all qubits would be a simple realization of a MUB measurement. Our methods are thus also applicable to such systems.

The above examples clearly demonstrate the wide applicability of our entanglement certification technique to a variety of physical systems, and highlights its potential for significantly impacting future experiments on high-dimensional entanglement in photonic and atomic platforms, and beyond.

#### A.8.2 Systematic errors

While there are no assumptions made about the state or how it is produced, the method introduced here intrinsically puts trust on the measurement devices to work correctly. Hence, a crucial part of the experiment is an in-depth characterization of the measurement method. While from a physical point of view, one would expect the crystal to predominantly produce perfectly correlated pairs due to a conservation of angular momentum, the real data features a significant amount of cross-talk and noise, ultimately diminishing the certified entanglement and dimensionality. On the other hand, non-perfect unbiasedness of the observables could even lead to classically correlated photons to appear entangled, the most extreme case being a measurement in two identical bases that while assumed to be unbiased, are actually the same. Furthermore, the coincidence counts in different settings may not correspond to the density matrix elements in the way assumed if the detector efficiency is different for the different bases, which could lead to either over- or under-estimation of correlations (and with it entanglement). These are all potential systematic errors that we want to address in this final section.

While the predominant source of crosstalk is due to imperfect alignment, our paradigm of state-independence also includes the notion of reference frames (i.e. we do not assume to have a perfect common reference frame) and this misalignment can only decrease observed correlations. In other words, alignment issues are essentially captured by local unitary operations and can never lead to an increase of correlations where there are none.

Upon inspecting the correspondence of coincidence counts to density matrix elements we noticed a significant impact of mode-dependent loss. The usual assumption that coincidence counts  $C_{ij}$  of  $N$  photon pairs per unit of time in basis elements  $i$  and  $j$  respectively are related to density matrix elements via

$$C_{ij} = N \langle ij | \rho | ij \rangle, \quad (\text{A.58})$$

implicitly assumes (1) a constant photon pair production rate and (2) a universal coupling efficiency that is independent of  $i$  and  $j$ . While the measured pair production rate fluctuations are low enough for that estimation to be valid, we actually do expect a strong mode-dependent loss. In the LG-basis we expect from theoretical computations that higher modes have a lower coupling efficiency in the single-mode fibers [192], which should lead to a systematic suppression of higher-mode coincidence counts and with it a systematic under-estimation of entanglement. The exact coupling efficiency, however, depends on many intricate details of the physical setup and any theoretical computation could increase systematic errors in unpredictable ways. In this section we thus introduce a general method that corrects for mode-dependent loss using only the singles and coincidences in the setup and will find application also in many other quantum optical setups. Denoting the singles per unit time in detector  $A/B$  as  $S_i^{A/B}$ , as well as the mode dependent loss factors as  $\eta_i^{A/B}$ , we conclude that:

$$C_{ij} = N \langle ij | \rho | ij \rangle \eta_i^A \eta_j^B \quad (\text{A.59})$$

as well as

$$S_i^{A/B} = N \langle i | \rho_{A/B} | i \rangle \eta_i^{A/B} \quad (\text{A.60})$$

Now if we define

$$M_{ij} := \frac{C_{ij}}{S_i^A S_j^B} = \frac{1}{N} \frac{\langle ij | \rho | ij \rangle}{\langle i | \rho_A | i \rangle \langle j | \rho_B | j \rangle} \quad (\text{A.61})$$

we can use the fact that

$$\sum_j M_{ij} \langle j | \rho_B | j \rangle = [M \vec{\rho}_B]_i = \frac{1}{N} \frac{\sum_j \langle ij | \rho | ij \rangle}{\langle i | \rho_A | i \rangle} = \frac{1}{N} \quad (\text{A.62})$$

This allows us to get  $N$ , as well as conclude that

$$\langle i | \rho_B | i \rangle = \sum_j (M)_{ij}^{-1} \frac{1}{N} \quad (\text{A.63})$$

Now all that is left is to insert this into the definition of  $M_{ij}$  to get

$$\langle ij | \rho | ij \rangle = \frac{M_{ij} (\sum_j (M)_{ij}^{-1}) (\sum_i (M^T)_{ij}^{-1})}{(\sum_i \sum_j (M)_{ij}^{-1})} \quad (\text{A.64})$$

The only assumptions remaining in this correction method are a constant pair production rate and that the majority of singles is generated by photon pairs. These assumptions can also be verified using the experimental data by checking that  $[M\vec{\rho}_B]_i^{-1} = N$  is indeed equally true for all  $i$ . Using this correction method we indeed find the expected effect: higher order modes in LG basis were significantly suppressed leading to artificially reduced coincidence counts. We account for this mode-dependent loss in our data, allowing us to more accurately estimate a target state and hence construct a more optimum tilted basis.

A second source of systematic error is the effect of imperfect measurements on the resulting fidelity bounds. As shown in Sec. A.8, the classical (one-photon) measurements made using our computer-generated holograms (CGHs) in the LG and the MUB bases are not perfect, with the MUB basis showing a lower visibility than the LG basis. In the two-photon experiment, this would manifest as additional counts appearing in the off-diagonals of the data matrices shown in the main text, which would in turn lower the measured fidelity bounds. In order to estimate this quantitatively, we proceed as follows.

First, we calculate the ideal state as obtained from diagonal measurements in the LG basis, by setting the off-diagonal (crosstalk) counts to zero and calculating the resulting density matrix. Second, we use this state to calculate the ideal experimental data one would obtain if measuring in the first MUB. Next, we simulate the imperfect measurements in MATLAB for each input state and hologram by multiplying the complex field amplitude by the hologram amplitude, and then calculating its overlap with a Gaussian fiber mode amplitude. The resulting probability matrices for the LG and MUB bases capture the resulting imperfections of the CGH measurement process. This process is repeated for each dimension considered in our experiment. We find that the imperfections in the LG measurement are almost negligible, while the visibility in the MUB drops as a function of dimension. We then calculate the effect of these hologram imperfections on the ideal two-photon experimental data calculated above.

A key factor that results in a lowering of the measured fidelity bound in experiment is the crosstalk due to imperfect alignment. We incorporate this crosstalk into our

fidelity calculation by using the LG basis data obtained in experiment, and the MUB data obtained from the above simulation. In this manner, both the effects of imperfect measurement and misalignments are captured in our systematic error-corrected fidelity bounds. Table A.1 lists the measured fidelities  $\tilde{F}(\rho, \Phi^+)$  from experiment, the simulated fidelities  $\tilde{F}_{s1}(\rho, \Phi^+)$  taking into account the effects of imperfect hologram measurements, and simulated fidelities  $\tilde{F}_{s2}(\rho, \Phi^+)$  additionally incorporating the effects of misalignment-induced crosstalk only in the LG basis (taken from the measured data). The effects of crosstalk on the MUB measurements cannot be added in independently of the simulated systematic error, but one can expect that it will lower the fidelities even more, ideally approaching the measured values  $\tilde{F}(\rho, \Phi^+)$ . Thus, imperfect measurements are always seen to result in an under-estimation of correlations, thus lowering our fidelities from their ideal expected values.



# B

## APPENDIX

### Supplemental Information of Chapter 2

In these appendices we provide material to complement the main text. In appendix B.1, we present the proofs of theorems 2.2 and 2.4, regarding behaviours and assemblages that cannot be expressed in terms of process matrices. In appendix B.2, we present the proofs of lemmas 2.1 and 2.2, regarding the realization of causal behaviours and assemblages in terms of causally separable process matrices. In appendix B.3, we show how to obtain the characterization of general and causal assemblages presented in definitions 2.5 and 2.6. In appendix B.4, we present the proof of theorem 2.6, showing a class of causally nonseparable process matrices that cannot be certified in a semi-device-independent way. In appendix B.5, we re-derive all concepts and results of the main text concerning certification of bipartite process matrices for the tripartite process matrices whose third party is always in the future of the other two. We start by defining all notions of certification for these tripartite process matrices and then explore each scenario (TTT, UUU, TTU, TUU, UTT, and UUT) in detail. This appendix contains technical results not used nor mentioned in the main text. In appendix B.6, we present the proof of theorem 2.7, i.e., we prove that the quantum switch cannot be certified to be causally nonseparable in the UUT scenario. Finally, in appendix B.7, we present our theoretical analysis of the quantum switch experiments of refs. [72] and [74].

## B.1 Behaviours and assemblages unattainable by process matrices

In this appendix, we prove theorem 2.2 and theorem 2.4, concerning behaviours and assemblages which cannot be obtained by process matrices. We start by presenting and proving the following lemma, which will be necessary for the proof of theorem 2.2.

**Lemma B.3.** *Let  $A \in \mathcal{L}(\mathcal{H}^1 \otimes \mathcal{H}^2)$  be a positive semidefinite operator. It holds that*

$$d \text{Tr}_2(A) \otimes \mathbb{1} - A \geq 0. \quad (\text{B.1})$$

where  $d = \min\{d_1, d_2\}$  and  $\mathbb{1}$  denotes the identity operator acting on  $\mathcal{H}^2$ .

*Proof.* Let us start with the case where  $A$  is a rank-1 operator, so that  $A = |\lambda\rangle\langle\lambda|$  for some unnormalized vector  $|\lambda\rangle$ . The Schmidt decomposition ensures that every bipartite vector  $|\lambda\rangle \in (\mathcal{H}^1 \otimes \mathcal{H}^2)$  can be written as  $|\lambda\rangle = \sum_{i=1}^d \lambda_i |ii\rangle$  for some real non-negative coefficients  $\{\lambda_i\}_{i=1}^d$  and  $d = \min\{d_1, d_2\}$ . We can then define a diagonal operator  $D : \mathcal{H}^1 \rightarrow \mathcal{H}^1$ ,  $D := \sqrt{d} \sum_{i=1}^d \lambda_i |i\rangle\langle i|$  such that  $|\lambda\rangle = D \otimes \mathbb{1} |\phi_d^+\rangle$  where  $|\phi_d^+\rangle := \sum_{i=1}^d \frac{1}{\sqrt{d}} |ii\rangle$  is a  $d$ -dimensional maximally entangled state acting on  $\mathcal{H}^1 \otimes \mathcal{H}^2$ .

Using the diagonal operator  $D$ , and  $D^\dagger = D$ , the partial trace  $\text{Tr}_2(A)$  can be written as

$$\text{Tr}_2(A) = \text{Tr}_2(D \otimes \mathbb{1} |\phi_d^+\rangle\langle\phi_d^+| D \otimes \mathbb{1}) \quad (\text{B.2})$$

$$= \frac{DD}{d}. \quad (\text{B.3})$$

Direct calculation of the left-hand side of inequality (B.1) leads to

$$d \text{Tr}_2(A) \otimes \mathbb{1} - A = d \frac{DD \otimes \mathbb{1}}{d} - D \otimes \mathbb{1} |\phi_d^+\rangle\langle\phi_d^+| D \otimes \mathbb{1} \quad (\text{B.4})$$

$$= D \otimes \mathbb{1} (\mathbb{1} \otimes \mathbb{1} - |\phi_d^+\rangle\langle\phi_d^+|) D \otimes \mathbb{1} \quad (\text{B.5})$$

$$\geq 0, \quad (\text{B.6})$$

where the last inequality holds since  $\mathbb{1} \otimes \mathbb{1} - |\phi_d^+\rangle\langle\phi_d^+| \geq 0$  and  $D \geq 0$ .

To prove the general case, note that we can write  $A$  as a sum of rank-1 operators,  $A = \sum_i |\lambda^{(i)}\rangle\langle\lambda^{(i)}|$ . Since for every  $i$  it holds that

$$d \text{Tr}_2(|\lambda^{(i)}\rangle\langle\lambda^{(i)}|) \otimes \mathbb{1} - |\lambda^{(i)}\rangle\langle\lambda^{(i)}| \geq 0, \quad (\text{B.7})$$

we also have that

$$d \text{Tr}_2 \left( \sum_i |\lambda^{(i)}\rangle\langle\lambda^{(i)}| \right) \otimes \mathbb{1} - \sum_i |\lambda^{(i)}\rangle\langle\lambda^{(i)}| \geq 0, \quad (\text{B.8})$$

and hence,

$$d \operatorname{Tr}_2(A) \otimes \mathbb{1} - A \geq 0. \quad (\text{B.9})$$

□

**Theorem 2.2.** *All process behaviours are valid behaviours, however, not all valid behaviours are process behaviours.*

*In particular, in the scenario where all parties have dichotomic inputs and outputs, any behaviour  $\{p(ab|xy)\}$  such that*

$$\frac{1}{4} \sum_{a,b,x,y} \delta_{a,y} \delta_{b,x} p(ab|xy) > 1 - \frac{1}{d+1}$$

*is not a process behaviour for process matrices with total dimension  $d_{A_I} d_{A_O} d_{B_I} d_{B_O} = d$ .*

*Proof.* It follows by definition that all process behaviours are valid behaviours, we now show that there exist valid behaviours that are not process behaviours for any finite dimension. Assume that there exists a process matrix  $W$  with total dimension  $d_{A_I} d_{A_O} d_{B_I} d_{B_O} = d$  and instruments  $\{A_{a|x}\}, \{B_{b|y}\}$  such that

$$p(ab|xy) = \operatorname{Tr}[W(A_{a|x} \otimes B_{b|y})] \quad (\text{B.10})$$

and

$$p_{\text{GYN}}^{\text{succ}} = \frac{1}{4} \sum_{abxy} \delta_{ay} \delta_{bx} p(ab|xy) > 1 - \frac{1}{d+1}, \quad (\text{B.11})$$

where  $\delta_{ij}$  is the Kronecker's delta function and  $p_{\text{GYN}}^{\text{succ}}$  is the probability of success achieved by this behaviour in the GYNI causal game (defined in ref. [69]).

Let then we define the GYNI operator

$$M := \frac{1}{4} \sum_{abxy} \delta_{ay} \delta_{bx} A_{a|x} \otimes B_{b|y}, \quad (\text{B.12})$$

so that

$$p_{\text{GYN}}^{\text{succ}} = \operatorname{Tr}(WM). \quad (\text{B.13})$$

Since  $W$  is a valid process matrix, it admits the decomposition

$$W = {}_{A_O}W + {}_{B_O}W - {}_{A_O}{}_{B_O}W, \quad (\text{B.14})$$

and by linearity we have

$$p_{\text{GYN}}^{\text{succ}} = \operatorname{Tr}({}_{A_O}WM) + \operatorname{Tr}({}_{B_O}WM) - \operatorname{Tr}({}_{A_O}{}_{B_O}WM). \quad (\text{B.15})$$

Since  ${}_{AO}W$  and  ${}_{BO}W$  are causally ordered process matrices, they cannot violate the GYNI causal inequality and respect [69]

$$\mathrm{Tr}(MW^{\mathrm{sep}}) \leq \frac{1}{2}. \quad (\text{B.16})$$

Also, it follows from lemma B.3 that

$$-\mathrm{Tr}({}_{AO}{}_{BO}WM) \leq -\frac{1}{d}\mathrm{Tr}(WM). \quad (\text{B.17})$$

We than have

$$p_{\text{GYNI}}^{\text{succ}} = \mathrm{Tr}(WM) = \mathrm{Tr}({}_{AO}WM) + \mathrm{Tr}({}_{BO}WM) - \mathrm{Tr}({}_{AO}{}_{BO}WM) \quad (\text{B.18})$$

$$\leq \frac{1}{2} + \frac{1}{2} - \frac{1}{d}\mathrm{Tr}(WM), \quad (\text{B.19})$$

which implies

$$p_{\text{GYNI}}^{\text{succ}} = \mathrm{Tr}(WM) \leq 1 - \frac{1}{d+1}, \quad (\text{B.20})$$

and bounds the maximal attainable value for process behaviours on the GYNI causal game, which can attain  $p_{\text{GYNI}}^{\text{succ}} = 1$  for general behaviours.

Note that this robust bound only holds for finite dimensional process matrices. In ref. [130], the authors have shown that a process matrix  $W$  cannot attain the *exact* maximal success probability in the GYNI causal game *i.e.*,  $p_{\text{GYNI}}^{\text{succ}} = \mathrm{Tr}(WM) = 1$ , even with infinite dimension.

□

**Theorem 2.4.** *All process assemblages are valid assemblages, however, not all valid assemblages are process assemblages.*

*In particular, in the scenario where Alice has dichotomic inputs and outputs, the general assemblage  $\{w_{a|x}\}$  given by  $w_{a|x} = |x\rangle\langle x| \otimes |a\rangle\langle a|$  is not a process assemblage.*

*Proof.* First we see that  $w_{a|x} = |x\rangle\langle x| \otimes |a\rangle\langle a|$  is a valid assemblage, since  $|x\rangle\langle x| \otimes |a\rangle\langle a| \geq 0$  for all  $a, x$ , and

$$\sum_a w_{a|x} = |x\rangle\langle x| \otimes \mathbb{1} \quad \forall x. \quad (\text{B.21})$$

Assume then that  $\{w_{a|x}\}$  is a process assemblage. Then, there must exist a process matrix  $W$  and instruments  $\{A_{a|x}\}$  such that

$$w_{a|x} = \mathrm{Tr}_{A_I A_O} [(A_{a|x} \otimes \mathbb{1}) W], \quad (\text{B.22})$$

and that for any set of instruments  $\{B_{b|y}\}$ ,

$$p(ab|xy) = \mathrm{Tr}(B_{b|y} w_{a|x}) = \mathrm{Tr} [(A_{a|x} \otimes B_{b|y}) W]. \quad (\text{B.23})$$

Consider now the valid set of instruments given by

$$B_{b|y} = |b\rangle\langle b| \otimes |y\rangle\langle y|. \quad (\text{B.24})$$

Behaviours obtained by this set of instruments and the assemblage  $\{w_{a|x}\}$  would be

$$p(ab|xy) = \text{Tr}(B_{b|y} w_{a|x}) \quad (\text{B.25})$$

$$= \delta_{b,x} \delta_{a,y} \quad \forall a, b, x, y, \quad (\text{B.26})$$

which according to theorem 2.2 cannot be obtained by any finite dimensional process matrix, contradicting the hypothesis that  $\{w_{a|x}\}$  is a process assemblage.  $\square$

## B.2 Behaviours and assemblages attainable by causally separable process matrices

In this appendix we prove lemma 2.1 and lemma 2.2, which we restate for the convenience of the reader.

**Lemma 2.1.** *A general behaviour is causal if and only if it is a process behaviour that can be obtained by a causally separable process matrix.*

*Proof.* To prove the if part we show that causally separable process matrices can only give rise to causal behaviours, regardless of the instruments performed on them.

We start by showing that a behaviour  $\{p(ab|xy)\}$  that comes from acting with any sets of instruments  $\{A_{a|x}\}$  and  $\{B_{b|y}\}$  on a process matrix that is causally ordered from Alice to Bob  $W^{A \prec B}$  is also causally ordered from Alice to Bob. Given  $\{p(ab|xy)\}$  that arises from  $W^{A \prec B}$  according to

$$p(ab|xy) = \text{Tr}[(A_{a|x} \otimes B_{b|y}) W^{A \prec B}] \quad (\text{B.27})$$

$$= \text{Tr}\left[(A_{a|x}^{A_I A_O} \otimes B_{b|y}^{B_I B_O}) W^{A_I A_O B_I} \otimes \frac{\mathbb{1}^{B_O}}{d_{B_O}}\right], \quad (\text{B.28})$$

one can check that Alice's marginal probability distributions,

$$\sum_b p(ab|xy) = \sum_b \text{Tr}\left[(A_{a|x}^{A_I A_O} \otimes B_{b|y}^{B_I B_O}) W^{A_I A_O B_I} \otimes \frac{\mathbb{1}^{B_O}}{d_{B_O}}\right] \quad (\text{B.29})$$

$$= \text{Tr}\left[(A_{a|x}^{A_I A_O} \otimes \mathbb{1}^{B_I})(W^{A_I A_O B_I}) \text{Tr}_{B_O}(\mathbb{1}^{A_I A_O} \otimes \frac{1}{d_{B_O}} \sum_b B_{b|y}^{B_I B_O})\right] \quad (\text{B.30})$$

$$= \frac{1}{d_{B_O}} \text{Tr}\left[(A_{a|x}^{A_I A_O} \otimes \mathbb{1}^{B_I}) W^{A_I A_O B_I}\right], \quad (\text{B.31})$$

are independent of  $y$  for all  $a, x$ . Hence,  $p(ab|xy) = p^{A \prec B}(ab|xy)$  is causally ordered from Alice to Bob. Equivalently,  $W^{B \prec A}$  implies  $\{p^{B \prec A}(ab|xy)\}$ .

Hence, behaviours that come from a causally separable process matrix according to

$$p(ab|xy) = \text{Tr} [(A_{a|x} \otimes B_{b|y}) W^{\text{causal}}] \quad (\text{B.32})$$

$$= \text{Tr} [(A_{a|x} \otimes B_{b|y}) (qW^{A \prec B} + (1-q)W^{B \prec A})] \quad (\text{B.33})$$

$$= qp^{A \prec B}(ab|xy) + (1-q)p^{B \prec A}(ab|xy) \quad (\text{B.34})$$

are causal by definition. In other words, causally separable process matrices can only generate causal behaviours.

To prove the only if part we show that all behaviours that are causal can be reproduced by performing some instruments on some causally separable process matrix.

Given a causal behaviour  $\{p^{\text{causal}}(ab|xy)\}$ , according to Bayes' rule and the non-signalling principle, one can decompose it in the following form:

$$p^{\text{causal}}(ab|xy) = q p^{A \prec B}(ab|xy) + (1-q)p^{B \prec A}(ab|xy) \quad (\text{B.35})$$

$$= q p_A^{A \prec B}(a|x)p_B^{A \prec B}(b|axy) + (1-q)p_B^{B \prec A}(b|y)p_A^{B \prec A}(a|bxy), \quad (\text{B.36})$$

so that  $\{p_A^{A \prec B}(a|x)\}$ ,  $\{p_B^{A \prec B}(b|axy)\}$ ,  $\{p_B^{B \prec A}(b|y)\}$ ,  $\{p_A^{B \prec A}(a|bxy)\}$ , and  $q$  are given quantities.

First, for the contribution that is causally ordered from Alice to Bob,  $p^{A \prec B}(ab|xy)$ , we construct instruments  $\{A_{a|x}^{A \prec B}\}$  and  $\{B_{b|y}^{A \prec B}\}$  according to

$$A_{a|x}^{A \prec B} = \mathbb{1}^{A_I} \otimes p_A^{A \prec B}(a|x)|ax\rangle\langle ax|^{A_O} \quad (\text{B.37})$$

$$B_{b|y}^{A \prec B} = \sum_{a,x} p_B^{A \prec B}(b|axy)|ax\rangle\langle ax|^{B_I} \otimes \frac{\mathbb{1}^{B_O}}{d_{B_O}}, \quad (\text{B.38})$$

and process matrix

$$W^{A \prec B} = \frac{\mathbb{1}^{A_I}}{d_{A_I}} \otimes |\Phi^+\rangle\langle\Phi^+|^{A_O B_I} \otimes \mathbb{1}^{B_O}, \quad (\text{B.39})$$

where  $|\Phi^+\rangle\langle\Phi^+|$  is the Choi operator of the identity channel, with  $|\Phi^+\rangle = \sum_{i=1}^d |ii\rangle$ . It can be checked that

$$\text{Tr} [(A_{a|x}^{A \prec B} \otimes B_{b|y}^{A \prec B}) W^{A \prec B}] = p_A^{A \prec B}(a|x)p_B^{A \prec B}(b|axy) \quad (\text{B.40})$$

$$= p^{A \prec B}(ab|xy), \quad (\text{B.41})$$

for every  $a, b, x, y$ . Hence, such construction can recover any behaviour that is causally ordered from Alice to Bob. Analogously, for  $\{p^{B \prec A}(ab|xy)\}$ , one has

$$A_{a|x}^{B \prec A} = \sum_{b,y} p_A^{B \prec A}(a|bxy)|by\rangle\langle by|^{A_I} \otimes \frac{\mathbb{1}^{A_O}}{d_{A_O}}, \quad (\text{B.42})$$

$$B_{b|y}^{B \prec A} = \mathbb{1}^{B_I} \otimes p_B^{B \prec A}(b|y)|by\rangle\langle by|^{B_O} \quad (\text{B.43})$$

$$W^{B \prec A} = \frac{\mathbb{1}^{B_I}}{d_{B_I}} \otimes |\Phi^+\rangle\langle\Phi^+|^{B_O A_I} \otimes \mathbb{1}^{A_O}. \quad (\text{B.44})$$

Now, for causal behaviours, which are convex combinations of causally ordered behaviours, we construct instruments and process matrices in such a way that each causal order acts on a complementary subspace. That is, we define the valid process matrix

$$W^{\text{causal}} = q W^{A \prec B} \otimes |00\rangle\langle 00|^{A'_I B'_I} + (1 - q) W^{B \prec A} \otimes |11\rangle\langle 11|^{A'_I B'_I}, \quad (\text{B.45})$$

by extending Alice's and Bob's input spaces according to  $\mathcal{H}^{A_I(B_I)} \rightarrow \mathcal{H}^{A_I(B_I)} \otimes \mathcal{H}^{A'_I(B'_I)}$ , and we define the valid instruments

$$A_{a|x} = A_{a|x}^{A \prec B} \otimes |0\rangle\langle 0|^{A'_I} + A_{a|x}^{B \prec A} \otimes |1\rangle\langle 1|^{A'_I} \quad (\text{B.46})$$

$$B_{b|y} = B_{b|y}^{A \prec B} \otimes |0\rangle\langle 0|^{B'_I} + B_{b|y}^{B \prec A} \otimes |1\rangle\langle 1|^{B'_I}. \quad (\text{B.47})$$

In this way, it is easy to check that orthogonal terms cancel out and we arrive at

$$\text{Tr} [(A_{a|x} \otimes B_{b|y}) W^{\text{causal}}] = qp^{A \prec B}(ab|xy) + (1 - q)p^{B \prec A}(ab|xy) \quad (\text{B.48})$$

$$= p^{\text{causal}}(ab|xy), \quad (\text{B.49})$$

recovering any causal behaviour.  $\square$

**Lemma 2.2.** *A general assemblage is causal if and only if it is a process assemblage that can be obtained from a causally separable process matrix.*

*Proof.* To prove the if part we show that a causally separable process matrix can only give rise to a causal assemblage regardless of Alice's sets of instruments. First we show that an assemblage that is generated by a process matrix that is ordered from Alice to Bob is also ordered from Alice to Bob.

$$B_O w_{a|x} =_{B_O} \text{Tr}_A [(A_{a|x} \otimes \mathbb{1}) W^{A \prec B}] \quad (\text{B.50})$$

$$= \text{Tr}_A [(A_{a|x} \otimes \mathbb{1})_{B_O} W^{A \prec B}] \quad (\text{B.51})$$

$$= \text{Tr}_A [(A_{a|x} \otimes \mathbb{1}) W^{A \prec B}] \quad (\text{B.52})$$

$$= w_{a|x}, \quad (\text{B.53})$$

since  $W^{A \prec B} =_{BO} W^{A \prec B}$ . Hence,  $\{w_{a|x}\} = \{w_{a|x}^{A \prec B}\}$  is causally ordered from Alice to Bob.

Next, we show that an assemblage that is generated by a process matrix that is ordered from Bob to Alice is also ordered from Bob to Alice.

$$\sum_a w_{a|x} = \sum_a \text{Tr}_A [(A_{a|x} \otimes \mathbb{1}) W^{B \prec A}] \quad (\text{B.54})$$

$$= \text{Tr}_A \left[ \left( \sum_a A_{a|x} \otimes \mathbb{1} \right)_{AO} W^{B \prec A} \right] \quad (\text{B.55})$$

$$= \text{Tr}_{AI} \left[ W^{B_I B_O A_I} \left( \text{Tr}_{AO} \sum_a A_{a|x} \otimes \mathbb{1}^{B_I B_O} \right) \right] \quad (\text{B.56})$$

$$= \text{Tr}_{AI} [W^{B_I B_O A_I} (\mathbb{1}^{A_I} \otimes \mathbb{1}^{B_I B_O})], \quad (\text{B.57})$$

which is independent of  $x$ . Hence,  $\{w_{a|x}\} = \{w_{a|x}^{B \prec A}\}$  is causally ordered from Bob to Alice.

Consequently, assemblages that come from causally separable process matrices according to

$$w_{a|x} = \text{Tr}_A [(A_{a|x} \otimes \mathbb{1}) W^{\text{causal}}] \quad (\text{B.58})$$

$$= \text{Tr}_A [(A_{a|x} \otimes \mathbb{1}) (q W^{A \prec B} + (1 - q) W^{B \prec A})] \quad (\text{B.59})$$

$$= q w_{a|x}^{A \prec B} + (1 - q) w_{a|x}^{B \prec A} \quad (\text{B.60})$$

are causal by definition. In other words, causally separable process matrices can only generate causal assemblages.

To prove the only if part we show that all assemblages that are causal can be reproduced by performing some instruments on some causally separable process matrix. Given a causal assemblage, it can be decomposed in  $\{w_{a|x}^{A \prec B}\}$  and  $\{w_{a|x}^{B \prec A}\}$  with some convex weight  $q$ .

With the contribution that is causally ordered from Alice to Bob,  $\{w_{a|x}^{A \prec B}\}$ , one can write its elements as  $w_{a|x}^{A \prec B} = \sigma_{a|x}^{B_I} \otimes \mathbb{1}^{B_O}$ , where  $\text{Tr} \sum_a \sigma_{a|x} = 1$ . We show that it can be recovered by instruments

$$A_{a|x}^{A \prec B} = \mathbb{1}^{A_I} \otimes \sigma_{a|x}^{T A_O}, \quad (\text{B.61})$$

where  $T$  is the transposition in the computational basis of  $\mathcal{H}^{AO}$ , and a causally ordered process matrix from Alice to Bob

$$W^{A \prec B} = \frac{\mathbb{1}^{A_I}}{d_{A_I}} \otimes |\Phi^+\rangle\langle\Phi^+|^{A_O B_I} \otimes \mathbb{1}^{B_O}, \quad (\text{B.62})$$

where  $|\Phi^+\rangle\langle\Phi^+|$  is the Choi operator of the identity channel, with  $|\Phi^+\rangle = \sum_{i=1}^d |ii\rangle$ . Indeed,

$$\mathrm{Tr}_{A_I A_O} \left[ (A_{a|x}^{A \prec B} \otimes \mathbb{1}^{B_I B_O}) W^{A \prec B} \right] = \mathrm{Tr}_{A_O} \left[ (\sigma_{a|x}^{T A_O} \otimes \mathbb{1}^{B_I B_O}) (|\Phi^+\rangle\langle\Phi^+|^{A_O B_I} \otimes \mathbb{1}^{B_O}) \right] \quad (\text{B.63})$$

$$= \sigma_{a|x}^{B_I} \otimes \mathbb{1}^{B_O} \quad (\text{B.64})$$

$$= w_{a|x}^{A \prec B}. \quad (\text{B.65})$$

On the other hand, with the contribution that is causally ordered from Bob to Alice,  $\{w_{a|x}^{B \prec A}\}$ , one can write  $\sum_a w_{a|x}^{B \prec A} = \rho^{B_I} \otimes \mathbb{1}^{B_O}$ , where  $\mathrm{Tr}\rho^{B_I} = 1$ . Since  $\rho \geq 0$ , it can be written as  $\rho = \sum_i \mu_i |i\rangle\langle i|$ , which is purified by  $|\psi\rangle = \sum_i \sqrt{\mu_i} |ii\rangle$ . We show that  $\{w_{a|x}^{B \prec A}\}$  can be recovered by instruments  $\{A_{a|x}\}$ ,

$$A_{a|x}^{B \prec A} = (\rho^{-\frac{1}{2} A'_I} \otimes \mathbb{1}^{A''_I}) w_{a|x}^{T A_I} (\rho^{-\frac{1}{2} A'_I} \otimes \mathbb{1}^{A''_I}) \otimes \frac{\mathbb{1}^{A_O}}{d_{A_O}}, \quad (\text{B.66})$$

where  $\rho^{-1}$  is the inverse of  $\rho$  on its support,  $\mathcal{H}^{A_I} = \mathcal{H}^{A'_I} \otimes \mathcal{H}^{A''_I}$ ,  $T$  is the transposition in the  $\{|i\rangle\}_i$  basis, and a causally ordered process matrix from Bob to Alice

$$W^{B \prec A} = |\psi\rangle\langle\psi|^{B_I A'_I} \otimes |\Phi^+\rangle\langle\Phi^+|^{B_O A''_I} \otimes \mathbb{1}^{A_O}. \quad (\text{B.67})$$

Indeed,

$$\mathrm{Tr}_{A_I A_O} \left[ (A_{a|x}^{B \prec A} \otimes \mathbb{1}^{B_I B_O}) W^{B \prec A} \right] = w_{a|x}^{B \prec A}. \quad (\text{B.68})$$

Finally, just like in the proof of lemma 2.1, by allowing the different causal orders to act in complementary subspaces, we can recover any convex combinations of causally ordered assemblages, i.e., causal assemblages, from causally separable process matrices.  $\square$

### B.3 Characterization theorems for general and causal assemblages

In this appendix, we prove the equivalence between the definition induced by eq. (2.24) and definition 2.5, that is, we show that the most general set of operators  $\{w_{a|x}\}$  that satisfies

$$p(ab|xy) = \mathrm{Tr}(B_{b|y} w_{a|x}) \quad \forall a, b, x, y, \quad (\text{B.69})$$

where  $\{p(ab|xy)\}$  is a general behaviour and  $\{B_{b|y}\}$  is a valid set of instruments, is a general assemblage of definition 2.5.

Next, we prove this equivalence to hold also between the definition induced by eq. (2.29) and definition 2.6, that is, we show that the most general set of operators  $\{w_{a|x}^{A \prec B}\}$  that satisfies

$$p^{A \prec B}(ab|xy) = \text{Tr}(B_{b|y} w_{a|x}^{A \prec B}) \quad \forall a, b, x, y, \quad (\text{B.70})$$

where  $\{p^{A \prec B}(ab|xy)\}$  is a behaviour that is causally ordered from Alice to Bob and  $\{B_{b|y}\}$  is a valid set of instruments, is an assemblage that is causally ordered from Alice to Bob, of definition 2.6, end equivalently from Bob to Alice.

Let us begin with the conditions of a general assemblage.

Non-negativity yields

$$p(ab|xy) \geq 0 \quad \forall B_{b|y} \geq 0 \iff w_{a|x} \geq 0 \quad \forall a, x, \quad (\text{B.71})$$

while normalization yields

$$\sum_{a,b} p(ab|xy) = \sum_{a,b} \text{Tr}(B_{b|y} w_{a|x}) = 1. \quad (\text{B.72})$$

Let  $B_y = \sum_b B_{b|y}$  be the Choi operator of a CPTP map. Then, following ref. [46], we can parametrize it according to

$$B_y =_{[1-B_O]} Y + \frac{\mathbb{1}}{d_{B_O}}, \quad (\text{B.73})$$

where  $Y$  is an hermitian operator and we define the map  ${}_{[1-B_O]} M = M -_{B_O} M$ .

Applying the parametrization, it follows that

$$\sum_a \text{Tr} \left[ \left( {}_{[1-B_O]} Y + \frac{\mathbb{1}}{d_{B_O}} \right) w_{a|x} \right] = 1. \quad (\text{B.74})$$

If  $Y = 0$ , we have

$$\text{Tr} \left( \frac{1}{d_{B_O}} \sum_a w_{a|x} \right) = 1 \iff \text{Tr} \sum_a w_{a|x} = d_{B_O}. \quad (\text{B.75})$$

If  $Y \neq 0$ , and using eq. (B.75), we have

$$\text{Tr} \left( {}_{[1-B_O]} Y \sum_a w_{a|x} \right) = 0. \quad (\text{B.76})$$

Due to the self-duality of the map  ${}_{[1-B_O]} \cdot$ ,

$$\text{Tr} \left( {}_{[1-B_O]} Y \sum_a w_{a|x} \right) = \text{Tr} \left( Y {}_{[1-B_O]} \sum_a w_{a|x} \right) = 0 \quad \forall Y \iff {}_{[1-B_O]} \sum_a w_{a|x} = 0. \quad (\text{B.77})$$

Hence,

$$\sum_a w_{a|x} =_{BO} \sum_a w_{a|x}. \quad (\text{B.78})$$

Together, eqs. (B.71), (B.75) and (B.78) define a general assemblage. It is simple to verify that general assemblages lead to valid probability distributions for all sets of instruments.

For assemblages that are causally ordered from Alice to Bob, we require first that they satisfy the conditions of a general assemblage, in order to lead to valid behaviours, and second that Alice's marginal probability distributions

$$\sum_b p^{A \prec B}(ab|xy) = \text{Tr} \left( \sum_b B_{b|y} w_{a|x}^{A \prec B} \right) \quad (\text{B.79})$$

are independent of  $y$ , so that the resulting behaviour is causally ordered from Alice to Bob. The implication is that

$$\sum_b p^{A \prec B}(ab|xy) = \text{Tr} \left[ \left( [1-B_O]Y + \frac{\mathbb{1}}{d_{BO}} \right) w_{a|x}^{A \prec B} \right] \quad (\text{B.80})$$

$$= \text{Tr} \left( [1-B_O]Y w_{a|x}^{A \prec B} \right) + \frac{1}{d_{BO}} \text{Tr} \left( w_{a|x}^{A \prec B} \right). \quad (\text{B.81})$$

will be independent of  $y$  if and only if

$$\text{Tr} \left( [1-B_O]Y w_{a|x}^{A \prec B} \right) = \text{Tr} \left( Y [1-B_O]w_{a|x}^{A \prec B} \right) = 0 \quad \forall Y \iff [1-B_O]w_{a|x}^{A \prec B} = 0. \quad (\text{B.82})$$

Hence,

$$w_{a|x}^{A \prec B} =_{BO} w_{a|x}^{A \prec B} \quad \forall a, x. \quad (\text{B.83})$$

Finally, for an assemblage that is causally ordered from Bob to Alice, it is also required that they satisfy the conditions of a general assemblage. To guarantee that Bob's marginal probability distributions

$$\sum_a p^{A \prec B}(ab|xy) = \text{Tr} \left( B_{b|y} \sum_a w_{a|x}^{A \prec B} \right) \quad (\text{B.84})$$

are independent of  $x$ , it is necessary and sufficient that

$$\sum_a w_{a|x}^{B \prec A} = \sum_a w_{a|x'}^{B \prec A} \quad \forall b, x, x', y. \quad (\text{B.85})$$

Together, eqs. (B.83) and (B.85) and the conditions of general assemblages define causally ordered assemblages. It is simple to verify that causal assemblages lead to causal probability distributions for all sets of instruments.

The technique used here to arrive at the definitions of general and causal assemblages is the same that will be used in all tripartite scenarios to define their respective general and causal assemblages.

## B.4 Causally nonseparable process matrices that cannot be certified in a semi-device-independent way

In this section we present the proof of theorem 2.6 from the main text.

**Theorem 2.6** (Device-dependent certifiable, semi-device-independent noncertifiable process matrix). *There exist causally nonseparable process matrices that, for any sets of instruments on Alice's side, always give rise to causal assemblages. That is, causally nonseparable process matrices that cannot be certified in a semi-device-independent way.*

In particular, let  $W \in \mathcal{L}(\mathcal{H}^{A_I A_O B_I B_O})$  be a process matrix and  $W^{T_A}$  be the partial transposition of  $W$  with respect to some basis in  $\mathcal{L}(\mathcal{H}^{A_I A_O})$  for Alice. If  $W^{T_A}$  is causally separable, the assemblages generated by  $w_{a|x} = \text{Tr}_A[(A_{a|x} \otimes \mathbb{1}^B) W]$  are causal for every set of instruments  $\{A_{a|x}\}$ .

*Proof.* We start our proof by showing that if  $W^{T_A}$  is causally separable then the assemblages generated by  $w_{a|x} = \text{Tr}_A[(A_{a|x} \otimes \mathbb{1}^B) W]$  are causal for every set of instruments  $\{A_{a|x}\}$ . Straightforward calculations shows that the transposition map is self-adjoint (i.e.  $\text{Tr}[A^T B] = \text{Tr}[A B^T]$ ,  $\forall A, B$ ), in particular, it is true that

$$\text{Tr}_A[(A_{a|x} \otimes \mathbb{1}) W^{T_A}] = \text{Tr}_A[(A_{a|x}^T \otimes \mathbb{1}) W] \quad \forall A_{a|x}. \quad (\text{B.86})$$

Now notice that every instrument can be written as a transposition of another instrument and we can define new valid instruments via  $A'_{a|x} := A_{a|x}^T$ , which allows us to recover any assemblages generated by  $W$  and instruments  $\{A_{a|x}\}$  with a causally separable process matrix. More precisely, since  $(A_{a|x}^T)^T = A_{a|x}$ , the mathematical identities

$$w_{a|x} = \text{Tr}_A[(A_{a|x} \otimes \mathbb{1}) W] \quad (\text{B.87})$$

$$= \text{Tr}_A[((A_{a|x}^T)^T \otimes \mathbb{1}) W] \quad (\text{B.88})$$

$$= \text{Tr}_A[(A_{a|x}^T \otimes \mathbb{1}) W^{T_A}] \quad (\text{B.89})$$

$$= \text{Tr}_A[(A'_{a|x} \otimes \mathbb{1}) W^{\text{sep}}] \quad (\text{B.90})$$

provide an explicit decomposition for the assemblage  $\{w_{a|x}\}$  in terms of a causally separable process matrix and valid instruments.

We finish our proof by referring to ref. [134], which presents several examples of process matrices  $W$  which are causally nonseparable but that  $W^{T_A}$  is causally separable.  $\square$

## B.5 Certification of tripartite process matrices

Here we generalize our approach to certification of indefinite causal order for a tripartite scenario where a process matrix is shared between parties Alice, Bob, and Charlie. In this particular tripartite scenario, Charlie is always in the future of Alice and Bob, so we only study tripartite process matrices that have the causal order  $(A, B) \prec C$ , however, the causal order between Alice and Bob may or may not be well defined. We will only consider this kind of tripartite scenario, the one which is appropriate for the study of the quantum switch processes. A semi-device-independent approach to more general tripartite scenarios, or multipartite scenarios, may be derived, in principle, from a straightforward generalization of our particular tripartite scenario. For more general multipartite scenarios under a device-dependent approach, we refer the reader to ref. [193], and for device-independent, ref. [194].

In the following, we explicitly extend all concepts, definitions, and results from the bipartite case presented in the main text to the tripartite case. We start by defining certification in all 6 inequivalent scenarios that arise from making different assumptions about the operations of each party: TTT (device-dependent), UUU (device-independent), TTU, TUU, UTT, and UUT (semi-device-independent).

**Definition B.9** (TTT (device-dependent) certification). *Given a tripartite behaviour  $\{p(abc|xyz)\}$  that arises from known instruments  $\{\bar{A}_{a|x}\}$  and  $\{\bar{B}_{b|y}\}$ , known POVMs  $\{\bar{M}_{c|z}\}$ , and an unknown tripartite process matrix, one certifies that this process matrix is causally nonseparable if, for some  $a, b, c, x, y, z$ ,*

$$p(abc|\bar{A}_{a|x}, \bar{B}_{b|y}, \bar{M}_{c|z}) \neq \text{Tr}[(\bar{A}_{a|x} \otimes \bar{B}_{b|y} \otimes \bar{M}_{c|z})W^{sep}], \quad (\text{B.91})$$

for all causally separable tripartite process matrices  $W^{sep}$ .

**Definition B.10** (UUU (device-independent) certification). *Given a tripartite behaviour  $\{p(abc|xyz)\}$  that arises from unknown instruments and an unknown tripartite process matrix, one certifies that this process matrix is causally nonseparable if, for some  $a, b, c, x, y, z$ ,*

$$p(abc|xyz) \neq \text{Tr}[(A_{a|x} \otimes B_{b|y} \otimes M_{c|z})W^{sep}], \quad (\text{B.92})$$

for all causally separable tripartite process matrices  $W^{sep}$ , all general instruments  $\{A_{a|x}\}$  and  $\{B_{b|y}\}$ , and all general POVMs  $\{M_{c|z}\}$ . A process matrix certified in such way is called UUU-noncausal, or device-independent noncausal.

**Definition B.11** (TTU (semi-device-dependent) certification). *Given a tripartite behaviour  $\{p(abc|xyz)\}$  that arises from known instruments  $\{\bar{A}_{a|x}\}$  and  $\{\bar{B}_{b|y}\}$  on*

Alice’s and Bob’s side, unknown POVMs on Charlie’s side, and an unknown tripartite process matrix, one certifies that this process matrix is causally nonseparable if, for some  $a, b, c, x, y, z$ ,

$$p(abc|\overline{A}_{a|x}, \overline{B}_{b|y}, z) \neq \text{Tr} [(\overline{A}_{a|x} \otimes \overline{B}_{b|y} \otimes M_{c|z})W^{sep}], \quad (\text{B.93})$$

for all causally separable tripartite process matrices  $W^{sep}$  and all general POVMs  $\{M_{c|z}\}$ . A process matrix certified in such way is called TTU-noncausal.

**Definition B.12** (TUU (semi-device-dependent) certification). Given a tripartite behaviour  $\{p(abc|xyz)\}$  that arises from known instruments  $\{\overline{A}_{a|x}\}$  on Alice’s, unknown instruments on Bob’s and Charlie’s side, and an unknown tripartite process matrix, one certifies that this process matrix is causally nonseparable if, for some  $a, b, c, x, y, z$ ,

$$p(abc|\overline{A}_{a|x}, y, z) \neq \text{Tr} [(\overline{A}_{a|x} \otimes B_{b|y} \otimes M_{c|z})W^{sep}], \quad (\text{B.94})$$

for all causally separable tripartite process matrices  $W^{sep}$ , all general instruments  $\{B_{b|y}\}$ , and all general POVMs  $\{M_{c|z}\}$ . A process matrix certified in such way is called TUU-noncausal.

**Definition B.13** (UTT (semi-device-dependent) certification). Given a tripartite behaviour  $\{p(abc|xyz)\}$  that arises from unknown instruments on Alice’s side, known instruments  $\{\overline{B}_{b|y}\}$  and  $\{\overline{M}_{c|z}\}$  on Bob’s and Charlie’s side, and an unknown tripartite process matrix, one certifies that this process matrix is causally nonseparable if, for some  $a, b, c, x, y, z$ ,

$$p(abc|x, \overline{B}_{b|y}, \overline{M}_{c|z}) \neq \text{Tr} [(A_{a|x} \otimes \overline{B}_{b|y} \otimes \overline{M}_{c|z})W^{sep}], \quad (\text{B.95})$$

for all causally separable tripartite process matrices  $W^{sep}$  and all general instruments  $\{A_{a|x}\}$ . A process matrix certified in such way is called UTT-noncausal.

**Definition B.14** (UUT (semi-device-dependent) certification). Given a tripartite behaviour  $\{p(abc|xyz)\}$  that arises from unknown instruments on Alice’s and Bob’s side, known POVMs  $\{\overline{M}_{c|z}\}$  on Charlie’s side, and an unknown tripartite process matrix, one certifies that this process matrix is causally nonseparable if, for some  $a, b, c, x, y, z$ ,

$$p(abc|x, y, \overline{M}_{c|z}) \neq \text{Tr} [(A_{a|x} \otimes B_{b|y} \otimes \overline{M}_{c|z})W^{sep}], \quad (\text{B.96})$$

for all causally separable tripartite process matrices  $W^{sep}$  and all general instruments  $\{A_{a|x}\}$  and  $\{B_{b|y}\}$ . A process matrix certified in such way is called UUT-noncausal.

### B.5.1 Device-dependent – TTT

Following ref. [46], a *tripartite process matrix*, with Charlie in the future of Alice and Bob, is an operator  $W \in \mathcal{L}(\mathcal{H}^{A_I} \otimes \mathcal{H}^{A_O} \otimes \mathcal{H}^{B_I} \otimes \mathcal{H}^{B_O} \otimes \mathcal{H}^{C_I})$  that satisfies

$$W \geq 0 \quad (\text{B.97})$$

$$\text{Tr } W = d_{A_O} d_{B_O} \quad (\text{B.98})$$

$${}_{A_I A_O C_I} W = {}_{A_I A_O B_O C_I} W \quad (\text{B.99})$$

$${}_{B_I B_O C_I} W = {}_{A_O B_I B_O C_I} W \quad (\text{B.100})$$

$${}_{C_I} W = {}_{A_O C_I} W + {}_{B_O C_I} W - {}_{A_O B_O C_I} W. \quad (\text{B.101})$$

A tripartite process matrix  $W^{A \prec B \prec C}$  is causally ordered from Alice to Bob to Charlie if it satisfies

$${}_{C_I} W^{A \prec B \prec C} = {}_{B_O C_I} W^{A \prec B \prec C} \quad (\text{B.102})$$

$${}_{B_I B_O C_I} W^{A \prec B \prec C} = {}_{A_O B_I B_O C_I} W^{A \prec B \prec C}, \quad (\text{B.103})$$

and a tripartite process matrix  $W^{B \prec A \prec C}$  is causally ordered from Bob to Alice to Charlie if it satisfies

$${}_{C_I} W^{B \prec A \prec C} = {}_{A_O C_I} W^{B \prec A \prec C} \quad (\text{B.104})$$

$${}_{A_I A_O C_I} W^{B \prec A \prec C} = {}_{A_I A_O B_O C_I} W^{B \prec A \prec C}. \quad (\text{B.105})$$

A tripartite process matrix  $W^{\text{sep}}$  is causally separable if it can be expressed as a convex combination of causally ordered process matrices of the kind  $W^{A \prec B \prec C}$  and  $W^{B \prec A \prec C}$ , i.e.,

$$W^{\text{sep}} := q W^{A \prec B \prec C} + (1 - q) W^{B \prec A \prec C} \quad (\text{B.106})$$

where  $0 \leq q \leq 1$  is a real number. A tripartite process matrix that does not satisfy eq. (B.106) is called causally nonseparable.

Just as in the bipartite case, all causally nonseparable tripartite process matrices can be certified in a device-dependent way for some choice of instruments, since tomographically complete instruments allow for the full characterization of the process matrix. Hence, theorem 2.1 also holds in the tripartite case.

### B.5.2 Device-independent – UUU

A *tripartite behaviour*  $\{p(abc|xyz)\}$  is a set of joint probability distributions, that is, a set in which each element  $p(abc|xyz)$  is a real number, such that

$$p(abc|xyz) \geq 0 \quad \forall a, b, c, x, y, z \quad (\text{B.107})$$

$$\sum_{a,b,c} p(abc|xyz) = 1 \quad \forall x, y, z, \quad (\text{B.108})$$

where  $a \in \{1, \dots, O_A\}$ ,  $b \in \{1, \dots, O_B\}$ , and  $c \in \{1, \dots, O_C\}$  label the outcomes and  $x \in \{1, \dots, I_A\}$ ,  $y \in \{1, \dots, I_B\}$ , and  $z \in \{1, \dots, I_C\}$  label the inputs.

A tripartite behaviour in which Charlie is in the future of Alice and Bob is a behaviour  $\{p(abc|xyz)\}$  that satisfies

$$\sum_c p(abc|xyz) = \sum_c p(abc|xyz') \quad \forall a, b, x, y, z, z', \quad (\text{B.109})$$

that is, a behaviour whose Alice and Bob's joint marginal does not depend on Charlie's inputs. We will only consider this type of behaviours and will refer to them as simply tripartite behaviours.

A tripartite behaviour is called a *tripartite process behaviour* if there exist a tripartite process matrix  $W^{A_I A_O B_I B_O C_I}$ , sets of instruments  $\{A_{a|x}^{A_I A_O}\}$  and  $\{B_{b|y}^{B_I B_O}\}$ , and a set of POVMs  $\{M_{c|z}^{C_I}\}$  such that

$$p^Q(abc|xyz) = \text{Tr} \left[ (A_{a|x}^{A_I A_O} \otimes B_{b|y}^{B_I B_O} \otimes M_{c|z}^{C_I}) W^{A_I A_O B_I B_O C_I} \right] \quad \forall a, b, c, x, y, z. \quad (\text{B.110})$$

Just like in the bipartite case, it can be shown that the tripartite process matrix is the most general operator that leads to valid tripartite behaviours when taken the trace with product instruments. Additionally, since not all bipartite behaviours can be realized by process matrices and the bipartite case is a particular case of the tripartite case, not all tripartite behaviours can be realized by process matrices as well. Hence, theorem 2.2 also holds in the tripartite case.

A tripartite behaviour  $\{p^{A \prec B \prec C}(abc|xyz)\}$  is causally ordered from Alice to Bob to Charlie if Alice's marginals do not depend on the inputs of Bob and Charlie, that is,

$$\sum_{b,c} p^{A \prec B \prec C}(abc|xyz) = \sum_{b,c} p^{A \prec B \prec C}(abc|xy'z') \quad \forall a, x, y, y', z, z', \quad (\text{B.111})$$

and a tripartite behaviour  $\{p^{B \prec A \prec C}(abc|xyz)\}$  is causally ordered from Bob to Alice to Charlie if Bob's marginals do not depend on the inputs of Alice and Charlie, that is,

$$\sum_{a,c} p^{B \prec A \prec C}(abc|xyz) = \sum_{a,c} p^{B \prec A \prec C}(abc|x'y'z') \quad \forall a, x, x', y, z, z'. \quad (\text{B.112})$$

A tripartite behaviour  $\{p^{\text{causal}}(abc|xyz)\}$  is causal if it can be written as a convex combination of causally ordered behaviours of the kind  $\{p^{A \prec B \prec C}(abc|xyz)\}$  and  $\{p^{B \prec A \prec C}(abc|xyz)\}$ , i.e.,

$$p^{\text{causal}}(abc|xyz) := qp^{A \prec B \prec C}(abc|xyz) + (1 - q)p^{B \prec A \prec C}(abc|xyz), \quad (\text{B.113})$$

where  $0 \leq q \leq 1$  is a real number. A tripartite behaviour that does not satisfy eq. (B.113) is called noncausal.

**Lemma B.4.** *A tripartite behaviour is causal if and only if it is a tripartite process behaviour that can be obtained by a causally separable tripartite process matrix.*

*Proof.* It can be straightforwardly checked that a behaviour that comes from a causally separable process matrix is causal, the tripartite case being analogous to the bipartite case (see lemma 2.1). To prove that all causal tripartite behaviours can be reproduced by a causally separable tripartite process matrix, we provide the explicit construction of instruments and process matrix below. Given a causal tripartite behaviour  $\{p^{\text{causal}}(abc|xyz)\}$ , one can write its decomposition into definite causal orders using  $\{p^{A \prec B \prec C}(abc|xyz)\}$ ,  $\{p^{B \prec A \prec C}(abc|xyz)\}$ , and  $q$ . According to Bayes' rule and the nonsignaling principle we can calculate the quantities

$$p^{A \prec B \prec C}(abc|xyz) = p_A^{A \prec B \prec C}(a|xyz)p_B^{A \prec B \prec C}(b|axyz)p_C^{A \prec B \prec C}(c|abxyz) \quad (\text{B.114})$$

$$= p_A^{A \prec B \prec C}(a|x)p_B^{A \prec B \prec C}(b|axy)p_C^{A \prec B \prec C}(c|abxyz). \quad (\text{B.115})$$

We use them to define instruments

$$A_{a|x}^{A \prec B \prec C} = \mathbb{1}^{A_I} \otimes p_A^{A \prec B \prec C}(a|x)|ax\rangle\langle ax|^{A_O} \quad (\text{B.116})$$

$$B_{b|y}^{A \prec B \prec C} = \sum_{a,x} p_B^{A \prec B \prec C}(b|axy)|ax\rangle\langle ax|^{B_I} \otimes |abxy\rangle\langle abxy|^{B_O} \quad (\text{B.117})$$

$$M_{c|z}^{A \prec B \prec C} = \sum_{a,b,x,y} p_C^{A \prec B \prec C}(c|abxyz)|abxy\rangle\langle abxy|^{C_I}, \quad (\text{B.118})$$

and the process matrix

$$W^{A \prec B \prec C} = \frac{\mathbb{1}^{A_I}}{d_{A_I}} \otimes |\Phi^+\rangle\langle\Phi^+|^{A_O B_I} \otimes |\Phi^+\rangle\langle\Phi^+|^{B_O C_I}. \quad (\text{B.119})$$

One can check that

$$\text{Tr} \left[ (A_{a|x}^{A \prec B \prec C} \otimes B_{b|y}^{A \prec B \prec C} \otimes M_{c|z}^{A \prec B \prec C}) W^{A \prec B \prec C} \right] = \quad (\text{B.120})$$

$$= p_A^{A \prec B \prec C}(a|x)p_B^{A \prec B \prec C}(b|axy)p_C^{A \prec B \prec C}(c|abxyz) \quad (\text{B.121})$$

$$= p^{A \prec B \prec C}(abc|xyz). \quad (\text{B.122})$$

Equivalently, the instruments and process matrix for the causal order  $B \prec A \prec C$  can be constructed, and by allowing each order to act on a complementary input subspace, just like in the proof of lemma 2.1, all causal tripartite behaviours can be recovered.  $\square$

**Theorem B.8.** *A tripartite process matrix is certified to be causally nonseparable in a device-independent way if and only if it can generate a noncausal tripartite behaviour for some choice of instruments for Alice and Bob and some choice of POVMs for Charlie.*

The proof is analogous to theorem 2.3.

### B.5.3 Semi-device-independent – TTU

We start the first semi-device-independent tripartite scenario by defining assemblages in the TTU scenario using the same reasoning and motivation as the bipartite case, explained in the main text.

**Definition B.15** (Process TTU-assemblage). *A process TTU-assemblage is a set of operators  $\{w_{c|z}^Q\}$ ,  $w_{c|z}^Q \in \mathcal{L}(\mathcal{H}^{A_I A_O} \otimes \mathcal{H}^{B_I B_O})$ , for which there exists a tripartite process matrix  $W^{A_I A_O B_I B_O C_I}$  and a set of POVMs  $\{M_{c|z}^{C_I}\}$  such that*

$$w_{c|z}^Q = \text{Tr}_{C_I} \left[ (\mathbb{1}^{A_I A_O} \otimes \mathbb{1}^{B_I B_O} \otimes M_{c|z}^{C_I}) W^{A_I A_O B_I B_O C_I} \right], \quad (\text{B.123})$$

for all  $c, z$ .

**Definition B.16** (General TTU-assemblage). *A general TTU-assemblage is a set of operators  $\{w_{c|z}\}$ ,  $w_{c|z} \in \mathcal{L}(\mathcal{H}^{A_I A_O} \otimes \mathcal{H}^{B_I B_O})$ , that satisfies*

$$w_{c|z} \geq 0 \quad \forall c, z \quad (\text{B.124})$$

$$\sum_c w_{c|z} = W^{A_I A_O B_I B_O} \quad \forall z, \quad (\text{B.125})$$

where  $W^{A_I A_O B_I B_O}$  is a valid bipartite process matrix.

Intuitively, behaviours are extracted from TTU-assemblages according to

$$p(abc|xyz) = \text{Tr} [(A_{a|x} \otimes B_{b|y}) w_{c|z}] \quad \forall a, b, c, x, y, z. \quad (\text{B.126})$$

Contrarily to the bipartite case, for which we proved that not all general assemblages can be realized by process matrices (theorem 2.4), for the TTU scenario, we prove that general and process TTU-assemblages are actually equivalent.

**Theorem B.9.** *A general TTU-assemblage is valid if and only if it is a process TTU-assemblage.*

*Proof.* By substituting the definition of a tripartite process matrix and POVMs into eq. (B.123) it is easy to check that the resulting assemblage satisfies definition B.16. To show that any valid TTU-assemblage can be obtained with process matrices and instruments, we give the following explicit construction. Given a general TTU-assemblage  $\{w_{c|z}\}$  and the general bipartite process matrix  $W = \sum_c w_{c|z}$ , we construct Charlie's POVMs  $\{M_{c|z}\}$  according to

$$M_{c|z} = W^{-\frac{1}{2}} w_{c|z}^T W^{-\frac{1}{2}}, \quad \forall c, z, \quad (\text{B.127})$$

where the transpose  $^T$  is taken in the basis in which  $W$  is diagonal. Notice that this implies  $\dim(\mathcal{H}^{C_I}) = \dim(\mathcal{H}^{A_I A_O} \otimes \mathcal{H}^{B_I B_O})$  for this particular construction. The sum

$\sum_c M_{c|z} = W^{-\frac{1}{2}} \sum_c w_{c|z}^T W^{-\frac{1}{2}} = W^{-\frac{1}{2}} W W^{-\frac{1}{2}} = \mathbb{1}$  for all  $z$  guarantees it is a valid set of POVMs.

Now, by writing the process matrix  $W$  in its diagonal basis,  $W = \sum_{ij} \mu_{ij} |ij\rangle\langle ij|$ , we can define its purification

$$|W^{ABC}\rangle = \sum_{ij} \sqrt{\mu_{ij}} |ij\ ij\rangle. \quad (\text{B.128})$$

The object  $W^{ABC} = |W^{ABC}\rangle\langle W^{ABC}|$  is a well defined tripartite process matrix. This is true particularly because the dimension of Charlie's output space is 1, that is, because it is in the future of Alice and Bob, and follows from the fact that  $\text{Tr}_C W^{ABC} = W$ . Hence,

$$\text{Tr}_C [(\mathbb{1}^A \otimes \mathbb{1}^B \otimes M_{c|z}^C) W^{ABC}] = \text{Tr}_C [(\mathbb{1}^{AB} \otimes W^{-\frac{1}{2}} w_{c|z}^T W^{-\frac{1}{2}}) W^{ABC}] \quad (\text{B.129})$$

$$= w_{c|z}, \quad (\text{B.130})$$

for all  $c, z$ . This concludes the proof that all TTU-assemblages can be realized with valid tripartite process matrices and a set of POVMs for Charlie.  $\square$

We now define our notion of causality for TTU-assemblages.

**Definition B.17** (Causal TTU-assemblage). *A TTU-assemblage is causally ordered from Alice to Bob to Charlie if it satisfies*

$$\sum_c w_{c|z}^{A \prec B \prec C} = W^{A \prec B} \quad \forall z, \quad (\text{B.131})$$

where  $W^{A \prec B}$  is a bipartite process matrix causally ordered from Alice to Bob, and equivalently from Bob to Alice.

A TTU-assemblage  $\{w_{c|z}^{\text{causal}}\}$  is causal if it can be expressed as a convex combination of causally ordered TTU-assemblages, i.e.,

$$w_{c|z}^{\text{causal}} := q w_{c|z}^{A \prec B \prec C} + (1 - q) w_{c|z}^{B \prec A \prec C} \quad \forall c, z, \quad (\text{B.132})$$

where  $0 \leq q \leq 1$  is a real number. A TTU-assemblage that does not satisfy eq. (B.132) is called a noncausal TTU-assemblage.

Notice that one can decide whether a TTU-assemblage is causal by means of SDP.

**Lemma B.5.** *A TTU-assemblage is causal if and only if it is a process TTU-assemblage that can be obtained from a causal tripartite process matrix.*

The proof is analogous to the one in theorem B.9. Notice that in this case  $W^{ABC}$  will be the purification of a causally ordered process matrix, and therefore, also causally ordered.

**Theorem B.10.** *A tripartite process matrix is certified to be causally nonseparable in a semi-device-independent TTU way if and only if it can generate a noncausal TTU-assemblage for some choice of POVMs for Charlie.*

The proof is analogous to theorem 2.5.

#### B.5.4 Semi-device-independent – TUU

The other semi-device-independent cases follow analogously. We continue with the TUU scenario.

**Definition B.18** (Process TUU-assemblage). *A process TUU-assemblage is a set of operators  $\{w_{bc|yz}\}$ ,  $w_{bc|yz} \in \mathcal{L}(\mathcal{H}^{A_I A_O})$ , for which there exist a tripartite process matrix  $W^{A_I A_O B_I B_O C_I}$ , a set of instruments  $\{B_{b|y}^{B_I B_O}\}$ , and a set of POVMs  $\{M_{c|z}^{C_I}\}$  such that*

$$w_{bc|yz}^Q = \text{Tr}_{B_I B_O C_I} \left[ (\mathbb{1}^{A_I A_O} \otimes B_{b|y}^{B_I B_O} \otimes M_{c|z}^{C_I}) W^{A_I A_O B_I B_O C_I} \right], \quad \forall b, c, y, z. \quad (\text{B.133})$$

**Definition B.19** (General TUU-assemblage). *A general TUU-assemblage is a set of operators  $\{w_{bc|yz}\}$ ,  $w_{bc|yz} \in \mathcal{L}(\mathcal{H}^{A_I A_O})$ , that satisfies*

$$w_{bc|yz} \geq 0 \quad \forall b, c, y, z \quad (\text{B.134})$$

$$\text{Tr} \sum_{b,c} w_{bc|yz} = d_{A_O} \quad \forall y, z \quad (\text{B.135})$$

$$\sum_c w_{bc|yz} = \sum_c w_{bc|yz'} \quad \forall b, y, z, z' \quad (\text{B.136})$$

$$\sum_{b,c} w_{bc|yz} =_{A_O} \sum_{b,c} w_{bc|yz} \quad \forall y, z. \quad (\text{B.137})$$

Intuitively, behaviours are extracted from TUU-assemblages according to

$$p(abc|xyz) = \text{Tr} (A_{a|x} w_{bc|yz}) \quad \forall a, b, c, x, y, z. \quad (\text{B.138})$$

In the bipartite case, we have proven that not all general bipartite assemblages can be realized by process matrices (are process bipartite assemblages), while in the previous tripartite case, TTU, we have proven that all general TTU-assemblages can indeed be realized by process matrices (are process TTU-assemblages). However, in this tripartite scenario, TUU, as well as in the remaining cases, UTT and UUT, it is not clear whether all general TUU-, UTT-, and UUT-assemblages can be realized by process matrices. We leave this problem as an open question.

Nevertheless, all process TUU-assemblages are valid general TUU-assemblages.

We now define our notion of causality for TUU-assemblages.

**Definition B.20** (Causal TUU-assemblage). *A TUU-assemblage is causally ordered from Alice to Bob to Charlie if it satisfies*

$$\sum_{b,c} w_{bc|yz}^{A \prec B \prec C} = \sum_{b,c} w_{bc|y'z'}^{A \prec B \prec C} \quad \forall y, y', z, z', \quad (\text{B.139})$$

and from Bob to Alice to Charlie if it satisfies

$$\sum_c w_{bc|yz}^{B \prec A \prec C} =_{AO} \sum_c w_{bc|yz}^{B \prec A \prec C} \quad \forall b, y, z. \quad (\text{B.140})$$

A TUU-assemblage  $\{w_{bc|yz}^{\text{causal}}\}$  is causal if it can be expressed as a convex combination of causally ordered TUU-assemblages, i.e.,

$$w_{bc|yz}^{\text{causal}} := q w_{bc|yz}^{A \prec B \prec C} + (1 - q) w_{bc|yz}^{B \prec A \prec C} \quad \forall b, c, y, z, \quad (\text{B.141})$$

where  $0 \leq q \leq 1$  is a real number. A TUU-assemblage that does not satisfy eq. (B.141) is called a noncausal TUU-assemblage.

Notice that one can decide whether a TUU-assemblage is causal by means of an SDP.

All causally separable tripartite process matrices lead to causal TUU-assemblages, for whatever choice of instruments. Whether all causal TUU-assemblages can be written in terms of causally separable process matrices is not clear, although for the particular case of assemblages that are causally ordered from Alice to Bob to Charlie, we can show this to be the case.

Here we present our explicit construction of any TUU-assemblage that is causally ordered from Alice to Bob to Charlie by a tripartite process matrix that is causally ordered from Alice to Bob to Charlie.

Given a TUU-assemblage  $\{w_{bc|yz}^{A \prec B \prec C}\}$ , we can define the state  $\rho$  such that  $\sum_{b,c} w_{bc|yz}^{A \prec B \prec C} = \rho^{A_I} \otimes \mathbb{1}^{AO}$ . Since  $\rho \geq 0$  it can be written as  $\rho = \sum_i \mu_i |i\rangle\langle i|$ , which can be purified by  $|\psi\rangle = \sum_i \sqrt{\mu_i} |ii\rangle$ . Then, let  $\{B_{b|y}^{A \prec B \prec C}\}$  and  $\{M_{c|z}^{A \prec B \prec C}\}$  be instruments

$$B_{b|y}^{A \prec B \prec C} = \mathbb{1}^{B_I} \otimes |by\rangle\langle by|^{B_O} \quad (\text{B.142})$$

$$M_{c|z}^{A \prec B \prec C} = \sum_{b,y} \left( \rho^{-\frac{1}{2} C'_I} \otimes \mathbb{1}^{C''_I} \right) w_{bc|yz}^{A \prec B \prec C} C'_I C''_I \left( \rho^{-\frac{1}{2} C'_I} \otimes \mathbb{1}^{C''_I} \right) \otimes |by\rangle\langle by|^{C''_I}, \quad (\text{B.143})$$

where  $\rho^{-1}$  be the inverse of  $\rho$  on its support, the transpose  ${}^T$  is taken on the  $\{|i\rangle\}_i$  basis and  $\mathcal{H}^{C_I} = \mathcal{H}^{C'_I} \otimes \mathcal{H}^{C''_I} \otimes \mathcal{H}^{C''_I}$ . Let

$$W^{A \prec B \prec C} = |\psi\rangle\langle\psi|^{C'_I A_I} \otimes |\Phi^+\rangle\langle\Phi^+|^{C''_I A_O} \otimes \frac{\mathbb{1}^{B_I}}{d_{B_I}} \otimes |\Phi^+\rangle\langle\Phi^+|^{C'''_I B_O} \quad (\text{B.144})$$

be a tripartite process matrix that is causally ordered from Alice to Bob to Charlie. Then, it is true that the assemblage  $\{w_{bc|yz}^{A \prec B}\}$  can be recovered by

$$\text{Tr}_{B_I B_O C_I} \left[ (\mathbb{1}^A \otimes B_{b|y}^{A \prec B \prec C} \otimes M_{c|z}^{A \prec B \prec C}) W^{A \prec B \prec C} \right] = w_{bc|yz}^{A \prec B \prec C}. \quad (\text{B.145})$$

### B.5.5 Semi-device-independent – UTT

Here we detail our concepts and definitions for the UTT scenario.

**Definition B.21** (Process UTT-assemblage). *A process UTT-assemblage is a set of operators  $\{w_{a|x}\}$ ,  $w_{a|x} \in \mathcal{L}(\mathcal{H}^{B_I B_O} \otimes \mathcal{H}^{C_I})$ , for which there exist a tripartite process matrix  $W^{A_I A_O B_I B_O C_I}$  and a set of instruments  $\{A_{a|x}^{A_I A_O}\}$  such that*

$$w_{a|x}^Q = \text{Tr}_{A_I A_O} \left[ (A_{a|x}^{A_I A_O} \otimes \mathbb{1}^{B_I B_O} \otimes \mathbb{1}^{C_I}) W^{A_I A_O B_I B_O C_I} \right], \quad \forall a, x. \quad (\text{B.146})$$

**Definition B.22** (General UTT-assemblage). *A general UTT-assemblage is a set of operators  $\{w_{a|x}\}$ ,  $w_{a|x} \in \mathcal{L}(\mathcal{H}^{B_I B_O} \otimes \mathcal{H}^{C_I})$ , that satisfies*

$$w_{a|x} \geq 0 \quad \forall a, x \quad (\text{B.147})$$

$$\text{Tr} \sum_a w_{a|x} = d_{B_O} \quad \forall x \quad (\text{B.148})$$

$${}_{C_I} \sum_a w_{a|x} = {}_{B_O C_I} \sum_a w_{a|x} \quad \forall x. \quad (\text{B.149})$$

Intuitively, behaviours are extracted from UTT-assemblages according to

$$p(abc|xyz) = \text{Tr} [w_{a|x}(B_{b|y} \otimes M_{c|z})] \quad \forall a, b, c, x, y, z. \quad (\text{B.150})$$

**Definition B.23** (Causal UTT-assemblage). *A UTT-assemblage is causally ordered from Alice to Bob to Charlie if it satisfies*

$${}_{C_I} w_{a|x}^{A \prec B \prec C} = {}_{B_O C_I} w_{a|x}^{A \prec B \prec C} \quad \forall a, x, \quad (\text{B.151})$$

and from Bob to Alice to Charlie if it satisfies

$${}_{C_I} \sum_a w_{a|x}^{B \prec A \prec C} = {}_{C_I} \sum_a w_{a|x'}^{B \prec A \prec C} \quad \forall x, x'. \quad (\text{B.152})$$

A UTT-assemblage  $\{w_{a|x}^{\text{causal}}\}$  is causal if it can be expressed as a convex combination of causally ordered UTT-assemblages, i.e.,

$$w_{a|x}^{\text{causal}} := q w_{a|x}^{A \prec B \prec C} + (1 - q) w_{a|x}^{B \prec A \prec C} \quad \forall a, x, \quad (\text{B.153})$$

where  $0 \leq q \leq 1$  is a real number. A UTT-assemblage that does not satisfy eq. (B.153) is called a noncausal UTT-assemblage.

Notice that one can decide whether a UTT-assemblage is causal by means of an SDP.

It is not clear whether all general UTT-assemblages can be realized by tripartite process matrices nor whether all causal UTT-assemblages can be realized by causally separable tripartite process matrices. Nevertheless, the set of general UTT-assemblages is an outer approximation of the set of process UUT-assemblages and the set of causal UUT-assemblages is an outer approximation of the set of process UTT-assemblages that come from causally separable process matrices. What is not clear is whether these approximations are tight.

### B.5.6 Semi-device-independent – UUT

The final tripartite semi-device independent scenario studied in this work is the UUT scenario.

**Definition B.24** (Process UUT-assemblage). *A process UUT-assemblage is a set of operators  $\{w_{ab|xy}\}$ ,  $w_{ab|xy} \in \mathcal{L}(\mathcal{H}^{C_I})$ , for which there exist a tripartite process matrix  $W^{A_I A_O B_I B_O C_I}$  and sets of instruments  $\{A_{a|x}^{A_I A_O}\}$ ,  $\{B_{b|y}^{B_I B_O}\}$  such that*

$$w_{ab|xy}^Q = \text{Tr}_{A_I A_O B_I B_O} \left[ (A_{a|x}^{A_I A_O} \otimes B_{b|y}^{B_I B_O} \otimes \mathbb{1}^{C_I}) W^{A_I A_O B_I B_O C_I} \right], \quad \forall a, b, x, y. \quad (\text{B.154})$$

**Definition B.25** (General UUT-assemblage). *A general UUT-assemblage is a set of operators  $\{w_{ab|xy}\}$ ,  $w_{ab|xy} \in \mathcal{L}(\mathcal{H}^{C_I})$ , that satisfies*

$$w_{ab|xy} \geq 0 \quad \forall a, b, x, y \quad (\text{B.155})$$

$$\text{Tr} \sum_{a,b} w_{ab|xy} = 1 \quad \forall x, y. \quad (\text{B.156})$$

Intuitively, behaviours are extracted from UUT-assemblages according to

$$p(abc|xyz) = \text{Tr} (w_{ab|xy} M_{c|z}) \quad \forall a, b, c, x, y, z. \quad (\text{B.157})$$

The set of general UUT-assemblages is an outer approximation of the set of process UUT-assemblages but it is not clear to us whether this approximation is tight, i.e., it is not clear whether or not all general UUT-assemblages can be obtained by tripartite process matrices.

**Definition B.26** (Causal UUT-assemblage). *A UUT-assemblage is causally ordered from Alice to Bob to Charlie if it satisfies*

$$\text{Tr} \sum_b w_{ab|xy}^{A \prec B \prec C} = \text{Tr} \sum_b w_{ab|xy'}^{A \prec B \prec C} \quad \forall a, x, y, y' \quad (\text{B.158})$$

and from Bob to Alice to Charlie if it satisfies

$$\mathrm{Tr} \sum_a w_{ab|xy}^{B \prec A \prec C} = \mathrm{Tr} \sum_a w_{ab|x'y}^{B \prec A \prec C} \quad \forall b, x, x', y. \quad (\text{B.159})$$

A UUT-assemblage  $\{w_{ab|xy}^{\text{causal}}\}$  is causal if it can be expressed as a convex combination of causally ordered UUT-assemblages, i.e.,

$$w_{ab|xy}^{\text{causal}} := q w_{ab|xy}^{A \prec B \prec C} + (1 - q) w_{ab|xy}^{B \prec A \prec C} \quad \forall a, b, x, y, \quad (\text{B.160})$$

where  $0 \leq q \leq 1$  is a real number. A UUT-assemblage that does not satisfy eq. (B.160) is called a noncausal UUT-assemblage.

Notice that one can decide whether a UUT-assemblage is causal by means of an SDP.

For the case of causal UUT-assemblages, we prove that our approximation is indeed tight, i.e., that all causal UUT-assemblages can be realized by causal tripartite process matrix, analogously to lemma 2.2 in the bipartite case and lemma B.5 in the TTU tripartite case.

**Lemma B.6.** *A UUT-assemblage is causal if and only if it is a process UUT-assemblage that can be obtained from a causal tripartite process matrix.*

*Proof.* We begin by showing that all UUT-assemblages that come from a causal process matrix are causal. Let  $\{w_{ab|xy}\}$  be such that

$$w_{ab|xy} = \mathrm{Tr}_{A_I A_O B_I B_O} \left[ (A_{a|x}^{A_I A_O} \otimes B_{b|y}^{B_I B_O} \otimes \mathbb{1}^{C_I}) W^{A \prec B \prec C} \right], \quad \forall a, b, x, y. \quad (\text{B.161})$$

Then, using eq. (B.102), which is  ${}_{C_I} W^{A \prec B \prec C} = {}_{B_O C_I} W^{A \prec B \prec C}$ , it is possible to deduce that  $\mathrm{Tr} \sum_b w_{ab|xy}$  is independent of  $y$ , and hence  $\{w_{ab|xy}\} = \{w_{ab|xy}^{A \prec B \prec C}\}$  is causally ordered. The equivalent is true for the order  $B \prec A \prec C$ .

To prove the only if part we show that every causal UUT-assemblage can be reproduced by acting with some instruments on a causal process matrix. Given a causal UUT-assemblage, it can be decomposed into  $\{w_{ab|xy}^{A \prec B \prec C}\}$  and  $\{w_{ab|xy}^{B \prec A \prec C}\}$  with some convex weight  $q$ .

From  $\{w_{ab|xy}^{A \prec B \prec C}\}$ , we define<sup>1</sup>

$$p^{A \prec B \prec C}(ab|xy) := \mathrm{Tr}(w_{ab|xy}^{A \prec B \prec C}) \quad \forall a, b, x, y, \quad (\text{B.162})$$

$$p_A^{A \prec B \prec C}(a|x) := \sum_b \mathrm{Tr}(w_{ab|xy}^{A \prec B \prec C}) \quad \forall a, x, \quad (\text{B.163})$$

$$\rho_{ab|xy} := \frac{w_{ab|xy}^{A \prec B \prec C}}{\mathrm{Tr}(w_{ab|xy}^{A \prec B \prec C})} \quad \forall a, b, x, y, \quad (\text{B.164})$$

---

<sup>1</sup>If  $\mathrm{Tr}(w_{ab|xy}^{A \prec B \prec C}) = 0$ , we define  $\rho_{ab|xy}$  as the null operator.

Using Bayes' rule we also have  $p_B^{A \prec B \prec C}(b|axy) = p^{A \prec B \prec C}(ab|xy)/p_A^{A \prec B \prec C}(a|x)$ . With this given quantities, we can construct instruments  $\{A_{a|x}^{A \prec B \prec C}\}$  and  $\{B_{b|y}^{A \prec B \prec C}\}$ ,

$$A_{a|x}^{A \prec B \prec C} = \mathbb{1}^{A_I} \otimes p_A^{A \prec B \prec C}(a|x)|ax\rangle\langle ax|^{A_O} \quad \forall a, x \quad (\text{B.165})$$

$$B_{b|y}^{A \prec B \prec C} = \sum_{a,x} p_B^{A \prec B \prec C}(b|axy)|ax\rangle\langle ax|^{B_I} \otimes \rho_{ab|xy}^T \quad \forall b, y, \quad (\text{B.166})$$

and also the process matrix

$$W^{A \prec B \prec C} = \frac{\mathbb{1}^{A_I}}{d_{A_I}} \otimes |\Phi^+\rangle\langle\Phi^+|^{A_O B_I} \otimes |\Phi^+\rangle\langle\Phi^+|^{B_O C_I}, \quad (\text{B.167})$$

and check that

$$\text{Tr}_{A_I A_O B_I B_O} \left[ (A_{a|x}^{A \prec B \prec C} \otimes B_{b|y}^{A \prec B \prec C} \otimes \mathbb{1}^{C_I}) W^{A \prec B \prec C} \right] = \quad (\text{B.168})$$

$$= p_A^{A \prec B \prec C}(a|x)p_B^{A \prec B \prec C}(b|axy)\rho_{ab|xy} \quad (\text{B.169})$$

$$= w_{ab|xy}^{A \prec B \prec C} \quad (\text{B.170})$$

for every  $a, b, x, y$ . Analogously, the same holds for  $\{w_{ab|xy}^{B \prec A \prec C}\}$ .

Finally, just like in the proof of lemma 2.1, by allowing the different causal orders to act in complementary subspaces, we can recover any convex combinations of causally ordered assemblages, i.e., causal assemblages.  $\square$

**Theorem B.11.** *A tripartite process matrix is certified to be causally nonseparable in a semi-device-independent UUT way if and only if it can generate a noncausal UUT-assemblage for some choice of instruments for Alice and Bob.*

The proof is analogous to theorem 2.5.

## B.6 Proof that the quantum switch processes are causal in the UUT scenario

In this appendix we prove theorem 2.7, which we restate for the convenience of the reader.

**Theorem 2.7.** *The quantum switch processes cannot be certified to be causally nonseparable on a semi-device-independent scenario where Alice and Bob are untrusted and Charlie is trusted (UUT).*

Moreover, any tripartite process matrix  $W \in \mathcal{L}(\mathcal{H}^{A_I A_O B_I B_O C_I})$ , with Charlie in the future of Alice and Bob, that satisfies

$$\text{Tr}[(A_{a|x}^{A_I A_O} \otimes B_{b|y}^{B_I B_O} \otimes \mathbb{1}^{C_I}) W^{A_I A_O B_I B_O C_I}] = qp^{A \prec B}(ab|xy) + (1-q)p^{B \prec A}(ab|xy), \quad (2.41)$$

for all  $a, b, x, y$ , where  $0 \leq q \leq 1$  is a real number, cannot be certified to be causally nonseparable in a UUT scenario.

*Proof.* Let

$$w_{ab|xy}^{\text{switch}} := \text{Tr}_{A_I A_O B_I B_O} \left[ (A_{a|x}^{A_I A_O} \otimes B_{b|y}^{B_I B_O} \otimes \mathbb{1}^{C_I}) W^{\text{switch}} \right] \quad (\text{B.171})$$

be a UUT-assemblage generated by the quantum switch. We define

$$p(ab|xy) := \text{Tr}(w_{ab|xy}^{\text{switch}}) \quad (\text{B.172})$$

and<sup>2</sup>

$$\rho_{ab|xy} := \frac{w_{ab|xy}^{\text{switch}}}{p(ab|xy)}. \quad (\text{B.173})$$

Given that

$$p(ab|xy) = \text{Tr}(w_{ab|xy}^{\text{switch}}) \quad (\text{B.174})$$

$$= \text{Tr} \left[ (A_{a|x}^{A_I A_O} \otimes B_{b|y}^{B_I B_O} \otimes \mathbb{1}^{C_I}) W^{\text{switch}} \right] \quad (\text{B.175})$$

$$= qp^{A \prec B}(ab|xy) + (1 - q)p^{B \prec A}(ab|xy), \quad (\text{B.176})$$

since the quantum switch is device-independent causal, one can write

$$w_{ab|xy}^{\text{switch}} = p(ab|xy)\rho_{ab|xy} \quad (\text{B.177})$$

$$= qp^{A \prec B}(ab|xy)\rho_{ab|xy} + (1 - q)p^{B \prec A}(ab|xy)\rho_{ab|xy}. \quad (\text{B.178})$$

One can verify that the first term satisfies

$$\text{Tr} \left( \sum_b p^{A \prec B}(ab|xy)\rho_{ab|xy} \right) = \sum_b p^{A \prec B}(ab|xy)\text{Tr}(\rho_{ab|xy}) = p(a|x) \quad (\text{B.179})$$

and

$$\text{Tr} \left( \sum_a p^{B \prec A}(ab|xy)\rho_{ab|xy} \right) = \sum_a p^{B \prec A}(ab|xy)\text{Tr}(\rho_{ab|xy}) = p(b|y). \quad (\text{B.180})$$

Hence,

$$p^{A \prec B}(ab|xy)\rho_{ab|xy} = w_{ab|xy}^{A \prec B} \quad (\text{B.181})$$

is an assemblage causally ordered from Alice to Bob and

$$p^{B \prec A}(ab|xy)\rho_{ab|xy} = w_{ab|xy}^{B \prec A} \quad (\text{B.182})$$

is an assemblage causally ordered from Bob to Alice. Consequently,

$$w_{ab|xy}^{\text{switch}} = qp_{ab|xy}^{A \prec B} + (1 - q)p_{ab|xy}^{B \prec A} \quad (\text{B.183})$$

is UUT causal for all instruments of Alice and Bob.

This proof holds for all process matrices  $W$  for which

$$p(ab|xy) = \text{Tr} \left[ (A_{a|x}^{A_I A_O} \otimes B_{b|y}^{B_I B_O} \otimes \mathbb{1}^{C_I}) W \right] \in \text{CAUSAL}. \quad (\text{B.184})$$

□

---

<sup>2</sup>If  $\text{Tr}(w_{ab|xy}^{\text{switch}}) = 0$ , we define  $\rho_{ab|xy}$  as the null operator.

## B.7 Analysis of experimental implementations of the quantum switch

In this appendix we present more details of our analysis of the experimental results involving the quantum switch reported in refs. [72, 74]. Among other assumptions, both papers consider a device-dependent scenario, where the analysis is made by assuming complete knowledge of all instruments involved (although no assumption is made about the process matrix). Here, we show that the experiment reported in both papers could have also certified indefinite causal order in a semi-device-independent scenario, where no assumptions are made about the instruments of one of the parties.

We remark that (device-dependent) causal witnesses, including the ones of refs. [72, 74], are derived assuming that all instruments are implemented perfectly. However, due to experimental imperfections, this assumption is seldom true. One might then obtain a noncausal – even nonprocess – behaviour simply because the implemented instruments are not the ideal ones, rather than the data being genuinely noncausal (or nonprocess). Indeed, we have analysed the experimental data<sup>3</sup> collected in ref. [72] and verified that if the instruments performed by Alice, Bob, and Charlie are trusted, the experimental behaviour is not a process behaviour. That is, there does not exist any process matrix  $W$ , causally separable or not, that can generate the experimental behaviour consistently with the assumed instruments<sup>4</sup>. To properly analyse experimental data we would need to allow some leeway in the instruments and probabilities, but developing the methods for that is beyond the scope of this paper.

Therefore, in order to avoid potential false positive results due to experimental error, we have considered the statistics one would have obtained in an ideal version of the experiment instead of considering the data collected on the actual experiments.

We start by analysing the device-dependent experiment described in ref. [74] by Goswami *et al.*, which considers an optical setup where three parties, Alice, Bob, and Charlie, have access to the reduced quantum switch process  $W_{\text{red}}$  (see eq. (2.42)). In this experiment, Alice and Bob can choose between the 8 different unitary operations<sup>5</sup> and Charlie performs a measurement in the  $\sigma_X$  basis on the control qubit state. Theoretically, the statistics of an ideal device-dependent TTT

---

<sup>3</sup>We have analysed the data points of this experiment without taking the error bars into consideration.

<sup>4</sup>Moreover, even if we drop the assumption about the knowledge of the instrument performed by Charlie, there is no process matrix which is consistent with the experimental data collected.

<sup>5</sup>Note that an unitary operation can be seen as an instrument with deterministic classical output.

(trusted-trusted-trusted) experiment would be given by the behaviour

$$p^{\text{ideal}1}(c|x, y) := \text{Tr} [(\overline{U}_x \otimes \overline{U}_y \otimes \overline{M}_c) W_{\text{red}}], \quad (\text{B.185})$$

where the instruments  $\{\overline{U}_x\}$  and  $\{\overline{M}_c\}$  are the ones described on the supplemental material of ref. [74]. As pointed in ref. [74], the behaviour  $\{p^{\text{ideal}1}(c|x, y)\}$  certifies indefinite causal order in a device-dependent way. Using the SDP formulation of the problem described in appendix B.5, we show that the noisy behaviour

$$p_\eta^{\text{ideal}1}(c|x, y) := (1 - \eta) p^{\text{ideal}1}(c|x, y) + \eta \frac{1}{2} \quad (\text{B.186})$$

where  $p_I(c|x, y) = \frac{1}{2}$  is a uniform probability distribution, cannot be described by a causally separable tripartite process matrix, i.e., for some  $c, x, y$ ,

$$p_\eta^{\text{ideal}1}(c|x, y) \neq \text{Tr} [(\overline{U}_x \otimes \overline{U}_y \otimes \overline{M}_c) W^{\text{sep}}], \quad (\text{B.187})$$

in the range  $\eta \in [0, 0.1989]$ . Hence, in this range of  $\eta$ , the behaviour  $\{p_\eta^{\text{ideal}1}(c|x, y)\}$  certifies indefinite causal order in a device-dependent (TTT) way.

In order to analyse the behaviour  $\{p^{\text{ideal}1}(c|x, y)\}$  in a semi-device-independent scenario, we drop the hypothesis that the measurement  $\{M_c\}$  performed by Charlie is trusted, working in a TTU (trusted-trusted-untrusted) scenario. Using the SDP formulation of the problem described in appendix B.5, we show that the noisy behaviour  $\{p_\eta^{\text{ideal}1}(c|x, y)\}$  cannot be described by a causal TTU-assemblage, i.e., for some  $c, x, y$ ,

$$p_\eta^{\text{ideal}1}(c|x, y) \neq \text{Tr} [(\overline{U}_x \otimes \overline{U}_y) w_c^{\text{causal}}], \quad (\text{B.188})$$

in the same range of  $\eta \in [0, 0.1989]$ . Hence, in this range of  $\eta$ , the behaviour  $\{p_\eta^{\text{ideal}1}(c|x, y)\}$  certifies indefinite causal order in a semi-device-independent (TTU) way.

Therefore, the experimental setup described in ref. [74] allows for certification of indefinite causal order with weaker hypotheses – in a semi-device-independent scenario. Using the machinery developed in this work, we could not show that the behaviour  $\{p^{\text{ideal}1}(c|x, y)\}$  can certify indefinite causal order in the UTT or TUU scenarios. However, since some of our SDP methods for the tripartite case may provide only an outer approximation of the sets of causal assemblages (see appendix B.5), we cannot discard the possibility of such certification either. We remark, nevertheless, that we have proven that it is possible to certify that the switch process is causally nonseparable in these scenarios for the right choice of instruments (see section 2.4).

The experiment performed in ref. [72] by Rubino *et al.* considers a four-partite version of the switch process which is slightly more general than the tripartite one

discussed in this paper. While we consider a tripartite switch process where the target state is embedded in the process, ref. [72] considers a four-partite switch process in which the target state can be chosen by a fourth party in the common past of all other parties. If we label the fourth party  $D_O$  (David output), the process matrix of the four-partite quantum switch is given by  $W_{\text{switch}}^4 := |w_{\text{switch}}^4\rangle\langle w_{\text{switch}}^4|$ , where

$$|w_{\text{switch}}^4\rangle := \frac{1}{\sqrt{2}} \left( |\Phi^+\rangle^{D_O A_I} |\Phi^+\rangle^{A_O B_I} |\Phi^+\rangle^{B_O C_I^t} |1\rangle^{C_I^c} + |\Phi^+\rangle^{D_O B_I} |\Phi^+\rangle^{B_O A_I} |\Phi^+\rangle^{A_O C_I^t} |0\rangle^{C_I^c} \right) \quad (\text{B.189})$$

and  $|\Phi^+\rangle := |00\rangle + |11\rangle$  is an unnormalized maximally entangled two-qubit state.

Reference [72] considers an optical setup where Alice, Bob, Charlie, and David have access to a reduced four-partite quantum switch process given by  $W_{\text{red}}^4 := \text{Tr}_{C_I^t}(W_{\text{switch}}^4)$ . In this experiment, David can choose between 3 different states  $\bar{\rho}_d$ , Alice can choose between 12 different dichotomic instruments  $\{\bar{A}_{a|x}\}$ , and Bob can choose between 10 unitary operations represented by the instruments  $\{\bar{U}_y\}$ . Theoretically, in an ideal device-dependent TTTT experiment, the statistics would be given by the behaviour

$$p^{\text{ideal2}}(a, c|x, y, d) := \text{Tr} \left[ (\bar{\rho}_d \otimes \bar{A}_{a|x} \otimes \bar{U}_y \otimes \bar{M}_c) W_{\text{red}}^4 \right], \quad (\text{B.190})$$

where the instruments  $\{\bar{\rho}_d\}$ ,  $\{\bar{A}_{a|x}\}$ ,  $\{\bar{U}_y\}$ , and  $\{\bar{M}_c\}$  are the ones described on the supplemental material of ref. [72].

In order to use the SDP formulations described in appendix B.5, we restrict our analysis to the particular case where David outputs the state  $\rho_1 = |0\rangle\langle 0|$ . That is, we analyse the behaviour

$$p^{\text{ideal2}}(a, c|x, y, 1) := \text{Tr} \left[ (\overline{|0\rangle\langle 0|} \otimes \bar{A}_{a|x} \otimes \bar{U}_y \otimes \bar{M}_c) W_{\text{red}}^4 \right]. \quad (\text{B.191})$$

Notice that when David is restricted to the choice of the state  $|0\rangle\langle 0|$ , the four-partite reduced switch process  $W_{\text{red}}^4$  relates to the tripartite reduced switch process  $W_{\text{red}}$  (eq. (2.42)) via the identity

$$\text{Tr}_{D_O} \left[ (|0\rangle\langle 0|^{D_O} \otimes \mathbb{1}^{A_I A_O} \otimes \mathbb{1}^{B_I B_O} \otimes \mathbb{1}^{C_I^c}) W_{\text{red}}^4 \right] = W_{\text{red}}. \quad (\text{B.192})$$

Moreover, if the behaviour  $\{p^{\text{ideal2}}(a, c|x, y, 1)\}$  allows for certification of indefinite causal order, the full behaviour  $\{p^{\text{ideal2}}(a, c|x, y, d)\}$  also allows for certification of indefinite causal order. We can, hence, use our tripartite machinery to certify indefinite causal order in this particular four-partite case.

Using the SDP formulation of the problem described in appendix B.5, we show that the noisy behaviour

$$p_\eta^{\text{ideal2}}(a, c|x, y, 1) := (1 - \eta) p^{\text{ideal2}}(a, c|x, y, 1) + \eta \frac{1}{4} \quad (\text{B.193})$$

where  $p_I(a, c|x, y) = \frac{1}{4}$  is a uniform probability distribution, cannot be described by a causally separable tripartite process matrix, i.e., for some  $a, c, x, y$ ,

$$p_\eta^{\text{ideal}2}(a, c|x, y, 1) \neq \text{Tr} [(\overline{A}_{a|x} \otimes \overline{U}_y \otimes \overline{M}_c) W^{\text{sep}}], \quad (\text{B.194})$$

in the range  $\eta \in [0, 0.2300)$ . Hence, in this range of  $\eta$ , the behaviour  $\{p_\eta^{\text{ideal}2}(a, c|x, y, 1)\}$  certifies indefinite causal order in a device-dependent (TTT) way.

We now analyse the behaviour  $p^{\text{ideal}2}$  in a semi-device-independent scenario, where we drop the hypothesis that the measurement  $\{M_c\}$  performed by Charlie is trusted, working in a TTU (trusted-trusted-untrusted) scenario. Using the SDP formulation of the problem described in appendix B.5, we show that the noisy behaviour  $\{p_\eta^{\text{ideal}1}(c|x, y)\}$  cannot be described by a causal TTU-assemblage, i.e., for some  $a, c, x, y$ ,

$$p_\eta^{\text{ideal}2}(a, c|x, y, 1) \neq \text{Tr} [(\overline{A}_{a|x} \otimes \overline{U}_y) w_c^{\text{causal}}], \quad (\text{B.195})$$

in the same range of  $\eta \in [0, 0.2300)$ . Hence, in this range of  $\eta$ , the behaviour  $\{p_\eta^{\text{ideal}2}(a, c|x, y, 1)\}$  certifies indefinite causal order in a semi-device-independent (TTU) way.

Therefore, the experimental setup described in ref. [72] allows for certification of indefinite causal order with weaker hypotheses – in a semi-device-independent scenario. Using the machinery developed in this work, we could not show that the behaviour  $\{p^{\text{ideal}2}(a, c|x, y, d)\}$  can certify indefinite causal order in other semi-device-independent scenarios. However, since we have only considered the case where the state  $\rho_z$  chosen by David is fixed and some of our SDP methods for the tripartite case may provide only an outer approximation of the sets of causal assemblages (see appendix B.5), we cannot discard the possibility of such certification either. We remark, nevertheless, that we have proven that it is possible to certify that the switch process is causally nonseparable in other semi-device-independent scenarios for the right choice of instruments (see section 2.4).



# Supplemental Information of Chapter 3

Here we present support material that complements the main text. It is structured as follows: Sec. C.1. Characterization theorem for general testers, Sec. C.2. Semidefinite programming and dual affine spaces, Sec. C.3. Computer-assisted proofs, and Sec. C.4. Sampling general channels and the typicality of the hierarchy between discrimination strategies.

## C.1 Characterization theorem for general testers

We start by demonstrating, for sake of completeness, the characterization of general one-copy testers that was presented in the main text, in the language of our paper. This result is already known and follows from Ref. [42].

**Theorem C.2.** *Let  $T = \{T_i\}_{i=1}^N$ ,  $T_i \in \mathcal{L}(\mathcal{H}^I \otimes \mathcal{H}^O)$ , called a general one-copy tester, be the most general set of operators that satisfy the relation*

$$p(i|C) = \text{Tr}(T_i C), \quad (\text{C.1})$$

*for all Choi operators of quantum channels  $C \in \mathcal{L}(\mathcal{H}^I \otimes \mathcal{H}^O)$ , where  $\{p(i|C)\}$  is a set of probability distributions. Let  $W := \sum_i T_i$ . Then,  $T = \{T_i\}$  is a set of operators that satisfy*

$$T_i \geq 0 \quad \forall i \quad (\text{C.2})$$

$$\text{Tr}(W) = d_O \quad (\text{C.3})$$

$$W = {}_O W. \quad (\text{C.4})$$

*Proof.* In order to guarantee that  $\{p(i|C)\}$  is a valid probability distribution, two conditions must be imposed: positivity and normalization.

Positivity:

$$p(i|C) = \text{Tr}(T_i C) \geq 0 \quad \forall i, C \geq 0 \iff T_i \geq 0 \quad \forall i. \quad (\text{C.5})$$

Normalization:

$$\sum_i p(i|C) = \text{Tr}\left(\sum_i T_i C\right) = \text{Tr}(W C) = 1 \quad \forall \text{ channels } C, \quad (\text{C.6})$$

where  $C$  is the Choi operator of a quantum channel, and therefore of a trace-preserving map, which can be parametrized as  $C = X -_O X + \frac{1}{d_O}$ , where  $X$  is an self-adjoint operator, using the same technique as in Appendix B of Ref. [46]. Then,

$$\text{Tr}[W(X -_O X + \frac{1}{d_O})] = 1 \quad \forall \text{self-adjoint } X. \quad (\text{C.7})$$

We can split this in two cases:  $X = 0$  and  $X \neq 0$ .

For  $X = 0$ :

$$\text{Tr}[W(X -_O X + \frac{1}{d_O})] = \frac{\text{Tr}(W)}{d_O} = 1 \iff \text{Tr}(W) = d_O. \quad (\text{C.8})$$

For  $X \neq 0$ :

$$\text{Tr}[W(X -_O X + \frac{1}{d_O})] = \text{Tr}[W(X -_O X)] + 1 = 1 \quad \forall X \neq 0 \quad (\text{C.9})$$

$$\iff \text{Tr}[W(X -_O X)] = 0 \quad \forall X \neq 0 \quad (\text{C.10})$$

$$\iff \text{Tr}[(W -_O W)X] = 0 \quad \forall X \neq 0 \quad (\text{C.11})$$

$$\iff W -_O W = 0. \quad (\text{C.12})$$

The equivalence between Eqs. (C.10) and (C.12) is given by the self-duality of the ‘trace-and-replace’ map, namely  $\text{Tr}(W_O X) = \text{Tr}(O W X)$ .

Together, conditions  $\text{Tr}(W) = d_O$  (Eq. (C.3)) and  $W =_O W$  (Eq. (C.4)) imply that  $W$  can be written as  $W = \sigma \otimes \mathbb{1}^O$ , where  $\sigma \in \mathcal{L}(\mathcal{H}^I)$  is a normalized quantum state.  $\square$

Now we prove a new characterization theorem, the one of general two-copy testers. In this case, we will need additional hypotheses. One is the hypothesis that a tester may not only be able to act on two copies of the same channel but also be able to act on two different, independent channels. This hypothesis is physically motivated in the sense that, if a general tester is a device in a quantum lab that can act on two copies of the same channel, then one should also be able to plug in two different channels and have it perform a meaningful physical operation. The second is that

these channels should be allowed to also act on auxiliary, potentially entangled, systems, and when a general tester acts upon part of these channels, the operation it performs should still result in a valid probability distribution. This last hypothesis is automatically satisfied in the one-copy case.

Formally, we have:

**Theorem C.3.** *Let  $T^{GEN} = \{T_i^{GEN}\}_{i=1}^N$ ,  $T_i^{GEN} \in \mathcal{L}(\mathcal{H}^{I_1} \otimes \mathcal{H}^{O_1} \otimes \mathcal{H}^{I_2} \otimes \mathcal{H}^{O_2})$ , called a general two-copy tester, be the most general set of operators that satisfy the relation*

$$p(i|C_A, C_B, \rho_{AB}) = \text{Tr} [(T_i^{GEN} \otimes \rho_{AB})(C_A \otimes C_B)], \quad (\text{C.13})$$

for all Choi operators of quantum channels  $C_A \in \mathcal{L}(\mathcal{H}^{I_1} \otimes \mathcal{H}^{O_1} \otimes \mathcal{H}^{aux_1})$  and  $C_B \in \mathcal{L}(\mathcal{H}^{I_2} \otimes \mathcal{H}^{O_2} \otimes \mathcal{H}^{aux_2})$ , and for all quantum states  $\rho_{AB} \in \mathcal{L}(\mathcal{H}^{aux_1} \otimes \mathcal{H}^{aux_2})$ , where  $\{p(i|C)\}$  is a set of probability distributions. Let  $W^{GEN} := \sum_i T_i^{GEN}$ . Then,  $T^{GEN} = \{T_i^{GEN}\}$  is a set of operators that satisfy

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

$$\text{Tr}(W^{GEN}) = d_{O_1} d_{O_2} \quad (\text{C.15})$$

$${}_{I_1 O_1} W^{GEN} = {}_{I_1 O_1 O_2} W^{GEN} \quad (\text{C.16})$$

$${}_{I_2 O_2} W^{GEN} = {}_{O_1 I_2 O_2} W^{GEN} \quad (\text{C.17})$$

$$W^{GEN} = {}_{O_1} W^{GEN} + {}_{O_2} W^{GEN} - {}_{O_1 O_2} W^{GEN}. \quad (\text{C.18})$$

*Proof.* Again, in order to guarantee that  $\{p(i|C_A, C_B)\}$  is a valid probability distribution, the conditions of positivity and normalization must be imposed.

Positivity:

$$p(i|C_A, C_B, \rho_{AB}) = \text{Tr} [(T_i^{GEN} \otimes \rho_{AB})(C_A \otimes C_B)] \geq 0 \quad \forall i, C_A \geq 0, C_B \geq 0, \rho_{AB} \geq 0 \quad (\text{C.19})$$

$$\iff T_i^{GEN} \geq 0 \quad \forall i. \quad (\text{C.20})$$

Normalization:

$$\begin{aligned} \sum_i p(i|C_A, C_B, \rho_{AB}) &= \text{Tr} [(\sum_i T_i^{GEN} \otimes \rho_{AB})(C_A \otimes C_B)] \\ &= \text{Tr} [(W^{GEN} \otimes \rho_{AB})(C_A \otimes C_B)] = 1 \end{aligned} \quad (\text{C.21})$$

$\forall$  channels  $C_A, C_B$  and states  $\rho_{AB}$ .

Notice that condition Eq. (C.21) is exactly the normalization condition that, in Appendix B of Ref. [46], defines  $W^{GEN}$  as a bipartite process matrix. Hence, it immediately follows from the proof contained therein that  $W^{GEN}$  must respect Eqs. (C.15)-(C.18).  $\square$

## C.2 Semidefinite programming formulation and dual affine spaces

In this section we present a method to obtain a dual problem formulation for a class of convex optimization problems which covers the SDP presented in our main text. This method employs ideas and techniques first presented in Ref. [148].

A subset of linear operators  $\mathcal{W} \subseteq \mathcal{L}(\mathcal{H})$  is said to be affine if for every set of real numbers  $\{w_i\}_i$  respecting  $\sum_i w_i = 1$ , and for every subset  $\{W_i\}_i \subseteq \mathcal{W}$  we have that  $(\sum_i w_i W_i) \in \mathcal{W}$ .

**Definition C.5** (Dual affine space [148]). *Let  $\mathcal{W} \subseteq \mathcal{L}(\mathcal{H})$  be a set of linear operators. The dual affine space  $\overline{\mathcal{W}}$  of  $\mathcal{W}$  is defined via*

$$\overline{W} \in \overline{\mathcal{W}} \text{ when } \text{Tr}(\overline{W}W) = 1, \forall W \in \mathcal{W}. \quad (\text{C.22})$$

If  $\mathcal{W} \subseteq \mathcal{L}(\mathcal{H})$  is the set of all quantum states, i.e., positive semidefinite operators  $W \in \mathcal{L}(\mathcal{H})$  such that  $\text{Tr}(W) = 1$ , the only operator  $\overline{W}$  such that  $\text{Tr}(\overline{W}W) = 1, \forall W \in \mathcal{W}$  is the identity operator. Hence, the dual affine space of set of quantum states has a single element which is the identity operator  $\mathbb{1}$  and corresponds to the normalisation constraint for quantum measurements.

If  $\mathcal{W}^{\text{PAR}} \subseteq \mathcal{L}(\mathcal{H}^I \otimes \mathcal{H}^O)$ , where  $\mathcal{H}^I = \bigotimes_{i=1}^k I_i$  and  $\mathcal{H}^O = \bigotimes_{i=1}^k O_i$  stands for the set of all parallel processes, i.e., positive semidefinite operators that can be written as  $W^{\text{PAR}} = \sigma^I \otimes \mathbb{1}^O$ , with  $\text{Tr}(\sigma) = 1$ , one can check that its dual affine space is given by a set of linear operators  $\overline{W}^{\text{PAR}}$  respecting  $\text{Tr}_O(\overline{W}^{\text{PAR}}) = \mathbb{1}^I$ , which is the set of quantum channels without the positivity condition.

If  $\mathcal{W}^{\text{SEQ}} \subseteq \mathcal{L}(\mathcal{H}^I \otimes \mathcal{H}^O)$  stands for the set of all sequential processes, Ref. [148] shows that its dual affine space  $\overline{\mathcal{W}}^{\text{SEQ}}$  is given by the set of Choi operators of  $k$ -partite channels with memory<sup>1</sup> [169] without the positivity constraint. In particular, for the two-slot case, an operator  $\overline{W}^{\text{SEQ}} \in \mathcal{L}(\mathcal{H}^{I_1} \otimes \mathcal{H}^{O_1} \otimes \mathcal{H}^{I_2} \otimes \mathcal{H}^{O_2})$  belongs to the dual affine space of the sequential processes if and only if  $\overline{W}^{\text{SEQ}}$  respects

$$O_2 \overline{W}^{\text{SEQ}} =_{I_2 O_2} \overline{W}^{\text{GEN}} \quad (\text{C.23})$$

$$O_1 I_2 O_2 \overline{W}^{\text{SEQ}} =_{I_1 O_1 I_2 O_2} \overline{W}^{\text{SEQ}} \quad (\text{C.24})$$

$$\text{Tr}(\overline{W}^{\text{SEQ}}) = d_{I_1} d_{I_2}. \quad (\text{C.25})$$

If  $\mathcal{W}^{\text{GEN}} \subseteq \mathcal{L}(\mathcal{H}^I \otimes \mathcal{H}^O)$  stands for the set of all general processes, Ref. [148] shows that its dual affine space  $\overline{\mathcal{W}}^{\text{GEN}}$  is given by the set of Choi operators of  $k$ -partite no-signalling channels [195, 196] without the positivity constraint. In particular, for

---

<sup>1</sup>Note that a  $k$ -partite channel with memory is formally equivalent to a quantum comb with  $k - 1$  slots [42]

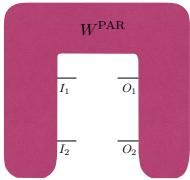
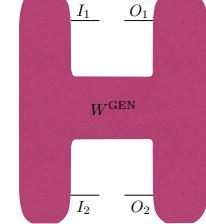
	PROCESS	DUAL AFFINE (CHANNEL)		
PARALLEL		$\text{Tr}(W^{\text{PAR}}) = d_{O_1} d_{O_2}$ $W^{\text{PAR}} =_{O_1 O_2} W^{\text{PAR}}$		$\text{Tr}(\overline{W}^{\text{PAR}}) = d_{I_1} d_{I_2}$ $o_1 o_2 \overline{W}^{\text{PAR}} =_{I_1 O_1 I_2 O_2} \overline{W}^{\text{PAR}}$
SEQUENTIAL		$\text{Tr}(W^{\text{SEQ}}) = d_{O_1} d_{O_2}$ $W^{\text{SEQ}} =_{O_2} W^{\text{SEQ}}$ $I_2 O_2 W^{\text{SEQ}} =_{O_1 I_2 O_2} W^{\text{SEQ}}$		$\text{Tr}(\overline{W}^{\text{SEQ}}) = d_{I_1} d_{I_2}$ $o_2 \overline{W}^{\text{SEQ}} =_{I_2 O_2} \overline{W}^{\text{SEQ}}$ $o_1 I_2 O_2 \overline{W}^{\text{SEQ}} =_{I_1 O_1 I_2 O_2} \overline{W}^{\text{SEQ}}$
GENERAL		$\text{Tr}(W^{\text{GEN}}) = d_{O_1} d_{O_2}$ $I_1 O_1 W^{\text{GEN}} =_{I_1 O_1 O_2} W^{\text{GEN}}$ $I_2 O_2 W^{\text{GEN}} =_{O_1 I_2 O_2} W^{\text{GEN}}$ $W^{\text{GEN}} =_{O_1} W^{\text{GEN}} +_{O_2} W^{\text{GEN}} -_{O_1 O_2} W^{\text{GEN}}$		$\text{Tr}(\overline{W}^{\text{GEN}}) = d_{I_1} d_{I_2}$ $o_1 \overline{W}^{\text{GEN}} =_{I_1 O_1} \overline{W}^{\text{GEN}}$ $o_2 \overline{W}^{\text{GEN}} =_{I_2 O_2} \overline{W}^{\text{GEN}}$

Figure C.1: Normalization constraints for parallel, sequential, and general two-slot processes, and for their dual affine spaces, which correspond to the normalization constraints for bipartite channels, bipartite channels with memory, and bipartite no-signalling channel respectively. Note that the dual affine space of a set  $\mathcal{W}^S$  may be intuitively visualized as the largest set of “quantum objects”  $\overline{\mathcal{W}}^S$  such that “connecting” objects from  $\mathcal{W}^S$  to objects from  $\overline{\mathcal{W}}^S$  always lead to the scalar number 1.

the two-slot case, an operator  $\overline{W}^{\text{GEN}} \in \mathcal{L}(\mathcal{H}^{I_1} \otimes \mathcal{H}^{O_1} \otimes \mathcal{H}^{I_2} \otimes \mathcal{H}^{O_2})$  belongs to the dual affine space of the general processes if and only if  $\overline{W}^{\text{GEN}}$  respects

$$o_2 \overline{W}^{\text{GEN}} =_{I_2 O_2} \overline{W}^{\text{GEN}} \quad (\text{C.26})$$

$$o_1 \overline{W}^{\text{GEN}} =_{I_1 O_1} \overline{W}^{\text{GEN}} \quad (\text{C.27})$$

$$\text{Tr}(\overline{W}^{\text{GEN}}) = d_{I_1} d_{I_2}. \quad (\text{C.28})$$

We have summarized the normalization constraints of parallel, sequential, and general processes and their respective dual affine spaces in Fig. C.1.

We now describe a method for obtaining the dual formulation of the SDPs presented in this paper based on the concept of dual affine spaces. In the main text we have defined the primal optimization problem as

$$\begin{aligned}
 & \text{given } \{p_i, C_i\} \\
 & \text{maximize } \sum_i p_i \text{Tr}(T_i^S C_i^{\otimes 2}) \\
 & \text{subject to } \{T_i^S\} \in \mathcal{T}^S,
 \end{aligned} \tag{C.29}$$

where  $\mathcal{T}^S$  is the set of all testers with strategy  $S$ . This problem can also be written as

$$\begin{aligned}
 & \text{given } \{p_i, C_i\} \\
 & \max \sum_i p_i \text{Tr}(T_i^S C_i^{\otimes 2}) \\
 & \text{s.t. } T_i^S \geq 0 \\
 & \quad \sum_i T_i^S \in \mathcal{W}^S
 \end{aligned} \tag{C.30}$$

where  $\mathcal{W}^S$  is set of all processes with strategy  $S$ .

We start this section by considering the above optimization problem for the case where the set  $\mathcal{W}^S$  is affine, which is the case for parallel, sequential, and general processes. For these strategies, we do not need to restrict ourselves to the case of  $k = 2$  copies of the input channel  $C_i$  but the method applies for any  $k \in \mathbb{N}$ . Note that the normalization constraints of separable processes do not form an affine set, for which reason the case of separable testers will be tackled later. We also point that the for  $k > 2$ , the definition of  $k$ -slots separable processes have several nuances and there is still no consensus on a single definition [193].

For finite dimensions, if  $\mathcal{W}$  is an affine set we have that  $\overline{\overline{\mathcal{W}}} = \mathcal{W}$ , i.e., the dual affine space of the dual affine space of  $\mathcal{W}$  is simply  $\mathcal{W}$ . Hence, for cases where  $\mathcal{W}$  is affine, the primal SDP presented in Eq. (C.30) can be written as:

$$\begin{aligned}
 & \text{given } \{p_i, C_i\} \\
 & \max \sum_i p_i \text{Tr}(T_i^S C_i^{\otimes k}) \\
 & \text{s.t. } T_i^S \geq 0 \\
 & \quad W^S := \sum_i T_i^S \\
 & \quad \text{Tr}(W^S \overline{W}^S) = 1, \quad \forall \overline{W}^S \in \overline{\mathcal{W}}^S,
 \end{aligned} \tag{C.31}$$

a formulation which has infinitely many constraints  $\left[ \text{Tr}(W^S \overline{W}^S) = 1, \forall \overline{W}^S \in \overline{\mathcal{W}}^S \right]$ . These infinitely many constraints can be made finite by writing  $\left[ \text{Tr}(W^S \overline{W}_j^S) = 1, \forall j \right]$

where  $\{\bar{W}_j^S\}_j$  is an affine basis for  $\bar{\mathcal{W}}^S$ , i.e., every  $\bar{W}^S \in \bar{\mathcal{W}}^S$  can be written as  $\bar{W}^S = \sum_j w_j \bar{W}_j^S$  for a set of coefficients  $\{w_j\}_j$  respecting  $\sum_j w_j = 1$ . The Lagrangian of the maximization problem can then be written as

$$L = \sum_i p_i \text{Tr}(C_i^{\otimes k} T_i^S) + \sum_i \text{Tr}(T_i^S \Gamma_i) + \sum_{ij} \left[ 1 - \text{Tr}(T_i^S \bar{W}_j^S) \right] \lambda_j. \quad (\text{C.32})$$

Hence, if  $\Gamma_i \geq 0$  and  $\{T_i^S\}_i$  is a tester,  $L \geq \sum_i p_i \text{Tr}(C_i^{\otimes k} T_i^S)$ . By re-arranging terms, the Lagrangian can be written as

$$L = \sum_i \text{Tr} \left[ T_i^S (p_i C_i^{\otimes k} + \Gamma_i) - \left( \sum_j \bar{W}_j^S \lambda_j \right) \right] + \sum_j \lambda_j. \quad (\text{C.33})$$

We then arrive at the dual problem by taking the supremum of the Lagrangian over the primal variables  $\{T_i^S\}_i$  and  $W^S$ . Finally, the solution of the dual problem will be given by the minimization over the dual variables  $\{\Gamma_i\}_i$  and  $\{\lambda_i\}_i$  under the constraint that  $\Gamma_i \geq 0, \forall i$ . The dual problem can be written as

$$\begin{aligned} & \text{given } \{p_i, C_i\} \\ & \text{minimize } \sum_j \lambda_j \\ & \text{s.t. } \Gamma_i \geq 0 \quad \forall i \\ & \quad p_i C_i^{\otimes k} + \Gamma_i + \sum_j \lambda_j \bar{W}_j^S = 0, \quad \forall i \end{aligned} \quad (\text{C.34})$$

Removing the  $\Gamma_i$  dummy variables, we obtain

$$\begin{aligned} & \text{given } \{p_i, C_i\} \\ & \text{min } \sum_j \lambda_j \\ & \text{s.t. } p_i C_i^{\otimes k} \leq \sum_j \lambda_j \bar{W}_j^S \quad \forall i. \end{aligned} \quad (\text{C.35})$$

The requirement of having an affine basis  $\{\bar{W}_j^S\}_j$  can be dropped by defining  $\lambda := \sum_j \lambda_j$  and  $\bar{W}^S := \sum_j \frac{\lambda_j \bar{W}_j^S}{\lambda}$  and noting that, by construction, for any choice of coefficient  $\lambda_j$ ,  $\bar{W}^S$  is an affine combination of the affine basis elements  $\{\bar{W}_j^S\}_i$ , hence  $\bar{W}^S$  necessarily belongs to  $\bar{\mathcal{W}}^S$ . We can then write

**given**  $\{p_i, C_i\}$

$$\begin{aligned} & \min \quad \lambda \\ & \text{s.t.} \quad p_i C_i^{\otimes k} \leq \lambda \bar{W}^S \\ & \quad \bar{W}^S \in \bar{\mathcal{W}}^S, \end{aligned} \tag{C.36}$$

where the dual affine space of the sets used in this work are explicitly presented in Fig. C.1.

Due to the product of variables  $\lambda$  and  $\bar{W}^S$ , the constraint  $p_i C_i^{\otimes k} \leq \lambda \bar{W}^S$  is not linear. This problem can be easily circumvented by noting that the elements of dual affine spaces have a fixed trace  $\text{Tr}(\bar{W}^S)$ . We can then “absorb” the variable  $\lambda$  into  $\bar{W}^S$  by defining  $\bar{W}^S := \lambda W^S$ .

For the case of separable testers, the primal problem can be formulated as

**given**  $\{p_i, C_i\}$

$$\begin{aligned} & \max \quad \sum_i p_i \text{Tr}(T_i^{\text{SEP}} C_i^{\otimes 2}) \\ & \text{s.t.} \quad T_i^{\text{SEP}} \geq 0 \\ & \quad W^{\text{SEP}} := \sum_i T_i^{\text{SEP}} = q W^{1 \leftarrow 2} + (1 - q) W^{2 \leftarrow 1} \\ & \quad W^{1 \leftarrow 2} \in \mathcal{W}^{1 \leftarrow 2}, \quad W^{2 \leftarrow 1} \in \mathcal{W}^{2 \leftarrow 1} \\ & \quad q \in [0, 1], \end{aligned} \tag{C.37}$$

where  $\mathcal{W}^{i \leftarrow j}$  is the set of sequential processes with slot  $i$  coming before slot  $j$ . The SDP described in Eqs. (C.37) can also be written as

**given**  $\{p_i, C_i\}$

$$\begin{aligned} & \max \quad \sum_i p_i \text{Tr}(T_i^{\text{SEP}} C_i^{\otimes 2}) \\ & \text{s.t.} \quad T_i^{\text{SEP}} \geq 0 \\ & \quad \sum_i T_i^{\text{SEP}} = W^{1 \leftarrow 2} + W^{2 \leftarrow 1} \\ & \quad \text{Tr}\left(W^{1 \leftarrow 2} \bar{W}_a^{1 \leftarrow 2}\right) = q, \quad \forall a \\ & \quad \text{Tr}\left(W^{2 \leftarrow 1} \bar{W}_b^{2 \leftarrow 1}\right) = 1 - q, \quad \forall b \\ & \quad W^{1 \leftarrow 2} \geq 0, \quad W^{2 \leftarrow 1} \geq 0 \end{aligned} \tag{C.38}$$

where the set  $\{\bar{W}_l^{i \leftarrow j}\}_l$  is an basis for the dual affine space of ordered processes.

The Lagrangian of the SDP presented in Eqs. (C.38) can be written as

$$L = \sum_i p_i \text{Tr} (C_i^{\otimes 2} T_i^{\text{SEP}}) + \sum_i \text{Tr} (T_i^{\text{SEP}} \Gamma_i) + \sum_i \text{Tr} [(T_i^{\text{SEP}} - W^{1 \leftarrow 2} - W^{2 \leftarrow 1}) H] \\ (C.39)$$

$$+ \sum_a \left[ q - \text{Tr} (W^{1 \leftarrow 2} \bar{W}_a^{1 \leftarrow 2}) \right] \lambda_a^{1 \leftarrow 2} + \sum_b \left[ (1 - q) - \text{Tr} (W^{2 \leftarrow 1} \bar{W}_b^{2 \leftarrow 1}) \right] \lambda_b^{2 \leftarrow 1} \\ (C.40)$$

$$+ \text{Tr}(W^{1 \leftarrow 2} \sigma^{1 \leftarrow 2}) + \text{Tr}(W^{2 \leftarrow 1} \sigma^{2 \leftarrow 1}). \\ (C.41)$$

By re-arranging terms we obtain

$$L = \text{Tr} [T_i^{\text{SEP}} (p_i C_i^{\otimes 2} + \Gamma_i + H)] \\ (C.42)$$

$$+ \text{Tr} [W^{1 \leftarrow 2} (\sigma^{1 \leftarrow 2} - H - \sum_a \bar{W}_a^{1 \leftarrow 2} \lambda_a^{1 \leftarrow 2})] + \text{Tr} [W^{2 \leftarrow 1} (\sigma^{2 \leftarrow 1} - H - \sum_b \bar{W}_b^{2 \leftarrow 1} \lambda_b^{2 \leftarrow 1})] \\ (C.43)$$

$$+ q \left( \sum_a \lambda_a^{1 \leftarrow 2} - \sum_b \lambda_b^{2 \leftarrow 1} \right) + \sum_b \lambda_b^{2 \leftarrow 1}. \\ (C.44)$$

Which leads to the dual problem

$$\begin{aligned} & \text{given } \{p_i, C_i\} \\ & \text{minimize } \sum_b \lambda_b^{2 \leftarrow 1} \\ & \text{s.t. } \Gamma_i \geq 0 \quad \forall i \\ & \quad \sigma^{1 \leftarrow 2} \geq 0, \quad \sigma^{2 \leftarrow 1} \geq 0 \\ & \quad q^0 \geq 0, \quad q^{1 \leftarrow 2} \geq 0 \\ & \quad \Gamma_i = -p_i C_i^{\otimes 2} - H, \quad \forall i \\ & \quad \sigma^{1 \leftarrow 2} = H + \sum_a \bar{W}_a^{1 \leftarrow 2} \lambda_a^{1 \leftarrow 2} \\ & \quad \sigma^{2 \leftarrow 1} = H + \sum_b \bar{W}_b^{2 \leftarrow 1} \lambda_b^{2 \leftarrow 1} \\ & \quad \sum_b \lambda_b^{2 \leftarrow 1} = \sum_a \lambda_a^{1 \leftarrow 2}. \end{aligned} \\ (C.45)$$

By removing the dummy variables we get

$$\begin{aligned}
 & \text{given } \{p_i, C_i\} \\
 & \text{minimize } \sum_b \lambda_b^{2 \prec 1} \\
 & \text{s.t. } p_i C_i^{\otimes 2} \leq -H, \quad \forall i \\
 & \quad -H \leq \sum_a \bar{W}_a^{1 \prec 2} \lambda_a^{1 \prec 2} \\
 & \quad -H \leq \sum_b \bar{W}_b^{2 \prec 1} \lambda_b^{2 \prec 1} \\
 & \quad \sum_b \lambda_b^{2 \prec 1} = \sum_a \lambda_a^{1 \prec 2}.
 \end{aligned} \tag{C.46}$$

As before we define  $\lambda := \sum_b \lambda_b^{2 \prec 1} = \sum_a \lambda_a^{1 \prec 2}$ ,  $\bar{W}^{1 \prec 2} := \sum_a \frac{\lambda_a^{1 \prec 2} \bar{W}_a^{1 \prec 2}}{\lambda}$  and  $\bar{W}^{2 \prec 1} := \sum_b \frac{\lambda_b^{2 \prec 1} \bar{W}_b^{2 \prec 1}}{\lambda}$ , and set  $-H \mapsto H$  to obtain the simplified SDP

$$\begin{aligned}
 & \text{given } \{p_i, C_i\} \\
 & \text{minimize } \lambda \\
 & \text{s.t. } p_i C_i^{\otimes 2} \leq H, \quad \forall i \\
 & \quad H \leq \lambda \bar{W}^{1 \prec 2} \\
 & \quad H \leq \lambda \bar{W}^{2 \prec 1} \\
 & \quad \bar{W}^{1 \prec 2} \in \bar{\mathcal{W}}^{1 \prec 2} \\
 & \quad \bar{W}^{2 \prec 1} \in \bar{\mathcal{W}}^{2 \prec 1}.
 \end{aligned} \tag{C.47}$$

### C.3 Computer-assisted proofs

In this section we provide a general algorithm that can be used to obtain a rigorous computer-assisted proof from numerical optimization packages which may use floating-point variables. Since floating-point variables use approximations to store real numbers, the constraints required by the optimization problem cannot be satisfied exactly. For instance, let  $C_{\text{float}} \in \mathcal{L}(\mathcal{H}^I \otimes \mathcal{H}^O)$  be a matrix with floating-point variables which is certified by a computer to respect the quantum channel constraints, i.e.,

$$C_{\text{float}} \geq 0 \tag{C.48}$$

$${}_O C_{\text{float}} =_{IO} C_{\text{float}} \tag{C.49}$$

$$\text{Tr}(C_{\text{float}}) = d_I. \tag{C.50}$$

Due to floating-point rounding errors, these constraints may be violated in a rigorous analysis, that is, they are satisfied only up to a numerical precision. For this reason,

numerical solutions involving floating-point variables or rounding approximations may lead to accuracy problems [197, 198]. In order to circumvent the floating-point accuracy issue, we provide an algorithm that, given a floating-point variable matrix which satisfies the constraints of a desired set, up to some numerical precision, we construct another matrix which does not make use of floating-point and satisfies the constraints of the desired set exactly. Here, by desired set we refer to six main sets consider in this work: parallel processes, sequential processes, general processes, and their dual affine spaces.

Before proceeding, we present an useful characterization of the aforementioned sets in a unified manner in terms of projections. More precisely, all these sets can be written as:  $C \in \mathcal{L}(\mathcal{H})$  belongs to the desired set  $\mathcal{C} \subseteq \mathcal{L}(\mathcal{H})$  if and only if<sup>2</sup>

$$C \geq 0 \quad (\text{C.51})$$

$$C = \tilde{P}(C) \quad (\text{C.52})$$

$$\text{Tr}(C) = \gamma, \quad (\text{C.53})$$

for a suitable linear space  $\mathcal{H}$ , for some linear projection map  $\tilde{P} : \mathcal{H} \rightarrow \mathcal{H}$ , i.e., some map  $\tilde{P}$  such that  $\tilde{P} \circ \tilde{P} = \tilde{P}$  and  $\tilde{P}(\mathbb{1}) = \mathbb{1}$ , and for some normalization coefficient  $\gamma$ . Here, the set  $\mathcal{C}$  is phrased in such a general way that it covers, for example, the set of quantum states, channels, combs, and processes, among others.

For instance, if the desired set  $\mathcal{C}$  is the set of quantum channels, we have that  $\mathcal{H} = \mathcal{H}_I \otimes \mathcal{H}_O$  and  $C \in \mathcal{C}$  if and only if

$$C \geq 0 \quad (\text{C.54})$$

$$C = \tilde{P}(C) = C -_O C +_{IO} C \quad (\text{C.55})$$

$$\text{Tr}(C) = \gamma = d_I. \quad (\text{C.56})$$

If the desired set is the set of two-slot parallel processes  $\mathcal{W}^{\text{PAR}}$ , we have that  $\mathcal{H} = \mathcal{H}_{I_1} \otimes \mathcal{H}_{O_1} \otimes \mathcal{H}_{I_2} \otimes \mathcal{H}_{O_2}$  and  $W \in \mathcal{W}^{\text{PAR}}$  if and only if

$$W \geq 0 \quad (\text{C.57})$$

$$W = \tilde{P}^{\text{PAR}}(W) =_{O_1 O_2} W \quad (\text{C.58})$$

$$\text{Tr}(W) = \gamma^{\text{PAR}} = d_{O_1} d_{O_2}. \quad (\text{C.59})$$

The projection maps  $\tilde{P}^S$  for the sets of processes  $\mathcal{W}^S$  and  $\tilde{P}^{\mathcal{S}}$  for the sets of dual affine spaces  $\overline{\mathcal{W}}^S$  used in this section are presented in Table C.1.

We now present Algorithm 1, which takes a linear operator  $C_{\text{float}}$  respecting the conditions of a set  $\mathcal{C}$  described by Eqs. (C.51)-(C.53) up to numerical precision and

---

<sup>2</sup>Note that when dual affine spaces are considered, the positivity constraints  $C \geq 0$  is not required.

Processes	
PARALLEL	$\tilde{P}^{\text{PAR}}(W) =_{O_1 O_2} W$
SEQUENTIAL	$\tilde{P}^{\text{SEQ}}(W) =_{O_2} W -_{I_2 O_2} W +_{O_1 I_2 O_2} W$
GENERAL	$\tilde{P}^{\text{GEN}}(W) =_{I_1 O_1 O_2} W -_{I_1 O_1} W +_{O_1 I_2 O_2} W$ $-_{I_2 O_2} W +_{O_1} W +_{O_2} W -_{O_1 O_2} W$
Dual affine space (Channels)	
PARALLEL	$\tilde{\bar{P}}^{\text{PAR}}(\bar{W}) = \bar{W} -_{O_1 O_2} \bar{W} +_{I_1 I_2 O_1 O_2} \bar{W}$
SEQUENTIAL	$\tilde{\bar{P}}^{\text{SEQ}}(\bar{W}) = \bar{W} -_{O_2} \bar{W} +_{I_2 O_2} \bar{W} -_{O_1 I_2 O_2} \bar{W} +_{I_1 O_1 I_2 O_2} \bar{W}$
GENERAL	$\tilde{\bar{P}}^{\text{GEN}}(\bar{W}) = \bar{W} -_{O_1} \bar{W} +_{I_1 O_1} \bar{W} -_{O_2} \bar{W} +_{I_2 O_2} \bar{W}$ $-_{O_1 I_2 O_2} \bar{W} -_{I_1 O_1 O_2} \bar{W} +_{O_1 O_2} \bar{W} +_{I_1 O_1 I_2 O_2} \bar{W}$

Table C.1: Projectors onto the linear space spanned by parallel, sequential, and general processes and their respective dual affine spaces. In all these cases,  $\mathcal{H} = \mathcal{H}_{I_1} \otimes \mathcal{H}_{O_1} \otimes \mathcal{H}_{I_2} \otimes \mathcal{H}_{O_2}$ . In addition, we remark that the trace constraint of Eq. (C.53) for processes is  $\text{Tr}(W) = d_{O_1} d_{O_2}$  and for their dual affine spaces, we have  $\text{Tr}(\bar{W}) = d_{I_1} d_{I_2}$ .

provide an operator  $C_{\text{OK}}$  which respects the conditions of  $\mathcal{C}$  exactly. Also, all the steps of our algorithm can be done without approximations or the use of numerical floating-point variables.

*Algorithm 1:*

1. Construct the non-floating-point matrix  $C_{\text{frac}}$  by truncating the matrix  $C_{\text{float}}$

This allows us to work with fractions and to avoid numerical imprecision.

2. Define the matrix  $C := \frac{C_{\text{frac}} + (C_{\text{frac}})^{\dagger}}{2}$  to obtain a self-adjoint matrix  $C$

Ensures that we are dealing with self-adjoint matrices

3. Project  $C$  into a valid subspace and obtain  $\tilde{P}(C)$

Ensures that the operator is in the valid linear subspace.

4. Find a coefficient  $\eta$  such that  $\tilde{D}_\eta(\tilde{P}(C)) := \eta\tilde{P}(C) + (1 - \eta)\mathbb{1}$  is positive semidefinite

Ensures positivity without leaving the valid subspace.

5. Output the operator  $C_{\text{OK}} = \gamma \frac{\tilde{D}_\eta(\tilde{P}(C))}{\text{Tr}[\tilde{D}_\eta(\tilde{P}(C))]}$  which lies in  $\mathcal{C}$

Ensures the trace condition, preserving positivity and without leaving the valid subspace.

One way to complete step 3 is to start with  $\eta = 1$  and check if the operator  $C$  is already positive semidefinite. If  $C$  is not positive semidefinite, we can slowly decrease the value of  $\eta$  and check if  $\tilde{D}_\eta(C)$  is positive definite. Checking if a matrix is positive semidefinite can be done efficiently by implementing the Cholesky decomposition algorithm and checking whether the algorithm leads to a valid Cholesky decomposition.

One can verify that the operator  $C_{\text{OK}}$  provided by the algorithm described above necessarily belongs to the desired valid set  $\mathcal{S}$  with the aid of the following theorem.

**Theorem C.4.** *Let  $\tilde{P} : \mathcal{L}(\mathcal{H}) \rightarrow \mathcal{L}(\mathcal{H})$  be a linear projector i.e.,  $\tilde{P} \circ \tilde{P} = \tilde{P}$ , which respects  $\tilde{P}(\mathbb{1}) = \mathbb{1}$ . Let  $\tilde{D}_\eta : \mathcal{L}(\mathcal{H}) \rightarrow \mathcal{L}(\mathcal{H})$  be an affine map defined by  $\tilde{D}_\eta(C) := \eta C + (1 - \eta)\mathbb{1}$ . It holds that*

$$\tilde{D}_\eta(\tilde{P}(C)) = \tilde{P}(\tilde{D}_\eta(\tilde{P}(C))) \quad (\text{C.60})$$

*Proof.*

$$\tilde{P}(\tilde{D}_\eta(\tilde{P}(C))) = \tilde{P}(\eta\tilde{P}(C) + (1 - \eta)\tilde{P}(\mathbb{1})) \quad (\text{C.61})$$

$$= \eta\tilde{P}(\tilde{P}(C)) + (1 - \eta)\tilde{P}(\tilde{P}(\mathbb{1})) \quad (\text{C.62})$$

$$= \eta\tilde{P}(C) + (1 - \eta)\mathbb{1} \quad (\text{C.63})$$

$$= \tilde{D}_\eta(\tilde{P}(C)). \quad (\text{C.64})$$

□

Algorithm 1 allows us to obtain upper bounds for the maximal probability of discriminating an ensemble of quantum channels. For the case in which the desired set  $\mathcal{C}$  is the set of dual affine spaces of processes  $\overline{\mathcal{W}}^{\mathcal{S}}$  for some strategy  $\mathcal{S}$ ,  $C_{\text{float}}$  is the floating-point matrix of a dual affine  $\overline{\mathcal{W}}_{\text{float}}^{\mathcal{S}}$ , that can be obtained using numerical convex optimization packages to solve the dual problem SDP, and map  $\tilde{P}$  is one of the projection maps  $\tilde{P}^{\mathcal{S}}$ , then Algorithm 1 will return a matrix  $\overline{W}_{\text{OK}}^{\mathcal{S}}$  that satisfies the constraints of the set  $\overline{\mathcal{W}}^{\mathcal{S}}$  exactly. A rigorous upper bound on the

maximal probability for discriminating the ensemble  $\{p_i, C_i\}_i$  is then given by the value  $p_{\text{upper}}$  such that  $p_i C_i^{\otimes k} \leq p_{\text{upper}} \overline{W_{\text{OK}}}$  for all  $i$ . Note that if the channels  $C_i$  are also represented with floating-point variables, one can also use Algorithm 1 to obtain exact channels  $C_{i,\text{OK}}$ .

In order to calculate lower bounds, we can use the primal SDP to obtain a set of  $\{T_{i,\text{float}}\}_{i=1}^N$  which satisfies the conditions of some desired class of tester up to some numerical precision. To tackle this situation, we present an algorithm to obtain a set of operators  $\{T_{i,\text{OK}}\}_{i=1}^N$  which satisfies the tester constraints exactly. Note that this algorithm also works for positive-operator valued measures (POVMs), instruments, and super-instruments, among others.

**Algorithm 2:**

1. Construct the non-floating-point matrix  $T_{i,\text{frac}}$  by truncating the matrix  $T_{i,\text{float}}$

This allows us to work with fractions and to avoid numerical imprecision.

2. Define the matrices  $T_i := \frac{T_{i,\text{frac}} + (T_{i,\text{frac}})^\dagger}{2}$  to obtain self-adjoint matrices  $T_i$

Ensures that we are dealing with self-adjoint matrices

3. Project  $W := \sum_{i=1}^N T_i$  into a valid subspace and obtain  $\tilde{P}(W)$

Ensures the operator  $W$  is in the valid linear subspace.

4. Define the extra-outcome tester element  $T_\emptyset := \tilde{P}(W) - W$

Useful step to later ensure the normalization constraints.

5. Find a coefficient  $\eta$  such  $\tilde{D}_\eta(T_\emptyset) \geq 0$  and  $\tilde{D}_\eta(T_i) \geq 0$  holds for every  $i$

Ensures positivity of all tester elements.

6. Define  $W_\eta := \left( \sum_{i=1}^N \tilde{D}_\eta(T_i) \right) + \tilde{D}_\eta(T_\emptyset)$

Defines a positive semidefinite operator such that  $W_\eta = \tilde{P}(W_\eta)$

7. Output the set  $T_{\text{OK}} := \left\{ \gamma \frac{\tilde{D}_\eta(T_i) + \frac{\tilde{D}_\eta(T_\emptyset)}{N}}{\text{Tr}(W_\eta)} \right\}_i$  which is a valid tester

Equally distributes the tester element  $\tilde{D}_\eta(T_\emptyset)$  between elements indexed by  $i$ .

Similarly to algorithm 1, one can verify that the set  $T_{\text{OK}}$  is a valid tester.

**Theorem C.5.** *The operator  $T_{\text{OK}}$  defined in step 6 of Algorithm 2 is a valid tester.*

*Proof.* By construction all tester elements

$$T_{i,\text{OK}} := \gamma \frac{\tilde{D}_\eta(T_i) + \frac{\tilde{D}_\eta(T_\emptyset)}{N}}{\text{Tr}(W_\eta)} \quad (\text{C.65})$$

are positive semidefinite, we then need to show that  $W_{\text{OK}} := \sum_i T_{i,\text{OK}}$  respects  $\tilde{P}(W_{\text{OK}}) = W_{\text{OK}}$  and  $\text{Tr}(W_{\text{OK}}) = \gamma$ . For that, note that

$$W_{\text{OK}} = \gamma \sum_{i=1}^N \frac{\tilde{D}_\eta(T_i) + \frac{\tilde{D}_\eta(T_\emptyset)}{N}}{\text{Tr}(W_\eta)} \quad (\text{C.66})$$

$$= \gamma \frac{W_\eta}{\text{Tr}(W_\eta)}. \quad (\text{C.67})$$

We can then guarantee that  $\text{Tr}(W_{\text{OK}}) = \gamma$  and

$$W_\eta = \left[ \sum_{i=1}^N \eta T_i + (1-\eta)\mathbb{1} \right] + \eta T_\emptyset + (1-\eta)\mathbb{1} \quad (\text{C.68})$$

$$= \eta W + (1+\eta)N\mathbb{1} + \eta \tilde{P}(W) - \eta W + (1-\eta)\mathbb{1} \quad (\text{C.69})$$

$$= \eta \tilde{P}(W) + (1-\eta)(N+1)\mathbb{1} \quad (\text{C.70})$$

$$= \eta \tilde{P}(W) + (1-\eta)(N+1)\tilde{P}(\mathbb{1}) \quad (\text{C.71})$$

$$= \tilde{P}(W_\eta). \quad (\text{C.72})$$

□

Algorithm 2 allows us to obtain lower bounds for the maximal probability of discriminating an ensemble of quantum channels. A floating-point set of matrices  $\{T_{i,\text{float}}^S\}_i$ , can be obtained via numerical convex optimization packages to solve the primal problem SDP. Then Algorithm 2 will return a set of matrices  $T_{\text{OK}}^S$  that satisfies the constraints of the set  $\mathcal{T}^S$  exactly. A rigorous lower bound on the maximal probability for discriminating ensemble  $\{p_i, C_i\}_i$  is then given by the value  $p_{\text{lower}} = \sum_{i=1}^N p_i \text{Tr}(C_i^{\otimes k} T_{i,\text{OK}}^S)$ .

We have implemented the algorithms presented in this section and the remaining code necessary for the calculation of the upper- and lower bounds presented in this paper. All code has been uploaded to an online repository [157]. The SDP optimization was implemented in MATLAB™ using the package cvx [199] and tested independently with the solvers MOSEK [200], SeDuMi [201], and SDPT3 [202]. The computer-assisted proof step used to obtain the exact upper and lower bounds was implemented in Mathematica™. All our code can be freely used, edited, and distributed under the MIT license [203].

---

Strategy gap	Number of pairs of channels (out of 100 000)
$P^{\text{PAR}} < P^{\text{SEQ}}$	99 955
$P^{\text{SEQ}} < P^{\text{SEP}}$	99 781
$P^{\text{SEP}} < P^{\text{GEN}}$	94 026
$P^{\text{PAR}} < P^{\text{SEQ}} < P^{\text{SEP}} < P^{\text{GEN}}$	94 015

Table C.2: The first column denotes between which strategies of channel discrimination a gap in performance was found and the second column denotes how many of the 100,000 pairs of channels that were sampled demonstrated such a gap.

#### C.4 Sampling general channels and the typicality of the hierarchy between discrimination strategies

Our method for generating a general channel goes as follows:

1. Fix input dimension  $d_I$  and output dimension  $d_O$ .
2. Uniformly sample a positive semidefinite matrix  $A$  of size  $(d_I d_O)$ -by- $(d_I d_O)$ , according to the Hilbert-Schmidt measure. This can be done, for example, using the function `RandomDensityMatrix` of the freely distributed MATLAB toolbox QETLAB [204].
3. Define  $C$  to be the projection of  $A$  on the subspace of valid quantum channels, according to

$$C = A - \frac{1}{d_O} A + \frac{1}{d_O}. \quad (\text{C.73})$$

4. Check whether  $C$  is a positive semidefinite matrix. If not, discard  $C$  and repeat the process. If yes, than  $C$  represents the Choi operator of a valid quantum channel  $\tilde{C} : \mathcal{L}(\mathcal{H}^I) \rightarrow \mathcal{L}(\mathcal{H}^O)$ .

We have sampled 100,000 pairs of general qubit-qubit channels using this method and computed, using our SDP methods, the maximal probability of discriminating these channels in an ensemble where both channels are equally probable, using parallel, sequential, separable, and general strategies. Our results are summarized in Table C.2. The first column denotes between which strategies a gap was found and the second column denotes how many of the 100,000 pairs of channels had such gap.

In particular, the last line of Table C.2, which shows that a strict hierarchy  $P^{\text{PAR}} < P^{\text{SEQ}} < P^{\text{SEP}} < P^{\text{GEN}}$  between all four strategies was found by 94,015 pairs of channels, implies that our method has around 94% probability of generating a pair of qubit-qubit channels that showcases this phenomenon.

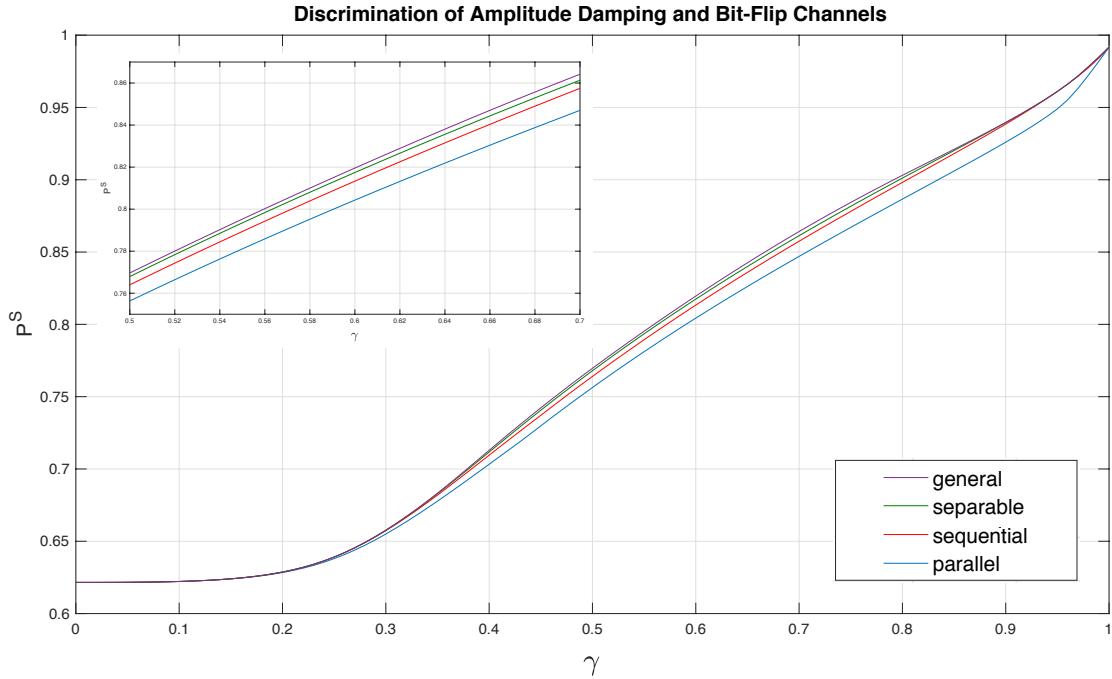


Figure C.2: Probability of successfully discriminating an amplitude damping channel and a bit-flip channel, in an equiprobable ensemble, using  $k = 2$  copies. The value of the decay parameter of the amplitude damping channel varies with the interval  $\gamma \in [0, 1]$ , while the flipping parameter of the bit-flip channel is fixed at  $\eta = 0.87$ . The four curves represent parallel, sequential, separable, and general strategies of channel discrimination. A clear gap between all four strategies is clearly visible in the picture-in-picture plot, with  $\gamma \in [0.5, 0.7]$ .

For the case of discriminating between amplitude damping channels and bit-flip channels, in order to show that the phenomenon of the advantage between different strategies is not unique to a specific choice of parameters, we plot on Fig. C.2 the probability of successful discrimination between an amplitude damping channel with decay parameter  $\gamma \in [0, 1]$  and a bit-flip channel with fixed flipping parameter  $\eta = 0.87$ . A clear gap between all four strategies can be clearly seen on the zoomed picture-in-picture, which plots only  $\gamma \in [0.5, 0.7]$ . Similar plots can be obtained for different values of  $\eta$ .

It is also true that a strict hierarchy between strategies of channel discrimination can be found when discriminating among two amplitude damping channels, in an equiprobable ensemble, with different decay parameters. Using our methods, we have calculated the probability of success for all four strategies, and would like to point out one interesting case of discrimination between one amplitude damping channel with  $\gamma_1 = 0.37$  and another with  $\gamma_2 = 0.87$ , which gives

$$\begin{aligned} \frac{8101}{10000} &< P^{\text{PAR}} < \frac{8102}{10000} \\ &< \frac{8161}{10000} < P^{\text{SEQ}} < \frac{8162}{10000} \\ &\quad < \frac{8166}{10000} < P^{\text{SEP}} < \frac{81665}{100000} \\ &\quad < \frac{8167}{10000} < P^{\text{GEN}} < \frac{8168}{10000}. \end{aligned} \tag{C.74}$$

Here, we confirm that there exists advantage in the discrimination of amplitude damping channels using sequential strategies over parallel strategies. Furthermore, we show that the case of discrimination among two amplitude damping channels is also an example of a complete hierarchy among all four strategies.



# APPENDIX D

## Supplemental Information of Chapter 4

Appendix D.1 presents the proof of Theorem 4.1, Appendix D.2 presents the proof of Examples 4.1 and 4.2, Appendix D.3 presents the proof of Example 4.3, Appendix D.4 presents the proof of Theorem 4.4, and finally Appendix D.5 contains further details on our numerical results.

Some of the sections in the Appendix will make use of the *link product* [205] 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

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

where  $(\cdot)^{T_2}$  stands for partial the 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.

### D.1 Proof of Theorem 4.1

We start this Section with Lemma D.1, which plays a main role in the proof of Theorem 4.1, and may be of independent interest. The theorems presented in this

section employ methods which are similar to the ones in Ref. [164, 206] which exploit the covariance of testers processes to parallelize strategies.

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

$$W^{IO} (\mathbb{1} \otimes U^{\otimes k})^{IO} = (\mathbb{1} \otimes U^{\otimes k})^{IO} W^{IO}, \quad (\text{D.2})$$

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

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

$$\text{Tr}(T_U^{PAR}|U\rangle\langle U|^{\otimes k}) = \text{Tr}(T_U|U\rangle\langle U|^{\otimes k}) \quad \forall U \in \{U\}_U. \quad (\text{D.3})$$

Moreover, this parallel tester can be written as  $T_U^{PAR} = \rho^{I'I} * M_U^{I'O}$  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

$$\rho^{I'I} := \sqrt{W}^{T^{I'I}} |\mathbb{1}\rangle\langle\mathbb{1}|^{I'I} \sqrt{W}^{T^{I'I}}, \quad (\text{D.4})$$

and  $\{M_U\}_U$  is a POVM defined by<sup>1</sup>

$$M_U^{I'O} := \sqrt{W}^{-1^{I'O}} {T_U}^{I'O} \sqrt{W}^{-1^{I'O}}. \quad (\text{D.5})$$

*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 normalisation condition follows from

$$\text{Tr}(\rho) = \text{Tr} \left( \sqrt{W}^{T^{I'I}} |\mathbb{1}\rangle\langle\mathbb{1}|^{I'I} \sqrt{W}^{T^{I'I}} \right) \quad (\text{D.6})$$

$$= \text{Tr} \left( W^{T^{I'I}} |\mathbb{1}\rangle\langle\mathbb{1}|^{I'I} \right) \quad (\text{D.7})$$

$$= \text{Tr} \left( W^{I'I} |\mathbb{1}\rangle\langle\mathbb{1}|^{T^{I'I}} \right) \quad (\text{D.8})$$

$$= \text{Tr} \left( W^{I'I} |\mathbb{1}\rangle\langle\mathbb{1}|^{I'I} \right) \quad (\text{D.9})$$

$$= 1, \quad (\text{D.10})$$

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 recognise that all operators  $M_U$  are compositions of positive semidefinite operators that add up to the identity, according to

---

<sup>1</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  $WW^{-1} =: \Pi_W$  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  $M_U^{I'O} := \sqrt{W}^{-1^{I'O}} {T_U}^{I'O} \sqrt{W}^{-1^{I'O}} + \frac{1}{N}(\mathbb{1} - WW^{-1})$ . With that, the proof written here also applies to the case where the operator  $W$  is not full-rank.

$$\sum_U M_U = \sqrt{W}^{-1} {}^{I' O} \sum_U T_U {}^{I' O} \sqrt{W}^{-1} {}^{I' O} \quad (\text{D.11})$$

$$= \sqrt{W}^{-1} {}^{I' O} W {}^{I' O} \sqrt{W}^{-1} {}^{I' O} \quad (\text{D.12})$$

$$= \mathbb{1} {}^{I' O}. \quad (\text{D.13})$$

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 od a linear map  $\tilde{C} : \mathcal{L}(\mathcal{H}^I) \rightarrow \mathcal{L}(\mathcal{H}^O)$ ,  $\rho^{I'I} \in \mathcal{L}(\mathcal{H}^{I'} \otimes \mathcal{H}^I)$  is a linear operator, it holds that  $\rho^{I'I} * C^{IO} = (\mathbb{1} \otimes \tilde{C}(\rho^{I'I})) {}^{I' O}$ . Also, if a diagonalisable operator  $W^{IO}$  commutes with  $(\mathbb{1} \otimes U^{\otimes k}) {}^{IO}$ , its positive semidefinite square root  $\sqrt{W}$  also commute with <sup>2</sup> $(\mathbb{1} \otimes U^{\otimes k}) {}^{IO}$ , hence we have

$$\sqrt{W} (\mathbb{1} \otimes U^{\otimes k}) = (\mathbb{1} \otimes U^{\otimes k}) \sqrt{W}. \quad (\text{D.14})$$

By taking the complex conjugation on both sides of Eq. (D.14) and exploiting the fact that  $\sqrt{W} = \sqrt{W}^\dagger$  implies  $\sqrt{W}^T = \sqrt{W}^*$ , it holds that

$$\sqrt{W}^T (\mathbb{1} \otimes U^{*\otimes k}) = (\mathbb{1} \otimes U^{*\otimes k}) \sqrt{W}^T. \quad (\text{D.15})$$

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

$$\rho^{I'I} * (|U\rangle\langle U|^{\otimes k T}) {}^{IO} = \rho^{I'I} * (|U^*\rangle\langle U^*|^{\otimes k}) {}^{IO} \quad (\text{D.16})$$

$$= [(\mathbb{1} \otimes U^{*\otimes k}) \rho (\mathbb{1} \otimes U^{T\otimes k})] {}^{I' O} \quad (\text{D.17})$$

$$= [(\mathbb{1} \otimes U^{*\otimes k}) \sqrt{W}^T |\mathbb{1}\rangle\langle \mathbb{1}| \sqrt{W}^T (\mathbb{1} \otimes U^{T\otimes k})] {}^{I' O} \quad (\text{D.18})$$

$$= [\sqrt{W}^T (\mathbb{1} \otimes U^{*\otimes k}) |\mathbb{1}\rangle\langle \mathbb{1}| (\mathbb{1} \otimes U^{T\otimes k}) \sqrt{W}^T] {}^{I' O} \quad (\text{D.19})$$

$$= (\sqrt{W}^T |U^*\rangle\langle U^*|^{\otimes k} \sqrt{W}^T) {}^{I' O}. \quad (\text{D.20})$$

---

<sup>2</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  $B$ .

We now finish the proof by verifying that

$$\mathrm{Tr} (T_U^{\mathrm{PAR}} |U\rangle\langle U|^{\otimes k}) = \mathrm{Tr} \left[ (\rho^{I'I} * M_U^{I'O}) |U\rangle\langle U|^{\otimes k} {}^{IO} \right] \quad (\mathrm{D}.21)$$

$$= (\rho^{I'I} * M_U^{I'O}) * (|U\rangle\langle U|^{\otimes k}) {}^{TIO} \quad (\mathrm{D}.22)$$

$$= M_U^{I'O} * \left( \rho^{I'I} * |U^*\rangle\langle U^*|^{\otimes k} {}^{IO} \right) \quad (\mathrm{D}.23)$$

$$= M_U^{I'O} * \left( \sqrt{W}^T |U^*\rangle\langle U^*|^{\otimes k} \sqrt{W}^T \right) {}^{I'O} \quad (\text{applying Eq. (D.20)}) \quad (\mathrm{D}.24)$$

$$= \mathrm{Tr} \left[ M_U^{I'O} \left( \sqrt{W}^T |U^*\rangle\langle U^*|^{\otimes k} \sqrt{W}^T \right) {}^{TIO} \right] \quad (\mathrm{D}.25)$$

$$= \mathrm{Tr} \left[ M_U^{I'O} \left( \sqrt{W} |U\rangle\langle U|^{\otimes k} \sqrt{W} \right) {}^{I'O} \right] \quad (\mathrm{D}.26)$$

$$= \mathrm{Tr} \left[ \left( \sqrt{W}^{-1} T_U \sqrt{W}^{-1} \right) {}^{I'O} \left( \sqrt{W} |U\rangle\langle U|^{\otimes k} \sqrt{W} \right) {}^{I'O} \right] \quad (\mathrm{D}.27)$$

$$= \mathrm{Tr} (T_U |U\rangle\langle U|^{\otimes k}) . \quad (\mathrm{D}.28)$$

□

Now we prove Theorem 4.1.

**Theorem 4.1.** *For ensembles composed of a uniform probability distribution and a set of unitary channels that form 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  $\{T_i^{\mathrm{GEN}}\}$ , there exists a parallel tester  $\{T_i^{\mathrm{PAR}}\}_i$ , such that

$$\sum_{i=1}^N \mathrm{Tr} (T_i^{\mathrm{PAR}} |U_i\rangle\langle U_i|^{\otimes k}) = \sum_{i=1}^N \mathrm{Tr} (T_i^{\mathrm{GEN}} |U_i\rangle\langle U_i|^{\otimes k}) . \quad (4.12)$$

A set of unitary operators  $\{U\}_U$ ,  $U \in \mathcal{L}(\mathbb{C}_d)$  forms a group up to a global phase if there exist real numbers  $\phi_i$  such that:

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

Before presenting the proof we recall that unitary operators which 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 channel is given by

$$\widetilde{U}'(\rho) = U'\rho U'^\dagger \quad (\text{D.29})$$

$$= e^{i\phi}e^{-i\phi}U\rho U^\dagger \quad (\text{D.30})$$

$$= U\rho U^\dagger \quad (\text{D.31})$$

$$= \widetilde{U}(\rho) \quad (\text{D.32})$$

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

$$|U\rangle\langle U| = |U'\rangle\langle U'|. \quad (\text{D.33})$$

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 will proceed as following, using the general tester  $\{T_U^{\text{GEN}}\}_U$ , we construct another general tester  $\{T_U\}_U$  which obeys

$$\frac{1}{N} \sum_{U \in \{U\}_U} \text{Tr}(T_U^{\text{GEN}} |U\rangle\langle U|^{\otimes k}) = \frac{1}{N} \sum_{U \in \{U\}_U} \text{Tr}(T_U |U\rangle\langle U|^{\otimes k}). \quad (\text{D.34})$$

Then, we prove that the general tester  $\{T_U\}_U$  we defined respects the hypothesis of Lemma D.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:

$$T_U := \frac{1}{N} \sum_{V \in \{U\}_U} \left( \mathbb{1}^I \otimes V^{\dagger \otimes k} \right)^{IO} T_{VU}^{\text{GEN}\, IO} \left( \mathbb{1}^I \otimes V^{\otimes k} \right)^{IO}, \quad (\text{D.35})$$

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. Notice 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

$$W := \sum_U T_U \tag{D.36}$$

$$= \sum_U \frac{1}{N} \sum_V (\mathbb{1} \otimes V^{\dagger \otimes k}) T_{VU}^{\text{GEN}} (\mathbb{1} \otimes V^{\otimes k}) \tag{D.37}$$

$$= \frac{1}{N} \sum_U \sum_V (\mathbb{1} \otimes V^{\dagger \otimes k}) T_{V(V^{-1}U)}^{\text{GEN}} (\mathbb{1} \otimes V^{\otimes k}) \tag{D.38}$$

$$= \frac{1}{N} \sum_U \sum_V (\mathbb{1} \otimes V^{\dagger \otimes k}) T_U^{\text{GEN}} (\mathbb{1} \otimes V^{\otimes k}) \tag{D.39}$$

$$= \frac{1}{N} \sum_V (\mathbb{1} \otimes V^{\dagger \otimes k}) \sum_U T_U^{\text{GEN}} (\mathbb{1} \otimes V^{\otimes k}) \tag{D.40}$$

$$= \frac{1}{N} \sum_V (\mathbb{1} \otimes V^{\dagger \otimes k}) W^{\text{GEN}} (\mathbb{1} \otimes V^{\otimes k}), \tag{D.41}$$

where  $W^{\text{GEN}} := \sum_U T_U^{\text{GEN}}$  and, in Eq. (D.38), 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.

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

$$C'^{IO} := \frac{1}{N} \sum_V (\mathbb{1} \otimes V^{\otimes k})^{IO} C^{IO} (\mathbb{1} \otimes V^{\dagger \otimes k})^{IO}, \tag{D.42}$$

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  $C = \bigotimes_{i=1}^k C_i^{I_i O_i}$ , we have

$$\text{Tr}(W^{IO} C^{IO}) = \frac{1}{N} \text{Tr} \left[ \sum_V (\mathbb{1} \otimes V^{\dagger \otimes k}) W^{\text{GEN}} (\mathbb{1} \otimes V^{\otimes k}) C \right] \tag{D.43}$$

$$= \frac{1}{N} \text{Tr} \left[ W^{\text{GEN}} \sum_V (\mathbb{1} \otimes V^{\otimes k}) C (\mathbb{1} \otimes V^{\dagger \otimes k}) \right] \tag{D.44}$$

$$= \text{Tr}(W^{IO} C'^{IO}) \tag{D.45}$$

$$= 1, \tag{D.46}$$

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,

$$\frac{1}{N} \sum_U \text{Tr}(T_U |U\rangle\langle U|^{\otimes k}) = \frac{1}{N^2} \sum_U \sum_V \text{Tr} \left[ (\mathbb{1} \otimes V^{\dagger \otimes k}) T_{VU}^{\text{GEN}} (\mathbb{1} \otimes V^{\otimes k}) |U\rangle\langle U|^{\otimes k} \right] \quad (\text{D.47})$$

$$= \frac{1}{N^2} \sum_U \sum_V \text{Tr} \left[ T_{VU}^{\text{GEN}} (\mathbb{1} \otimes V^{\otimes k}) |U\rangle\langle U|^{\otimes k} (\mathbb{1} \otimes V^{\dagger \otimes k}) \right] \quad (\text{D.48})$$

$$= \frac{1}{N^2} \sum_U \sum_V \text{Tr} \left( T_{VU}^{\text{GEN}} |VU\rangle\langle VU|^{\otimes k} \right) \quad (\text{D.49})$$

$$= \frac{1}{N^2} \sum_U \sum_V \text{Tr} \left( T_{V(V^{-1}U)}^{\text{GEN}} |V(V^{-1}U)\rangle\langle V(V^{-1}U)|^{\otimes k} \right) \quad (\text{D.50})$$

$$= \frac{1}{N^2} \sum_U \sum_V \text{Tr} \left( T_U^{\text{GEN}} |U\rangle\langle U|^{\otimes k} \right) \quad (\text{D.51})$$

$$= \frac{1}{N} \sum_U \text{Tr} \left( T_U^{\text{GEN}} |U\rangle\langle U|^{\otimes k} \right). \quad (\text{D.52})$$

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

$$(\mathbb{1} \otimes U^{\otimes k}) W (\mathbb{1} \otimes U^{\dagger \otimes k}) = (\mathbb{1} \otimes U^{\otimes k}) \frac{1}{N} \sum_V (\mathbb{1} \otimes V^{\dagger \otimes k}) W^{\text{GEN}} (\mathbb{1} \otimes V^{\otimes k}) (\mathbb{1} \otimes U^{\dagger \otimes k}) \quad (\text{D.53})$$

$$= \frac{1}{N} \sum_V \left( \mathbb{1} \otimes (UV^{\dagger})^{\otimes k} \right) W^{\text{GEN}} \left( \mathbb{1} \otimes (VU^{\dagger})^{\otimes k} \right) \quad (\text{D.54})$$

$$= \frac{1}{N} \sum_V \left[ \mathbb{1} \otimes (U(VU)^{\dagger})^{\otimes k} \right] W^{\text{GEN}} \left[ \mathbb{1} \otimes ((VU)U^{\dagger})^{\otimes k} \right] \quad (\text{D.55})$$

$$= \frac{1}{N} \sum_V (\mathbb{1} \otimes V^{\dagger \otimes k}) W^{\text{GEN}} (\mathbb{1} \otimes V^{\otimes k}) \quad (\text{D.56})$$

$$= W. \quad (\text{D.57})$$

Hence, we have that

$$W^{IO} (\mathbb{1} \otimes U^{\otimes k})^{IO} = (\mathbb{1} \otimes U^{\otimes k})^{IO} W^{IO}, \quad (\text{D.58})$$

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

$$\text{Tr} \left( T_U^{\text{PAR}} |U\rangle\langle U|^{\otimes k} \right) = \text{Tr} \left( T_U |U\rangle\langle U|^{\otimes k} \right) \quad \forall U \in \{U\}_U, \quad (\text{D.59})$$

and therefore, by applying Eq. (D.52), we have

$$\frac{1}{N} \sum_U \text{Tr}\left(T_U^{\text{PAR}} |U\rangle\langle U|^{\otimes k}\right) = \frac{1}{N} \sum_U \text{Tr}\left(T_U^{\text{GEN}} |U\rangle\langle U|^{\otimes k}\right), \quad (\text{D.60})$$

concluding our proof.  $\square$

## D.2 Proof of Examples 4.1 and 4.2

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 start by proving Example 4.1 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 4.1.** *The ensemble composed by a uniform probability distribution and  $N = 4$  qubit-qubit unitary channels given by  $U_1 = \mathbb{1}$ ,  $U_2 = \sigma_x$ ,  $U_3 = \sigma_y$ , and  $U_4 = \sqrt{\sigma_z}$ , in a discrimination task that allows for  $k = 2$  copies, can be discriminated under a sequential strategy success probability  $P^{\text{SEQ}} = 1$  while any parallel strategy copies yields  $P^{\text{PAR}} < 1$ .*

*Proof.* Before starting the proof we note that the four Bell states can be written as:

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

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

$$(\mathbb{1} \otimes \sqrt{\sigma_z}) |\phi^+\rangle = \frac{|00\rangle + i|11\rangle}{\sqrt{2}} \quad (\text{D.61})$$

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  $\mathcal{H}^{\text{aux}}$  to be isomorphic to  $\mathcal{H}^{I_1}$  and prepare the initial state  $\rho \in \mathcal{L}(\mathcal{H}^{I_1} \otimes \mathcal{H}^{\text{aux}})$  as

$$\rho^{I_1 \text{aux}} := |\phi^+\rangle\langle\phi^+|^{I_1 \text{aux}}. \quad (\text{D.62})$$

The state  $\rho^{I_1\text{aux}}$  will then be subjected to the first copy of a unitary operation  $U_i$ , leading to the state

$$(U_i \otimes \mathbb{1}^{\text{aux}}) \rho^{I_1\text{aux}} (U_i \otimes \mathbb{1}^{\text{aux}})^\dagger = (U_i \otimes \mathbb{1}^{\text{aux}}) |\phi^+\rangle\langle\phi^+|^{I_1\text{aux}} (U_i \otimes \mathbb{1}^{\text{aux}})^\dagger. \quad (\text{D.63})$$

We are now in the situation where we can act on the system. We then perform a projective measurement with POVM elements given by

$$M_{\psi^+} := |\psi^+\rangle\langle\psi^+| \quad (\text{D.64})$$

$$M_{\psi^-} := |\psi^-\rangle\langle\psi^-| \quad (\text{D.65})$$

$$M_\phi := |\phi^+\rangle\langle\phi^+| + |\phi^-\rangle\langle\phi^-| \quad (\text{D.66})$$

and Lüders instruments, that is, after the measurements, the output quantum system is given by  $\widetilde{M}_i(\rho) = \sqrt{M_i}\rho\sqrt{M_i}^\dagger$  with probability  $\text{Tr}(\widetilde{M}_i(\rho)) = \text{Tr}(\rho M_i)$ . It can be checked that, if  $U_i = \sigma_x$ , one obtains the outcome associated to  $M_{\psi^+}$  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 to  $M_\phi$ , the unitary  $U_i$  can be either  $\mathbb{1}$  or  $\sqrt{\sigma_z}$ .

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

$$(\mathbb{1} \otimes \mathbb{1})^2 |\phi^+\rangle = |\phi^+\rangle. \quad (\text{D.67})$$

If  $U_i = \sqrt{\sigma_z}$  after the second use of the unitary  $U_i$  the state of the system is

$$(\mathbb{1} \otimes \sqrt{\sigma_z})^2 |\phi^+\rangle = |\phi^-\rangle. \quad (\text{D.68})$$

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  $\tilde{E}$ , 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  $\rho^{I_1\text{aux}} := |\phi^+\rangle\langle\phi^+|^{I_1\text{aux}}$ . Now, instead of using an instrument, we define a second auxiliary  $\mathcal{H}^{\text{aux}'} \cong \mathcal{H}^{I_2} \otimes \mathcal{H}^{\text{aux}}$

and a quantum encoder channel  $\tilde{E} : \mathcal{L}(\mathcal{H}^{O_1} \otimes \mathcal{H}^{\text{aux}}) \rightarrow \mathcal{L}(\mathcal{H}^{I_2} \otimes \mathcal{H}^{\text{aux}} \otimes \mathcal{H}^{\text{aux}'})$  as

$$\begin{aligned}\tilde{E}(\rho) &= \left( \sqrt{M_\phi} \rho \sqrt{M_\phi}^\dagger \right) \otimes M_\phi^{\text{aux}}, \\ &+ \left( \sqrt{M_{\psi^+}} \rho \sqrt{M_{\psi^+}}^\dagger \right) \otimes M_{\psi^+}^{\text{aux}}, \\ &+ \left( \sqrt{M_{\psi^-}} \rho \sqrt{M_{\psi^-}}^\dagger \right) \otimes M_{\psi^-}^{\text{aux}}.\end{aligned}\quad (\text{D.69})$$

We finish our sequential tester construction by presenting quantum measurement given by operators  $N_i \in \mathcal{L}(\mathcal{H}^{O_2} \otimes \mathcal{H}^{\text{aux}} \otimes \mathcal{H}^{\text{aux}'})$ ,

$$N_1 := |\phi^+\rangle\langle\phi^+| \otimes M_\phi \quad (\text{D.70})$$

$$N_2 := |\phi^-\rangle\langle\phi^-| \otimes M_\phi \quad (\text{D.71})$$

$$N_3 := \mathbb{1} \otimes M_{\psi^+} \quad (\text{D.72})$$

$$N_4 := \mathbb{1} \otimes M_{\psi^-}. \quad (\text{D.73})$$

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

In order to show that the probability  $P^{\text{PAR}}$  of discriminating these unitaries with  $k = 2$  copies in a parallel strategy is strictly less than one, we make use of the dual problem formulation of the SDP presented in (4.11). Reference [45] shows that the dual problem formulation for the case of parallel strategies reads

$$\begin{aligned}&\text{given } \{p_i, C_i\} \\&\text{minimize } \lambda \\&\text{subject to } p_i C_i^{\otimes 2} \leq \lambda \bar{W} \quad \forall i,\end{aligned}\quad (\text{D.74})$$

where  $\bar{W} \in \mathcal{L}(\mathcal{H}^I \otimes \mathcal{H}^O)$  is the Choi state of a quantum channel that maps  $\mathcal{H}^I$  to  $\mathcal{H}^O$ , that is,  $\bar{W} \geq 0$  and  $\text{Tr}_O(\bar{W}) = \mathbb{1}^I$ . 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 respect

$$\frac{1}{4} |U_i\rangle\langle U_i|^{\otimes 2} \leq \lambda \bar{W} \quad \text{for } i \in \{1, 2, 3, 4\}. \quad (\text{D.75})$$

Using the computer assisted method presented in Ref. [45], we obtain an operator  $\bar{W}$  which satisfies all the quantum channel conditions exactly and that for  $\lambda = \frac{9741}{1000}$ , the inequality (D.75) holds. Hence,  $P^{\text{PAR}} \leq \frac{9741}{1000}$ . In the online repository in Ref. [171] we present a Mathematica<sup>TM</sup> notebook that can be used to verify that  $\bar{W}$  is a valid Choi state of a quantum channel.  $\square$

We now prove Example 4.2 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. In the following,  $\sigma_x, \sigma_y, \sigma_z$  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 4.2.** Let  $\{U_i\}_i := \{\mathbb{1}, \sigma_x, \sigma_y, \sigma_z, H, \sigma_x H, \sigma_y H, \sigma_z H\}$  be a tuple of eight unitary channels that forms a group up to a global phase, and let  $\{p_i\}_i$  be a tuple of probabilities in which each element  $p_i$  is proportional the  $i$ -th digit of the number  $\pi \approx 3.1415926$ , that is,  $\{p_i\}_i := \{\frac{3}{31}, \frac{1}{31}, \frac{4}{31}, \frac{1}{31}, \frac{5}{31}, \frac{9}{31}, \frac{2}{31}, \frac{6}{31}\}$ . For the ensemble  $\{p_i, U_i\}_i$ , 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. [45]. 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

$$\frac{8196}{10000} < P^{\text{PAR}} < \frac{8197}{10000} < P^{\text{SEQ}} < \frac{8198}{10000}, \quad (\text{D.76})$$

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 in Ref. [171], along with a Mathematica<sup>TM</sup> notebook file which shows that this set of unitaries forms a group.  $\square$

### D.3 Proof of Example 4.3

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 4.3 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. 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  $|_{+P}\rangle := \frac{1}{5}(3|0\rangle + 4|1\rangle)$  and  $|_{-P}\rangle := \frac{1}{5}(4|0\rangle - 3|1\rangle)$ .

**Example 4.3.** For the ensemble composed by a uniform probability distribution and  $N = 4$  qubit-qubit unitary channels given by  $U_1 = \sqrt{\sigma_x}$ ,  $U_2 = \sqrt{\sigma_z}$ ,  $U_3 = \sqrt{H_P}$ , and  $U_4 = \sqrt{H_y}$ , 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 form a strict hierarchy  $P^{\text{PAR}} < P^{\text{SEQ}} < P^{\text{GEN}}$ .

*Proof.* The proof follows from direct application of the computer-assisted methods presented in Ref. [45]. 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. Also, 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 presenting theorem is publicly available at the online repository in Ref. [171].

The computed bounds for the maximal probability of successful discrimination are:

$$\begin{aligned} \frac{9570}{1000} &< P^{\text{PAR}} < \frac{9571}{1000} \\ &< \frac{9876}{1000} < P^{\text{SEQ}} < \frac{9877}{1000} \\ &< \frac{9881}{1000} < P^{\text{GEN}} < \frac{9882}{1000}, \end{aligned} \tag{D.77}$$

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.  $\square$

## D.4 Proof of Theorem 4.4

In this Section, we prove Theorem 4.4 from the main text, which concerns the inability of switch-like strategies to outperform sequential strategies on channel discrimination tasks that involve only unitary operations.

**Theorem 4.4.** *The action of the switch-like process on  $k$  copies of a unitary operation  $U$  can be equivalently described by a sequential process that acts on  $k$  copies of the same unitary operation.*

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

$$\sum_{i=1}^N p_i \text{Tr} (T_i^{\text{SL}} |U_i\rangle\langle U_i|^{\otimes k}) = \sum_{i=1}^N p_i \text{Tr} (T_i^{\text{SEQ}} |U_i\rangle\langle U_i|^{\otimes k}). \tag{4.15}$$

In order to provide a better intuition on this result, before presenting the formal definition of switch-like process with  $k$  slots and proving Theorem 4.4 in full generality, we present a proof for the  $k = 2$  case which is illustrated in Fig. D.1.

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

$$\begin{aligned} \mathcal{W}_{\text{SL}}(U_1, U_2) := & |0\rangle\langle 0|^c \otimes V_{02} (U_2 \otimes \mathbb{1}) V_{01} (U_1 \otimes \mathbb{1}) V_{00} \\ & + |1\rangle\langle 1|^c \otimes V_{12} (U_1 \otimes \mathbb{1}) V_{11} (U_2 \otimes \mathbb{1}) V_{10}, \end{aligned} \quad (\text{D.78})$$

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

$$\begin{aligned} \mathcal{W}_{\text{SL}}(U, U) = & |0\rangle\langle 0|^c \otimes V_{02} (U \otimes \mathbb{1}) V_{01} (U \otimes \mathbb{1}) V_{00} \\ & + |1\rangle\langle 1|^c \otimes V_{12} (U \otimes \mathbb{1}) V_{11} (U \otimes \mathbb{1}) V_{10}. \end{aligned} \quad (\text{D.79})$$

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

$$V_{0i}^{\text{ctrl}} := |0\rangle\langle 0|^c \otimes V_{0i} + |1\rangle\langle 1|^c \otimes \mathbb{1}. \quad (\text{D.80})$$

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

$$V_{1i}^{\text{ctrl}} := |0\rangle\langle 0|^c \otimes \mathbb{1} + |1\rangle\langle 1|^c \otimes V_{1i}. \quad (\text{D.81})$$

We first note that due to 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

$$\begin{aligned} V_{12}^{\text{ctrl}} V_{02}^{\text{ctrl}} (U \otimes \mathbb{1}) \cdot V_{11}^{\text{ctrl}} V_{01}^{\text{ctrl}} (U \otimes \mathbb{1}) \cdot V_{10}^{\text{ctrl}} V_{00}^{\text{ctrl}} = & \\ = & (|0\rangle\langle 0|^c \otimes V_{02} + |1\rangle\langle 1|^c \otimes V_{12}) (U \otimes \mathbb{1}) \cdot \\ & (|0\rangle\langle 0|^c \otimes V_{01} + |1\rangle\langle 1|^c \otimes V_{11}) (U \otimes \mathbb{1}) \cdot \\ & (|0\rangle\langle 0|^c \otimes V_{00} + |1\rangle\langle 1|^c \otimes V_{10}) (U \otimes \mathbb{1}) \quad (\text{D.82}) \\ = & \mathcal{W}_{\text{SL}}(U, U). \quad (\text{D.83}) \end{aligned}$$

This shows that, when  $U_1 = U_2 = U$ , a 2-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. D.1 for an illustration.

**Definition D.1** (Switch-like superchannel). *Let  $\{\pi\}_\pi$ ,  $\pi \in \{0, \dots, k! - 1\}$  be a set where 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  $\{U_i\}_{i=1}^k$ ,  $U_i : \mathcal{H}^{I_i} \rightarrow \mathcal{H}^{O_i}$  as

$$\begin{aligned} \mathcal{W}_{SL}(U_1, \dots, U_k) &:= \\ &\sum_{\pi=0}^{k!-1} |\pi\rangle\langle\pi| \otimes [V_{\pi k} (U_{\sigma_\pi(k)} \otimes \mathbb{1}^{aux}) V_{\pi(k-1)} (U_{\sigma_\pi(k-1)} \otimes \mathbb{1}^{aux}) V_{\pi(k-2)} \dots (U_{\sigma_\pi(1)} \otimes \mathbb{1}^{aux}) V_{\pi 0}], \end{aligned} \quad (\text{D.84})$$

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

$$V_{\pi 0} : \mathcal{H}^{P_t} \otimes \mathcal{H}^{aux} \rightarrow \mathcal{H}^{I_{\sigma_\pi(1)}} \otimes \mathcal{H}^{aux} \quad (\text{D.85})$$

$$V_{\pi n} : \mathcal{H}^{I_{\sigma_\pi(i)}} \otimes \mathcal{H}^{aux} \rightarrow \mathcal{H}^{O_{\sigma_\pi(i+1)}} \otimes \mathcal{H}^{aux} \quad \text{for } n \in \{1, \dots, k-1\} \quad (\text{D.86})$$

$$V_{\pi k} : \mathcal{H}^{I_{\sigma_\pi(k)}} \otimes \mathcal{H}^{aux} \rightarrow \mathcal{H}^{F_t} \otimes \mathcal{H}^{aux} \quad (\text{D.87})$$

Here we have defined the switch-like superchannel only by its action on unitary operations, without explicitly stating how the switch-like superchannel acts on general quantum operations nor its process  $W^{\text{SL}} \in \mathcal{L}(\mathcal{H}^P \otimes \mathcal{H}^I \otimes \mathcal{H}^O \otimes \mathcal{H}^F)$ . In order to prove Theorem 4.4, and for the main purpose of this paper, knowing the action of switch-like superchannels only on unitary operations 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  $W^{\text{SL}} := |U_{\text{SL}}\rangle\langle U_{\text{SL}}|$  where

$$U_{\text{SL}} := \bigoplus_{\pi} V_{\pi k} V_{\pi k-1} \dots V_{\pi 1} V_{\pi 0}. \quad (\text{D.88})$$

Following Lemma 1 in Ref. [56] (see also Theorem 2 of Ref. [54]), one can verify that the process  $W^{\text{SL}}$  acts on unitary operators accordingly to the switch-like superchannel, as presented on Definition D.1.

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

$$V_n^{\text{ctrl}} := \sum_{\pi=0}^{k!-1} |\pi\rangle\langle\pi| \otimes V_{\pi n} \quad \forall n \in \{0, \dots, k\}, \quad (\text{D.89})$$

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

$$V_k^{\text{ctrl}} (U \otimes \mathbb{1}) V_{(k-1)}^{\text{ctrl}} (U \otimes \mathbb{1}) V_{(k-2)}^{\text{ctrl}} \dots (U \otimes \mathbb{1}) V_0^{\text{ctrl}} = \mathcal{W}_{SL}(U, \dots, U). \quad (\text{D.90})$$

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

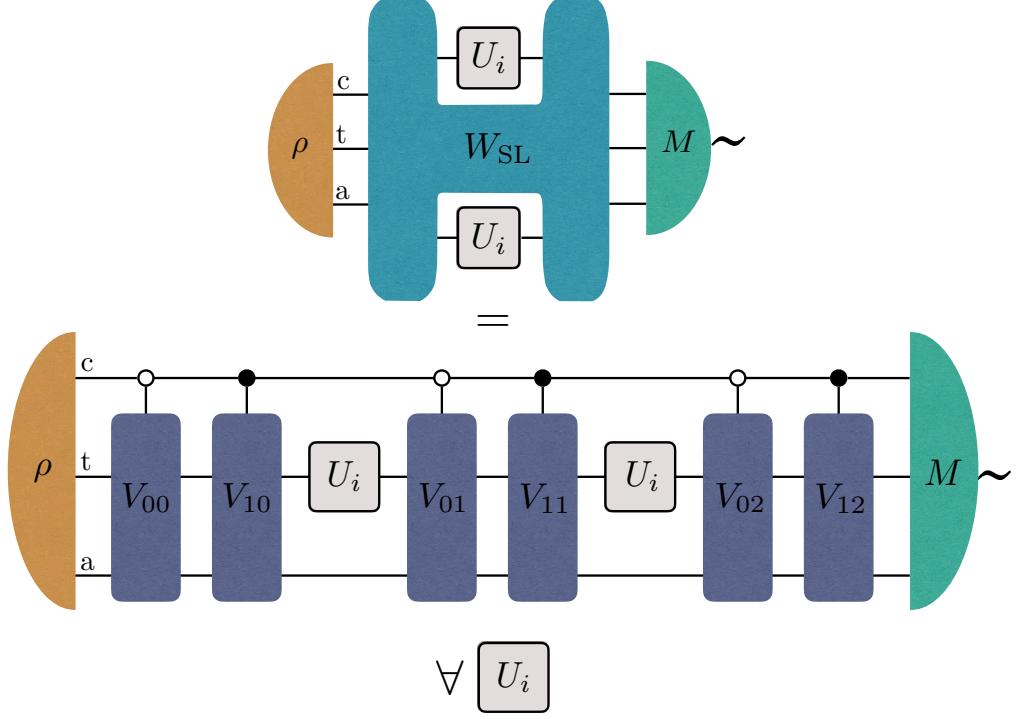


Figure D.1: A 2-copy sequential strategy, that attains the same probability of successful discrimination of any 2-copy switch-like strategy, for all sets of unitaries  $\{U_i\}_{i=1}^N$ . Line “c” represents a control system, “t”, a target system, and “a”, an auxiliary system. Both strategies can be straightforwardly extended to  $k$  copies.

Since every sequential quantum circuit can be written as an ordered process  $W^{\text{SEQ}} \in \mathcal{L}(\mathcal{H}^P \otimes \mathcal{H}^I \otimes \mathcal{H}^O \otimes \mathcal{H}^F)$  [205], when  $k$  identical unitary operators  $U$  are plugged into the process  $W^{\text{SEQ}}$ , the output operation is described by

$$W^{\text{SEQ}} * |U\rangle\langle U|^{\otimes k} = W^{\text{SL}} * |U\rangle\langle U|^{\otimes k}, \quad (\text{D.91})$$

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

$$T_i^{\text{SL}} := \text{Tr}_{PF}[(\rho \otimes \mathbb{1})W^{\text{SL}}(\mathbb{1} \otimes M_i)] \quad (\text{D.92})$$

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

$$T_i^{\text{SEQ}} = \text{Tr}_{PF}[(\rho \otimes \mathbb{1})W^{\text{SEQ}}(\mathbb{1} \otimes M_i)] \quad (\text{D.93})$$

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

$$\text{Tr}(T_i^{\text{SL}} |U\rangle\langle U|^{\otimes k}) = \text{Tr}(T_i^{\text{SEQ}} |U\rangle\langle U|^{\otimes k}), \quad (\text{D.94})$$

ensuring that there is always a sequential tester which performs as well as any switch-like one.  $\square$

Uniformly sampling qubit-qubit unitary channels		
$N$	$k = 2$	$k = 3$
2	$\mathbf{P}^{\text{PAR}} = \mathbf{P}^{\text{SEQ}} = P^{\text{GEN}}$	$\mathbf{P}^{\text{PAR}} = \mathbf{P}^{\text{SEQ}} = P^{\text{GEN}}$
3	$P^{\text{PAR}} = P^{\text{SEQ}} = P^{\text{GEN}}$	$P^{\text{PAR}} < P^{\text{SEQ}} = P^{\text{GEN}}$
4	$P^{\text{PAR}} < P^{\text{SEQ}} = P^{\text{GEN}}$	$P^{\text{PAR}} < P^{\text{SEQ}} < P^{\text{GEN}}$
$\vdots$	$\vdots$	$\vdots$
9	$P^{\text{PAR}} < P^{\text{SEQ}} = P^{\text{GEN}}$	$P^{\text{PAR}} < P^{\text{SEQ}} < P^{\text{GEN}}$
$\vdots$	$\vdots$	
25	$P^{\text{PAR}} \approx P^{\text{SEQ}} = P^{\text{GEN}}$	

Table 4.1: Summary of numerical findings.  $N$  denotes the number of unitary channels in the ensemble and  $k$  denotes the number of copies. The bold equalities on row  $N = 2$  mark analytical results [143]. A strict inequality between the maximal probabilities of success of different strategies in a certain scenario indicates that examples of ensembles that exhibit such gap were encountered. An equality indicates that, for all sampled ensembles, the maximal probabilities of success of different strategies were equal, up to numerical precision.

## D.5 Numerical Findings

In this section we elaborate further on our numerical analysis.

In order to find more examples of the advantage of sequential and general strategies for unitary channel discrimination, we applied the following sampling method. To create ensembles  $\{p_i, U_i\}_{i=1}^N$ , with  $p_i = 1/N \forall i$ , we uniformly sampled sets of up to  $N$  unitary operators according to the Haar measure. Most ensembles were composed with qubit unitaries, although some qutrit examples were also analysed.

Then, we evaluated semidefinite programmes (SDP) to compute the maximal probability of successful discrimination of each ensemble under parallel, sequential, and general strategies, using  $k = 2$  and  $k = 3$  copies. We analysed ensembles of up to  $N = 25$  unitaries.

For qubit unitaries, and  $k = 2$  copies, ensembles of  $N = 2$  and  $N = 3$  unitaries were sampled 50 000 times, ensembles of  $N = 4$  and  $N = 5$  unitaries were sampled 25 000 times, and ensembles of between  $N = 6$  and  $N = 25$  unitaries were sampled 1 000 times. Still for qubit unitaries but for  $k = 3$  copies, ensembles of  $N = 2$  unitaries were sampled 10 000 times, ensembles of  $N = 3$  unitaries were sampled 5 000 times, and ensembles of between  $N = 4$  and  $N = 9$  unitaries were sampled 500 times.

For  $k = 2$ , the maximal probability of success of general and sequential strategies

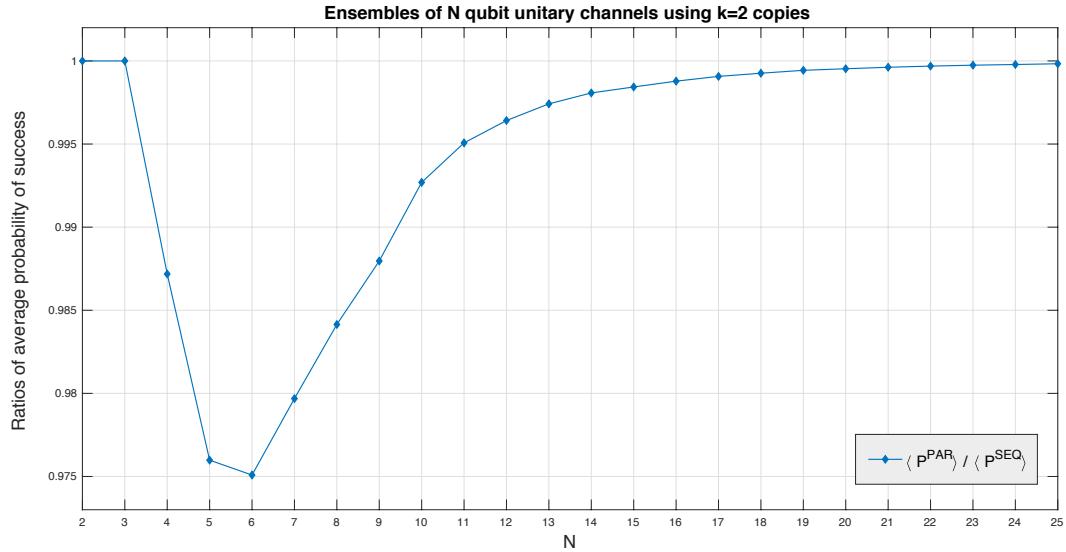


Figure D.2: Ratios of the averages of the maximal probability of success of parallel and sequential strategies using  $k = 2$  copies. Ensembles of  $N \in \{2, \dots, 25\}$  qubit unitaries.

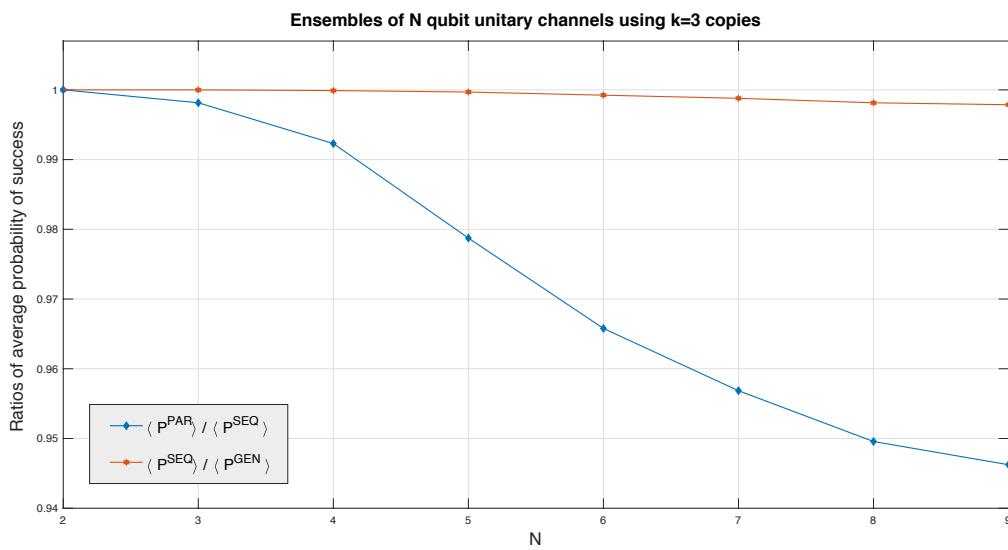


Figure D.3: Ratios of the averages of the maximal probability of success of parallel and sequential strategies, and of sequential and general strategies, using  $k = 3$  copies. Ensembles of  $N \in \{2, \dots, 9\}$  qubit unitaries.

coincided, up to numerical precision, for *all* ensembles. The same was true between parallel and sequential strategies for all ensembles of  $N = 3$  qubit unitaries. For  $k = 3$ , sequential and general strategies coincided for all ensembles of  $N = 2$  and  $N = 3$  unitaries qubit. In all other scenarios, gaps were observed.

We then calculated the averages of the maximal probability of success for each  $N$ ,  $k$ , and strategy. We plot the ratio of these averages for  $k = 2$  in Fig. D.2 and for  $k = 3$  in Fig. D.3. In Fig D.2, only the  $\langle P^{\text{PAR}} \rangle / \langle P^{\text{SEQ}} \rangle$  ratio is plotted, while  $\langle P^{\text{SEQ}} \rangle / \langle P^{\text{GEN}} \rangle$  is omitted, since  $\langle P^{\text{SEQ}} \rangle$  and  $\langle P^{\text{GEN}} \rangle$  coincide. In Fig. D.3, both ratios are plotted.

The qutrit unitary results are not plotted. Only scenarios of  $k = 2$  were analysed, with  $N = 2$  and  $N = 3$ , each sampled 350 and 100 times, respectively. A gap between sequential and separable strategies was found for  $N = 3$ , and no gaps between parallel, sequential, and general strategies was found in the other cases.

These results are also summarized in Table 4.1 in the main text, repeated here for convenience of the reader.



# Bibliography

- [1] F. Jazaeri, A. Beckers, A. Tajalli, and J. Sallese, A review on quantum computing: From qubits to front-end electronics and cryogenic mosfet physics, *IEEE* , 15–25 (2019), arXiv:1908.02656 [quant-ph].
- [2] N. Gisin and R. Thew, Quantum communication, *Nature Photonics* **1**, 165 (2007), arXiv:quant-ph/0703255.
- [3] H.-K. Lo, M. Curty, and K. Tamaki, Secure quantum key distribution, *Nature Photonics* **8**, 595 (2014), arXiv:1505.05303 [quant-ph].
- [4] U. Vazirani and T. Vidick, Fully device-independent quantum key distribution, *Phys. Rev. Lett.* **113**, 140501 (2014), arXiv:1210.1810 [quant-ph].
- [5] C. Hempel *et al.*, Quantum chemistry calculations on a trapped-ion quantum simulator, *Phys. Rev. X* **8**, 031022 (2018), arXiv:1803.10238 [quant-ph].
- [6] M. K. Joshi *et al.*, Quantum information scrambling in a trapped-ion quantum simulator with tunable range interactions, *Phys. Rev. Lett.* **124**, 240505 (2020), arXiv:2001.02176 [quant-ph].
- [7] M. Krenn, J. Handsteiner, M. Fink, R. Fickler, R. Ursin, M. Malik, and A. Zeilinger, Twisted light transmission over 143 km, *Proceedings of the National Academy of Sciences* **113**, 13648–13653 (2016), arXiv:1606.01811 [quant-ph].
- [8] S. Wengerowsky *et al.*, Passively stable distribution of polarisation entanglement over 192 km of deployed optical fibre, *npj Quantum Information* , 5 (2020), arXiv:1907.04864 [quant-ph].
- [9] F. Steinlechner, S. Ecker, M. Fink, B. Liu, J. Bavaresco, M. Huber, T. Scheidl, and R. Ursin, Distribution of high-dimensional entanglement via an intra-city free-space link, *Nat. Commun.* **8**, 15971 (2017), arXiv:1612.00751 [quant-ph].
- [10] E. Anisimova *et al.*, A low-noise single-photon detector for long-distance free-space quantum communication, arXiv eprint (2020), arXiv:2007.12768 [quant-ph].
- [11] S. Philipp *et al.*, Q3Sat: quantum communications uplink to a 3U CubeSat—feasibility & design, *EPJ Quantum Technology* , 4 (2018), arXiv:1711.03409 [quant-ph].
- [12] S.-K. Liao *et al.*, Satellite-relayed intercontinental quantum network, *Phys. Rev. Lett.* **120**, 030501 (2018), arXiv:1801.04418 [quant-ph].
- [13] J. L. O’Brien, A. Furusawa, and J. Vučković, Photonic quantum technologies, *Nature Photonics* **3**, 687 (2009), arXiv:1003.3928 [quant-ph].

- [14] J. Wang, F. Sciarrino, A. Laing, and M. G. Thompson, Integrated photonic quantum technologies, *Nature Photonics* **14**, 273 (2020), arXiv:2005.01948 [quant-ph].
- [15] F. Schäfer, T. Fukuhara, S. Sugawa, Y. Takasu, and Y. Takahashi, Tools for quantum simulation with ultracold atoms in optical lattices, *Nature Reviews Physics* **2**, 411 (2020), arXiv:2006.06120 [quant-ph].
- [16] J. Wang *et al.*, Multidimensional quantum entanglement with large-scale integrated optics, *Science* **360**, 285–291 (2018), arXiv:1803.04449 [quant-ph].
- [17] K. S. Kumar, A. Vepsäläinen, S. Danilin, and G. S. Paraoanu, Stimulated raman adiabatic passage in a three-level superconducting circuit, *Nat. Commun.* **7**, 10628 (2016), arXiv:1508.02981 [quant-ph].
- [18] C. Schaeff, R. Polster, M. Huber, S. Ramelow, and A. Zeilinger, Experimental access to higher-dimensional entangled quantum systems using integrated optics, *Optica* **2**, 523–529 (2015), arXiv:1502.06504 [quant-ph].
- [19] A. J. Gutiérrez-Esparza *et al.*, Detection of nonlocal superpositions, *Phys. Rev. A* **90**, 032328 (2014).
- [20] M. Krenn, M. Malik, M. Erhard, and A. Zeilinger, Orbital angular momentum of photons and the entanglement of Laguerre–Gaussian modes, *Phil. Trans. R. Soc. A* **375**, 20150442 (2017), arXiv:1607.05114 [quant-ph].
- [21] A. Vaziri, G. Weihs, and A. Zeilinger, Experimental two-photon, three-dimensional entanglement for quantum communication, *Phys. Rev. Lett.* **89**, 240401 (2002).
- [22] G. G. Kulkarni, R. Sahu, O. S. Magana-Loaiza, R. W. Boyd, and A. Jha, Single-shot measurement of the orbital-angular-momentum spectrum of light, *Nat. Comm.* **8**, 1054 (2017), arXiv:1401.3512 [quant-ph].
- [23] H. de Riedmatten, I. Marcikic, H. Zbinden, and N. Gisin, Creating high dimensional time-bin entanglement using mode-locked lasers, *Quant. Inf. Comp.* **2**, 425–433 (2002), arXiv:quant-ph/0204165.
- [24] R. T. Thew, A. Acín, H. Zbinden, and N. Gisin, Bell-type test of energy-time entangled qutrits, *Phys. Rev. Lett.* **93**, 010503 (2004), arXiv:quant-ph/0402048.
- [25] A. Martin *et al.*, Quantifying photonic high-dimensional entanglement, *Phys. Rev. Lett.* **118**, 110501 (2017), arXiv:1701.03269 [quant-ph].
- [26] A. K. Jha, M. Malik, and R. W. Boyd, Exploring Energy-Time Entanglement Using Geometric Phase, *Phys. Rev. Lett.* **101**, 180405 (2008), arXiv:1711.05079 [quant-ph].
- [27] F. Steinlechner *et al.*, Distribution of high-dimensional entanglement via an intra-city free-space link, *Nat. Commun.* **8**, 15971 (2017), arXiv:1612.00751 [quant-ph].
- [28] J. T. Barreiro, N. K. Langford, N. A. Peters, and P. G. Kwiat, Generation of Hyperentangled Photon Pairs, *Phys. Rev. Lett.* **95**, 260501 (2005), arXiv:quant-ph/0507128.

- [29] N. J. Cerf, M. Bourennane, A. Karlsson, and N. Gisin, Security of quantum key distribution using  $d$ -level systems, *Phys. Rev. Lett.* **88**, 127902 (2002), arXiv:quant-ph/0107130.
- [30] J. Barrett, A. Kent, and S. Pironio, Maximally nonlocal and monogamous quantum correlations, *Phys. Rev. Lett.* **97**, 170409 (2006), arXiv:quant-ph/0605182.
- [31] S. Gröblacher, T. Jennewein, A. Vaziri, G. Weihs, and A. Zeilinger, Experimental quantum cryptography with qutrits, *New J. Phys.* **8**, 75 (2006), arXiv:quant-ph/0511163.
- [32] R. Horodecki, P. Horodecki, M. Horodecki, and K. Horodecki, Quantum entanglement, *Rev. Mod. Phys.* **81**, 865–942 (2009), arXiv:quant-ph/0702225.
- [33] C. Eltschka and J. Siewert, Quantifying entanglement resources, *J. Phys. A: Math. Theor.* **47**, 424005 (2014), arXiv:1402.6710 [quant-ph].
- [34] M. Huber and M. Pawłowski, Weak randomness in device independent quantum key distribution and the advantage of using high dimensional entanglement, *Phys. Rev. A* **88**, 032309 (2013), arXiv:1301.2455 [quant-ph].
- [35] I. Gerhardt, Q. Liu, A. Lamas-Linares, J. Skaar, C. Kurtsiefer, and V. Makarov, Full-field implementation of a perfect eavesdropper on a quantum cryptography system, *Nature Communications* **2**, 349 (2011), arXiv:1011.0105 [quant-ph].
- [36] M. Krenn, M. Huber, R. Fickler, R. Lapkiewicz, S. Ramelow, and A. Zeilinger, Generation and confirmation of a (100x100)-dimensional entangled quantum system, *P. Natl. Acad. Sci. USA* **111**, 6243–6247 (2014), arXiv:1306.0096 [quant-ph].
- [37] M. Erhard, M. Malik, M. Krenn, and A. Zeilinger, Experimental GHZ entanglement beyond qubits (2017), arXiv:1708.03881 [quant-ph].
- [38] A. C. Dada, J. Leach, G. S. Buller, M. J. Padgett, and E. Andersson, Experimental high-dimensional two-photon entanglement and violations of generalised bell inequalities, *Nat. Phys.* **7**, 677–680 (2011), arXiv:1104.5087 [quant-ph].
- [39] M. Kues *et al.*, On-chip generation of high-dimensional entangled quantum states and their coherent control, *Nature* **546**, 622 (2017).
- [40] J. Bavaresco, N. Herrera Valencia, C. Klöckl, M. Pivoluska, P. Erker, N. Friis, M. Malik, and M. Huber, Measurements in two bases are sufficient for certifying high-dimensional entanglement, *Nature Physics* **14**, 1032 (2018), arXiv:1709.07344 [quant-ph].
- [41] G. Chiribella, G. M. D’Ariano, and P. Perinotti, Transforming quantum operations: Quantum supermaps, *Europhysics Letters* **83**, 30004 (2008), arXiv:0804.0180 [quant-ph].
- [42] G. Chiribella, G. M. D’Ariano, and P. Perinotti, Theoretical framework for quantum networks, *Phys. Rev. A* **80**, 022339 (2009), arXiv:0904.4483 [quant-ph].
- [43] 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 [quant-ph].
- [44] O. Oreshkov, F. Costa, and Č. Brukner, Quantum correlations with no causal order, *Nat. Commun.* **3**, 1092 (2012), arXiv:1105.4464 [quant-ph].

- [45] J. Bavaresco, M. Murao, and M. T. Quintino, Strict hierarchy between parallel, sequential, and indefinite-causal-order strategies for channel discrimination, arXiv preprint (2020), [arXiv:2011.08300 \[quant-ph\]](https://arxiv.org/abs/2011.08300).
- [46] M. Araújo, C. Branciard, F. Costa, A. Feix, C. Giarmatzi, and Č. Brukner, Witnessing causal nonseparability, *New J. Phys.* **17**, 102001 (2015), [arXiv:1506.03776 \[quant-ph\]](https://arxiv.org/abs/1506.03776).
- [47] J. Wechs, H. Dourdent, A. A. Abbott, and C. Branciard, Quantum circuits with classical versus quantum control of causal order, arXiv preprint (2021), [arXiv:2101.08796 \[quant-ph\]](https://arxiv.org/abs/2101.08796).
- [48] T. Purves and A. J. Short, Quantum theory cannot violate a causal inequality, arXiv e-prints (2021), [arXiv:2101.09107 \[quant-ph\]](https://arxiv.org/abs/2101.09107).
- [49] Wikipedia, [https://en.wikipedia.org/wiki/Arago\\_spot#History](https://en.wikipedia.org/wiki/Arago_spot#History).
- [50] Wikipedia, [https://en.wikipedia.org/wiki/Dirac\\_sea#Origins](https://en.wikipedia.org/wiki/Dirac_sea#Origins).
- [51] Wikipedia, [https://en.wikipedia.org/wiki/Gravitational\\_lens#History](https://en.wikipedia.org/wiki/Gravitational_lens#History).
- [52] Wikipedia, [https://en.wikipedia.org/wiki/Gravitational\\_wave#History](https://en.wikipedia.org/wiki/Gravitational_wave#History).
- [53] Wikipedia, [https://en.wikipedia.org/wiki/Higgs\\_boson#History](https://en.wikipedia.org/wiki/Higgs_boson#History).
- [54] 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 \[quant-ph\]](https://arxiv.org/abs/1611.08535).
- [55] O. Oreshkov, Time-delocalized quantum subsystems and operations: on the existence of processes with indefinite causal structure in quantum mechanics, *Quantum* **3**, 206 (2019), [arXiv:1801.07594 \[quant-ph\]](https://arxiv.org/abs/1801.07594).
- [56] W. Yokojima, M. T. Quintino, A. Soeda, and M. Murao, Consequences of preserving reversibility in quantum superchannels, arXiv preprint (2020), [arXiv:2003.05682 \[quant-ph\]](https://arxiv.org/abs/2003.05682).
- [57] G. Chiribella, Perfect discrimination of no-signalling channels via quantum superposition of causal structures, *Phys. Rev. A* **86**, 040301 (2012), [arXiv:1109.5154 \[quant-ph\]](https://arxiv.org/abs/1109.5154).
- [58] 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 \[quant-ph\]](https://arxiv.org/abs/1508.07840).
- [59] 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 \[quant-ph\]](https://arxiv.org/abs/1605.07372).
- [60] M. Araújo, F. Costa, and Č. Brukner, Computational advantage from quantum-controlled ordering of gates, *Phys. Rev. Lett.* **113**, 250402 (2014), [arXiv:1401.8127 \[quant-ph\]](https://arxiv.org/abs/1401.8127).
- [61] 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), [arXiv:1810.06944 \[quant-ph\]](https://arxiv.org/abs/1810.06944).
- [62] E. Prugovečki, Information-theoretical aspects of quantum measurement, *International Journal of Theoretical Physics* **16**, 321–331 (1977).

- [63] C. M. Caves, C. A. Fuchs, and R. Schack, Unknown quantum states: The quantum de Finetti representation, *Journal of Mathematical Physics* **43**, 4537–4559 (2002), arXiv:quant-ph/0104088.
- [64] G. M. D’Ariano, M. de Laurentis, M. G. A. Paris, A. Porzio, and S. Solimeno, Quantum tomography as a tool for the characterization of optical devices, *Journal of Optics B: Quantum and Semiclassical Optics* **4**, S127 (2002), arXiv:quant-ph/0110110.
- [65] M. Horodecki, P. Horodecki, and R. Horodecki, Separability of mixed states: necessary and sufficient conditions, *Phys. Lett. A* **223**, 1–8 (1996), arXiv:quant-ph/9605038.
- [66] C. Branciard, Witnesses of causal nonseparability: an introduction and a few case studies, *Scientific Reports* **6**, 26018 (2016), arXiv:1603.00043 [quant-ph].
- [67] J. S. Bell, On the Einstein Podolsky Rosen paradox, *Physics* **1**, 195–200 (1964).
- [68] N. Brunner, D. Cavalcanti, S. Pironio, V. Scarani, and S. Wehner, Bell nonlocality, *Rev. Mod. Phys.* **86**, 419–478 (2014), arXiv:1303.2849 [quant-ph].
- [69] 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].
- [70] O. Oreshkov and C. Giarmatzi, Causal and causally separable processes, *New J. Phys.* **18**, 093020 (2016), arXiv:1506.05449 [quant-ph].
- [71] L. M. Procopio, A. Moqanaki, M. Araújo, F. Costa, I. Alonso Calafell, E. G. Dowd, D. R. Hamel, L. A. Rozema, Č. Brukner, and P. Walther, Experimental superposition of orders of quantum gates, *Nat. Commun.* **6**, 7913 (2015), arXiv:1412.4006 [quant-ph].
- [72] G. Rubino, L. A. Rozema, A. Feix, M. Araújo, J. M. Zeuner, L. M. Procopio, Č. Brukner, and P. Walther, Experimental verification of an indefinite causal order, *Science Advances* **3**, 3 (2017), arXiv:1608.01683 [quant-ph].
- [73] G. Rubino, L. A. Rozema, F. Massa, M. Araújo, M. Zych, Č. Brukner, and P. Walther, Experimental entanglement of temporal orders, *arXiv e-prints* (2017), arXiv:1712.06884 [quant-ph].
- [74] K. Goswami, C. Giarmatzi, M. Kewming, F. Costa, C. Branciard, J. Romero, and A. G. White, Indefinite causal order in a quantum switch, *Phys. Rev. Lett.* **121**, 090503 (2018), arXiv:1803.04302 [quant-ph].
- [75] K. Goswami, Y. Cao, G. A. Paz-Silva, J. Romero, and A. G. White, Increasing communication capacity via superposition of order, *Phys. Rev. Research* **2**, 033292 (2020), arXiv:1807.07383 [quant-ph].
- [76] K. Wei *et al.*, Experimental quantum switching for exponentially superior quantum communication complexity, *Phys. Rev. Lett.* **122**, 120504 (2019), arXiv:1810.10238 [quant-ph].
- [77] Y. Guo, X.-M. Hu, Z.-B. Hou, H. Cao, J.-M. Cui, B.-H. Liu, Y.-F. Huang, C.-F. Li, G.-C. Guo, and G. Chiribella, Experimental transmission of quantum information using a superposition of causal orders, *Phys. Rev. Lett.* **124**, 030502 (2020), arXiv:1811.07526 [quant-ph].

- [78] K. Goswami, Y. Cao, G. A. Paz-Silva, J. Romero, and A. G. White, Increasing communication capacity via superposition of order, *Phys. Rev. Research* **2**, 033292 (2020), arXiv:1807.07383 [quant-ph].
- [79] G. Rubino *et al.*, Experimental quantum communication enhancement by superposing trajectories, *Phys. Rev. Research* **3**, 013093 (2021), arXiv:2007.05005 [quant-ph].
- [80] M. M. Taddei *et al.*, Computational advantage from the quantum superposition of multiple temporal orders of photonic gates, *PRX Quantum* **2**, 010320 (2021), arXiv:2002.07817 [quant-ph].
- [81] H. M. Wiseman, S. J. Jones, and A. C. Doherty, Steering, entanglement, nonlocality, and the Einstein-Podolsky-Rosen paradox, *Phys. Rev. Lett.* **98**, 140402 (2007), arXiv:quant-ph/0612147.
- [82] R. Uola, A. C. S. Costa, H. C. Nguyen, and O. Gühne, Quantum steering, *Rev. Mod. Phys.* **92**, 015001 (2020), arXiv:1903.06663 [quant-ph].
- [83] C. W. Helstrom, Quantum detection and estimation theory, *Journal of Statistical Physics* **1**, 231–252 (1969).
- [84] H.-K. Lo, M. Curty, and K. Tamaki, Secure quantum key distribution, *Nat. Photon.* **8**, 595–604 (2014), arXiv:1505.05303 [quant-ph].
- [85] C. H. Bennett, P. W. Shor, J. A. Smolin, and A. V. Thapliyal, Entanglement-assisted capacity of a quantum channel and the reverse shannon theorem, *IEEE T. Inform. Theory* **48**, 2637–2655 (2002), arXiv:quant-ph/0106052.
- [86] C. H. Bennett, G. Brassard, and N. D. Mermin, Quantum cryptography without Bell’s theorem, *Phys. Rev. Lett.* **68**, 557–559 (1992).
- [87] A. Acín, N. Brunner, N. Gisin, S. Massar, S. Pironio, and V. Scarani, Device-independent security of quantum cryptography against collective attacks, *Phys. Rev. Lett.* **98**, 230501 (2007), arXiv:quant-ph/0702152.
- [88] U. Vazirani and T. Vidick, Fully device-independent quantum key distribution, *Phys. Rev. Lett.* **113**, 140501 (2014), arXiv:1210.1810 [quant-ph].
- [89] B. E. Anderson, H. Sosa-Martinez, C. A. Riofrío, I. H. Deutsch, and P. S. Jessen, Accurate and Robust Unitary Transformations of a High-Dimensional Quantum System, *Phys. Rev. Lett.* **114**, 240401 (2015), arXiv:1410.3891 [quant-ph].
- [90] K. S. Kumar, A. Vepsäläinen, S. Danilin, and G. S. Paraoanu, Stimulated Raman adiabatic passage in a three-level superconducting circuit, *Nat. Commun.* **7**, 10628 (2016), arXiv:1508.02981 [quant-ph].
- [91] R. A. Bertlmann and P. Krammer, Bloch vectors for qudits, *J. Phys. A: Math. Theor.* **41**, 235303 (2008), arXiv:0806.1174 [quant-ph].
- [92] C. H. Bennett, D. P. Di Vincenzo, J. A. Smolin, and W. K. Wootters, Mixed-state entanglement and quantum error correction, *Phys. Rev. A* **54**, 3824 (1996), arXiv:quant-ph/9604024.

- [93] W. K. Wootters, Entanglement of formation of an arbitrary state of two qubits, *Phys. Rev. Lett.* **80**, 2245–2248 (1998), arXiv:quant-ph/9709029.
- [94] P. Erker, M. Krenn, and M. Huber, Quantifying high dimensional entanglement with two mutually unbiased bases, *Quantum* **1**, 22 (2017), arXiv:1512.05315 [quant-ph].
- [95] M. Piani and C. Mora, Class of PPT bound entangled states associated to almost any set of pure entangled states, *Phys. Rev. A* **75**, 012305 (2007), arXiv:quant-ph/0607061.
- [96] R. Fickler, R. Lapkiewicz, M. Huber, M. P. J. Lavery, M. J. Padgett, and A. Zeilinger, Interface between path and orbital angular momentum entanglement for high-dimensional photonic quantum information, *Nat. Commun.* **5**, 4502 (2014), arXiv:1402.2423 [quant-ph].
- [97] V. Arrizón, U. Ruiz, R. Carrada, and L. A. González, Pixelated phase computer holograms for the accurate encoding of scalar complex fields, *J. Opt. Soc. Am. A* **24**, 3500–3507 (2007).
- [98] W. K. Wootters and B. D. Fields, Optimal state-determination by mutually unbiased measurements, *Annals Phys.* **191**, 363 (1989).
- [99] A. Coladangelo, K. T. Goh, and V. Scarani, All Pure Bipartite Entangled States can be Self-Tested, *Nat. Commun.* **8**, 15485 (2017), arXiv:1611.08062 [quant-ph].
- [100] G. C. G. Berkhout, M. P. J. Lavery, J. Courtial, M. W. Beijersbergen, and M. J. Padgett, Efficient sorting of orbital angular momentum states of light, *Phys. Rev. Lett.* **105**, 153601 (2010).
- [101] M. Mirhosseini, M. Malik, Z. Shi, and R. W. Boyd, Efficient separation of the orbital angular momentum eigenstates of light, *Nat. Commun.* **4**, 2781 (2013), arXiv:1306.0849 [quant-ph].
- [102] T. Brougham and S. M. Barnett, Mutually unbiased measurements for high-dimensional time-bin-based photonic states, *Europhys. Lett.* **104**, 30003 (2013), arXiv:1311.2773 [quant-ph].
- [103] J. Mower, Z. Zhang, P. Desjardins, C. Lee, J. H. Shapiro, and D. Englund, High-dimensional quantum key distribution using dispersive optics, *Phys. Rev. A* **87**, 062322 (2013), arXiv:1210.4501 [quant-ph].
- [104] C. Spengler, M. Huber, S. Brierley, T. Adaktylos, and B. C. Hiesmayr, Entanglement detection via mutually unbiased bases, *Phys. Rev. A* **86**, 022311 (2012), arXiv:1202.5058 [quant-ph].
- [105] D. Giovannini, J. Romero, J. Leach, A. Dudley, A. Forbes, and M. J. Padgett, Characterization of high-dimensional entangled systems via mutually unbiased measurements, *Phys. Rev. Lett.* **110**, 143601 (2013), arXiv:1212.5825 [quant-ph].
- [106] D. S. Tasca, L. Rudnicki, R. S. Aspden, M. J. Padgett, P. H. Souto Ribeiro, and S. P. Walborn, Testing for entanglement with periodic coarse graining, *Phys. Rev. A* **97**, 042312 (2018), arXiv:1506.01095 [quant-ph].
- [107] E. C. Paul, D. S. Tasca, Ł. Rudnicki, and S. P. Walborn, Detecting entanglement of continuous variables with three mutually unbiased bases, *Phys. Rev. A* **94**, 012303 (2016), arXiv:1604.07347 [quant-ph].

- [108] D. S. Tasca, P. Sánchez, S. P. Walborn, and Ł. Rudnicki, Mutual Unbiasedness in Coarse-grained Continuous Variables, *Phys. Rev. Lett.* **120**, 040403 (2018), arXiv:1709.00137 [quant-ph].
- [109] J. Schneeloch and G. A. Howland, Quantifying high-dimensional entanglement with einstein-podolsky-rosen correlations, *Phys. Rev. A* **97**, 042338 (2018), arXiv:1709.03626 [quant-ph].
- [110] D. Sauerwein, C. Macchiavello, L. Maccone, and B. Kraus, Multipartite correlations in mutually unbiased bases, *Phys. Rev. A* **95**, 042315 (2017), arXiv:1701.07412 [quant-ph].
- [111] M. Malik, M. Erhard, M. Huber, M. Krenn, R. Fickler, and A. Zeilinger, Multi-photon entanglement in high dimensions, *Nat. Photon.* **10**, 248–252 (2016), arXiv:1509.02561 [quant-ph].
- [112] M. Krenn, M. Malik, R. Fickler, R. Lapkiewicz, and A. Zeilinger, Automated search for new quantum experiments, *Phys. Rev. Lett.* **116**, 090405 (2016), arXiv:1509.02749 [quant-ph].
- [113] M. M. Wolf, D. Perez-Garcia, and C. Fernandez, Measurements incompatible in quantum theory cannot be measured jointly in any other no-signaling theory, *Phys. Rev. Lett.* **103**, 230402 (2009), arXiv:0905.2998 [quant-ph].
- [114] D. Mayers and A. Yao, Self testing quantum apparatus, *Quantum Information and Computation* **4**, 273–286 (2004), arXiv:quant-ph/0307205.
- [115] C.-E. Bardyn, T. C. H. Liew, S. Massar, M. McKague, and V. Scarani, Device-independent state estimation based on bell’s inequalities, *Phys. Rev. A* **80**, 062327 (2009), arXiv:0907.2170 [quant-ph].
- [116] A. Aspect, P. Grangier, and G. Roger, Experimental Realization of Einstein-Podolsky-Rosen-Bohm *Gedankenexperiment*: A New Violation of Bell’s Inequalities, *Phys. Rev. Lett.* **49**, 91–94 (1982).
- [117] B. Hensen, *et al.*, Loophole-free Bell inequality violation using electron spins separated by 1.3 kilometres, *Nature* **526**, 682–686 (2015), arXiv:1508.05949 [quant-ph].
- [118] M. Giustina, *et al.*, Significant-loophole-free test of Bell’s theorem with entangled photons, *Phys. Rev. Lett.* **115**, 250401 (2015), arXiv:1511.03190 [quant-ph].
- [119] L. K. Shalm, *et al.*, Strong loophole-free test of local realism, *Phys. Rev. Lett.* **115**, 250402 (2015), arXiv:1511.03189 [quant-ph].
- [120] A. Acín, N. Brunner, N. Gisin, S. Massar, S. Pironio, and V. Scarani, Device-independent security of quantum cryptography against collective attacks, *Phys. Rev. Lett.* **98**, 230501 (2007), arXiv:quant-ph/0702152.
- [121] A. Acín, S. Massar, and S. Pironio, Randomness versus nonlocality and entanglement, *Phys. Rev. Lett.* **108**, 100402 (2012), arXiv:1107.2754 [quant-ph].
- [122] M. T. Quintino, T. Vértesi, and N. Brunner, Joint Measurability, Einstein-Podolsky-Rosen Steering, and Bell Nonlocality, *Phys. Rev. Lett.* **113**, 160402 (2014), arXiv:1406.6976 [quant-ph].

- [123] R. Uola, T. Moroder, and O. Ghne, Joint measurability of generalized measurements implies classicality, *Phys. Rev. Lett.* **113**, 160403 (2014), arXiv:1407.2224 [quant-ph].
- [124] C. Branciard, E. G. Cavalcanti, S. P. Walborn, V. Scarani, and H. M. Wiseman, One-sided device-independent quantum key distribution: Security, feasibility, and the connection with steering, *Phys. Rev. A* **85**, 010301 (2012), arXiv:1109.1435 [quant-ph].
- [125] K. Kraus, A. Bhm, J. D. Dollard, and W. H. Wootters, *States, Effects, and Operations: Fundamental Notions of Quantum Theory. Lectures in Mathematical Physics at the University of Texas at Austin* (Cambridge University Press, 1983).
- [126] J. de Pillis, Linear transformations which preserve hermitian and positive semidefinite operators, *Pacific Journal of Mathematics* **23**, 129–137 (1967).
- [127] A. Jamiołkowski, Linear transformations which preserve trace and positive semidefiniteness of operators, *Reports on Mathematical Physics* **3**, 275–278 (1972).
- [128] M.-D. Choi, Completely positive linear maps on complex matrices, *Linear Algebra and its Applications* **10**, 285–290 (1975).
- [129] F. Riesz, Sur une espce de gomtrie analytique des systmes de fonctions sommables, *C. R. Acad. Sci. Paris* **144**, 1409–1411 (1907).
- [130] C. Budroni, *et al.*, (in preparation).
- [131] L. A. Khalfin and B. S. Tsirelson, Quantum and quasi-classical analogs of Bell inequalities, *Symposium on the Foundations of Modern Physics* , 441–460 (1985).
- [132] P. Rastall, Locality, Bell’s theorem, and quantum mechanics, *Foundations of Physics* **15**, 963–972 (1985).
- [133] S. Popescu and D. Rohrlich, Quantum nonlocality as an axiom, *Foundations of Physics* **24**, 379–385 (1994).
- [134] A. Feix, M. Arajo, and Č. Brukner, Causally nonseparable processes admitting a causal model, *New J. Phys.* **18**, 083040 (2016), arXiv:1604.03391 [quant-ph].
- [135] M. F. Pusey, Negativity and steering: A stronger Peres conjecture, *Phys. Rev. A* **88**, 032313 (2013), arXiv:1305.1767 [quant-ph].
- [136] E. Schrödinger, Discussion of probability relations between separated systems, *Proceedings of the Cambridge Philosophical Society* **31**, 555 (1935).
- [137] N. Gisin, Stochastic quantum dynamics and relativity, *Helvetica Physica Acta* **62**, 363–371 (1989).
- [138] L. P. Hughston, R. Jozsa, and W. K. Wootters, A complete classification of quantum ensembles having a given density matrix, *Phys. Lett. A* **183**, 14–18 (1993).
- [139] <https://github.com/jessicabavaresco/sdi-causality>.

- [140] A. Y. Kitaev, Quantum computations: algorithms and error correction, *Russian Mathematical Surveys* **52**, 1191–1249 (1997).
- [141] D. Aharonov, A. Kitaev, and N. Nisan, Quantum circuits with mixed states, *Proceedings of the Thirtieth Annual ACM Symposium on Theory of Computing*, 20–30 (1998).
- [142] A. Acín, Statistical distinguishability between unitary operations, *Phys. Rev. Lett.* **87**, 177901 (2001), arXiv:quant-ph/0102064.
- [143] G. Chiribella, G. M. D'Ariano, and P. Perinotti, Memory effects in quantum channel discrimination, *Phys. Rev. Lett.* **101**, 180501 (2008), arXiv:0803.3237 [quant-ph].
- [144] 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].
- [145] V. Katariya and M. M. Wilde, Evaluating the advantage of adaptive strategies for quantum channel distinguishability, *arXiv preprint* (2020), arXiv:2001.05376 [quant-ph].
- [146] 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 [quant-ph].
- [147] G. Chiribella, Optimal networks for quantum metrology: semidefinite programs and product rules, *New Journal of Physics* **14**, 125008 (2012), arXiv:1207.6172 [quant-ph].
- [148] G. Chiribella and D. Ebler, Optimal quantum networks and one-shot entropies, *New Journal of Physics* **18**, 093053 (2016), arXiv:1606.02394 [quant-ph].
- [149] H. Yuen, R. Kennedy, and M. Lax, Optimum testing of multiple hypotheses in quantum detection theory, *IEEE Transactions on Information Theory* **21**, 125–134 (1975).
- [150] H. Peyrl and P. A. Parrilo, Computing sum of squares decompositions with rational coefficients, *Theoretical Computer Science* **409**, 269–281 (2008).
- [151] S. M. Rump, Verification methods: Rigorous results using floating-point arithmetic, *Acta Numerica* **19**, 287–449 (2010).
- [152] S. Pirandola, R. Laurenza, C. Lupo, and J. L. Pereira, Fundamental limits to quantum channel discrimination, *npj Quantum Information* **5**, 50 (2019), arXiv:1803.02834 [quant-ph].
- [153] Q. Zhuang and S. Pirandola, Ultimate limits for multiple quantum channel discrimination, *Phys. Rev. Lett.* **125**, 080505 (2020), arXiv:2007.14566 [quant-ph].
- [154] J. L. Pereira and S. Pirandola, Bounds on amplitude-damping-channel discrimination, *Phys. Rev. A* **103**, 022610 (2021), arXiv:2009.04783 [quant-ph].
- [155] M. Rexiti and S. Mancini, Discriminating qubit amplitude damping channels, *arXiv preprint* (2020), arXiv:2009.01000 [quant-ph].
- [156] 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].

- [157] [https://github.com/mteq/channel\\_discrimination](https://github.com/mteq/channel_discrimination).
- [158] M. Hayashi, *Quantum Hypothesis Testing and Discrimination of Quantum States* (Springer Berlin Heidelberg, 2006) pp. 69–91.
- [159] M. G. Paris, Quantum estimation for quantum technology, *International Journal of Quantum Information* **7**, 125–137 (2009), arXiv:0804.2981 [quant-ph].
- [160] D. Deutsch and R. Jozsa, Rapid solution of problems by quantum computation, *Proceedings of the Royal Society of London. Series A: Mathematical and Physical Sciences* **439**, 553–558 (1992).
- [161] L. K. Grover, A framework for fast quantum mechanical algorithms, *Proceedings of the 30th annual ACM symposium on Theory of computing* , 53–62 (1998), arXiv:quant-ph/9711043.
- [162] D. R. Simon, On the power of quantum computation, *Proceedings 35th Annual Symposium on Foundations of Computer Science* , 116–123 (1994).
- [163] A.Chefles, A. Kitagawa, M. Takeoka, M. Sasaki, and J. Twamley, Unambiguous discrimination among oracle operators, *Journal of Physics A Mathematical General* **40**, 10183–10213 (2007), arXiv:quant-ph/0702245.
- [164] 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.
- [165] 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.
- [166] G. Chiribella, G. M. D’Ariano, and M. Roetteler, Identification of a reversible quantum gate: assessing the resources, *New Journal of Physics* **15**, 103019 (2013), arXiv:1306.0719 [quant-ph].
- [167] M. Ziman and M. Sedlák, Single-shot discrimination of quantum unitary processes, *Journal of Modern Optics* **57**, 253–259 (2010), arXiv:1003.1488 [quant-ph].
- [168] G. Gutoski and J. Watrous, Toward a general theory of quantum games, *Proceedings of the 39th annual ACM symposium on Theory of computing* , 565–574 (2006), arXiv:quant-ph/0611234.
- [169] D. Kretschmann and R. F. Werner, Quantum channels with memory, *Phys. Rev. A* **72**, 062323 (2005), arXiv:quant-ph/0502106.
- [170] J. Barrett, R. Lorenz, and O. Oreshkov, Cyclic quantum causal models, *Nature Communications* **12**, 885 (2021), arXiv:2002.12157 [quant-ph].
- [171] [https://github.com/mteq/unitary\\_channel\\_discrimination](https://github.com/mteq/unitary_channel_discrimination).
- [172] T. Kraft, S. Designolle, C. Ritz, N. Brunner, O. Gühne, and M. Huber, Quantum entanglement in the triangle network, *arXiv preprint* (2020), arXiv:2002.03970 [quant-ph].
- [173] M. Navascués, E. Wolfe, D. Rosset, and A. Pozas-Kerstjens, Genuine network multipartite entanglement, *Phys. Rev. Lett.* **125**, 240505 (2020), arXiv:2002.02773 [quant-ph].

- [174] C. Branciard, N. Gisin, and S. Pironio, Characterizing the nonlocal correlations created via entanglement swapping, *Phys. Rev. Lett.* **104**, 170401 (2010), arXiv:0911.1314 [quant-ph].
- [175] D. Cavalcanti, M. L. Almeida, V. Scarani, and A. Acín, Quantum networks reveal quantum nonlocality, *Nat. Commun.* **2**, 184 (2011), arXiv:1010.0900 [quant-ph].
- [176] T. Fritz, Beyond Bell's theorem: correlation scenarios, *New Journal of Physics* **14**, 103001 (2012), arXiv:1206.5115 [quant-ph].
- [177] M.-O. Renou, E. Bäumer, S. Boreiri, N. Brunner, N. Gisin, and S. Beigi, Genuine quantum nonlocality in the triangle network, *Phys. Rev. Lett.* **123**, 140401 (2019), arXiv:1905.04902 [quant-ph].
- [178] W. F. Stinespring, Positive Functions on C\*-algebras, *Proc. Amer. Math. Soc.* **6**, 211 (1955).
- [179] P. G. L. Dirichlet, Sur la convergence des séries trigonométriques qui servent à représenter une fonction arbitraire entre des limites données, *Journal für die reine und angewandte Mathematik* **4**, 157 (1829).
- [180] M. Huber and J. I. de Vicente, Structure of multidimensional entanglement in multipartite systems, *Phys. Rev. Lett.* **110**, 030501 (2013), arXiv:1210.6876 [quant-ph].
- [181] M. Huber, M. Perarnau-Llobet, and J. I. de Vicente, Entropy vector formalism and the structure of multidimensional entanglement in multipartite systems, *Phys. Rev. A* **88**, 042328 (2013), arXiv:1307.3541 [quant-ph].
- [182] J. L. W. V. Jensen, Sur les fonctions convexes et les inégalités entre les valeurs moyennes, *Acta Math.* **30**, 175–193 (1906).
- [183] D. N. Klyshko, A simple method of preparing pure states of an optical field, of implementing the Einstein–Podolsky–Rosen experiment, and of demonstrating the complementarity principle, *Sov. Phys. Usp.* **31**, 74–85 (1988).
- [184] M. Karpiński, M. Jachura, L. J. Wright, and B. J. Smith, Bandwidth manipulation of quantum light by an electro-optic time lens, *Nat. Photonics* **11**, 53–57 (2016), arXiv:1604.02459 [quant-ph].
- [185] J. Carolan *et al.*, Universal linear optics, *Science* **349**, 711–716 (2015), arXiv:1505.01182 [quant-ph].
- [186] G. Vallone, R. Ceccarelli, F. De Martini, and P. Mataloni, Hyperentanglement of two photons in three degrees of freedom, *Phys. Rev. A* **79**, 030301(R) (2009), arXiv:0810.4461 [quant-ph].
- [187] W.-B. Gao *et al.*, Experimental demonstration of a hyper-entangled ten-qubit Schrödinger cat state, *Nat. Phys.* **6**, 331–335 (2010), arXiv:0809.4277 [quant-ph].
- [188] X.-L. Wang *et al.*, Experimental ten-photon entanglement, *Phys. Rev. Lett.* **117**, 210502 (2016), arXiv:1605.08547 [quant-ph].
- [189] X.-L. Wang *et al.*, 18-qubit entanglement with six photons' three degrees of freedom, *Phys. Rev. Lett.* **120**, 260502 (2018), arXiv:1801.04043 [quant-ph].

- [190] C. Song *et al.*, 10-qubit entanglement and parallel logic operations with a superconducting circuit, *Phys. Rev. Lett.* **119**, 180511 (2017), arXiv:1703.10302 [quant-ph].
- [191] N. Friis *et al.*, Observation of entangled states of a fully controlled 20-qubit system, *Phys. Rev. X* **8**, 021012 (2018), arXiv:1711.11092 [quant-ph].
- [192] H. Qassim, F.-M. Miatto, J.-P. Torres, M.-J. Padgett, E. Karimi, and R.-W. Boyd, Limitations to the determination of a Laguerre–Gauss spectrum via projective, phase-flattening measurement, *Journal of the Optical Society of America B* **31**, A20–A23 (2014), arXiv:1401.3512 [quant-ph].
- [193] 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 [quant-ph].
- [194] A. A. Abbott, J. Wechs, F. Costa, and C. Branciard, Genuinely multipartite noncausality, *Quantum* **1**, 39 (2017), arXiv:1708.07663 [quant-ph].
- [195] D. Beckman, D. Gottesman, M. A. Nielsen, and J. Preskill, Causal and localizable quantum operations, *Phys. Rev. A* **64**, 052309 (2001), arXiv:quant-ph/0102043.
- [196] T. Eggeling, D. Schlingemann, and R. F. Werner, Semicausal operations are semilocalizable, *EPL (Europhysics Letters)* **57**, 782–788 (2002), arXiv:quant-ph/0104027.
- [197] Wikipedia, [https://en.wikipedia.org/wiki/Floating-point\\_arithmetic](https://en.wikipedia.org/wiki/Floating-point_arithmetic).
- [198] <https://floating-point-gui.de>.
- [199] CVX: Matlab Software for Disciplined Convex Programming, <http://cvxr.com/cvx/>.
- [200] The MOSEK optimization toolbox for MATLAB, <https://www.mosek.com>.
- [201] SeDuMi, <http://sedumi.ie.lehigh.edu/>.
- [202] SDPT3 — a MATLAB software for semidefinite-quadratic-linear programming, <https://blog.nus.edu.sg/mattohkc/softwares/sdpt3/>.
- [203] The MIT License, <https://opensource.org/licenses/MIT>.
- [204] N. Johnston, QETLAB: A MATLAB Toolbox for Quantum Entanglement, <http://www.qetlab.com>.
- [205] G. Chiribella, G. M. D’Ariano, and P. Perinotti, Quantum circuit architecture, *Phys. Rev. Lett.* **101**, 060401 (2008), arXiv:0712.1325 [quant-ph].
- [206] A. Bisio, G. M. D’Ariano, P. Perinotti, and M. Sedlák, Optimal processing of reversible quantum channels, *Physics Letters A* **378**, 1797–1808 (2014), arXiv:1308.3254 [quant-ph].