The Technical University of Catalonia

PhD Thesis

Bell nonlocality and causal networks

Author:
Emanuel-Cristian
Boghiu Crihan

Supervisor:
Prof. Dr. Antonio Acín

June 5, 2024

Abstract

(English) Understanding the cause and effect relationships behind observed correlations is central to how we reason and interact with the world. Causal relationships help us make sense of the patterns we observe and predict what interventions in nature might lead to a desired outcome. These patterns can be mathematically framed as the joint probability distribution of a set of classical random variables which capture information gathered from the environment. This information may range from abstract data, like survey response statistics, to physical events, such as the probability of triggering a photon detector. A fundamental question is that of causal compatibility: Are the observed correlations compatible with a given causal explanation? A causal explanation can be expressed in terms of causal models, which can be systematically studied with the tools provided by the field of causal inference. Causal models consist of observable random variables with known probability distributions and latent variables with unknown distributions which, together, explain observed correlations through causal influences, that is, functional relationships between the values of these variables. Quantum theory—one of the most accurate theories at a fundamental level—is inherently probabilistic. Measurement results are, therefore, represented as random variables. This naturally leads to causal analysis: Which cause and effect relationships can explain observed measurement statistics in a quantum experiment? One of the simplest quantum experiments is that of two distant parties performing space-like separated, independently chosen measurements on a shared quantum state. In 1964, John Bell showed that in this experiment quantum theory predicts correlations that defy any classical common-cause explanation through a result known as Bell's Theorem. This phenomenon is known as Bell nonlocality. This thesis aims to operationally characterize the fundamental differences between classical and quantum

theories within causal scenarios beyond Bell's common-cause scenario. Such an understanding may eventually help integrate quantum phenomena into a coherent, conceptually clear framework of causality. Towards this goal, we explore how classical and quantum causal models diverge in operational tasks in specific causal scenarios. We focus on simple scenarios that go beyond Bell's, while seeking to discover new forms of quantum advantage that are fundamentally different from traditional Bell nonlocality. Our goal is to link these new forms of quantum advantage to different nonclassical features of quantum theory and study their potential applications. A critical component of this research is testing for the causal compatibility of specific correlations with a given causal model. As such, an important part of this thesis is dedicated to expanding and refining the scope of current methods for testing causal compatibility.

(Catalan) Entendre les relacions de causa i efecte darrere de les correlacions observades és central per a la nostra manera de raonar i interactuar amb el món. Les relacions de causa i efecte ens ajuden a donar sentit als patrons que observem i predir quines intervencions en la naturalesa podrien portar a un resultat desitjat. Aquests patrons es poden emmarcar matemàticament com la distribució de probabilitat conjunta d'un conjunt de variables aleatòries clàssiques que capturen informació recollida de l'entorn. Aquesta informació pot variar des de dades abstractes, com estadístiques de respostes d'enquestes, fins a esdeveniments físics, com la probabilitat d'activar un fotodetector. Una pregunta fonamental és la de la compatibilitat causal: Són les correlacions observades compatibles amb una explicació causal donada? Una explicació causal pot expressar-se en termes de models causals, els quals poden ser estudiats sistemàticament amb les eines proporcionades pel camp de la inferència causal. Els models causals consisteixen en variables aleatòries observables amb distribucions de probabilitat conegudes i variables latents amb distribucions desconegudes que, juntes, expliquen les correlacions observades a través d'influències causals, és a dir, relacions funcionals entre els valors d'aquestes variables. La teoria quàntica—una de les teories més precises a nivell fonamental—és inherentment probabilística. Per tant, els resultats de les mesures es representen com a variables aleatòries. Això porta naturalment a l'anàlisi causal: Quines relacions de causa i efecte poden explicar els resultats obtinguts en un experiment quàntic? Un dels experiments quàntics més simples és aquell en què dos laboratoris distants

realitzen mesures triades independentment i separades espacialment sobre un estat quàntic compartit. El 1964, John Bell va demostrar que en aquest experiment la teoria quàntica prediu correlacions que desafien qualsevol explicació clàssica en termes de causes comunes. Aquest resultat és conegut com el Teorema de Bell, i el fenomen com no localitat de Bell. Aquesta tesi té com a objectiu caracteritzar operacionalment les diferències fonamentals entre les teories clàssiques i quàntiques dins d'escenaris causals més enllà de l'escenari de causa comuna de Bell. Aquest enteniment eventualment pot ajudar a integrar fenòmens quàntics en un marc de causalitat coherent i conceptualment clar. Per assolir aquest objectiu, explorem com els models causals clàssics i quàntics divergeixen en tasques operacionals en escenaris causals específics. Ens centrem en escenaris simples que van més enllà dels de Bell, mentre busquem descobrir noves formes d'avantatge quàntic que són fonamentalment diferents de la no localitat de Bell tradicional. El nostre objectiu és vincular aquestes noves formes d'avantatge quàntic amb diferents propietats no clàssiques de la teoria quàntica i estudiar les seves aplicacions potencials. Un component crític d'aquesta investigació és provar la compatibilitat causal de correlacions específiques amb un model causal donat. Així doncs, una part important d'aquesta tesi està dedicada a expandir i refinar l'abast dels mètodes actuals per provar la compatibilitat causal.

(Castilian Spanish) Entender las relaciones de causa y efecto detrás de las correlaciones observadas es central para nuestra forma de razonar e interactuar con el mundo. Las relaciones de causa y efecto nos ayudan a dar sentido a los patrones que observamos y predecir qué intervenciones en la naturaleza podrían llevar a un resultado deseado. Estos patrones pueden enmarcarse matemáticamente como la distribución de probabilidad conjunta de un conjunto de variables aleatorias clásicas que capturan información recogida del entorno. Esta información puede variar desde datos abstractos, como estadísticas de respuestas de encuestas, hasta eventos físicos, como la probabilidad de activar un fotodetector. Una pregunta fundamental es la de la compatibilidad causal: ¿Son las correlaciones observadas compatibles con una explicación causal dada? Una explicación causal puede expresarse en términos de modelos causales, los cuales pueden ser estudiados sistemáticamente con las herramientas proporcionadas por el campo de la inferencia causal. Los modelos causales consisten en variables aleatorias observables

con distribuciones de probabilidad conocidas y variables latentes con distribuciones desconocidas que, juntas, explican las correlaciones observadas a través de influencias causales, es decir, relaciones funcionales entre los valores de estas variables. La teoría cuántica—una de las teorías más precisas a nivel fundamental—es inherentemente probabilística. Por lo tanto, los resultados de las mediciones se representan como variables aleatorias. Esto lleva naturalmente al análisis causal: ¿Qué relaciones de causa y efecto pueden explicar las resultados obtenidos en un experimento cuántico? Uno de los experimentos cuánticos más simples es aquel en el que dos laboratorios distantes realizan mediciones independientemente elegidas y espacialmente separadas sobre un estado cuántico compartido. En 1964, John Bell demostró que en este experimento la teoría cuántica predice correlaciones que desafían cualquier explicación clásica en terminos de causas comunes. Este resultado es conocido como el Teorema de Bell, y el fenómeno como no localidad de Bell. Esta tesis tiene como objetivo caracterizar operacionalmente las diferencias fundamentales entre las teorías clásicas y cuánticas dentro de escenarios causales más allá del escenario de causa común de Bell. Tal entendimiento eventualmente puede ayudar a integrar fenómenos cuánticos en un marco de causalidad coherente y conceptualmente claro. Hacia este objetivo, exploramos cómo los modelos causales clásicos y cuánticos divergen en tareas operacionales en escenarios causales específicos. Nos enfocamos en escenarios simples que van más allá de los de Bell, mientras buscamos descubrir nuevas formas de ventaja cuántica que son fundamentalmente diferentes de la no localidad de Bell tradicional. Nuestro objetivo es vincular estas nuevas formas de ventaja cuántica con diferentes propiedades no clásicas de la teoría cuántica y estudiar sus potenciales aplicaciones. Un componente crítico de esta investigación es probar la compatibilidad causal de correlaciones específicas con un modelo causal dado. Como tal, una parte importante de esta tesis está dedicada a expandir y refinar el alcance de los métodos actuales para probar la compatibilidad causal.

Acknowledgements

În primul rând, doresc să îmi exprim recunoștința față de **familia** mea pentru sprijinul constant, răbdarea și încrederea arătate pe parcursul studiilor mele doctorale.

También quiero agradecer a mis amigos de Zaragoza: **Miguel, Amador, Gabriela, Dani, Jorge, David, Vicky, Victor, Fer, Lucía, Sergio, Mikel, Pablo, Marcos**. He valorado mucho poder reunirme y ponerme al día con cada uno de vosotros al volver a casa.

In Barcelona there's an enormous list of people that influenced me positively. I want to thank **Toni** for being very nice and supportive during my PhD. He was very understanding during the COVID lockdown, gave me liberty to pursue my own ideas and allowed me to travel extensively. **Joe**: You have naturally adapted a mentor role throughout my PhD. I am grateful for all the things I've learned from you, both professionally and personally. "I" am especially grateful for introducing "me" to meditation, which really changed "my" perspective towards many things for the better. **Alex**: Even though you left one month after I started my PhD, you've been a constant all throughout. I'm inspired by your niceness, humbleness, and working spirit! **Zahra, Felix**: You were very nice to me at the start of my PhD; thank you for letting me pester you with all of my questions! **Máté, Vicky**: You quickly became a core of the QIT group, and left a big hole once you left :(. **Matteo, Jacopo, Paolo, Erika, Dario, Binz, Arturo**: I really enjoyed spending time with all of you. I'm especially thankful to Matteo, Jacopo and Paolo for helping me broaden my perspective on many things. **Alexia, Moha, Tamás**: You're very nice humble wholesome people, and a source of inspiration! **Gaël**: I really enjoyed our conversations about life and politics. **Miqueleta**: It was very fun to procrastinate together in the office all these years :). **Anna, Luke, Fionnuala, Hyppo**: Thank you for all the ramen and moviegoing, and for giving me an excuse to be more nerdy!

I'm very grateful to **Arturo** for introducing me to ICONS! It was a

wonderful experience, and I've learned a lot from being part of a student chapter. I'm also thankful to everyone at ICFO that collaborated with ICONS and made organising things easy. **Michael, Costi, Niccolò, Stefano, Joana, Lukas, Jessica, Ian, Ediz, Marta**: Thank you for the support and all the nice times we had together, in the flat during COVID and outside! Costi, I promise I'll finally start going to therapy once I have money :). **Egle**: I appreciate the vibe you bring both in the flat flat and outside; you and manzanillas are now inseparable in my mind! **Loïc, Max, Jan, Eliza, Sandra, Giorgio**: I enjoyed all the time we've spent together and all your support! Loïc, thank you for motivating me to go bouldering!

Outside Barcelona, I'm very grateful for my stay at Perimeter, and to especially to **Elie** for making it possible! I'm really impressed by your passion for science and how well you transmit it to others. I've learned a lot from you! I'm particularly grateful to **Hlér, Jacopo, Marina** for making Waterloo home, and to **Dani, Maria, Barbara, Bruno, Pedro, Arad, Adrian, Seraphim** for all the times we hung out, all the squash and all the conversations we've had!

I'm also grateful for the quantum information and quantum foundations community! **Carla, Armin, Ravi, Sebastién, Martin, Sophie, Flavien, Marco-Tulio, Victor, Hlér, Vilasini, Lídia**: You're wonderful people, humble, chill and hard working, and you make the quantum information field a good place to do research in. I'm particularly grateful to **Lídia** for organising amazing, inclusive summer schools. These allowed me to meet wonderful people, and proved to be a great source of motivation during my PhD!

E.-C. B. acknowledges financial support from the Spanish State Research Agency through the "Severo Ochoa" program for Centers of Excellence in R&D (CEX2019-000910-S), from Fundació Cellex, Fundació Mir-Puig, and from Generalitat de Catalunya through the CERCA program. This thesis also received funding from the "Presidencia de la Agencia Estatal de Investigación" within the "Convocatoria de tramitación anticipada, correspondiente al año 20XX, de las ayudas para contratos predoctorales (Ref. PRE2019-088482) para la formación de doctores contemplada en el Subprograma Estatal de Formación del Programa Estatal de Promoción del Talento y su Empleabilidad en I+D+i, en el marco del Plan Estatal de Investigación Científica y Técnica y de Innovación 2017-2020, cofinanciado por el Fondo Social Europeo."

List of publications

Peer-reviewed publications forming part of this thesis:

- Paolo Abiuso, Tamás Kriváchy, **Emanuel-Cristian Boghiu**, Marc-Olivier Renou, Alejandro Pozas-Kerstjens, and Antonio Acín, *"Single-photon nonlocality in quantum networks"*, Phys. Rev. Research 4, L012041—Published 29 March 2022
- **Emanuel-Cristian Boghiu**, Flavien Hirsch, Pei-Sheng Lin, Marco Túlio Quintino, Joseph Bowles, *"Device-independent and semi-device-independent entanglement certification in broadcast Bell scenarios"*, SciPost Phys. Core 6, 028 (2023)—Published 11 April 2023
- **Emanuel-Cristian Boghiu**, Elie Wolfe, and Alejandro Pozas-Kerstjens, *"Inflation: a Python library for classical and quantum causal compatibility"*, Quantum 7, 996 (2023)—Published 4 May 2023

Peer-reviewed publications not forming part of this thesis:

- Lluís Arola-Fernández, Sergio Faci-Lázaro, Per Sebastian Skardal, **Emanuel-Cristian Boghiu**, Jesús Gómez-Gardeñes and Alex Arenas, *"Emergence of explosive synchronization bombs in networks of oscillators"*, Commun Phys 5, 264 (2022)—Published 29 October 2022

Manuscripts under review not forming part of this thesis:

- **Emanuel-Cristian Boghiu**, Jesús Clemente-Gallardo, Jorge A. Jover-Galtier, David Martínez-Crespo, *"Hybrid quantum-classical control problems"*—Submitted to Communications in Analysis and Mechanics: New trends in Geometric Mechanics on 1 March 2024

Manuscripts under preparation not forming part of this thesis:

- Maria Ciudad-Alanon, **Emanuel-Cristian Boghiu**, Paolo Abiuso, Elie Wolfe, *"Navigating the Nuances of Novel Network Nonlocality"*—Under preparation

List of computational contributions

List of open-acess code contributions as a result of this thesis:

- Emanuel-Cristian Boghiu, Elie Wolfe and Alejandro Pozas-Kerstjens (2022), *"Inflation: Python implementations of the Inflation Technique for causal inference (v1.0.0)"*, Zenodo, https://doi.org/10.5281/zenodo.7305545

  The package is hosted on GitHub[1], with a documentation website[2], distributed on the Python Package Index[3] and also published in a peer-reviewed journal [BWPK23].

- The code to reproduce the results from SciPost Phys. Core 6, 028 (2023) is hosted online on GitHub[4].

[1] https://github.com/ecboghiu/inflation
[2] https://ecboghiu.github.io/inflation/_build/html/index.html
[3] https://pypi.org/project/inflation/
[4] https://github.com/josephbowles/broadcast_nl

List of symbols and abbreviations

DI	**D**evice **I**ndependent
DAG	**D**irected **A**cyclic **G**raph
LP	**L**inear **P**rogramming
SDP	**S**emidefinite **P**rogramming
$\mathcal{L}$	Local set of correlations
$\mathcal{NS}$	No-signalling set of correlations
$\mathcal{Q}$	Quantum set of correlations

Contents

Abstract iii

Acknowledgements vii

List of publications ix

List of computational contributions xi

List of symbols and abbreviations xiii

1 Overview 1
- 1.1 Motivation . . . 1
- 1.2 Main thesis contributions . . . 5

2 Preliminaries 13
- 2.1 Definition of a causal scenario . . . 13
- 2.2 Classical, quantum and no-signaling processes . . . 19
  - 2.2.1 Classical process . . . 19
  - 2.2.2 A short review of finite-dimensional quantum theory . . . 20
  - 2.2.3 Quantum process . . . 27
  - 2.2.4 No-signaling process . . . 27
- 2.3 Bell's common cause scenario . . . 27
  - 2.3.1 Classical common cause . . . 28
  - 2.3.2 Quantum common cause . . . 36
  - 2.3.3 No-signaling common cause . . . 48

3 Beyond common cause scenarios 53
- 3.1 3-party chain scenario . . . 53
- 3.2 Inflation of the 3-party chain scenario . . . 55
  - 3.2.1 Classical preparations . . . 59

3.2.2 Quantum preparations . . . 64
3.2.3 No-signaling preparations . . . 66
3.3 Triangle scenario . . . 67
3.4 Inflation of the triangle scenario . . . 72
3.4.1 Classical preparations . . . 74
3.4.2 Quantum preparations . . . 75
3.4.3 No-signaling preparations . . . 77

4 Single photon nonlocality in quantum networks 81
4.1 Background . . . 81
4.2 The triangle network . . . 83
4.3 Witnessing single-photon nonlocality . . . 86
4.4 Noise tolerance and machine learning analysis . . . 88
4.5 Discussion . . . 91

5 Broadcasting of quantum states 93
5.1 Introduction . . . 93
5.1.1 Bell nonlocality in broadcast scenarios . . . 96
5.2 Novel results and methods for broadcast nonlocality . . . 99
5.2.1 Promoting standard Bell inequalities to the broadcast scenario . . . 99
5.2.2 Robustness to detection inefficiencies . . . 102
5.3 Activation of non-signaling genuine multipartite nonlocality . 104
5.3.1 Broadcast activation without a broadcast channel . . 109
5.4 Device-independent entanglement certification . . . 110
5.5 Broadcast steering . . . 112
5.5.1 Standard quantum steering . . . 112
5.5.2 Steering in the broadcast scenario . . . 114
Other potential notions of broadcast steering . . . 115
5.5.3 Broadcast steering with the two-qubit isotropic state . 116
5.6 Discussion . . . 117

6 Tests for causal compatibility 119
6.1 The library . . . 120
6.1.1 Requirements and installation . . . 120
6.1.2 Components . . . 121
6.1.3 Reductions of the feasible region . . . 123
6.2 Main functionality . . . 124

6.2.1 Feasibility problems and extraction of certificates . . . 125
6.2.2 Optimization of Bell operators . . . 129
6.2.3 Bounds on critical parameter values . . . 130
6.3 Further features . . . 131
6.3.1 Standard NPA hierarchy . . . 131
6.3.2 Scenarios with partial information . . . 132
6.3.3 Scenarios beyond networks . . . 132
6.3.4 Feasibility based on distribution supports . . . 134
6.4 Additional library information . . . 135
6.4.1 Computational considerations . . . 135
6.4.2 Future extensions . . . 136
6.4.3 Documentation for Inflation . . . 137
6.4.4 Contribution guidelines . . . 138
6.5 Concluding remarks . . . 138

7 Outlook **141**

A Feasibility in convex programming **177**
A.1 Feasibility as optimization . . . 177
A.2 Farkas' lemma and certificates of infeasibility . . . 178

B Single photon nonlocality **183**
B.1 Noiseless output distribution . . . 183
B.2 Nonlocality of the noiseless distribution . . . 186
B.2.1 Constraints on local models simulating p_t . . . 188
Proof that $X_2^B \cap X_2^C = \emptyset$, $X_0^B \cap X_0^C = \emptyset$: . . . 189
Proof that $X_0^B \cup X_0^C = X$: . . . 189
Proof that $X_0^B \cap X_2^B = \emptyset$, $X_0^C \cap X_2^C = \emptyset$: . . . 189
Proof that all sets X_0^P, Y_0^P, Z_0^P have probability 1/2: 190
B.2.2 Breaking up the coarse-graining . . . 191
B.2.3 Testing $q_t(i, j, k, s)$ using linear programming . . . 194
B.3 Noisy optical realization . . . 196
B.4 Generalization to chains of N parties . . . 199
B.4.1 Constraints on the sets . . . 201
B.4.2 Constraints on the local coarse grained strategy . . . 203
B.4.3 Breaking the coarse-graining and finding linear constraints . . . 203

C Broadcasting appendices **207**
C.1 Heuristic method for device-independent entanglement certification . . . 207
C.2 Efficient method for computing the LHS bound of steering inequalities . . . 208
C.3 Heuristic search for certifying broadcast steering of bipartite states . . . 209
C.3.1 Optimizing a steering inequality . . . 210
C.4 Proof of the lifting *ansatz* in Ineq. (5.10) . . . 212

Chapter 1

Overview

1.1 Motivation

Understanding the cause and effect relationships behind observed correlations is central to how we reason and interact with the world. Causal relationships help us make sense of the patterns we observe and predict what interventions in nature might lead to a desired outcome. These patterns can be mathematically framed as the joint probability distribution of a set of classical random variables which capture information gathered from the environment. This information may range from abstract data, like survey response statistics, to physical events, such as the probability of triggering a photon detector. A fundamental question is that of causal compatibility: *Are the observed correlations compatible with a given causal explanation?* A causal explanation can be expressed in terms of causal models, which can be systematically studied with the tools provided by the field of causal inference [Pea09]. Causal models consist of observable random variables with known probability distributions and latent variables with unknown distributions which, together, explain observed correlations through causal influences, i.e., functional relationships between the values of these variables.

Quantum theory—one of the most accurate theories at a fundamental level—is inherently probabilistic. Measurement results are, therefore, represented as random variables. This naturally leads to causal analysis: *Which cause and effect relationships can explain observed measurement statistics in a quantum experiment?* One of the simplest quantum experiments is that of two distant parties performing space-like separated, independently chosen measurements on a shared quantum state. Physical principles such as

no-superdeterminism and local causality imply that, due to space-like separation, any correlation between the measurement results must be coincidental, rather than causal [WC17b]. This type of coincidental link, known as a spurious correlation, occurs when two variables are correlated but have no direct causal influence on each other. Instead, this correlation can be explained by a third variable, known as a *common-cause* variable, which influences both variables in question. In 1964, John Bell showed that in this experiment quantum theory predicts correlations that defy any classical common-cause explanation through a result known as Bell's Theorem [Bel64]. Bell's Theorem describes how quantum theory violates so-called Bell inequalities, which describe upper bounds on the strength of correlations achievable via classical common-causes; this phenomenon is known as *Bell nonlocality*. The term "nonlocality" reflects the prevalent view within the scientific community that nature is not superdeterministic and therefore, that it is the principle of local causality that is taken to be incompatible with quantum theory.

Overview of Bell nonlocality. Exploring the mismatch between quantum theory and local causality gave rise to an extensive field of research focused on understanding this incompatibility [Bru+14]. Significant progress has been made in comprehending the role of various nonclassical features of quantum theory in producing Bell nonlocal correlations, such as *measurement incompatibility* and *quantum entanglement*. Measurement incompatibility describes quantum measurements or properties that cannot be assigned a well-defined value at the same time, a concept with no counterpart in classical theories. Quantum entanglement refers to correlations between quantum systems strong enough that it appears as if these systems can affect each other instantaneously, even when separated by large distances. This behaviour, colloquially known as "spooky action at a distance", led Einstein, Podolsky and Rosen to argue in 1935 that quantum theory must be an incomplete theory of reality [EPR35].

Some key advancements in the field include understanding that quantum entanglement and Bell nonlocality correspond to different types of resources, as recently formalized in Ref. [Sch+23], or that quantum joint measurability—a generalization of quantum measurement incompatibility—is equivalent to the phenomenon of "spooky action at a distance", formally known as EPR steering [WJD07; CS16; Uol+20]. Another important development was the creation of *device-independent* (DI) protocols. The term "device-independent"

reflects the fact that these protocols do not require any assumptions about the inner workings of the devices producing the observed correlations. This is appealing from a practical perspective, as one can certify the correct functioning of a device without needing to trust its manufacturer. Bell nonlocality is central to various DI protocols such as DI quantum key distribution [Eke91; Ac07], DI certified randomness generation [AM16] and self-testing [MY98; MY04; ŠB20]. Self-testing, in particular, constitutes one of the strongest types of certification as it allows for the DI certification of a quantum device's internal operations—the implemented quantum states and measurements—based solely on the observed correlations.

Bell's Theorem has undergone a long history of experimental scrutiny. This started in the 1970s, with the experimental proposal of Clauser, Horne, Shimony and Holt [Cla+69] which sparked the first experimental test of Bell nonlocality [FC72]. This test, while important, relied on additional assumptions beyond those intrinsic to Bell's Theorem. Over the years, the experimental tests grew in precision and sophistication (for example, Refs. [AGR81; AGR82; Wei+98] among others), culminating in a series of experiments in 2015 that were designed to eliminate any extra assumptions, and address all the loopholes that had been a subject of debate in earlier experiments [Giu+15a; Sha+15; Hen+15]. These experiments lead to a broad consensus within the scientific community about the empirical validation of Bell's Theorem. In 2022, Alain Aspect, John Clauser and Anton Zeilinger were awarded the Nobel Prize in Physics for their contributions to the experimental verification of Bell's Theorem [Nev22]. For a more in-depth review of Bell nonlocality, see Ref. [Bru+14].

Beyond Bell's Theorem. Bell's Theorem has been crucial for demonstrating the inadequacy of classical notions of causality to explain phenomena at nature's most fundamental level. However, it does not provide guidance on *how* to understand causality at this fundamental level. In classical settings, the concept of causality is well-defined, where causal influences are seen as functional relationships between classical variables, and probabilities represent incomplete knowledge about these variables. This clarity, however, does not extend to quantum contexts. Here, the interpretation of probability, or amplitudes, remains ambiguous and largely depends on one's preferred interpretation of quantum theory. Furthermore, quantum theory allows for causal phenomena with no classical counterparts. For instance, quantum

theory seems to be compatible with a superposition of causal orders [OCB12; Chi+13], the significance of which is difficult to understand, and where fundamental questions like its physical realizability are still a matter of debate within the scientific community [Ore19; VR23].

There exist proposals which expand the classical notions of causality to include quantum phenomena, such as the framework of *quantum causal models* [All+17; BLO20; BLO21]. These proposals, while mathematically consistent, do not fully solve all conceptual issues surrounding the notion of causality in quantum theory. However, they provide a framework in which to systematically explore the full range of quantum behaviour. Given that our interaction with the world is in terms of classical information, it is meaningful to study the classical information produced by quantum causal models, that is, classical probability distributions. These distributions can be contrasted against those that can be generated by classical theories in a similar "causal scenario".

The concept of a "causal scenario" may be articulated independently of any specific theory through what is known as the *operational framework* [DCP17]. The operational framework describes processes directly controllable in a laboratory, such as preparations, transformations, and measurements that can be performed on physical systems, and the probability distributions that are observable when connecting such processes together in a circuit. A quantum causal model is understood as such a circuit, with its elements, or processes, following the rules of quantum theory. By comparing the correlations achievable in the same circuit structure but with classical processes as elements, a fair comparison can be made between quantum and classical theories.

Comparing classical and quantum correlations can reveal new forms of *quantum advantage*, defined as operational tasks where quantum theory outperforms classical theories. These tasks can be expressed as a game between distant players, each equipped with a device in the circuit, that can be won with a higher probability via quantum as opposed to classical processes in the circuit. An intriguing question is whether all the nonclassical features of quantum theory can lead to quantum advantage. In other words: *Can the fundamental difference between quantum and classical theories be operationally characterized through causal scenarios?* Such a characterization would improve our understanding of causality at the fundamental level. This improved understanding may lead to progress on other fronts, such as in

developing a theory of quantum gravity, where causal order can potentially be subject to quantum uncertainty [Har09]. It may also shed light on controversial issues such as the measurement problem or, as suggested by recent evidence [WC20b; WC20a; Ren+21], potentially inspire new physical principles, based on causality, that may uniquely define quantum theory among all possible theories.

Vision of this thesis. This thesis aims to operationally characterize the fundamental differences between classical and quantum theories within causal scenarios. Such an understanding may eventually help integrate quantum phenomena into a coherent, conceptually clear framework of causality. Towards this goal, we explore how classical and quantum causal models diverge in operational tasks in specific causal scenarios. We focus on simple scenarios that go beyond Bell's, while seeking to discover new forms of quantum advantage that are fundamentally different from traditional Bell nonlocality. Our goal is to link these new forms of quantum advantage to different nonclassical features of quantum theory and study their potential applications. A critical component of this research is testing for the causal compatibility of specific correlations with a given causal model. As such, an important part of this thesis is dedicated to expanding and refining the scope of current methods for testing causal compatibility.

1.2 Main thesis contributions

In this thesis, we start by considering two families of causal scenarios that lie in the middle ground between the common-cause scenario and general scenarios:

- **Network scenarios.** In a typical common-cause scenario, a single state is prepared and distributed among various parties, each equipped with a measurement device, and where the parties are sufficiently separated such that their measurement events are space-like separated. In contrast, network scenarios involve multiple, independent state preparations. These preparations are distributed among subsets of the distant parties but, crucially, not all parties need to have access to every state preparation. This differs from common-cause scenarios, where the state preparations are accessible to all involved parties.

- **Broadcast scenarios.** Common-cause and network scenarios involve only "trivial" transformations, i.e., those that can be included as part of the preparations or measurements. Broadcast scenarios involve a single state preparation together with channels which distribute, or "broadcast", shares of the initial prepared state to multiple parties, which themselves can implement space-like separated measurements. These parties themselves may further broadcast their shares of the state to other parties and so forth.

Individually, these scenarios allow for the focused study of specific aspects of general causal scenarios, while when composed together, they can describe for the whole spectrum of causal behaviour. In **Chapter 2** we provide formal **definitions** of these and more general causal scenarios and present the primary **methods** used for their characterization.

In **Chapter 4**, we analyse quantum advantage in **network scenarios**. The study of network scenarios has recently become a subfield of its own [Tav+22]. In part, the growing interest in networks is due to their ability to manifest nonclassical phenomena with no counterpart in common-cause scenarios. One example is given by the phenomenon of *entanglement swapping*, whereby entanglement is distributed between initially independent, distant parties in the network [Żuk+93]. Additionally, networks allow for the generation of *nonclassicality "without inputs"*, i.e. without requiring local, independent measurement choices [Fri12; Ren+19a]. This constitutes a new phenomenon, as common-cause scenarios cannot lead to nonclassical correlations without such measurement choices. Furthermore, the ongoing technological development of *large-scale quantum networks* motivates more practical concerns [Kim08; WEH18; KW19]. These include the development of novel applications for quantum networks, or the creation of new, noise-robust methods for certifying the proper functioning of quantum networks in real-world settings.

We focus our study on **single-photon nonlocality** in network scenarios. A quantum state is said to be nonlocal if it can lead to Bell nonlocal correlations. We focus particularly on the following single-photon entangled state:

$$|\psi^+\rangle_{AB} = \frac{|01\rangle_{AB} + |10\rangle_{AB}}{\sqrt{2}}, \tag{1.1}$$

which is obtained by sending a single photon into a balanced beamsplitter, where $|10\rangle_{AB}$ (resp. $|01\rangle_{AB}$) represents the photon occupying optical mode A (resp. B). The resulting state is entangled in the two spatial modes. Whether this state can lead to nonlocal correlations has been a matter of intense debate (e.g., [TWC91; YS92a; YS92b; Har94; Ger96; Vai95; AV00; Hes+04; D'A+06; BCB13; Mor+13a; Don+14; Das+21; GPS21]). An affirmative answer can be given either when the mode populations can be transferred to *excitations of massive particles* [Ger96; Vai95; AV00] or, if restricted to purely optical means, when using *active measurements* [TWC91; Har94; BCB13; Das+21] which locally insert extra photons in the modes during the measurement process. It can be argued that without active measurements, measuring the state $|\psi^+\rangle_{AB}$ allows for the parties to deduce the path of the photon, thereby decohering the state. This implies that *passive measurements*, such as phase shifters, beamsplitters and photodetectors, cannot be used to produce Bell nonlocal correlations from a single-photon entangled state.

Our contributions. The main result our work is to show that *single-photon entangled states, when placed in a network, can generate nonclassical correlations through passive measurements alone*, which arguably is not possible in the Bell scenario. Moreover, we manifest a strong form of nonclassicality with no equivalent in the Bell scenario, namely, without the need for independent measurement choices. We achieve this by arranging three copies of the state $|\psi^+\rangle_{AB}$ within a *triangle causal network* (cf. Chapter 4 for more details). The network's structure induces indeterminancy in the photon's path, necessary for exploiting the superposition of the modes. The combination of passive measurements only, and the lack of need of independent measurement choices, arguably makes this of the *simplest experimental proposals* for demonstrating the nonlocality of a single-photon entangled state. Our published work can be accessed at the following reference:

> Paolo Abiuso, Tamás Kriváchy, **Emanuel-Cristian Boghiu**, Marc-Olivier Renou, Alejandro Pozas-Kerstjens, and Antonio Acín, *"Single-photon nonlocality in quantum networks"*, Phys. Rev. Research 4, L012041—Published 29 March 2022

Subsequent works. Following the publication of our work, there have

been some important advancements in the study of the triangle causal network. For instance, Ref. [Pol+23] reports the experimental verification of nonclassical correlations without measurement choices in a triangle network. The triangle network in this experiment is built with one quantum state entangled in polarization degrees of freedom and two classical states. Another recent experiment, presented in Ref. [Wan+24], uses three entangled pairs of photons and entangling measurements—the so-called "elegant" joint measurement [Gis19]—and report machine-learning based evidence of nonlocality in the experiment. Other works include the development of partial self-testing techniques that do not require measurement choices [SBB23], improvements in methods for causal compatibility [PKGR23a], studies of particularly symmetric quantum distributions [GG23] and simpler versions of the triangle scenario [Bor+23b]. Additionally, there has been a focus on the analysis of post-quantum correlations in the triangle scenario [PK+23], along with similar extensions and experimental implementations [Cao+22; Mao+22]. Particularly relevant is a recent preprint which presents novel noise-robust proofs of nonclassicality for the triangle network [Bor+23a].

In **Chapter 5**, we study fundamental applications of **broadcast scenarios**. We focus on their potential to manifest nonclassical correlations by leveraging quantum phenomena that cannot produce nonclassicality in common-cause scenarios. For example, it is known that certain quantum states, despite being entangled, have a so-called *local hidden variable (LHV) model* [Wer89; ADA14; Hir+17]. A LHV model for an entangled quantum state provides a classical explanation, in terms of common causes, for any possible correlation observed when performing local, independently chosen measurements on the state. There also exist examples of incompatible quantum measurements—the other resource for Bell nonlocality—which cannot lead to nonclassical correlations [BV18; HQB18; Gho+23]. The existence of LHV models proves that not all of nonclassical phenomena in quantum theory can necessarily lead to Bell nonlocal correlations.

A key development was the discovery of the phenomenon of **activation of nonlocality**. This phenomenon describes how states with LHV models can nevertheless lead to nonclassical correlations—be "activated"—in more complex causal scenarios. These scenarios can be broadly categorized into *single-copy* and *multiple-copy* scenarios. Single-copy scenarios achieve activation through a single copy of the quantum state and multiple

rounds of measurements, such as local filtering operations [Pop95; Hir+16] or more generally, time-ordered sequential measurements [Ż+98; Gal+14]. In contrast, in multiple-copy scenarios, nonlocality is activated through the simultaneous joint measurement of multiple copies of the quantum state, either in parallel [NV11; Pal12] or distributed in a network of distant parties [Sen+05; Cav+11], aided by filtering operations. The broadcast scenario, initially introduced in Ref. [BHC21], constitutes a novel scenario that can activate entangled states with LHV models using a single copy of the state, a broadcasting channel and local measurements.

Our contributions. We expand on the work of Ref. [BHC21] in several directions. We start by constructing Bell inequalities specific to the broadcast scenario, and show how broadcasting can lead to stronger notions of activation of nonlocality. In particular, we exploit these ideas to show that entangled states admitting a LHV model for the most general class of quantum measurements can lead to genuine tripartite nonlocal correlations, as defined in Ref. [Ban+13]. We then analyse DI entanglement certification in the broadcast scenario, and show through convex optimization techniques that DI entanglement certification is possible for the two-qubit Werner state [Wer89] in almost its entire range of entanglement. Lastly, we extend the concept of EPR steering to the broadcast scenario, and present new examples of activation for the two-qubit isotropic state. Our published work can be accessed at the following reference:

> **Emanuel-Cristian Boghiu**, Flavien Hirsch, Pei-Sheng Lin, Marco Túlio Quintino, Joseph Bowles, *"Device-independent and semi-device-independent entanglement certification in broadcast Bell scenarios"*, SciPost Phys. Core 6, 028 (2023)—Published 11 April 2023

Subsequent works. Concerning LHV models, Ref. [Des+23] expands the family of known bipartite and tripartite entangled states with LHV models. Regarding the broadcast scenario, the preprint of Ref. [Rod+23] uncovers new Bell inequalities via the exclusivity graph approach. Another recent preprint reports the experimental confirmation of activation of nonlocality via the broadcasting of quantum states [VA+23]. This work highlights the potential real-world applicability of the broadcasting technique, particularly for protocols fuelled by noisy quantum states with LHV models, that is, states not viable in standard DI protocols based on Bell nonlocality.

In **Chapter 6** we turn our attention towards **methods for testing causal compatibility in general causal scenarios**. Recent developments lead to the creation of numerous such methods, such as those that utilize conditional independence relations [BGP10; Bra+12; Ros+16; Tav+14; Cha16; PK+19], entropic constraints [WC17a], correlation matrices [Å+20; Kel+20; Kra+21; BR22], Finner inequalities [Ren+19b; Shu+23; Luo21], token-counting [Ren+19a; RB22c; RB22a] or the principle of device replication [WSF19; NW20a; Wol+21]. For a complete review of these techniques, we point to Ref. [Tav+22]. We focus on the method based on the principle of device replication, known as the **inflation technique** [WSF19; NW20a; Gis+20a; Wol+21], as it is the most promising method to tackle causal compatibility with full generality.

The inflation technique considers *gedankenexperiments* where one has access to multiple copies of the elements that connect together to define the causal scenario under study. The copies of these elements are assembled into a new, expanded "inflated" scenario. This inflated scenario includes subsets of elements connected in a manner which mirrors the connections of the original scenario. Tests for causal compatibility in the inflated scenario are simpler due to the ability to exploit additional symmetries of the scenario. The inflation technique can be used for causal compatibility with classical [WSF19; NW20a], quantum [Wol+21] and even beyond-quantum sources of correlations [WSF19; Gis+20a]. Inflation methods are, in general, **technically complex to implement** due to the large amount of symbolic computations initially required. As a result, current implementations are restricted to small, specific networks and particular inflated causal scenarios (see, e.g., the computational appendices of [PKGT22; PKGR23a]).

Our contributions. We develop a **Python library** for implementing the inflation hierarchies of Ref. [Wol+21; NW20a]. The library can test for causal compatibility with network scenarios with classical, quantum and no-signaling sources of correlations, and also optimize linear functions of the correlations. In addition, we expand the scope of the original technique of Ref. [Wol+21] to test for possibilistic causal compatibility with networks, that is, to test whether a partition of all the possible events into a group of strictly possible and strictly impossible events is realizable in the causal network. To promote ease of use, contributions and further development, the library was developed following best software development practices, including modular

architecture, a comprehensive testing suite and detailed documentation. Our published work can be accessed at the following reference:

> **Emanuel-Cristian Boghiu**, Elie Wolfe, and Alejandro Pozas-Kerstjens, *"Inflation: a Python library for classical and quantum causal compatibility"*, Quantum 7, 996 (2023)—Published 4 May 2023

The package is hosted on GitHub[1], with a documentation website[2] and it is also distributed on the Python Package Index[3].

Subsequent works. Following the publication of our library, it has been used in different works, such as in estimating the volume of sets of correlations in networks [Cam+23] and for evaluating the compatibility of distributions with the triangle causal scenario [PK+23].

Lastly, in **Chapter 7** we conclude this thesis with an overview of important open problems in the field of quantum causality.

[1] https://github.com/ecboghiu/inflation
[2] https://ecboghiu.github.io/inflation/_build/html/index.html
[3] https://pypi.org/project/inflation/

Chapter 2

Preliminaries

In this chapter we introduce the basic concepts and definitions that will be used throughout this dissertation. We present the concept of a causal scenario within the operational framework based on the approach of Ref. [DCP17]. We then study the sets of correlations achievable in the common-cause scenario (to be defined) under different assumptions on the nature of the common-cause.

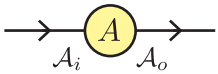
2.1 Definition of a causal scenario

We define the concept of a causal scenario in the operational framework, together with its pictorial representation in terms of directed-acyclic-graphs (to be defined), following the approach of Ref. [DCP17]. We adapt their terminology and notation to fit the conventions used in this thesis.

The operational framework can be understood as an extension of probability theory [DCP17, Ch. 3], just as probability theory can be seen as an extension of logic [Cox01; Jay03]. At the heart of probability theory lies the concept of *event* and *probability*. The operational approach further enhances this view through the idea of *connectivity of events*. An event can be described, without loss of generality, as a moment when classical information about a *physical process* is obtained through a *physical device*. For example, this could involve acquiring some information about a process by noting whether a display shows "0" or "1" during the run of an experiment. Physical processes act upon and transform physical systems. The system before the modification is the *input* to the process, while the system after the modification is the *output* of the process. It is important to note that the input and output may be different types of systems; for example, the

input could be the spin of a particle represented as a qubit, and the output could be a qutrit encoded in the modes of the electromagnetic field.

Definition 2.1.1 (Physical process)**.** A physical process $\mathcal{A}$ is a transformation of an *input* system, $\mathcal{A}_i$, into an *output* system, $\mathcal{A}_o$, accompanied by an associated set of events, $\{A_a\}_{a\in N}$ where $N = \{1, \ldots, n\}$ for some integer n.[1] The set of events represents some classical information about the process obtained through an experimental device. We define a *random variable*, A, which maps events to their integer index, $A(A_e) = e$. Probabilities will be specified over the values of the random variable, rather than over events. Graphically, we depict it as follows:



where the process is identified through its random variable, A, and wires represent systems, with arrows indicating whether the system is an input or output system.

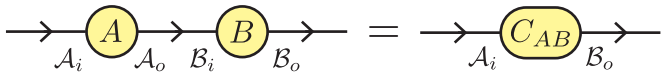
Definition 2.1.2 (Hidden/latent/deterministic process)**.** A physical process is hidden or latent whenever its associated set of events has a single element, i.e., it is deterministic. Graphically, it can be represented as follows:

where we use the colour grey and dashed lines to represent that it is a hidden process. We have no classical information about a hidden process except that it occurs, and how it connects with other processes. Note that here we use "deterministic" not in the sense that the outcomes follow a probability distribution that is deterministic in a sample space of various events, but rather that the sample space contains a single, trivial, event.

The connectivity of events relates to how processes and systems compose together.

Definition 2.1.3 (Sequential process composition)**.** Given a process with input $\mathcal{A}_i$, output $\mathcal{A}_o$ and events $\{A_a\}_{a\in N_A}$ and another with input $\mathcal{B}_i$, output $\mathcal{B}_o$ and events $\{B_b\}_{b\in N_B}$, their sequential composition is a process with input $\mathcal{A}_i$, output $\mathcal{B}_o$ and events $\{C_{(a,b)}\}_{(a,b)\in N_A\times N_B}$ provided the system types of $\mathcal{A}_o$ and $\mathcal{B}_i$ are equal. Graphically, we depict it as follows:

[1] A physical process is known as a *test* in Ref. [DCP17].

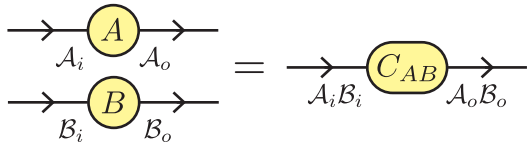


where $\mathcal{A}_o = \mathcal{B}_i$ and where C_{AB} denotes that the random variable C is indexed by the outcomes of A and B.

Definition 2.1.4 (System composition)**.** Two physical systems, $\mathcal{A}$ and $\mathcal{B}$ can be composed together into a joint system $\mathcal{AB}$.

Remark 2.1.1. A process with several inputs (respectively, outputs) can be seen as a single input (resp. output) process through system composition.

Definition 2.1.5 (Parallel process composition)**.** Given a process with input $\mathcal{A}_i$, output $\mathcal{A}_o$ and events $\{A_a\}_{a\in N_A}$ and another with input $\mathcal{B}_i$, output $\mathcal{B}_o$ and events $\{B_b\}_{b\in N_B}$, their parallel composition is a process with input $\mathcal{A}_i\mathcal{B}_i$, output $\mathcal{A}_o\mathcal{B}_o$ and events $\{C_{(a,b)}\}_{(a,b)\in N_A\times N_B}$. Graphically, it can be represented as follows:



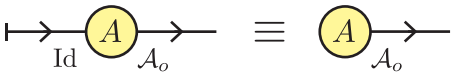
where C_{AB} denotes that the random variable C is indexed by the outcomes of A and B.

More definitions can be given concerning systems and processes, such as operational equivalence of physical systems and processes, but they are not relevant for this thesis.

Next, we introduce the concept of the trivial system. This allows us to define two special types of processes, namely, preparations and measurements.

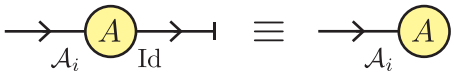
Definition 2.1.6 (Trivial system, Id)**.** The trivial system is characterized as either the absence of a system, or a system that transmits classical information. It acts as the identity element in the composition of systems. In graphical representation, it is depicted either by a wire intersected by an orthogonal segment or by the absence of a wire, indicating no input or output system. For examples, see Definitions 2.1.7 and 2.1.8.

Definition 2.1.7 (Preparation process)**.** A preparation process is defined as a process where the input system is the trivial system. Graphically, it can be represented as follows:



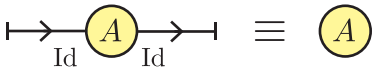
either with the input wire as the identity (left) or absent altogether (right).

Definition 2.1.8 (Measurement process)**.** A measurement process is defined as a process where the output system is the trivial system. Graphically, it can be represented as follows:



either with the output wire as the identity (left) or absent altogether (right).

Definition 2.1.9 (Pure randomness)**.** When a process has both its input and output as the trivial system, this indicates randomness that is causally independent of the rest of the circuit. Graphically, we represent it as follows:



The existence of this type of randomness is taken as an axiom in this thesis.

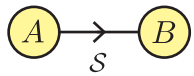
We are now ready to define a causal scenario.

Definition 2.1.10 (Causal scenario/causal structure)**.** A causal scenario, often referred to as a causal structure, corresponds to a **closed circuit** created through the composition of physical processes. A circuit is considered closed if the only unconnected systems are the trivial ones. These processes can be conceptualizes as nodes, and the systems as edges in a directed graph, represented by $\mathcal{G}$. The direction of these edges indicates the inputs and outputs linked to each process. Additionally, a causal scenario is associated with a **joint probability distribution** $p(A{=}a, B{=}b, \ldots)$ which relates to the events $A, B, \ldots$ tied to the processes within the causal scenario. Whenever it leads to no confusion, the joint probability will be expressed succinctly only in terms of the values of the random variables, $p(ab\ldots)$. In this thesis, the term "probability distribution" is interchangeably used with "**correlation**".

Definition 2.1.11 (Acyclic causal scenario)**.** Acyclic causal scenarios are characterized by the absence of closed loops. These scenarios correspond to

directed acyclic graphs (DAGs). Acyclic scenarios avoid causal paradoxes akin to the "grandfather paradox", where an effect can be its own cause. Throughout this thesis, we will focus only on acyclic causal scenarios.

Example 2.1.1 (Prepare and measure scenario)**.** One of the simplest causal scenarios is that of a preparation followed by a measurement, as shown in the following illustration:



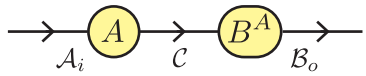
We define causality in terms of statistical independence of events.

Axiom (Causality)**.** The marginal probability of an event in a process is independent of any subsequent processes connected to it. This forbids signaling from future events to the past, where the arrow of time is indicated by the input and output system labels.

Example 2.1.2 (Causality in the prepare and measure scenario)**.** In the scenario of Example 2.1.1, the causality axiom is expressed as the marginal of A being independent of the posterior choice of measurement process. That is, $p(a) := \sum_b p(ab) = \sum_{b'} p(ab')$ where $p(ab')$ represents the probability of a scenario with measurement process B' instead of B.

That the probability distribution of the first process event is independent of the choice of posterior process allows for conditioning the choice of a process on a previous event.

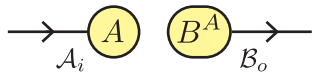
Definition 2.1.12 (Conditioned process)**.** Given a process with input $\mathcal{A}_i$, output $\mathcal{A}_o$ and events $\{A_a\}_{a\in N_A}$ and another process with input $\mathcal{B}_i$, output $\mathcal{B}_o$ and events $\{B_b^a\}_{b\in(N_B)_a}$ for all $a \in N_A$, the conditioned process is a process with input $\mathcal{A}_i$, output $\mathcal{B}_o$ and events $\{(A_a, B_b^a)\}_{(a,b)\in\cup_a\{\{a\}\times(N_B)_a\}}$. Graphically, it can be represented as follows:



where B^A denotes that the set of events of B is conditioned on the events of A and $\mathcal{C} = \mathcal{A}_o = \mathcal{B}_i$. The conditioned process B^A can equivalently be understood as a set of processes, and where the outcome of A determines which specific process from the set is implemented.

Remark 2.1.2. A conditioned process simplifies to a sequential composition whenever the number of events of the conditioned process does not depend on the conditioned variable, i.e., $|(N_B)_a| = |(N_B)_{a'}|$ for all $a, a' \in N_A$ in Definition 2.1.12.

Example 2.1.3 (Measure and prepare scenario)**.** A special type of conditioned process is one where the system connecting process A and B^A is the trivial system, $\mathcal{C} = \mathrm{Id}$:



This corresponds to the preparation of a new system conditioned on a previous measurement result.

The causality axiom also allows for post-selection of preparations, which allows for deterministic preparations of physical systems.

Definition 2.1.13 (Post-selection)**.** Post-selection refers to discarding certain observed events from the experimental data table from which the probabilities are constructed.

Definition 2.1.14 (Local post-selection)**.** A local post-selection is one where the decision to accept or discard an observed event is made independently of the criteria applied at a different node in the causal scenario. Such local post-selections need not occur after the main experiment ("post"); they can also be simulated during the experiment through local filtering techniques. In this context, local filtering means that the output system of a process is passed on as input to a subsequent process only if a specific event is observed. If this event is not observed, the experimental run is discarded.

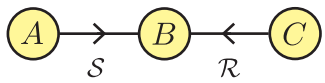
Definition 2.1.15 (Deterministic preparation)**.** In the prepare-and-measure scenario of Example 2.1.1 one may post-select on the preparation A keeping only a subset of the possible outcomes. When only one outcome is kept, this can be understood as a deterministic state preparation (also known as heralding). This is allowed by the causality axiom, as the probability of preparation is independent of any posterior process connected to it.

Remark 2.1.3 (Arbitrary preparation probability). Post-selection allows for more than just deterministic state preparation; it also enables the simulation

of a specific frequency of preparation, $p'(a)$, based on the natural frequency of the process $p(a)$. In a given causal scenario, these probabilities can be effectively simulated through local post-selection. Therefore, we can without loss of generality consider just the joint probability distribution conditioned on the observable preparations of the causal scenario.

Example 2.1.4 (Conditioned distribution in the prepare and measure scenario)**.** In the prepare-and-measure scenario of Example 2.1.1, per Remark 2.1.3, the joint probability may be written as $p(ab) = p(b|a)p(a)$. Given that the probability of preparation $p(a)$ can be fixed arbitrarily, it is sufficient to consider only the joint distribution conditioned on the preparation, $p(b|a)$. While the amount of data is the same, the usefulness of considering just the conditioned distribution will be apparent when exploring different causal scenarios.

Example 2.1.5 (Double prepare and measure scenario)**.** Consider the following causal scenario, where B is a process that performs a joint measurement on two independent state preparations:



such that $b = 0$ if $a = c$ and $b = 1$ if $a \neq c$. This is reflected through the lack of a connected one-way path between A and C. While the full probability table respects the causality axiom, i.e., $\sum_b p(abc) = p(ac) = p(a)p(c)$, a post-selection on the values of B induces correlations between A and C which are not compatible with the causal scenario. On the other hand, local post-selections on A and C lead to probability tables that respect the causal constraints of the scenario.

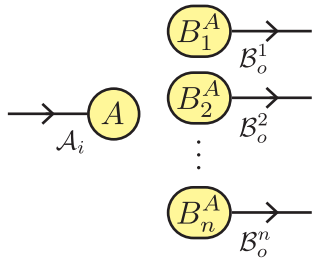
2.2 Classical, quantum and no-signaling processes

Now we will describe different types of processes: classical, quantum and no-signaling processes.

2.2.1 Classical process

Using the notion of a conditioned process (Definition 2.1.12), we define a classical process as any physical process with input $\mathcal{A}_i$ and output $\mathcal{B}_o$

which is operationally indistinguishable from a measure and prepare process (Example 2.1.3), which prepares a state on $\mathcal{B}_o$ conditioned on a measurement of $\mathcal{A}_i$. This captures the intuition that any classical process can be thought of as a sequence of three steps: measuring the exact state of the input, simulating the output of the process based on this measurement and then physically preparing the simulated output state. Furthermore, if the process has not only one output but a collection of output systems, $\{\mathcal{B}_o^i\}_i$, then a classical process will implement parallel preparations on $\{\mathcal{B}_o^i\}_i$, conditioned on the result of the measurement of the input:



This also captures the intuition that a joint classical system consists of a statistical mixture of locally well-defined classical systems. Additionally, it implies that the output of a classical process can be arbitrarily cloned.

In order to define a quantum process, we first give an overview of the basic elements of quantum theory while also introducing the notation used throughout the thesis.

2.2.2 A short review of finite-dimensional quantum theory

Quantum theory is built on three main elements: *states*, *transformations* and *measurements*, which are defined within the framework of Hilbert spaces. We denote the set of operators acting on a Hilbert space, $\mathcal{H}$, by $L(\mathcal{H})$, and the set of positive semidefinite operators by $\mathrm{PSD}(\mathcal{H})$.

Quantum states. A quantum state, ρ, is a positive semidefinite operator, written as $\rho \succeq 0$, with trace one, $\mathrm{tr}\,\rho = 1$, acting on a Hilbert space $\mathcal{H}$.

- *Purity.* A state is **pure** if its rank is equal to one. Pure states can be uniquely represented by a projector onto a normalised vector, $|\psi\rangle\langle\psi|$,

where $\langle\psi|\psi\rangle = 1$, $\langle\psi| := (|\psi\rangle)^\dagger$ and $(\cdot)^\dagger$ represents the Hermitian adjoint. Non-pure states are called **mixed**. Through the spectral decomposition theorem, any mixed state can be expressed as a statistical mixture of pure states:

$$\rho = \sum_x p(x)|x\rangle\langle x|, \tag{2.1}$$

where $\{|x\rangle\}_x$ defines an orthonormal basis, and $p(x) : \mathbb{N} \to \mathbb{R}$ satisfies $\sum_x p(x) = 1$ and $p(x) \geq 0$ for all x.

- *Convexity.* The set of quantum states is convex. This means that if ρ_1 and ρ_2 are quantum states acting on the same Hilbert space, then $t\rho_1 + (1-t)\rho_2$ for any t in the range $[0, 1]$ is also a valid quantum state.

- *Composition.* States ρ_A and ρ_B, each acting on different Hilbert spaces $\mathcal{H}_A$, $\mathcal{H}_B$, can be combined to form a joint state $\rho_{AB} = \rho_A \otimes \rho_B$ acting on the joint Hilbert space, $\mathcal{H}_{AB} = \mathcal{H}_A \otimes \mathcal{H}_B$.

- *Separability.* A joint state acting on $\mathcal{H}_{AB}$ is considered separable across the bipartition $(A|B)$ if it can be expressed as a convex combination of product states acting on $\mathcal{H}_A$ and $\mathcal{H}_B$:

$$\rho_{AB} = \sum_i t_i \, \rho_A^i \otimes \rho_B^i \quad \sum_i t_i = 1, \, t_i \geq 0 \; \forall i. \tag{2.2}$$

  A state that is not separable is said to be **entangled**.

- *Purification.* The concept of purification describes how any mixed state acting on $\mathcal{H}_S$ can be seen as a pure state acting on a larger Hilbert space $\mathcal{H}_S \otimes \mathcal{H}_A$. The larger Hilbert space includes an auxiliary system $\mathcal{H}_A$ which we have no control over. Having "no control" over part of a Hilbert space is modelled through the *partial trace*, defined as the unique linear operator $\mathrm{tr}_A : \mathcal{H}_S \otimes \mathcal{H}_A \to \mathcal{H}_S$ such that $\mathrm{tr}((M \otimes \mathbb{1}_A)N) = \mathrm{tr}(M(\mathrm{tr}_A N))$ [Par12]. Therefore, the purification property can be formally written as follows:

$$\rho_S = \mathrm{tr}_A \, |\psi\rangle\langle\psi|_{SA}, \tag{2.3}$$

  where ρ is an arbitrary state acting on $\mathcal{H}_S$ and $|\psi\rangle_{SA} \in \mathcal{H}_S \otimes \mathcal{H}_A$. From the spectral decomposition of the state, $\rho_S = \sum_i p(x)|x\rangle\langle x|_S$, we can build the purified state as $|\psi\rangle_{SA} = \sum_x \sqrt{p(x)} \, |x\rangle_S \otimes |x\rangle_A$

where $\{|x\rangle_A\}_x$ defines an orthonormal basis on $\mathcal{H}_A$. The minimal dimension of $\mathcal{H}_A$ for a purification to exist depends on the rank of the quantum state. It can be shown that all purifications are equivalent, as they only differ by a local change of basis on the auxiliary system, i.e., $|\psi\rangle_{SA} = \mathbb{1}_S \otimes U_A|\psi\rangle'_{SA}$ where $|\psi\rangle'_{SA}$ is another purification that satisfies Eq. (2.3) and U_A is an orthonormal change of basis matrix.

The properties of convexity, composition and separability, although presented for two states and two subsystems, can be easily generalized to any number of states and subsystems.

Quantum transformations. A quantum map, $\Lambda : L(\mathcal{H}_I) \to L(\mathcal{H}_O)$, is a linear map between operators, $\Lambda(\alpha_1\rho_1 + \alpha_2\rho_2) = \alpha_1\Lambda(\rho_1) + \alpha_2\Lambda(\rho_2)$, where α_1, α_2 are scalars and ρ_1, ρ_2 are quantum states acting on $\mathcal{H}_I$.

- *Positivity.* A map is said to be *positive* if it transforms positive semidefinite operators into positive semidefinite operators, $\rho \succeq 0 \Rightarrow \Lambda(\rho) \succeq 0$.
- *Complete positivity (CP).* A map is said to be *completely positive* if the output of the channel is positive even when it acts locally on a state on a bigger Hilbert space, i.e., $\sigma \succeq 0 \Rightarrow \Lambda \otimes \mathbb{1}_A(\sigma) \succeq 0$ with $\sigma \in \mathrm{PSD}(\mathcal{H} \otimes \mathcal{H}_A)$, for any $\mathcal{H}_A$.
- *Trace preservation (TP).* A map is said to be trace preserving if $\mathrm{tr}(\Lambda(\rho)) = \mathrm{tr}(\rho)$.

For the next properties, we will restrict to endomorphisms, that is, $\mathcal{H}_I = \mathcal{H}_O$.

- *Unitality.* A map is said to be unital if it maps the identity to itself, $\Lambda(\mathbb{1}) = \mathbb{1}$.
- *Unitary.* A unitary map is one which corresponds to a orthonormal change of basis, $\Lambda(\rho) = U\rho U^\dagger$ where $UU^\dagger = U^\dagger U = \mathbb{1}$. A unitary is a completely positive and trace preserving (CPTP) map.
- *Channels.* Consider a mixed state, ρ, acting on a space $\mathcal{H}_S$. We may prepare in our laboratory an additional state ρ_E on an auxiliary space $\mathcal{H}_E$. Given perfect control over the auxiliary space, we can assume the state we prepare to be pure, $\rho_E = |0\rangle\langle 0|_E$ where $|0\rangle_E$ is an arbitrary

pure state on $\mathcal{H}_E$. Now we can jointly evolve $\mathcal{H}_S$ and $\mathcal{H}_E$ through a unitary, and then discard the auxiliary system E:

$$\Lambda(\rho) = \mathrm{tr}_E\, U^{SE}(\rho \otimes |0\rangle\langle 0|_E)(U^{SE})^\dagger. \tag{2.4}$$

Equivalently, one may absorb the state on the auxiliary space into the unitary defining an isometry $V^{S\to SE} := U^{SE}(\mathbb{1}_S \otimes |0\rangle_E)$ such that:

$$\Lambda(\rho) = \mathrm{tr}_E\, V^{S\to SE}\rho(V^{S\to SE})^\dagger.$$

Maps of this form are called *channels*, and they are CPTP maps. Note that we assume that we do not have control over the purification of ρ; if we did, we would take ρ to be pure. Importantly, all CPTP maps are also channels, and thus may be expressed as in Eq. (2.4) (cf. the Choi-Kraus theorem [Cho75a; Cho75c]).

- *Operator sum representation.* Let $M_a := (\mathbb{1}_S \otimes \langle a|_E)\, U^{SE}\, (\mathbb{1}_S \otimes |0\rangle_E)$ where $|a\rangle_E$ is an orthonormal basis on $\mathcal{H}_E$. It is easy to check the following identity:
$$\Lambda(\rho) = \sum_a M_a \rho M_a^\dagger. \tag{2.5}$$
This is the operator sum representation of a channel. The set of operators $\{M_a\}_a$ are referred to as *Kraus operators* and they satisfy $\sum_a M_a^\dagger M_a = \mathbb{1}_S$.

- *Dual channel.* We can define the dual channel, $\Lambda^\dagger(\rho) = \sum_a M_a^\dagger \rho M_a$. The dual channel is always unital, $\Lambda^\dagger(\mathbb{1}) = \mathbb{1}$, due to the property $\sum_a M_a^\dagger M_a = \mathbb{1}$. The dual channels satisfies $\mathrm{tr}[\Lambda(M)N] = \mathrm{tr}[M\Lambda^\dagger(N)]$.

- *Purification.* We have shown how given the unitary description of Eq. (2.4), we can build the operator sum representation (2.5). The reverse can also be done, namely, given a collection of Kraus operators $\{M_a\}_a$, one can build an isometry $V^{S\to SE} = \sum_a M_a \otimes |a\rangle_E$ where $\{|a\rangle_E\}_a$ is an orthonormal basis on $\mathcal{H}_E$ such that:
$$\Lambda(\rho) = \sum_a M_a \rho M_a^\dagger = \mathrm{tr}_E\, V^{S\to SE}\rho(V^{S\to SE})^\dagger. \tag{2.6}$$

- *Arbitrary dimensions.* More general channels between arbitrary Hilbert spaces, $\Lambda : L(\mathcal{H}_I) \to L(\mathcal{H}_O)$, can be realized through a bigger channel

acting on the tensor product of the input and output Hilbert spaces, $\Lambda^C : L(\mathcal{H}_I) \otimes L(\mathcal{H}_O) \to L(\mathcal{H}_I) \otimes L(\mathcal{H}_O)$, and feeding a junk state on the "output" part of the input space, and "forgetting" the "input" part of the output Hilbert space:

$$\Lambda(\rho) = \mathrm{tr}_I \left(\Lambda^C(\rho \otimes |0\rangle\langle 0|_O) \right). \tag{2.7}$$

If $\{M_i^C\}_i$ are the Kraus operators of Λ^C and we express them in a basis that is the product of an orthonormal basis $\{|\alpha\rangle_I\}_\alpha$ of $\mathcal{H}_I$ and $\{|\beta\rangle_O\}_\alpha$ of $\mathcal{H}_O$:

$$M_i^C = \sum_{\alpha,\beta,\alpha',\beta'} c^i_{(\alpha,\beta),(\alpha',\beta')} |\alpha,\beta\rangle\langle\alpha',\beta'|_{IO},$$

then the Kraus operators of Λ are $\{M_{(i,e)}\}_{(i,e)}$ where

$$M_{(i,e)} = \sum_{\beta,\alpha'} c^i_{(e,\beta),(\alpha',0)} |\beta\rangle_O\langle\alpha'|_I.$$

- *Choi-Jamiołkowski isomorphism.* Just as an operator M is fully specified through its action on a basis $\{|i\rangle\}_i$, a channel $\Lambda : \mathcal{H}_I \to \mathcal{H}_O$ is fully specified by its action on a complete basis of the space of operators in the input space, such as $\{|i\rangle\langle i'|\}_{i,i'}$ where $\{|i\rangle\}_i$ is a basis of $\mathcal{H}_I$. We can collect in an operator the image under Λ of all the basis elements $\{|i\rangle\langle i'|\}_{i,i'}$:

$$C^\Lambda = \sum_{i,i'} |i\rangle\langle i'|_I \otimes \Lambda(|i\rangle\langle i'|_I), \tag{2.8}$$

where $C^\Lambda \in \mathcal{H}_I \otimes \mathcal{H}_O$ is called Choi state of Λ. Given a Choi state, the image of ρ can be computed as follows:

$$\Lambda(\rho) = \mathrm{tr}_I \, C^\Lambda(\rho^T \otimes \mathbb{1}_O), \tag{2.9}$$

where $(\cdot)^T$ denotes the transpose. The Choi state is a positive operator, $C^\Lambda \succeq 0$, and furthermore, when tracing out the output system we get the identity, $\mathrm{tr}_O \, C^\Lambda = \mathbb{1}_I$, which expresses that there can be no signaling backwards in time. The mapping between the standard description of an operator map and its Choi state representation is known as the Choi-Jamiołkowski isomorphism [Cho75b].

- *Quantum instrument.* We define a quantum instrument Λ^E as a channel which keeps a record of the auxiliary system E after the global unitary in the definition of a channel in Eq. (2.4):

$$\Lambda^E(\rho) = \sum_a M_a \rho M_a^\dagger \otimes |a\rangle\langle a|_E. \tag{2.10}$$

Note that tracing out E in (2.10) recovers the standard definition of a channel. We may consider a fine-graining of the auxiliary system and trace out only parts of the information content of the auxiliary system:

$$\begin{aligned}
\Lambda^{E_2}(\rho) &= \mathrm{tr}_{E_1} \Lambda^{E_1 E_2}(\rho) && (2.11)\\
&= \sum_{a_1,a_2} M_{a_1,a_2} \rho M_{a_1,a_2}^\dagger \otimes |a_1, a_2\rangle\langle a_1, a_2|_{P_1 P_2} && (2.12)\\
&= \sum_{a_2} \left(\mathrm{tr}_{E_1} \sum_{a_1} M_{a_1,a_2} \rho M_{a_1,a_2}^\dagger \otimes |a_1\rangle\langle a_1|_{E_1} \right) \otimes |a_2\rangle\langle a_2|_{E_2} && (2.13)\\
&= \sum_{a_2} \Lambda_{a_2}^{E_1}(\rho) \otimes |a_2\rangle\langle a_2|_{E_2}. && (2.14)
\end{aligned}$$

Note that the set of maps $\{\Lambda_{a_2}^{E_1}\}_{a_2}$ are collectively trace-preserving, but not necessarily individually.

Quantum measurements. A quantum measurement is implemented by interacting a measurement probe P with the system S, and then reading information from the probe. Note that this is precisely the definition of a quantum instrument (2.10):

$$\Lambda^P(\rho) = \sum_a M_a \rho M_a^\dagger \otimes |a\rangle\langle a|_P. \tag{2.15}$$

Now we will describe the rules for measuring the state of the probe.

- *Born rule.* Given a pure state $|\psi\rangle \in \mathcal{H}$, the probability of measuring if it is state $|a\rangle$ where $\{|a\rangle\}_a$ defines an orthonormal basis of $\mathcal{H}$ is given by the Born rule:

$$p(a|\psi) = |\langle a|\psi\rangle|^2 = \mathrm{tr}[(|a\rangle\langle a|)(|\psi\rangle\langle\psi|)]. \tag{2.16}$$

If we have a mixed state, which can always be written as a statistical mixture of orthogonal pure states $\rho = \sum_x p(x)|x\rangle\langle x|$, then the probability of reading outcome a is a statistical mixture of the probabilities of reading outcome a of each pure state $|x\rangle\langle x|$:

$$p(a|\rho) = \sum_x p(x)p(a|x) = \sum_x p(x)\,\mathrm{tr}[(|a\rangle\langle a|)(|x\rangle\langle x|)] = \mathrm{tr}(|a\rangle\langle a|)\rho. \tag{2.17}$$

The Born rule may also be used for measuring if only *part* of a system is in a given pure state. In the case of the quantum instrument (2.15), the probability of the probe being in state $|a\rangle_P$ is given by:

$$p(a|\Lambda^P(\rho)) = \mathrm{tr}(\mathbb{1}_S \otimes |a\rangle)(\mathbb{1}_S \otimes \langle a|)\Lambda^P(\rho) = \\ = \mathrm{tr}\, M_a \rho M_a^\dagger = \mathrm{tr}\, M_a^\dagger M_a \rho, \tag{2.18}$$

where in the last step we used the cyclicity of the trace, $\mathrm{tr}\, AB = \mathrm{tr}\, BA$.

- *As an operator measure.* The probability of the probe being in state $|a\rangle_E$ is given by $\mathrm{tr}\, M_a^\dagger M_a \rho$. Recall that the Kraus operators satisfy $\sum_a M_a^\dagger M_a = \mathbb{1}_S$. Let use define $P_a \coloneqq M_a^\dagger M_a$. Then, the set $\{P_a\}_a$ defines an positive operator valued measure (POVM). We may identify a quantum measurement as a POVM. For orthogonal projectors, that is, $P_a P_{a'} = \delta_{aa'} P_a$ (with $\delta_{aa'}$ being the Kronecker delta function, i.e., $\delta_{aa'} = 0$ if $a = a'$ and $\delta_{aa'} = 1$ otherwise), the set $\{P_a\}_a$ constitutes a Projection-Valued Measure (PVM).

- *State update.* After reading that the probe is in state $|a\rangle_E$, our description of the state after the application of the instrument is updated to one where the state of the probe is well-defined:

$$\rho \to \frac{M_a \rho M_a^\dagger}{\mathrm{tr}\, M_a^\dagger M_a \rho} \otimes |a\rangle\langle a|_E. \tag{2.19}$$

  Note that now the joint state of the probe and the system is a product state.

- *Purification.* Analogous to the purification of mixed states, POVMs can be "purified" to a PVM through the Naimark's dilation (see Ref. [Wil13, Ch. 5.2.2]). For any measurement $\{P_a\}_{a\in N}$ on $\mathcal{H}_S$, there is an isometry

$V : \mathcal{H}_S \to \mathcal{H}_S \otimes \mathcal{H}_A$ defined by $V = \sum_i M_i \otimes |i\rangle$ such that:

$$P_a \equiv M_a^\dagger M_a = V^\dagger \left(\mathbb{1}_S \otimes |a\rangle\langle a|_A\right) V, \tag{2.20}$$

where $\{|a\rangle_A\}_a$ defines an orthonormal basis on $\mathcal{H}_A$, and $\{|a\rangle\langle a|_A\}_a$ defines a PVM on $\mathcal{H}_A$.

2.2.3 Quantum process

A quantum process corresponds to a *general quantum instrument*, which is a quantum channel that takes as input a quantum state, and outputs another quantum state accompanied by a classical variable. This classical variable represents the classical outcomes we observe in the laboratory.

2.2.4 No-signaling process

A no-signaling process is any process that satisfies the **Axiom of Causality**. Often this imposes constraints on the set of joint distributions achievable in a given causal scenario. This is exemplified in the next section, which presents causal scenarios that will appear throughout the thesis.

2.3 Bell's common cause scenario

Bell's scenario, shown in Figure 2.1, involves the deterministic preparation[2] by a process Λ of two physical systems, $\mathcal{S}_A$ and $\mathcal{S}_B$, that are then submitted to causally-disconnected measurement processes, A and B, which themselves take as input the outputs of processes X and Y. In this section we study how the possible joint probabilities over the events of the scenarios depend on assumptions made about the processes (whether they are classical, quantum or no-signaling). Per Remark 2.1.3, instead of the full joint probability distribution $p(A{=}a, B{=}b, X{=}x, Y{=}y)$ we can work with the conditional distribution $p(A{=}a, B{=}b|X{=}x, Y{=}y)$, written as $p(ab|xy)$ in simplified notation. Note that since Λ has a single outcome, it does not appear in the probability distribution.

[2]Recall that in this thesis, a deterministic preparation is merely a process whose associated classical event has a single outcome, i.e., there is no information available about the process except how it is wired with the other processes in the circuit.

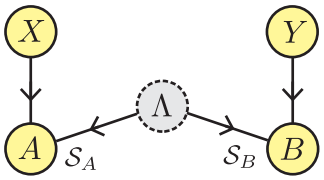
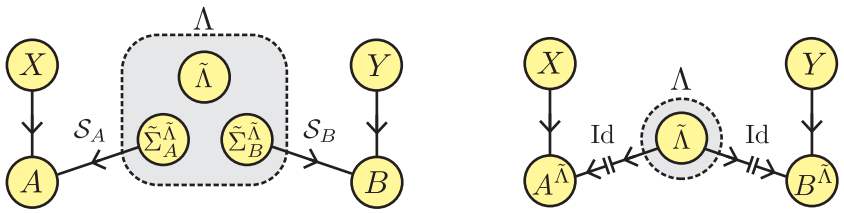


FIGURE 2.1: Bell's common cause scenario in which two physical systems $\mathcal{S}_A$ and $\mathcal{S}_B$ are prepared, and then measured independently by processes A and B which depend on X and Y.



(A) General deterministic classical process (B) Process sharing a random variable $\tilde{\Lambda}$

FIGURE 2.2: (2.2a) Bell's common cause scenario of Figure 2.1 with the assumption that the deterministic process Λ is a general (deterministic, i.e., single observable event) classical process, per Definition 2.2.1. (2.2b) A classical process which distributes a random variable $\tilde{\Lambda}$, and where the output systems are trivial systems.

2.3.1 Classical common cause

If the process Λ is classical, then as shown in Figure 2.2a, it is simulable with a measure and prepare scenario where on each output system $\mathcal{S}_A$, $\mathcal{S}_B$, there is a local state preparation $\tilde{\Sigma}_A^{\tilde{\Lambda}}$, $\tilde{\Sigma}_B^{\tilde{\Lambda}}$, conditioned on a measurement of the input $\tilde{\Lambda}$. Given that the input of Λ is the trivial system, then $\tilde{\Lambda}$ is a pure random variable, cf. Definition 2.1.9. Note that Λ and $\tilde{\Lambda}$ are independent random variables; one has a single outcome and is accessible experimentally, while the other is internal and there is no bound on the number of possible outcomes. The choice of using similar names, distinguished only by a tilde, is deliberate, aiming to facilitate the association of the internal random variable with its corresponding deterministic process.

Shared randomness. We will use q to denote the joint probability distribution over the events of Figure 2.2a, which includes internal variables $\tilde{\Lambda}$, $\tilde{\Sigma}_A^{\tilde{\Lambda}}$, $\tilde{\Sigma}_B^{\tilde{\Lambda}}$, whereas we will use p to refer to the experimentally accessible joint distribution over the events of Figure 2.1. The two are related, as when marginalizing over the unobservable events $\tilde{\Lambda}, \tilde{\Sigma}_A^{\tilde{\Lambda}}$, $\tilde{\Sigma}_B^{\tilde{\Lambda}}$ we recover the joint distribution p:

$$p(ab|xy) = \int_{\tilde{\lambda}\in N_{\tilde{\Lambda}}} \int_{\tilde{\sigma}_A \in N_A^{\tilde{\lambda}}} \int_{\tilde{\sigma}_B \in N_B^{\tilde{\lambda}}} q(ab\tilde{\lambda}\tilde{\sigma}_A\tilde{\sigma}_B|xy)\, \mathrm{d}\tilde{\lambda}\, \mathrm{d}\tilde{\sigma}_A\, \mathrm{d}\tilde{\sigma}_B \tag{2.21}$$

where $N_{\tilde{\Lambda}}$, $N_A^{\tilde{\lambda}}$ and $N_B^{\tilde{\lambda}}$ refer to the set of possible outcomes of $\tilde{\Lambda}$, $\tilde{\Sigma}_A^{\tilde{\Lambda}}$, $\tilde{\Sigma}_B^{\tilde{\Lambda}}$. Using the chain rule of conditional probabilities, we expand $q(ab\tilde{\lambda}\tilde{\sigma}_A\tilde{\sigma}_B|xy)$ as follows:

$$\begin{aligned} q(ab\tilde{\lambda}\tilde{\sigma}_A\tilde{\sigma}_B|xy) = q(a|xyb\tilde{\sigma}_A\tilde{\sigma}_B\tilde{\lambda})q(b|xy\tilde{\sigma}_A\tilde{\sigma}_B\tilde{\lambda})\cdot \\ \cdot\, q(\tilde{\sigma}_A|xy\tilde{\sigma}_B\tilde{\lambda})q(\tilde{\sigma}_B|xy\tilde{\lambda})q(\tilde{\lambda}|xy) \end{aligned} \tag{2.22}$$

Per the causality axiom, any variables that are not connected by a directed path in the causal scenario are statistically independent:

$$q(a|xyb\tilde{\sigma}_A\tilde{\sigma}_B\tilde{\lambda}) = q(a|x\tilde{\sigma}_A\tilde{\lambda}) \tag{2.23}$$

$$q(b|xy\tilde{\sigma}_A\tilde{\sigma}_B\tilde{\lambda}) = q(b|y\tilde{\sigma}_B\tilde{\lambda}) \tag{2.24}$$

$$q(\tilde{\sigma}_A|xy\tilde{\sigma}_B\tilde{\lambda}) = q(\tilde{\sigma}_A|\tilde{\lambda}) \tag{2.25}$$

$$q(\tilde{\sigma}_B|xy\tilde{\lambda}) = q(\tilde{\sigma}_B|\tilde{\lambda}) \tag{2.26}$$

$$q(\tilde{\lambda}|xy) = q(\tilde{\lambda}). \tag{2.27}$$

We may now rewrite Eq. (2.21):

$$p(ab|xy) = \int_{\tilde{\lambda}\in N_{\tilde{\Lambda}}} \underbrace{\left(\int_{\tilde{\sigma}_A\in N_A^{\tilde{\lambda}}} q(a|x\tilde{\sigma}_A\tilde{\lambda})q(\tilde{\sigma}_A|\tilde{\lambda})\,\mathrm{d}\tilde{\sigma}_A\right)}_{q(a|x\tilde{\lambda})}\cdot \tag{2.28}$$

$$\cdot \underbrace{\left(\int_{\tilde{\sigma}_B\in N_B^{\tilde{\lambda}}} q(b|y\tilde{\sigma}_B\tilde{\lambda})q(\tilde{\sigma}_B|\tilde{\lambda})\,\mathrm{d}\tilde{\sigma}_B\right)}_{q(b|y\tilde{\lambda})} q(\tilde{\lambda})\,\mathrm{d}\tilde{\lambda} \tag{2.29}$$

$$= \int_{\tilde{\lambda}\in N_{\tilde{\Lambda}}} q(a|x\tilde{\lambda})\,q(b|y\tilde{\lambda})\,q(\tilde{\lambda})\,\mathrm{d}\tilde{\lambda}\,. \tag{2.30}$$

The state preparations $\tilde{\Sigma}_A^{\tilde{\Lambda}}$, $\tilde{\Sigma}_B^{\tilde{\Lambda}}$ have been absorbed into the local response functions $q(a|x\tilde{\lambda})$, $q(b|y\tilde{\lambda})$. This means that whether the systems $\mathcal{S}_A$, $\mathcal{S}_B$ are prepared by Λ or later, as part of the measurement processes A and B, is not relevant. The key function of Λ in this context is to coordinate these local state preparations through the shared random variable $\tilde{\Lambda}$. As such, a classical process can be conceptually simplified to the idea of *distribution of classical information*, or "shared randomness". This is depicted in Figure 2.2b, where the systems being sent to A and B are trivial, emphasizing the role of shared randomness rather than any specific physical transmission.

Local causality. Regardless of whether Λ is a classical process, if it contains an internal variable $\tilde{\Lambda}$, the following equation is universally true: $p(ab|xy) = \int_{\tilde{\lambda}\in N_{\tilde{\Lambda}}} q(ab|xy\tilde{\lambda})\,q(\tilde{\lambda})\,\mathrm{d}\tilde{\lambda}$. The key defining property of a classical process is the factorization of the response function upon conditioning on the shared random variable $\tilde{\Lambda}$, as demonstrated in Eq. (2.30):

$$q(ab|xy\tilde{\lambda}) = q(a|x\tilde{\lambda})\,q(b|y\tilde{\lambda}).$$

That is, knowledge of the value of $\tilde{\Lambda}$ "explains" the correlations between A and B in the sense that conditioning on $\tilde{\Lambda}$ removes the apparent correlation. This is known as *Reichenbach's common cause principle* [Rei91]. Reichenbach's common cause principle together with the assumption of relativistic causality—the notion that causes must precede effects in all possible frames of reference—form the principle of *local causality* [WC17b].

Local determinism. The variable $\tilde{\Lambda}$ does not have a predefined number of outcomes, that is, its cardinality is unspecified, as long as the set of outcomes is a measurable set. This flexibility allows for a simplification of Eq. (2.30). Knowledge of $\tilde{\Lambda}$ fully determines the local response functions $q(a|x\tilde{\lambda})$, $q(b|y\tilde{\lambda})$. These response functions can always be simulated through the generation of a uniform random variable plus a deterministic function through what is known as *inverse transform sampling*. For example, consider sampling a discrete outcome a according to the distribution $q(a|x\tilde{\lambda})$. One may do this by first sampling a continuous uniform random variable M_A with values between zero and one, $M = \mu_A \in [0, 1]$. Then one needs to determine k such that $\sum_{a=0}^{a=k-1} q(a|x\tilde{\lambda}) \leq \mu_A \leq \sum_{a=0}^{a=k} q(a|x\tilde{\lambda})$ and return $a = k$. Averaged over μ_A, this method will return outcome $a = k$ with probability $q(k|x\tilde{\lambda})$. Inverse transform sampling therefore provides a systematic way in which to sample from a known probability distribution together with a uniform random variable.

Given that there is no limitation on the number of outcomes of $\tilde{\Lambda}$, we can define a larger joint variable $\tilde{\Lambda}' = (\tilde{\Lambda}, M_A, M_B)$. With this variable, both A and B can determine their outcome deterministically by using the values of the second and third component of $\tilde{\Lambda}'$. As such, Eq. (2.30) can be rewritten as a convex combination of local deterministic response functions:

$$p(ab|xy) = \int_{\tilde{\lambda}' \in N_{\tilde{\Lambda}'}} D(a|x\tilde{\lambda}')\, D(b|y\tilde{\lambda}')\, q(\tilde{\lambda}')\, \mathrm{d}\tilde{\lambda}, \tag{2.31}$$

where $D(a|x\tilde{\lambda}')$, $D(b|y\tilde{\lambda}')$ are conditional probability distributions where a single outcome occurs with unit probability. Mathematically, they may be described using the Kronecker delta function as follows:

$$D(a|x\tilde{\lambda}') = \delta_{f_{\tilde{\lambda}'}(x),a} \quad D(b|y\tilde{\lambda}') = \delta_{g_{\tilde{\lambda}'}(y),b},$$

where f maps each value of the input x one of the possible outcomes a, $f : N_X \to N_A$, and $\{f_{\tilde{\lambda}'}\}_{\tilde{\lambda}'}$ represents a collection of such mappings, each indexed by $\tilde{\lambda}'$. There are exactly $|N_X|^{|N_A|}$ unique assignments, which can be indexed through a superscript as $\{f^i\}_{i=1}^{i=|N_X|^{|N_A|}}$. Similar logic applies to g and the deterministic response of process B.

We may now group the set of values of $\tilde{\Lambda}'$ into subsets where for every value in the set, the deterministic assignment is the same. That is:

$$N_{\tilde{\Lambda}'} = \bigcup_{ij} N^{ij}_{\tilde{\Lambda}'} \quad \text{where} \quad N^{ij}_{\tilde{\Lambda}'} = \{\tilde{\lambda}' | f_{\tilde{\lambda}'} = f^i \text{ and } g_{\tilde{\lambda}'} = g^j\}. \tag{2.32}$$

Now Eq. (2.31) simplifies as follows:

$$p(ab|xy) = \sum_{i=1}^{|N_X|^{|N_A|}} \sum_{j=1}^{|N_Y|^{|N_B|}} \int_{\tilde{\lambda}' \in N^{ij}_{\tilde{\Lambda}'}} D(a|x\tilde{\lambda}')\, D(b|y\tilde{\lambda}')\, q(\tilde{\lambda}')\, \mathrm{d}\tilde{\lambda} \tag{2.33}$$

$$= \sum_{i=1}^{|N_X|^{|N_A|}} \sum_{j=1}^{|N_Y|^{|N_B|}} D(a|xi)\, D(b|yj) \underbrace{\int_{\tilde{\lambda}' \in N^{ij}_{\tilde{\Lambda}'}} q(\tilde{\lambda}')\, \mathrm{d}\tilde{\lambda}}_{:=q_{ij}} \tag{2.34}$$

$$= \sum_{i=1}^{|N_X|^{|N_A|}} \sum_{j=1}^{|N_Y|^{|N_B|}} D(a|xi)\, D(b|yj)\, q_{ij}, \tag{2.35}$$

where $D(a|xi) = \delta_{f^i(x),a}$, $D(b|yj) = \delta_{g^j(y),b}$ and q_{ij} represents a measure of the set of values of $\tilde{\Lambda}'$ which assign the deterministic response functions described by f^i and g^j.

One important consequence of this derivation is that we have shown that a finite cardinality $\tilde{\Lambda}$ is sufficient to reproduce all correlations achievable with a common cause of continuous cardinality, as we may simply set the values of $\tilde{\Lambda}$ to be (i, j) in (2.33) which occur with probability q_{ij}.

Unpacking. The unboundedness of the internal variables may be used in yet a different approach called "unpacking", first introduced in Ref. [NW20b]. Consider the preparations to contain as many copies of the prepared state as different possible values of the inputs:

$$\tilde{\Sigma}^{\tilde{\Lambda}}_A \to \left((\tilde{\Sigma}^{\tilde{\Lambda}}_A)_1, \ldots, (\tilde{\Sigma}^{\tilde{\Lambda}}_A)_{|N_X|}\right), \tag{2.36}$$

$$\tilde{\Sigma}^{\tilde{\Lambda}}_B \to \left((\tilde{\Sigma}^{\tilde{\Lambda}}_B)_1, \ldots, (\tilde{\Sigma}^{\tilde{\Lambda}}_B)_{|N_Y|}\right). \tag{2.37}$$

We may imagine A to contain N_X copies of the original process $A \to \{A_1, \ldots A_{|N_X|}\}$, which are applied in parallel to the N_X preparations. This process will have N_X outcomes with a local response functions $q(a_i|\tilde{\lambda})$. When

reporting an outcome given an input value $X = x$, A simply reports the outcome of the x-th process applied to the x-th preparation, and similarly for B. This corresponds to a marginalization over the joint distribution averaged over $\tilde{\Lambda}$:

$$p(ab|xy) = \sum_{\{a_i\}_i \setminus a_x} \sum_{\{b_j\}_j \setminus b_y} \int_{\tilde{\lambda}} \prod_{\alpha=1}^{|N_X|} q(a_\alpha|\tilde{\lambda}) \prod_{\beta=1}^{|N_X|} q(b_\beta|\tilde{\lambda})\, q(\tilde{\lambda})\, \mathrm{d}\tilde{\lambda} \tag{2.38}$$

$$= \sum_{\{a_i\}_i \setminus a_x} \sum_{\{b_j\}_j \setminus b_y} q(a_1, \ldots, a_x, \ldots, a_{|N_X|}, b_1, \ldots, b_y, \ldots, b_{|N_Y|}) \tag{2.39}$$

$$= \sum_{\bar{a} \setminus a_x} \sum_{\bar{b} \setminus b_y} q(\bar{a}, \bar{b}) \tag{2.40}$$

where we use bar variables:

$$\bar{a} = (a_1, \ldots, a_x, \ldots, a_{|N_X|}), \quad \bar{b} = (b_1, \ldots, b_y, \ldots, b_{|N_Y|}),$$

to simplify the expressions.[3] In the step (2.38)→(2.39) we used that in the absence of inputs, a classical common cause explanation spans the whole space of probability distributions:

$$\forall p(a_1, \ldots, a_n)\, \exists\, q(\lambda), q(a_j|\lambda) \text{ s.t. } p(a_1, \ldots, a_n) = \int_\lambda \prod_i q(a_i|\lambda)\, q(\lambda)\, \mathrm{d}\lambda. \tag{2.41}$$

This is true as any probability distribution $p(a_1, \ldots, a_n)$ can be trivially written as a convex combination of deterministic probability distributions. One may choose $\prod_i q(a_i|\lambda)$ to enumerate the different deterministic distributions for different values of λ, and fix $q(\lambda)$ such that $p(a_1, \ldots, a_n) = \int_{\lambda \in D_{(a_1,\ldots,a_n)}} q(\lambda)\, \mathrm{d}\lambda$, where $D_{(a_1,\ldots,a_n)}$ is the set of λ that specify $\prod_i q(a_i|\lambda)$ to be deterministic with outcome $(a_1, \ldots, a_n)$.

Through unpacking, therefore, testing for compatibility of $p(ab|xy)$ with a classical common-cause model simplifies to finding a joint distribution $q(\bar{a}, \bar{b})$

[3] We make a deliberate choice to slightly abuse notation in order to simplify the mathematical expressions, though this should not lead to confusion. For example, there is an inconsistency when we combine tuples with a set substraction symbol, e.g., $\bar{a} \setminus a_x = \{a_i\}_i \setminus a_x$. Furthermore, we equate a function that takes two tuples as arguments to a function applied to the combined elements of these tuples, e.g., $q(\bar{a}, \bar{b}) = q(a_1, \ldots, a_{|N_X|}, b_1, \ldots, b_{|N_Y|})$.

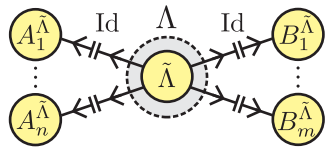


FIGURE 2.3: "Unpacked" processes A and B, where copies of the original process of Figure 2.2b are applied to n and m preparations of their respective states. When $n = |N_X|$ and $m = |M_Y|$, one can simulate given statistics $p(ab|xy)$ by returning the outcome of the measurement of the x-th and y-th preparations.

which marginalizes to the observed distribution per Eq. (2.40). Unpacking is shown in Figure 2.3, and it will be relevant later on during the thesis. The equivalence with a model in terms of local deterministic response functions is formalized through Fine's Theorem [Fin82].

Set of classical common-cause distributions. Now we will consider only the local deterministic description of Eq. (2.35), but one may equivalently work with unpacking as in Eq. (2.40). To determine whether some joint probability $p(ab|xy)$ is compatible with a common cause model where the process Λ is classical, one needs to find values q_{ij} such that $q_{ij} \geq 0$ and $\sum_{ij} q_{ij} = 1$ and which satisfy Eq. (2.35):

$$\begin{array}{ll} \text{find} & q_{ij} \\ \text{s.t.} & p(ab|xy) = \sum_i D(a|xi)\, D(b|yj)\, q_{ij} \\ & \sum q_{ij} = 1 \\ & q_{ij} \geq 0 \end{array} \tag{2.42}$$

This is a **linear programming** (LP) problem. If a solution does not exist to Eq. (2.42), then the joint probability $p(ab|xy)$ is not compatible with a classical common cause process Λ. As detailed in Appendix A, Farkas' lemma ensures that finding a set of values $\{\eta_0\} \cup \{\eta_{abxy}\}_{abxy}$ such that:

$$\sum_{abxy} \eta_{abxy} p(ab|xy) + \eta_0 > 0, \tag{2.43}$$

$$\sum_{abxy} \eta_{abxy} D(a|xi) D(b|yj) + \eta_0 \leq 0, \quad \forall\ i, j \tag{2.44}$$

proves that the program of Eq. (2.42) is infeasible.[4] By averaging expression (2.44) over q_{ij} we arrive at what is known as a **Bell inequality**:

$$\mathcal{I}(p(ab|xy)) \coloneqq \sum_{abxy} \eta_{abxy} p(ab|xy) + \eta_0 \leq 0, \tag{2.45}$$

which holds for $p(ab|xy) = \sum_{ij} D(a|xi)D(b|yj)q_{ij}$. That is, a Bell inequality is an algebraic bound on an affine function of the joint distribution $p(ab|xy)$ that holds for distributions compatible with a classical common cause explanation.

We define the **local set**[5] as the set of all such probabilities:

$$\mathcal{L} = \left\{ p(ab|xy) \,\middle|\, \exists q_{ij} \text{ s.t. } p(ab|xy) = \sum_{ij} D(a|xi)D(b|yj)q_{ij}, \right. \\ \left. \sum_{ij} q_{ij} = 1,\, q_{ij} \geq 0 \,\forall i,j \right\}. \tag{2.46}$$

A given correlation $p(ab|xy)$ is called **local** if it belongs to $\mathcal{L}$; otherwise it is said to be **nonlocal**.

The structure of $\mathcal{L}$ is that of a **polytope** living in the vector space of conditional joint probabilities, $\mathbb{R}^{|N_A||N_B||N_X||N_Y|}$, with vertices $V_{ij} = D(a|xi)D(b|yj)$. A Bell inequality can be interpreted geometrically as a hyperplane which either does not intersect with the local set, or is only tangent to it. Therefore, any nonlocal distribution $p(ab|xy)$ can be certified to be such by specifying a hyperplane which separates it from $\mathcal{L}$, i.e., a Bell inequality which is violated by $p(ab|xy)$. The **local bound** of a Bell inequality is the optimal value of the affine function $\mathcal{I}(p(ab|xy))$ over distributions

[4] Note that we flip the sign in Farkas' lemma with respect to the proof of Appendix A for convenience.

[5] The terminology comes historically comes from compatibility or incompatibility with the principle of local causality.

in $\mathcal{L}$. Such bound may be computed using a modification of LP (2.42):

$$\max_{p(ab|xy)\in\mathcal{L}} \mathcal{I}(p(ab|xy)) = \begin{array}{ll} \max & \mathcal{I}(p(ab|xy)) \\ \text{s.t.} & p(ab|xy) = \sum_i D(a|xi)\, D(b|yj)\, q_{ij} \\ & \sum q_{ij} = 1 \\ & q_{ij} \geq 0 \end{array} . \tag{2.47}$$

A distribution $p(ab|xy)$ is said to **violate a Bell inequality** if the value of the affine function $\mathcal{I}(p(ab|xy))$ is greater than the local bound. Note that talking about minimums or maximums of Bell inequalities is arbitrary, as this convention can be changed through a sign flip.

2.3.2 Quantum common cause

If the process Λ is a quantum process, then it corresponds to a generalised quantum instrument. However, since the input of Λ is trivial, this is just the preparation of an arbitrary quantum state. Therefore, the joint probabilities achievable when Λ is a quantum process are given by the Born rule:

$$p(ab|xy) = \mathrm{tr}\left[\rho_\Lambda \left(A_{a|x} \otimes B_{b|y}\right)\right], \tag{2.48}$$

where $\{A_{a|x}\}_a$ and $\{B_{b|y}\}_b$ define on Hilbert spaces $\mathcal{H}_A$ and $\mathcal{H}_B$, for all the values of x and y respectively, and ρ_Λ is a mixed quantum state on $\mathcal{H}_A \otimes \mathcal{H}_B$. Notably, we make no assumption on the dimensions $\mathcal{H}_A$, $\mathcal{H}_B$.

Quantum set of correlations. We define the *quantum set of correlations*, $\mathcal{Q}$, as the set of all probability distributions achievable through expression (2.48):

$$\mathcal{Q} = \left\{ p(ab|xy) \,\middle|\, p(ab|xy) = \mathrm{tr}\left[\rho_\Lambda \left(A_{a|x} \otimes B_{b|y}\right)\right] \right\}, \tag{2.49}$$

where ρ_Λ, $A_{a|x}$, $B_{b|y}$ have the same meaning as in Eq. (2.48).

The set $\mathcal{Q}$ is convex, as for any two distributions $p_1(ab|xy)$, $p_2(ab|xy)$ built from states ρ_1, ρ_2 and measurements $\{A^1_{a|x}\}_{xa}$, $\{B^1_{b|y}\}_{yb}$, $\{A^2_{a|x}\}_{xa}$, $\{B^2_{b|y}\}_{yb}$, the distribution $t\, p_1(ab|xy) + (1-t)\, p_2(ab|xy)$ belongs to $\mathcal{Q}$ for any $t \in [0,1]$. This can be proven by enlarging the Hilbert space to include two flags F_1,

F_2 which allows A and B to synchronize their strategies to be either the one that generates $p_1(ab|xy)$ with frequency t or the one that generates $p_2(ab|xy)$ with frequency $(1-t)$. One can simply check that through the following states and measurements:

$$\rho = t|00\rangle\langle 00|_{F_1F_2} \otimes \rho_1 + (1-t)|11\rangle\langle 11|_{F_1F_2} \otimes \rho_2, \tag{2.50}$$

$$A_{a|x} = |0\rangle\langle 0|_{F_1} \otimes A^1_{a|x} + |1\rangle\langle 1|_{F_1} \otimes A^2_{a|x}, \tag{2.51}$$

$$B_{b|y} = |0\rangle\langle 0|_{F_2} \otimes B^1_{b|y} + |1\rangle\langle 1|_{F_2} \otimes B^2_{b|y}, \tag{2.52}$$

and Born's rule, one generates the distribution $t\,p_1(ab|xy) + (1-t)\,p_2(ab|xy)$.

Bell's Theorem. Bell's Theorem states that the local set is strictly included in the quantum set:

$$\mathcal{L} \subsetneq \mathcal{Q}.$$

This can be proven by specifying a point in $\mathcal{Q}$ which is not in $\mathcal{L}$. One such point can be built using the following states and measurements with two outcomes:

$$\rho_\Lambda = |\Psi^+\rangle\langle\Psi^+| \qquad \text{where} \qquad |\Psi^+\rangle = \frac{|00\rangle + |11\rangle}{\sqrt{2}} \tag{2.53}$$

$$A_{0|0} = |+\rangle\langle +| \qquad \text{where} \qquad |+\rangle = \frac{|0\rangle + |1\rangle}{\sqrt{2}} \tag{2.54}$$

$$A_{0|1} = |0\rangle\langle 0| \tag{2.55}$$

$$B_{0|0} = |v\rangle\langle v| \qquad \text{where} \qquad |v\rangle = \frac{(1+\sqrt{2})|0\rangle + |1\rangle}{\sqrt{(1+\sqrt{2})^2+1}} \tag{2.56}$$

$$B_{0|} = |v'\rangle\langle v'| \qquad \text{where} \qquad |v'\rangle = \frac{|0\rangle + (1+\sqrt{2})|1\rangle}{\sqrt{(1+\sqrt{2})^2+1}} \tag{2.57}$$

The other measurement operators can be deduced from the completeness relation. The measurements correspond to PVMs whose effects, for each outcome, project onto the eigenvectors of the operators σ_x, σ_z, $\frac{1}{\sqrt{2}}(-\sigma_x - \sigma_z)$, $\frac{1}{\sqrt{2}}(-\sigma_x + \sigma_z)$, respectively, where $\{\sigma_x, \sigma_y, \sigma_z\}$ are the Pauli matrices:

$$\sigma_x = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \quad \sigma_y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} \quad \sigma_z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \tag{2.58}$$

The joint probability $p(ab|xy)$ constructed from the above states and measurement is nonlocal, and its nonlocality can be certified through the following Bell inequality:

$$p(00|00)+p(00|01)+p(00|10)-p(00|11)-p_A(0|0)-p_B(0|0)+1\leq 0 \quad (2.59)$$

where

$$p_A(a|x) \coloneqq \sum_b p(ab|xy) \quad p_B(b|y) \coloneqq \sum_a p(ab|xy).$$

Note that due to the axiom of causality, $p_A(a|x)$ does not functionally depend on the values of Y and neither $p_B(b|y)$ on the values of X. By using correlators, defined as follows:

$$\langle A_x\rangle \coloneqq p_A(0|x)-p_A(1|x) \quad (2.60)$$

$$\langle B_y\rangle \coloneqq p_B(0|y)-p_B(1|y) \quad (2.61)$$

$$\langle A_xB_y\rangle \coloneqq \sum_{ab}(-1)^{a+b}p(ab|xy), \quad (2.62)$$

the inequality (2.59) takes the form:

$$\langle A_0B_0\rangle+\langle A_0B_1\rangle+\langle A_1B_0\rangle-\langle A_1B_1\rangle\leq 2. \quad (2.63)$$

This is the standard form of the Clauser-Horne-Shimony-Holt (CHSH) inequality [Cla+69]. Evaluated with the strategy of (2.53)—(2.57), it achieves a value of $2\sqrt{2}$, which is also the maximum value it can achieve over distributions in $\mathcal{Q}$. This is known as the **Tsirelson bound** or **quantum bound**.

Given an arbitrary distribution $p(ab|xy)$, is it possible to determine whether it belongs to the quantum set, $\mathcal{Q}$, analogously to the linear program method (2.42) for testing membership in the local set? To determine whether $p(ab|xy)$ is a member of the quantum set one needs to find states and measurements that reproduce the distribution. This is very difficult as the dimension of the Hilbert space is unbounded. One strategy consists of relaxing the problem, and looking for necessary but not sufficient conditions for $p(ab|xy)$ to belong to the quantum set of correlations. One such relaxation is the hierarchy of Navascués-Pironio-Acín (NPA) based on moment matrices [NPA08], which we will introduce following the presentation of Ref. [Mor+13b].

If $p(ab|xy)$ belongs to $\mathcal{Q}$, then there exists a state ρ_Λ and measurement operators $\{A_{a|x}\}_{xa}$, $\{B_{b|y}\}_{yb}$ which simulate $p(ab|xy)$ using Born's rule. For any finite dimensional Hilbert space, one may define "locality" not through a tensor product structure, but through commutativity. As such, Born's rule becomes:

$$p(ab|xy) = \operatorname{tr} \rho_\Lambda A_{a|x} B_{b|y} \quad \text{s.t.} \quad [A_{a|x}, B_{b|y}] = 0 \ \ \forall x, a, y, b. \tag{2.64}$$

where $[A, B] \coloneqq AB - BA$. This is known as the commuting-operator formalism [Slo20]. Let $\tilde{\mathcal{Q}}$ be the set of correlations achievable through expressions of the form (2.64). Whether $\mathcal{Q} = \tilde{\mathcal{Q}}$ is known as Tsirelson's problem [Tsi93] which Tsirelson himself solved for finite dimension quantum systems [Tsi06], proving that they are indeed equal. For infinite dimensional systems, it is now believed that the tensor-product and commuting-operator frameworks are inequivalent [Ji+22; CQK23]. Given that in this thesis we will work only with finite-dimensional quantum systems, we will use the commuting-operator framework in the remainder of this section (i.e., the support of $A_{a|x}$ and $B_{b|y}$, which are taken to commute, is the joint Hilbert space $\mathcal{H}_A \otimes \mathcal{H}_B$).

Similarly to how the unboundedness of the shared randomness in the classical common cause model can be used to make the local response functions as certain as possible, i.e., deterministic (Eq. (2.35)), we can use the unboundedness of the Hilbert space dimension to remove as much uncertainty from the states and measurements as possible, that is, purify both the state and the measurements. This is done by enlarging the total Hilbert space with an auxiliary space used for purification: $\mathcal{H}_A \otimes \mathcal{H}_B \to \mathcal{H}_A \otimes \mathcal{H}_B \otimes \mathcal{H}_{\text{aux}}$. We will henceforth assume that we are working with pure states and PVMs, i.e., we write the probability as $p(ab|xy) = \langle\psi|A_{a|x}B_{b|y}|\psi\rangle$ where $\{A_{a|x}\}_a$, $\{B_{b|y}\}_b$ are PVMs for all x, y.

NPA hierarchy. Let $\mathcal{O} = \{\mathbb{1}_{AB}\} \cup \{A_{a|x}\}_{xa} \cup \{B_{b|y}\}_{yb}$ the set of all the measurements operators, together with the identity operator, which we will call an alphabet. We many now define a set $\mathcal{S}$ of monomials built from the alphabet $\mathcal{O}$. We can use a set $\mathcal{S}$ to build a completely positive map, $\Gamma^{\mathcal{S}}$, with Kraus operators $M_i = \sum_s |s\rangle\langle i|\mathcal{S}_s$. Then:

$$\Gamma(\rho_\Lambda) = \sum_i M_i \rho_\Lambda M_i^\dagger = \sum_{st} |s\rangle\langle t| \operatorname{tr}\left[\rho_\Lambda \mathcal{S}_s^\dagger \mathcal{S}_t\right]. \tag{2.65}$$

We shall refer to expectation values of monomials as *moments*, and simplify notation using angled brackets: $\langle M \rangle \coloneqq \operatorname{tr} \rho_\Lambda M$. If ρ_Λ is a positive semidefinite operator, then $\Gamma^{\mathcal{S}}(\rho_\Lambda)$ is also positive semidefinite:

$$\rho_\Lambda \succeq 0 \Rightarrow \Gamma^{\mathcal{S}}(\rho_\Lambda) \succeq 0.$$

While $\Gamma^{\mathcal{S}}(\rho_\Lambda)$ itself need not be a quantum state, it can always be renormalized such that its trace is equal to one. As such, one may think of $\Gamma^{\mathcal{S}}$ as a quantum channel that could be in principle implemented in a laboratory. The $(0,0)$ element of $\Gamma^{\mathcal{S}}(\rho_\Lambda)$ encodes the normalization of the state:

$$\langle 0|\Gamma^{\mathcal{S}}(\rho_\Lambda)|0\rangle = \operatorname{tr} \rho_\Lambda.$$

Generally, most entries (s,t) of the moment matrix can only be known if we have complete knowledge of the states and measurements, but some depend only on the joint probability $p(ab|xy)$, such as:

$$\langle 0|\Gamma^{\mathcal{S}}(\rho_\Lambda)|1\rangle = \operatorname{tr} \rho_\Lambda A_{0|0} = p_A(0|0).$$

Such entries can be assigned a known numerical value. In general, the moment matrix can be decomposed into a sum over entries which depend on the probability and ones which do not:

$$\Gamma^{\mathcal{S}}(\rho_\Lambda) =: \Gamma = \Gamma_0 + \sum_i x_i \Gamma_i, \tag{2.66}$$

where Γ_0 is a constant matrix whose value may depend on $p(ab|xy)$, and Γ_i are matrices that encode the coefficients of the moment x_i in the moment matrix.

Let us consider a simple example. If $\mathcal{S} = \{\mathbb{1}, A_{0|1}, B_{0|0}, A_{0|0}A_{0|1}\}$, then the moment matrix is given by:

$$\Gamma = \begin{array}{c} \\ \mathbb{1} \\ A_{0|1} \\ B_{0|0} \\ A_{0|0}A_{0|1} \end{array} \begin{array}{c} \begin{array}{cccc} \mathbb{1} & A_{0|1} & B_{0|0} & A_{0|0}A_{0|1} \end{array} \\ \begin{pmatrix} 1 & p_A(0|1) & p_B(0|0) & \langle A_{0|0}A_{0|1}\rangle \\ & p_A(0|1) & p(00|10) & \langle A_{0|1}A_{0|0}A_{0|1}\rangle \\ & & p_B(0|0) & \langle B_{0|0}A_{0|0}A_{0|1}\rangle \\ & & & \langle A_{0|1}A_{0|0}A_{0|1}\rangle \end{pmatrix} \end{array}. \tag{2.67}$$

As the moment matrix is hermitian—as it can be checked from Eq. (2.65)—we

only show the upper triangular elements. Note that we used commutativity $[A_{a|x}, B_{b|y}] = 0$, orthogonality and that the measurement operators are projectors to algebraically simplify the monomials. Noticeably, some entries of the moment matrix are fully determined by the probability while others, such as $\langle A_{0|0}A_{0|1}\rangle$ or $\langle B_{0|0}A_{0|0}A_{0|1}\rangle$, are expectation values that can only be determined with full knowledge of the states and measurements. However, if $p(ab|xy) \in \mathcal{Q}$, then we know that there must exist states and measurements that give rise to $p(ab|xy)$, and that therefore there must exist values for expectation values, e.g., $\langle A_{0|0}A_{0|1}\rangle$ or $\langle B_{0|0}A_{0|0}A_{0|1}\rangle$, such that $\Gamma \succeq 0$. This defines a *semidefinite program* (SDP):

$$\begin{array}{ll} \text{find} & x_1, x_2, x_3 \in \mathbb{C} \\ \text{s.t.} & \Gamma = \Gamma_0 + \sum_i x_i \Gamma_i \succeq 0 \end{array} \tag{2.68}$$

where

$$\Gamma_0 = \begin{pmatrix} 1 & p_A(0|1) & p_B(0|0) & 0 \\ & p_A(0|1) & p(00|10) & 0 \\ & & p_B(0|0) & 0 \\ & & & 0 \end{pmatrix} \qquad \Gamma_1 = \begin{pmatrix} 0 & 0 & 0 & 1 \\ & 0 & 0 & 0 \\ & & 0 & 0 \\ & & & 0 \end{pmatrix} \tag{2.69}$$

$$\Gamma_2 = \begin{pmatrix} 0 & 0 & 0 & 0 \\ & 0 & 0 & 1 \\ & & 0 & 0 \\ & & & 1 \end{pmatrix} \qquad \Gamma_3 = \begin{pmatrix} 0 & 0 & 0 & 0 \\ & 0 & 0 & 0 \\ & & 0 & 1 \\ & & & 0 \end{pmatrix} \tag{2.70}$$

If the program (2.68) is infeasible, then we reached a contradiction and therefore, $p(ab|xy) \notin \mathcal{Q}$. If however, the program (2.68) is feasible, this test is inconclusive, as the semidefinite positivity of Γ is a necessary but not sufficient condition for $p(ab|xy)$ to belong to $\mathcal{Q}$. In general, the solution to SDP (2.68) is complex, however, for any solution Γ to the SDP, we also have, from convexity, that the real part of the solution matrix, $\mathrm{Re}(\Gamma) = \frac{\Gamma+\Gamma^*}{2}$, is also a solution. Therefore, without loss of generality we may consider only the SDP over real numbers:

$$\begin{array}{ll} \text{find} & x_1, x_2, x_3 \in \mathbb{R} \\ \text{s.t.} & \Gamma = \Gamma_0 + \sum_i x_i \Gamma_i \succeq 0 \end{array} \tag{2.71}$$

as infeasibility of SDP (2.71) implies infeasibility of SDP (2.68). If a distribution $p(ab|xy)$ leads to an infeasible SDP (2.71), as with linear programs,

we can use Farkas' lemma (cf. Appendix A) to extract a Bell-type inequality which is valid for all $p(ab|xy) \in \mathcal{Q}^{\mathcal{S}}$.

Instead of proving at the level of the SDP that real matrices are sufficient, one may also use flags, such as in (2.50)—(2.52), to coordinate A and B to both use either the quantum state and measurements ρ_Λ, $\{A_{a|x}\}_{xa}$, $\{B_{b|y}\}_{yb}$ or their complex conjugate, ρ_Λ^*, $\{A_{a|x}^*\}_{xa}$, $\{B_{b|y}^*\}_{yb}$, which when averaging over the flag with probability $\frac{1}{2}$ will be operationally indistinguishable from using real states and measurement operators [Ren+21]:

$$\rho_\Lambda \to \frac{|{+i},{+i}\rangle\langle{+i},{+i}|_{F_1F_2} \otimes \rho_\Lambda + |{-i},{-i}\rangle\langle{-i},{-i}|_{F_1F_2} \otimes \rho_\Lambda^*}{2}, \tag{2.72}$$

$$A_{a|x} \to |{+i}\rangle\langle{+i}|_{F_1} \otimes A_{a|x} + |{-i}\rangle\langle{-i}|_{F_1} \otimes A_{a|x}^*, \tag{2.73}$$

$$B_{b|y} \to |{+i}\rangle\langle{+i}|_{F_2} \otimes B_{b|y} + |{-i}\rangle\langle{-i}|_{F_2} \otimes B_{b|y}^*, \tag{2.74}$$

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

Let $\mathcal{Q}^{\mathcal{S}}$ be the set of all distributions $p(ab|xy)$ which are compatible with the moment matrix relaxation based on the set of monomials $\mathcal{S}$. The NPA hierarchy consists of a recipe for building sets $\mathcal{S}_n$ such that the sets of correlations $\mathcal{Q}^{\mathcal{S}_n}$ provide tighter approximations to the quantum set as n increases while ultimately converging to it, $\mathcal{Q}^{\mathcal{S}_1} \supseteq \mathcal{Q}^{\mathcal{S}_2} \supseteq \cdots \supseteq \lim_{n\to\infty} \mathcal{Q}^{\mathcal{S}_n} = \mathcal{Q}$.[6] These sets are the n-fold products of elements of $\mathcal{O}$: $\mathcal{S}_n = \{\mathcal{O}_{i_1}\mathcal{O}_{i_2}\cdots\mathcal{O}_{i_n}\}_{i_1,i_2,\ldots,i_n}$. One may algebraically simplify the set by removing elements of $\mathcal{S}_n$ which are equivalent under substitutions derived from commutations and orthogonality of operators for different outcomes. One may also use the completeness relation of POVMs to exclude operators corresponding to the last outcome from the alphabet $\mathcal{O}$.

Given that $\mathcal{Q} \subseteq \mathcal{Q}^{\mathcal{S}_n}$, the NPA hierarchy may be used to compute upper bounds on the Tsirelson bound of a Bell inequality, $\mathcal{I}(p(ab|xy))$, as follows:

$$\begin{aligned} \max_{p(ab|xy)\in\mathcal{Q}} \mathcal{I}(p(ab|xy)) &\leq \max_{p(ab|xy)\in\mathcal{Q}^{\mathcal{S}_n}} \mathcal{I}(p(ab|xy)) = \\ &= \begin{array}{ll} \max & \mathcal{I}(p(ab|xy)) \\ \text{s.t.} & \Gamma = \Gamma_0 + \sum_{abxy} p(ab|xy)\Gamma_{abxy} + \sum_i x_i\Gamma_i \succeq 0 \\ & x_i \in \mathbb{R}, p(ab|xy) \in \mathbb{R} \end{array} \end{aligned} \tag{2.75}$$

[6] Recall that strictly speaking it converges to the quantum set of correlations in the commuting-operator framework.

where $p(ab|xy)$ are now variables in the SDP.

DI certification of entanglement. The NPA hierarchy is "device independent", as no assumptions are made about the states and measurements, except that they follow the rules of quantum theory.[7] However, we may consider semi-device independent scenarios where more assumptions are made about the inner workings of the (quantum) process Λ and the processes A and B. For example, we may consider the set of quantum correlations arising from measuring separable states:

$$\mathcal{Q}_{\text{SEP}} = \{p(ab|xy) \,|\, p(ab|xy) = \operatorname{tr}\left[\rho_\Lambda \cdot \left(A_{a|x} \otimes B_{b|y}\right)\right], \\ \rho_\Lambda = \int_\lambda q(\lambda)\rho_A^\lambda \otimes \rho_B^\lambda \,\mathrm{d}\lambda\}. \quad (2.76)$$

It is easy to see that $\mathcal{Q}_{\text{SEP}} = \mathcal{L}$:

$$\begin{aligned} p(ab|xy) &= \operatorname{tr}\left[\rho_\Lambda \cdot \left(A_{a|x} \otimes B_{b|y}\right)\right] && (2.77)\\ &= \operatorname{tr}\left[\left(\int_\lambda q(\lambda)\rho_A^\lambda \otimes \rho_B^\lambda \,\mathrm{d}\lambda\right) \cdot \left(A_{a|x} \otimes B_{b|y}\right)\right] && (2.78)\\ &= \int_\lambda \underbrace{(\operatorname{tr}\rho_A^\lambda A_{a|x})}_{=:q(a|x\lambda)} \underbrace{(\operatorname{tr}\rho_B^\lambda B_{b|y})}_{=:q(b|y\lambda)} \,\mathrm{d}\lambda && (2.79)\\ &= \int_\lambda q(a|x\lambda)q(b|y\lambda)\,\mathrm{d}\lambda. && (2.80) \end{aligned}$$

Since $\mathcal{Q}_{\text{SEP}} = \mathcal{L}$, any instance of nonlocality is immediately a *device-independent certification of entanglement.*

The set $\mathcal{Q}_{\text{SEP}}$ is fully characterizable through linear programming techniques. While linear programs can be solved in time that is polynomial in the number of variables, the number of vertices of the local polytope scales exponentially with the number of inputs, $|N_A|^{|N_X|}|N_B|^{|N_Y|}$. As such, it is of interest to develop new techniques that may be solved more efficiently, either when increasing the cardinalities of the inputs or outputs, or when considering a common cause between more than two processes (known as "multipartite scenarios").

[7] We use quotation marks—"device independent"—as full device independence is also theory independent, whilst here we assume quantum theory to be true.

All-commuting NPA hierarchy. One approach is to discover necessary but not sufficient conditions for separability that may be more scalable. If a state is separable, then it can be prepared locally with the aid of shared randomness, i.e., it is the output of a classical common cause process. The output of a classical process can be cloned arbitrarily, so each process can locally prepare as many copies of the state as needed (this the same as unpacking, as previously defined, and related to notions such as symmetric extensions of quantum states [DPS02; DPS04]). In particular, A may prepare as many copies as the cardinality of the input X, N_X, and measure the x-th local copy when $X = x$. Since the measurement operators $\{A_{a|x}\}_{xa}$ act on disjoint subspaces for different x, they commute. As such, separability of the quantum state implies that for any general measurement $\{A_{a|x}\}_{xa}$, there exists a protocol where the measurements commute for different pairs of values of x while reproducing the same expectation values. A similar protocol can be implemented in process B. One may then build a hierarchy of moment matrix relaxations of $\mathcal{Q}_{\text{SEP}}$ where all operators commute with each other, and which converges to $\mathcal{Q}_{\text{SEP}}$ at a finite level n of the hierarchy. This relaxation is used in Ref. [Bac+17] for device-independent certification of entanglement, and in this thesis in Chapter 6.

PPT-NPA. A well-known necessary but not sufficient criteria for separability is that of positivity under the partial transpose (PPT):

$$\rho \text{ separable} \Rightarrow \rho^{T_A} \succeq 0, \tag{2.81}$$

where $(\cdot)^{T_A}$ denotes that partial transpose, which acts on a basis of operators as $(|i_1 i_2\rangle\langle j_1 j_2|)^{T_A} = |j_1 i_2\rangle\langle i_1 j_2|$. One may define the set of PPT-quantum correlations, which are joint distributions that arise from measuring a PPT state:

$$\mathcal{Q}_{\text{PPT}} = \{p(ab|xy) \,|\, p(ab|xy) = \operatorname{tr}\left[\rho_\Lambda \left(A_{a|x} \otimes B_{b|y}\right)\right], \rho_\Lambda^{T_A} \succeq 0\}. \tag{2.82}$$

One can easily check the following mathematical identity:

$$\operatorname{tr} M^{T_A} N = \sum M_{j_1 i_2, i_1, j_2} N_{j_1 j_2, i_1, i_2} = \overbrace{=}^{j_1 \leftrightarrow i_1} \sum M_{i_1 i_2, j_1, j_2} N_{i_1 j_2, j_1, i_2} =: \operatorname{tr} M N^{T_A}. \tag{2.83}$$

Consider an expectation value evaluated on the state ρ^{T_A}:

$$\begin{aligned}\operatorname{tr} \rho_\Lambda^{T_A} A_{i_1} A_{i_2} \cdots A_{i_n} \otimes B_{j_1} \cdots B_{j_m} = \\ = \operatorname{tr} \rho_\Lambda (A_{i_1} A_{i_2} \cdots A_{i_n})^T \otimes B_{j_1} \cdots B_{j_m} = \\ = \operatorname{tr} \rho_\Lambda A_{i_n}^* A_{i_{n-1}}^* \cdots A_{i_1}^* \otimes B_{j_1} \cdots B_{j_m}, \end{aligned} \tag{2.84}$$

where $i_1, i_2, \ldots, i_n$ and $j_1, j_2, \ldots, j_n$ represent tuples of values $(a|x)$ and $(b|y)$ respectively. Using flags, as in (2.72)—(2.74), one may take the state and measurements to be real, $A_{a|x}^* = A_{a|x}$. Furthermore, we can move from the tensor-product formalism to the commuting-operator formalism:

$$\langle A_{i_1} A_{i_2} \ldots A_{i_n} B_{j_1} \ldots B_{j_m} \rangle_{\rho_\Lambda^{T_A}} = \langle A_{i_n} A_{i_{n-1}} \ldots A_{i_1} B_{j_1} \ldots B_{j_m} \rangle_{\rho_\Lambda}. \tag{2.85}$$

Now, given a set of monomials $\mathcal{S}$ we may build two moment matrices

$$\Gamma := \Gamma^{\mathcal{S}}(\rho_\Lambda) \tag{2.86}$$

$$\tilde{\Gamma} := \Gamma^{\mathcal{S}}(\rho_\Lambda^{T_A}) \tag{2.87}$$

whose moments are connected by Eq. (2.85). Since $\Gamma^{\mathcal{S}}$ is a completely positive map, and $\rho_\Lambda^{T_A} \succeq 0$ by hypothesis, then $\Gamma^{\mathcal{S}}(\rho_\Lambda^{T_A}) \succeq 0$. This leads to a SDP which is feasible if the correlations $p(ab|xy)$ can be generated by measuring a state that is PPT:

$$\begin{array}{ll} \text{find} & x_i \in \mathbb{R}\ \forall i \\ \text{s.t.} & \Gamma = \Gamma_0 + \sum_i x_i \Gamma_i \succeq 0 \\ & \tilde{\Gamma} = \tilde{\Gamma}_0 + \sum_i x_i \tilde{\Gamma}_i \succeq 0 \end{array} \tag{2.88}$$

If SDP (2.88) is infeasible, then the underlying state ρ_Λ giving rise to $p(ab|xy)$ must be entangled. Some moments x_i are present both in Γ and in Γ_{PPT}, while others are not (for those, either $\Gamma_i = 0$ or $\tilde{\Gamma}_i = 0$). Ref. [Mor+13b] provides a construction of sets $\mathcal{S}_n^{\text{PPT}}$ such that the set of distributions compatible with the SDP relaxation (2.88) converges to the set of PPT correlations, $\mathcal{Q}_{\text{PPT}}^{\mathcal{S}_1} \supseteq \mathcal{Q}_{\text{PPT}}^{\mathcal{S}_2} \supseteq \cdots \supseteq \lim_{n\to\infty} \mathcal{Q}_{\text{PPT}}^{\mathcal{S}_n} = \mathcal{Q}_{\text{PPT}}$. We use this hierarchy in Chapter 5 for semi-device independent entanglement certification.

NPA as NCPOP. The NPA hierarchy can be seen as a particular example of a *non-commuting polynomial optimization problem* (NCPOP) [BKP+16;

NPA08]. A NCPOP consists of optimizing a polynomial of non-commuting variables subject to polynomial constraints.

A non-commuting polynomial is built from an alphabet $\mathcal{O}$ of variables which do not commute. We also equip our description with an involution such that $(X_1X_2\cdots X_n)^* = X_n^*X_{n_1}^*\cdots X_1^*$ where X_i are elements of $\mathcal{O}$. A non-commuting polynomial is a formal sum of monomials built from $\mathcal{O}$. For example, if $\mathcal{O} = \{1, X, Y, Z\}$, a non-commuting polynomial is an expression such as $f(\mathcal{O}) = XY + Y^2Z + 1$. The formal polynomial f may be evaluated on a tuple of operators, such as $\bar{O} = (\mathbb{1}, \sigma_x, \sigma_y, \sigma_z)$, by substituting the i-th element of $\mathcal{O}$ with the i-th operator in $\bar{O}$: $f(\bar{O}) = \sigma_x\sigma_y + \sigma_y^2\sigma_z + \mathbb{1}$. We may now impose polynomial constraints, such as for example $g(\mathcal{O}) = Y^2 - 1 \succeq 0$, which are understood as all evaluations of g on tuples of operators to be positive semidefinite, $g(\bar{O}) \succeq 0\ \forall \bar{O}$. Equality constraints, $g(\mathcal{O}) = 0$, may be expressed as two inequalities, $g(\mathcal{O}) \succeq 0, -g(\mathcal{O}) \succeq 0$. An optimization function is a linear function $L(f(\mathcal{O}))$ which takes a numerical value when $f(\mathcal{O})$ is evaluated on a tuple of operators $\bar{O}$. For example, we can consider an eigenvalue maximization problem:

$$\begin{array}{ll} \max_{|\psi\rangle,\bar{O}} & \langle\psi|f(\bar{O})|\psi\rangle \\ \text{s.t.} & g_\alpha(\bar{O}) \succeq 0 \quad \alpha = 1, \ldots, p \\ & \langle\psi|\psi\rangle = 1 \end{array} \tag{2.89}$$

which aims to determine the maximum possible eigenvalue of $f(\bar{O})$ subject to constraints $g_\alpha(\bar{O}) \succeq 0$.

We can express the problem of determining the quantum bound of a Bell inequality, $\max_{p(ab|xy)\in\mathcal{Q}} \mathcal{I}(p(ab|xy))$, as an eigenvalue optimization problem using the linearity of the trace in Born's rule:

$$\begin{array}{ll} \max_{|\psi\rangle,A_{a|x},B_{b|y}} & \langle\psi|\left(\eta_0\mathbb{1} + \sum_{abxy}\eta_{abxy}A_{a|x}B_{b|y}\right)|\psi\rangle \\ \text{s.t.} & [A_{a|x}, B_{b|y}] = 0, \\ & \sum_a A_{a|x} = \mathbb{1}, \quad \sum_b B_{b|y} = \mathbb{1}, \\ & A_{a|x}A_{xa'} = \delta_{aa'}A_{a|x}, \\ & B_{b|y}B_{yb'} = \delta_{bb'}B_{b|y}, \\ & A_{a|x} \succeq 0, \quad B_{b|y} \succeq 0, \\ & \langle\psi|\psi\rangle = 1. \end{array} \tag{2.90}$$

Membership of $p(ab|xy)$ in $\mathcal{Q}$ can be cast as a similar NCPOP by having a

trivial objective function in (2.89) together with the additional constraints $p(ab|xy) = \operatorname{tr} \rho A_{a|x} B_{b|y}$. Solving NCPOP (2.89) is difficult as we put no bound on the possible tuples of operators $\bar{O}$ we can consider. However, the problem can be relaxed by considering **moment matrix relaxations** of the problem (in what continues, we follow the presentation of Ref. [IR22]).

As with NPA, let $\mathcal{S}$ be a set of monomials in $\mathcal{O}$, including the identity 1, and let $\vec{\mathcal{S}}$ be the vector whose i-th element is the i-th element of $\mathcal{S}$.[8] We define a matrix of moments Γ:

$$\Gamma = \langle\psi|\vec{\mathcal{S}} \cdot \vec{\mathcal{S}}^\dagger|\psi\rangle, \tag{2.91}$$

where $(\cdot)^\dagger = ((\cdot)^*)^T$. It is easy to see that the matrix Γ is positive semidefinite:

$$\vec{v}^\dagger \Gamma \vec{v} = \langle\psi|\vec{v}^\dagger \vec{\mathcal{S}} \cdot \vec{\mathcal{S}}^\dagger \vec{v}|\psi\rangle = \langle\psi|PP^*|\psi\rangle \geq 0 \ \forall \vec{v} \in \mathbb{C}^{|\mathcal{S}|}, \tag{2.92}$$

where $P = \vec{v}^\dagger \vec{\mathcal{S}}$, as $\langle\psi|P(\bar{O})P(\bar{O})^*|\psi\rangle = \langle\psi'|\psi'\rangle \geq 0$ for any choice of operators $\bar{O}$, where $|\psi'\rangle \coloneqq P(\bar{O})^*|\psi\rangle$.

Let us consider a set $\mathcal{M} = \{\vec{\mathcal{S}}_i \vec{\mathcal{S}}_j^*\}_{ij}$. Then if $f(\mathcal{O})$ can be written as a linear combination of elements of $\mathcal{M}$, the objective function $\langle\psi|f(\mathcal{O})|\psi\rangle$ may be expressed as a linear combination of entries of Γ:

$$f(\mathcal{O}) = \sum_{ij} c_{ij} \mathcal{M}_{ij} \Rightarrow \langle\psi|f(\mathcal{O})|\psi\rangle = \sum_{ij} c_{ij} \Gamma_{ij}. \tag{2.93}$$

Furthermore, the constraints $g_\alpha(\mathcal{O}) \succeq 0$ can be also be expressed in terms of a **localizing matrix**, $\Gamma_{g_\alpha} = \langle\psi|(\vec{\mathcal{S}} g_\alpha) \cdot \vec{\mathcal{S}}^\dagger|\psi\rangle$, which is positive semidefinite on any choice of operators $\bar{O}$ since $g_\alpha(\bar{O}) \succeq 0$:

$$\vec{v}^\dagger \Gamma_{g_\alpha} \vec{v} = \langle\psi|\vec{v}^\dagger(\vec{\mathcal{S}} g_\alpha) \cdot \vec{\mathcal{S}}^\dagger \vec{v}|\psi\rangle = \langle\psi|P g_\alpha P^*|\psi\rangle \geq 0 \ \forall \vec{v} \in \mathbb{C}^{|\mathcal{S}|}. \tag{2.94}$$

Given a NCPOP (2.89), and a generating set $\mathcal{S}$, we define a SDP relaxation of it as follows:

$$\begin{array}{ll} \max & \sum_{ij} c_{ij} \Gamma_{ij} \\ \text{s.t.} & \Gamma = \Gamma_0 + \sum_i x_i \Gamma_i \succeq 0 \\ & \Gamma_{g_\alpha} = (\Gamma_{g_\alpha})_0 + \sum_i x_i (\tilde{\Gamma_{g_\alpha}})_i \succeq 0 \quad \alpha = 1, \ldots, p \\ & \Gamma_{00} = 1 \end{array} \tag{2.95}$$

[8] In full generality, $\mathcal{S}$ could be a set of polynomials in $\mathcal{O}$, however, a subset of a monomial basis of the space of non-commuting polynomials is sufficient.

where $(\Gamma_0)_{00} = 1$, which captures that $\langle\psi|\psi\rangle = 1$. Note that this constitutes a relaxation as instead of optimizing over sets of operators $|\psi\rangle, \bar{O}$ as in NCPOP (2.89), we are optimizing over values $x_i = \langle\mathcal{M}_i\rangle$ of expectation values of monomials. There does not necessarily need to exist $|\psi\rangle, \bar{O}$ that realize values the x_i found by the SDP (2.95). However, sometimes this is possible through what is known as the Gelfand-Naimark-Segal (GNS) construction, through a judicious choice of a set $\mathcal{S}$ or a limit of sets, $\lim_{n\to\infty} \mathcal{S}_n$ [NPA08].

It is noteworthy that in practice, when utilizing the same generating set $\mathcal{S}$ for both the moment matrix Γ and localizing matrices Γ_{g_α}, several moments x_i may appear in Γ_{g_α} but not in Γ. Therefore, for constructing localizing matrices, in practice it is preferred to use the largest set $\mathcal{S}'$ that avoids generating monomials in Γ_{g_α} not present in Γ.

Lastly, often constraints g_α may be eliminated through symbolic substitutions, which is usually the case for problems in quantum information theory. For example, if the objective polynomial is $f(\mathcal{O}) = X^2 + YX + XY$ with constraint $g(\mathcal{O}) = XY - YX = 0$, we need not build a localizing matrix for $g(\mathcal{O}) = 0$, as it can be used to rewrite the objective as $f'(\mathcal{O}) = X^2 + 2XY$.

For a more thorough review of NPA as a NCPOP, we point to Ref. [IR22], which also introduces Sums-Of-Squares relaxations of NCPOPs, which is the dual perspective to that of relaxations based on moment matrices.

2.3.3 No-signaling common cause

If no assumptions are made about the process Λ except that it satisfies the axiom of causality—no-signaling between space-like separated processes—then we say we are dealing with a no-signaling process, or a more-than-quantum or post-quantum source of correlations Λ. The axiom of causality implies that the marginal distribution for each process is independent of the input of the other process. We define the **no-signaling set** as the set of distributions which satisfy no-signaling constraints:

$$\mathcal{NS} \coloneqq \Bigg\{ p(ab|xy) \;\Big|\; \sum_b p(ab|xy) =: p_A(a|xy) = p_A(a|x)\,\forall y, \quad \sum_a p(ab|xy) =: p_B(b|xy) = p_B(b|y)\,\forall x \Bigg\}. \tag{2.96}$$

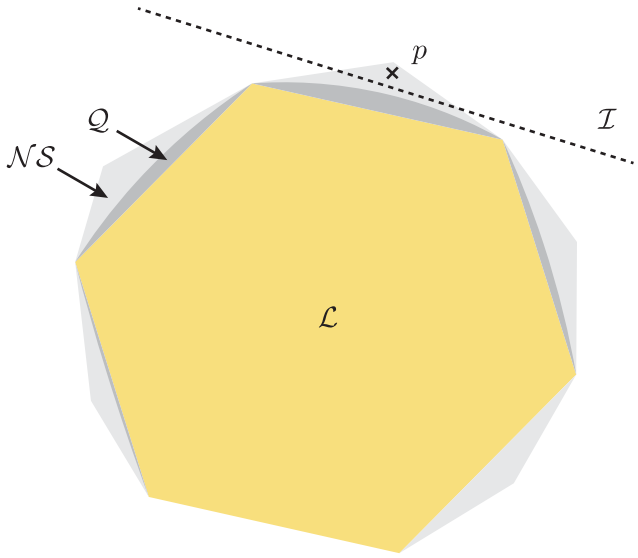


FIGURE 2.4: The sets of correlations $\mathcal{L}$, $\mathcal{Q}$, and $\mathcal{NS}$ achievable in a common-cause scenario under different assumptions on the process Λ. We also represent a Bell inequality $\mathcal{I}$ as a hyperplane which separates a point $p := p(ab|xy)$ from the appropriate set of correlations. In this case, $\mathcal{I}$ separates p from both $\mathcal{L}$ and $\mathcal{Q}$.

A slice of the higher-dimensional sets of local, quantum and no-signaling correlations is graphically depicted in Figure 2.4.

That the quantum set is strictly included in the no-signaling set is shown by Popescu and Rohrlich in their seminal work [Tsi93; PR94; Pop14], where they provide an example of a distribution which satisfies the axiom of causality but is nevertheless *incompatible* with quantum theory. This distribution is known as the "PR box", and it is over events with two outputs and two inputs ($a, b, x, y \in \{0, 1\}$):

$$p^{PR}(ab|xy) = \begin{cases} \frac{1}{2} & a \neq b, x = 1, y = 1 \\ \frac{1}{2} & a = b, x \neq 1, y \neq 1 \\ 0 & \text{otherwise} \end{cases} . \tag{2.97}$$

It achieves the maximum algebraic bound of the CHSH inequality of Eq.

(2.59), namely 4. It does so as it manifests perfect correlation and anti-correlation of the outputs in terms of the inputs x, y. This correlation, however, is only present when considering the joint statistics; locally the marginal distributions of the PR box are uniformly random, thus they are no-signaling. One may see the distribution resulting from the quantum strategy of Eqs. (2.53)—(2.57), $\tilde{p}(ab|xy)$, as a noisy version of the PR box: $\tilde{p}(ab|xy) = \frac{1}{\sqrt{2}} p^{PR}(ab|xy) + (1 - \frac{1}{\sqrt{2}})\frac{1}{4}$. As such, $\tilde{p}(ab|xy)$ is the closest approximation in nature[9] of the PR box.

In addition, PR boxes can be used to help distant parties infer strictly more classical bits of information about each other than the ones physically communicated [Dam05]. This inspired physical principles such as information causality, which limit the information that can be inferred about distant parties in such scenarios [Paw+09], and may potentially uniquely identify quantum theory among all possible theories.

Collins-Gisin (CG) parametrization. There is a useful parameterization of the set of no-signaling distributions. Let us consider the case of binary inputs and outputs. Then the set of distributions is fully parameterized by considering a set of 32 parameters $\{p(ab|xy)\}_{a,b,x,y\in\{0,1\}}$ together with normalization and no-signaling constraints:

$$\{p(ab|xy)\}_{a,b,x,y\in\{0,1\}} \quad + \quad \text{NS constraints} \quad + \quad \text{normalization.} \tag{2.98}$$

We may however reduce the number of parameters from 32 to 9:

$$\{1\} \cup \{p_A(0|x)\}_{x\in\{0,1\}} \cup \{p_B(0|y)\}_{y\in\{0,1\}} \cup \{p(00|xy)\}_{x,y\in\{0,1\}} + \\ + \quad \text{normalization inequalities.} \tag{2.99}$$

The no-signaling constraints have disappeared as they are implicitly used when labelling marginal distributions only by the input of the relevant party, e.g., $p_A(0|x)$ does not depend on y. The number 1 is written as an explicit parameter as it may have a different value when the distribution $p(ab|xy)$ is subnormalized. Regarding the normalization constraints, these have been included by removing events that involve the last outcome (a=1, b=1) of each party. Events with last outcomes can be expressed in terms of CG parameters, but they need to be positive in order to specify a valid joint

[9] Assuming quantum theory is the most precise theory in its regimes of applicability.

probability distribution:

$$p(11|xy) = 1 - p_A(0|x) - p_B(0|y) + p(00|xy) \geq 0\,. \tag{2.100}$$

The "normalization inequalities" of expression (2.99) include all expressions such as those of Eq. (2.100). This representation is also known as Collins-Gisin (CG) notation, introduced in Ref. [CG04]. We use CG notation in Chapter 6.

Chapter 3

Beyond common cause scenarios

In this chapter we study causal structures involving more than one independent system preparation, that is, several independent common-causes. We present a novel unified description of the inflation method for characterizing the compatibility of correlations with a given causal structure.

3.1 3-party chain scenario

The Bell common-cause scenario of Section 2.3 presents a single prepared system shared among various measurement devices. Here we consider another scenario with two independent preparations of bipartite systems that undergo three measurement processes, A, B and C, forming a chain-like causal

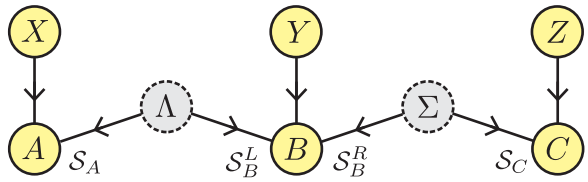


FIGURE 3.1: The three party chain scenario, also known as the "bilocal scenario", where three parties receive bipartite systems that are the output of two independent processes Λ and Σ. The input systems of B are labelled as "left" (L), $\mathcal{S}_B^L$, and "right" (R), $\mathcal{S}_B^R$.

structure, as shown in Figure 3.1. This scenario is also known as the "bilocal scenario" in the literature.[1]

If the preparations Λ, Σ are classical, the set of distributions achievable in the 3-party chain-like scenario, through unpacking as presented in Section 2.3, is given by:

$$\mathcal{L}^{\text{3-chain}} = \{p(abc|xyz) \,|\, p(abc|xyz) = \sum_{\bar{a}\setminus a_x} \sum_{\bar{b}\setminus b_y} \sum_{\bar{c}\setminus c_z} q(\bar{a},\bar{b},\bar{c}) \quad \text{s.t.} \quad \sum_{\bar{a},\bar{b},\bar{c}} q(\bar{a},\bar{b},\bar{c}) = 1,\ q(\bar{a},\bar{b},\bar{c}) \geq 0,\ q(\bar{a},\bar{c}) = q(\bar{a})q(\bar{c})\}, \quad (3.1)$$

where recall that bar notation is used to denote tuples of unpacked variables, $\bar{a} \coloneqq (a_1, \ldots, a_{|N_X|})$, $\bar{b} \coloneqq (b_1, \ldots, b_{|N_Y|})$ and $\bar{c} \coloneqq (c_1, \ldots, c_{|N_Z|})$, and that we use standard notation for marginalized distributions, $q(\bar{a},\bar{c}) \coloneqq \sum_{\bar{b}} q(\bar{a},\bar{b},\bar{c})$, $q(\bar{a}) \coloneqq \sum_{\bar{b}\bar{c}} q(\bar{a},\bar{b},\bar{c})$, $q(\bar{c}) \coloneqq \sum_{\bar{a}\bar{b}} q(\bar{a},\bar{b},\bar{c})$.[2]

Noticeably, the constraint $q(\bar{a},\bar{c}) = q(\bar{a})q(\bar{c})$ is a quadratic constraint. As such, determining membership of a distribution $p(abc|xyz)$ in the set $\mathcal{L}^{\text{3-chain}}$ is no longer a convex optimization program, but a *quadratic programming* (QP) problem; this implies that $\mathcal{L}^{\text{3-chain}}$ is not a convex set of distributions. For small instances, QP problems may be solved using branch-and-bound methods available through commercial software, such as GUROBI [Gur22], or open-source software, such as the BMIBNB solver available through YALMIP [Löf04].[3]

[1] Since the concept of a causal scenario as presented in this thesis makes no reference to any underlying theory, we choose a nomenclature that makes no reference to local causality.

[2] The set $\mathcal{L}^{\text{3-chain}}$ may also be defined, through the arguments of Section 2.3, in terms of local deterministic strategies, as distributions that satisfy $p(abc|xyz) = \sum_{\alpha\beta\gamma} q_{\alpha\beta\gamma} D(a|x\alpha)D(b|y\beta)D(c|z\gamma)$ where $\sum_{\alpha\beta\gamma} q_{\alpha\beta\gamma} = 1$, $q_{\alpha\beta\gamma} \geq 0$ and $q_{\alpha\gamma} = q_\alpha q_\gamma$.

[3] More precisely, QPs involve constraints of the form $x^T Q x + Ax + b \geq 0$, where Q is a semidefinite positive matrix, A a constant matrix and b a constant vector. While not all non-convex causal constraints are of the form of a QP, they can all be mapped to a QP through the addition of extra auxiliary variables. For example, the constraint $x = yzw$ can be mapped to quadratic two constraints $(x = y\alpha, \alpha = yz)$ through the addition of the variable α.

If Λ, Σ prepare quantum states, the set of quantum correlations is given by Born's rule:

$$\mathcal{Q}^{\text{3-chain}} = \{p(abc|xyz) \mid p(abc|xyz) = \text{tr}(\rho_\Lambda \otimes \rho_\Sigma)(A_{a|x} \otimes B_{b|y} \otimes C_{c|z})\}, \quad (3.2)$$

where the meaning of the operators is the same as in the quantum common cause scenario of Section 2.3. The quantum state is fixed by the causal constraint of the scenario to be a product state, $\rho_\Lambda \otimes \rho_\Sigma$. As such, the quantum set of correlations $\mathcal{Q}^{\text{3-chain}}$ is not convex. Arguments to prove convexity that use flags to construct a quantum strategy as in Eqs. (2.50)—(2.52) will construct a quantum state that is no longer a product state, thus the resulting distribution will be outside the set $\mathcal{Q}^{\text{3-chain}}$.

If Λ, Σ are no-signaling processes, the only constraint imposed by the axiom of causality is that of no-signaling and the statistical independence of A and C upon marginalizing B:

$$\mathcal{NS}^{\text{3-chain}} = \{p(abc|xyz) \mid p(abc|xyz) \in \mathcal{NS}^{ABC} \text{ and } p(ac|xz) = p(a|x)p(c|z)\}. \quad (3.3)$$

Statistical independence is also non-convex constraint, thus the set $\mathcal{NS}^{\text{3-chain}}$ is non-convex.

3.2 Inflation of the 3-party chain scenario

We will now present a relaxation of the sets of correlations—the "inflation" method [WSF19; NW20a; Wol+21]—which relies on the principle of *device replication*. The rationale behind device replication is rooted in the practical consideration that any experimental setup requires the preparation of a device which implements a specific physical process. However, one may prepare more than one such device, and wire them in any compatible configuration. For example, the device Σ in the 3-party chain scenario may be prepared twice, and the two preparations may be wired with the other elements in the circuit as in Figure 3.2. Such scenarios are called "inflations" of the original scenario.

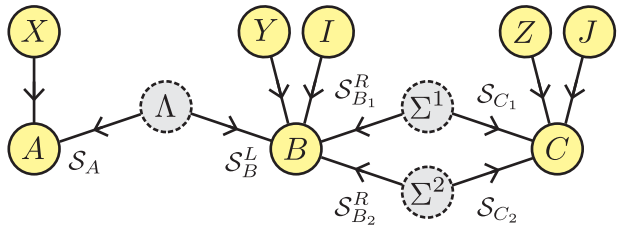


FIGURE 3.2: A causal scenario corresponding to an inflation of the 3-party chain scenario of Figure 3.1. The device implementing process Σ is manufactured two times, implementing processes Σ^1 and Σ^2 which are operationally indistinguishable. The inner working of the devices A, B and C is assumed to be the same, except that B and C, which have access to the copies of Σ, receive extra inputs, I and J respectively, which specify whether the processes take as input Σ^1 or Σ^2. In this inflation given that there are only two copies of Σ and one copy of Λ, the settings I, J take binary values, $I = i \in \{1, 2\}$, $J = j \in \{1, 2\}$.

The inflated scenario may be envisioned as an independent causal scenario of its own. However, one can consider strategies and protocols that connect the distributions realizable in the inflated scenario to those realizable in the original scenario. For example, while B in the inflated scenario receives three input systems $(\mathcal{S}^L_B, \mathcal{S}^R_{B_1}, \mathcal{S}^R_{B_2})$, as opposed to just $(\mathcal{S}^L_B, \mathcal{S}^R_B)$, one possible inflated process B is the one which implements the original process of the base scenario on either $(\mathcal{S}^L_B, \mathcal{S}^R_{B_1})$ or $(\mathcal{S}^L_B, \mathcal{S}^R_{B_2})$. We introduce an extra input I to the inflated process which can be used to specify on which input systems the process is applied to, with $I = i$ representing the choice of systems $(\mathcal{S}^L_B, \mathcal{S}^R_{B_i})$. Similarly, J acts as an input to C, enabling the selection of the system $\mathcal{S}_{C_j}$ upon which the original process from the base scenario is implemented.

Following this protocol, one can produce a distribution over the inflated scenario, $q(abc|x, (y, i), (z, j))$, that is connected to that of the original scenario, $p(abc|xyz)$, for different values of I and J:

$$q(abc|x, (y, i), (z, j)) = \begin{cases} p(abc|xyz) & i = j \\ p(ab|xy)p(c|z) & i \neq j \end{cases}. \tag{3.4}$$

Whenever $i \neq j$, it is the case that B and C do not measure the same copy of the source Σ, thus they are causally disconnected.

In addition, the preparations Σ^1, Σ^2 are operationally indistinguishable, therefore index labels are arbitrary. We denote by $\mathcal{G}$ the group of symmetries of the inflated scenario arising from the operational equivalence of preparations. In this example, as we only have two copies of Σ, the symmetry group is the permutation group of two elements, $\mathcal{G} = \text{Perm}(\{1, 2\})$. The distribution over the inflated graph must be invariant under the action of the group, which in this inflation corresponds to a simultaneous relabelling of the I, J indices:

$$q(abc|x, (y, i), (z, j)) = q(abc|x, (y, \pi(i)), (z, \pi(j))) \quad \forall\ \pi \in \mathcal{G} \tag{3.5}$$

Whenever $p(abc|xyz)$ is realizable in the 3-partite chain scenario, through the above protocol a distribution $q(abc|x, (y, i), (z, j))$ satisfying (3.4) and (3.5) can be produced in the inflated scenario:

$$\begin{aligned} p(abc|xyz) \in \mathcal{D}^{\text{3-chain}} \Rightarrow \\ \Rightarrow \exists\ q(abc|x, (y, i), (z, j)) \in \text{Inf}^{n_\Lambda=1, n_\Sigma=2}[\mathcal{D}^{\text{3-chain}}], \end{aligned} \tag{3.6}$$

where $\mathcal{D}^{\text{3-chain}}$ is the set of distributions realizable in the 3-party chain scenario (for some assumptions on the nature of the processes Λ, Σ; to be considered shortly) and $\text{Inf}^{n_\Lambda=1, n_\Sigma=2}[\mathcal{D}^{\text{3-chain}}]$ denotes the set of distributions realizable in the inflated scenario corresponding to one copy of Λ, two copies of Σ, and satisfying (3.4), (3.5).

Through the contrapositive of (3.6), if there does not exist a distribution $q(abc|x, (y, i), (z, j))$ compatible with (3.4) and (3.5), $p(abc|xyz)$ is incompatible with the 3-party chain scenario:

$$\begin{aligned} \nexists\ q(abc|x, (y, i), (z, j)) \in \text{Inf}^{n_\Lambda=1, n_\Sigma=2}[\mathcal{D}^{\text{3-chain}}] \Rightarrow \\ p(abc|xyz) \notin \mathcal{D}^{\text{3-chain}}. \end{aligned} \tag{3.7}$$

It is important to remark that the set $\text{Inf}^{n_\Lambda=1, n_\Sigma=2}[\mathcal{D}^{\text{3-chain}}]$ is strictly speaking more challenging to characterize than $\mathcal{D}^{\text{3-chain}}$. This is due to its non-convexity and to its larger dimension following the addition of the extra inputs I and J. We may however relax the non-linear constraints defining the sets, which corresponds to considering the convex hull of the set of

distributions. The convex hull of the set of correlations of the 3-party chain scenario is the same as that of a 3-party common-cause scenario,

$$\mathrm{Conv}(\mathcal{D}^{\text{3-chain}}) = \mathcal{D}^{ABC|XYZ},$$

where $\mathcal{D}^{ABC|XYZ}$ is the set of distributions achievable with a common-cause scenario with three processes A, B, C with inputs X, Y, Z. The convexification of the inflated scenario, however, is that of a common-cause (of a scenario with more inputs, I and J) *together with extra linear constraints*, (3.4) and (3.5), which are a consequence of the causal independence of Λ and Σ:

$$\begin{aligned}\mathrm{Conv}(\mathrm{Inf}^{n_\Lambda=1,n_\Sigma=2}[\mathcal{D}^{\text{3-chain}}]) := \\ \{q(abc|x,(y,i),(z,j)) \mid \\ q(abc|x,(y,i),(z,j)) \in \mathcal{D}^{ABC|X(Y,I)(Z,J)} \text{ s.t. } (3.4),(3.5)\}.\end{aligned} \tag{3.8}$$

Clearly, if a distribution exists on the inflated scenario, it also exists on its convexification. We may thus update the contrapositive (3.7) to:

$$\begin{aligned}\nexists\, q(abc|x,(y,i),(z,j)) \in \mathrm{Conv}(\mathrm{Inf}^{n_\Lambda=1,n_\Sigma=2}[\mathcal{D}^{\text{3-chain}}]) \Rightarrow \\ \Rightarrow p(abc|xyz) \notin \mathcal{D}^{\text{3-chain}}.\end{aligned} \tag{3.9}$$

This captures the *essence of the inflation technique*. Namely, inflated scenarios introduce extra symmetries which "survive" convexification, and which are a consequence of the non-convex properties of the original set of correlations. These constitute necessary but not sufficient conditions—expressed in terms of convex sets—for a distribution $p(abc|xyz)$ to be realizable in the original scenario of interest. The number of copies of preparations is known as the *level* of the inflation technique, and it may be a tuple (n_Λ, n_Σ) or a unique number such that $n = n_\Lambda = n_\Sigma$.

In what follows, we describe different methods for characterizing the set $\mathrm{Conv}(\mathrm{Inf}^{n_\Lambda=1,n_\Sigma=2}[\mathcal{D}^{\text{3-chain}}])$ under different assumptions about the type of processes Λ, Σ. The generalization to arbitrary numbers of copies of Λ, Σ is straightforward, but we will not present it in this section.

3.2.1 Classical preparations

If the preparations Λ, Σ, are classical processes, then as reasoned in Section 2.3, one may fully unpack the random variables. To simplify the graphical representation further on, it is useful to consider an intermediate unpacking. For example, for B, we "repack" the variable Y, leaving two unpacked variables $B_{I=0}$, $B_{I=1}$ with inputs Y_1, Y_2:

$$\underbrace{B_{Y=0,I=0},\ldots,B_{Y=|N_Y|,I=0}}_{=:B_{I=0}},\ \underbrace{B_{Y=0,I=1},\ldots,B_{Y=|N_Y|,I=1}}_{=:B_{I=1}}. \tag{3.10}$$

With full generality, the unpacked variables are connected to all the input systems of B, e.g., $B_{I=1}$ can be connected to Σ^2 and $B_{I=2}$ to Σ^1. However, we may restrict to the protocol in which I acts as a flag fixing a particular wiring of the input systems. This justifies removing the causal connections from $B_{I=1}$ to Σ^2 and from $B_{I=2}$ to Σ^1.[4] Similar considerations hold for C. This results in the scenario of Figure 3.3. Note that we only use hybrid unpacking for the graphical representation; in the formulaic description we use full unpacking.

It is noteworthy that Λ is now being shared between three processes, as opposed to two as in the original scenario. As such, a $(n_\Lambda{=}\,1, n_\Sigma{=}\,2)$ inflation makes use of the classical nature of Λ, whereas the type of process Σ is not constrained in any way.[5] An inflation which considers several copies of both preparations, Λ and Σ, through unpacking will lead to scenarios that exploit the classical nature of both processes.

Lastly, note that while we only duplicated a preparation device, through unpacking we arrived at duplicated measurement processes B_1, B_2 and C_1, C_2, with their own independent inputs Y_1, Y_2 and Z_1, Z_2. The unpacked distribution $q(\bar{a},\bar{b}_{I=1},\bar{b}_{I=2},\bar{c}_{J=1},\bar{c}_{J=2}) =: q(\bar{a},\bar{b}_1,\bar{b}_2,\bar{c}_1,\bar{c}_2)$ connects with

[4] Recall that without this protocol, we cannot connect the distribution over the inflated scenario with that over the original scenario. An alternative to removing said connections from the causal scenario is to make a fine-tuning argument, and define processes B_1 and B_2 to be such that the factorization (3.4) holds. However, it is more convenient to depict this fine-tuning assumption graphically by removing the appropriate connections.

[5] This observation lead to the definition of "full network nonlocality" of Ref. [PKGT22].

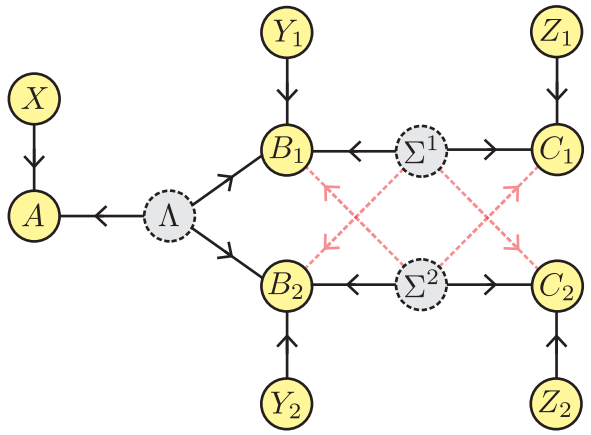


FIGURE 3.3: The scenario resulting from unpacking inputs I and J in the inflation of Figure 3.2. The interpretation of I and J as inputs that determine a wiring has been used to remove the connections—denoted by a dashed red arrows in the figure—from Σ^2 to B_1 and Σ^1 to B^2, and similarly, from Σ^2 to C_1 and Σ^1 to C_2.

$q(abc|x,(y,i),(z,j))$ as follows:

$$\begin{aligned}
q(abc|x,(y,1),(z,1)) &\coloneqq \sum_{\bar{a}\setminus a_x}\sum_{\bar{b}_1\setminus b_y}\sum_{\bar{b}_2}\sum_{\bar{c}_1\setminus c_z}\sum_{\bar{c}_2} q(\bar{a},\bar{b}_1,\bar{b}_2,\bar{c}_1,\bar{c}_2),\\
q(abc|x,(y,1),(z,2)) &\coloneqq \sum_{\bar{a}\setminus a_x}\sum_{\bar{b}_1\setminus b_y}\sum_{\bar{b}_2}\sum_{\bar{c}_1}\sum_{\bar{c}_2\setminus c_z} q(\bar{a},\bar{b}_1,\bar{b}_2,\bar{c}_1,\bar{c}_2),\\
q(abc|x,(y,2),(z,1)) &\coloneqq \sum_{\bar{a}\setminus a_x}\sum_{\bar{b}_1}\sum_{\bar{b}_2\setminus b_y}\sum_{\bar{c}_1\setminus c_z}\sum_{\bar{c}_2} q(\bar{a},\bar{b}_1,\bar{b}_2,\bar{c}_1,\bar{c}_2),\\
q(abc|x,(y,2),(z,2)) &\coloneqq \sum_{\bar{a}\setminus a_x}\sum_{\bar{b}_1}\sum_{\bar{b}_2\setminus b_y}\sum_{\bar{c}_1}\sum_{\bar{c}_2\setminus c_z} q(\bar{a},\bar{b}_1,\bar{b}_2,\bar{c}_1,\bar{c}_2).
\end{aligned} \tag{3.11}$$

Consider that there exists an unpacked distribution which marginalizes to $q(abc|x,(y,i),(z,j))$ and which satisfies the appropriate symmetries (3.5). Invariance under relabelling of Σ^i implies invariance under the joint swap $(I=1,J=1)\leftrightarrow(I=2,J=2)$. Let us consider $q(abc|x,(y,1),(z,1))$ and $q(abc|x,(y,2),(z,2))$. Since they are equal, averaging them does not change

their value:

$$q(abc|x,(y,1),(z,1)) = \frac{q(abc|x,(y,1),(z,1)) + q(abc|x,(y,2),(z,2))}{2} =$$
$$= \frac{1}{2}\left(\sum_{\bar{a}\setminus a_x}\sum_{\bar{b}_1\setminus b_y}\sum_{\bar{b}_2}\sum_{\bar{c}_1\setminus c_z}\sum_{\bar{c}_2} q(\bar{a},\bar{b}_1,\bar{b}_2,\bar{c}_1,\bar{c}_2) + \right.$$
$$\left. + \sum_{\bar{a}\setminus a_x}\sum_{\bar{b}_1}\sum_{\bar{b}_2\setminus b_y}\sum_{\bar{c}_1}\sum_{\bar{c}_2\setminus c_z} q(\bar{a},\bar{b}_1,\bar{b}_2,\bar{c}_1,\bar{c}_2)\right) =$$
$$= \sum_{\bar{a}\setminus a_x}\sum_{\bar{b}_1\setminus b_y}\sum_{\bar{b}_2}\sum_{\bar{c}_1\setminus c_z}\sum_{\bar{c}_2} \underbrace{\frac{q(\bar{a},\bar{b}_1,\bar{b}_2,\bar{c}_1,\bar{c}_2) + q(\bar{a},\bar{b}_2,\bar{b}_1,\bar{c}_2,\bar{c}_1)}{2}}_{=:\tilde{q}(\bar{a},\bar{b}_1,\bar{b}_2,\bar{c}_1,\bar{c}_2)} \tag{3.12}$$

where in the last step we relabelled jointly the indices $(I{=}1, J{=}1) \leftrightarrow (I{=}2, J{=}2)$ of the second sum, which can be done with full generality. Note that $\tilde{q}(\bar{a},\bar{b}_1,\bar{b}_2,\bar{c}_1,\bar{c}_2)$ is invariant under the joint swap of the parties (B_1, C_1) and (B_2, C_2). Therefore, if there exists a (potentially asymmetric) unpacked distribution that marginalizes to a symmetric $q(abc|x,(y,i),(z,j))$, one can construct a symmetric unpacked distribution $\tilde{q}$ which marginalizes to the same distribution. We have thus lifted the symmetries of Eq. (3.5), described at the level of $q(abc|x,(y,i),(z,j))$, to the level of the unpacked distribution. This can be done without loss of generality as long as we do not consider non-convex constraints. The convex constraints implied by the inflated causal scenario are the following:

$$\sum_{\bar{b}_1} q(\bar{a},\bar{b}_1,\bar{b}_2,\bar{c}_1,\bar{c}_2) = q(\bar{a},\bar{b}_2,\bar{c}_2)\, q(\bar{c}_1), \tag{3.13}$$

$$\sum_{\bar{b}_2} q(\bar{a},\bar{b}_1,\bar{b}_2,\bar{c}_1,\bar{c}_2) = q(\bar{a},\bar{b}_1,\bar{c}_1)\, q(\bar{c}_2). \tag{3.14}$$

It is easy to check that the symmetrized unpacked distribution $\tilde{q}$ need not satisfy them. For example:

$$\sum_{\bar{b}_1} \tilde{q}(\bar{a},\bar{b}_1,\bar{b}_2,\bar{c}_1,\bar{c}_2) = \frac{q_{AB_2C_2}(\bar{a},\bar{b}_2,\bar{c}_2)q_{C_1}(\bar{c}_1) + q_{AB_1C_1}(\bar{a},\bar{b}_2,\bar{c}_2)q_{C_2}(\bar{c}_1)}{2}$$
$$\neq \tilde{q}(\bar{a},\bar{b}_2,\bar{c}_2)\, \tilde{q}(\bar{c}_1).$$

While non-convex constraints are relevant for characterizing the set of distributions on the inflated scenario, $\mathrm{Inf}^{n_\Lambda=1,n_\Sigma=2}[\mathcal{L}^{\text{3-chain}}]$, they are not relevant when considering its convexification, $\mathrm{Conv}(\mathrm{Inf}^{n_\Lambda=1,n_\Sigma=2}[\mathcal{L}^{\text{3-chain}}])$. Therefore, we may without loss of generality consider symmetrized unpacked distributions.

Now we may define a program that looks for the existence of an unpacked distribution satisfying the inflation symmetries and which marginalizes to (products of marginals of) the distribution of interest:

$$
\begin{aligned}
\text{find}\quad & \tilde{q}(\bar{a},\bar{b}_1,\bar{b}_2,\bar{c}_1,\bar{c}_2)\\
\text{s.t.}\quad & \tilde{q}(\bar{a},\bar{b}_1,\bar{b}_2,\bar{c}_1,\bar{c}_2)\geq 0 \quad \forall \bar{a},\bar{b}_1,\bar{b}_2,\bar{c}_1,\bar{c}_2\\
& \textstyle\sum_{\bar{a},\bar{b}_1,\bar{b}_2,\bar{c}_1,\bar{c}_2}\tilde{q}(\bar{a},\bar{b}_1,\bar{b}_2,\bar{c}_1,\bar{c}_2)=1\\
& q(abc|x,(y,i),(z,j)) \coloneqq \sum_{\bar{a}\setminus a_x}\sum_{\bar{b}_i\setminus b_y}\sum_{\{\bar{b}_\alpha\}_{\alpha\neq i}}\sum_{\bar{c}_j\setminus c_z}\sum_{\{\bar{c}_\beta\}_{\beta\neq j}}\tilde{q}(\bar{a},\bar{b}_1,\bar{b}_2,\bar{c}_1,\bar{c}_2)\\
& q(abc|x,(y,i),(z,j)) = \begin{cases} p(abc|xyz) & i=j\\ p(ab|xy)p(c|z) & i\neq j\end{cases}\\
& \tilde{q}(\bar{a},\bar{b}_1,\bar{b}_2,\bar{c}_1,\bar{c}_2)=\tilde{q}(\bar{a},\bar{b}_2,\bar{b}_1,\bar{c}_2,\bar{c}_1)\quad \forall \bar{a},\bar{b}_1,\bar{b}_2,\bar{c}_1,\bar{c}_2
\end{aligned}
\tag{3.15}
$$

Note that the assumption that $\tilde{q}$ satisfies the inflation symmetries can be used to significantly reduce the number of variables in the linear program.

We re-emphasize that the convexified problem corresponds to a scenario which has no causal independence constraints, thus it is equivalent to a common-cause scenario with extra symmetries where only marginals of the total joint distribution are known. The convexified scenario is depicted in Figure 3.4.

As the program (3.15) is a linear programming problem, we may use Farkas' lemma (Appendix A) to extract a certificate in case of infeasibility. Such certificate will be a linear combination of the constant part of the linear constraints of LP (3.15), which are non-linear in terms of the original distribution $p(abc|xyz)$. As such, the certificates of infeasibility that arise from inflation methods are *polynomial* in $p(abc|xyz)$. For a concrete example, consider a distribution resulting from a wiring of two PR boxes [Bra+12],

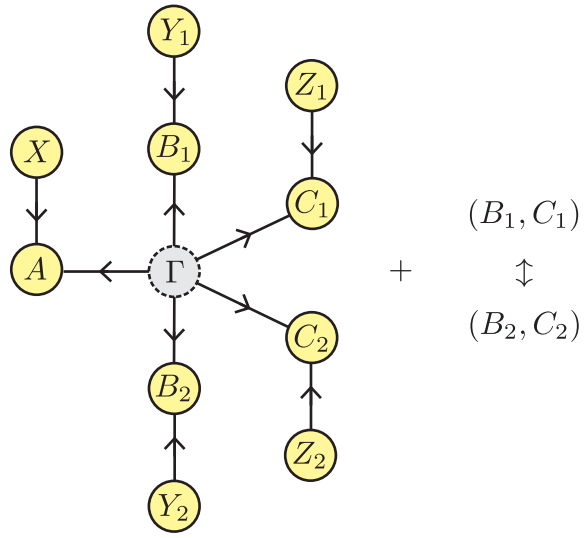


FIGURE 3.4: The convexification of the scenario of Figure 3.3 which results from removing the non-convex constraints of program (3.15). The symmetries derived from inflation, that is, the invariance under the joint swap $(B_1, C_1) \leftrightarrow (B_2, C_2)$, survive the convexification.

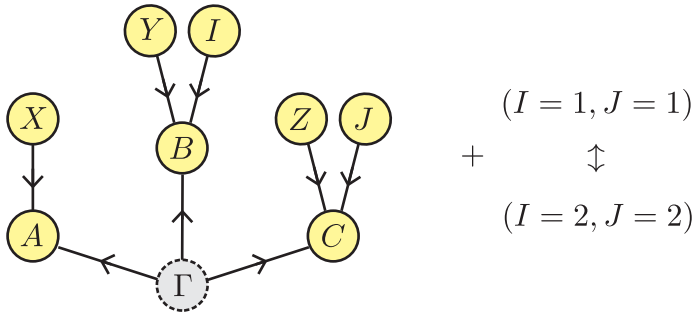


FIGURE 3.5: The convexification of the scenario of Figure 3.3 together with symmetries derived from inflation, that is, the invariance under the joint swap of indexes $(I = 1, J = 1) \leftrightarrow (I = 2, J = 2)$. This is used for characterizing the quantum set of correlations in the 3-party chain scenario.

mixed with the uniform distribution:

$$p^v_{\text{2-PR}}(abc|xyz) = \frac{1 + v(-1)^{a+b+c+xy+yz}}{8}. \tag{3.16}$$

It is known that the distribution $p^v_{\text{2-PR}}(abc|xyz)$ is compatible with classical preparations for $0 \leq v \leq \frac{1}{4}$. The inflation considered so far in this section detects $p^v_{\text{2-PR}}(abc|xyz)$ to be incompatible with the 3-party chain scenario with classical preparations for $v > \frac{1}{2}$. Running the LP (3.15) with $p(abc|xyz) = p^{1/2+\epsilon}_{\text{2-PR}}(abc|xyz)$ with $\epsilon = 10^{-5}$ we find that there is no solution; with Farkas' lemma we extract the following certificate of infeasibility:

$$\begin{aligned}&(pC(0|0) - 1)\,(p(000|001) + p(000|101) - p_{BC}(00|01) - p_A(0|1)\,p_C(0|1)) + \\ &p_C(0|1)\,(p(000|110) - p(000|010) + p_{AB}(00|01) - p_{AB}(00|11)) \leq 0,\end{aligned} \tag{3.17}$$

which is nonlinear in the original observed distribution. Higher levels of inflation can certify lower visibilities v as being infeasible; for $(n_\Lambda = 2, n_\Sigma = 2)$, we can certify the range $v \gtrapprox 0.3$ as being incompatible with the classical 3-party chain scenario. In the limit of infinite number of preparation copies, $n = n_\Lambda = n_\Sigma$ the inflation technique for classical preparations is known to converge [WSF19], i.e.,

$$\lim_{n\to\infty} \text{Conv}(\text{Inf}^{n_\Lambda=n, n_\Sigma=n}[\mathcal{L}^{\text{3-chain}}]) = \mathcal{L}^{\text{3-chain}}.$$

3.2.2 Quantum preparations

If the preparations Λ, Σ are quantum processes, we may characterize the set of correlations $\text{Conv}(\text{Inf}^{n_\Lambda=1, n_\Sigma=2}[\mathcal{Q}^{\text{3-chain}}])$ with the following modification of the NPA hierarchy. We build the standard alphabet, including also the inputs I and J:

$$\mathcal{O} = \{\mathbb{1}\} \cup \{A_{a|x}\}_{xa} \cup \{B^{I=1}_{b|y}\}_{yb} \cup \{B^{I=2}_{b|y}\}_{yb} \cup \{C^{J=1}_{c|z}\}_{zc} \cup \{C^{J=2}_{c|z}\}_{zc},$$

and a generating set $\mathcal{S}$ of monomials in $\mathcal{O}$. Note that while C's operators commute for different values of J, $[C^{J=1}_{c|z}, C^{J=2}_{c|z}] = 0$, this is not the case for B:

$$[B^{I=1}_{b|y}, B^{I=2}_{b|y}] \neq 0, \tag{3.18}$$

as the operators of B overlap by acting on the same share of the state prepared by Λ. The moment matrix has elements (α, β) given by:

$$\Gamma^{\mathcal{O}}_{\alpha\beta}(\rho) = \operatorname{tr} \rho \mathcal{S}^{\dagger}_{\alpha} \mathcal{S}_{\beta}.$$

The distribution $q(abc|x,(y,i),(z,j))$ corresponds to the following moments (we simplify the notation by writing $B^{I=i}_{b|y} =: B^{i}_{yb}$ and similarly for C):

$$q(abc|x,(y,i),(z,j)) = \langle A_{a|x} B^{i}_{b|y} C^{j}_{c|z} \rangle = \begin{cases} p(abc|xyz) & i = j \\ p(ab|xy)p(c|z) & i \neq j \end{cases}. \tag{3.19}$$

There exist more moments, however, that connect to $q(abc|x,(y,i),(z,j))$, which are built from monomials in which all the operators commute. Moments of sequences of all-commuting positive semidefinite operators are themselves non-negative, as such they define a probability distribution. For this inflated scenario, the following are the only two patterns of sequences of maximum length of all-commuting operators:

$$q_1(ab_1c_1c_2|xy_1z_1z_2) \coloneqq \langle A_{a|x} B^{1}_{y_1 b_1} C^{1}_{z_1 c_1} C^{2}_{z_1 c_2} \rangle = \tag{3.20}$$

$$= q_1(ab_1c_1|x,(y_1,1),(z_1,1))\, q_1(c_2|(z_2,2)) \tag{3.21}$$

$$q_2(ab_2c_1c_2|xy_2z_1z_2) \coloneqq \langle A_{a|x} B^{2}_{y_2 b_2} C^{1}_{z_1 c_1} C^{2}_{z_1 c_2} \rangle = \tag{3.22}$$

$$= q_2(ab_2c_2|x,(y_2,2),(z_1,2))\, q_2(c_1|(z_1,1)) \tag{3.23}$$

Note that q_1 and q_2 correspond to a **partial unpacking** of the distribution $q(abc|x,(y,i),(z,j))$:

$$q(abc|x,(y,i{=}1),(z,j)) = \sum_{\{c_\alpha\}_{\alpha\neq j}} q_1(ab_1c_1c_2|xy_1z_1z_2) \tag{3.24}$$

$$q(abc|x,(y,i{=}2),(z,j)) = \sum_{\{c_\alpha\}_{\alpha\neq j}} q_2(ab_2c_1c_2|xy_2z_1z_2) \tag{3.25}$$

In addition, since for different values of the settings z_1, z_2, the operators of C_1 and C_2 may not commute, $[C^{J}_{zc}, C^{J}_{z'c'}] \neq 0$ for $z \neq z'$, the distribution cannot be unpacked over z_1, z_2 (similarly for B and unpacking over y_1, y_2). We may interpret q_1 and q_2 as being distributions over two different inflations:

Analogous constraints hold for other moments which cannot be associated with the known distribution, such as:

$$\langle A_{00}B^1_{00}B^1_{10}C^2_{00}C^2_{10}\rangle = \langle A_{00}B^1_{00}B^1_{10}\rangle\langle C^2_{00}C^2_{10}\rangle. \tag{3.26}$$

As these types of constraints are non-convex, they do not form part of the definition of the convexified set $\mathrm{Conv}(\mathrm{Inf}^{n_\Lambda=1,n_\Sigma=2}[\mathcal{Q}^{\text{3-chain}}])$, and thus we need not consider them.

Lastly, we need to consider the action of the symmetry group $\mathcal{G}$ on the set of moments. This corresponds to a simultaneous index relabelling of the values of I and J consistent with a swap in the preparations Σ^1, Σ^2. For example, the following moments are equivalent:

$$\langle A_{00}B^1_{00}B^1_{10}C^2_{00}C^2_{10}\rangle \overset{(I=1,J=1)\leftrightarrow(I=2,J=2)}{=} \langle A_{00}B^2_{00}B^2_{10}C^1_{00}C^1_{10}\rangle. \tag{3.27}$$

For the classical case, in Eq. (3.12), we showed how given an unpacked distribution giving rise to a symmetric $q(abc|x,(y,i),(z,j))$, we may generally construct a symmetric unpacked distribution. By using the convexity of the set of solutions of a SDP, we can show a similar result. For any moments $\{x_i\}_i$ giving rise to a symmetric $q(abc|x,(y,i),(z,j))$, we can build a solution $\{x^*_i\}_i$ that satisfies the inflation symmetries by averaging the action of the group of symmetries $\mathcal{G}$ on the solution $\{x_i\}_i$:

$$x^*_i = \frac{1}{|\mathcal{G}|}\sum_{g\in\mathcal{G}} g(x_i), \tag{3.28}$$

where $|\mathcal{G}|$ is the number of elements in the group. Therefore, we may without loss of generality assume that all moments in the NPA relaxation—including those that are not connected to $q(abc|x,(y,i),(z,j))$—satisfy the inflation symmetries.

3.2.3 No-signaling preparations

If Λ and Σ satisfy only no-signaling constraints, distributing the prepared systems across more measurement devices than in the initial setup becomes infeasible. This is possible when the preparations are classical systems as classical information can be arbitrarily copied, thus it can correlate equally an indefinite number of measurement devices. This process is referred to

as "fanning out" the preparation. A *fanout inflation* is one which fans out preparations, such as the one of Figure 3.3, where the system prepared by Λ is being distributed among three measurement devices, $(A,\ B_1,\ B_2)$, as opposed to just (A, B) as in the baseline scenario. Conversely, a *non-fanout inflation* is one which does not expand the distribution of the preparation beyond the initial setup.

The protocol in which I and J are interpreted as fixing a certain wiring between the preparations and the measurement devices is still compatible with non-fanout inflations and with certain unpackings. For example, for a fixed value of I, one may unpack over J as observed when considering quantum inflation in Eqs. (3.24) and (3.25). These partial unpackings correspond to looking for distributions q_1 and q_2 compatible with the inflation of Figure 3.6. This inflation is "trivial" as it includes the original scenario, thus it does not add new constraints. As a consequence, we cannot develop convex programming tests for membership in $\mathcal{NS}^{\text{3-chain}}$, as the convexification of inflation scenarios are in essence equivalent to just the common-cause no-signaling set of correlations, with no extra symmetries.

The non-convex nature of the sets of distributions is graphically depicted in Figure 3.7, together with a representation of a polynomial Bell inequality such as that of Eq. (3.17).

3.3 Triangle scenario

The last scenario that we will consider is that of three preparations Λ, Σ and Γ of bipartite systems subject to measurements A, B and C in a triangle-like causal structure. This is known as the triangle scenario, and it is depicted in Figure 3.8. Note that in this scenario the measurement devices A, B and C do not have independent inputs X, Y and Z as in the other scenarios considered so far. The observed joint distribution over this scenario is then $p(abc)$. We will now define the three sets of correlations under the assumptions of classical, quantum and no-signaling preparations.

The classical set of correlations is given by:

$$p(abc) = \int_{\tilde{\lambda}\in\mathbb{R}} \int_{\tilde{\sigma}\in\mathbb{R}} \int_{\tilde{\gamma}\in\mathbb{R}} q(a|\tilde{\lambda}\,\tilde{\sigma}) q(b|\tilde{\lambda}\,\tilde{\gamma}) q(c|\tilde{\sigma}\,\tilde{\gamma}) q(\tilde{\lambda}) q(\tilde{\sigma}) q(\tilde{\gamma})\, \mathrm{d}\tilde{\lambda}\mathrm{d}\tilde{\sigma}\mathrm{d}\tilde{\gamma} \tag{3.29}$$

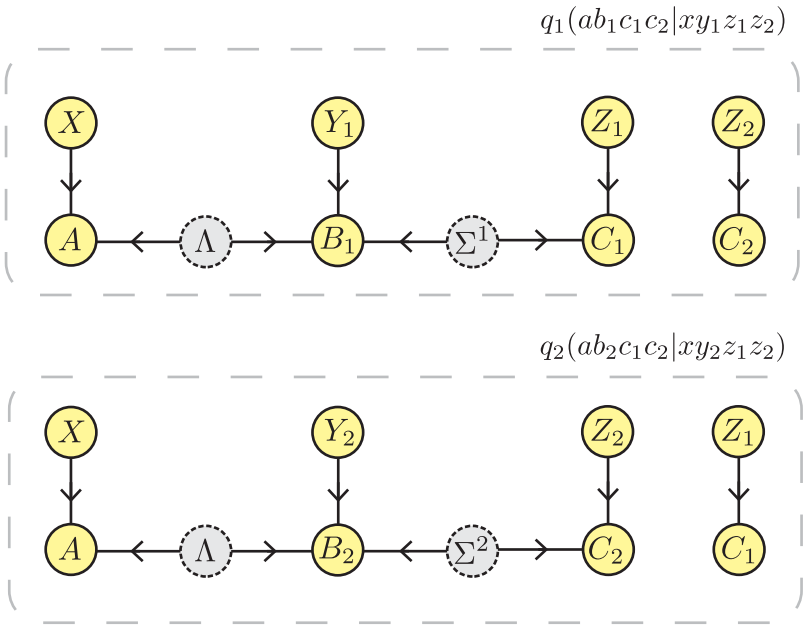


FIGURE 3.6: A visual depiction of the causal constraints the partial unpackings described by Eqs. (3.24) and (3.25) must satisfy. Note that this is circular: by inflating the original scenario, we arrive at the several scenarios which are essentially equivalent to the original scenario if we do not impose non-convex constraints. Therefore, non-fanout inflation (see text for definition) does not lead to any novel constraints that survive convexification.

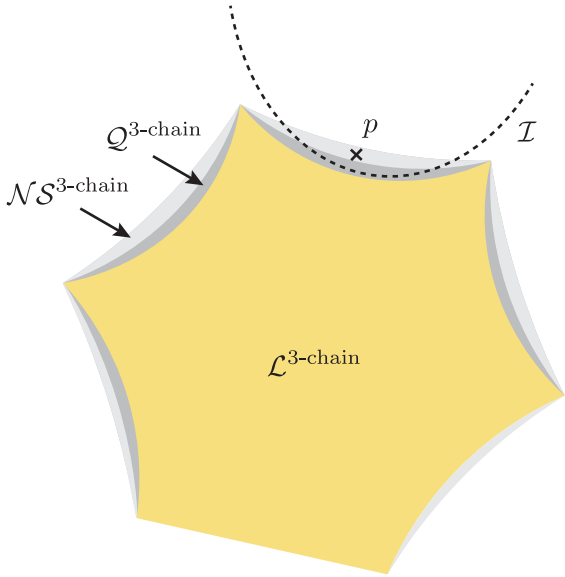


FIGURE 3.7: A visual depiction the non-convex nature of the sets of distributions $\mathcal{L}^{\text{3-chain}}$, $\mathcal{Q}^{\text{3-chain}}$ and $\mathcal{NS}^{\text{3-chain}}$. See in text Eqs. (3.1), (3.2) and ((3.3)) for the definition of these sets. The vertices of the set $\mathcal{L}^{\text{3-chain}}$ are the same as those of the classical common-cause set, $\mathcal{L}^{ABC|XYZ}$. From inflation methods one may extract polynomial bell inequalities $\mathcal{I}(p(abc|xyz) \leq 0$, where $\mathcal{I}(p(abc|xyz))$ is a polynomial in $p(abc|xyz)$. For an example, see the Bell inequality of Eq. (3.17).

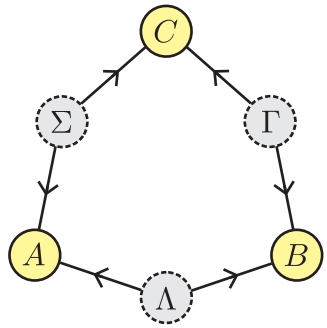


FIGURE 3.8: The triangle causal scenario where the bipartite systems prepared by three preparations Λ, Σ and Γ are submitted to measurements A, B and C that each receive two subsystems. Note that in this scenario, as opposed to common-cause scenarios or the 3-party chain scenario, there are no inputs X, Y, Z to the processes A, B, C.

Using the same derivation as for the common-cause scenario, we can use inverse transform sampling to simulate the response functions $q(a|\tilde{\lambda}\tilde{\sigma})$, $q(b|\tilde{\lambda}\tilde{\gamma})$, $q(c|\tilde{\sigma}\tilde{\gamma})$ using a deterministic function and a uniform random variable. For this, we enlarge the latent variables to include a uniform random variable and denote the enlarged variables with primes, $\tilde{\Lambda}'$, $\tilde{\Sigma}'$, $\tilde{\Gamma}'$. In addition, we can perform a change of variables to the cumulative distribution function for each latent variable:

$$\tilde{\lambda}'' = \int_{-\infty}^{\tilde{\lambda}''} q(\tilde{\lambda}')\,\mathrm{d}\tilde{\lambda}', \qquad \tilde{\sigma}'' = \int_{-\infty}^{\tilde{\sigma}''} q(\tilde{\sigma}')\,\mathrm{d}\tilde{\sigma}', \qquad \tilde{\gamma}'' = \int_{-\infty}^{\tilde{\gamma}''} q(\tilde{\gamma}')\,\mathrm{d}\tilde{\gamma}',$$

where the new variables are denoted with double prime. This is an invertible change of variable, $\tilde{\lambda}'' = f(\tilde{\lambda}')$, $\tilde{\sigma}'' = g(\tilde{\sigma}')$, $\tilde{\gamma}'' = h(\tilde{\gamma}')$, where f, g, h are injective functions. Note that $\mathrm{d}\tilde{\lambda}'' = \frac{df}{d\tilde{\lambda}'}(\tilde{\lambda}')\,\mathrm{d}\tilde{\lambda}' = q(\tilde{\lambda}')\,\mathrm{d}\tilde{\lambda}'$ and similarly for the other variables. This allows us to absorb the marginals $q(\tilde{\lambda}')$, $q(\tilde{\sigma}')$, $q(\tilde{\gamma}')$

into the new variables. Altogether:

$$p(abc) = \int_{\tilde{\lambda},\tilde{\sigma},\tilde{\gamma}\in\mathbb{R}^3} q(a|\tilde{\lambda}\tilde{\sigma})q(b|\tilde{\lambda}\tilde{\gamma})q(c|\tilde{\sigma}\tilde{\gamma})q(\tilde{\lambda})q(\tilde{\sigma})q(\tilde{\gamma})\,\mathrm{d}\tilde{\lambda}\mathrm{d}\tilde{\sigma}\mathrm{d}\tilde{\gamma} \tag{3.30}$$

$$\downarrow \text{(change to deterministic response functions)}$$

$$= \int_{\tilde{\lambda}',\tilde{\sigma}',\tilde{\gamma}'\in\mathbb{R}^3} q(a|\tilde{\lambda}'\tilde{\sigma}')q(b|\tilde{\lambda}'\tilde{\gamma}')q(c|\tilde{\sigma}'\tilde{\gamma}')q(\tilde{\lambda}')q(\tilde{\sigma}')q(\tilde{\gamma}')\,\mathrm{d}\tilde{\lambda}'\mathrm{d}\tilde{\sigma}'\mathrm{d}\tilde{\gamma}' \tag{3.31}$$

$$\downarrow \text{(change to cumulative variables)}$$

$$= \int_{\tilde{\lambda}'',\tilde{\sigma}'',\tilde{\gamma}''\in[0,1]^3} q(a|f^{-1}(\tilde{\lambda}'')g^{-1}(\tilde{\sigma})'')\cdot \cdot q(b|f^{-1}(\tilde{\lambda}'')h^{-1}(\tilde{\gamma}''))\,q(c|g^{-1}(\tilde{\sigma}'')h^{-1}(\tilde{\gamma}''))\,\mathrm{d}\tilde{\lambda}''\mathrm{d}\tilde{\sigma}''\mathrm{d}\tilde{\gamma}'' \tag{3.32}$$

$$\downarrow \text{(redefine } q \text{ to absorb } f^{-1}, g^{-1}, h^{-1})$$

$$= \int_{\tilde{\lambda}'',\tilde{\sigma}'',\tilde{\gamma}''\in[0,1]^3} q(a|\tilde{\lambda}''\tilde{\sigma}'')q(b|\tilde{\lambda}''\tilde{\gamma}'')q(c|\tilde{\sigma}''\tilde{\gamma}'')\mathrm{d}\tilde{\lambda}''\mathrm{d}\tilde{\sigma}''\,\mathrm{d}\tilde{\gamma}''. \tag{3.33}$$

We have now arrived, without loss of generality, at a simplified description of the classical set of correlations in the triangle scenario:

$$\mathcal{L}^{\text{triangle}} = \{p(abc)\,|\,p(abc) = \int_{\tilde{\lambda}'',\tilde{\sigma}'',\tilde{\gamma}''\in[0,1]^3} q(a|\tilde{\lambda}''\tilde{\sigma}'')\,q(b|\tilde{\lambda}''\tilde{\gamma}'')\,q(c|\tilde{\sigma}''\tilde{\gamma}'')\,\mathrm{d}\tilde{\lambda}''\,\mathrm{d}\tilde{\sigma}''\,\mathrm{d}\tilde{\gamma}''\}. \tag{3.34}$$

In this description, the values of the latent variables are continuous and restricted to a unit cube, while the response functions are deterministic. Note that in the 3-party chain scenario, it was possible to coarse grain the distributions and to arrive at a relatively simple description polynomial description of the non-convex set of correlations (cf. Eq. (3.1)). This is not so straightforward in the triangle scenario. There are techniques to restrict the cardinality of the latent variables through Carathéodory's theorem (see Ref. [Wol]), however we will not cover this in this dissertation. Note that even though the response functions are deterministic, the other measurement devices cannot predict the response of the rest from only two values for the latent variables. For example, while A has access to $\tilde{\lambda}''$ and $\tilde{\sigma}''$, they do not know the value of $\tilde{\gamma}''$, needed to predict the outcome of B and C.

The quantum set of correlations is given by

$$\mathcal{Q}^{\text{triangle}} = \{p(abc|xyz) \,|\, p(abc|xyz) = \\ \text{tr}(\rho_\Lambda \otimes \rho_\Sigma \otimes \rho_\Gamma)(A_{a|x} \otimes B_{b|y} \otimes C_{c|z})\}. \quad (3.35)$$

Lastly, given that there are no inputs X, Y and Z for A, B and C, it is not immediately clear how the axiom of causality is expressed in this scenario at the level of correlations. We defer the definition of $\mathcal{NS}^{\text{triangle}}$ to a later section.

3.4 Inflation of the triangle scenario

We consider an inflation of the triangle scenario which includes two copies of the preparations Λ, Σ and Γ. This inflation is depicted in Figure 3.9. The inputs I^A, I^B and I^C describe which copy of the preparations each measurement uses. Their value is a tuple of two numbers:

$$I^A = (i^A_\Sigma, i^A_\Lambda), \qquad I^B = (i^B_\Lambda, i^B_\Gamma), \qquad I^C = (i^C_\Gamma, i^C_\Sigma),$$

where $i^A_\Sigma, i^A_\Lambda, i^B_\Lambda, i^B_\Gamma, i^C_\Gamma, i^C_\Sigma \in \{1, 2\}$. The fundamental object is the joint distribution over the inflated scenario given these inputs:

$$q(abc|(i^A_\Sigma, i^A_\Lambda), (i^B_\Lambda, i^B_\Gamma), (i^C_\Gamma, i^C_\Sigma)) = \\ = \begin{cases} p(abc) & i^A_\Sigma = i^C_\Sigma,\ i^A_\Lambda = i^B_\Lambda,\ i^C_\Sigma = i^B_\Sigma \\ p(ac)p(b) & i^A_\Sigma = i^C_\Sigma,\ i^A_\Lambda \neq i^B_\Lambda,\ i^C_\Sigma \neq i^B_\Sigma \\ p(ab)p(c) & i^A_\Sigma \neq i^C_\Sigma,\ i^A_\Lambda = i^B_\Lambda,\ i^C_\Sigma \neq i^B_\Sigma \\ p(a)p(bc) & i^A_\Sigma \neq i^C_\Sigma,\ i^A_\Lambda \neq i^B_\Lambda,\ i^C_\Sigma = i^B_\Sigma \\ p(a)p(b)p(c) & i^A_\Sigma \neq i^C_\Sigma,\ i^A_\Lambda \neq i^B_\Lambda,\ i^C_\Sigma \neq i^B_\Sigma \\ \text{unknown} & \text{otherwise} \end{cases}. \quad (3.36)$$

The other combinations of indices cannot be connected to the distribution $p(abc)$.

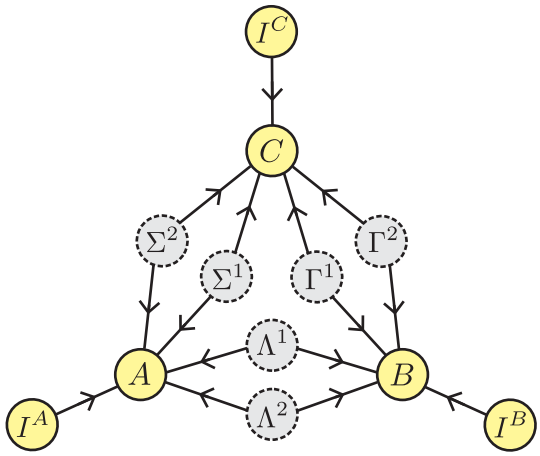


FIGURE 3.9: An inflation of the triangle causal scenario with 2 copies of Λ, Σ and Γ. The inputs I^A, I^B and I^C describe which copy each measurement device uses.

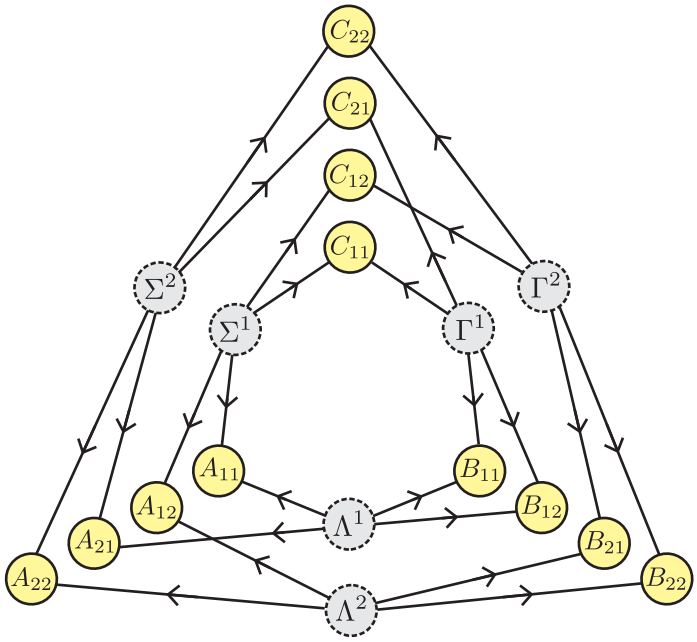


FIGURE 3.10: The fully unpacked scenario of the inflation of Figure 3.9, known as the web inflation [WSF19].

3.4.1 Classical preparations

If the preparations are classical, we may unpack the distribution q over all the inputs, whilst "removing" connections between unpacked variables to reflect independence constraints, as in the 3-party chain scenario inflation of Figure 3.3. The fully unpacked distribution marginalizes to $q(abc|(i^A_\Sigma, i^A_\Lambda), (i^B_\Lambda, i^B_\Gamma), (i^C_\Gamma, i^C_\Sigma))$ as follows:

$$q(abc|(i^A_\Sigma, i^A_\Lambda), (i^B_\Lambda, i^B_\Gamma), (i^C_\Gamma, i^C_\Sigma)) = \sum_{\{a_{ii'}\}_{ii'\setminus a_{(i^A_\Sigma, i^A_\Lambda)}}} \sum_{\{b_{ii'}\}_{ii'\setminus b_{(i^B_\Lambda, i^B_\Gamma)}}} \sum_{\{c_{ii'}\}_{ii'\setminus c_{(i^C_\Gamma, i^C_\Sigma)}}} q(a_{11}\, a_{12}\, a_{21}\, a_{22}\, b_{11}\, b_{12}\, b_{21}\, b_{22}\, c_{11}\, c_{12}\, c_{21}\, c_{22}). \quad (3.37)$$

The unpacked distribution must be compatible with the unpacked causal scenario of Figure 3.10, which is also known as the "web inflation" [WSF19].

In the general case in which there are $(n_\Lambda, n_\Sigma, n_\Gamma)$ replicas of each prepared system, the group determining the inflation symmetries is generated by the combination of permutations for each set of replicated states:

$$\mathcal{G} = \langle \mathrm{Perm}(\{1, \ldots, n_\Lambda\}) \times \mathrm{Perm}(\{1, \ldots, n_\Sigma\}) \times \mathrm{Perm}(\{1, \ldots, n_\Gamma\}) \rangle,$$

where $\mathcal{G} = \langle S \rangle$ denotes that $\mathcal{G}$ is the smallest group containing all the elements of S. The group acts on the conditioned distribution through a simultaneous relabelling of the indices $(i^A_\Lambda, i^B_\Lambda)$, (i^A_Σ, i^C_Σ) and (i^B_Γ, i^C_Γ) respectively. For example, swapping copy number 1 and 2 of Σ changes the corresponding indices as follows:

$$(i^A_\Sigma = 1, i^C_\Sigma = 1) \to (i^A_\Sigma = 2, i^C_\Sigma = 2), \quad (3.38)$$

$$(i^A_\Sigma = 2, i^C_\Sigma = 1) \to (i^A_\Sigma = 1, i^C_\Sigma = 2). \quad (3.39)$$

Given that the conditioned distribution is invariant under such relabellings, we can replace it with an average over the group action:

$$q(abc|(i^A_\Sigma, i^A_\Lambda), (i^B_\Lambda, i^B_\Gamma), (i^C_\Gamma, i^C_\Sigma)) \to \\ \to \frac{1}{|\mathcal{G}|} \sum_{g\in\mathcal{G}} q(abc|(\pi^\Sigma_g(i^A_\Sigma), \pi^\Lambda_g(i^A_\Lambda)), (\pi^\Lambda_g(i^B_\Lambda), \pi^\Gamma_g(i^B_\Gamma)), (\pi^\Gamma_g(i^C_\Gamma), \pi^\Sigma_g(i^C_\Sigma))), \tag{3.40}$$

where for a group element g we have associated permutations π^Λ_g, π^Σ_g, π^Γ_g for each corresponding tuples of indices. The average over the group action can be lifted to the fully unpacked distribution as in the 3-party chain scenario (compare with Eq. (3.12)):

$$\begin{aligned} q(a_{11}\, a_{12}\, a_{21}\, a_{22}\, b_{11}\, b_{12}\, b_{21}\, b_{22}\, c_{11}\, c_{12}\, c_{21}\, c_{22}) = \\ = q(\ldots, a_{(i^A_\Sigma, i^A_\Lambda)}, \ldots, b_{(i^B_\Lambda, i^B_\Gamma)}, \ldots, c_{(i^C_\Gamma, i^C_\Sigma)}, \ldots) \to \\ \to \frac{1}{|\mathcal{G}|} \sum_{g\in\mathcal{G}} q(\ldots, a_{(\pi^\Sigma_g(i^A_\Sigma), \pi^\Lambda_g(i^A_\Lambda))}, \ldots, b_{(\pi^\Lambda_g(i^B_\Lambda), \pi^\Gamma_g(i^B_\Gamma))}, \ldots, \\ \ldots, c_{(\pi^\Gamma_g(i^C_\Gamma), \pi^\Sigma_g(i^C_\Sigma))}, \ldots). \end{aligned} \tag{3.41}$$

This symmetrization is incompatible with other non-convex causal constraints, such as:

$$q(a_{11}\, a_{12}\, c_{21}\, c_{22}) = q(a_{11}\, a_{12})\, q(c_{21}\, c_{22}), \tag{3.42}$$
$$q(b_{11}\ b_{21}\, c_{12}\, c_{22}) = q(b_{11}\ b_{21})\, q(c_{11}\, c_{12}), \tag{3.43}$$
$$q(a_{12}\, a_{22}\, b_{11}\, b_{12}) = q(a_{11}\, a_{12})\, q(b_{11}\, b_{12}), \tag{3.44}$$

however such constraints are not relevant when considering the convexified set $\mathrm{Conv}(\mathrm{Inf}^{n_\Lambda=2, n_\Sigma=2, n_\Gamma=2}[\mathcal{L}^{\mathrm{triangle}}])$. One now can establish a LP such as that of the 3-party chain scenario in expression (3.15) whose infeasibility proves that the distribution $p(abc)$ is incompatible with the triangle scenario with classical sources of correlations.

3.4.2 Quantum preparations

If the preparations Λ, Σ, Γ are quantum processes, we may characterize the set of correlations $\mathrm{Conv}(\mathrm{Inf}^{n_\Lambda, n_\Sigma, n_\Gamma}[\mathcal{Q}^{\mathrm{triangle}}])$ with a modification to the

NPA hierarchy analogous to that used for the 3-party chain scenario. This involves constructing an alphabet which incorporates also the inputs I^A, I^B and I^C:

$$\mathcal{O} = \{\mathbb{1}\} \cup \bigcup_{i^A \in N_{I^A}} \{A^{i^A}_{a|x}\}_{xa} \cup \bigcup_{i^B \in N_{I^B}} \{B^{i^B}_{b|y}\}_{yb} \cup \bigcup_{i^C \in N_{I^C}} \{C^{i^C}_{c|z}\}_{yb},$$

where i^A, i^B and i^C represent tuples of indices that correspond to each subsystem accessible by the measurement device in the original setup.

Consider the measurement operators of A. They will commute for the same values of input x and outcome a whenever their inflation indices have no overlap:

$$[A^{(i^A_\Sigma, i^A_\Lambda)}_{a|x}, A^{(j^A_\Sigma, j^A_\Lambda)}_{a|x}] = 0 \qquad \text{if} \qquad i^A_\Sigma \neq j^A_\Sigma \quad i^A_\Lambda \neq j^A_\Lambda. \tag{3.45}$$

More concisely, we write $i^A \cap j^A = \varnothing$ to denote that the tuples i^A and j^A have different values for all entries. Similar commutation rules hold for B and C.

We can now construct a generating set $\mathcal{S}$ of monomials in $\mathcal{O}$ which determines a moment matrix. The distribution $q(abc|(i^A_\Sigma, i^A_\Lambda), (i^B_\Lambda, i^B_\Gamma), (i^C_\Gamma, i^C_\Sigma))$ connects with the following entries of the moment matrix:

$$q(abc|(i^A_\Sigma, i^A_\Lambda), (i^B_\Lambda, i^B_\Gamma), (i^C_\Gamma, i^C_\Sigma)) = \langle A^{(i^A_\Sigma, i^A_\Lambda)}_{a|x} B^{(i^B_\Lambda, i^B_\Gamma)}_{b|y} C^{(i^C_\Gamma, i^C_\Sigma)}_{c|z} \rangle. \tag{3.46}$$

The group acts on an arbitrary moment by jointly relabelling the appropriate indices according to the symmetry group element g:

$$\begin{aligned} &\langle A^{i^A_1}_{\alpha_1} \cdots A^{i^A_n}_{\alpha_n} B^{i^B_1}_{\beta_1} \cdots B^{i^B_m}_{\beta_m} C^{i^C_1}_{\gamma_1} \cdots C^{i^C_o}_{\gamma_o} \rangle \to \\ &\qquad \to \langle A^{\pi_g(i^A_1)}_{\alpha_1} \cdots A^{\pi_g(i^A_n)}_{\alpha_n} B^{\pi_g(i^B_1)}_{\beta_1} \cdots B^{\pi_g(i^B_m)}_{\beta_m} C^{\pi_g(i^C_1)}_{\gamma_1} \cdots C^{\pi_g(i^C_o)}_{\gamma_o} \rangle, \end{aligned} \tag{3.47}$$

where n, m, o are arbitrary integers. The values of these integers depend on the maximum degree of the monomials that appear in the NPA relaxation, which in turn depend on the choice of $\mathcal{S}$. We simplified the index description for readability, where $\{\alpha_i\}_i$, $\{\beta_i\}_i$, $\{\gamma_i\}_i$ describe input and output tuples, and $\{i^A_\alpha\}_\alpha$, $\{i^B_\alpha\}_\alpha$, $\{i^C_\alpha\}_\alpha$ describe the tuples of copy indices. The action of the group is not fully specified at the level of the individual entries of the tuples as it similar as in previous expressions such as (3.40) and (3.42).

Through an analogous reasoning to the one of the 3-party chain scenario, given a solution to the SDP of the NPA relaxation, we may build other solutions to the SDP through the symmetry group action. As such, we may without loss of generality assume that all moments are invariant under the action of the inflation symmetry group. We do not explicitly write the SDP as it is very similar to other formulations presented in this thesis.

3.4.3 No-signaling preparations

As mentioned in the presentation of the triangle scenario (Section 3.3) the axiom of causality seems to impose no constraints on the observed distribution $p(abc)$. However, through inflation, one can build scenarios where causality restricts the distribution on the inflated graph non-trivially. This is possible in the triangle scenario, whereas it is not possible in the 3-party chain scenario (recall Figure 3.6).

To do so, we may analyse the combinations of indices which fix a non-fanout wiring, i.e., where a prepared system is not distributed to more measurement devices than in the original scenario. Such combinations of indices can be unpacked, as the wiring protocol itself ensures that the distribution is jointly measurable over the different wiring inputs I^A, I^B, I^C. Let us consider an example. In the level 2 inflation of the triangle scenario, one compatible unpacking is the following:

$$q(a_{11}\, a_{22}\, b_{11}\, b_{22}\, c_{12}\, c_{21}), \tag{3.48}$$

together with all its index relabellings per the inflation symmetries. This unpacking corresponds to a hexagonal inflation, as shown in Figure 3.11. This unpacking connects with the conditioned distribution as follows:

$$q(abc|(i^A_\Sigma{=}1, i^A_\Lambda{=}1), (i^B_\Lambda{=}1, i^B_\Gamma{=}1), (i^C_\Gamma{=}1, i^C_\Sigma{=}2)) = \\ = \sum_{a_{22}b_{22}c_{21}} q(a_{11}\, a_{22}\, b_{11}\, b_{22}\, c_{12}\, c_{21}), \tag{3.49}$$

$$q(abc|(i^A_\Sigma{=}1, i^A_\Lambda{=}1), (i^B_\Lambda{=}2, i^B_\Gamma{=}2), (i^C_\Gamma{=}1, i^C_\Sigma{=}2)) = \\ = \sum_{a_{22}b_{11}c_{21}} q(a_{11}\, a_{22}\, b_{11}\, b_{22}\, c_{12}\, c_{21}), \tag{3.50}$$

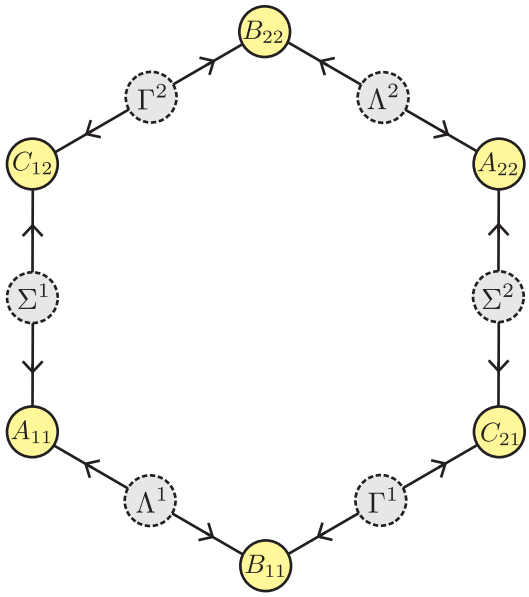


FIGURE 3.11: One possible non-fanout unpacking of the inflation of the triangle scenario of Figure 3.9, known as the hexagon inflation [Gis+20b]. This inflation contains also other inflated graphs such as the "cut inflation" of Figure 3.12.

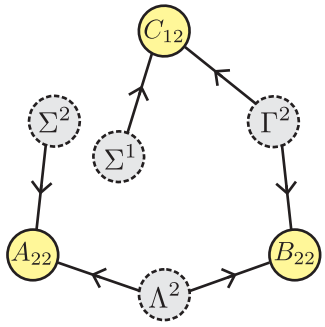


FIGURE 3.12: One possible non-fanout unpacking of inflation of the triangle scenario of Figure 3.9, known as the cut inflation [WSF19]. This inflated scenario is contained in the hexagon inflation of Figure 3.11.

$$q(abc|(i^A_\Sigma{=}2, i^A_\Lambda{=}2), (i^B_\Lambda{=}2, i^B_\Gamma{=}2), (i^C_\Gamma{=}1, i^C_\Sigma{=}2)) = \text{etc.}$$

We do not write the rest of the input combinations as the pattern is clear. As done so far in this dissertation, we may lift the symmetries at the level of the unpacked distribution by averaging over the group action. This enables us to assume without loss of generality that unpacked distributions such as $q(a_{11}, a_{22}, b_{11}, b_{22}, c_{12}, c_{21})$ are invariant under the action of the group provided we do not consider non-convex constraints, that is, provided we work with the convexified set $\text{Conv}(\text{Inf}^{n_\Lambda=2, n_\Sigma=2, n_\Gamma=2}[\mathcal{NS}^{\text{triangle}}])$. One can then write a linear program analogous to the one of classical preparations (recall LP (3.15)) which looks for the existence of an unpacked distribution that marginalizes to the distribution conditioned on compatible inputs.

It is important to remark that not all combinations of inputs in the conditioned distribution $q(abc|(i^A_\Sigma, i^A_\Lambda), (i^B_\Lambda, i^B_\Gamma), (i^C_\Gamma, i^C_\Sigma))$ are compatible. In non-fanout inflation, the inputs I^A, I^B and I^C cannot be freely chosen locally, as there is a global consistency condition, namely, that preparations cannot be fanout. Combinations of inputs which fanout preparations are not merely unknown variables, rather they are ill-defined; as such, they are not present in the linear programming relaxation. One may then interpret I^A, I^B and I^C not as locally independent choices for A, B and C, but rather as meta-parameters which determine the topology of the experiment, which may be determined before the run of the experiment in a correlated fashion (to satisfy the global consistency condition).

Lastly, observe that the hexagon inflation of Figure 3.11 contains other inflations such as the cut inflation of Figure 3.12. The hexagon inflation itself is sub-contained in the web inflation for classical preparations (see Figure 3.10). As such, when considering all possible compatible combinations of inputs I^A, I^B, I^C, one simultaneously tests for all possible inflation scenarios that are achievable with $(n_\Lambda, n_\Sigma, n_\Gamma)$ copies of each preparation. This observation is merely to clarify that the thought process when using the inflation method need not be "start by drawing an inflated scenario", but rather, choose the number of copies of each preparation, and then consider unpackings for all possible compatible wirings. In doing so, one automatically considers all possible inflation scenarios for a given inflation level $(n_\Lambda, n_\Sigma, n_\Gamma)$. This is the approach we take in the Python package we developed which implements the inflation method, which we explain in Chapter 6.

Remark on symmetries and non-convex constraints. In the current presentation of the inflation method, we have opted for a unifying approach in which the marginal distribution, for example, $q(abc|(i^A_\Sigma, i^A_\Lambda), (i^B_\Lambda, i^B_\Gamma), (i^C_\Gamma, i^C_\Sigma))$, is the fundamental object satisfying the inflation symmetries. The lifting of the symmetries to the unpacked distribution under different assumptions on the nature of the preparations can be done without loss of generality as long as non-convex constraints are not considered. We are not suggesting that non-convex constraints and symmetries cannot be addressed together; rather, we note a potential conflict between them in this unified framework. Nevertheless, investigating the effects of non-convex constraints, either by considering the complete quadratic program tied to an inflation or by softening the non-convex constraints using polynomial optimization techniques, and their compatibility with symmetries, presents an interesting research direction.

In this chapter we presented a cohesive description of the inflation technique where the fundamental inflation is that which adds extra inputs, and the other types of inflation (fanout, non-fanout and quantum) are derived depending on the assumptions on the nature of the preparations. This unified description may pave the way for a proof of convergence of the non-fanout and quantum inflation hierarchies, which currently is absent.

Chapter 4

Single photon nonlocality in quantum networks

This chapter is based on Ref. [Abi+22]:
Paolo Abiuso, Tamás Kriváchy, **Emanuel-Cristian Boghiu**, Marc-Olivier Renou, Alejandro Pozas-Kerstjens, and Antonio Acín, *"Single-photon nonlocality in quantum networks"*, Phys. Rev. Research 4, L012041—Published 29 March 2022

4.1 Background

As shown in the preliminaries, classical common cause models cannot account for all the predictions of quantum theory. Bell nonlocality manifests in so-called Bell tests, defined by the correlations obtained when performing appropriate local measurements on a well-chosen entangled state (cf. Section 2.3.2). Bell tests have been performed in many different systems, from massive particles [Hen+15] to photons [Giu+15b; Sha+15], and using many different degrees of freedom, such as electronic levels, polarization, orbital angular momentum or time bins. In most of these realizations the relevant degrees of freedom used to encode the entanglement are transmitted to each distant observer by a physical carrier, such as, for instance, a photon.

In this chapter we are interested in whether single-particle quantum states can display nonlocal correlations with no classical analogue. In particular, we consider the question in the context of single-photon entanglement, that

is, the state

$$|\psi^+\rangle_{AB} = \frac{1}{\sqrt{2}}(|01\rangle_{AB} + |10\rangle_{AB}), \tag{4.1}$$

obtained when sending a single photon into a balanced beamsplitter. Here $|01\rangle_{AB}$ (resp. $|10\rangle_{AB}$) represents the situation in which the photon is sent to the right party B (resp. the left party A). The resulting state therefore consists of only one photon and entanglement is encoded in the two optical spatial modes.

Is the state (4.1) nonlocal? A state is understood to be nonlocal if it can lead to Bell nonlocal correlations. This question has been intensively debated in the quantum foundations and quantum optics community, e.g. [TWC91; Har94; Ger96; Vai95; AV00; Hes+04; D'A+06; BCB13; Mor+13a; Don+14; Das+21; YS92a; YS92b; GPS21]. In principle, a positive answer is provided by the following simple argument [Ger96; Vai95; AV00]: the two optical modes can be transferred to the population of two energy levels of two distant massive particles. Single-photon entanglement is therefore mapped into two-particle entanglement and a Bell test can now be implemented. The question is subtler when considering only optical means. To obtain a nonlocal correlation, the two observers need to use local *active measurements* involving local oscillators creating extra local photons [TWC91; Har94; BCB13; Das+21]: without such active measurements, measuring the information content of the state (4.1) allows the observers to deduce if they received the photon sent by the source, destroying the indeterminacy in the photon path, that is, the coherences in (4.1). Then, the correlations become classically simulable. One might be tempted to conclude that the observation of nonlocal effects in the single-photon entangled state by passive optical means, that is, phase shifters, beamsplitters and photodetectors, is impossible.

The main result of our work is to show that this is not the case; one can indeed reveal the nonlocality of state (4.1) with only *passive measurements*. To do so, we go beyond standard Bell tests and consider setups defined by more sophisticated causal scenarios, more particularly, network scenarios. Network scenarios are causal structures involving several independent sources, each being distributed to a subset of the parties involved in the scenario, according to a structure defined by a network [Tav+22]. It understood that these networks offer new possibilities to design quantum experiments with no classical analogue [BGP10; Fri12; Fri16; Cha+17; VH+19; Ren+19a]. We show how three copies of single-photon entangled states placed in a

triangle causal network (cf. Section 3.3 in the Preliminaries; Figure 4.1 contains a depiction of the experimental proposal) can exhibit non-classical correlations. The main idea is to exploit the topology of the network to reintroduce indeterminacy in the photon path, necessary to exploit the coherences of these states. Interestingly, the obtained setup is not only passive in terms of the implemented measurements, but also because it does not require any active choice of measurements. Our setup contains no classical inputs, and observers perform a single measurement on their received shares. These characteristics make the proposal, arguably, the simplest experimental demonstration of the nonlocality of the single-photon entangled state, as well as the first experimental proposal for genuine network nonlocality [Ren+19a].

Beyond the fundamental motivation, our results may also be relevant from an applied point of view. Correlations with no classical analogue are the main resource for device-independent applications. For instance, the security of DI protocols for quantum random number generation [Col07; Pir+10] and quantum key distribution [Ac07] is based on the observation of Bell inequality violations. For that, the simplest way of producing entangled states is through Spontaneous Parametric Down Conversion (SPDC). Entanglement can be encoded on different degrees of freedom of the resulting two photons. However, the state produced by SPDC is a mixture of the desired entangled state and vacuum [CV+15]. In fact, a heralded preparation of a two-photon maximally entangled state is challenging [SB03]. In turn, single-photon entanglement can be easily prepared in a heralded way: an arbitrarily good approximation to it can be obtained when detecting photons in one of the two modes resulting from the SPDC process and sending the non-measured mode into a balanced beamsplitter (see Appendix B.3). Moreover, this form of entanglement does not require the control of any other light degrees of freedom, such as, for example, polarization or orbital angular momentum. Therefore, the design of simple setups to generate correlations with no classical analogue from this state opens new avenues for the implementation of DI protocols.

4.2 The triangle network

The considered Bell-type experiment consists of a triangle causal network (cf. Section 3.3 in the Preliminaries) where three measurement devices A,

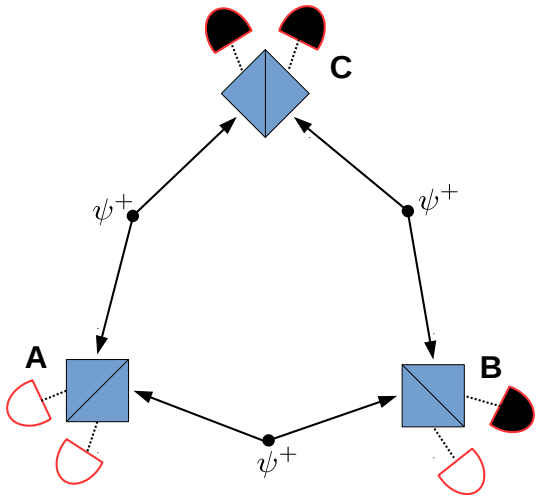


FIGURE 4.1: Schematics of the proposed quantum optical experiment. A, B and C share single-photon entangled states $|\psi^+\rangle = (|01\rangle + |10\rangle)/\sqrt{2}$ prepared by the sources. Each party receives two optical modes that are mixed on a beamsplitter, the resulting output modes being measured by photodetectors. In the specific experimental instance depicted here, A does not detect any photon, B has one detector firing, and C has both detectors firing.

B and C, receive states prepared by three sources, see Figure 4.1. These states are measured producing outcomes a, b and c with probability $p(abc)$. Recall that if the observed correlations $p(abc)$ are compatible with classical preparations, they must have a decomposition of the following form:

$$p(abc) = \int_{\gamma,\sigma,\lambda\in[0,1]^3} q(a|\sigma\lambda)q(b|\lambda\gamma)q(c|\gamma\sigma)\,\mathrm{d}\gamma\mathrm{d}\sigma\mathrm{d}\lambda. \tag{4.2}$$

The causal model therefore consists of classical variables γ, σ and λ distributed by the preparations and local deterministic response functions producing the measurement outcomes. In analogy with standard Bell tests, we define probability distribution $p(abc)$ that can be written as Eq. (4.2) as causally classical or, simpler, *local*.

A quantum description of the experiment compatible with the causal network replaces the random variables by quantum states ρ_γ, ρ_σ and ρ_λ and the local response functions by quantum measurements. Therefore, quantum probabilities compatible with the triangle network have the form:

$$p(abc) = \mathrm{Tr}\left[(\rho_\gamma \otimes \rho_\sigma \otimes \rho_\lambda)(A_a \otimes B_b \otimes C_c)\right], \tag{4.3}$$

where A_a is a POVM, and similarly for B and C. We slightly abuse the notation in Eq. (4.3) by not specifying the tensor products and different Hilbert spaces in which the different operators act. We say that a quantum experiment, defined by states and measurements producing the outcome distribution $p(abc)$ according to Eq. (4.3), is *nonlocal* whenever this distribution cannot be described by a classical model (4.2). Our goal in what follows is to provide a nonlocal quantum experiment in the triangle network using only single-photon entangled states, beamsplitters and photodetectors.

The essential idea of the experimental proposal is depicted in Figure 4.1: three parties A, B, C share, for each pair AB, BC, CA, the single photon entangled state $|\psi^+\rangle$, see Eq. (4.1). The initial state is thus:

$$|\psi^+\rangle_{A_2B_1} \otimes |\psi^+\rangle_{B_2C_1} \otimes |\psi^+\rangle_{C_2A_1} \equiv |\Psi^+\rangle_{A_1A_2B_1B_2C_1C_2}\,. \tag{4.4}$$

Each party then receives its two optical inputs on modes X_1X_2 ($X = A, B, C$) and mixes them with a beamsplitter, which induces a unitary transformation $\mathcal{B}_{X_1X_2}(t, \phi)$ parametrized by its transmissivity t and phase ϕ. All parties use the same value for t, and the phases are all null for simplicity in the following (cf. Appendix B.1).

After passing through the beamsplitters, the photons end up in photodetectors. For each mode X_i, the operators describing a perfectly efficient photodetection correspond to the projectors onto the vacuum state $D^{\square}_{X_i} = |0\rangle\langle 0|_{X_i}$ (detector off) and the projector on its orthogonal complement $D^{\blacksquare}_{X_i} = \mathbb{1}_{X_i} - |0\rangle\langle 0|_{X_i}$ (detector firing). Indeed, we assume that the detectors do not resolve the number of photons but only their presence. The measurement obtained by mixing two modes with the beamsplitter and the ideal photodetectors can be accordingly expressed as a POVM for each party

(here $\mathcal{B}_{X_1X_2} = \mathcal{B}_{X_1X_2}(t,0)$):

$$\begin{aligned}
{\Pi_t^{(0)}}_{X_1X_2} &= \mathcal{B}^\dagger_{X_1X_2}(D^{\square}_{X_1} \otimes D^{\square}_{X_2})\mathcal{B}_{X_1X_2}, \\
{\Pi_t^{(L)}}_{X_1X_2} &= \mathcal{B}^\dagger_{X_1X_2}(D^{\blacksquare}_{X_1} \otimes D^{\square}_{X_2})\mathcal{B}_{X_1X_2}, \\
{\Pi_t^{(R)}}_{X_1X_2} &= \mathcal{B}^\dagger_{X_1X_2}(D^{\square}_{X_1} \otimes D^{\blacksquare}_{X_2})\mathcal{B}_{X_1X_2}, \\
{\Pi_t^{(2)}}_{X_1X_2} &= \mathcal{B}^\dagger_{X_1X_2}(D^{\blacksquare}_{X_1} \otimes D^{\blacksquare}_{X_2})\mathcal{B}_{X_1X_2},
\end{aligned} \tag{4.5}$$

where the measurement labels stand respectively for no photon counts (0), a count in the left detector (L), a count in the right detector (R), or counts in both detectors (2). The crucial point is that when $t \neq 0$, the L and R measurements actually detect superpositions of photons in the incoming modes (see details in Appendix B.1).

The quantum experiment described here results in the output distribution

$$\begin{aligned}
p_t(abc) &= \mathrm{Tr}[|\Psi^+\rangle\langle\Psi^+| \, (\Pi_t^{(a)} \otimes \Pi_t^{(b)} \otimes \Pi_t^{(c)})] \\
a, b, c &\in \{0, L, R, 2\}
\end{aligned} \tag{4.6}$$

which depends on the transmissivity t of the beamsplitters used by the parties and whose exact expression can be found in the Appendix B.1.

4.3 Witnessing single-photon nonlocality

The first main result of this work is that

The distribution p_t obtained from the experiment described in Figure 4.1 (cf. previous section), is nonlocal (at least) for values of the beamsplitter transmissivity in the intervals $t \in (0, 0.215)$ and $t \in (0.785, 1)$.

We give in the following a sketch of the proof, which is analytical and detailed in Appendix B.2. As expressed in Eq. (4.2), any local model is specified by deterministic triangle-local response functions that map all the points of the cube $[0,1]^3$ to the observed outputs:

$$\{\gamma, \sigma, \lambda\} \to \{a(\sigma, \lambda), b(\lambda, \gamma), c(\gamma, \sigma)\}. \tag{4.7}$$

We were able to identify strict constraints that need to be satisfied by all possible classical causal models simulating the considered experimental

output $p_t(abc)$ in the triangle network. In particular, we exploited the cyclic symmetry and null components of the distribution. For example, all outputs of the form (here χ represents any of L or R) $\{(000), (00\chi), (2\chi\chi), (22\chi)\}$, or any of their permutations, have zero probability, due to the fact that there are initially 3 photons in the network, of which at most 2 can end up in the same photodetector. That is, in each run of the experiment the total number of clicks in the detectors must be 2 or 3. By taking all the relevant properties of p_t into account, one can identify constraints that need to be satisfied by any classical strategy, specified by the response functions (4.7), aiming at reproducing p_t. In fact, while the exact form of the response functions remains in general unknown, some of its marginals can be expressed in terms of the output p_t. These relevant marginals are nothing other than linear constraints on the response functions, parametrized by t. Together with standard normalization and positivity constraints, these define a Linear Program. The feasibility of such Linear Program is, by definition, necessary for the existence of such local response functions. Therefore, when infeasible, no local model exists to simulate our experiment proposal. Results show that the Linear Program is infeasible for $t \in (0.785, 1)$ and $t \in (0, 0.215)$, proving the claims of this section.

The techniques we used are similar to those introduced in [Ren+19a] and generalized in [RB22b]. However, their findings cannot be applied directly to our scenario. The reason behind this is that the works [Ren+19a; RB22b] are based on a token-counting approach to some physical "tokens" that are: *i*) generated from the sources, *ii*) distributed to the parties in a coherent superposition of different ways, and *iii*) counted at the output. In our experiment the physical tokens are the photons, which however can be miscounted at the output, as more than one could enter in the same photodetector. For these reasons, in the proof of Appendix B.2 we had to extend these techniques so that they could be applied to our setup. As part of the proof, we showed that our distribution is nonlocal if and only if the distribution proposed in [Ren+19a], which we dub p'_t, is nonlocal as well. In Ref. [PKGR23b], they prove nonlocality of p'_t for discrete points in the range $t \in (0.5, 0.785)$ as well. Nonlocality of p'_t in such interval has been conjectured already [Kri+20]. Given the above mentioned equivalence between the nonlocality of p_t and p'_t proven in this work, this would imply that the proposed ideal experiment is nonlocal for all transmissivities except $t \in \{0.0, 0.215, 0.5, 0.785, 1.0\}$, which are known to have local models (cf. [Ren+19a]).

4.4 Noise tolerance and machine learning analysis

After proving the nonlocality of the outputs of the ideal noiseless experiment, we analyzed the robustness of our results against typical noise errors, by modelling imperfections which occur in experimental realizations of the optical network presented in Figure 4.1. Therefore, the resulting output distribution, $p_t^{Q,T,\nu}(abc)$ depends on additional noise parameters quantifying: the impurity of the generated single-photon entangled state (Q), the transmissivity of the optical channels (T) of the network, and the efficiency of the final photodetectors (ν). It follows that:

$$p_t^{Q=0,T=1,\nu=1}(abc) \equiv p_t(abc), \tag{4.8}$$

that is, with no impurity, and perfect transmission and detection, we recover the idealized experiment. The details of the modelling employed are deferred to the Appendix B.3.

Inevitably, part of the key properties and symmetries of $p_t(abc)$ disappear as soon as noise is introduced in the network. This makes the analytic approach unworkable in this case. Consequently, in order to estimate the tolerance to the noises introduced above, we resorted to a technique recently introduced in [Kri+20]: there, a feed-forward neural network is shaped with the same topology of the causal network under study, and it is then asked to reproduce the target distribution $p_t^{(Q,T,\nu)}$. Each output of the neural network is thus *literally* an instance of a classical model (which can be therefore described by Eq. (4.2) in our case) trying to reproduce $p_t^{(Q,T,\nu)}$. For a fixed target distribution, the neural network is trained by minimizing the Euclidean distance from the neural network's local model to the target. When the target distribution is inside the local set, a sufficiently large neural network should be capable of learning it. Instead, a large distance between the machine's best guess and the target is taken as an indication of nonlocality. What it means to be "large" enough can be somewhat arbitrary, since some nonlocal behaviors are extremely close to the local set (as is the case here), and additionally the neural network's model is not guaranteed to converge to the optimal solution as it can get stuck in local minima during training. In order to gain deeper insight into the boundary between locality and nonlocality we examine transitions of the learning algorithm's behaviour when adding noise to the target distribution, and retraining the machine

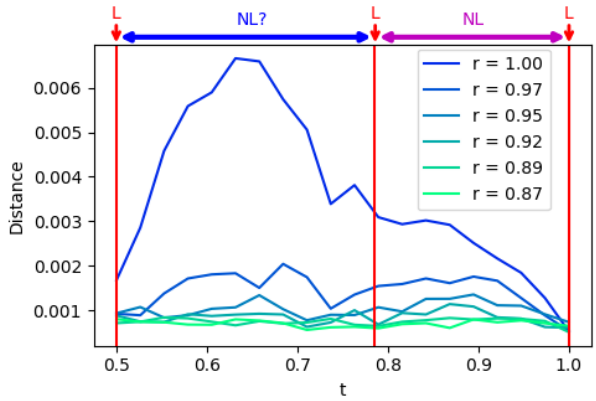


FIGURE 4.2: Euclidean distance of machine learned local models to the target distributions $p_t(abc)$, for various levels of artificial noise on the singlets (4.1) (visibilities r of Werner states $r\,|\psi^+\rangle\langle\psi^+| + (1-r)\mathbb{1}/4$). With red vertical lines we depict the transmissivities t at which analytic local models exist ($t \in \{0.5, 0.785, 1\}$). At the top of the figure a purple line shows the regime where we have proven nonlocality, while the blue line shows the regime where we conjecture nonlocality, based on these numerics and the relation to the distribution in Ref. [Ren+19a], which was studied numerically in Ref. [Kri+20].

independently for each target distribution. The very noisy case is guaranteed to be local and the machine learning results on those give a reference to which we can compare the nonlocal regime. By definition, this technique does not certify nonlocality analytically, but has been shown to be reliable and efficient from the point of view of computational resources [Kri+20].

The results of the analysis are summarized in Figs. 4.2 and 4.3, where we consider only $t \geq 0.5$ because of the symmetry of the experiment when mirroring the beamsplitters $t' = 1 - t$. For the noiseless distribution (perfect visibility $r = 1$ in Figure 4.2), the neural network's best guess is distant from the experimental output, corroborating the analytical proof of nonlocality for $t \in (0.785, 1)$. At the same time the neural network hints at the locality of the output distribution for $t = 0.5$ and $t = 1$, which clearly have local strategies. A local model exists as well for $t \sim 0.785$ (cf. Ref. [Ren+19a],

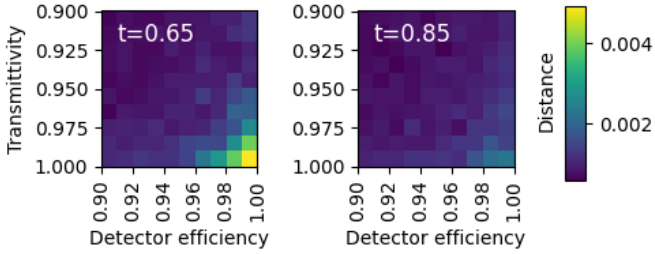


FIGURE 4.3: Euclidean distance of machine learned local models from the noisy distribution $p_t^{(Q,T,\nu)}$ under an experimentally realistic noise model for $t = 0.65$ (left) and $t = 0.85$ (right), with $Q = 0.006875$ for both.

Appendix B.2) where the neural network struggles to get closer; however, note that the distance of 0.003 achieved there is already *very* close to the local set. Moreover, the same machine indicates (seemingly even stronger) nonlocality in the range $t \in (0.5, 0.785)$, in line with the conjecture of Ref. [Kri+20] and the results of Ref. [PKGR23b].

The noise robustness is, however, small. In Figure 4.2 an artificial noise is considered by adding a Werner state visibility to the source (4.1) of ideal experiment ($Q = 0, T = 1, \nu = 1$). The neural network seems to indicate that the points that are "most nonlocal" are $t \sim 0.85$ in the proven region (purple interval in Figure 4.2), and $t \sim 0.65$ in the conjectured region (blue interval). For these two points we tested the tolerance to the physical noises introduced above, see Figure 4.3: choosing $Q \simeq 0,7\%$ (cf. Appendix B.3), the neural network tries to learn $p_t^{(Q,T,\nu)}$ for different values of the transmissivity T and detector efficiency ν. Results show that nonlocality is more robust for $t = 0.65$, where it is lost when $T \lesssim 95\%$ or $\nu \lesssim 95\%$.

All data was obtained by representing each of the three response function $(q(a|\lambda\sigma), q(b|\lambda\gamma), q(c|\gamma\sigma))$ by a multilayer perceptron of depth 4 and width 20 with rectified linear activation functions. For each target distribution we retrained the neural network independently 30 times and kept the smallest distance among those.

4.5 Discussion

In this chapter, we have shown how single-photon entangled states can be used to generate an outcome distribution with no classical analogue in a triangle causal network. The considered setup only requires passive optical elements, namely beam-splitters, phase shifters and photodetectors, and involves a single measurement choice per observer. Our results not only challenge the current understanding of the nonlocal properties of single-photon entanglement, but also open new perspective for the use of this form of entanglement for quantum information applications, as they provide the first proposal of an experimental demonstration of genuine network nonlocality.

We have shown that the nonlocality of such proposal has (small) noise-tolerance to natural noises that can arise in its implementation, through a machine learning analysis. Such approach is however not exact, and it remains an open question to prove nonlocality in the noisy regime by other means, for example, certifying it by inflation techniques [WSF19], which would be crucial for an experimental implementation.

Finally, in Appendix B.4 we show that our main result on the nonlocality of the ideal experimental proposal in the triangle network can be extended to any ring network with $N \geq 3$ parties, although increasing the number of parties does not improve the detectability of nonlocality in the proposed experiment with our current techniques.

Chapter 5

Broadcasting of quantum states

This chapter is based on Ref. [Bog+23]:
Emanuel-Cristian Boghiu, Flavien Hirsch, Pei-Sheng Lin, Marco Túlio Quintino, Joseph Bowles, *"Device-independent and semi-device-independent entanglement certification in broadcast Bell scenarios"*, SciPost Phys. Core 6, 028 (2023)—Published 11 April 2023

5.1 Introduction

In this chapter, we study how Bell nonlocality and other notions of non-classicality such as EPR steering [WJD07; CS16; Uol+20; Jev+15; CV16; Bow+16] relate to quantum entanglement. Like Bell nonlocality, EPR steering is a form of non-classicality exhibited by entangled quantum states, and relates to the fact that a measurement made on one subsystem of an entangled state has the ability to influence or "steer" the distant quantum state of another subsystem. EPR steering is similar to Bell nonlocality but involves more specific assumptions about one of the measurement devices, rendering it easier to demonstrate in experiments [Sau+10; Wit+12].

One of the most basic questions one can ask is which entangled states can manifest these forms of non-classicality. It is known that entanglement alone is not sufficient to observe neither Bell nonlocality nor EPR steering, since some mixed entangled states are known to admit so-called local hidden-variable, or local hidden state models [Wer89; Bar02; Qui+15]. A deeper understanding of this question is not only interesting from a foundational

perspective, but also from a technological perspective, given their connection to quantum information technologies.

An important discovery in this respect was that of activation. The essential idea is the following: some quantum states that show only classical behaviour in simple causal scenarios (i.e., common-cause scenarios) can have their non-classicality activated, or revealed, by subjecting the state to a more complex measurement protocol. This broadens the spectrum of entangled states showing non-classical correlations and rekindles the hope of proving Bell nonlocality or EPR steering of all entangled states. There are a number of different techniques that have been shown to activate quantum states (see [BHC21] for a more detailed discussion). In Refs. [Pop95; Hir+13], it is shown that Bell nonlocality can be activated by applying local filters to the state before a Bell test. This can be seen as a specific case of the more general sequential measurement scenario, where a sequence of time-ordered measurements is made on the local subsystems of the state [Gal+14]. Afterwards, it was shown that activation of Bell nonlocal and EPR-steering is also possible by taking multiple copies of the state, and performing joint measurements on the local subsystems [Pal12; QBH16]. This method appears to be more powerful than the sequential scenario [Hir+16], which is perhaps to be expected given the additional resources and entanglement granted by the multiple copies.

Recently, a new technique based on broadcasting was discovered and shown to lead to Bell nonlocality activation [BHC21] (see also [TRC19] for a prior related work which inspired the definition of broadcast nonlocality). In this scenario (cf. Chapter 5.1.1) one or more of the local subsystems is broadcast to a number of additional measurement devices (see Figure 5.2). The entanglement present in the original state is thus spread between a larger number of systems, and interestingly, this can be used to activate the Bell nonlocality of the original state. The broadcast scenario also appears to be significantly more powerful than the sequential measurement scenario: for instance, for the two-qubit Werner state, broadcasting leads to activation of Bell nonlocality for significantly lower visibilities [BHC21]. This has practical implications, since although stronger examples of activation are known by using many copies, the broadcast scenario requires the manipulation of a single copy of the state per experimental round, does not require joint measurements, and may thus admit a simpler implementation.

In this chapter we build on this initial work, and prove a number of

new results related to nonlocality, device-independent (DI) and semi-DI entanglement certification, which we summarize in the following:

- **Bell nonlocality in broadcast scenarios**—We give two methods to construct Bell inequalities tailored to the broadcast Bell scenario, starting from a Bell inequality in the standard scenario. We also study detector inefficiencies in the broadcast scenario. For the case of the two-qubit maximally entangled state $|\Phi^+\rangle = [|00\rangle + |11\rangle]/\sqrt{2}$, we show how one can demonstrate Bell nonlocality with lower detection inefficiencies than in the standard scenario.

- **Stronger activation through broadcasting**—We prove a stronger notion of activation than previously shown in [BHC21]. More precisely, we show that through broadcasting, it is possible to convert a state with a local hidden-variable (LHV) model for general (POVM) measurements, to a state this exhibits genuinely multipartite nonlocal correlations. This is the most extreme "jump" in Bell nonlocality class that has been demonstrated using a single quantum state. Such a result highlights the extent to which notions of locality in the standard scenario, such as the existence of a LHV model, are unable to capture the strongly nonlocal properties of entangled states.

- **Device-independent entanglement certification**—We investigate device-independent entanglement certification in the broadcast scenario. In Ref. [BHC21] it was shown that broadcasting allows entanglement certification for noise thresholds much lower than previously known. For the case of the isotropic state of two qubits (local unitary equivalent to the two-qubit Werner state [Wer89]),

$$\rho(\alpha) = \alpha|\Phi^+\rangle\langle\Phi^+| + (1-\alpha)\mathbb{1}/4, \tag{5.1}$$

  it was shown that DI entanglement certification is possible for visibilities $\alpha > \frac{1}{2}$. Here, we show that this can in fact be extended to visibilities greater than $\alpha > 0.338$, using a numerical technique based on the NPA hierarchy. Since the state is entangled for $\alpha > \frac{1}{3}$, this is essentially the entire range of entanglement, and we suspect that with more computational power this could be proven for values of α arbitrarily close to $\frac{1}{3}$. This suggests the possibility of designing device-independent protocols with much greater tolerance to noise,

which is highly desirable given the high experimental requirements that hinder device-independent protocols.

- **Broadcast steering**—We extend the definition of EPR steering to the broadcast scenario, and study the phenomenon using the two-qubit isotropic state, showing that broadcast steering is possible for visibilities greater than 0.4945 when broadcasting to two parties and visibility greater than 0.4679 when broadcasting to three parties. This is below the threshold of $\frac{1}{2}$ below which the state has a local hidden state model in the standard scenario [Wer89; WJD07], and is the first example of the activation of steering using a single copy of this state.

5.1.1 Bell nonlocality in broadcast scenarios

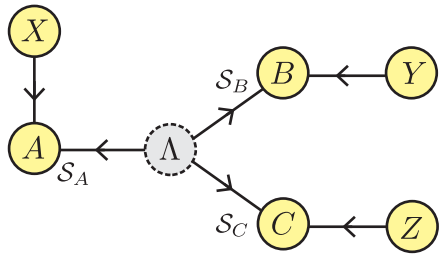
In the "broadcast" scenario, introduced in Ref. [BHC21], a common cause Λ, instead of outputting a three-partite system sent to A, B and C, as in Figure 5.1a, outputs a bipartite system which undergoes a further transformation Σ, before undergoing measurements B and C, as shown in Figure 5.1. Broadcasting reflects the assumption that the final three-partite system is generated from an initial bipartite system, which is useful for investigating properties of bipartite systems.

The different sets of correlations in the broadcast scenario depend on the assumptions made on the processes Λ, Σ. If Λ, Σ are both classical with unbounded cardinality, then Σ may copy Λ and forward it to B, C, therefore the distributions achievable in the all-classical broadcast scenario are indistinguishable from those achievable through a classical three-way common cause, denoted by $\mathcal{L}^{ABC|XYZ}$. The three-way common cause scenario is depicted in Figure 5.1a. If both processes are quantum instruments, the set of quantum correlations in the broadcast scenario is given by

$$\begin{aligned}\mathcal{Q}^{A,(B,C)|X,(Y,Z)} = \{p(abc|xyz)\,|\\ p(abc|xyz) = \mathrm{tr}(\mathbb{1}_A \otimes \Sigma_{B_0\to BC})(\rho_\Lambda)(A_{a|x}\otimes B_{b|y}\otimes C_{c|z})\}\end{aligned} \tag{5.2}$$

where we use parenthesis to denote the broadcast parties, and where $\Sigma_{B_0\to BC}$ is a quantum channel (a quantum instrument with a single classical outcome) and the rest of the operators have a similar meaning as in the quantum common cause scenario of Section 2.3. Examples of such channels are approximate cloning [Wer98; Sca+05] or broadcasting of quantum information

(A) Common-cause scenario



(B) Broadcast scenario

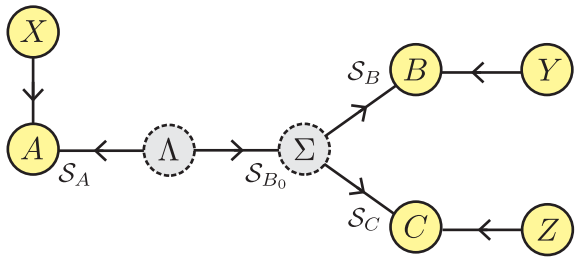


FIGURE 5.1: (5.1a) A three-way common cause scenario with the process Λ outputting a three-partite system that is shared with A, B and C. The local, quantum and no-signaling sets are denoted with superscripts, $\mathcal{L}^{ABC|XYZ}$, $\mathcal{Q}^{ABC|XYZ}$ and $\mathcal{NS}^{ABC|XYZ}$, to distinguish them from the analogous sets in the standard Bell scenario. (5.1b) The broadcast scenario, which differs from Bell's scenario as one of the output systems of the common cause Λ is submitted to another latent process Σ before being transmitted to B and C. This represents the assumption that the tripartite joint state shared between A, B and C is built from an original bipartite state prepared by Λ.

as in Ref. [YHD11], although the channel need not be of this form. If the dimension of the Hilbert spaces is unbounded, the set of quantum correlations $\mathcal{Q}^{A,(B,C)|X,(Y,Z)}$ is the same as the one achievable through a quantum three-way common cause, $\mathcal{Q}^{ABC|XYZ}$. To see this, given any three-partite quantum state ρ with support on $\mathcal{H}_A \otimes \mathcal{H}_B \otimes \mathcal{H}_C$, one may coarse grain the Hilbert spaces, $\mathcal{H}_{B_0} = \mathcal{H}_B \otimes \mathcal{H}_C$, and treat ρ as a bipartite system on $\mathcal{H}_A \otimes \mathcal{H}_{B_0}$. Then the action of Σ is simply that of physically separating $\mathcal{H}_B$ and $\mathcal{H}_C$. If Λ, Σ are both no-signaling resources, then the set of broadcast-no-signaling correlations is the same as that of a three-way no-signaling common cause, $\mathcal{NS}^{A,(B,C)|X,(Y,Z)} \equiv \mathcal{NS}^{ABC|XYZ}$.

It is more interesting to consider a hybrid scenario, where Λ is classical process and Σ is no-signaling. We call the set of correlations achievable with such processes the broadcast-local set, $\mathcal{L}^{A,(B,C)|X,(Y,Z)}$, and it can be shown that it is given by all distributions that satisfy the following decomposition:

$$\mathcal{L}^{A,(B,C)|X,(Y,Z)} = \{p(abc|xyz) \,|\, p(abc|xyz) = \sum_{\alpha} D(a|x\alpha) p_{\alpha}^{\mathcal{NS}}(bc|yz) \\ \text{s.t. } p_{\alpha}^{\mathcal{NS}}(bc|yz) \in \mathcal{NS}\, \forall \alpha, \sum_{bc} p_{\alpha}^{\mathcal{NS}}(bc|yz) =: p_{\alpha} \leq 1\, \forall \alpha\}, \quad (5.3)$$

where the notation has the same meaning as in the Section 2.3. The last constraint captures that for each α, $p_{\alpha}^{\mathcal{NS}}(bc|yz)$ is a subnormalized probability distribution which satisfies no-signaling between B and C, but also that the value of α is independent of the inputs Y and Z. This is more interesting, as in Ref. [BHC21], the authors show how a bipartite entangled quantum state that can never lead to Bell nonlocal correlations in the standard Bell scenario can nevertheless be used to generate correlations that are broadcast-nonlocal, i.e., that are outside $\mathcal{L}^{A,(B,C)|X,(Y,Z)}$ but inside $\mathcal{Q}^{A,(B,C)|X,(Y,Z)}$. This demonstrates that while broadcasting is not "interesting" from the perspective of the span of achievable correlations—it is identical to the span of a three-way common cause scenario—it nevertheless allows for the DI certification of properties of bipartite quantum states when such certification is not possible in scenarios such as Bell's scenario.

In a similar vein, one can also consider a broadcast scenario where the broadcast channel broadcasts the state into multiple parties, as in example 5.2a of Figure 5.2 or both sides (A and B) perform broadcast channels before the Bell test, as in Figure 5.2b. The corresponding definitions of broadcast nonlocality in these scenarios follow the same logic as above, by

allowing measurement devices that share a common broadcast channel to share non-signaling resources.

5.2 Novel results and methods for broadcast nonlocality

5.2.1 Promoting standard Bell inequalities to the broadcast scenario

Here we give a method to construct Bell inequalities tailored to the broadcast scenario, starting from Bell inequalities defined in the standard scenario. In the broadcast scenario, it is shown that one can activate nonlocality for the isotropic state, $\rho_\alpha = \alpha \left|\Phi^+\right\rangle \left\langle\Phi^+\right| + (1-\alpha)\mathbb{1}/4$, for $\alpha > \frac{1}{\sqrt{3}}$ [BHC21]. This is certified by the inequality:

$$\begin{aligned}\mathcal{I} = &\langle A_0B_0C_0\rangle + \langle A_0B_1C_1\rangle + \langle A_1B_1C_1\rangle - \langle A_1B_0C_0\rangle \\ &+ \langle A_0B_0C_1\rangle + \langle A_0B_1C_0\rangle + \langle A_1B_0C_1\rangle - \langle A_1B_1C_0\rangle \\ &+2\langle A_2B_1\rangle - 2\langle A_2B_0\rangle \leq 4\,,\end{aligned} \tag{5.4}$$

written in the standard correlator notation, where:

$$\langle A_xB_yC_z\rangle = \sum_{a,b,c=\pm1} abc\, p(abc|xyz), \tag{5.5}$$

$$\langle A_xB_y\rangle = \sum_{a,b=\pm1} ab\, p(ab|xy), \tag{5.6}$$

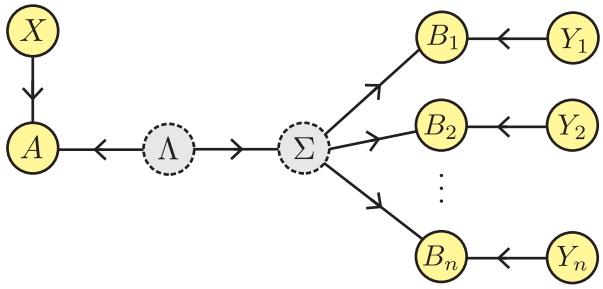
$$\langle A_x\rangle = \sum_{a=\pm1} a\, p(a|x). \tag{5.7}$$

Similar definitions hold for $\langle A_xC_z\rangle$ and $\langle B_yC_z\rangle$, and for $\langle B_y\rangle$ and $\langle C_z\rangle$. The best quantum violation found for this inequality is $4\sqrt{3}$ with the measurements and channel given in Ref. [BHC21] and using the isotropic state with $\alpha = 1$.

Ineq. (5.4) can be restructured as follows:

$$\begin{aligned}&\langle \mathrm{CHSH}\left[A_0, A_1, C_0, C_1\right](B_0 + B_1)\rangle + \\ &\qquad + \mathcal{L}_{\mathrm{CHSH}}\langle A_2\,(B_1 - B_0)\rangle \leq 2\mathcal{L}_{\mathrm{CHSH}}\,,\end{aligned} \tag{5.8}$$

(A) n-party broadcast scenario



(B) Symmetric 2-party broadcast scenario

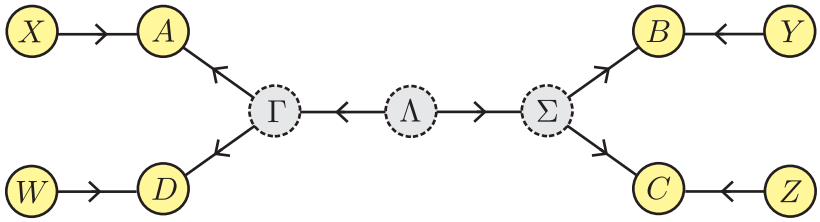


FIGURE 5.2: Generalizations of the broadcast scenario of Figure 5.1b. One (or more) of the local systems is broadcast via the application of a quantum channel, resulting in a multipartite state, sent to distant parties. Local measurements are then performed on this state, and the resulting statistics are used to rule out a local hidden-variable description for the original bipartite state.

where:

$$\mathrm{CHSH}\left[A_0, A_1, C_0, C_1\right] \coloneqq (A_0 - A_1)C_0 + (A_0 + A_1)C_1, \tag{5.9}$$

and $\mathcal{L}_{\mathrm{CHSH}}$ denotes the local bound of the CHSH inequality in the standard Bell scenario, that is, $\mathcal{L}_{\mathrm{CHSH}} = 2$. Here, we slightly abuse notation so that for example $\langle A_2 \left(B_1 - B_0\right)\rangle$ is understood as $\langle A_2 B_1\rangle - \langle A_2 B_0\rangle$. The form of (5.8) suggests the following recipe for promoting any two-outcome Bell inequality $\mathcal{I}\left[A_0, \ldots, A_m, C_0, \ldots, C_k\right]$ to the broadcast scenario through the following *ansatz*:

$$\langle \mathcal{I}\left[A_0, \ldots, A_m, C_0, \ldots, C_k\right] \left(B_0 + B_1\right)\rangle + \\ + \mathcal{L}_{\mathcal{I}} \langle A_{m+1} \left(B_1 - B_0\right)\rangle \leq 2\mathcal{L}_{\mathcal{I}}\,, \tag{5.10}$$

where $\mathcal{L}_{\mathcal{I}}$ is the local bound of $\mathcal{I}$ in the standard scenario. We prove in Appendix C.4 that (5.10) is valid so long as the Bell inequality $\mathcal{I}$ does not contain any 1-body correlator terms $\langle A_x\rangle$ for device A.

The same procedure can also be applied to the 4-partite symmetric broadcast scenario of Figure 5.2. In [BHC21, Sec. 4.2] an inequality is given for this scenario which can be written:

$$\langle \mathrm{CHSH}\left[A_0, A_1, B_0, B_1\right] C_0 \left(D_0 + D_1\right)\rangle + \\ + \mathcal{L}_{\mathrm{CHSH}} \langle C_1 \left(D_1 - D_0\right)\rangle \leq 2\mathcal{L}_{\mathrm{CHSH}}, \tag{5.11}$$

where A and D are on the left side and B and C are on the right side, as in Figure 5.2b. This inequality is also violated by the isotropic state for $\alpha > \frac{1}{\sqrt{3}}$. Similarly to above, this suggests the construction:

$$\langle \mathcal{I}\left[A_0, \ldots, A_m, B_0, \ldots, B_k\right] C_0 \left(D_0 + D_1\right)\rangle + \\ + \mathcal{L}_{\mathcal{I}} \langle C_1 \left(D_1 - D_0\right)\rangle \leq 2\mathcal{L}_{\mathcal{I}}\,. \tag{5.12}$$

We prove in Appendix C.4 that (5.12) is valid so long as the Bell inequality does not contain any 1-body correlator terms $\langle A_x\rangle$ or $\langle B_y\rangle$.

We now apply these constructions to two well-known Bell inequalities in the standard scenario, and study noise resistance with respect to the isotropic state (5.1). The two Bell inequalities we consider are (i) the chained Bell

inequality [Pea70; BC90] in the case of both measurements having inputs of cardinality 3, and (ii) the elegant Bell inequality [BG03], where A has a trichotomic input and B has an input of cardinality 4. These inequalities are both maximally violated by the maximally entangled state ($\alpha = 1$), and are violated by the isotropic state for (i) $\alpha > 4/(6\cos\frac{\pi}{6}) \approx 0.7698$, and (ii) $\alpha > 6/(4\sqrt{3}) \approx 0.8660$ respectively. Via a numerical see-saw optimization in the broadcast scenario, we have found that the corresponding inequalities (5.10) are violated in the broadcast scenario for visibilities (i) $\alpha > 0.6100$ and (ii) $\alpha > 0.6799$. Surprisingly, the same bounds are obtained using the construction (5.12). Notice that for both examples this visibility is below $\alpha = 1/K_3$ where $0.683 < 1/K_3 < 0.697$ and K_3 is Grothendieck's constant [Gro56; Hir+17; DBV17]. This means that both examples show activation of nonlocality of the isotropic state in the range in which it has a projective-LHV model [AGT06]. These examples however do not improve on the $\alpha > \frac{1}{\sqrt{3}}$ visibility achieved via the CHSH inequality. It would be interesting to investigate further if other Bell inequalities (probably with more input settings) could be used to show activation of the isotropic state below this threshold.

5.2.2 Robustness to detection inefficiencies

The isotropic state, $\rho_\alpha = \alpha \left|\Phi^+\right\rangle \left\langle\Phi^+\right| + (1-\alpha)\frac{\mathbb{1}}{4}$, is a simple model for a noisy quantum state, with α representing the probability of applying a depolarizing channel to one half of a pair of maximally entangled qubits. From an experimental perspective, there are other interesting notions of noise. Notably, detectors are not ideal, and they often fail to register an outcome. Thus, it is important to study the robustness of nonlocality with respect to detector inefficiencies. Let us first consider the standard Bell scenario and let η represent the detection efficiency, the probability of the detector working correctly. We take all detectors to have the same detection efficiency, and assume no detection events of different detectors are statistically independent. Here we consider scenarios with binary inputs taking values in $\{0, 1\}$ and binary outcomes taking values in $\{+1, -1\}$. When a no detection event occurs, an outcome in $\{+1, -1\}$ is chosen deterministically as a function of the measurement input of the detector in the round. Mathematically, this is described by two functions $f_A(x), f_B(y) : \{0, 1\} \to \pm 1$ which give the corresponding outputs given a failure event for each party and their input in

that round. Given ideal statistics $p(ab|xy)$ —computed with the noiseless quantum state and measurements— the noisy statistics $P^\eta(a,b|x,y)$ are given by:

$$\begin{aligned} P^\eta(ab\,|\,xy) = \eta^2\, p(ab|xy) \\ + \eta(1-\eta)\left[\delta_{f_A(x),a}\, p(b|y) + \delta_{f_B(y),b}\, p(a|x)\right] \\ + (1-\eta)^2 \delta_{f_A(x),a}\delta_{f_B(y),b} \quad , \end{aligned} \tag{5.13}$$

where $\delta_{i,j}$ is the Kronecker delta function. The critical detection threshold η_c is defined as the lowest η such that, for all $\eta' > \eta$, $P^{\eta'}(a,b|x,y)$ is outside the local set. In this scenario, the best known critical visibility for the two-qubit maximally entangled state is $\eta = 0.8214$ achieved using a Bell inequality with four settings per party [BG08]. This is very close to the critical detection efficiency of $\eta = 2(\sqrt{2}-1) \simeq 0.8284$ resulting from the CHSH Bell inequality.

Here, we show that by using a single copy of the maximally entangled state in the broadcast scenario, one can achieve a significantly lower critical detection efficiency of $\eta_c = 0.7355$. We consider the tripartite broadcast scenario of Figure 5.1b. In this case, we have three measurement devices, and we assume again the same detection efficiency η for each device. We similarly consider a strategy in which the detectors output either ± 1 when a failure event occurs, and describe this choice of strategy by three functions $f_A(x), f_B(y), f_C(z) : \{0,1\} \to \pm 1$. The statistics given a detection efficiency η are thus:

$$\begin{aligned} P^\eta(abc\,|\,xyz) = \eta^3\, p(abc|xyz) \\ + \eta^2(1-\eta)\left[\delta_{f_A(x),a}\, p(bc|yz) + \delta_{f_B(y),b}\, p(ac|xz) + \delta_{f_C(z),c}\, p(ab|xy)\right] \\ + \eta(1-\eta)^2\left[\delta_{f_A(x),a}\delta_{f_B(y),b}\, p(c|z) + \delta_{f_A(x),a}\delta_{f_C(z),c}\, p(c|z) + \right. \\ \left. + \delta_{f_B(y),b}\delta_{f_C(z),c}\, p(a|x)\right] + (1-\eta)^3 \delta_{f_A(x),a}\delta_{f_B(y),b}\delta_{f_C(z),c}\,. \end{aligned} \tag{5.14}$$

To find the noiseless statistics that give $\eta_c = 0.7355$, we use the see-saw algorithm from [BHC21, Appx. B]. This algorithm optimizes the robustness of the isotropic state with respect to the visibility parameter α and returns a corresponding Bell inequality valid in the broadcast scenario. We extract the channel and measurements after the algorithm converges and build the ideal

statistics $p(abc|xyz)$ using the noiseless isotropic state (i.e. the maximally entangled two-qubit state). Then we find η such that $P^{\eta}(abc|xyz)$ saturates the local bound of the returned Bell inequality, and we do this process over all possible detector strategies f_A, f_B and f_C. The lowest efficiency found is $\eta_c = 0.7355$, certified by the following inequality:

$$\langle \mathrm{CHSH}\,[A_0, A_1, B_0, B_1]\, C_1\rangle - \langle \mathrm{CHSH}\,[A_0, A_1, B_0, B_1]\rangle + 2(\langle C_1\rangle - 1) \leq 0, \tag{5.15}$$

where $\mathrm{CHSH}\,[A_0, A_1, B_0, B_1] \coloneqq (A_0 - A_1)B_0 + (A_0 + A_1)B_1$. Here, one adopts a detector failure strategy such that $f_A(x) = -1\ \forall x$, $f_B(y) = 1\ \forall y$, $f_C(z) = 1\ \forall z$. We remark that the best detection efficiency is not achieved for the inequality that gives the best visibility (*i.e.,* Eq. (5.4)). For this inequality, the best critical efficiency we found is $\eta_c = 0.7997$.

We note here that depending on the experimental implementation of the considered scenario, one will not only have inefficiencies coming from detector imperfections (losses, etc.), but also from transmission of the state, and in the case of the broadcast scenario, potential inefficiencies coming from the implementation of the channel are expected.

Previous studies of detector inefficiencies focus on the noise from the detectors and neglect other sources of noise. Therefore, we have adopted the same approach so that we can meaningfully benchmark our results against the existing literature, e.g., [Ban+13], [Ebe93], and [Bru+07].

Our result can be compared with the results of [MBV23], where improved bounds are found by using multiple copies of the maximally entangled two-qubit state achieving $\eta_c = 0.8086$, $\eta_c = 0.7399$ and $\eta_c = 0.6929$ for two, three, and four copies of the state respectively (and local quantum measurements). Note that that our bound is strictly better than that achieved with three copies of the state, while using a single copy in our scenario (plus a channel). It would be interesting to study if the techniques presented in [MBV23] could be used to find lower efficiency thresholds by applying our strategy in parallel with the use of several copies of the maximally entangled state.

5.3 Activation of non-signaling genuine multipartite nonlocality

In Refs. [Ban+13; Gal+12], the authors analyse different notions of genuine multipartite Bell nonlocality and introduce the concept of non-signaling

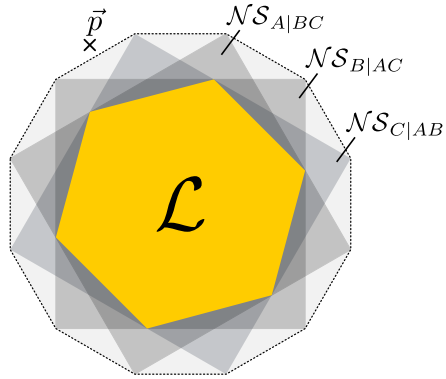


FIGURE 5.3: A vector $\vec{p}$ corresponding to a behaviour $\{p(abc|xyz)\}$ is non-signaling bilocal if it is inside the convex hull of all bipartite local polytopes (dashed line), where distant parties respect the non-signaling constraints. Behaviours which are not non-signaling bilocal are referred here as NS genuine tripartite Bell nonlocal. Here, we show that using the broadcast scenario bipartite quantum states with a LHV model for all POVMs can lead to NS genuine tripartite nonlocality.

bilocality, which is intimately related to the idea of broadcast nonlocality presented here. Following Ref. [Ban+13], a tripartite behaviour with probabilities $p(abc|xyz)$ is non-signaling bilocal (NS_2-local) if it can be written as:

$$\begin{aligned} p(abc|xyz) = t_1 \sum_\lambda q_A(\lambda) q_A(a|x\lambda) q^{\mathcal{NS}}_{BC}(bc|yz\lambda) + \\ t_2 \sum_\lambda q_B(\lambda) q_B(b|y\lambda) q^{\mathcal{NS}}_{AC}(ac|xz\lambda) + \\ (1 - t_1 - t_2) \sum_\lambda q_C(\lambda) q_C(c|z\lambda) q^{\mathcal{NS}}_{AB}(ab|xy\lambda), \end{aligned} \tag{5.16}$$

where $t_1, t_2 \in [0, 1]$, all functions are probability distributions and $q^{\mathcal{NS}}_{BC}(bc|yz\lambda)$, $q^{\mathcal{NS}}_{AC}(ac|xz\lambda)$, $q^{\mathcal{NS}}_{AB}(ab|xy\lambda)$ are non-signaling behaviours for all λ. Behaviours that are not NS_2-local are then referred to as non-signaling (NS) genuine multipartite nonlocal, see Figure 5.3.

We recall from Eq. (5.3) that a tripartite behaviour with probabilities $p(abc|xyz)$ is broadcast local if it can be written as:

$$p(abc|xyz) = \sum_{\lambda} q(\lambda) q_A(a|x\lambda) q_{BC}^{\mathcal{NS}}(bc|yz\lambda). \tag{5.17}$$

By direct inspection, we then see that any NS genuine multipartite nonlocal behaviour (that is, under definition Eq. (5.16)) is also broadcast nonlocal. Indeed, the set of non-signaling bilocal behaviours may be viewed as the convex hull of the set of broadcast local in every possible bipartition.

We now present a bipartite state which admits a local hidden-variable model for all possible local measurements but can lead to correlations that are nonlocal in the broadcast scenario. Additionally, we will show that, despite being Bell local in bipartite scenarios, this state displays NS genuine multipartite nonlocality, following the definition of Eq. (5.16). Consider the following family of two-qubit states:

$$\rho_{\mathrm{POVM}}(\alpha, \chi) := \frac{1}{2}\rho(\alpha, \chi) + \frac{1}{2}\rho_A \otimes |0\rangle\langle 0|, \tag{5.18}$$

where:

$$\rho(\alpha, \chi) := \alpha \left|\psi_\chi\right\rangle \left\langle\psi_\chi\right| + (1-\alpha)\frac{\mathbb{1}}{2} \otimes \rho_\chi^B, \tag{5.19}$$

$$|\psi_\chi\rangle := \cos\chi|00\rangle + \sin\chi|11\rangle, \tag{5.20}$$

and:

$$\rho_\chi^B := \mathrm{Tr}_A \left|\psi_\chi\right\rangle \left\langle\psi_\chi\right|, \quad \rho_A := \mathrm{Tr}_B\, \rho(\alpha, \chi). \tag{5.21}$$

As shown in Ref. [Bow+16], if $\cos^2(2\chi) \geq \frac{2\alpha-1}{(2-\alpha)\alpha^3}$, the state $\rho_{\mathrm{POVM}}(\alpha, \chi)$ admits a local hidden-variable model for all local POVMs performed by Alice and Bob.

We will make use of the inequalities for NS genuine tripartite nonlocality. Reference [Ban+13] listed all non-signaling bilocal Bell inequalities in a tripartite scenario where each party has two inputs and two outputs. Using the optimization methods detailed in Appendix B.1 of Ref. [BHC21], we have analysed all these inequalities to check whether there exist a channel

$\Sigma_{B_0\to BC}$ and local quantum measurements such that the tripartite state:

$$\rho_{ABC} := \mathbb{1} \otimes \Sigma_{B_0\to BC}[\rho_{\mathrm{POVM}}(\alpha,\chi)], \tag{5.22}$$

leads to NS-bilocal nonlocality. We have identified that the inequality 16 of Ref. [Ban+13]:

$$\begin{aligned} -2\langle C_0\rangle + \langle A_1B_0\rangle + \langle A_0B_1\rangle - \langle A_0B_0\rangle - \langle A_1B_1\rangle + 2\langle A_1C_1\rangle + 2\langle B_1C_1\rangle \\ + \langle A_1B_0C_0\rangle - \langle A_0B_0C_0\rangle + \langle A_0B_1C_0\rangle + \langle A_1B_1C_0\rangle \leq 4, \end{aligned}$$

can be used to show that the state $\rho_{\mathrm{POVM}}(\alpha,\chi)$ is NS-genuine tripartite nonlocal (hence, also broadcast nonlocal) in a region where it admits a LHV model for general POVMs. In Figure 5.4 we present the (α,χ) values for which $\rho_{\mathrm{POVM}}(\alpha,\chi)$ is guaranteed to have a local hidden-variable model (shaded region) and, for each χ, the lowest visibility α for which the state violates a NS_2 inequality (dashed blue line). In the intersection of these two regions, the POVM local state (5.18) can therefore be transformed to a NS genuinely multipartite nonlocal state via the application of a broadcast channel. We note that although standard bipartite nonlocality has been activated from bipartite POVM-local states before, this is the first example of the creation of genuine multipartite nonlocality using a single copy of such a state.

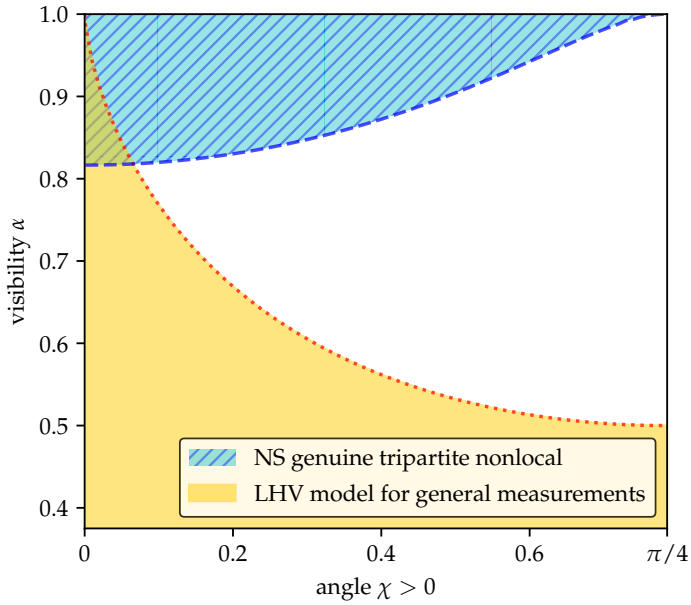


FIGURE 5.4: In the shaded (yellow) region, the state $\rho_{\text{POVM}}(\alpha, \chi)$ admits a local hidden-variable model for all POVM measurements. Also, above the (blue) dashed line the state $\rho_{\text{POVM}}(\alpha, \chi)$, with $\chi > 0$, violates the $\mathcal{NS}_2$ genuine tripartite inequality number 16 of Ref. [Ban+13]. We can see that for small values of χ, there is a range of visibility α such that the state $\rho_{\text{POVM}}(\alpha, \chi)$ is bipartite Bell local in the standard scenario but broadcast nonlocal and NS genuine tripartite nonlocal.

5.3.1 Broadcast activation without a broadcast channel

A reinterpretation of the above results also allows us to construct an example of broadcast activation without a broadcast channel. That is, we consider scenario 5.1b of Figure 5.2, and take the broadcast channel $\Sigma_{B_0\to BC}$ to be the identity channel:

$$\mathbb{1}\otimes\Sigma_{B_0\to BC}[\rho_{AB_0}]=\rho_{AB_0}. \tag{5.23}$$

Let us now take the state ρ_{POVM} defined in Eq. (5.18), in the activation region of Figure 5.4 (i.e., anywhere in the yellow-blue intersection). We then apply the channel $\Sigma_{B_0\to BC}$ which leads to activation of NS genuine multipartite nonlocality, the final state thus being:

$$\rho_\Sigma=\mathbb{1}\otimes\Sigma_{B_0\to BC}[\rho_{\text{POVM}}]. \tag{5.24}$$

Note that, since local quantum channels cannot create Bell nonlocality from states admitting a LHV model for all POVMs[1] [Bar02], ρ_Σ has a POVM LHV model on the partition $A|BC$:

$$\rho_\Sigma\to\text{Bell local for all POVMs on } A|BC. \tag{5.25}$$

However, our previous result shows that ρ_Σ is $\mathcal{NS}$ genuine tripartite nonlocal, and thus broadcast nonlocal too, using scenario (a) of Figure 5.2 and inequality (5.23):

$$\rho_\Sigma\to\text{broadcast nonlocal on } A|BC. \tag{5.26}$$

From this perspective, starting with ρ_Σ as a bipartite state on $(A|BC)$ we obtain "activation" by performing the identity channel, i.e., no broadcast channel, and by understanding Bob and Charlie as distinct parties (meaning here that they are restricted to local measurements quantum mechanically and non-signaling strategies classically).

[1] Essentially because applying the dual channel on the measurements (instead of applying the channel on the state) gives rise to the same behaviour, implying a model for all POVMs still holds for that final behaviour.

5.4 Device-independent entanglement certification

Here we apply the broadcast scenario to the task of DI entanglement certification. In [BHC21] it was shown that DI entanglement of the two-qubit isotropic state (5.1) is possible for $\alpha > \frac{1}{2}$, significantly lower than previous best known bound $\alpha \approx 0.6964$ [DBV17]. This result gave promising evidence that DI entanglement certification may be possible in the entire range $\alpha > \frac{1}{3}$ in which the state is entangled. In this section, we give strong evidence this is the case, by showing that DI entanglement certification is possible for $\alpha > 0.338$. To do this, we make use of SDP tools [VB96]. As we discuss below, given the proximity of 0.338 to $\frac{1}{3}$, we suspect that this value could be improved to any visibility arbitrarily close to $\frac{1}{3}$ with more computational resources.

The broadcast scenario we consider is the four party scenario shown in Figure 5.2b). Each local subsystem of the state of the source is broadcasted to two additional parties through channels Σ and Γ. If one considers an arbitrary separable state at the source:

$$\rho_{SEP} = \int q(\lambda)\, \sigma_\lambda^{A_0} \otimes \sigma_\lambda^{B_0}\, d\lambda, \tag{5.27}$$

with $q(\lambda)$ a normalized probability density, then after the application of the broadcast channels, the most general state shared between the four parties is:

$$\rho_{ABCD} = \int q(\lambda)\, \sigma_\lambda^{AD} \otimes \sigma_\lambda^{BC}\, d\lambda, \tag{5.28}$$

where $\sigma_\lambda^{BD} = \Gamma_{A_0\to AD}(\sigma_\lambda^{A_0})$ and $\sigma_\lambda^{BC} = \Sigma_{B_0\to BC}(\sigma_\lambda^{B_0})$ and $\Sigma_{B_0\to BC}$ and $\Sigma_{A_0\to AD}$ are the quantum channels describing the broadcasting. Local measurements performed on this state lead to behaviours of the form:

$$p(abcd|xyzw) = \mathrm{Tr}\left[\rho_{ABCD} A_{a|x} \otimes B_{b|y} \otimes C_{c|z} \otimes D_{d|w}\right] \tag{5.29}$$

$$= \int q(\lambda)\, \mathrm{Tr}\left[\sigma_\lambda^{AD} A_{a|x} \otimes B_{b|y}\right] \mathrm{Tr}\left[\sigma_\lambda^{BC} C_{c|z} \otimes D_{d|w}\right]\, d\lambda \tag{5.30}$$

$$= \int q(\lambda)\, p^{\mathcal{Q}}_{AD}(ad|xw\lambda)\, p^{\mathcal{Q}}_{BC}(bc|yz\lambda)\, d\lambda, \tag{5.31}$$

where for each λ, $p^{\mathcal{Q}}_{AD}(ad|xw\lambda)$ and $p^{\mathcal{Q}}_{BC}(bc|yz\lambda)$ are behaviours from the quantum set of correlations. Alternatively, the behaviours (5.31) are the most general that can be obtained by making local measurements on a state which is separable with respect to the bipartition AD vs BC. Since these behaviours are the most general that can be obtained from a separable source state, it follows that if a decomposition (5.31) cannot be found, the source must be entangled, and this therefore constitutes a DI certification of the entanglement of the source.

The question remains, however, of how to prove that a given behaviour does not admit a decomposition (5.31). This is complicated by the fact that the states and the measurements can in principle act on infinite dimensional Hilbert spaces. In order to tackle this, we will make use of a semi-definite programming technique introduced in [Mor+13b] and based on the NPA hierarchy [NPA08]. For a fixed number of inputs and outputs, let us denote the set of behaviours admitting a decomposition (5.31) by $\mathcal{Q}_{AD|BC}$ so that $p \in \mathcal{Q}_{AD|BC}$ if and only if (5.31) is satisfied. Furthermore, let us denote by $\mathcal{Q}^{PPT}_{AD|BC}$ the set of correlations obtained by using states that admit a positive partial transpose (PPT) with respect to the bipartition AB vs CD. Since separable states are PPT, it follows that $\mathcal{Q}_{AD|BC} \subseteq \mathcal{Q}^{PPT}_{AD|BC}$ and thus $p \notin \mathcal{Q}^{PPT}_{AD|BC} \implies p \notin \mathcal{Q}_{AD|BC}$. A certificate that $p \notin \mathcal{Q}^{PPT}_{AD|BC}$ is therefore a device-independent proof of entanglement.

Such a certificate may be obtained by using a modification of the PPT-NPA hierarchy presented for a partition A vs B in the preliminaries in Chapter 2.3.2, and which we extend to the partition AD vs BC. One builds the NPA hierarchy for 4 measurement devices A, B, C and D.

Now the separability of ρ_{ABCD} across the bipartition $(AD|BC)$ implying that it is PPT across the same bipartition leads to the following constraints at the level of the moments of the NPA relaxation (cf. Eq. (2.85) from the preliminaries):

$$\begin{aligned}\langle A_{i_1} \dots A_{i_{n-1}} A_{i_n} B_{j_1} \dots B_{j_m} C_{k_1} \dots C_{k_o} D_{l_1} \dots D_{l_{p-1}} D_{l_p} \rangle_{\rho_\Lambda^{T_{AD}}} = \\ = \langle A_{i_n} A_{i_{n-1}} \dots A_{i_1} B_{j_1} \dots B_{j_m} C_{k_1} \dots C_{k_o} D_{l_p} D_{l_{p-1}} \dots D_{l_1} \rangle_{\rho_\Lambda}. \quad (5.32)\end{aligned}$$

One may now define a SDP program analogous to that of SDP (2.88) which whose infeasibility certifies that a correlation $p(abcd|xyzw) \notin \mathcal{Q}^{PPT}_{AD|BC}$.

Using the above, we were able to prove device-independent entanglement

certification for ρ_α for $\alpha > 0.338$. To achieve this, we used a heuristic optimization procedure described in Appendix C.1. The precise strategy involves each of the measurement devices having inputs of cardinality $3 = N_{|X|} = N_{|Y|} = N_{|Z|} = N_{|W|}$. The numerical values of the measurement and channels, as well as the corresponding Bell inequality that certifies this visibility, can be found in the GitHub repository for the article [Git]. Although we could not obtain a proof that $\alpha > \frac{1}{3}$ implies the possibility of a DI entanglement certification in the broadcast scenario, our numerical analysis strongly suggests that all entangled two-qubit Werner states can be DI certified in the broadcast scenario. This result suggests that an analytic proof of DI entanglement certification for $\alpha > \frac{1}{3}$ may be within reach, and would be an exciting avenue of future research. Given that this is exactly the separability limit, one may even hope that broadcasting could activate DI entanglement certification for all entangled states.

5.5 Broadcast steering

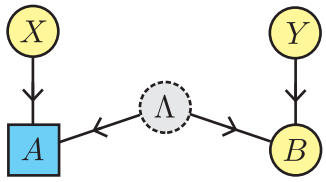
In this section, we introduce a new broadcast scenario which is based on EPR-steering from one measurement device B to another, A. This represents a scenario where the measurement A is completely characterized, but no assumption is made on the device B. In this broadcast steering scenario we present a novel example of activation of EPR-steering correlations, which does not rely on previous methods such as local filtering [Qui+15] or the multi-copy regime [QBH16].

5.5.1 Standard quantum steering

Before presenting the EPR-steering broadcast scenario, we review the concept of standard EPR-steering [WJD07]. For a more detailed introduction, we recommend the review articles [CS16; Uol+20]. We consider a scenario where a bipartite state ρ is measured by two devices A and B, and B implements a set of local POVMs described by $\{B_{b|y}\}$. When B performs the measurement labelled by y and obtains the outcome b, the physical system accessible by A is described by its *assemblage*, a set of unnormalized states defined by:

$$\sigma_{b|y} := \mathrm{Tr}_B(\mathbb{1} \otimes B_{b|y}\rho), \tag{5.33}$$

(A) Standard steering



(B) Broadcast steering

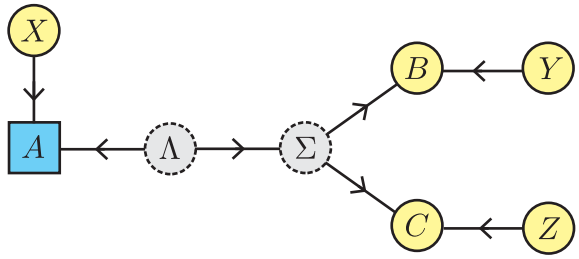


FIGURE 5.5: Illustration of a steering scenario where the measurement A is trusted to be completely characterized (hence, represented by a square and by the colour blue). Figure 5.5a represents the standard steering scenario and 5.5b represents the case where a broadcast channel sends the system to two measurement devices which only have access to non-signaling resources.

where Tr_B denotes the partial trace over Bob's subsystem and the $\mathrm{Tr}(\sigma_{b|y})$ corresponds to the probability of obtaining the output b given input y for Bob. An assemblage admits a local hidden-state (LHS) model if it can be written as :

$$\sigma_{b|y} = \int q(\lambda)\sigma_\lambda q_B(b|y\lambda)\, d\lambda, \tag{5.34}$$

where λ stands for a hidden variable and $\{q(\lambda)\}_\lambda$ and $\{q_B(b|y,\lambda)\}_b$ are probability distributions. We thus say that an assemblage $\sigma_{b|y}$ is *steerable* if it does not admit a LHS decomposition of the form (5.34).

5.5.2 Steering in the broadcast scenario

In the simplest broadcast scenario, we start with bipartite state ρ_{AB_0}, which after channel $\Sigma_{B_0\to BC}$ is mapped to state ρ_{ABC}, which is then subject to measurement devices A, B and C which have inputs X, Y and Z. The question is then whether the statistics observed by A can be explained by a local hidden-state model. That is, can the assemblage:

$$\sigma_{bc|yz} = \mathrm{Tr}_{BC}\left(\mathbb{1}\otimes B_{b|y}\otimes C_{c|z}\left[\mathbb{1}\otimes\Sigma_{B_0\to BC}(\rho)\right]\right), \tag{5.35}$$

be written as:

$$\sigma_{bc|yz} = \int q(\lambda)\sigma_\lambda p^{\mathcal{NS}}_{BC}(bc|yz\lambda)\, d\lambda, \tag{5.36}$$

where $p^{\mathcal{NS}}_{BC}(bc|yz\lambda)$ is an arbitrary non-signaling distribution over B and C, for each value λ. We refer to a violation of (5.36) as *broadcast steering*, and similarly to broadcast nonlocality, such a violation cannot be explained by the transformation device alone, as long as it generates non-signaling resources only.

In this chapter, we also consider the scenario where the broadcast channel Σ maps the space B_0 to a tripartite space $B\otimes C\otimes D$. That is, after the channel the state is a four-partite state ρ_{ABCD}, subjected to measurement devices A, B, C and D. We then consider an assemblage:

$$\sigma_{bcd|yzw} = \mathrm{Tr}_{BCD}\left(\mathbb{1}\otimes B_{b|y}\otimes C_{c|z}\otimes D_{d|w}\left[\mathbb{1}\otimes\Sigma_{B_0\to BCD}(\rho)\right]\right), \tag{5.37}$$

which is broadcast steerable if it can be written as:

$$\sigma_{bcd|yzw} = \int \Pi(\lambda)\, \sigma_\lambda\, p^{\mathcal{NS}}_{BCD}(bcd|yzw\lambda)\, d\lambda, \tag{5.38}$$

where $p^{\mathcal{NS}}_{BCD}(bcd|yzw\lambda)$ is an arbitrary non-signaling behaviour for each λ.

Before finishing this subsection, we remark that, in a standard steering scenario (steering from B to A), the main hypothesis is that the device A is trusted to be fully characterized. In the broadcast steering case, we need an additional hypothesis, which is that the output of the channel Σ may be any no-signaling resource, and that B and C implement local measurements, that is, they may be strongly correlated through a common cause, but the devices cannot influence one another. Nevertheless, this hypothesis can be imposed in a physical and fair way by ensuring that the devices after the broadcast channel are in sufficiently separated at the time of the measurements.

Other potential notions of broadcast steering

We presented broadcast steering for the case where A is the trusted device. In principle, one could consider other natural configurations for defining broadcast steering:

- B and C are the trusted devices: in that case, one wonders whether the assemblage:

  $$\sigma_{a|x} = \mathrm{Tr}_A(A_{a|x} \otimes \mathbb{1}_B \otimes \mathbb{1}_C[\mathbb{1} \otimes \Sigma_{B_0 \to BC}(\rho)]), \tag{5.39}$$

  can be written as:

  $$\sigma_{a|x} = \int q(\lambda)\, \sigma_\lambda\, q_A(a|x\lambda)\, d\lambda. \tag{5.40}$$

  Note however that, this corresponds to standard steering, where B and C can be seen as a single device. Thus, this scenario is trivial and no activation is possible in this case.

- Hybrid cases: either A and B are trusted devices, or only C is trusted. In either case, the absence of a LHS model does not imply anything about ρ_{AB_0}, since it could be explained by (standard) bipartite steering between B and C.

Since these two other approaches lead to trivial definitions, we focus on the definition described by (5.36) where only device A is trusted to perform characterized measurements.

5.5.3 Broadcast steering with the two-qubit isotropic state

We now present some steering activation results in broadcast scenarios by carefully analysing the two-qubit isotropic state[2]:

$$\rho_\alpha = \alpha|\phi^+\rangle\langle\phi^+| + (1-\alpha)\frac{\mathbb{1}}{4}. \tag{5.41}$$

The isotropic state represents a maximally entangled state which undergoes white noise. Due to its symmetry, simplicity, and experimental relevance, the isotropic state is often used as a benchmark for several tasks in quantum information. In the standard steering scenario, the two-qubit isotropic state was only shown to be steerable for visibility $\alpha > \frac{1}{2}$ [WJD07]. Moreover, the two-qubit isotropic state has a LHS model for projective measurements when $\alpha \leq \frac{1}{2}$ [WJD07], and there is evidence that it also has a LHS model for general POVMs when $\alpha \leq \frac{1}{2}$ [Bav+17; Cha+18]. The results presented in this subsection were obtained with the help of the heuristic search described in the Appendix C.3 found in the GitHub online repository [Git].

Two broadcasted devices—We first consider a scenario where there is broadcasting to two devices B and C, as in Figure 5.1b. When B and C have dichotomic inputs, that is, $y \in \{0,1\}$ and $z \in \{0,1\}$, we could find a channel Σ and measurements $\{B_{b|y}\}$, $\{C_{c|z}\}$, to certify broadcast steering for $\alpha > 0.5616$. We also investigated the scenario where B and C have three dichotomic measurements each, i.e., $y \in \{0,1,2\}$ and $z \in \{0,1,2\}$. In this case, we detected broadcast steering up to $\alpha > 0.4945$.

Three broadcasted devices—We now consider the scenario where there are three devices B, C, and D, after the broadcast channel as in Figure 5.2a. We focused on the scenario where B, C and D have dichotomic inputs.. Since the vertices of the 3-partite non-signaling polytope performing two dichotomic measurements were explicitly obtained in Ref. [PBS11], we can use these vertices to run (a straightforward extension of) our heuristic procedure

[2]We remark that since the two-qubit isotropic state is local-unitary equivalent to the two-qubit Werner state, all results presented in this subsection also hold for the two-qubit Werner state.

presented in Appendix C.3. This allowed us to certify that the two-qubit isotropic state exhibits broadcast steering for $\alpha > 0.4678$, showcasing an even stronger example of steering activation.

Note that the latter two results are example of activation of steering (relative to projective measurements), since the isotropic state admits a LHS model in the range $\alpha \leq \frac{1}{2}$. These are the first examples of single-copy activation of steering for this class of states.

5.6 Discussion

The relationship between quantum entanglement and Bell nonlocality plays a major role in understanding quantum correlations and the development of device-independent protocols. In a seminal paper, Werner showed that entangled states may admit a local hidden-variable model and cannot lead to Bell nonlocal correlations in the standard Bell scenario [Wer89]. What seemed to be definite proof that some entangled states cannot lead to nonlocality is today recognized as only a first (fundamental) step. Over the past years, natural extensions of Bell scenarios revealed that states admitting local hidden-variable models may also display nonlocal correlations [Pop95] and we are forced to accept that the relationship between entanglement and nonlocality is far from being fully understood.

In this chapter we investigated entanglement and nonlocality scenarios where broadcasted systems reveal strong correlations which are hidden in the standard Bell test. From a foundational perspective, we provided novel examples of how to activate the nonlocality of entangled states which admit local hidden-variable models. We presented an example of bipartite local states leading to genuine multipartite nonlocality, introduced the concept of broadcast device-independent entanglement certification and the concept of broadcast steering. From a more practical aspect, we developed analytical and computational methods to analyse entanglement and nonlocality in broadcast scenarios. Our findings advance the discussion on whether entanglement can lead to nonlocality, and we hope that the methods presented here may pave the way for network-based and broadcast-based device-independent protocols.

All the code can be found in the GitHub online repository [Git] and can be freely used under the MIT licence [Lic].

Chapter 6

Tests for causal compatibility

This chapter is based on Ref. [BWPK23]:
Emanuel-Cristian Boghiu, Elie Wolfe, and Alejandro Pozas-Kerstjens, *"Inflation: a Python library for classical and quantum causal compatibility"*, Quantum 7, 996 (2023)—Published 4 May 2023

In this chapter, we focus on improving existing methods for testing for causal compatibility of observed correlations with a given causal structure. A particularly successful and versatile tool for testing causal compatibility is the inflation method [WSF19; Gis+20a; Wol+21], which consists of a series of increasingly strict necessary conditions that can be tested via linear or semidefinite programming (cf. Sections 3.2 and 3.4 of the preliminaries). Despite its broad applicability within and outside the field of quantum nonlocality, available implementations of the inflation technique are typically limited in terms of the type of causal structures it applies to, or in the type of inflations considered (see, e.g., [PKGT22; PKGR23b]). This means that researchers must code their own programs every time they seek to analyse a different structure or try a different solution, adding an extra level of difficulty to the application of the technique.

In this chapter we discuss the package Inflation [BWPK], an open-source library, written in Python, that implements the inflation framework for causal compatibility. It allows both for solving feasibility problems (i.e., answering the question *"can I generate this distribution in this causal scenario?"*) and for bounding optimal values of functionals over distributions compatible with a causal structure. These include all network scenarios considered in the field of quantum nonlocality [Tav+22], and structures that have recently gained attention in that field, such as the so-called instrumental scenario [VH+19].

Currently, Inflation implements both the quantum inflation hierarchy of Ref. [Wol+21] and the fanout and non-fanout inflations of Ref. [WSF19; Gis+20b]. The library can also be used, by setting the corresponding flags, for assessing compatibility with distributions generated when latent nodes represent mixes sources of correlations, namely either quantum or classical (framed within the quantum inflation hierarchy, thereby extending the ideas in [Bac+17; SG11]) or no-signaling or classical sources of correlations. In addition, one may also easily consider any network together with a global source of classical shared randomness.

This chapter presents the package and illustrates simple use cases, which are extended in the library documentation. The fundamental ideas behind the inflation technique have already been presented in Sections 3.2 and 3.4 of the preliminaries. In Section 6.1 we show how to get started with the library and describe its main components and features. In Sections 6.2 and 6.3 we demonstrate with code snippets the different types of problems that can be addressed with Inflation. We discuss further software details and library information in Section 6.4 including future development, contribution guidelines, and planned maintenance and support, and we provide some concluding remarks in Section 6.5.

6.1 The library

6.1.1 Requirements and installation

Inflation is a Python library that can be installed on Mac, Windows, and Linux operating systems via `pip` by executing the instruction below at a command line.

```
pip install inflation
```

The core requirements of Inflation are NumPy [HMW+20] (used for general numerical procedures), SymPy [MSP+17] (used to make the input format more user-friendly), and SciPy [VGO+20] (used for handling sparse matrices). It can also use Numba [LPS15] as a just-in-time compiler to speed up core calculations. For solving the generated relaxations, the library uses the MOSEK Fusion API [ApS19] to solve the linear and semidefinite programming problems. It also allows for writing the problem in a human-readable form as a comma-separated values file, to a MATLAB-compatible .mat file,

or to SDPA data format for further manipulation in other interfaces such as Yalmip [Löf04].

To test installation, one can run the following.

```
import inflation
inflation.about()
```

These lines print basic information about the version of Inflation installed and the versions of installed dependencies.

```
# Inflation: Implementations of the Inflation Technique for Causal
    Inference
# ========================================================================
# Authored by: Emanuel-Cristian Boghiu, Elie Wolfe and Alejandro Pozas
    -Kerstjens

# Inflation Version:        0.1
#
# Core Dependencies
# -----------------
# NumPy Version:  1.23.1
# SciPy Version:  1.8.1
# SymPy Version:  1.11.1
# Numba Version:  0.56.2
# Mosek Version:  10.0.20
#
# Python Version: 3.10.4
# Platform Info:  Windows (AMD64)
```

The source code for Inflation is hosted on GitHub at

https://github.com/ecboghiu/inflation

and is distributed with an open-source software licence: GNU GPL version 3.0. More details about the software, packaging information, and guidelines for contributing to Inflation are included in Section 6.4.

6.1.2 Components

There are two main layers in Inflation. First, the basic characterization of the causal scenario and its desired inflation are stored in an `InflationProblem`. This includes the directed-acyclic-graph (DAG) describing the causal structure, the number of inputs and outputs of each of its visible nodes, and the number of copies of each latent node in the desired inflation. If the causal structure furthermore contains visible-to-visible connections, then the procedures in, e.g., [NW20a, Sec. V], [Wol+21, Sec. V] and [LGG23, Sec. IV.C] are executed in order to find the network and suitable constraints that generate the same distributions as the original causal structure.

The second layer takes an `InflationProblem` object and sets up and solves the compatibility or optimization problem of interest. Currently, the library supports the quantum inflation hierarchy described in Ref. [Wol+21] via the `InflationSDP` object, and the fanout and non-fanout inflations of Refs. [WSF19; Gis+20b] via the `InflationLP` object.

In the quantum inflation hierarchy, the characterization of the set of probability distributions is given by a list of operators, in the spirit of the NPA hierarchy [NPA07; NPA08; PNA10]. This is input to the `InflationSDP` object via the function `generate_relaxation()`, which admits either a generic list-of-lists notation for arbitrary lists of operators, or string-based notations for operator sets that are routinely used in the literature: `npa#` for the sets describing the NPA hierarchy (denoted as $\mathcal{T}_n$ in Ref. [NPA07] and as $\mathcal{S}_n$ in Ref. [NPA08]), `local#` for the so-called local levels (denoted as $\mathcal{L}_n$ in [Wol+21, App. C], see also Ref. [Mor+13b]), and `physical#` for the sets of operators of bounded length whose expectation value is non-negative for any quantum state.

The function `generate_relaxation()` automatically imposes the equality constraints that are derived from the invariance of the inflation under permutation of copies of a same original element (this is, it imposes the constraints described in [Wol+21, Eq. (10)]). Furthermore, it identifies which marginals of distributions in the inflation must coincide with marginals of distributions in the original scenario. After this, the user can specify either a probability distribution over the visible nodes for a feasibility problem (i.e., to determine whether the distribution can be identified as incompatible with the causal structure) using the function `set_values()`, or a combination of operators whose expectation value will be optimized, by using the function `set_objective()`. When using `set_values()`, the user can choose to set also the so-called linearized polynomial constraints, which constrain the set of compatible distributions further at the expense of obtaining certificates with more limited applicability [PK19; PKGR23b].

In the fanout and non-fanout inflation hierarchies instead of inputting a list of operators, the package automatically generates all combinations of events (in Collins-Gisin notation, cf. Section 2.3.3) which are compatible with the nature of the preparations. As such, when using `InflationLP` it is not necessary to call `generate_relaxation()`. With an instance of `InflationLP` one may use methods such as `set_values()` to fix an observed distribution or `set_objective()` to specify a linear functional to optimize

over the convex relaxation. If we consider the triangle scenario with an inflation of level 2 with all classical preparations, then the package will automatically generate the Web inflation presented in Section 3.4; if the preparations are restricted to all be no-signaling, then this will generate the hexagon inflation presented in Section 3.4. In a hybrid scenario with both classical and no-signaling preparations, the library will generate the most general inflation which fans out the classical preparations, and does not fanout the no-signaling preparations.

6.1.3 Reductions of the feasible region

In general, the fact that one must consider multiple copies of the elements in the original causal structure leads to a large computational load when storing and solving the relevant problems. Inflation implements a number of additional constraints not included in the original definitions of the hierarchies, either automatically or at the user's choice, that give a tighter relaxation for the same level of computational resources. These constraints are:

Non-negativity of physical moments. This is a feature only relevant for the implementations of inflation that characterize distributions generated by measuring quantum systems. Recall that implementations of quantum inflation require the use of the NPA hierarchy [NPA07] in order to assess whether a compatible inflation exists. The main object in the NPA hierarchy are the so-called moment matrices, whose rows and columns are indexed by products of (a priori unknown) projection operators. Each cell of the moment matrix contains the expectation of the product of the operators in the row and the operators in the column under an also unknown quantum state. Despite all elements being unknown a priori, one can derive constraints for some of them in certain situations. For instance, it is known that eigenpotent operators have a non-negative expectation value under any possible quantum state, and thus these non-negativity constraints are always imposed in Inflation. In fact, using operator products which have a non-negative expectation value under any quantum state in the generating set for `InflationSDP` leads to drastic reductions in problem size for certain problems, as we explicitly show in Section 6.2.1.

Sandwich-nonnegativity. The operator products described above have a non-negative expectation value also if they appear *sandwiched* between another product of operators and its conjugate. Indeed, if O_2 is a product of operators that has non-negative expectation value with any state, then $\langle\psi|O_1^\dagger O_2 O_1|\psi\rangle = \langle\psi'|O_2|\psi'\rangle \geq 0$ for any O_1, $|\psi\rangle$. An example of these products are those that correspond to subsequent measurements on a same state, whose expectation values represent the probabilities of sequences of outcomes. This feature is also automatically imposed when generating an instance of `InflationSDP`.

Linearized polynomial constraints. Both in the inflation methods for compatibility with quantum (based on semidefinite programming) and classical and generalized physical models (based on linear programming), it is possible to transform certain non-linear relations between the unknowns in the problems into linear ones when one assesses the compatibility of a given distribution with the scenario. When a subgraph of the inflation contains more than one connected components, and (at least) one of these components can be associated a numerical value from the distribution under scrutiny (these are known as injectable components in the terminology of Ref. [WSF19]), the variables associated to the subgraph can be related to those corresponding to its non-injectable connected components via linear relations, reducing the feasible region. Linearized polynomial constraints are used, for instance, in Ref. [PKGR23b] in the case of inflation for compatibility with classical models. For compatibility with quantum models, [PK19, App. D.2] contains a demonstration of its advantage in several scenarios. It should be noted that when using linearized polynomial constraints, in case of infeasibility, the dual of the semidefinite (or linear) program can only serve as a certificate of infeasibility for the tested distribution, given that for other distributions the feasible region reduction under linearized polynomial constraints can be different [PKGR23b]. Inflation allows imposing linearized polynomial constraints by setting the flag `use_lpi_constraints=True` in `set_values()`.

6.2 Main functionality

We showcase here the user experience in using Inflation to solve a series of problems that routinely appear in causal inference scenarios. All these

examples can be downloaded as ready-to-run Jupyter notebooks from the repository of the library (see Section 6.4.3).

6.2.1 Feasibility problems and extraction of certificates

The first example we will consider is that of demonstrating that a particular distribution can not be generated in a specific network arrangement when the preparations distribute quantum systems, and the measurement devices (the visible nodes) implement quantum measurements on the systems received. We illustrate this type of problems by showing that the so-called W distribution, defined as

$$P_W(a,b,c) = \begin{cases} \frac{1}{3} & \text{if } a+b+c=1, \\ 0 & \text{otherwise} \end{cases}, \tag{6.1}$$

where $a, b, c \in \{0, 1\}$, is incompatible with the quantum triangle causal scenario.

The first step is to specify the original scenario and its desired inflation by creating an instance of `InflationProblem`. This requires specifying (i) the DAG of the original scenario as a dictionary where the keys are parent nodes and the values are lists of the corresponding children, (ii) the number of possible outcomes and number of possible measurement settings of each of the visible nodes in the scenario, and (iii) the inflation levels, which represent the number of copies of each latent node that will be considered in the inflated scenario:

```
dag = {"rho_AB": ["A", "B"],
       "rho_BC": ["B", "C"],
       "rho_AC": ["A", "C"]}
nr_outputs = [2, 2, 2]
nr_inputs  = [1, 1, 1]
nr_copies  = [2, 2, 2]
classical_sources = []
scenario   = InflationProblem(dag, nr_outputs, nr_inputs,
nr_copies, order=['A', 'B', 'C'], classical_sources=
classical_sources)
```

The order of the parties is specified via the `order` optional argument, and the order of the sources is taken to be the same as insertion order in the `dag` dictionary. Which sources, or preparations, are classical is specified

by the `classical_sources` argument, which is a list of the names of the preparations, as specified in the keys of `dag`, which are classical in nature.

The generated instance of `InflationProblem` is fed to an `InflationSDP` instance. This object controls all the features related to the semidefinite relaxation of the problem considered. For instance, one can easily specify the relaxation obtained when using as generating monomials all products of operators of length at most 2 (this is, what is known as the second level in the NPA hierarchy [NPA07; NPA08]):

```
sdp = InflationSDP(scenario)
sdp.generate_relaxation("npa2")
```

Similarly, one may also feed the generated instance of `InflationProblem` to `InflationLP`:

```
lp = InflationLP(scenario)
```

Note that it is not necessary to call `generate_relaxation()` when instantiating an `InflationLP` object. In the code snippets that follow we will mostly use an `InflationSDP` instance, but most of the examples works similarly with an `InflationLP` instance.

We need to set the constraints corresponding to observing the target probability distribution in marginals of the inflation distribution that can be identified with marginals of the original scenario. If one wants to do this for all marginals, this can be achieved with the function `set_distribution()`, but if more granularity is needed one must use the function `set_values()` instead (see Section 6.3.2 for an explicit example using this function). The distribution must be input as a multidimensional NumPy array:

$$\texttt{P[out}_1\texttt{,...,out}_n\texttt{,in}_1\texttt{,...,in}_n\texttt{]},$$

where each cell contains the corresponding probability, and where the i-th output and input, $\texttt{out}_i$ and $\texttt{in}_i$, correspond to the party in the i-th position in the `order` keyword argument.

```
sdp.set_distribution(P_W)
```

Then, running `sdp.solve()` executes the semidefinite program and stores its status in `sdp.status`. For the problem at hand, this is `infeasible`, meaning that the W distribution of Eq. (6.1) does not admit a second-order (because of our specification of `nr_copies`) quantum inflation of the triangle scenario, and thus by the inflation arguments, it cannot be generated in the

triangle causal scenario when the latent nodes represent sources of bipartite quantum states.

Once the problem is determined to be infeasible, linear and semidefinite programming provides a certificate of such infeasibility. These certificates of infeasibility, which witness incompatibility with a given inflation, can be transformed into polynomial Bell-like inequalities that witness distributions incompatible with the original causal scenario (see, e.g., [Wol+21, Sec. VII.B]). Inflation automatically computes these certificates and provides them in a variety of useful forms. For instance, running

```
sdp.certificate_as_probs(clean=True)
```

produces as output a SymPy object of the following form

$$-0.476p_A(0|0)^2 + 0.059p_A(0|0)p_{AB}(00|00) + \\ + 0.476p_A(0|0)p_{ABC}(000|000) + \cdots + 0.563,$$

which signals distributions that produce a negative value as incompatible with the triangle scenario. This object can be further manipulated easily with SymPy's built-in functions.

Feasibility as optimization. In terms of numerical stability, it is often advised to frame feasibility problems as optimization problems. This is specially relevant when the distribution whose compatibility we are testing is close to the boundary of the feasible set, where analytically feasible problems can be reported as infeasible.

Inflation allows for this framing in the case of quantum inflation by optimizing the smallest eigenvalue of the problem's moment matrix[1]. This is achieved by passing one argument at solving time:

```
sdp.solve(feas_as_optim=True)
```

The optimal (largest) value of the smallest eigenvalue of the moment matrix is stored in `sdp.objective_value` and by inspecting its value one can determine whether the original problem is feasible (if `sdp.objective_value` ≥ 0) or not (if `sdp.objective_value` < 0). Furthermore, the quantity in `sdp.objective_value` can provide a rudimentary notion of distance to the

[1] If the largest value that the smallest eigenvalue can achieve is negative, it means that the moment matrix associated to the problem cannot be made positive-semidefinite, and thus the corresponding feasibility problem is infeasible.

feasible set, and help estimating how the size of the feasible set changes when adding extra constraints to the problem or when changing the generating set. Importantly, the extraction of certificates is unaffected, and can still be performed even when treating feasibility problems as optimization problems. In the case of a linear programming relaxation, Inflation reframes all linear constraints as a greater than or equal to zero inequality constraints, $Ax + b \geq 0$, and then maximises λ such that $Ax + b \geq \lambda$. If the result of the opimization is 0, then the original problem was feasible and infeasible in any other scenario.

Physical moments as generating set. It is known that, when considering the distribution in Eq. (6.1) subject to white noise,

$$\nu P_W + (1 - \nu)\frac{1}{8}, \quad \nu \in [0, 1], \tag{6.2}$$

quantum inflation of order 2 can certify its incompatibility with the triangle with quantum latent nodes at least down to $\nu = 0.8038$. This result was obtained in Ref. [Wol+21] by using a moment matrix of size 1175×1175. By using as monomials indexing the rows and columns of the moment matrix only those with non-negative expectation values under any quantum state as mentioned in Sec. 6.1, one can recover the same ν_{crit} but with a much smaller moment matrix, of size 287×287. Due to its notable gains in memory (and its consequent gains in speed), using the non-negative monomials in the generating set is made very easy in Inflation. In order to recover the result mentioned above with the much smaller moment matrix, one just needs to run:

```
genset = sdp.build_columns("physical2", max_monomial_length
=4)
sdp.generate_relaxation(genset)
v = 0.8039
sdp.set_distribution(v*P_W + (1-v)/8)
sdp.solve()
# infeasible
```

This improvement, is most notable when dealing with problems that involve distributions without settings. In our experience, in the case where the parties in the problem have a choice of different measurements to perform on the states received by the sources, gains are more moderate.

6.2.2 Optimization of Bell operators

The second large class of problems that can be solved within Inflation is obtaining bounds on expectation values of Bell operators. For example, let us consider Mermin's operator [Mer90]:

$$\text{Mermin} = \langle A_1 B_0 C_0 \rangle + \langle A_0 B_1 C_0 \rangle + \langle A_0 B_0 C_1 \rangle - \langle A_1 B_1 C_1 \rangle,$$

where $\langle A_x B_y C_z \rangle = \sum_{a,b,c \in \{0,1\}} (-1)^{a+b+c} p(a,b,c|x,y,z)$. Ref. [Wol+21] bounds this quantity in the quantum triangle scenario by using its second-order inflation and a generating set composed of the union of the second level in the associated NPA hierarchy and the first local level, obtaining that its maximum value cannot be larger than 3.085.

In order to reproduce these results in Inflation, one needs to pass the corresponding generating set to `generate_relaxation()` and set the objective function. The former can be easily done since `generate_relaxation()` also admits an explicit list of operators as argument:

```
npa2   = sdp.build_columns("npa2",   symbolic=True)
local1 = sdp.build_columns("local1", symbolic=True)

npa2_union_local1 = set(npa2).union(set(local1))

sdp.generate_relaxation(list(npa2_union_local1))
```

Setting the objective is done via the function `set_objective()`, which admits as input a SymPy object that is a polynomial of the operators involved in the problem

```
mmnts = sdp.measurements
A0, B0, C0, A1, B1, C1 = (1 - 2*mmnts[party][0][x][0]
                                  for x     in [0, 1]
                                  for party in [0, 1, 2])
Mermin = A1*B0*C0 + A0*B1*C0 + A0*B0*C1 - A1*B1*C1

sdp.set_objective(objective=Mermin, direction="max")
sdp.solve()
print(sdp.objective_value)
# 3.0851...
```

Note that, as mentioned in [Wol+21, Sec. VII.C.2], one can also optimize polynomial expressions of distributions compatible with the original scenario as long as they can be written as linear combinations of products of the operators in the inflation. For example, one can optimize the mean

squared distance to a target distribution in a second-order inflation by using operators corresponding to non-overlapping inflation copies. In Inflation, these operators are stored in the second dimension of `sdp.measurements`.

6.2.3 Bounds on critical parameter values

A third large class of problems of interest in quantum information theory that can be solved in Inflation is the calculation of critical values of some parameter that characterize the family of distributions under scrutiny. Examples of this are the estimation of the maximum amount of noise [PK+19], the maximum angle between the measurements of a party [PKGT22], or the maximum probability of detection failure [Abi+22] beyond which nonlocality cannot be certified. This type of problems can be handled within Inflation by using the function `max_within_feasible`. This function takes as input an instance of an `InflationSDP` that characterizes the set of feasible distributions, and a mapping (in the form of a Python dictionary) from cells in the corresponding moment matrix to arbitrary symbolic expressions depending on the variable to be optimized.

Currently Inflation features two ways for obtaining critical parameter values, which are specified through the `method` flag in `max_within_feasible`. The first one, `method="bisection"`, is a bisection algorithm, which takes increasingly small steps in the direction of the critical value of the parameter, taking $n = \lceil \log_2 \Delta - \log_2 \varepsilon \rceil$ iterations to reach a solution within accuracy ε (where Δ is the width of the interval that the variable is constrained to lie in). The second method, `method="dual"`, exploits the certificates of infeasibility in order to reduce the number of iterations required. The certificates of infeasibility are surfaces that always leave the feasible region in the half-space that takes positive values. This second method, instead of modifying the parameter by a fixed value, chooses as next candidate the value that lies in the boundary of the certificate, typically leading to fewer evaluations of semidefinite programs. Furthermore, it is possible to use the functions `set_values()` and `set_distribution()` as shortcuts to obtain complete dictionaries of assignments that are stored in `InflationSDP.known_moments`.

As an illustration of the use of `max_within_feasible`, the following example computes, after defining the relevant `InflationSDP`, the critical visibility for the W distribution of Eq. (6.1):

```
from inflation import max_within_feasible
from sympy import Symbol
```

```
v = Symbol("v")
sdp.set_distribution(v*P_W + (1-v)/8)
max_within_feasible(sdp, sdp.known_moments, "dual", bounds
=[0, 1])
# 0.8038...
```

6.3 Further features

In this section we collect additional functionality of Inflation when restricting to specific kinds of problems and situations. All the functionality described in this section can be combined seamlessly, both among them and with that described earlier.

6.3.1 Standard NPA hierarchy

Since multipartite Bell scenarios (namely those where all parties receive states distributed by the same source) are particular instances of causal scenarios, Inflation can also be used for analysing standard multipartite quantum correlations using the NPA hierarchy [NPA07; NPA08]. In order to do so, one can call `InflationProblem` without specifying its `dag` argument. For instance, the code to optimize the CHSH inequality in the bipartite Bell scenario with quantum preparations is:

```
scenario = InflationProblem(outcomes_per_party=[2, 2],
                            settings_per_party=[2, 2])
# UserWarning: The DAG must be a non-empty dictionary with
parent variables as keys and lists of children as values.
Defaulting to one global source.
sdp = InflationSDP(scenario)
sdp.generate_relaxation("npa1")
mmnts = sdp.measurements
A0, B0, A1, B1 = (1 - 2*mmnts[party][0][x][0] for x     in
[0, 1]
                                                for party in
[0, 1])
CHSH = A0*(B0+B1) + A1*(B0-B1)
sdp.set_objective(CHSH)
sdp.solve()
print(sdp.objective_value)
# 2.8284...
```

By using an instance of `InflationLP` one may similarly explore the Bell scenario with classical bipartite preparations.

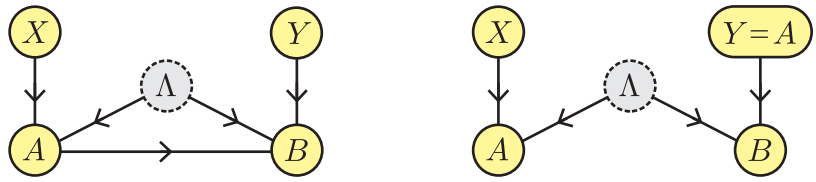
6.3.2 Scenarios with partial information

Inflation can also handle scenarios where not all the information about a particular distribution in the original scenario is known. An important example is the analysis of cryptographic scenarios, where the honest parties may know what is their joint distribution but they can not know what is the joint distribution with a potential adversary. One simple such scenario is considered in Ref. [Wol+21, Sec. VIII]. Specifying particular elements of a distribution in an `InflationSDP` object is achieved via the use of the function `set_values()`, which admits as input a dictionary where the keys are the variables to be assigned numerical quantities, and the corresponding values are the quantities themselves. In order to address the problem in Ref. [Wol+21, Sec. VIII] in Inflation, one would write:

```
p0 = sum(probability[0,b,0,0,0] for b in range(4))
mmnts = sdp.measurements
A = mmnts[0][0]
B = mmnts[0][0][0]
C = mmnts[2][0]
E = mmnts[3][0][0]
sdp.set_objective(A[0][0]*C[0][0]*E[0]/p0 - E[0])
known_values = {}
for a, b, c, x, z in numpy.ndindex(1, 3, 1, 2, 2):
    known_values[A[x][a]*B[b]*C[z][c]] = probability[a, b,
c, x, z]
# Proceed analogously for all other known quantities
sdp.set_values(known_values)
```

6.3.3 Scenarios beyond networks

So far, all the examples described have involved causal scenarios known as networks. These are bipartite DAGs with a layer of visible variables denoting the parties outcomes, and a layer of both latent and visible variables that denote the sources of physical systems and the measurements performed by the parties on them, respectively. Importantly, there are no connections between the nodes in each of the layers. Not all DAGs fall in this category, and some non-network DAGs have received considerable attention recently



(A) The instrumental scenario (B) Interruption of the instrumental scenario

FIGURE 6.1: 6.1a The instrumental scenario written as a DAG. Because of the arrow originating at A and pointing to B, the scenario is not a network. 6.1b The interruption of 6.1a. This is a network scenario, where the input to B and the output of A are restricted to coincide. The similarity of this scenario with the Bell scenario has motivated its investigation within quantum information theory.

in the literature. The most important example is the so-called instrumental scenario [Pea09], which has been extensively studied in the quantum information literature [VH+19; GMC20; Agr+20; Agr+22] (see Figure 6.1a). Inflation is capable of handling with causal inference problems in these scenarios, by internally considering equivalent network-type scenarios such as that depicted in Figure 6.1b (see [NW20a; Wol+21] for the details on the equivalence). The user experience in considering these problems is no different to that of considering causal inference over network-type DAGs. As an illustration, the following snippet recovers the bounds of Bonet's inequality in [VH+19, Eq. (23)].

```
from sympy import Symbol as Sym
inst = InflationProblem(dag={"rhoAB": ["A", "B"],
                             "A": ["B"]},
                        outcomes_per_party=(2, 2),
                        settings_per_party=(3, 1))
sdp = InflationSDP(inst)
sdp.generate_relaxation("local1")
objective = Sym("pAB(00|00)")
objective += Sym("pA(1|0)")    - Sym("pAB(10|00)")  # pAB
(11|0)
objective += Sym("pAB(00|10)") + Sym("pAB(10|10)")  # pB
(0|1)
objective += Sym("pA(0|2)")    - Sym("pAB(00|20)")  # pAB
(01|2)
sdp.set_objective(objective)
```

```
sdp.solve()
# 2.2071...
```

6.3.4 Feasibility based on distribution supports

The violation of Bell-type inequalities is the main method for the certification of non-classical behaviour. However, there exist even simpler certifications of non-classicality, that instead rely on possibilistic arguments: instead of setting bounds on combinations of the elements of compatible probability distributions, they only assume whether certain events are possible (positive probability) or impossible (zero probability) in order to reach a contradiction. These certificates are known as Hardy-type paradoxes [MF12].

Inflation can handle proofs of causal incompatibility with arbitrary DAGs based on possibilistic arguments. In the case of classical and/or no-signaling inflation, the corresponding linear programming relaxations may be written as find $q = q_{\text{const}} \oplus q_{(\geq 0)}$ such that $Aq = 0, q_{\text{const}} = 1, q_{(\geq 0)} \geq 0$ where q is the joint distribution over the inflated graph in Collins-Gisin notation. In this case q is expressed as the direct sum of the normalization variable in Collins-Gisin notation, q_{const}, and the rest of the variables which are non-negative, $q_{(\geq 0)}$. We may now consider a possibilistic constraint, whereby some elements of $q_{(\geq 0)}$ are restricted to be either exactly zero or strictly positive, while the rest are non-negative. We may thus write the possibilistic LP as find $q = q_{\text{const}} \oplus q_{(=0)} \oplus q_{(>0)} \oplus q_{(\geq 0)}$ such that $Aq = 0, q_{\text{const}} = 1, q_{(=0)} = 0, q_{(\geq 0)} > 0, q_{(\geq 0)} \geq 0$. Consider that this linear program is feasible, and let q^* be a solution satisfying the possibilistic constraints. Let us now consider the smallest of the variables constrained to be strictly positive, $\epsilon = \min(q^*_{(>0)})$, and let us define $\bar{q}^* := q^*/\epsilon$. This new rescales solution will satisfy the following constraints: $A\bar{q}^* = 0, \bar{q}^*_{\text{const}} \geq 1, \bar{q}^*_{(=0)} = 0, \bar{q}^*_{(\geq 0)} \geq 1, \bar{q}^*_{(\geq 0)} \geq 0$. Notice that the strict positivity constraint $q_{(\geq 0)} > 0$ is mapped under rescaling by ϵ to an inequality constraint, $\bar{q}^*_{(\geq 0)} \geq 1$. We may now define a rescaled possibilistic LP which looks for $\bar{q}$ satisfying the previous constraints; if the program is infeasible, then there cannot exist a q satisfying the possibilistic LP either. Furthermore, in case of feasibility of the rescaled possibilistic LP, one may descale $\bar{q}$ by the value of the rescaled normalization variable, $\bar{q} \to q := \bar{q}/\bar{q}_{\text{const}}$, which constructs a feasible solution to the original, possibilistic LP. As such, for possibilistic LP, rescaling is not a relaxation, but an equivalent transformation.

Similarly, in the case quantum inflation, it assesses whether an inflation exists (with commuting or non-commuting operators, respectively) where the probability elements inside the support are constrained to lie in the interval $[1, \infty)$ while those outside the support are given the value 0. This represents a re-scaled version of a standard quantum inflation moment matrix, $\bar{\Gamma} = \Gamma/\epsilon$, that does not suffer from floating-point instabilities when determining if a probability element is outside the support or has assigned a very small value. To get back the original moment matrix Γ, which contains probabilities as some of its cells, one just needs to divide $\bar{\Gamma}$ by the numerical value in its top-left corner, which corresponds to the re-scaled value of the expectation value for the identity, $\bar{\Gamma}_{00} = \langle \mathbb{1} \rangle / \epsilon$.

In order to deal with possibilistic feasibility problems in Inflation one sets the argument `supports_problem=True` when instantiating `InflationSDP` or `InflationSDP`. As an example, in order to recover the result that the distribution from [VH+19, Eq. (19)] is incompatible with the scenario of Figure 6.1a, one would run the following code.

```
sdp = InflationSDP(inst, supports_problem=True)
sdp.generate_relaxation("local1")
sdp.set_distribution(P_Eq19)
sdp.solve()
# "infeasible"
```

Note that, in contrast with other functionalities discussed in this section, optimization of objective functions is not possible when assessing the feasibility of distribution supports.

6.4 Additional library information

Here we provide further information concerning the development of the Inflation library.

6.4.1 Computational considerations

Inflation has been developed with speed and efficiency in mind. It uses just-in-time compilation through Numba [LPS15] in order to speed up core calculations, and dictionary caching to avoid needless function calls. This results in all examples in the documentation being executable on a standard laptop with 8 GB of RAM. Moment matrices of around 200 columns and

1500 free scalar variables can be generated in 5 seconds and the SDP solved in 3 seconds; those of around 2000 columns and around 2500 variables can be generated in 8 minutes and the SDP solved in 10 minutes, and those of around 20000 columns can be generated in 17 hours. In this last case, however, the SDP is too large to be solvable on a traditional desktop computer.[2] In order to solve these larger problems, a promising venue to pursue is using symmetries to block diagonalize the moment matrix. In the library documentation we provide an example using the MATLAB software `RepLAB` [RMMB21].

6.4.2 Future extensions

Inflation is a general technique that comprises a collection of routines specific to different types of physical systems. Moreover, the fact that it is a young technique makes it not unreasonable to expect that refinements and alternative hierarchies will be developed in the future. For these reasons, Inflation is built having modularity in mind, so that new functionalities are easy to implement.

The main feature of the current implementation of Inflation is the characterization of sets of quantum correlations in networks and certain non-network causal structures. Moreover, it is also capable of handling the characterization of sets of classical correlations, and it contains the necessary equipment to handle simple non-network scenarios. Subsequent releases will be focused on consolidating these capabilities, as well as adding functionalities to improve user experience and increase the range of problems and scenarios it can handle. Planned additions to the library include:

Arbitrary causal scenarios. One of the most exciting features of inflation methods is their ability to handle scenarios with latent-to-latent, visible-to-visible and visible-to-latent connections. By the application of unpacking and exogenization algorithms (see Refs. [NW20a; Wol+21] for their descriptions), probability distributions compatible with any causal structure encoded in a DAG can be transformed into and analysed as distributions compatible with an equivalent network-form DAG and satisfying additional equality constraints. While the library already allows handling scenarios with visible-to-visible connections, in future versions we plan to add support for scenarios

[2] Tested on a PC with an Intel Core i9-10900X CPU @ 3.70GHz and 32 GB of RAM.

with visible-to-latent connections. The automatic handling of this type of networks will mostly be integrated in the `InflationProblem` object, although it is expected that, due to the need of handling differently the cases of classical and quantum latent nodes (see [Wol+21, Figure 8]), processing is needed also further down in the pipeline.

Localizing matrices. The library currently focuses on operator constraints relevant for standard causal compatibility problems, namely it assumes Projection-Valued-Measurements (PVMs). One however might consider further restricting the algebra of operators, e.g., by assuming Pauli algebras, or consider more general algebras relevant for applications such as computing device-independent lower bounds on the conditional von Neumann entropy [BFF21], relevant for several device-independent protocols, or operator sub-algebras as used in convergent quantum inflation hierarchies [LGG23]. These constraints will be able to be imposed, in future versions of the library, either through so-called *localizing matrices*, or through repeated substitutions when computing monomial canonical forms if the constraints are simple enough, or by imposing extra symmetries to the semidefinite program.

Interfacing. Future plans include adding support for other optimizers widely available such as SDPT3 [TTT99], CVXPY [DB16], SCS [O'D+21] and Gurobi [Gur22] (this last one is especially interesting since it is not restricted to linear and semidefinite programming problems), and translating the problems generated to forms compatible with other optimization libraries such as PICOS [SS22] and CVXOPT [ADV15]. Interfacing with other tools, such as `SDPSymmetryReduction.jl` [BK22] for reducing the memory and computational load of the problems via exploitation of symmetries, and scalar extension [PK+19] for imposing the independence of variables in the inflation scenarios, will also prove useful. Currently, it is possible to export the problem in a MATLAB-compatible form that can be directly read out by `RepLAB` [RMMB21] in order to block diagonalize it.

6.4.3 Documentation for Inflation

The documentation of the library, which includes a user's guide and an API glossary, can be found online at https://ecboghiu.github.io/inflation.

The user's guide contains more information on the installation and the topics covered in this manuscript, as well as subjects not covered here; for example, more applications, tips on improving performance, and in-depth tutorials. The API glossary is automatically generated from the documentation comments written in the code and contains information about the public functions and classes defined.

6.4.4 Contribution guidelines

We welcome contributions to Inflation from the larger community interested in causality, quantum nonlocality, and software for quantum information theory. Contributions can come in the form of feedback about the library, feature requests, bug fixes, or code contributions (pull requests). Feedback and feature requests can be done by opening an issue on the Inflation GitHub repository. Bug fixes and other pull requests can be done by forking the Inflation source code, making changes, and then opening a pull request to the Inflation GitHub repository. Pull requests are peer-reviewed by Inflation's core developers to provide feedback and/or request changes.

Contributors are expected to adhere to Inflation development practices including style guidelines and unit tests. Tests are written with the UnitTest Python framework and are implemented outside the module. To test installation or changes, one can download the source code from the repository, and use standard UnitTest functions. For example, executing the following in a Unix terminal in the test folder runs all the tests:

```
python -m unittest -v
```

More details can be found in the Contribution guidelines documentation.

6.5 Concluding remarks

We have presented the first open-source implementation of the inflation framework for causal compatibility. Its focus is put in user experience and modularity, with the goal of being easy to use off the shelf while allowing for modifications needed by expert users.

While the current core implements the quantum inflation technique of Ref. [Wol+21], the library can be used off the shelf to characterize the sets of classical and quantum correlations in any network-type DAG, and non-network scenarios with visible-to-visible connections. After briefly mentioning

the principles behind inflation, we described the main components of the library and techniques to achieve tighter characterizations, and we illustrated its use in multiple problems of interest. Finally, we outlined different ways in which the library can be extended to accommodate problems of interest for the broad community interested in quantum nonlocality and causality, and described additional software information including support and contribution guidelines.

Chapter 7

Outlook

In this thesis we explored quantum advantage in causal structures beyond Bell's scenario:

- We gave a proposal for an experimental setup where a triangle causal network device-independently reveals the non-classicality of a single-photon entangled state using passive optical measurements alone, which is not possible in common-cause scenarios.
- We went beyond network-type scenarios by studying the broadcast scenario, which includes transformations in addition to preparations and measurements. In the broadcast scenario, we have strengthened known results for activation of nonlocality, extended the notion of EPR steering to the broadcast scenario and provided the first examples of activation of EPR steering in this scenario. In addition, we demonstrated that broadcasting enhances robustness of device-independent entanglement certification against white noise in the quantum state.
- We developed the Python library `Inflation` for testing causal compatibility of observed probability distributions with arbitrary causal scenarios via the inflation method. In addition, we made several improvements to the inflation technique, namely, we generalized it to hybrid scenarios with mixed classical, quantum or no-signaling preparations and expanded its range of applicability to include possibilistic proofs of nonclassicality.

These results highlight that already relatively small causal structures—in terms of number of nodes and preparations involved—provide an advantage over common-cause scenarios. Structures beyond networks, which involve

transformations, seem particularly promising to explore due to their capability to activate nonlocality. In contrast, network-type scenario have yet to demonstrate this capability. This motivates the study of more complex causal structures, involving more nodes and transformations, and the nonclassical phenomena emerging from these structures.

It is particularly important to understand when such nonclassical phenomena constitute "interesting" or "genuinely novel" examples of nonclassicality. For instance, if these phenomena can be traced back to standard Bell nonlocality, it can be argued that the more complex causal structure is not being fully utilized to achieve novel results. In addition, such a sophisticated structure might be considered an unnecessary complication if it does not contribute to extending beyond standard Bell nonlocality. In the following, we give an overview of various existing definitions and concepts related to genuinely novel nonclassicality, and highlight some interesting research directions.

What constitutes genuinely novel nonclassicality? We understand by "standard" nonclassicality the presence of correlations between measurement devices which are incompatible with classical common-cause explanations, that is, Bell nonlocality. In network-type causal scenarios, there are several independent common-causes correlating various measurement devices.

Let us consider the 3-party chain scenario involving two independent preparations of bipartite systems, with one measurement device having access to both preparations (cf. Figure 3.1). One quantum strategy involves preparing one maximally entangled two-qubit state and one arbitrary separable state; the quantum state is measured with the appropriate quantum observables such as to maximally violate the CHSH inequality (cf. Eq. (2.63)), whereas arbitrary measurements are performed on the separable state. While through this strategy one may generate correlations incompatible with the 3-party chain scenarios with classical preparations, arguably this nonclassicality is not novel, as it can be traced back to Bell nonlocality.

Interestingly, through a similar protocol, a well known Bell inequality characterizing the 3-party chain scenario (the "IJ" inequality [Bra+12]) can be maximally violated [PKGT22]. This shows that a violation of this inequality does not necessarily imply that the nonclassicality present in the correlations is novel. This motivated the definition of **full network nonlocality** of Ref. [PKGT22]. Full network nonlocality is a theory-independent

definition, as nonclassicality is not limited to quantum sources of correlations. Furthermore, it leads to linear programming tests based on the inflation technique [WSF19; NW20a] which, if they fail, certify that a given correlation is fully network nonlocal. These tests are noise robust, facilitating the experimental demonstration of full network nonlocality [Gu+23].

It can be argued that full network nonlocality does not capture all the desiderata for a definition of genuinely novel nonclassicality. For example, in the 3-party chain scenario one may consider two parallel Bell tests, where the device with access to the two preparations can perform a classical processing of the two outcomes of the parallel Bell tests. Such a protocol requires all sources to be nonclassical, however, it is simply a wiring of two Bell tests, and thus in some sense, does not constitute a novel example of nonclassicality.

This motivated the definition of **genuine network quantum nonlocality** [Šu+22], which is the set of quantum correlations realizable in a network where the measurement devices are only allowed to perform local measurements on the preparations they receive, and local wirings of their local measurement settings and outcomes. Some examples of correlations have been proven to be genuine network quantum nonlocal, by certifying, through self-testing [ŠB20], the use of entangling measurements [RKB18; BSS18] or the use of non-entangling measurements that cannot be implemented with local operations and local wirings [ŠB23].

Characterizing this set of genuine network quantum nonlocal correlations is very complex, and no general test exists to determine whether a given correlation is genuine network quantum nonlocal. The existing techniques based on self-testing, while powerful forms of certification, are technically challenging to prove and expensive in terms of experimental noise sensitivity, as self-testing typically requires close to ideal quantum sources to demonstrate. As such, it is of interest to develop more noise robust tests of genuine network quantum nonlocality. A particularly interesting research direction is to explore whether it is possible to use quantum inflation hierarchy [Wol+21], or the PPT hierarchy [Mor+13b], to device-independently certify the use of entangling measurements in a network-type causal scenario.

Other examples of new forms of nonclassicality have been proposed, where nonclassical correlations manifest in networks without the need for independent measurement choices [Ren+19a; PKGR23b; Abi+22]. This differs from standard Bell-type nonclassicality, where choosing measurement settings independently is critical to reveal nonclassical behaviour. However,

as Tobias Fritz demonstrated in Ref. [Fri12, Thm. 2.16], Bell nonlocality *can* be used to produce this form of network nonclassicality. Fritz provided an explicit example in a triangle network structure (cf. Fig. 3.8) where only one of the three preparations needs to be nonclassical to generate this form of nonclassicality. Currently, there is no established method nor commonly accepted definition to determine whether a given correlation in a network, where measurement choices are absent, is attributable merely to some underlying Bell-type nonclassicality. The self-testing method, while useful for proving genuine quantum network nonlocality, depends on having measurement choices and thus is not applicable in this context.

Lastly, any example of activation of nonlocality is a clear sign of quantum advantage, as the causal structure is exploited to achieve nonclassical correlations with resources that cannot generate nonclassicality in Bell scenarios. Causal scenarios such as broadcasting, which include transformations, can manifest activation. Network-type scenarios with many measurement devices are also known to manifest activation of nonlocality, however the scalability of the existing inflation techniques is currently inadequate for investigating these scenarios, as they require at least 21 measurement devices [Cav+11].

How to explore larger causal networks? The number of variables in the convex relaxation of fanout inflation scales superexponentially with the number of copies of each preparation [NW20a, Sec. 4.2]. Consequently there is a need to develop more scalable methods for evaluating causal compatibility. For quantum inflation, symmetries are currently used to reduce the number of variables in the convex relaxation. Further exploitation of these symmetries is possible, for instance, through transforming the basis to block-diagonalise the moment matrix (see Ref. [IR22] for a reference on symmetry exploitation within semidefinite programming). One may explore adding additional assumptions which make the inflation method more scalable, at the cost of losing generality and restricting the strength of the infeasibility certificates. For example, one may consider scenarios with stronger symmetries, such as *all* preparations being identical.

Alternatively, one may rely on heuristic methods to explore correlations in quantum causal scenarios, such as those based on machine learning-based constructions of local models [Kri+20]. One may use such heuristic techniques in combination with the inflation technique (cf. the see-saw

heuristic optimisation described in Appendix C.3) to search for examples of activation of nonlocality in general causal structures.

How to connect quantum advantage with practical applications? In this thesis we explored quantum advantage described in terms of an operational task where quantum resources outperform classical resources. However, these operational tasks need not be beneficial for technological purposes. For instance, despite various examples of activation of nonlocality, none have yet demonstrated a practical advantage for device-independent applications. An important open question is thus whether a quantum state with a local model can be used, via nonlocality activation techniques, to generate, for example, a non-zero key rate in DI quantum key distribution or certify DI private randomness generation. Towards this goal, the broadcast scenario seems promising as it uses a relatively small number of components in order to achieve activation of nonlocality.

Bibliography

[Å+20] Johan Åberg et al. “Semidefinite Tests for Quantum Network Topologies”. In: *Phys. Rev. Lett.* 125 (11 2020), p. 110505. DOI: 10.1103/PhysRevLett.125.110505. URL: https://link.aps.org/doi/10.1103/PhysRevLett.125.110505.

[Abi+22] Paolo Abiuso et al. “Single-photon nonlocality in quantum networks”. In: *Phys. Rev. Res.* 4 (1 Mar. 2022), p. L012041. DOI: 10.1103/PhysRevResearch.4.L012041. URL: https://link.aps.org/doi/10.1103/PhysRevResearch.4.L012041.

[Ac07] Antonio Acín et al. “Device-Independent Security of Quantum Cryptography against Collective Attacks”. In: *Phys. Rev. Lett.* 98 (23 June 2007), p. 230501. DOI: 10.1103/PhysRevLett.98.230501. URL: https://link.aps.org/doi/10.1103/PhysRevLett.98.230501.

[ADA14] Remigiusz Augusiak, Maciej Demianowicz, and Antonio Acín. “Local hidden–variable models for entangled quantum states”. In: *Journal of Physics A: Mathematical and Theoretical* 47.42 (Oct. 2014), p. 424002. DOI: 10.1088/1751-8113/47/42/424002. URL: https://doi.org/10.1088%2F1751-8113%2F47%2F42%2F424002.

[ADV15] Martin S. Andersen, Joachim Dahl, and Lieven Vandenberghe. *CVXOPT: Python software for convex optimization*. http://cvxopt.org/. May 2015.

[Agr+20] Iris Agresti et al. “Experimental device-independent certified randomness generation with an instrumental causal structure”. In: *Commun. Phys.* 3 (2020), p. 110. DOI: 10.1038/s42005-020-0375-6. arXiv: 1905.02027. URL: https://www.nature.com/articles/s42005-020-0375-6#citeas.

[Agr+22] Iris Agresti et al. “Experimental test of quantum causal influences”. In: *Sci. Adv.* 8.8 (2022), eabm1515. DOI: 10.1126/sciadv.abm1515. arXiv: 2108.08926. URL: https://www.science.org/doi/abs/10.1126/sciadv.abm1515.

[AGR81] Alain Aspect, Philippe Grangier, and Gérard Roger. “Experimental Tests of Realistic Local Theories via Bell’s Theorem”. In: *Phys. Rev. Lett.* 47 (7 Aug. 1981), pp. 460–463. DOI: 10.1103/PhysRevLett.47.460. URL: https://link.aps.org/doi/10.1103/PhysRevLett.47.460.

[AGR82] Alain Aspect, Philippe Grangier, and Gérard Roger. “Experimental Realization of Einstein Podolsky Rosen Bohm Gedankenexperiment: A New Violation of Bell’s Inequalities”. In: *Phys. Rev. Lett.* 49 (2 July 1982), pp. 91–94. DOI: 10.1103/PhysRevLett.49.91. URL: https://link.aps.org/doi/10.1103/PhysRevLett.49.91.

[AGT06] Antonio Acín, Nicolas Gisin, and Benjamin Toner. “Grothendieck’s constant and local models for noisy entangled quantum states”. In: *Phys. Rev. A* 73.6 (June 2006). ISSN: 1094-1622. DOI: 10.1103/physreva.73.062105. URL: http://dx.doi.org/10.1103/PhysRevA.73.062105.

[All+17] John-Mark A. Allen et al. “Quantum Common Causes and Quantum Causal Models”. In: *Physical Review X* 7.3 (July 2017). ISSN: 2160-3308. DOI: 10.1103/physrevx.7.031021. URL: http://dx.doi.org/10.1103/PhysRevX.7.031021.

[AM16] Antonio Acín and Lluis Masanes. “Certified randomness in quantum physics”. In: *Nature* 540.7632 (Dec. 2016), pp. 213–219. ISSN: 1476-4687. DOI: 10.1038/nature20119. arXiv: 1708.00265 [quant-ph]. URL: https://doi.org/10.1038/nature20119.

[ApS19] MOSEK ApS. *MOSEK Fusion API for Python*. https://docs.mosek.com/latest/pythonfusion/index.html. 2019.

[AV00] Yakir Aharonov and Lev Vaidman. “Nonlocal aspects of a quantum wave”. In: *Phys. Rev. A* 61 (5 Apr. 2000), p. 052108. DOI: 10.1103/PhysRevA.61.052108. URL: https://link.aps.org/doi/10.1103/PhysRevA.61.052108.

[Bac+17] Flavio Baccari et al. "Efficient Device-Independent Entanglement Detection for Multipartite Systems". In: *Phys. Rev. X* 7 (2 June 2017), p. 021042. DOI: 10.1103/PhysRevX.7.021042. arXiv: 1612.08551. URL: https://link.aps.org/doi/10.1103/PhysRevX.7.021042.

[Ban+13] Jean-Daniel Bancal et al. "Definitions of multipartite nonlocality". In: *Phys. Rev. A* 88.1 (2013), p. 014102. DOI: 10.1103/physreva.88.014102. arXiv: 1112.2626.

[Bar02] Jonathan Barrett. "Nonsequential positive-operator-valued measurements on entangled mixed states do not always violate a Bell inequality". In: *Phys. Rev. A* 65.4, 042302 (Apr. 2002), p. 042302. DOI: 10.1103/PhysRevA.65.042302. arXiv: quant-ph/0107045 [quant-ph].

[Bav+17] Jessica Bavaresco et al. "Most incompatible measurements for robust steering tests". In: *Phys. Rev. A* 96.2, 022110 (Aug. 2017), p. 022110. DOI: 10.1103/PhysRevA.96.022110. arXiv: 1704.02994 [quant-ph].

[BC90] Samuel L. Braunstein and Carlton M. Caves. "Wringing out better Bell inequalities". In: *Annals of Physics* 202.1 (1990), pp. 22–56. DOI: https://doi.org/10.1016/0003-4916(90)90339-P.

[BCB13] Jonatan Bohr Brask, Rafael Chaves, and Nicolas Brunner. "Testing nonlocality of a single photon without a shared reference frame". In: *Phys. Rev. A* 88 (1 July 2013), p. 012111. DOI: 10.1103/PhysRevA.88.012111. URL: https://link.aps.org/doi/10.1103/PhysRevA.88.012111.

[Bel64] John S. Bell. "On the Einstein-Podolsky-Rosen paradox". In: *Physics Physique Fizika* 1 (3 Nov. 1964), pp. 195–200. DOI: 10.1103/PhysicsPhysiqueFizika.1.195. URL: https://link.aps.org/doi/10.1103/PhysicsPhysiqueFizika.1.195.

[Ben+12] A. J. Bennet et al. "Arbitrarily Loss-Tolerant Einstein-Podolsky-Rosen Steering Allowing a Demonstration over 1 km of Optical Fiber with No Detection Loophole". In: *Phys. Rev. X* 2 (3 2012), p. 031003. DOI: 10.1103/PhysRevX.2.031003. URL: https://link.aps.org/doi/10.1103/PhysRevX.2.031003.

[BFF21] Peter Brown, Hamza Fawzi, and Omar Fawzi. *Device-independent lower bounds on the conditional von Neumann entropy*. 2021. arXiv: 2106.13692 [quant-ph].

[BG03] Helle Bechmann-Pasquinucci and Nicolas Gisin. "Intermediate states in quantum cryptography and Bell inequalities". In: *Phys. Rev. A* 67.6, 062310 (June 2003), p. 062310. DOI: 10.1103/PhysRevA.67.062310. arXiv: quant-ph/0208031 [quant-ph].

[BG08] Nicolas Brunner and Nicolas Gisin. "Partial list of bipartite Bell inequalities with four binary settings". In: *Physics Letters A* 372.18 (Apr. 2008), pp. 3162–3167. DOI: 10.1016/j.physleta.2008.01.052. arXiv: 0711.3362 [quant-ph].

[BGP10] Cyril Branciard, Nicolas Gisin, and Stefano Pironio. "Characterizing the Nonlocal Correlations Created via Entanglement Swapping". In: *Phys. Rev. Lett.* 104 (17 Apr. 2010), p. 170401. DOI: 10.1103/PhysRevLett.104.170401. URL: https://link.aps.org/doi/10.1103/PhysRevLett.104.170401.

[BHC21] Joseph Bowles, Flavien Hirsch, and Daniel Cavalcanti. "Single-copy activation of Bell nonlocality via broadcasting of quantum states". In: *Quantum* 5 (July 2021), p. 499. ISSN: 2521-327X. DOI: 10.22331/q-2021-07-13-499. arXiv: 2007.16034. URL: http://dx.doi.org/10.22331/q-2021-07-13-499.

[BK22] Daniel Brosch and Etienne de Klerk. "Jordan symmetry reduction for conic optimization over the doubly nonnegative cone: theory and software". In: *Optim. Methods Softw.* 37.6 (2022), pp. 2001–2020. DOI: 10.1080/10556788.2021.2022146. arXiv: 2001.11348. URL: https://doi.org/10.1080/10556788.2021.2022146.

[BKP+16] Sabine Burgdorf, Igor Klep, Janez Povh, et al. *Optimization of polynomials in non-commuting variables*. Vol. 2. Springer, 2016.

[BLO20] Jonathan Barrett, Robin Lorenz, and Ognyan Oreshkov. *Quantum Causal Models*. 2020. arXiv: 1906.10726 [quant-ph].

[BLO21] Jonathan Barrett, Robin Lorenz, and Ognyan Oreshkov. "Cyclic quantum causal models". In: *Nature communications* 12.1 (2021), p. 885.

[Bog+23] Emanuel-Cristian Boghiu et al. "Device independent and semi device independent entanglement certification in broadcast Bell scenarios". In: *SciPost Phys. Core* 6 (2023), p. 028. DOI: 10.21468/SciPostPhysCore.6.2.028. URL: https://scipost.org/10.21468/SciPostPhysCore.6.2.028.

[Bor+23a] Sadra Boreiri et al. *Noise-robust proofs of quantum network nonlocality*. 2023. arXiv: 2311.02182 [quant-ph].

[Bor+23b] Sadra Boreiri et al. "Towards a minimal example of quantum nonlocality without inputs". In: *Phys. Rev. A* 107 (6 2023), p. 062413. DOI: 10.1103/PhysRevA.107.062413. URL: https://link.aps.org/doi/10.1103/PhysRevA.107.062413.

[Bow+16] Joseph Bowles et al. "Sufficient criterion for guaranteeing that a two-qubit state is unsteerable". In: *Phys. Rev. A* 93.2, 022121 (Feb. 2016), p. 022121. DOI: 10.1103/PhysRevA.93.022121. arXiv: 1510.06721 [quant-ph].

[Boy20] Robert W Boyd. *Nonlinear optics*. Academic press, 2020.

[BR22] Salman Beigi and Marc-Olivier Renou. "Covariance Decomposition as a Universal Limit on Correlations in Networks". In: *IEEE Transactions on Information Theory* 68.1 (2022), pp. 384–394. DOI: 10.1109/TIT.2021.3119651.

[Bra+12] Cyril Branciard et al. "Bilocal versus nonbilocal correlations in entanglement-swapping experiments". In: *Phys. Rev. A* 85 (3 Mar. 2012), p. 032119. DOI: 10.1103/PhysRevA.85.032119. arXiv: 1112.4502. URL: https://link.aps.org/doi/10.1103/PhysRevA.85.032119.

[Bru+07] Nicolas Brunner et al. "Detection Loophole in Asymmetric Bell Experiments". In: *Phys. Rev. Lett.* 98 (22 May 2007), p. 220403. DOI: 10.1103/PhysRevLett.98.220403. URL: https://link.aps.org/doi/10.1103/PhysRevLett.98.220403.

[Bru+14] Nicolas Brunner et al. "Bell nonlocality". In: *Reviews of Modern Physics* 86.2 (Apr. 2014), pp. 419–478. DOI: 10.1103/revmodphys.86.419. URL: https://doi.org/10.1103%2Frevmodphys.86.419.

[BSS18] Jean-Daniel Bancal, Nicolas Sangouard, and Pavel Sekatski. "Noise-Resistant Device-Independent Certification of Bell State Measurements". In: *Phys. Rev. Lett.* 121 (25 2018), p. 250506. DOI: 10.1103/PhysRevLett.121.250506. URL: https://link.aps.org/doi/10.1103/PhysRevLett.121.250506.

[BV04] Stephen P. Boyd and Lieven Vandenberghe. *Convex optimization*. Cambridge university press, 2004.

[BV18] Erika Bene and Tamás Vértesi. "Measurement incompatibility does not give rise to Bell violation in general". In: *New journal of Physics* 20.1 (2018), p. 013021.

[BWPK] Emanuel-Cristian Boghiu, Elie Wolfe, and Alejandro Pozas-Kerstjens. "Source code for `inflation`". Zenodo **7305544** (2022). DOI: 10.5281/zenodo.7305544.

[BWPK23] Emanuel-Cristian Boghiu, Elie Wolfe, and Alejandro Pozas-Kerstjens. "Inflation: a Python library for classical and quantum causal compatibility". In: *Quantum* 7 (May 2023), p. 996. ISSN: 2521-327X. DOI: 10.22331/q-2023-05-04-996. URL: https://doi.org/10.22331/q-2023-05-04-996.

[Cam+23] Giulio Camillo et al. *Estimating the volumes of correlations sets in causal networks*. 2023. arXiv: 2311.08574 [quant-ph].

[Cao+22] Huan Cao et al. "Experimental Demonstration that No Tripartite-Nonlocal Causal Theory Explains Nature's Correlations". In: *Phys. Rev. Lett.* 129 (15 2022), p. 150402. DOI: 10.1103/PhysRevLett.129.150402. URL: https://link.aps.org/doi/10.1103/PhysRevLett.129.150402.

[Cas+20] Patrik Caspar et al. "Heralded distribution of single-photon path entanglement". In: *arXiv preprint arXiv:2004.09465* (2020).

[Cav+11] Daniel Cavalcanti et al. "Quantum networks reveal quantum nonlocality". In: *Nature Communications* 2, 184 (Feb. 2011), p. 184. DOI: 10.1038/ncomms1193. arXiv: 1010.0900 [quant-ph].

[CG04] Daniel Collins and Nicolas Gisin. "A relevant two qubit Bell inequality inequivalent to the CHSH inequality". In: *Journal of Physics A: Mathematical and General* 37.5 (Jan. 2004), 1775–1787. ISSN: 1361-6447. DOI: 10.1088/0305-4470/37/5/021. URL: http://dx.doi.org/10.1088/0305-4470/37/5/021.

[Cha16] Rafael Chaves. "Polynomial Bell Inequalities". In: *Phys. Rev. Lett.* 116 (1 2016), p. 010402. DOI: 10.1103/PhysRevLett.116.010402. URL: https://link.aps.org/doi/10.1103/PhysRevLett.116.010402.

[Cha+17] Rafael Chaves et al. "Quantum violation of an instrumental test". In: *Nature Physics* 14.3 (Dec. 2017), 291–296. ISSN: 1745-2481. DOI: 10.1038/s41567-017-0008-5. URL: http://dx.doi.org/10.1038/s41567-017-0008-5.

[Cha+18] H. Chau Nguyen et al. "Quantum steering with positive operator valued measures". In: *Journal of Physics A Mathematical General* 51.35, 355302 (Aug. 2018), p. 355302. DOI: 10.1088/1751-8121/aad115. arXiv: 1706.08166 [quant-ph].

[Chi+13] Giulio Chiribella et al. "Quantum computations without definite causal structure". In: *Phys. Rev. A* 88 (2 Aug. 2013), p. 022318. DOI: 10.1103/PhysRevA.88.022318. URL: https://link.aps.org/doi/10.1103/PhysRevA.88.022318.

[Cho75a] Man-Duen Choi. "Completely positive linear maps on complex matrices". In: *Linear algebra and its applications* 10.3 (1975), pp. 285–290.

[Cho75b] Man-Duen Choi. "Completely positive linear maps on complex matrices". In: *Linear Algebra and its Applications* 10.3 (1975), 285–290. DOI: 10.1016/0024-3795(75)90075-0. URL: http://dx.doi.org/10.1016/0024-3795(75)90075-0.

[Cho75c] Man-Duen Choi. "Positive semidefinite biquadratic forms". In: *Linear Algebra and its applications* 12.2 (1975), pp. 95–100.

[Cla+69] John F. Clauser et al. "Proposed experiment to test local hidden-variable theories". In: *Physical review letters* 23.15 (1969), p. 880.

[Col07] Roger Colbeck. *Quantum and Relativistic Protocols For Secure Multi-Party Computation*. 2007. eprint: Ph.D.thesis,UniversityofCambridge,arXiv:0911.3814.

[Cou18] Christophe Couteau. "Spontaneous parametric down-conversion". In: *Contemporary Physics* 59.3 (2018), pp. 291–304.

[Cox01] Richard T Cox. *Algebra of probable inference*. Johns Hopkins University Press, 2001.

[CQK23] Adán Cabello, Marco Túlio Quintino, and Matthias Kleinmann. *Logical possibilities for physics after MIP*=RE*. 2023. arXiv: 2307.02920 [quant-ph].

[CS12] Andreas Christ and Christine Silberhorn. "Limits on the deterministic creation of pure single-photon states using parametric down-conversion". In: *Physical Review A* 85.2 (2012), p. 023829.

[CS16] Daniel Cavalcanti and Paul Skrzypczyk. "Quantum steering: a review with focus on semidefinite programming". In: *Reports on Progress in Physics* 80.2 (Dec. 2016), p. 024001. DOI: 10.1088/1361-6633/80/2/024001. URL: https://doi.org/10.1088%2F1361-6633%2F80%2F2%2F024001.

[CV+15] Valentina Caprara Vivoli et al. "Challenging preconceptions about Bell tests with photon pairs". In: *Phys. Rev. A* 91 (1 Jan. 2015), p. 012107. DOI: 10.1103/PhysRevA.91.012107. URL: https://link.aps.org/doi/10.1103/PhysRevA.91.012107.

[CV16] H. Chau Nguyen and Thanh Vu. "Necessary and sufficient condition for steerability of two-qubit states by the geometry of steering outcomes". In: *EPL (Europhysics Letters)* 115.1 (July 2016), p. 10003. DOI: 10.1209/0295-5075/115/10003. arXiv: 1604.03815 [quant-ph].

[D'A+06] Milena D'Angelo et al. "Tomographic test of Bell's inequality for a time-delocalized single photon". In: *Phys. Rev. A* 74 (5 Nov. 2006), p. 052114. DOI: 10.1103/PhysRevA.74.052114. URL: https://link.aps.org/doi/10.1103/PhysRevA.74.052114.

[Dam05] Wim van Dam. *Implausible Consequences of Superstrong Nonlocality*. 2005. arXiv: quant-ph/0501159 [quant-ph].

[Das+21] Tamoghna Das et al. "Can single photon excitation of two spatially separated modes lead to a violation of Bell inequality via weak-field homodyne measurements?" In: *New Journal of Physics* 23.7 (July 2021), p. 073042. ISSN: 1367-2630. DOI: 10.1088/1367-2630/ac0ffe. URL: http://dx.doi.org/10.1088/1367-2630/ac0ffe.

[DB16] Steven Diamond and Stephen Boyd. "CVXPY: A Python-embedded modeling language for convex optimization". In: *J. Mach. Learn. Res.* 17.83 (2016), pp. 1–5. DOI: 10.48550/arXiv.1603.00943. arXiv: 1603.00943. URL: http://jmlr.org/papers/v17/15-408.html.

[DBV17] Péter Diviánszky, Erika Bene, and Tamás Vértesi. "Qutrit witness from the Grothendieck constant of order four". In: *Phys. Rev. A* 96.1 (July 2017). ISSN: 2469-9934. DOI: 10.1103/physreva.96.012113. URL: http://dx.doi.org/10.1103/PhysRevA.96.012113.

[DCP17] Giacomo Mauro D'Ariano, Giulio Chiribella, and Paolo Perinotti. *Quantum Theory from First Principles: An Informational Approach*. Cambridge University Press, 2017. DOI: 10.1017/9781107338340.

[DCPSM04] Andrew Doherty C., Pablo A. Parrilo, and Federico Spedalieri M. "Complete family of separability criteria". In: *Physical Review A* 69.2 (Feb. 2004). DOI: 10.1103/physreva.69.022308. URL: https://doi.org/10.1103%2Fphysreva.69.022308.

[Des+23] Sébastien Designolle et al. "Improved local models and new Bell inequalities via Frank-Wolfe algorithms". In: *Physical Review Research* 5.4 (Oct. 2023). ISSN: 2643-1564. DOI: 10.11

03/physrevresearch.5.043059. URL: http://dx.doi.org/10.1103/PhysRevResearch.5.043059.

[Don+14] Gaia Donati et al. "Observing optical coherence across Fock layers with weak-field homodyne detectors". In: *Nature Communications* 5.1 (2014), p. 5584. DOI: 10.1038/ncomms6584. URL: https://doi.org/10.1038/ncomms6584.

[DPS02] A. C. Doherty, Pablo A. Parrilo, and Federico M. Spedalieri. "Distinguishing Separable and Entangled States". In: *Phys. Rev. Lett.* 88 (18 2002), p. 187904. DOI: 10.1103/PhysRevLett.88.187904. URL: https://link.aps.org/doi/10.1103/PhysRevLett.88.187904.

[DPS04] Andrew C. Doherty, Pablo A. Parrilo, and Federico M. Spedalieri. "Complete family of separability criteria". In: *Phys. Rev. A* 69 (2 2004), p. 022308. DOI: 10.1103/PhysRevA.69.022308. URL: https://link.aps.org/doi/10.1103/PhysRevA.69.022308.

[Ebe93] Philippe H. Eberhard. "Background level and counter efficiencies required for a loophole-free Einstein-Podolsky-Rosen experiment". In: *Phys. Rev. A* 47 (2 Feb. 1993), R747–R750. DOI: 10.1103/PhysRevA.47.R747. URL: https://link.aps.org/doi/10.1103/PhysRevA.47.R747.

[Eke91] Artur K. Ekert. "Quantum cryptography based on Bell's theorem". In: *Phys. Rev. Lett.* 67 (6 Aug. 1991), pp. 661–663. DOI: 10.1103/PhysRevLett.67.661. URL: https://link.aps.org/doi/10.1103/PhysRevLett.67.661.

[EPR35] Albert Einstein, Boris Podolsky, and Nathan Rosen. "Can Quantum-Mechanical Description of Physical Reality Be Considered Complete?" In: *Phys. Rev.* 47 (10 May 1935), pp. 777–780. DOI: 10.1103/PhysRev.47.777. URL: https://link.aps.org/doi/10.1103/PhysRev.47.777.

[FC72] Stuart J. Freedman and John F. Clauser. "Experimental test of local hidden-variable theories". In: *Physical Review Letters* 28.14 (1972), p. 938.

[Fin82] Arthur Fine. "Hidden Variables, Joint Probability, and the Bell Inequalities". In: *Phys. Rev. Lett.* 48 (5 1982), pp. 291–295. DOI: 10.1103/PhysRevLett.48.291. URL: https://link.aps.org/doi/10.1103/PhysRevLett.48.291.

[Fri12] Tobias Fritz. "Beyond Bell's theorem: correlation scenarios". In: *New Journal of Physics* 14.10 (Oct. 2012), p. 103001. ISSN: 1367-2630. DOI: 10.1088/1367-2630/14/10/103001. URL: http://dx.doi.org/10.1088/1367-2630/14/10/103001.

[Fri16] Tobias Fritz. "Beyond Bell's Theorem II: Scenarios with arbitrary causal structure". In: *Comm. Math. Phys.* 341 (2 2016), pp. 391–434. DOI: 10.1007/s00220-015-2495-5. URL: https://link.springer.com/article/10.1007/s00220-015-2495-5.

[Fuk+11] Daiji Fukuda et al. "Titanium-based transition-edge photon number resolving detector with 98% detection efficiency with index-matched small-gap fiber coupling". In: *Optics express* 19.2 (2011), pp. 870–875.

[Gal+12] Rodrigo Gallego et al. "Operational Framework for Nonlocality". In: *Phys. Rev. Lett.* 109.7, 070401 (Aug. 2012), p. 070401. DOI: 10.1103/PhysRevLett.109.070401. arXiv: 1112.2647 [quant-ph].

[Gal+14] Rodrigo Gallego et al. "Nonlocality in sequential correlation scenarios". In: *New Journal of Physics* 16.3 (Mar. 2014), p. 033037. ISSN: 1367-2630. DOI: 10.1088/1367-2630/16/3/033037. URL: http://dx.doi.org/10.1088/1367-2630/16/3/033037.

[Ger96] Christopher C. Gerry. "Nonlocality of a single photon in cavity QED". In: *Phys. Rev. A* 53 (6 June 1996), pp. 4583–4586. DOI: 10.1103/PhysRevA.53.4583. URL: https://link.aps.org/doi/10.1103/PhysRevA.53.4583.

[GG23] Antoine Girardin and Nicolas Gisin. "Violation of the Finner inequality in the four-output triangle network". In: *Phys. Rev. A* 108 (4 2023), p. 042213. DOI: 10.1103/PhysRevA.108.042213. URL: https://link.aps.org/doi/10.1103/PhysRevA.108.042213.

[Gho+23] Priya Ghosh et al. *Measurement incompatibility at remote entangled parties is insufficient for Bell nonlocality in two-input and two-output setting*. 2023. arXiv: 2312.15705 [quant-ph].

[Gis19] Nicolas Gisin. "Entanglement 25 years after quantum teleportation: Testing joint measurements in quantum networks". In: *Entropy* 21.3 (2019), p. 325.

[Gis+20a] Nicolas Gisin et al. "Constraints on nonlocality in networks from no-signaling and independence". In: *Nat. Commun.* 11.1 (May 2020), p. 2378. ISSN: 2041-1723. DOI: 10.1038/s41467-020-16137-4. arXiv: 1906.06495. URL: https://doi.org/10.1038/s41467-020-16137-4.

[Gis+20b] Nicolas Gisin et al. "Constraints on nonlocality in networks from no-signaling and independence". In: *Nature communications* 11.1 (2020), p. 2378.

[Git] *https://github.com/josephbowles/broadcast_nl*. URL: https://github.com/josephbowles/broadcast_nl.

[Giu+15a] Marissa Giustina et al. "Significant-Loophole-Free Test of Bell's Theorem with Entangled Photons". In: *Phys. Rev. Lett.* 115 (25 Dec. 2015), p. 250401. DOI: 10.1103/PhysRevLett.115.250401. URL: https://link.aps.org/doi/10.1103/PhysRevLett.115.250401.

[Giu+15b] Marissa Giustina et al. "Significant-Loophole-Free Test of Bell's Theorem with Entangled Photons". In: *Phys. Rev. Lett.* 115 (25 Dec. 2015), p. 250401. DOI: 10.1103/PhysRevLett.115.250401. URL: https://link.aps.org/doi/10.1103/PhysRevLett.115.250401.

[GMC20] Mariami Gachechiladze, Nikolai Miklin, and Rafael Chaves. "Quantifying Causal Influences in the Presence of a Quantum Common Cause". In: *Phys. Rev. Lett.* 125 (23 Dec. 2020), p. 230401. DOI: 10.1103/PhysRevLett.125.230401. arXiv: 2007.01221. URL: https://link.aps.org/doi/10.1103/PhysRevLett.125.230401.

[GPS21] Valentin Gebhart, Luca Pezzè, and Augusto Smerzi. "Genuine multipartite nonlocality with causal-diagram postselection". In: *arXiv preprint arXiv:2104.10069* (2021).

[Gro56] Alexandre Grothendieck. *Résumé de la théorie métrique des produits tensoriels topologiques*. Soc. de Matemática de São Paulo, 1956. URL: http://cm2vivi2002.free.fr/AG-biblio/AG-22.pdf.

[Gu+23] Xue-Mei Gu et al. "Experimental Full Network Nonlocality with Independent Sources and Strict Locality Constraints". In: *Phys. Rev. Lett.* 130 (19 2023), p. 190201. DOI: 10.1103/PhysRevLett.130.190201. URL: https://link.aps.org/doi/10.1103/PhysRevLett.130.190201.

[Gur22] Gurobi Optimization, LLC. *Gurobi Optimizer Reference Manual*. https://www.gurobi.com. 2022.

[Har09] Lucien Hardy. "Quantum Gravity Computers: On the Theory of Computation with Indefinite Causal Structure". In: *Quantum Reality, Relativistic Causality, and Closing the Epistemic Circle*. Springer Netherlands, 2009, 379–401. ISBN: 9781402091070. DOI: 10.1007/978-1-4020-9107-0_21. URL: http://dx.doi.org/10.1007/978-1-4020-9107-0_21.

[Har94] Lucien Hardy. "Nonlocality of a Single Photon Revisited". In: *Phys. Rev. Lett.* 73 (17 Oct. 1994), pp. 2279–2283. DOI: 10.1103/PhysRevLett.73.2279. URL: https://link.aps.org/doi/10.1103/PhysRevLett.73.2279.

[Hen+15] Bas Hensen et al. "Loophole-free Bell inequality violation using electron spins separated by 1.3 kilometres". In: *Nature* 526.7575 (2015), pp. 682–686. DOI: 10.1038/nature15759. URL: https://doi.org/10.1038/nature15759.

[Hes+04] Björn Hessmo et al. "Experimental Demonstration of Single Photon Nonlocality". In: *Phys. Rev. Lett.* 92 (18 May 2004), p. 180401. DOI: 10.1103/PhysRevLett.92.180401. URL: https://link.aps.org/doi/10.1103/PhysRevLett.92.180401.

[Hir+13] Flavien Hirsch et al. "Genuine Hidden Quantum Nonlocality". In: *Phys. Rev. Lett.* 111.16, 160402 (Oct. 2013), p. 160402. DOI: 10.1103/PhysRevLett.111.160402. arXiv: 1307.4404 [quant-ph].

[Hir+16] Flavien Hirsch et al. “Entanglement without hidden nonlocality”. In: *New Journal of Physics* 18.11, 113019 (Nov. 2016), p. 113019. DOI: 10.1088/1367-2630/18/11/113019. arXiv: 1606.02215 [quant-ph].

[Hir+17] Flavien Hirsch et al. “Better local hidden variable models for two-qubit Werner states and an upper bound on the Grothendieck constant KG(3)”. In: *Quantum* 1 (Apr. 2017), p. 3. ISSN: 2521-327X. DOI: 10.22331/q-2017-04-25-3. arXiv: 1609.06114 [quant-ph]. URL: http://dx.doi.org/10.22331/q-2017-04-25-3.

[HMW+20] Charles R. Harris, K. Jarrod Millman, Stéfan J. van der Walt, et al. “Array programming with NumPy”. In: *Nature* 585.7825 (Sept. 2020), pp. 357–362. DOI: 10.1038/s41586-020-2649-2. URL: https://doi.org/10.1038/s41586-020-2649-2.

[HQB18] Flavien Hirsch, Marco Túlio Quintino, and Nicolas Brunner. “Quantum measurement incompatibility does not imply Bell nonlocality”. In: *Phys. Rev. A* 97 (1 2018), p. 012129. DOI: 10.1103/PhysRevA.97.012129. URL: https://link.aps.org/doi/10.1103/PhysRevA.97.012129.

[IR22] Marie Ioannou and Denis Rosset. *Noncommutative polynomial optimization under symmetry*. 2022. arXiv: 2112.10803 [quant-ph].

[Jay03] Edwin T Jaynes. *Probability theory: The logic of science*. Cambridge university press, 2003.

[Jev+15] Sania Jevtic et al. “Einstein-Podolsky-Rosen steering and the steering ellipsoid”. In: *Journal of the Optical Society of America B Optical Physics* 32.4 (Apr. 2015), A40. DOI: 10.1364/JOSAB.32.000A40. arXiv: 1411.1517 [quant-ph].

[Ji+22] Zhengfeng Ji et al. *MIP*=RE*. 2022. arXiv: 2001.04383 [quant-ph].

[JM05] Nick S. Jones and Lluís Masanes. “Interconversion of nonlocal correlations”. In: *Phys. Rev. A* 72.5, 052312 (Nov. 2005), p. 052312. DOI: 10.1103/PhysRevA.72.052312. arXiv: quant-ph/0506182 [quant-ph].

[Kel+20] Aditya Kela et al. “Semidefinite Tests for Latent Causal Structures”. In: *IEEE Transactions on Information Theory* 66.1 (2020), pp. 339–349. DOI: 10.1109/TIT.2019.2935755.

[Kim08] H. J. Kimble. “The quantum internet”. In: *Nature* 453.7198 (June 2008), 1023–1030. ISSN: 1476-4687. DOI: 10.1038/nature07127. URL: http://dx.doi.org/10.1038/nature07127.

[Kra+21] Tristan Kraft et al. “Characterizing quantum networks: Insights from coherence theory”. In: *Phys. Rev. A* 103 (5 2021), p. 052405. DOI: 10.1103/PhysRevA.103.052405. URL: https://link.aps.org/doi/10.1103/PhysRevA.103.052405.

[Kri+20] Tamás Kriváchy et al. “A neural network oracle for quantum nonlocality problems in networks”. In: *npj Quantum Information* 6.1 (2020), pp. 1–7. DOI: https://doi.org/10.1038/s41534-020-00305-x. URL: https://doi.org/10.1038/s41534-020-00305-x.

[KW19] Wojciech Kozlowski and Stephanie Wehner. “Towards Large-Scale Quantum Networks”. In: *Proceedings of the Sixth Annual ACM International Conference on Nanoscale Computing and Communication*. NANOCOM ’19. ACM, Sept. 2019. DOI: 10.1145/3345312.3345497. URL: http://dx.doi.org/10.1145/3345312.3345497.

[LGG23] Laurens T. Ligthart, Mariami Gachechiladze, and David Gross. “A Convergent Inflation Hierarchy for Quantum Causal Structures”. In: *Communications in Mathematical Physics* 401.3 (Aug. 2023), pp. 2673–2714. ISSN: 1432-0916. DOI: 10.1007/s00220-023-04697-7. URL: https://doi.org/10.1007/s00220-023-04697-7.

[Lic] *The MIT License*. URL: https://opensource.org/licenses/MIT.

[LMN08] Adriana E Lita, Aaron J Miller, and Sae Woo Nam. “Counting near-infrared single-photons with 95% efficiency”. In: *Optics express* 16.5 (2008), pp. 3032–3040.

[Löf04] Johann Löfberg. "YALMIP: A Toolbox for Modeling and Optimization in MATLAB". In: *Proceedings of the CACSD Conference*. Taipei, Taiwan, 2004. URL: https://yalmip.github.io/.

[LPS15] Siu Kwan Lam, Antoine Pitrou, and Stanley Seibert. "Numba: A LLVM-Based Python JIT Compiler". In: *Proceedings of the Second Workshop on the LLVM Compiler Infrastructure in HPC*. LLVM '15. Austin, Texas: Association for Computing Machinery, 2015. ISBN: 9781450340052. DOI: 10.1145/2833157.2833162. URL: https://doi.org/10.1145/2833157.2833162.

[Luo21] Ming-Xing Luo. *Network configuration theory for all networks*. 2021. arXiv: 2107.05846 [quant-ph].

[Mao+22] Ya-Li Mao et al. "Test of Genuine Multipartite Nonlocality". In: *Phys. Rev. Lett.* 129 (15 2022), p. 150401. DOI: 10.1103/PhysRevLett.129.150401. URL: https://link.aps.org/doi/10.1103/PhysRevLett.129.150401.

[MBV23] István Márton, Erika Bene, and Tamás Vértesi. "Bounding the detection efficiency threshold in Bell tests using multiple copies of the maximally entangled two-qubit state carried by a single pair of particles". In: *Phys. Rev. A* 107 (2 Feb. 2023), p. 022205. DOI: 10.1103/PhysRevA.107.022205. arXiv: 2103.10413 [quant-ph]. URL: https://link.aps.org/doi/10.1103/PhysRevA.107.022205.

[Mer90] N. David Mermin. "Quantum mysteries revisited". In: *Amer. J. Phys.* 58.8 (Aug. 1990), pp. 731–734. DOI: 10.1119/1.16503. URL: https://doi.org/10.1119/1.16503.

[MF12] Shane Mansfield and Tobias Fritz. "Hardy's Non-locality Paradox and Possibilistic Conditions for Non-locality". In: *Found. Phys.* 42.5 (Mar. 2012), pp. 709–719. DOI: 10.1007/s10701-012-9640-1. arXiv: 1105.1819. URL: https://doi.org/10.1007%2Fs10701-012-9640-1.

[Mil+11] Aaron J Miller et al. "Compact cryogenic self-aligning fiber-to-detector coupling with losses below one percent". In: *Optics express* 19.10 (2011), pp. 9102–9110.

[Mor+13a] Olivier Morin et al. "Witnessing Trustworthy Single-Photon Entanglement with Local Homodyne Measurements". In: *Phys. Rev. Lett.* 110 (13 Mar. 2013), p. 130401. DOI: 10.1103/PhysRevLett.110.130401. URL: https://link.aps.org/doi/10.1103/PhysRevLett.110.130401.

[Mor+13b] Tobias Moroder et al. "Device-Independent Entanglement Quantification and Related Applications". In: *Phys. Rev. Lett.* 111.3 (July 2013). ISSN: 1079-7114. DOI: 10.1103/physrevlett.111.030501. URL: http://dx.doi.org/10.1103/PhysRevLett.111.030501.

[Mos+19] Maria Moshkova et al. "High-performance superconducting photon-number-resolving detectors with 86% system efficiency at telecom range". In: *JOSA B* 36.3 (2019), B20–B25.

[MSP+17] Aaron Meurer, Christopher P. Smith, Mateusz Paprocki, et al. "SymPy: symbolic computing in Python". In: *PeerJ Comput. Sci.* 3 (Jan. 2017), e103. ISSN: 2376-5992. DOI: 10.7717/peerj-cs.103. URL: https://doi.org/10.7717/peerj-cs.103.

[MY04] Dominic Mayers and Andrew Yao. "Self Testing Quantum Apparatus". In: *Quantum Info. Comput.* 4.4 (July 2004), 273–286. ISSN: 1533-7146.

[MY98] Dominic Mayers and Andrew Yao. "Quantum Cryptography with Imperfect Apparatus". In: *Proceedings of the 39th Annual Symposium on Foundations of Computer Science*. FOCS '98. USA: IEEE Computer Society, 1998, p. 503. ISBN: 0818691727.

[Nev22] Eva Nevelius. *The Nobel Prize in Physics 2022 [Press release]*. https://www.nobelprize.org/prizes/physics/2022/summary/. Accessed: 2023-10-10. Oct. 4, 2022.

[Nga+15] Lutfi Arif Ngah et al. "Ultra-fast heralded single photon source based on telecom technology". In: *Laser & Photonics Reviews* 9.2 (2015), pp. L1–L5.

[NPA07] Miguel Navascués, Stefano Pironio, and Antonio Acín. "Bounding the Set of Quantum Correlations". In: *Phys. Rev. Lett.* 98 (1 2007), p. 010401. DOI: 10.1103/PhysRevLett.98.010401. arXiv: quant-ph/0607119. URL: https://link.aps.org/doi/10.1103/PhysRevLett.98.010401.

[NPA08] Miguel Navascués, Stefano Pironio, and Antonio Acín. "A convergent hierarchy of semidefinite programs characterizing the set of quantum correlations". In: *New Journal of Physics* 10.7, 073013 (July 2008), p. 073013. DOI: 10.1088/1367-2630/10/7/073013. arXiv: 0803.4290 [quant-ph].

[NTH12] Chandra M Natarajan, Michael G Tanner, and Robert H Hadfield. "Superconducting nanowire single-photon detectors: physics and applications". In: *Superconductor science and technology* 25.6 (2012), p. 063001.

[NV11] Miguel Navascués and Tamás Vértesi. "Activation of Nonlocal Quantum Resources". In: *Phys. Rev. Lett.* 106 (6 Feb. 2011), p. 060403. DOI: 10.1103/PhysRevLett.106.060403. URL: https://link.aps.org/doi/10.1103/PhysRevLett.106.060403.

[NW20a] Miguel Navascués and Elie Wolfe. "The Inflation Technique Completely Solves the Causal Compatibility Problem". In: *J. Causal Inference* 8.1 (2020), pp. 70 –91. DOI: 10.1515/jci-2018-0008. arXiv: 1707.06476. URL: https://www.degruyter.com/view/journals/jci/8/1/article-p70.xml.

[NW20b] Miguel Navascués and Elie Wolfe. "The Inflation Technique Completely Solves the Causal Compatibility Problem". In: *Journal of Causal Inference* 8.1 (Sept. 2020), pp. 70–91. ISSN: 2193-3677. DOI: 10.1515/jci-2018-0008. URL: http://dx.doi.org/10.1515/jci-2018-0008.

[OCB12] Ognyan Oreshkov, Fabio Costa, and Časlav Brukner. "Quantum correlations with no causal order". In: *Nature communications* 3.1 (2012), p. 1092.

[O'D+21] Brendan O'Donoghue et al. *SCS: Splitting Conic Solver*. https://github.com/cvxgrp/scs. Nov. 2021.

[Ore19] Ognyan Oreshkov. "Time-delocalized quantum subsystems and operations: on the existence of processes with indefinite causal structure in quantum mechanics". In: *Quantum* 3 (Dec. 2019), p. 206. ISSN: 2521-327X. DOI: 10.22331/q-2019-12-02-206. URL: https://doi.org/10.22331/q-2019-12-02-206.

[Pal12] Carlos Palazuelos. "Superactivation of Quantum Nonlocality". In: *Physical Review Letters* 109.19 (Nov. 2012). ISSN: 1079-7114. DOI: 10.1103/physrevlett.109.190401. URL: http://dx.doi.org/10.1103/PhysRevLett.109.190401.

[Par12] M. G. A. Paris. "The modern tools of quantum mechanics: A tutorial on quantum states, measurements, and operations". In: *The European Physical Journal Special Topics* 203.1 (Apr. 2012), 61–86. ISSN: 1951-6401. DOI: 10.1140/epjst/e2012-01535-1. URL: http://dx.doi.org/10.1140/epjst/e2012-01535-1.

[Paw+09] Marcin Pawłowski et al. "Information causality as a physical principle". In: *Nature* 461.7267 (Oct. 2009), 1101–1104. ISSN: 1476-4687. DOI: 10.1038/nature08400. URL: http://dx.doi.org/10.1038/nature08400.

[PBS11] Stefano Pironio, Jean-Daniel Bancal, and Valerio Scarani. "Extremal correlations of the tripartite no-signaling polytope". In: *Journal of Physics A Mathematical General* 44.6, 065303 (Feb. 2011), p. 065303. DOI: 10.1088/1751-8113/44/6/065303. arXiv: 1101.2477 [quant-ph].

[Pea09] Judea Pearl. *Causality*. 2nd ed. Cambridge University Press, 2009. DOI: 10.1017/CBO9780511803161.

[Pea70] Philip M. Pearle. "Hidden-Variable Example Based upon Data Rejection". In: *Phys. Rev. D* 2 (8 Oct. 1970), pp. 1418–1425. DOI: 10.1103/PhysRevD.2.1418. URL: https://link.aps.org/doi/10.1103/PhysRevD.2.1418.

[Pir+10] S. Pironio et al. "Random numbers certified by Bell's theorem". In: *Nature* 464.7291 (2010), pp. 1021–1024. DOI: 10.1038/nature09008. URL: https://doi.org/10.1038/nature09008.

[PK19] Alejandro Pozas-Kerstjens. "Quantum information outside quantum information". PhD thesis. Universitat Politécnica de Catalunya, 2019. URL: http://hdl.handle.net/10803/667696.

[PK+19] Alejandro Pozas-Kerstjens et al. “Bounding the Sets of Classical and Quantum Correlations in Networks”. In: *Phys. Rev. Lett.* 123 (14 Oct. 2019), p. 140503. DOI: 10.1103/PhysRevLett.123.140503. arXiv: 1904.08943. URL: https://link.aps.org/doi/10.1103/PhysRevLett.123.140503.

[PK+23] Alejandro Pozas-Kerstjens et al. “Post-quantum nonlocality in the minimal triangle scenario”. In: *New Journal of Physics* 25.11 (Nov. 2023), p. 113037. ISSN: 1367-2630. DOI: 10.1088/1367-2630/ad0a16. URL: http://dx.doi.org/10.1088/1367-2630/ad0a16.

[PKGR23a] Alejandro Pozas-Kerstjens, Nicolas Gisin, and Marc-Olivier Renou. “Proofs of Network Quantum Nonlocality in Continuous Families of Distributions”. In: (2023). DOI: 10.1103/PHYSREVLETT.130.090201.

[PKGR23b] Alejandro Pozas-Kerstjens, Nicolas Gisin, and Marc-Olivier Renou. “Proofs of Network Quantum Nonlocality in Continuous Families of Distributions”. In: *Phys. Rev. Lett.* 130 (9 Feb. 2023), p. 090201. DOI: 10.1103/PhysRevLett.130.090201. arXiv: 2203.16543. URL: https://link.aps.org/doi/10.1103/PhysRevLett.130.090201.

[PKGT22] Alejandro Pozas-Kerstjens, Nicolas Gisin, and Armin Tavakoli. “Full Network Nonlocality”. In: *Phys. Rev. Lett.* 128 (1 Jan. 2022), p. 010403. DOI: 10.1103/PhysRevLett.128.010403. arXiv: 2105.09325. URL: https://link.aps.org/doi/10.1103/PhysRevLett.128.010403.

[PNA10] Stefano Pironio, Miguel Navascués, and Antonio Acín. “Convergent relaxations of polynomial optimization problems with non-commuting variables”. In: *SIAM J. Optim.* 20 (2010), pp. 2157–2180. DOI: 10.1137/090760155. arXiv: 0903.4368.

[Pol+23] Emanuele Polino et al. “Experimental nonclassicality in a causal network without assuming freedom of choice”. In: (2023). DOI: 10.1038/S41467-023-36428-W.

[Pop14] Sandu Popescu. “Nonlocality beyond quantum mechanics”. In: *Nature Physics* 10.4 (2014), pp. 264–270.

[Pop95] Sandu Popescu. "Bell's Inequalities and Density Matrices: Revealing "Hidden" Nonlocality". In: *Phys. Rev. Lett.* 74 (14 Apr. 1995), pp. 2619–2622. DOI: 10.1103/PhysRevLett.74.2619. URL: https://link.aps.org/doi/10.1103/PhysRevLett.74.2619.

[PR94] Sandu Popescu and Daniel Rohrlich. "Quantum nonlocality as an axiom". In: *Foundations of Physics* 24.3 (1994), pp. 379–385.

[QBH16] Marco T. Quintino, Nicolas Brunner, and Marcus Huber. "Superactivation of quantum steering". In: *Phys. Rev. A* 94 (Nov. 2016), p. 062123. DOI: 10.1103/physreva.94.062123. arXiv: 1610.01037 [quant-ph].

[Qui+15] Marco T. Quintino et al. "Inequivalence of entanglement, steering, and Bell nonlocality for general measurements". In: *Phys. Rev. A* 92 (3 Sept. 2015), p. 032107. DOI: 10.1103/PhysRevA.92.032107. arXiv: 1501.03332 [quant-ph]. URL: http://link.aps.org/doi/10.1103/PhysRevA.92.032107.

[RB22a] Marc-Olivier Renou and Salman Beigi. "Network nonlocality via rigidity of token counting and color matching". In: *Phys. Rev. A* 105 (2 2022), p. 022408. DOI: 10.1103/PhysRevA.105.022408. URL: https://link.aps.org/doi/10.1103/PhysRevA.105.022408.

[RB22b] Marc-Olivier Renou and Salman Beigi. "Network nonlocality via rigidity of token counting and color matching". In: *Physical Review A* 105.2 (Feb. 2022). ISSN: 2469-9934. DOI: 10.1103/physreva.105.022408. URL: http://dx.doi.org/10.1103/PhysRevA.105.022408.

[RB22c] Marc-Olivier Renou and Salman Beigi. "Nonlocality for Generic Networks". In: *Phys. Rev. Lett.* 128 (6 2022), p. 060401. DOI: 10.1103/PhysRevLett.128.060401. URL: https://link.aps.org/doi/10.1103/PhysRevLett.128.060401.

[Red+19] Dileep V Reddy et al. "Exceeding 95% system efficiency within the telecom c-band in superconducting nanowire single photon detectors". In: *CLEO: QELS_Fundamental Science*. Optical Society of America. 2019, FF1A–3.

[Rei91] Hans Reichenbach. *The direction of time*. Vol. 65. Univ of California Press, 1991.

[Ren+19a] Marc-Olivier Renou et al. "Genuine Quantum Nonlocality in the Triangle Network". In: *Phys. Rev. Lett.* 123 (14 Sept. 2019), p. 140401. DOI: 10.1103/PhysRevLett.123.140401. URL: https://link.aps.org/doi/10.1103/PhysRevLett.123.140401.

[Ren+19b] Marc-Olivier Renou et al. "Limits on Correlations in Networks for Quantum and No-Signaling Resources". In: *Phys. Rev. Lett.* 123 (7 2019), p. 070403. DOI: 10.1103/PhysRevLett.123.070403. URL: https://link.aps.org/doi/10.1103/PhysRevLett.123.070403.

[Ren+21] Marc-Olivier Renou et al. "Quantum theory based on real numbers can be experimentally falsified". In: *Nature* 600.7890 (Dec. 2021), pp. 625–629. DOI: 10.1038/s41586-021-04160-4. URL: https://doi.org/10.1038%2Fs41586-021-04160-4.

[RKB18] Marc Olivier Renou, J ędrzej Kaniewski, and Nicolas Brunner. "Self-Testing Entangled Measurements in Quantum Networks". In: *Phys. Rev. Lett.* 121 (25 2018), p. 250507. DOI: 10.1103/PhysRevLett.121.250507. URL: https://link.aps.org/doi/10.1103/PhysRevLett.121.250507.

[RMMB21] Denis Rosset, Felipe Montealegre-Mora, and Jean-Daniel Bancal. "RepLAB: A Computational/Numerical Approach to Representation Theory". In: *Quantum Theory and Symmetries*. Ed. by M. B. Paranjape et al. Cham: Springer International Publishing, 2021, pp. 643–653. ISBN: 978-3-030-55777-5.

[Rod+23] Giovanni Rodari et al. *Characterizing Hybrid Causal Structures with the Exclusivity Graph Approach*. 2023. arXiv: 2401.00063 [quant-ph].

[Ros+16] Denis Rosset et al. "Nonlinear Bell Inequalities Tailored for Quantum Networks". In: *Phys. Rev. Lett.* 116 (1 2016), p. 010403. DOI: 10.1103/PhysRevLett.116.010403. URL: https://link.aps.org/doi/10.1103/PhysRevLett.116.010403.

[Sau+10] Dylan J. Saunders et al. "Experimental EPR-steering using Bell-local states". In: *Nature Physics* 6.11 (Nov. 2010), pp. 845–849. DOI: 10.1038/nphys1766. arXiv: 0909.0805 [quant-ph].

[SB03] Cezary Sliwa and Konrad Banaszek. "Conditional preparation of maximal polarization entanglement". In: *Phys. Rev. A* 67 (3 Mar. 2003), p. 030101. DOI: 10.1103/PhysRevA.67.030101. URL: https://link.aps.org/doi/10.1103/PhysRevA.67.030101.

[ŠB20] Ivan Šupić and Joseph Bowles. "Self-testing of quantum systems: a review". In: *Quantum* 4 (Sept. 2020), p. 337. ISSN: 2521-327X. DOI: 10.22331/q-2020-09-30-337. URL: https://doi.org/10.22331/q-2020-09-30-337.

[ŠB23] Ivan Šupić and Nicolas Brunner. "Self-testing nonlocality without entanglement". In: *Physical Review A* 107.6 (June 2023). ISSN: 2469-9934. DOI: 10.1103/physreva.107.062220. URL: http://dx.doi.org/10.1103/PhysRevA.107.062220.

[SBB23] Pavel Sekatski, Sadra Boreiri, and Nicolas Brunner. "Partial Self-Testing and Randomness Certification in the Triangle Network". In: *Phys. Rev. Lett.* 131 (10 2023), p. 100201. DOI: 10.1103/PhysRevLett.131.100201. URL: https://link.aps.org/doi/10.1103/PhysRevLett.131.100201.

[SC15] Paul Skrzypczyk and Daniel Cavalcanti. "Loss-tolerant Einstein-Podolsky-Rosen steering for arbitrary-dimensional states: Joint measurability and unbounded violations under losses". In: *Phys. Rev. A* 92 (2 2015), p. 022354. DOI: 10.1103/PhysRevA.92.022354. URL: https://link.aps.org/doi/10.1103/PhysRevA.92.022354.

[Sca+05] Valerio Scarani et al. "Quantum cloning". In: *Reviews of Modern Physics* 77.4 (Oct. 2005), pp. 1225–1256. DOI: 10.1103/RevModPhys.77.1225. arXiv: quant-ph/0511088 [quant-ph].

[Sch+23] David Schmid et al. "Understanding the interplay of entanglement and nonlocality: motivating and developing a new branch of entanglement theory". In: *Quantum* 7 (Dec. 2023), p. 1194.

ISSN: 2521-327X. DOI: 10.22331/q-2023-12-04-1194. URL: https://doi.org/10.22331/q-2023-12-04-1194.

[Sen+05] Aditi Sen(de) et al. "Entanglement swapping of noisy states: A kind of superadditivity in nonclassicality". In: *Phys. Rev. A* 72.4, 042310 (Oct. 2005), p. 042310. DOI: 10.1103/PhysRevA.72.042310. arXiv: quant-ph/0311194 [quant-ph].

[SG11] Greg ver Steeg and Aram Galstyan. "A Sequence of Relaxations Constraining Hidden Variable Models". In: *Proceedings of the Twenty-Seventh Conference on Uncertainty in Artificial Intelligence*. UAI'11. Barcelona, Spain: AUAI Press, 2011, 717–726. ISBN: 9780974903972. DOI: 10.48550/arXiv.1106.1636. arXiv: 1106.1636. URL: https://arxiv.org/abs/1106.1636.

[Sha+15] Lynden K. Shalm et al. "Strong Loophole-Free Test of Local Realism". In: *Phys. Rev. Lett.* 115 (25 Dec. 2015), p. 250402. DOI: 10.1103/PhysRevLett.115.250402. URL: https://link.aps.org/doi/10.1103/PhysRevLett.115.250402.

[Shu+23] Manish Kumar Shukla et al. "Correlations in Quantum Network Topologies Created with Cloning". In: *Mathematics* 11.11 (May 2023), p. 2440. ISSN: 2227-7390. DOI: 10.3390/math11112440. URL: http://dx.doi.org/10.3390/math11112440.

[Slo20] William Slofstra. "Tsirelson's problem and an embedding theorem for groups arising from non-local games". In: *Journal of the American Mathematical Society* 33.1 (2020), pp. 1–56.

[SS22] Guillaume Sagnol and Maximilian Stahlberg. "PICOS: A Python interface to conic optimization solvers". In: *J. Open Source Softw.* 7.70 (Feb. 2022), p. 3915. ISSN: 2475-9066. DOI: 10.21105/joss.03915.

[Tav+14] Armin Tavakoli et al. "Nonlocal correlations in the star-network configuration". In: *Phys. Rev. A* 90 (6 2014), p. 062109. DOI: 10.1103/PhysRevA.90.062109. URL: https://link.aps.org/doi/10.1103/PhysRevA.90.062109.

[Tav+22] Armin Tavakoli et al. "Bell nonlocality in networks". In: *Reports on Progress in Physics* 85.5 (Mar. 2022), p. 056001. ISSN: 1361-6633. DOI: 10.1088/1361-6633/ac41bb. URL: http://dx.doi.org/10.1088/1361-6633/ac41bb.

[TRC19] Tassius Temistocles, Rafael Rabelo, and Marcelo Terra Cunha. "Measurement compatibility in Bell nonlocality tests". In: *Phys. Rev. A* 99.4, 042120 (Apr. 2019), p. 042120. DOI: 10.1103/PhysRevA.99.042120. arXiv: 1806.09232 [quant-ph].

[Tsi06] Boris Tsirelson. "Bell inequalities and operator algebras". In: *Open Quantum Problems* (2006).

[Tsi93] Boris S Tsirelson. "Some results and problems on quantum Bell-type inequalities". In: *Hadronic Journal Supplement* 8.4 (1993), pp. 329–345.

[TTT99] Kim-Chuan Toh, Michael J. Todd, and Reha H. Tütüncü. "SDPT3 — A MATLAB software package for semidefinite programming". In: *Optim. Metods Softw.* 11.1-4 (1999), pp. 545–581. DOI: 10.1080/10556789908805762. URL: https://doi.org/10.1080/10556789908805762.

[TWC91] Sze M. Tan, Dan F. Walls, and Matthew J. Collett. "Nonlocality of a single photon". In: *Phys. Rev. Lett.* 66 (3 Jan. 1991), pp. 252–255. DOI: 10.1103/PhysRevLett.66.252. URL: https://link.aps.org/doi/10.1103/PhysRevLett.66.252.

[Uol+20] Roope Uola et al. "Quantum steering". In: *Reviews of Modern Physics* 92.1 (Mar. 2020). DOI: 10.1103/revmodphys.92.015001. URL: https://doi.org/10.1103%2Frevmodphys.92.015001.

[VA+23] Luis Villegas-Aguilar et al. *Nonlocality activation in a photonic quantum network*. 2023. arXiv: 2309.06501 [quant-ph].

[Vai95] Lev Vaidman. "Nonlocality of a Single Photon Revisited Again". In: *Phys. Rev. Lett.* 75 (10 Sept. 1995), pp. 2063–2063. DOI: 10.1103/PhysRevLett.75.2063. URL: https://link.aps.org/doi/10.1103/PhysRevLett.75.2063.

[VB96] Lieven Vandenberghe and Stephen Boyd. "Semidefinite Programming". In: *SIAM Review* 38.1 (1996), pp. 49–95. DOI: 10.1137/1038003. eprint: https://doi.org/10.1137/1038003. URL: https://doi.org/10.1137/1038003.

[VGO+20] Pauli Virtanen, Ralf Gommers, Travis E. Oliphant, et al. "SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python". In: *Nat. Methods* 17 (2020), pp. 261–272. DOI: 10.1038/s41592-019-0686-2.

[VH+19] Thomas Van Himbeeck et al. "Quantum violations in the Instrumental scenario and their relations to the Bell scenario". In: *Quantum* 3 (Sept. 2019), p. 186. ISSN: 2521-327X. DOI: 10.22331/q-2019-09-16-186. URL: http://dx.doi.org/10.22331/q-2019-09-16-186.

[VR23] Vilasini Venkatesh and Renato Renner. *Embedding cyclic causal structures in acyclic space-times: no-go results for indefinite causality*. 2023. arXiv: 2203.11245 [quant-ph].

[Wan+24] Ning-Ning Wang et al. *Experimental genuine quantum nonlocality in the triangle network*. 2024. arXiv: 2401.15428 [quant-ph].

[WC17a] Mirjam Weilenmann and Roger Colbeck. "Analysing causal structures with entropy". In: *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences* 473.2207 (Nov. 2017), p. 20170483. ISSN: 1471-2946. DOI: 10.1098/rspa.2017.0483. URL: http://dx.doi.org/10.1098/rspa.2017.0483.

[WC17b] Howard M. Wiseman and Eric G. Cavalcanti. "Causarum Investigatio and the Two Bell's Theorems of John Bell". In: *Quantum [Un]Speakables II: Half a Century of Bell's Theorem*. Ed. by Reinhold Bertlmann and Anton Zeilinger. Cham: Springer International Publishing, 2017, pp. 119–142. ISBN: 978-3-319-38987-5. DOI: 10.1007/978-3-319-38987-5_6. URL: https://doi.org/10.1007/978-3-319-38987-5_6.

[WC20a] Mirjam Weilenmann and Roger Colbeck. "Self-Testing of Physical Theories, or, Is Quantum Theory Optimal with Respect to Some Information-Processing Task?" In: *Phys. Rev. Lett.*

125 (6 Aug. 2020), p. 060406. DOI: 10.1103/PhysRevLett.125.060406. URL: https://link.aps.org/doi/10.1103/PhysRevLett.125.060406.

[WC20b] Mirjam Weilenmann and Roger Colbeck. "Toward correlation self-testing of quantum theory in the adaptive Clauser-Horne-Shimony-Holt game". In: *Phys. Rev. A* 102 (2 Aug. 2020), p. 022203. DOI: 10.1103/PhysRevA.102.022203. URL: https://link.aps.org/doi/10.1103/PhysRevA.102.022203.

[WEH18] Stephanie Wehner, David Elkouss, and Ronald Hanson. "Quantum internet: A vision for the road ahead". In: *Science* 362.6412 (2018), eaam9288. DOI: 10.1126/science.aam9288. eprint: https://www.science.org/doi/pdf/10.1126/science.aam9288. URL: https://www.science.org/doi/abs/10.1126/science.aam9288.

[Wei+98] Gregor Weihs et al. "Violation of Bell's Inequality under Strict Einstein Locality Conditions". In: *Phys. Rev. Lett.* 81 (23 Dec. 1998), pp. 5039–5043. DOI: 10.1103/PhysRevLett.81.5039. URL: https://link.aps.org/doi/10.1103/PhysRevLett.81.5039.

[Wer89] Reinhard F. Werner. "Quantum states with Einstein-Podolsky-Rosen correlations admitting a hidden-variable model". In: *Phys. Rev. A* 40 (8 Oct. 1989), pp. 4277–4281. DOI: 10.1103/PhysRevA.40.4277. URL: https://link.aps.org/doi/10.1103/PhysRevA.40.4277.

[Wer98] Reinhard F. Werner. "Optimal cloning of pure states". In: *Phys. Rev. A* 58 (3 Sept. 1998), pp. 1827–1832. DOI: 10.1103/PhysRevA.58.1827. URL: https://link.aps.org/doi/10.1103/PhysRevA.58.1827.

[Wil13] Mark M Wilde. *Quantum information theory*. Cambridge university press, 2013.

[Wit+12] Bernhard Wittmann et al. "Loophole-free Einstein-Podolsky-Rosen experiment via quantum steering". In: *New Journal of Physics* 14.5, 053030 (May 2012), p. 053030. DOI: 10.1088/1367-2630/14/5/053030. arXiv: 1111.0760 [quant-ph].

[WJD07] Howard M. Wiseman, Steve J. Jones, and Andrew C. Doherty. "Steering, Entanglement, Nonlocality, and the Einstein Podolsky Rosen Paradox". In: *Phys. Rev. Lett.* 98.14, 140402 (Apr. 2007), p. 140402. DOI: 10.1103/PhysRevLett.98.140402. arXiv: quant-ph/0612147 [quant-ph].

[Wol] In: *Quantum Information and Computation* 18.11 & 12 (Sept. 2018). ISSN: 1533-7146. DOI: 10.26421/qic18.11-12. URL: http://dx.doi.org/10.26421/QIC18.11-12.

[Wol+21] Elie Wolfe et al. "Quantum Inflation: A General Approach to Quantum Causal Compatibility". In: *Phys. Rev. X* 11 (2 May 2021), p. 021043. DOI: 10.1103/PhysRevX.11.021043. arXiv: 1909.10519. URL: https://link.aps.org/doi/10.1103/PhysRevX.11.021043.

[WSF19] Elie Wolfe, Robert W. Spekkens, and Tobias Fritz. "The Inflation Technique for Causal Inference with Latent Variables". In: *Journal of Causal Inference* 7.2 (2019), p. 20170020. DOI: doi:10.1515/jci-2017-0020. URL: https://doi.org/10.1515/jci-2017-0020.

[YHD11] Jon Yard, Patrick Hayden, and Igor Devetak. "Quantum Broadcast Channels". In: *IEEE Transactions on Information Theory* 57.10 (Oct. 2011), 7147–7162. ISSN: 1557-9654. DOI: 10.1109/tit.2011.2165811. arXiv: quant-ph/0603098 [quant-ph]. URL: http://dx.doi.org/10.1109/TIT.2011.2165811.

[YS92a] Bernard Yurke and David Stoler. "Bell's-inequality experiments using independent-particle sources". In: *Phys. Rev. A* 46 (5 Sept. 1992), pp. 2229–2234. DOI: 10.1103/PhysRevA.46.2229. URL: https://link.aps.org/doi/10.1103/PhysRevA.46.2229.

[YS92b] Bernard Yurke and David Stoler. "Einstein-Podolsky-Rosen effects from independent particle sources". In: *Phys. Rev. Lett.* 68 (9 Mar. 1992), pp. 1251–1254. DOI: 10.1103/PhysRevLett.68.1251. URL: https://link.aps.org/doi/10.1103/PhysRevLett.68.1251.

[Ż+98] Marek Żukowski et al. “Generalized quantum measurements and local realism”. In: *Phys. Rev. A* 58.3 (Sept. 1998), pp. 1694–1698. DOI: 10.1103/PhysRevA.58.1694. arXiv: quant-ph/9608035 [quant-ph].

[Zhu+20] Di Zhu et al. “Resolving Photon Numbers Using a Superconducting Nanowire with Impedance-Matching Taper”. In: *Nano Letters* 20.5 (2020), pp. 3858–3863.

[Żuk+93] Marek Żukowski et al. ““Event-ready-detectors” Bell experiment via entanglement swapping”. In: *Physical Review Letters* 71.26 (1993), p. 4287.

[Šu+22] Ivan Šupić et al. “Genuine network quantum nonlocality and self-testing”. In: *Physical Review A* 105.2 (Feb. 2022). ISSN: 2469-9934. DOI: 10.1103/physreva.105.022206. URL: http://dx.doi.org/10.1103/PhysRevA.105.022206.

Appendix A

Feasibility in convex programming

A feasibility problem is a type of optimization problem where the goal is to determine whether a set of constraints can be satisfied or not. The objective function is not relevant in this type of problem, and the focus is solely on finding a feasible solution that satisfies the constraints.

Often, the constraints are continuously parameterized, and one is interested in finding the range of parameters for which the constraints are feasible. Close to the threshold in parameter space where the constraints become infeasible, numerical instabilities may appear. These are manifested as either the solver failing to find a feasible solution, or the solver finding a feasible solution where it should be infeasible. This is due to the fact that the solver works with finite precision. One approach to avoid these instabilities is to reformulate the feasibility problem as an optimization problem.

A.1 Feasibility as optimization

Consider a generic linear program with constraints $Ax = b, x \geq 0$. One equivalent formulation as an optimization problem is as follows:

$$\begin{array}{lll} \text{find} & x & \\ \text{s.t.} & Ax - b = 0 & \\ & x \geq 0 & \end{array} \quad \longleftrightarrow \quad \begin{array}{ll} \max & \lambda \\ \text{s.t.} & -\lambda\vec{1} \geq Ax - b \geq \lambda\vec{1} \\ & x \geq \lambda\vec{1},\ \lambda \in \mathbb{R} \end{array}$$

where $\vec{1}$ is a vector of ones. If the solution of the optimization program, λ^*, is strictly negative $\lambda^* < 0$, then the original program is infeasible. If,

however, the optimal solution is nonnegative, $\lambda^* \geq 0$, the original program is feasible.

As another example, consider a semidefinite programming feasibility problem, where one wants to find whether a matrix $F = F_0 + \sum x_i F_i$ can be made positive semidefinite, where F_0, F_i are fixed. This can be formulated as an optimization problem in the following manner:

$$\begin{array}{lll} \text{find} & x_i & \\ \text{s.t.} & F = F_0 + \sum x_i F_i \succeq 0 & \\ & x_i \in \mathbb{R} & \end{array} \quad \longleftrightarrow \quad \begin{array}{lll} \max & \lambda & \\ \text{s.t.} & F = F_0 + \sum x_i F_i \succeq \lambda \mathbb{1} & . \\ & x_i \in \mathbb{R}, \lambda \in \mathbb{R} & \end{array}$$

The solution λ^* is the maximum smallest eigenvalue of F over x_i. Strictly negative values, $\lambda^* < 0$, imply that the matrix F cannot be made positive semidefinite, thus the original program is infeasible. If the solution is $\lambda^* = -\infty$, then the only way to make the original problem feasible is by removing the constraint $F \succeq 0$. All other values of λ^* correspond to feasible programs.

A.2 Farkas' lemma and certificates of infeasibility

Farkas' lemma is of central importance in convex programming, as it allows one to certify that a given convex program will be infeasible for a wide choice of parameters without having to solve the program. This is particularly useful in the context of quantum information theory, as it allows for the construction of entanglement witnesses [DCPSM04], the construction of Bell inequalities [Bru+14] or bounding the set of quantum correlations [NPA08] amongst other applications.

Linear programming. The standard formulation of Farkas' lemma states that a linear program with constraints $Ax = b, x \geq 0$ is infeasible if and only if there exists a vector y such that $A^T y \geq 0$ and $b \cdot y < 0$.

Proof. This can be shown directly from the dual problem to the feasibility problem expressed as optimization. Following the standard approach from Ref. [BV04], we derive the dual formulation. We start by writing the Lagrangian of the optimization problem for the linear program:

$$L(x, \lambda, \mu, \nu, \gamma) = \lambda + \mu \cdot (Ax - b - \lambda \vec{1}) + \nu \cdot (-Ax + b + \lambda \vec{1}) + \gamma \cdot (x - \lambda \vec{1}),$$

which for $\mu \geq 0, \nu \geq 0, \gamma \geq 0$, and for λ, x satisfying the constraints of the original problem, gives the upper bound

$$L(x, \lambda, \mu, \nu, \gamma) \geq \lambda \,.$$

In particular, this upper bound holds for the optimal value of the optimization program, λ^*, which we can get by maximising both sides over x, λ satisfying the constraints of the problem:

$$g(\mu, \nu, \gamma) := \max_{x,\lambda} L(x, \lambda, \mu, \nu, \gamma) \geq \lambda^*,$$

where:

$$g(\mu, \nu, \gamma) = \begin{cases} \nu \cdot b - \mu \cdot b & \text{if } \begin{array}{c} 1 - \mu \cdot \vec{1} + \nu \cdot \vec{1} - \gamma \cdot \vec{1} = 0 \\ A^T\mu - A^T\nu + \gamma = 0 \end{array} \\ \infty & \text{otherwise} \end{cases} .$$

This follows by grouping together the terms in the Lagrangian by primal variables. Now to get the tightest upper bound on the optimal value of the optimization program, λ^*, we minimise the upper bound $g(\mu, \nu, \gamma)$:

$$\begin{aligned} \min_{\mu,\nu,\gamma} \quad & (\nu - \mu) \cdot b \\ \text{s.t.} \quad & -A^T(\nu - \mu) + \gamma = 0 \\ & 1 - \mu \cdot \vec{1} + \nu \cdot \vec{1} - \gamma \cdot \vec{1} = 0 = 0 \\ & \mu \geq 0, \nu \geq 0, \gamma \geq 0 \,. \end{aligned}$$

By introducing the variable $y := \nu - \mu$, and replacing $\nu = y + \mu$, and removing the slack variable γ thereby converting equality constraints to inequality constraints, we arrive at the dual problem:

$$\min_{y,\nu,\gamma} \quad b \cdot y \tag{A.1}$$

$$\text{s.t.} \quad A^T y \geq 0 \tag{A.2}$$

$$1 + y \cdot \vec{1} \geq 0 \,. \tag{A.3}$$

Any y such that $b \cdot y < 0$ and $A^T y \geq 0$, proves that the original program is infeasible, as this certifies that the optimal value of optimization program

is strictly negative, thus the original program is infeasible. This is the statement of Farkas' lemma for linear programs.

Note that any y satisfying $b \cdot y \leq 0$ and $A^T y \geq 0$ can be rescaled to satisfy $1 + y \cdot \vec{1} \geq 0$ through $y_i \to y_i / |\sum y_i|$.

Bell inequalities. As a particularly relevant example, consider the problem of determining whether a particular probability distribution $p(ab|xy)$ admits a local hidden variable model. This is formulated as the following feasibility problem:

$$\begin{array}{ll} \text{find} & q_{ij}, \\ \text{s.t.} & p(ab|xy) = \sum_i D(a|xi)\, D(b|yj)\, q_{ij}, \\ & \sum q_{ij} = 1, \\ & q_{ij} \geq 0. \end{array}$$

where $D(a|xi)$, $D(b|yj)$ are all possible deterministic functions labeled by i, j. This connects to the standard form $Ax = b, x \geq 0$ by writing:

$$b = \begin{bmatrix} p(00|00) \\ p(00|01) \\ \vdots \\ p(11|11) \\ 1 \end{bmatrix}, \quad A = \begin{bmatrix} D \\ \vec{1}^T \end{bmatrix},$$

where the matrix D has elements $D_{(a,b,x,y),(i,j)} = D(a|xi)D(b|yj)$. Farkas' lemma then states that the existence of a vector $y = (y_{abxy}, y_0)$ such that:

$$\begin{aligned} &\sum_{abxy} y_{abxy} p(ab|xy) + y_0 < 0, \\ &\sum_{abxy} y_{abxy} D(a|xi) D(b|yj) + y_0 \geq 0 \quad \forall\ i, j, \end{aligned}$$

certifies that the program is infeasible, and that thus $p(ab|xy)$ does not admit a local hidden variable model. This is the standard form of Bell inequalities [Bru+14].

The geometric intuition is the following: the vector y defines a hyperplane in the space of probability distributions such that it does not intersect the polytope of local hidden variable models. Therefore, if a probability

distribution $p(ab|xy)$ lies on the other side of the hyperplane then it cannot admit a local hidden variable model.

Semidefinite programming. We can derive analogues of Farkas' lemma for semidefinite programs by following the same steps as in the previous paragraph. The Lagrangian of the semidefinite feasibility problem written as an optimization problem is:

$$\begin{aligned} L(\lambda, x_i, Z) &= \lambda + \operatorname{Tr} Z \left(F_0 + \sum x_i F_i - \lambda \mathbb{1} \right) \\ &= \operatorname{Tr} Z F_0 + \lambda(1 - \operatorname{Tr} Z) + \sum x_i (\operatorname{Tr} Z F_i) \,. \end{aligned}$$

It holds that:

$$g(Z) := \max_{\lambda, x_i} L(\lambda, x_i, Z) = \begin{cases} \operatorname{Tr} Z F_0 & \text{if } \begin{array}{c} 1 - \operatorname{Tr} Z = 0 \\ \operatorname{Tr} Z F_i = 0 \end{array} \\ \infty & \text{otherwise} \end{cases} .$$

for $Z \succeq 0$ and x_i, λ satisfying the constraints of the problem. The dual is then:

$$\begin{aligned} \min_{Z} \quad & \operatorname{Tr} Z F_0 \\ \text{s.t.} \quad & \operatorname{Tr} Z = 1, \quad \operatorname{Tr} Z F_i = 0, \quad Z \succeq 0 \,. \end{aligned}$$

Therefore any $Z \succeq 0$ such that $\operatorname{Tr} Z F_0 < 0$ and $\operatorname{Tr} Z F_i = 0$ proves that the original program is infeasible, as this certifies that the optimal value of optimization program λ^* is strictly negative. For any such Z, we can rescale it to satisfy $\operatorname{Tr} Z = 1$ through $Z \to Z / \operatorname{Tr} Z$.

Appendix B

Single photon nonlocality

B.1 Noiseless output distribution

Here we derive the form of the noiseless output distribution:

$$p_t^{Q=0,T=1,\nu=1}(abc) \equiv p_t(abc),$$

produced when all the elements of the optical scheme described in Section B.2 of this Appendix are perfect.

The initial state shared among the parties is:

$$|\psi^+\rangle\langle\psi^+|_{A_2B_1} \otimes |\psi^+\rangle\langle\psi^+|_{B_2C_1} \otimes |\psi^+\rangle\langle\psi^+|_{C_2A_1}$$
$$\text{with} \quad |\psi^+\rangle = \frac{|01\rangle + |10\rangle}{\sqrt{2}}. \quad \text{(S1)}$$

The action of a beamsplitter with transmissivity t and phase ϕ is described in terms of the input and output optical modes with creation operators $a_i^\dagger$ as:

$$\begin{pmatrix} a_2^\dagger \\ a_1^\dagger \end{pmatrix}_{in} = \begin{pmatrix} \sqrt{t} & -e^{-i\phi}\sqrt{1-t} \\ e^{i\phi}\sqrt{1-t} & \sqrt{t} \end{pmatrix} \begin{pmatrix} a_2^\dagger \\ a_1^\dagger \end{pmatrix}_{out}. \quad \text{(S2)}$$

Consequently, the corresponding unitary induced by the transformation can be derived in the Fock basis by expressing $|mn\rangle_{X_2X_1} \equiv \frac{a_2^{\dagger m}}{\sqrt{m!}} \frac{a_1^{\dagger n}}{\sqrt{n!}} |00\rangle_{X_2X_1}$, to

obtain

$$|10\rangle_{in} = \sqrt{t}|10\rangle_{out} - e^{-i\phi}\sqrt{1-t}|01\rangle_{out}, \tag{S3}$$
$$|01\rangle_{in} = \sqrt{t}|01\rangle_{out} + e^{i\phi}\sqrt{1-t}|10\rangle_{out}, \tag{S4}$$
$$|00\rangle_{in} = |00\rangle_{out}, \tag{S5}$$
$$|11\rangle_{in} = (2t-1)|11\rangle_{out} - e^{-i\phi}\sqrt{2t(1-t)}|02\rangle_{out} + e^{i\phi}\sqrt{2t(1-t)}|20\rangle_{out}. \tag{S6}$$

Accordingly, the POVM (4.6) can be written as

$$\Pi_t^{(0)} = |00\rangle\langle 00|, \tag{S7}$$
$$\Pi_t^{(R)} = |\chi_r\rangle\langle\chi_r| + 2t(1-t)|11\rangle\langle 11|, \tag{S8}$$
$$\Pi_t^{(L)} = |\chi_l\rangle\langle\chi_l| + 2t(1-t)|11\rangle\langle 11|, \tag{S9}$$
$$\Pi_t^{(2)} = (2t-1)^2|11\rangle\langle 11|, \tag{S10}$$

where $|\chi_r\rangle = \sqrt{t}|01\rangle - e^{i\phi}\sqrt{1-t}|10\rangle$, $|\chi_l\rangle = \sqrt{t}|10\rangle + e^{-i\phi}\sqrt{1-t}|01\rangle$, and where we truncated the Hilbert space considering that the input state consists only of combinations of vacuum and a single-photon excitation. Therefore, each party has four possible outputs $a, b, c \in \{0, L, R, 2\}$, standing for no detector counts □□, a count in the left detector ■□, a count in the right detector □■, or counts in both detectors ■■, respectively, described by the POVM above.

The resulting network output

$$p_t(abc) = \mathrm{Tr}[(\,|\psi^+\rangle\langle\psi^+|_{A_2B_1} \otimes |\psi^+\rangle\langle\psi^+|_{B_2C_1} \otimes |\psi^+\rangle\langle\psi^+|_{C_2A_1}\,)\cdot \\ \cdot\,(\Pi_t^{(a)}{}_{A_1A_2} \otimes \Pi_t^{(b)}{}_{B_1B_2} \otimes \Pi_t^{(c)}{}_{C_1C_2})] \tag{S11}$$

has multiple constraints due to the cyclic symmetry of the experiment, due to all the parties using the same value for the beamsplitter transmissivity t (S2), as well as photon number conservation. For example, all outputs of

the form (here χ represents any of L or R):

$$p_t(000) = 0\ , \quad p_t(00\chi) = 0\ , \qquad \text{(too few photons would be detected)} \tag{S12}$$

$$p_t(2\chi\chi) = 0\ , \quad p_t(22\chi) = 0\ , \qquad \text{(too many photons would be detected)} \tag{S13}$$

are null, due to the fact that there are initially 3 photons in the network, of which at most 2 can end up in the same photodetector.

The non-zero probabilities are, modulo the cyclic symmetry, in the form $p(0\chi\chi)$, $p(02\chi)$, $p(0\chi 2)$, $p(\chi\chi\chi)$, and are summarized, in order, in the following.

$$p_t(0LL) = p_t(0RR) = \frac{1}{4}t(1-t), \tag{S14}$$

$$p_t(0RL) = \frac{1}{2}t(1-t)^2, \tag{S15}$$

$$p_t(0LR) = \frac{1}{2}t^2(1-t), \tag{S16}$$

$$p_t(02R) = \frac{1}{8}(2t-1)^2 t, \qquad p_t(02L) = \frac{1}{8}(2t-1)^2(1-t), \tag{S17}$$

$$p_t(0R2) = \frac{1}{8}(2t-1)^2(1-t), \qquad p_t(0L2) = \frac{1}{8}(2t-1)^2 t, \tag{S18}$$

$$\begin{aligned} p_t(RRL) =& \frac{1}{8}t(1-t)(1+2\cos(\Phi)\sqrt{t(1-t)}), \\ p_t(LLR) =& \frac{1}{8}t(1-t)(1-2\cos(\Phi)\sqrt{t(1-t)}), \end{aligned} \tag{S19}$$

$$p_t(LLL) = \frac{1}{8}(1-3t(1-t)+2t^{\frac{3}{2}}(1-t)^{\frac{3}{2}}\cos(\Phi)), \tag{S20}$$

$$p_t(RRR) = \frac{1}{8}(1-3t(1-t)-2t^{\frac{3}{2}}(1-t)^{\frac{3}{2}}\cos(\Phi)), \tag{S21}$$

where $\Phi \equiv \phi_A + \phi_B + \phi_C$.

In what follows, we take $\Phi = 0$, as the range of values of t for which the distribution is proven to be nonlocal decreases when $\Phi \neq 0$ (that is, the following analysis can be performed for an arbitrary value of Φ, and the

interval of values of t for which p_t is nonlocal is maximized when $\Phi = 0$). Also, note that $\Phi = \phi_A + \phi_B + \phi_C$ can be tuned locally by any of the parties.

B.2 Nonlocality of the noiseless distribution

To prove the nonlocality of the ideal noiseless distribution p_t presented above, we take an approach inspired by the one presented in [Ren+19a]. There, a quantum distribution is proposed, which is based on the same input state in the triangle network (we report it in our notation):

$$|\psi^+\rangle_{A_2B_1} \otimes |\psi^+\rangle_{B_2C_1} \otimes |\psi^+\rangle_{C_2A_1} \equiv |\Psi^+\rangle_{A_1A_2B_1B_2C_1C_2}\,, \tag{S22}$$

with $|\psi^+\rangle = \frac{|01\rangle+|10\rangle}{\sqrt{2}}$, and the following POVM on the two modes X_2X_1 of each party $X = A, B, C$ (again, we use a notation that makes the comparison easier with the experiment proposed in the present thesis):

$$\Pi_t'^{(0)} = |00\rangle\langle 00|, \quad \Pi_t'^{(R)} = |\chi_r\rangle\langle\chi_r|, \quad \Pi_t'^{(L)} = |\chi_l\rangle\langle\chi_l|, \quad \Pi_t'^{(2)} = |11\rangle\langle 11|, \tag{S23}$$

where $|\chi_r\rangle = \sqrt{t}|01\rangle - \sqrt{1-t}|10\rangle$ and $|\chi_l\rangle = \sqrt{t}|10\rangle + \sqrt{1-t}|01\rangle$ (here we put all the phases ϕ_x to zero, as mentioned above). The output distribution of our experiment is not equivalent to that of [Ren+19a], as our POVM consists, as described in Sec. B.1, of:

$$\begin{aligned}\Pi_t^{(0)} &= |00\rangle\langle 00|, \quad \Pi_t^{(R)} = |\chi_r\rangle\langle\chi_r| + 2t(1-t)|11\rangle\langle 11|,\\ \Pi_t^{(L)} &= |\chi_l\rangle\langle\chi_l| + 2t(1-t)|11\rangle\langle 11|, \quad \Pi_t^{(2)} = (2t-1)^2|11\rangle\langle 11|.\end{aligned} \tag{S24}$$

Notice that both POVMs Π and Π' are a coarse graining of the measurement

$$\begin{aligned}\Pi_t''^{(0)} = |00\rangle\langle 00|, \quad \Pi_t''^{(R1)} = |\chi_r\rangle\langle\chi_r|, \quad \Pi_t''^{(R2)} = 2t(1-t)|11\rangle\langle 11|,\\ \Pi_t''^{(L)} = |\chi_l\rangle\langle\chi_l|, \quad \Pi_t''^{(L2)} = 2t(1-t)|11\rangle\langle 11|, \quad \Pi_t''^{(2)} = (2t-1)^2|11\rangle\langle 11|.\end{aligned} \tag{S25}$$

The POVM Π'' is the one that would be obtained from the scheme described in the main text if the photodetectors were able to resolve photon numbers, and has thus six possible outputs (cf.(S7)—(S10)). Accordingly, it is possible to define distributions p_t, p_t', p_t'', obtained from the state (S22) and applying

(respectively) Π_t, Π'_t, Π''_t at each party modes X_2X_1, i.e.:

$$p_t(abc) = \mathrm{Tr}[|\Psi^+\rangle\langle\Psi^+|_{A_1A_2B_1B_2C_1C_2} (\Pi_t^{(a)}{}_{A_1A_2} \otimes \Pi_t^{(b)}{}_{B_1B_2} \otimes \Pi_t^{(c)}{}_{C_1C_2})] \tag{S26}$$

$$p'_t(abc) = \mathrm{Tr}[|\Psi^+\rangle\langle\Psi^+|_{A_1A_2B_1B_2C_1C_2} (\Pi_t'^{(a)}{}_{A_1A_2} \otimes \Pi_t'^{(b)}{}_{B_1B_2} \otimes \Pi_t'^{(c)}{}_{C_1C_2})] \tag{S27}$$

$$p''_t(abc) = \mathrm{Tr}[|\Psi^+\rangle\langle\Psi^+|_{A_1A_2B_1B_2C_1C_2} (\Pi_t''^{(a)}{}_{A_1A_2} \otimes \Pi_t''^{(b)}{}_{B_1B_2} \otimes \Pi_t''^{(c)}{}_{C_1C_2})] \tag{S28}$$

Surprisingly, we prove that p_t, p'_t, and p''_t, have the same range of nonlocality for the parameter t. That is, for a fixed t, if one among p_t, p'_t, p''_t, is classically reproducible in the triangle network, then all of them are. At the same time, the infeasibility of one among p_t, p'_t, p''_t, implies the infeasibility of all of them. From the physical point of view, this means that the possibility of performing perfect number-resolving photodetection does not enhance the "nonlocality" of the output distribution of our ideal experiment, although it may improve its resistance to noise.

To prove the nonlocal equivalence (in the triangle network) of the three distributions p_t, p'_t, p''_t we proceed as follows:

$$\text{feasibility } p'_t \Rightarrow \text{feasibility } p''_t \Rightarrow \text{feasibility } p_t \Rightarrow \text{feasibility } p'_t\,, \tag{S29}$$

where by "feasibility" we mean the feasibility of classically simulating the distribution with a local model, as from Eq. (4.2). The first two implications follow immediately, without assumptions on the input state $|\Psi\rangle$, from simple properties of the POVMs involved. Indeed:

- The POVM Π'' can be obtained as a fine-graining of Π' via a probabilistic splitting of $\Pi'^{(2)}$ in three outcomes $\Pi''^{(L2)}$, $\Pi''^{(R2)}$, $\Pi''^{(2)}$, which is just a classical local post-processing of the original projector $|11\rangle\langle 11|$.
- The POVM Π is a local coarse-graining of Π'' and thus p_t is classically simulable whenever p''_t is.

The last implication requires more effort and we prove it in the following subsections. To do so, we identify constraints on local strategies simulating p_t and show that these are the same as those needed to simulate p'_t, as from [Ren+19a] (cf. following derivations and Paragraph B.2.3).

B.2.1 Constraints on local models simulating p_t

We start by assuming that there exists a classical model that simulates the output distribution p_t of the experiment proposed in the main text, and we find the constraints that it has to respect. That is, we assume that indeed p_t (which is summarized in Sec. B.1), can be written as

$$p_t(abc) = \int_{\sigma,\lambda,\gamma\in[0,1]^3} q(a|\sigma\lambda)q(b|\lambda\gamma)q(c|\gamma\sigma)\,\mathrm{d}\gamma\mathrm{d}\sigma\mathrm{d}\lambda. \tag{S30}$$

Recall from Section 3.3 of the preliminaries that the local response functions may be taken to be deterministic without loss of generality, that is:

$$q(a|\sigma,\lambda) = \delta(a - A(\sigma,\lambda)), \tag{S31}$$

where $A(\sigma,\lambda)$ is some deterministic response function. Let X, Y and Z denote the set of possible γ, σ and λ respectively. Let us define

$$\begin{aligned}
X_0^B &= \{\gamma \,|\, \exists\,\lambda : B(\lambda,\gamma) = 0\}, & X_2^B &= \{\gamma \,|\, \exists\,\lambda : B(\lambda,\gamma) = 2\},\\
X_2^C &= \{\gamma \,|\, \exists\,\sigma : C(\gamma,\sigma) = 2\}, & X_0^C &= \{\gamma \,|\, \exists\,\sigma : C(\gamma,\sigma) = 0\},\\
Y_0^C &= \{\sigma \,|\, \exists\,\gamma : C(\gamma,\sigma) = 0\}, & Y_2^C &= \{\sigma \,|\, \exists\,\gamma : C(\gamma,\sigma) = 2\},\\
Y_2^A &= \{\sigma \,|\, \exists\,\lambda : A(\sigma,\lambda) = 2\}, & Y_0^A &= \{\sigma \,|\, \exists\,\lambda : A(\sigma,\lambda) = 0\},\\
Z_0^A &= \{\lambda \,|\, \exists\,\sigma : A(\sigma,\lambda) = 0\}, & Z_2^A &= \{\lambda \,|\, \exists\,\sigma : A(\sigma,\lambda) = 2\},\\
Z_2^B &= \{\lambda \,|\, \exists\,\gamma : B(\lambda,\gamma) = 2\}, & Z_0^B &= \{\lambda \,|\, \exists\,\gamma : B(\lambda,\gamma) = 0\}.
\end{aligned} \tag{S32}$$

In short, set X_i^P is the set of γ's for which party P can potentially obtain output i, and similarly for the Y_i^P and Z_i^P sets for σ's and λ's, respectively.

We coarse-grain the possible outcomes by grouping outcomes L and R as χ, which means that the possible outcomes are now $a, b, c \in \{0, \chi, 2\}$. Then, according to Sec. B.1, the set of outcomes with nonzero probability in our setup are (up to permutations):

$$abc \in \{\chi\chi\chi, 0\chi\chi, 0\chi 2\}. \tag{S33}$$

Observe that:

- two 0's never appear at the same time, nor two 2's,
- 2 only appears together with exactly one χ and one 0.

These properties are simply due to the fact that the number of photons is conserved, and that at most two photons can end up in the same photodetector. Already from these observations we obtain some structure on the previously defined sets in three steps. We demonstrate the steps for the $\{X_i^P\}_{i,P}$ sets, but they can be done with the $\{Y_i^P\}_{i,P}$ and $\{Z_i^P\}_{i,P}$ similarly.

Proof that $X_2^B \cap X_2^C = \emptyset$, $X_0^B \cap X_0^C = \emptyset$:

This is a direct consequence of the previous observation (Eq. S33). There cannot be 4 photons among 4 parties or 0 photons in total for two parties.

Proof that $X_0^B \cup X_0^C = X$:

Assume by contradiction that $\exists\ \gamma^* \in X \setminus \left(X_0^B \cup X_0^C\right)$. Then by definition $\forall \sigma, \lambda$:

$$B(\lambda, \gamma^*) \in \{\chi, 2\},$$
$$C(\gamma^*, \sigma) \in \{\chi, 2\}.$$

Observe that when $\gamma = \gamma^*$, Alice must not answer $a = 2$, due to (S33). However, to do this, since she does not know the value of γ, Alice must *always* not answer $a = 2$. A similar conclusion can be drawn for the other parties, due to the cyclic symmetry. This, however, leads to a contradiction since parties can in general output 2, e.g., $p_t(a = 2) \neq 0$.

Proof that $X_0^B \cap X_2^B = \emptyset$, $X_0^C \cap X_2^C = \emptyset$:

Assume by contradiction that $\exists \gamma^* \in X_0^B \cap X_2^B$. Then $\exists \gamma_1, \gamma_2$ such that:

$$B(\gamma_1, \gamma^*) = 0,$$
$$B(\gamma_2, \gamma^*) = 2.$$

Charlie does not know λ, so if $\gamma = \gamma^*$, he knows he must answer χ for *any* σ, since that is the only symbol consistent with both 0 and 2. Thus, we have that:

$$\forall \sigma : \quad C(\gamma^*, \sigma) = \chi.$$

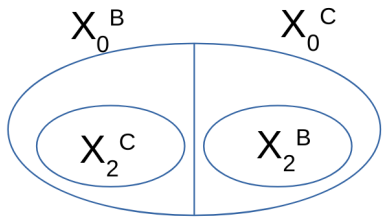


FIGURE S1: The relation of the sets $\{X_i^P\}_{i,P}$ to each other.

Say Alice receives $\lambda = \gamma_2$. Alice does not know whether $\gamma = \gamma^*$ or not. Thus, her response must be one that is consistent with the scenario that $\gamma = \gamma^*$. Because of Charlie's response being χ, this implies that for any σ she must answer $a = 0$, that is:

$$\forall \sigma : \quad A(\sigma, \gamma_2) = 0.$$

This means, by definition, that $Y_0^A = Y$. This implies, after doing steps 1 and 2 for the sets $\{Y_i^P\}_{i,P}$, that $Y_0^C = \emptyset$. However, since $p_t(c = 0) \neq 0$, we arrive at a contradiction.

Proof that all sets X_0^P, Y_0^P, Z_0^P have probability $1/2$:

The previous constraints B.2.1-B.2.1 on the sets X_i^P are can be summarized as in Fig. S1. We now give a partial quantitative assessment on the size of these sets. Note that by the definition of the sets Y_0^A and Z_0^A we have:

$$\frac{1}{4} = p_t(a = 0) \leq p(\sigma \in Y_0^A, \lambda \in Z_0^A) = p(Y_0^A)p(Z_0^A), \tag{S34}$$

where in the last step we used the statistical independence of the hidden variables. At the same time, by using the inequality $ab \leq ((a + b)/2)^2$ we have:

$$p(Y_0^A)p(Y_0^C) \leq \left(\frac{p(Y_0^A) + p(Y_0^C)}{2} \right)^2 = \frac{1}{4}.$$

Combining the two we see that $p(Y_0^C) \leq p(Z_0^A)$. Repeating the same argument (cyclically) for the other parties we get:

$$p(Y_0^C) \leq p(Z_0^A) \leq p(X_0^B) \leq p(Y_0^C),$$

which implies that they are all equal. Using also (S34) it is clear that all sets with $i = 0$ are equally probable with probability $\frac{1}{2}$, i.e.:

$$p(Y_0^C) = p(Z_0^A) = p(X_0^B) = p(Y_0^C) = p(Y_0^A) = p(Z_0^B) = p(X_0^C) = p(Y_0^A) = \frac{1}{2}. \tag{S35}$$

Equation (S35) combined with (S34) tells us that Alice, when receiving from Y_0^A on one side and from Z_0^A on the other, will *deterministically* output 0. The same holds for the other parties (Bob when receiving from X_0^B and Z_0^B, and Charlie when receiving from X_0^B and Y_0^B). This consideration combined with the previous ones and the definition of the sets (S32), yields a constrained picture of all possible classical models that simulate the coarse graining of p_t in the triangle network. This is illustrated in Fig. S2 where:

- Alice outputs 0 when receiving from Y_0^A and Z_0^A.
- Alice outputs χ when receiving from Y_0^A and Z_0^B, or when receiving from Y_0^C and Z_0^A (in both cases Alice cannot output $a = 2$ because of property B.2.1).
- Alice outputs either χ or 2 when receiving from Y_0^C and Z_0^B, (further structure can be given using the sets Z_2^B and X_2^B).

Bob and Charlie follow similar strategies when cycling the indices.

B.2.2 Breaking up the coarse-graining

Now, the main idea is the following: If there exists a local model for $p_t(a, b, c)$ as from Fig. S2, then there should exist a distribution $q_t(i, j, k, s)$ representing the parties collective response function $(i, j, k = L, R)$ when the hidden variables γ, σ, λ come from $X_0^C \times Y_0^A \times Z_0^B$ (s=0) or $S_1 = X_0^B \times Y_0^C \times Z_0^A$ (s=1). We cannot directly derive $q_t(i, j, k, s)$

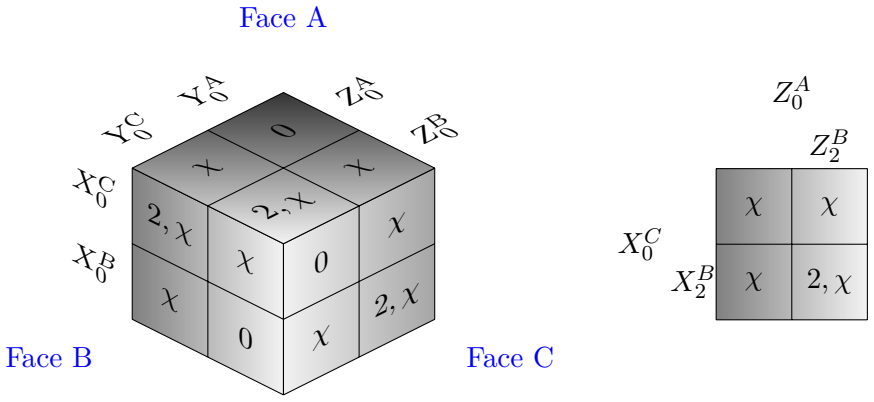


FIGURE S2: (Left) Classical strategies visualized on a cube. The edges of the cube represent the interval $[0, 1]$, on which the hidden variables $\{\gamma, \sigma, \lambda\}$ distributed, and the sets (S32) are represented, all having probability $1/2$ due to (S35). The labels on faces are the possible responses of a given party. (Right) The $2, \chi$ part on face B, for example, can be further decomposed using X_2^B and Z_2^B.

from $p_t(a, b, c)$, however, we can derive its marginals (see below). These marginals will be incompatible for some values of transmissivity t. For these situations, thus, we can deduce that there does not exist a local model for $p_t(a, b, c)$. Additionally, the marginals constraints on $q_t(i, j, k, s)$, are the same as in [Ren+19a] for the distribution p'_t, meaning that the classical feasibility of p_t implies the classical feasibility of p'_t, as stated in (S29).

To start, consider the two sets $S_0 = X_0^C \times Y_0^A \times Z_0^B$ and $S_1 = X_0^B \times Y_0^C \times Z_0^A$. Note that $S_0 \cap S_1 = \emptyset$ and the events $\chi\chi\chi$ can happen if and only if $(\gamma, \sigma, \lambda) \in S_0 \cup S_1$. We define:

$$q_t(i, j, k, s) = p(a = i, b = j, c = k, (\gamma, \sigma, \lambda) \in S_s \,|\, (\gamma, \sigma, \lambda) \in S_0 \cup S_1), \quad \text{(S36)}$$

where the indices i, j, k are each either L or R, and the index s is either 0 or 1. This is a probability distribution, since if $(\gamma, \sigma, \lambda) \in S_0 \cup S_1$, then it must be either in S_0 or S_1, and all parties must output either L or R

(hence normalization and positivity are satisfied). Using the definition of conditional probability and the fact that the sets S_0 and S_1 have probability 1/8 (cf. B.2.1 and Fig. S2), we see that:

$$q_t(i,j,k,s) = 4p(a=i, b=j, c=k, (\gamma,\sigma,\lambda) \in S_s). \tag{S37}$$

Marginalizing over s gives us:

$$\boxed{q_t(i,j,k) = 4p_t(a=i, b=j, c=k)}, \tag{S38}$$

the value of which is given by the parameters of the model, e.g., the transmissivity.

Next we would like to express other marginals, for example, $q_t(i,s) \equiv \sum_{jk} q_t(i,j,k,s)$, as a function of the target probability distribution. To do this, first note that if $b=0$ then $\gamma \in Z_0^B, \lambda \in X_0^B$ and *either* $\sigma \in Y_0^A$ *or* $\sigma \in Y_0^C$. Note that next to a 0 output we can only have the other two parties answering $\{\chi, 2\}$ or $\{\chi,\chi\}$. For $q_t(i,s)$ we are, however, interested in the probabilities of $a=i$, therefore we break up the χ in Alice's response. In terms of probabilities this means

$$\begin{aligned} p_t(a{=}i, b=0, c=\chi) + p_t(a{=}i, b=0, c=2) = \\ = p_t(a{=}i, (\gamma,\sigma,\lambda) \in X_0^B \times Y_0^A \times Z_0^B) + p_t(a{=}i, (\gamma,\sigma,\lambda) \in X_0^B \times Y_0^C \times Z_0^B) \end{aligned} \tag{S39}$$

where we used colours to simplify the reading, separating the sets in a local strategy on which Bob bases his choice (in blue), from those to which he has no access (in red). From now on we use a shorthand for expressions like this, indicating, for example, $(\gamma,\sigma,\lambda) \in X_0^B \times Y_0^C \times Z_0^B$ simply as $X_0^B Y_0^C Z_0^B$.

Next, consider the sum where we force Alice to output i, but Bob and Charlie can either output 0 or χ. In other words we are focusing on the $\chi 0\chi, \chi 02, \chi\chi 0, \chi 20$ outputs, breaking coarse-graining $\chi \to L, R$ only in Alice's case. Define the quantity D_A^i as:

$$\begin{aligned} D_A^i := p_t(a=i, b=0, c=\chi) + p_t(a=i, b=0, c=2) + \\ - \left[p_t(a=i, b=\chi, c=0) + p_t(a=i, b=2, c=0\right]. \end{aligned} \tag{S40}$$

A few manipulations show that

$$\begin{aligned} D_A^i = p_t(a=i, X_0^B Y_0^A Z_0^B) + p_t(a=i, X_0^B Y_0^C Z_0^B) + \\ - \left[p_t(a=i, X_0^C Y_0^C Z_0^A) + p_t(a=i, X_0^C Y_0^C Z_0^B)\right] = \\ = p_t(a=i, \underline{\mathbf{X_0^C} Y_0^A Z_0^B}) + p_t(a=i, \mathbf{X_0^C} Y_0^C Z_0^B) + \\ - p_t(a=i, \underline{\mathbf{X_0^B} Y_0^C Z_0^A}) - p_t(a=i, X_0^C Y_0^C Z_0^B) = \\ = p_t(a=i, \underline{S_0}) - p_t(a=i, \underline{S_1}), \end{aligned}$$

where we first used (S39) (and a similar expression for $c=0$), and then that Alice does not have access to γ, so the probabilities stay the same under the swap of X_0^B for $\mathbf{X_0^C}$ and X_0^C for $\mathbf{X_0^B}$. Finally, we identified S_0 and S_1 in the relevant expressions. Hence, we could express the differences of $q(i, s=0)$ and $q(i, s=1)$ as an expression of known terms. We also know that the sum is

$$\begin{aligned} p_t(a=i, S_0) + p_t(a=i, S_1) = p(a=i, b=\chi, c=\chi) = \\ = \sum_{j,k=L,R} p(a=i, b=j, c=k). \end{aligned} \tag{S41}$$

Combining the two we get that:

$$\boxed{q(i, s=0) = 2p_t(a=i, b=\chi, c=\chi) + 2D_A^i}, \tag{S42}$$

$$\boxed{q(i, s=1) = 2p_t(a=i, b=\chi, c=\chi) - 2D_A^i}. \tag{S43}$$

B.2.3 Testing $q_t(i,j,k,s)$ using linear programming

We sum up here the marginal properties (boxed equations in the previous section B.2.2) of the distribution $q_t(i,j,k,s)$ found above. These properties are linear constraints on the vector $q_t(i,j,k,s)$ which are parametrized by the transmissivity t. A linear program can be implemented to verify if a distribution a $q_t(i,j,k,s)$ compatible with these marginals exists.

Constraint 0 (normalization). First of all,

$$q_t(i,j,k,s) \geq 0 \ \forall i,j,k,s \qquad \text{and} \qquad \sum_{i,j,k,s} q_t(i,j,k,s) = 1 \tag{S44}$$

that is, it truly represents a probability vector.

Constraint 1. Then:

$$\sum_s q_t(i,j,k,s) = 4p_t(a=i,b=j,c=k), \tag{S45}$$

with (cf. Sec. B.1):

$$\begin{aligned} 4p_t(RRL) &= \frac{1}{2}t(1-t)(1+2\sqrt{t(1-t)}), \\ 4p_t(RLL) &= \frac{1}{2}t(1-t)(1-2\sqrt{t(1-t)}), \qquad \text{(S46)} \\ 4p_t(LLL) &= \frac{1}{2}((1-t)^{\frac{3}{2}}+t^{\frac{3}{2}})^2, \qquad \text{(S47)} \\ 4p_t(RRR) &= \frac{1}{2}((1-t)^{\frac{3}{2}}-t^{\frac{3}{2}})^2, \qquad \text{(S48)} \end{aligned}$$

and cyclic combinations (meaning only the number of Ls and Rs matters).

Constraint 2. This constraint is actually a consequence of Constraint 1, but we write it for completeness:

$$\sum_{s,j,k} q_t(i,j,k,s) = 4p_t(a=i,\chi,\chi) \quad \text{and cyclic cases } i\to j\to k\to i. \tag{S49}$$

Constraint 3.

$$\begin{aligned} &\sum_{j,k} q_t(i,j,k,s=0) - q_t(i,j,k,s=1) = \\ &\quad = 4\big[p_t(a=i,0,\chi) + p(a=i,0,2) - p(a=i,\chi,0) - p(a=i,2,0)\big] \end{aligned} \tag{S50}$$

and cyclic combinations. This last constraint can be made explicit (cf. Sec. B.1):

$$\begin{aligned} i=L \to 4\big[&p_t(a=i,0,\chi) + p_t(a=i,0,2) \\ &- p_t(a=i,\chi,0) - p_t(a=i,2,0)\big] = \frac{1}{2} - t, \end{aligned} \tag{S51}$$

$$i = R \to 4\big[p_t(a = i, 0, \chi) + p_t(a = i, 0, 2) - p_t(a = i, \chi, 0) - p_t(a = i, 2, 0)\big] = t - \frac{1}{2}. \quad \text{(S52)}$$

Relation to Ref. [Ren+19a] and equivalence between p_t and p'_t The constraints defining the linear program above, can be translated to be the same constraints of a linear program found in Ref. [Ren+19a], where the distribution p'_t (S27) is considered (in [Ren+19a] t is identified as u^2). Specifically, both the distributions p_t and p'_t are local if a solution $\bar{q}_t(i, j, k, t)$ to the same linear program exists and can be generated via a local model (cf. [Ren+19a]). This proves that the local feasibility of p_t is equivalent to that of p'_t. At the same time, the existence of $\bar{q}_t(i, j, k, t)$ is a *necessary condition* for the local feasibility of p_t. This means that when the linear program fails to find a solution, the nonlocality of p_t is certified, while if a solution is found, this does not directly imply the locality of p_t.

The Linear Program resulting from the constraint above is infeasible for $t \in (0, 0.215)$ and $t \in (0.785, 1)$.

B.3 Noisy optical realization

As introduced in the main material, after proving the nonlocality of the idealized experiment, in this section we give the modelling details of the imperfections that can arise in the different elements of the optical network presented in Fig. 4.1, when realized experimentally. We focused on:

a. the impurity of the generated single-photon entangled state (Q),
b. the transmissivity of the optical channels (T) of the network, and
c. the efficiency of the final photodetectors (ν).

Our results (see Chapter 4) indicate that the noise tolerance with respect to these parameters is of the order of few percentage points, which makes the proposal very stringent from the experimental point of view, but possible on a table-optical experiment with high-efficient detectors.

Source imperfections. Firstly, we considered a realistic process of creation for the single photon entangled state $|\psi^+\rangle = (|01\rangle + |10\rangle)/\sqrt{2}$. This is generated by a single photon sent onto a 50:50 beamsplitter. Typical sources achieve the heralding of single photons from two-photon states created in a

SPDC process, followed by the detection of one of the two photons [Boy20; Cou18].

An externally controlled laser pulses at high frequency on a $\chi^{(2)}$ non-linear crystal. For each pulse, the crystal consequently outputs a two-mode squeezed vacuum state $|\Psi\rangle \propto \sum_n q^n |nn\rangle$. Then, photodetection is performed on one of the two modes. Conditioning on a detection allows to isolate a very good approximation of the one-photon Fock state on the unmeasured mode [CS12]. The trade-off between probability of heralding and quality (fidelity to target) of the heralded state is strongly conditioned by the photodetector efficiency and ability to resolve photon number, as well as the characteristics of the crystal and the laser power, which tune the value of q [CS12]. Here we chose typical currently achievable values for the SPDC, which we assume to have $q = 0.01$ and 10MHz frequency of the pulses [Cas+20]. The heralding is simulated by currently available number-resolving photodetectors which we assume to have 8-photon resolution achieved with an array of $M = 8$ single photon detectors pixels, having each a $\eta = 70\%$ efficiency, well in the range of present technologies [Mos+19; Zhu+20]. Conditioning on the firing of a single pixel in the detector, the resulting state in the unmeasured mode can be approximated by:

$$\varrho \sim (1 - Q) |1\rangle \langle 1| + Q |2\rangle \langle 2| + \mathcal{O}(Q^2), \tag{S53}$$

where $Q \propto q$ is the ratio between the chance of obtaining a single pixel firing due to a double-photon hitting the detector, and the chance of obtaining a single pixel fire due to a single photon, that is:

$$Q = \frac{q^2 \left(\frac{1}{M}(1 - (1 - \eta)^2) + 2\frac{M-1}{M}\eta(1 - \eta) \right)}{q\eta}. \tag{S54}$$

Note that the probability of heralding is $q\eta$ and thus for the three sources (of the experiment proposed in the main text) to be heralded at the same time, the corresponding total experimental repetition rate is of approximately $q^3\eta^3 10\text{MHz} \sim 1\text{Hz}$ [1]. Considering the imperfect state ϱ (S53), propagated through a 50:50 beamsplitter, the resulting true source shared by each couple

[1] Notice that lasers pulsed at GHz rates have been used recently [Nga+15], which would result in an experimental repetition rate of $\sim$ KHz.

in the triangle network is:

$$\rho = (1 - Q) \left|\psi^+\right\rangle \left\langle\psi^+\right| + Q \left|\varphi\right\rangle \left\langle\varphi\right| + \mathcal{O}(Q^2). \tag{S55}$$

With the above-mentioned values of q, η, and M, it results $Q = 0.006875$.

Notice that the same single-photon preparation could be done with simple, non-number-resolving (NNR) photodetection. In such a case the value of Q (which we remind, is the ratio between the chance of the detector clicking due to a double-photon, and the chance of a click due to a single-photon), would be:

$$Q^{(\mathrm{NNR})} = \frac{q^2(1 - (1 - \eta)^2)}{q\eta} = q(2 - \eta), \tag{S56}$$

where η is the efficiency of the detectors. We see that in such a case Q is bounded to be larger than q, for example with the same values above ($q = 0.01$, $\eta = 70\%$), one obtains $Q^{(\mathrm{NNR})} = 0.013$, essentially double what can be obtained with number-resolving detectors. This is not a huge limitation per se, as we can rescale q to make $Q^{(\mathrm{NNR})}$ smaller. At the same time, halving q makes the total repetition rate of the experiment ($\propto q^3\eta^3$) decrease by one order of magnitude.

Finally, let us notice how basing our proposal on the single-photon state $|\psi^+\rangle \propto |01\rangle + |10\rangle$ is crucial in our scenario. A unitarily equivalent state is the two-photon state $\propto |HV\rangle + |VH\rangle$, which encodes the information in the polarization degree of freedom. However the creation of such state from a SPDC source typically needs the heralding of the 6-photons term $|33\rangle$ from $\sum_n q^n |nn\rangle$ (and 4 photodetectors per source) [SB03]. This means that even in an ideal scenario in which all detectors have unit efficiency, the probability of heralding the correct state would be $\sim q^3$, and for the whole experiment with 3 sources, q^9, compared to q^3 for our single-photon proposal. For a 1% error in the source, we chose $q = 0.01$, which is translated into 12 orders of magnitude of difference in the heralding rate.

Losses in the channels. Secondly, loss might happen during the transmission along the channels that form the sides of the triangle network of Fig. 4.1, before the local POVM performed by the parties. We denote by T the transmissivity of these optical channels. The resulting correction due to

photon loss can be computed as:

$$\delta\rho = \sum_{n,X_i} K_{X_i}^{(n)} \rho K_{X_i}^{(n)\dagger}, \tag{S57}$$

where Kraus operators of the form:

$$K_{X_i}^{(n)} = \sqrt{(1-T)}\sqrt{n}\,|n-1\rangle_{X_i}\langle n|\,, \tag{S58}$$

act on each of the six modes X_i, and the sum is truncated to $n = 1, 2$ (given the support of input state (S55)). In fact, as we work in the regime of low losses, we only keep the first-order terms in $1 - T$ in Eq. (S57).

Detectors Finally, the photodetectors used at the vertices of the triangle (Fig. 4.1) do not resolve photon number, and are assumed to have a finite, high efficiency ν, thus modelled, at first order in $1 - \nu$ as:

$$\begin{aligned} D^{\square}(\nu) &= |0\rangle\langle 0| + (1-\nu)\,|1\rangle\langle 1|\,, \\ D^{\blacksquare}(\nu) &= \mathbb{1} - (1-\nu)\,|1\rangle\langle 1| - |0\rangle\langle 0|\,. \end{aligned} \tag{S59}$$

Notice that high efficiencies close to 100% have been reached by modern photodetection systems [NTH12; LMN08; Mil+11; Fuk+11; Red+19].

B.4 Generalization to chains of N parties

In this section, we sketch a generalization of the experiment presented in the main text (which is proposed in the triangle scenario), to a chain of N parties in a circular network. For such case, we generalize the procedure carried out through Sec. B.2 which proves the existence of a range of transmissivities for which the network output is nonlocal.

The generalized experiment is described as follows: N parties A_i share a copy of the single photon state $|\psi^+\rangle = \frac{|01\rangle+|10\rangle}{\sqrt{2}}$ for each couple of neighbouring parties A_iA_{i+1} with $i = 1, \ldots, N$ (the total network is circular and thus we identify $N + 1 \equiv 1$). Each party consequently receives two input

modes containing at most 1 photon, and performs the same measurement described in the main text (4.6), and detailed in Sec. B.1, consisting in a local mixing of the modes with a beamsplitter of transmissivity t, followed by photodetection on both modes. All the parties choose the same value for t and the photodetectors do not resolve the number of photons, thus being described by projective measurements on vacuum and its orthogonal complement $M^{\square} = |0\rangle\langle 0|$, $M^{\blacksquare} = \mathbb{1} - |0\rangle\langle 0|$. Consequently, the resulting output distribution is given by:

$$p_t(a_1, \ldots, a_n) = \mathrm{Tr}\left[\left(\bigotimes_{i=1}^{N} \psi^{+}_{A_i^{(R)} A_{i+1}^{(L)}}\right)\left(\bigotimes_{j=1}^{N} \Pi^{(a_j)}_{t\, A_j^{(L)} A_j^{(R)}}\right)\right] \qquad a_j = 0, L, R, 2, \quad \text{(S60)}$$

where the state $\psi^{+} \equiv |\psi^{+}\rangle\langle\psi^{+}|$ is shared between each "right mode" of the ith party ($A_i^{(R)}$) and the "left mode" ($A_{i+1}^{(L)}$) of the following, and each party performs the POVM operationally described above, corresponding to Π_t (4.6) (detailed in Eqs (S7)-(S10)) on its two modes.

We now put constraints on any possible local strategy aiming at reproducing the same statistical output of p_t in the circular network. That is we assume p_t can be written as:

$$p_t(a_1 \ldots a_N) = \int p(a_1|\alpha_{N1}\alpha_{12}) p(a_2|\alpha_{12}\alpha_{23}) \ldots \\ \ldots p(a_N|\alpha_{(N-1)N}\alpha_{N1})\, \mathrm{d}\alpha_{12}\mathrm{d}\alpha_{23} \ldots \mathrm{d}\alpha_{N1}, \quad \text{(S61)}$$

where a_i is the output of party A_i, which is based on a local response on the hidden variables $\{\alpha_{i(i+1)}, \alpha_{(i-1)i}\}$ shared with his left and right neighbours. In the coarse grained scenario, parties can output $0, \chi, 2$ as before (χ is the coarse graining of $\{L, R\}$, cf. Sec. B.2), representing the outcomes with 0, 1, or 2 photodetectors firing respectively at each party station. Following Sec. B.2 we define the equivalent of the sets (S32), accompanying the formal definitions with an intuitive notation and explanation of the underlying local model; the sets are represented by arrows that intuitively suggest the direction of "classical photons" in a corresponding local hidden variable model. The following definitions are pictured in Figure S3. We have formally,

for the set of sources $\alpha_{(k-1)k}$ between A_{k-1} and A_k ,

$$(\nrightarrow^k) := \{\alpha_{(k-1)k} \,|\, \exists\, \alpha_{k(k+1)} : A_k(\alpha_{(k-1)k}, \alpha_{k(k+1)}) = 0\} \tag{S62}$$

This is the set allowing A_k to output 0 for some of the hidden variables that come from the other side. (S63)

That is, *classical photons are not sent to A_k from the left.*

$$({}^{k-1}\nleftarrow) := \{\alpha_{(k-1)k} \,|\, \exists\, \alpha_{(k-2)(k-1)} : A_{k-1}(\alpha_{(k-2)(k-1)}, \alpha_{(k-1)k}) = 0\} \tag{S64}$$

This is the set allowing A_{k-1} to output 0 for some of the hidden variables that come from the other side. (S65)

That is, *classical photons are not sent to A_{k-1} from the right.*

$$(\rightarrow^k) := \{\alpha_{(k-1)k} \,|\, \exists\, \alpha_{k(k+1)} : A_k(\alpha_{(k-1)k}, \alpha_{k(k+1)}) = 2\} \tag{S66}$$

This is the set allowing A_k to output 2 for some of the hidden variables that come from the other side. (S67)

That is, *some classical photons are sent to A_k from the left.*

$$({}^{k-1}\leftarrow) := \{\alpha_{(k-1)k} \,|\, \exists\, \alpha_{(k-2)(k-1)} : A_{k-1}(\alpha_{(k-2)(k-1)}, \alpha_{(k-1)k}) = 2\} \tag{S68}$$

This is the set allowing A_{k-1} to output 2 for some of the hidden variables that come from the other side. (S69)

That is, *some classical photons are sent to A_{k-1} from the right.*

B.4.1 Constraints on the sets

We here derive in this generalized N-party scenario the constraints on any local model reproducing p_t corresponding to those obtained for the triangle network (B.2.1 to B.2.1).

As depicted in Fig. S3 we have, firstly:

$$(\nrightarrow^k) \cap ({}^{k-1}\nleftarrow) = \emptyset, \tag{S70}$$

because otherwise two neighbouring parties A_{k-1}, A_k, would be allowed to output 0 at the same time, which is in contrast with the output of p_t (the photon shared between two parties ends up in one of their detectors).

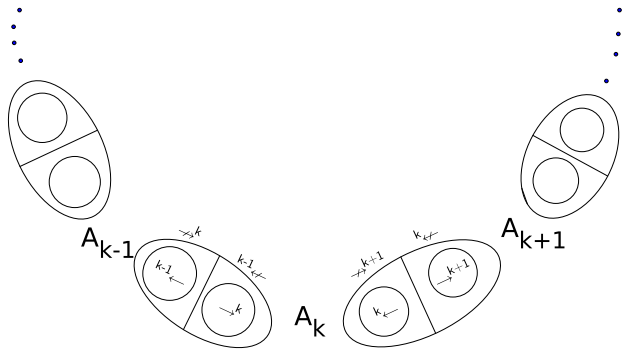


FIGURE S3: Generalized setting with N parties A_k, and representation of the sets (S62)-(S68) describing a local strategy that simulates p_t (cf. Eqs. (S60) and (S61)).

Secondly:

$$(\nrightarrow^{k}) \cup ({}^{k-1}\nleftarrow) = 1, \tag{S71}$$

meaning that, together, the two sets form the total set of sources $\alpha_{(k-1)k}$ between A_{k-1} and A_k. This is proven as a consequence of the fact that at least one between A_{k-1} and A_k must be allowed to output 0 (otherwise there would be a non-zero probability of more than N photodetectors firing, as in $\{a_{k-1} = \chi, a_k = \chi, a_{k+1} = 2, \chi, \chi, \chi \dots\}$). The initial total number of photons is N, therefore this cannot happen.

Thirdly we have:

$$(\rightarrow^{k}) \subseteq ({}^{k-1}\nleftarrow)\ , \tag{S72}$$

$$({}^{k-1}\leftarrow) \subseteq (\nrightarrow^{k})\ . \tag{S73}$$

This is true again because otherwise an event like $\{a_{k-1} = \chi, a_k = 2, a_{k+1} = \chi, \chi, \chi, \dots\}$ or $\{a_{k-1} = 2, a_{k-1} = \chi, a_{k+1} = \chi, \chi, \chi, \dots\}$ would have nonzero probability. All the above constraints are derived out of photon number conservation (note that in our optical setup, if we do not resolve the number of photons, sometimes we may lose track of some of them when they end up in the same detector, which is why we are not able to say that the above equations are equalities, but just inclusions).

Now, it is also true that:

$$\text{all the } (\twoheadrightarrow^k) \text{ and } ({}^k\twoheadleftarrow) \text{ sets have probability equal to } \frac{1}{2} \forall k \,. \tag{S74}$$

This can be proven by using the definitions as:

$$(\twoheadrightarrow^k) * ({}^k\twoheadleftarrow) \geq p(a_k = 0) = \frac{1}{4} = \left(\frac{({}^k\twoheadleftarrow) + (\twoheadrightarrow^{k+1})}{2} \right)^2 \geq ({}^k\twoheadleftarrow) * (\twoheadrightarrow^{k+1}) \tag{S75}$$

which implies $(\twoheadrightarrow^k) \geq (\twoheadrightarrow^{k+1})$, but such inequality can be cycled until obtaining $(\twoheadrightarrow^k) \geq (\twoheadrightarrow^k)$, which entails that all the inequalities are actually equalities.

B.4.2 Constraints on the local coarse grained strategy

Given (S74), we have that the parties will output deterministically 0 when allowed from both sides, as they have to simulate $p(0) = \frac{1}{4}$. Summing up we have:

$$(\twoheadrightarrow^k) A_k ({}^k\twoheadleftarrow) \quad \Rightarrow A_k \text{ outputs } 0, \tag{S76}$$

$$({}^{k-1}\twoheadleftarrow) A_k ({}^k\twoheadleftarrow) \quad \text{or} \quad (\twoheadrightarrow^k) A_k (\twoheadrightarrow^{k+1}) \quad \Rightarrow A_k \text{ outputs } \chi, \tag{S77}$$

$$({}^{k-1}\twoheadleftarrow) A_k (\twoheadrightarrow^{k+1}) \quad \Rightarrow A_k \text{ outputs } \chi \text{ or } 2. \tag{S78}$$

B.4.3 Breaking the coarse-graining and finding linear constraints

Here we repeat and generalize the scheme presented in B.2.2 to give linear constraints on a subset of the local response functions. We define $q_t(i_1, i_2, ..., i_N, s)$, analogously to (S36) as the probability of outputting $\{\chi_{i_1}, \chi_{i_2}, ..., \chi_{i_N}, \}$ given sources γs coming from left ($s = 0$) or right ($s = 1$) part of the sets drawn in Fig. S3, i.e.

$$q_t(i_1, i_2, ..., i_N, s) = \\ = \begin{cases} p_t(a_1 = \chi_{i_1}, a_2 = \chi_{i_2}, ..., a_N = \chi_{i_N}, (\alpha_{12}, \alpha_{23}, \dots) \in S_0) & s = 0 \\ p_t(a_1 = \chi_{i_1}, a_2 = \chi_{i_2}, ..., a_N = \chi_{i_N}, (\alpha_{12}, \alpha_{23}, \dots) \in S_1) & s = 1 \end{cases} \tag{S79}$$

where we formally define the above mentioned sets as:

$$S_0 := \times_j(\twoheadrightarrow^j) \equiv (\twoheadrightarrow^1) \times (\twoheadrightarrow^2) \times \cdots \times (\twoheadrightarrow^N) ,$$
$$S_1 := \times_j({}^j\twoheadleftarrow) \equiv ({}^1\twoheadleftarrow) \times ({}^2\twoheadleftarrow) \times \cdots \times ({}^N\twoheadleftarrow) . \tag{S80}$$

Given that the configurations of sources $\times_j(\twoheadrightarrow^j)$ and $\times_j({}^j\twoheadleftarrow)$ are the only ones allowing possible outputs being all χ, q satisfies the following equality involving one of its marginal distributions:

$$\sum_{s=0,1} q_t(i_1, i_2, ..., i_N, t) = p_t(\chi_{i_1}, \chi_{i_2}, ..., \chi_{i_N}) . \tag{S81}$$

We now consider instead the marginal on i_k and t:

$$q_t(i_k, s) = \sum_{i_1,...,i_{k-1},i_{k+1},...,i_N} q_t(i_1, i_2, ..., i_N, s).$$

This satisfies:

$$q_t(i_k, 0) - q_t(i_k, 1) = \frac{p_t(a_k = \chi_i, a_{k+1} = 0) - p_t(a_{k-1} = 0, a_k = \chi_i)}{2^{N-3}}. \tag{S82}$$

The proof of this equation is formalized as follows:

$$\begin{aligned}
&q_t(i_k, 0) - q_t(i_k, 1) \\
&\quad = p_t(a_k = \chi_{i_k}, \times_j(\twoheadrightarrow^j)) - p_t(a_k = \chi_{i_k}, \times_j({}^j\twoheadleftarrow)) \\
&\quad\quad = p_t\left(a_k = \chi_{i_k}, (\times_{j\neq k+2}(\twoheadrightarrow^j)) \times ({}^{k+1}\twoheadleftarrow)\right) \\
&\quad\quad - p_t\left(a_k = \chi_{i_k}, (\times_{j\neq k-2}({}^j\twoheadleftarrow)) \times (\twoheadrightarrow^{k-1})\right) \\
&\quad = \frac{1}{2^{N-3}}\Big[p_t\left(a_k = \chi_{i_k}, (\twoheadrightarrow^k) \times (\twoheadrightarrow^{k+1}) \times ({}^{k+1}\twoheadleftarrow)\right) \\
&\quad\quad - p_t\left(a_k = \chi_{i_k}, (\twoheadrightarrow^{k-1}) \times ({}^{k-1}\twoheadleftarrow) \times ({}^{k}\twoheadleftarrow)\right)\Big] \\
&\quad\quad = \frac{p_t(a_k = \chi_i, a_{k+1} = 0) - p_t(a_{k-1} = 0, a_k = \chi_i)}{2^{N-3}}.
\end{aligned} \tag{S83}$$

The first equality above simply follows from the definition of $q_t(i_k, s)$ for $s = 0, 1$. The second equality is obtained by noticing that all sets $(\twoheadrightarrow^j)$ and $({}^j\twoheadleftarrow)$ have probability $1/2$, and that the output a_k does not depend on the source shared between A_{k+1} and A_{k+2}, nor it depends on the source shared

between A_{k-2} and A_{k-1}. The third equality is obtained by tracing out the probability of $N-3$ of the sets which were included in the previous lines. Finally, the last inequality is implied by property (S76).

The above constraints on q_t coincide with the ones derived in the Appendix C of [Ren+19a]. There, it is proven that it is always possible to choose the value of the transmissivity t such that no solution can be found for $q_t(i_1, i_2, \ldots, i_N, s)$ satisfying the linear constraints (S81) and (S82). Therefore, for those values t the output p_t of the experiment is proven to be nonlocal.

Appendix C

Broadcasting appendices

C.1 Heuristic method for device-independent entanglement certification

To search for a witness for $\rho_{\alpha>0.338}$, we employ the following heuristic optimization.

1. Pick random (projective) measurements $\{A_{a|x}\}$, $\{B_{b|y}\}$, $\{C_{c|z}\}$, $\{D_{d|w}\}$, and channels $\sigma_\lambda^{AB} = \Omega_{A_0\to AB}[\sigma_\lambda^{B_0}]$, $\sigma_\lambda^{CD} = \Omega_{C_0\to CD}[\sigma_\lambda^{C_0}]$.
2. Find α^* such that the resulting correlation from ρ_α is on the boundary of $\mathcal{Q}_{AB|CD}^{PPT,1}$. This can be done via a semi-definite programming described below.
3. Extract corresponding inequality F.
4. For state ρ_{α^*}, optimize the inequality F over all POVMs $\{A_{a|x}\}$, $\{B_{b|y}\}$, $\{C_{c|z}\}$, $\{D_{d|w}\}$ and channels $\sigma_\lambda^{AB} = \Omega_{A_0\to AB}[\sigma_\lambda^{B_0}]$, $\sigma_\lambda^{CD} = \Omega_{C_0\to CD}[\sigma_\lambda^{C_0}]$.
5. Repeat point 2-4 until two successive values of α^* are identical.

In order to find the value such that ρ_{α^*} is on the boundary of $\mathcal{Q}_{AB|CD}^{PPT,1}$, one can run the following SDP

$$\begin{aligned} &\text{maximize } \alpha \\ &\text{s.t. } \Gamma(p(abcd|xyzw)) \succeq 0 \end{aligned}$$

hierarchy by employing the corresponding moment matrix in the above optimization.

C.2 Efficient method for computing the LHS bound of steering inequalities

We now describe the method for computing LHS bounds of steering inequalities used in our work. A similar formula was previously used in Refs. [SC15] and [Ben+12]. We derive a proof here for completeness:

For an assemblage, $\sigma_{a|x}$ a steering inequality is of the form:

$$\sum_{a,x} \mathrm{Tr}(F_{a|x}\sigma_{a|x}) \leq L \tag{S1}$$

where $F_{a|x}$ are matrices of the same dimension of $\sigma_{a|x}$, and L is the LHS bound of the inequality, that is, the maximal value attained with a LHS assemblage. Formally:

$$L = \max_{\sigma_\lambda} \left\{ \sum_{a,x} \mathrm{Tr}(F_{a|x}\sigma_{a|x}) \;\middle|\; \sigma_{a|x} = \sum_\lambda \sigma_\lambda D(a|x\lambda),\ \sigma_\lambda \geq 0,\ \mathrm{Tr}\left(\sum_\lambda \sigma_\lambda\right) = 1 \right\} \tag{S2}$$

where λ runs over all deterministic strategies $D(a|x\lambda)$. From equation (S2), one can see that the LHS bound L can be computed with a SDP optimization (linear objective function and SDP conditions of the variables σ_λ). However, one can devise a more efficient formula to compute it. Let us define $M_k := \sum_{a,x} F_{a|x} D(a|xk)$ and consider the inequality applied to an unsteerable assemblage:

$$\sum_{a,x} \mathrm{Tr}\left(F_{a|x}\sigma^{\mathrm{LHS}}_{a|x}\right) = \sum_{a,x} \mathrm{Tr}\left(F_{a|x} \sum_k \sigma_k D(a|xk)\right) \tag{S3}$$

$$= \sum_k \mathrm{Tr}\left(\sigma_k \sum_{a,x} F_{a|x} D(a|xk)\right) \tag{S4}$$

$$= \sum_k \mathrm{Tr}\left(\sigma_k M_k\right) = \sum_k p_k \,\mathrm{Tr}\left(\hat{\sigma_k} M_k\right) \tag{S5}$$

$$\leq \max_k \mathrm{Tr}\left(\hat{\sigma_k} M_k\right) \leq \max_k \lambda_M(M_k) \tag{S6}$$

where $\hat{\sigma}$ means $\mathrm{Tr}(\hat{\sigma}) = 1$ and $\lambda_M(A)$ means the largest eigenvalue of A. Moreover, one can see that the bound is tight (it can be achieved by setting all σ_k to 0 but the one corresponding to the M_k with the maximal largest eigenvalue, which is set to the projector onto the corresponding eigenvector). All in all we have that:

$$L = \max_k \lambda_M(M_k)\,. \tag{S7}$$

C.3 Heuristic search for certifying broadcast steering of bipartite states

Here we describe how we searched for interesting examples of steering in the broadcast scenario. For convenience, we consider the scenario featuring broadcasting to 3 devices, see the scenario of figure 5.2a. Note however that it extends straightforwardly to more devices. Let us consider a family of state of the form:

$$\rho_v = v\rho_{NL} + (1-v)\rho_{SEP}, \tag{S8}$$

where ρ_{NL} is typically a Bell nonlocal state while ρ_{SEP} is separable, and the linear parameter $0 \leq v \leq 1$. For example, the isotropic state of two qubits is of that form:

$$\rho_\alpha = \alpha|\phi^+\rangle\langle\phi^+| + (1-\alpha)\frac{\mathbb{1}}{4}. \tag{S9}$$

Here the goal is to find the smallest possible v such that the state exhibits broadcast steering. We used the following procedure:

1. Pick random (projective) measurements $\{B_{b|y}\}$, $\{C_{c|z}\}$ and channel $\Omega_{B_0\to BC}$.
2. Find v^* such that the resulting assemblage using state ρ_v is broadcast steerable. This can be done via a semi-definite programming described below.
3. Extract corresponding steering inequality F.
4. For state ρ_{v^*}, optimize steering inequality F over all POVMs $\{B_{b|y}\}$, $\{C_{c|z}\}$ and channels $\Omega_{B_0\to BC}$.

5. Repeat point 2-4 until two successive values of v^* are identical.

In order to find the value such that ρ_v is broadcast steerable for fixed measurements and channel (step 2), one can run the following SDP:

$$\begin{aligned}&\text{maximise } v\\ &\text{s.t. } \mathrm{Tr}_{BC}(\mathbb{1}\otimes B_{b|y}\otimes C_{c|z}[\mathbb{1}\otimes\Omega_{B_0\to BC}(\rho_v)]) = \sum_k \sigma_k D_{\mathcal{NS}}(bc|yzk),\\ &\quad \sigma_k \geq 0, \mathrm{Tr}(\sum_k \sigma_k) = 1,\end{aligned}$$

where the σ_k (together with v) are the SDP variables and the $D_{\mathcal{NS}}(bc|yzk)$ are the extremal non-signalling strategies between B and C. The dual variables of the equality constraints of this SDP provide a witness F, that is, a steering inequality of the form:

$$\sum_{a,x} \mathrm{Tr}(F_{a|x}\sigma_{a|x}) \leq L, \tag{S10}$$

here $F_{a|x}$ are matrices of the same dimension of $\sigma_{a|x}$, and L is the LHS bound of the inequality, that is, the maximal value attained with a LHS assemblage. Formally:

$$L = \max_{\sigma_\lambda} \left\{ \sum_{a,x} \mathrm{Tr}(F_{a|x}\sigma_{a|x}) \;\middle|\; \sigma_{a|x} = \sum_\lambda \sigma_\lambda D(a|x\lambda),\ \sigma_\lambda \geq 0,\ \mathrm{Tr}\left(\sum_\lambda \sigma_\lambda\right) = 1 \right\} \tag{S11}$$

where λ runs over all deterministic strategies $D(a|x\lambda)$. A formula to compute the LHS bound L of such an inequality is given in Appendix C.2. An algorithm to maximize such a steering inequality (step 4) over measurements and channels is given in Appendix C.3.1.

C.3.1 Optimizing a steering inequality

Assume one wants to maximize the violation of a steering inequality characterized by operators $F_{bc|yz}$ for a fixed state ρ_{AB_0} and over channels $\Omega_{B_0\to BC}$

and POVMs $\{B_{b|y}\}$, $\{C_{c|z}\}$. This means one wants to maximize:

$$\mathrm{Tr}\left[\sum_{b,c,y,z} F_{bc|yz}\,\mathrm{Tr}_{B_0BC}(\mathbb{1}_A \otimes B_{b|y} \otimes C_{c|z}\,[\mathbb{1}_A \otimes \Omega_{B_0\to BC}(\rho_{AB_0})]\right]. \quad \text{(S12)}$$

Both the objective function and the constraints on the variables are thus nonlinear, making the naive parametrization and optimization potentially inefficient. One can instead decompose the optimization on several subsets of variables, such that each optimization can be performed efficiently (aka see-saw optimization). Here, we used the following procedure:

1. Fix randomly POVMs $\{C_{c|z}\}$ and channel $\Omega_{B_0\to BC}$.
2. Optimize the inequality with respect to POVMs $\{B_{b|y}\}$, update variables accordingly.
3. Optimize the inequality with respect to POVMs $\{C_{c|z}\}$, update variables accordingly.
4. Optimize the inequality with respect to channels $\Omega_{B_0\to BC}$, update variables accordingly.
5. Repeat point 2 - 4 until two successive values of the inequality are equal (up to some desired precision).

The motivation for such a heuristic is that steps 2-4 can be written as single-shot SDPs. Indeed, for step 2 the constraints are $B_{b|y} \geq 0$ and $\sum_b B_{b|y} = \mathbb{1}$, and the objective function is linear. Step 3 is similar. For step 4, we can use the Choi-Jamiolkowski isomorphism [Cho75b]: the action of the map $\Omega_{B_0\to BC}$ on some state σ_{B_0} can be written as:

$$\Omega_{B_0\to BC}(\sigma_{B_0}) = \mathrm{Tr}_1(\rho_\Omega(\sigma_{B_0}^T \otimes \mathbb{1}_{BC})) \quad \text{(S13)}$$

where $\rho_\Omega \equiv d \cdot \mathbb{1} \otimes \Omega[|\Phi^+\rangle\langle\Phi^+|]$ is called the Choi state of the map $\Omega_{B_0\to BC}$ (where $|\Phi^+\rangle$ is the maximally entangled state of local dimension $d = \dim(\mathcal{H}_{B_0})$).

For valid channels, the Choi state satisfies $\rho_\Omega \geq 0$ and $\mathrm{Tr}_{BC}(\rho_\Omega) = \mathbb{1}_{B_0}$. The Choi-Jamiolkowski isomorphism ensures that for each state satisfying these two constraints, there is a unique corresponding channel. We can thus

use the variable ρ_Ω, which can be treated as a SDP variable, to solve step 4. One can indeed write the steering inequality as a linear function of ρ_Ω:

$$\mathrm{Tr}\left[\sum_{b,c,y,z} F_{bc|yz}\,\mathrm{Tr}_{B_0BC}((\mathbb{1}_A\otimes\mathbb{1}_{B_0}\otimes B_{b|y}\otimes C_{c|z})\,(\mathbb{1}_A\otimes\rho_\Omega)(\rho_{AB_0}^{T_{B_0}}\otimes\mathbb{1}_{BC}))\right]. \tag{S14}$$

Therefore, each step of the aforementioned procedure can be efficiently carried, since single-shot SDPs provide global optimums in polynomial time. In practice, we indeed observe that the entire see-saw optimization converges to what seems to be the global maximum in a few dozens seconds, for two-qubit states on bipartite and tripartite broadcast steering scenarios.

C.4 Proof of the lifting *ansatz* in Ineq. (5.10)

We rewrite Ineq. (5.10) for convenience:

$$\langle\mathcal{I}\,[A_0,\ldots,A_m,C_0,\ldots,C_k]\,(B_0+B_1)\rangle+\mathcal{L}_\mathcal{I}\langle A_{m+1}\,(B_1-B_0)\rangle\leq 2\mathcal{L}_\mathcal{I}\,. \tag{S15}$$

To prove it, we follow the same logic as in [BHC21, Sec. 4.1]. For the set of broadcast local distributions, the extremal strategies[1] consist of a deterministic strategy for A (this already implies $\langle A_xB_yC_z\rangle=\langle A_x\rangle\langle B_yC_z\rangle$), and, for B and C either a local deterministic strategy or a nonlocal extremal strategy.

Assuming a local deterministic strategy for B and C, this further implies $\langle B_yC_z\rangle=\langle B_y\rangle\langle C_z\rangle$. Ineq. (5.10) becomes:

$$\langle\mathcal{I}\,[A_0,\ldots,A_m,C_0,\ldots,C_k]\rangle\langle B_0+B_1\rangle+\mathcal{L}_\mathcal{I}\langle A_{m+1}\rangle\langle B_1-B_0\rangle\leq 2\mathcal{L}_\mathcal{I}\,.$$

For any deterministic strategy, the values of the 1-body correlators are extremal, *i.e.*, $\langle B_y\rangle,\langle C_z\rangle\in\{+1,-1\}$. As such, either $\langle B_0+B_1\rangle=0$ and $\langle B_0-B_1\rangle=\pm2$, or vice versa. Assuming the first case, then the first term is zero, and it is direct to see that the bound is satisfied. It is also easy to check that this is true for the other case.

[1] We use "probability distribution", "strategy" and "behaviour" interchangeably.

Assume now a nonlocal extremal strategy for B and C. The correlators in the second term do not involve C, and as such, it factorizes as follows:

$$\mathcal{L}_{\mathcal{I}}\langle A_{m+1}(B_1 - B_0)\rangle = \mathcal{L}_{\mathcal{I}}\langle A_{m+1}\rangle(\langle B_1\rangle - \langle B_0\rangle)\,.$$

It has been shown [JM05, Table. II] that for any 2-output non-local extremal distribution with 2 inputs for one device and any number of inputs for the other device in a 2-device setting (e.g., a Bell test), the marginals for the device with 2 inputs are all equal to $\frac{1}{2}$. This means that $\langle B_0\rangle = \langle B_1\rangle = 0$, which implies that the second term is zero.

Regarding the first term, expand $\mathcal{I}$ as a linear combination of correlators:

$$\begin{aligned}\langle \mathcal{I}\left[A_0,\ldots,A_m,C_0,\ldots,C_k\right](B_0+B_1)\rangle = \\ = \sum_{i=0}^{m}\sum_{j=0}^{k} M_{ij}\langle A_i\rangle\langle (B_0+B_1)C_j\rangle + \\ + \sum_{i=0}^{m}\nu_i\langle A_i\rangle(\langle B_0\rangle + \langle B_1\rangle) + \sum_{j=0}^{k}\mu_j\langle (B_0+B_1)C_j\rangle \quad \text{(S16)}\end{aligned}$$

Notice that since $\langle B_0\rangle + \langle B_1\rangle = 0$, any contribution from the $\langle A_i\rangle$ terms in the inequality vanishes, which might affect the bound of the inequality. From henceforth, assume that $\mathcal{I}$ contains no 1-body correlator terms for A (i.e., $\nu_i = 0$). Then we can absorb B_0 and B_1 into the C's in the following sense:

$$\begin{aligned}\langle \mathcal{I}\left[A_0,\ldots,A_m,C_0,\ldots,C_k\right](B_0+B_1)\rangle = \langle \mathcal{I}\left[A_0,\ldots,A_m,B_0C_0,\ldots,B_0C_k\right]\rangle + \\ \langle \mathcal{I}\left[A_0,\ldots,A_m,B_1C_0,\ldots,B_1C_k\right]\rangle\,.\end{aligned}$$

Now, from $\langle A_xB_yC_z\rangle = \langle A_x\rangle\langle B_yC_z\rangle$ and since $-1 \le \langle A_x\rangle \le 1$ and $-1 \le \langle B_yC_z\rangle \le 1$ one has :

$$\begin{aligned}\langle \mathcal{I}\left[A_0,\ldots,A_m,B_yC_0,\ldots,B_yC_k\right]\rangle \le \\ \le \max_{|\langle A_x\rangle|,|\langle B_yC_z\rangle|\le 1} \mathcal{I}\left[\langle A_0\rangle,\ldots,\langle A_m\rangle,\langle B_yC_0\rangle,\ldots,\langle B_yC_k\rangle\right] = \\ = \max_{|\langle A_x\rangle|,|\langle C_z\rangle|\le 1} \mathcal{I}\left[\langle A_0\rangle,\ldots,\langle A_m\rangle,\langle C_0\rangle,\ldots,\langle C_k\rangle\right] = \mathcal{L}_{\mathcal{I}}. \quad \text{(S17)}\end{aligned}$$

which implies the bound of $2\mathcal{L}_{\mathcal{I}}$.

This concludes the proof of equation (5.10). The other lifting *ansatz* concerns the 4-partite symmetric broadcast scenario:

$$\langle \mathcal{I}\left[A_0,\ldots,A_m,B_0,\ldots,B_k\right]\left(D_0+D_1\right)C_0\rangle + \mathcal{L}_{\mathcal{I}}\langle C_1\left(D_1-D_0\right)\rangle \leq 2\mathcal{L}_{\mathcal{I}}\,.$$

If all devices have a local deterministic strategy, then the expression simplifies to:

$$\langle \mathcal{I}\left[A_0,\ldots,A_m,B_0,\ldots,B_k\right]\rangle\langle C_0\rangle\langle D_0+D_1\rangle + \mathcal{L}_{\mathcal{I}}\langle C_1\rangle\langle D_1-D_0\rangle\,.$$

Since either $\langle D_0+D_1\rangle = \pm 2$ and $\langle D_1-D_0\rangle = 0$ or vice-versa, the $2\mathcal{L}_{\mathcal{I}}$ bound follows using the reasoning from the previous proof in this section.

If A and D share a NS resource and B and C do a local deterministic strategy, then the second term in the inequality is zero because $\langle D_0\rangle = \langle D_1\rangle = 0$. The first term simplifies to:

$$\langle \mathcal{I}\left[A_0(D_0+D_1),\ldots,A_m(D_0+D_1),B_0,\ldots,B_k\right]\rangle\langle C_0\rangle\,.$$

Because of the reasoning from the previous proof, this is upper-bounded by $2\mathcal{L}_{\mathcal{I}}$. Notice that here we need to assume, as in the previous proof, that the Bell expression $\mathcal{I}\left[A_0,\ldots,A_m,B_0,\ldots,B_k\right]$ has no 1-body correlator terms for C.

If A and D have a local deterministic strategy and B and C share a no-signaling resource, notice that $\langle C_1\rangle = 0$, therefore, the second term vanishes. The first one becomes:

$$\langle \mathcal{I}\left[A_0,\ldots,A_m,B_0C_0,\ldots,B_kC_0\right]\rangle\langle D_0+D_1\rangle\,.$$

Now for all values of $\langle D_0+D_1\rangle \in \{0,\pm 2\}$ the $2\mathcal{L}_{\mathcal{I}}$ bound is satisfied. Here we need to assume that $\mathcal{I}\left[A_0,\ldots,A_m,B_0,\ldots,B_k\right]$ has no 1-body correlator terms for A.

Lastly, we consider the case where A and D share a NS resource and B and C also share a NS resource. Notice that in this case, the 4-body correlator still factorizes between the two pairs because of the definition of broadcast nonlocality, $\langle A_iB_jC_kD_l\rangle = \langle A_iD_l\rangle\langle B_jC_k\rangle$. The second term of the inequality vanishes and the first one, because of the factorization, can be written as:

$$\langle \mathcal{I}\left[A_0(D_0+D_1),\ldots,A_m(D_0+D_1),B_0C_0,\ldots,B_kC_0\right]\rangle\,.$$

It is also clear from the arguments in the previous proof that this is upper bounded by $2\mathcal{L}_{\mathcal{I}}$.