# NONCLASSICALITY IN NOISY QUANTUM NETWORKS

BY

BRIAN DOPKINS DOOLITTLE

DISSERTATION

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Physics
in the Graduate College of the
University of Illinois Urbana-Champaign, 2023

Urbana, Illinois

Doctoral Committee:

Associate Professor Bryan K. Clark, Chair
Associate Professor Eric A. Chitambar, Director of Research
Professor Paul G. Kwiat
Professor Karin A. Dahmen

# Abstract

Quantum networks are rapidly being developed using the noisy quantum devices available today. As quantum networks scale, noise will lead to significant challenges in quantum network characterization, design, and automation, challenges that classical methods may be ill equipped to tackle. Moreover, the advantage and value of quantum networks is not well understood in the presence of noise, making it difficult to justify the cost of quantum network development for real-world applications.

In this dissertation, we define operational nonclassicality as a quantifier of quantum advantage in general multipoint communication networks and describe a procedure for deriving operational tests of nonclassicality in general communication networks. Then, we develop a quantum-hardware-compatible variational optimization framework for optimizing quantum networks to exhibit nonclassicality. In a wide range of communication network topologies, including nonsignaling networks, multiaccess networks, broadcast networks, and interference networks, we derive operational tests that witness nonclassicality. We then use our variational framework to optimize various quantum resource configurations for maximal performance against these operational tests of nonclassicality. In all communication network topologies, we find examples where quantum resources lead to observable violations of the nonclassicality witnesses, implying that quantum resources provide a strict advantage over classical resources. Furthermore, we investigate how the presence of noise diminishes these advantages, and by extension, the value of quantum resources. Finally, we demonstrate that our variational optimization techniques can be deployed on quantum hardware and applied well beyond the scope of finding nonclassical quantum behaviors.

In conclusion, we find that nearly all quantum resource configurations in communication networks can provide operational advantage as witnessed by operational tests of nonclassicality. These nonclassical network behaviors show novel ways that quantum physics defies the classical assumptions of locality, causality, and realism, but nonclassicality can also be used to test and certify quantum resources in communication networks. Furthermore, nonclassicality can also provide advantages in information security and distributed computing. We assert that variational quantum optimization techniques are well-suited to design and automation tasks in quantum networks. The advantages of these methods are that they are hardware agnostic, do not require full network characterization, and can optimize quantum systems against their inherent and unknown noise models. Thus, we introduce variational quantum networking as an engineering paradigm for designing and automating noisy quantum networks. In many ways, variational quantum networking circumvents the challenges of characterizing, designing, and automating noisy quantum networks.

*Dedication:*

# *Leonard A. Dopkins*

*In loving memory of my grandfather, who always supported my academic pursuits.*

*(03/25/1929 — 03/24/2023)*

# Acknowledgments

I thank my academic advisor, Eric Chitambar, for guidance and support over the past five years. I thank each of my colleagues in my research group, Haneul Kim, Ian George, Louis Schatzki, Xinan Chen, Sarah Hagen, and Amanda Gatto for all of their useful discussions and collaborations. I thank my Illinois faculty coauthors and collaborators Marius Junge and Felix Leditzky for their research support and contributions. I thank my industry collaborators from Xanadu, Tom Bromley and Nathan Killoran, for their mentorship and support in research. I thank my experimental collaborator, Colin Lualdi, for helping to integrate my theoretical work into the Kwiat lab. I thank my collaborators at Argonne National Lab, Jeff Larson, Zain Saleem, and Danny Chen for opening up interesting new research avenues. I thank Thaddeus Ladd and Tyler Keating for their support and mentorship during my summer internship at HRL Labs. I thank my former colleagues at athenaHealth for helping to shape my software engineer skills, which have proven to be invaluable. I thank all of my friends in the Champaign-Urbana community. I thank my grandparents, parents, siblings, spouse, and pets for their unconditional support. I acknowledge the University of Illinois physics department for their academic support. I acknowledge the support of the QISE-Net fellowship provided by the National Science Foundation under Award DMR-1747426. I acknowledge the support of NSF Award # 2016136. I acknowledge the use of IBM Quantum services for this work.

# Table of contents

# List of Abbreviations

| | |
|---|---|
| NISQ | Noisy Intermediate-Scale Quantum. |
| SDP | Semi-Definite Programming. |
| POVM | Positive Operator-Valued Measure. |
| PVM | Projection-Valued Measure. |
| CPTP | Completely-Positive Trace-Preserving. |
| DAG | Directed Acyclic Graph |
| LOCC | Local Operations and Classical Communication |
| VQO | Variational Quantum Optimization. |
| CHSH | Clauser, Horne, Shimony, and Holt. |
| RAC | Random Access Code. |
| GPU | Graphics Processing Unit. |
| MAN | Multiple Access Network. |
| BCN | Broadcast Network. |
| IN | Interference Network. |
| CIN | Compressed Interference Network. |
| BFN | Butterfly Network. |
| DIQKD | Device-Independent Quantum Key Distribution. |
| DIQRNG | Device-Independent Quantum Random Number Generation. |

# Part I

# Introduction

# Chapter 1

# Introduction

## 1.1 Motivation

A new era dawns in science, technology, and society—an era in which we harness the power of quantum physics to revolutionize the ways we sense, process, secure, and communicate information. Such optimism is shared globally as nations and venture capitalists alike accelerate their investments into quantum information science and technology [1]–[8]. In this vision, quantum information technologies exploit the laws of quantum physics to ascertain an advantage over existing, *classical* information technologies that we rely on constantly for our computing and communications needs. Moreover, quantum advantages are expected to broadly impact society through scientific advancements, economic growth, and national security enhancements [9]. Furthermore, the profound impact and benefit to humanity that quantum physics might offer is indicated by the recent 2022 Nobel Prize in Physics being awarded to Aspect, Clauser, and Zeilinger for their "experiments with entangled photons, establishing the violations of Bell inequalities, and pioneering quantum information science" [10]. Indeed, science, technology, and society are in the midst of a quantum revolution.

From a physical standpoint, it is interesting to question why quantum physics seems to offer an information processing advantage and what this advantage implies about the natural world. In principle, if quantum information technologies have an advantage, then there should be some quantum physical mechanism that leads to this advantage. According to quantum theory, the mechanisms of advantage are *superposition*, *entanglement*, and *measurement contextuality*, which have been observed to lead to information processing advantages. These strange physical mechanisms were initially met with skepticism [11] because they implied that nature was not *locally realistic*. That is when two entangled quantum particles are space-like separated and measured in a particular way, then their measurement statistics can be correlated in such a way that cannot be reproduced by a similar classical system [12], [13]. However such *nonlocal* correlations were later demonstrated to exist in nature [14]–[17]. Overall, it appears that the information processing advantages provided by quantum system are closely related to quantum systems' ability to exhibit stronger-than-classical correlations. How this advantage gets applied is then a matter of innovation.

From an engineering standpoint, quantum information technologies apply quantum physics to solve problems in computing, sensing, and communications. Quantum computers are theorized to provide significant advantages in simulation tasks [18]–[22], which may lead to improved algorithms for designing drugs [23] or materials [24], as well as the ability to simulate new systems in particle and high-energy physics. Similarly,

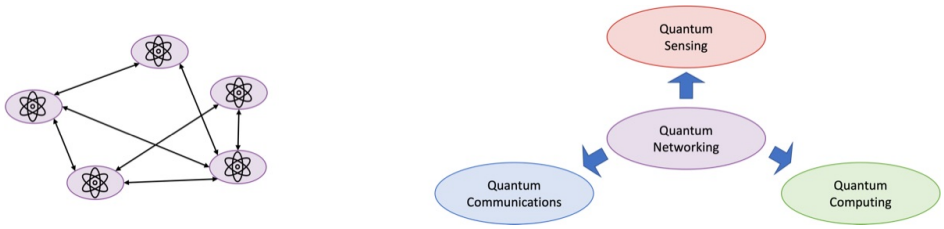



Figure 1.1: (Left) A quantum network. The purple ovals are quantum devices and the arrows represent quantum communication between the devices. (Right) Quantum networks are a foundational technology that enable quantum sensors, computers, and communications to scale.

certain quantum algorithms have been identified that provide exponential improvements over the best existing algorithms [25]. Examples include searching over big data sets [26], or factoring large numbers [27]. In sensing applications, quantum systems can be used to sense their environment with greater precision than existing techniques [28]. Some quantum sensing techniques can already be applied to practical applications such as improving sensitivity gravitational wave detectors [29]. In the future, quantum sensing techniques are expected to improve telescopes [30], [31], clock synchronization [32], and biomedical imaging applications [33], [34]. Finally, quantum communications technologies can help improve the efficiency of data transmission, improve information security, or be used transmit quantum systems [35]–[37].

To realize a practical quantum advantage, these quantum technologies will need to be scaled such that they can handle problems that are more complex. To effectively scale quantum technologies, it will be important to develop quantum communication networks that link quantum devices together using quantum communication. In essence, the quantum information technology can then be distributed amongst many devices. This network approach to scaling is commonly applied in examples of classical information technologies. Indeed, many of today's cutting edge applications in information technology rely on networks to operate. For instance, self-driving cars rely upon a network of sensors to navigate their environments, exascale super computers rely on networks of computers across which to distribute their processing, and smartphones rely upon wireless communication networks to connect users to applications on the world-wide web. In each of these examples, the network serves as the means by which technologies are scaled to handle more complex problems. In a similar manner, modular quantum devices can be integrated into a larger system called a quantum network. Such quantum communication networks should then be thought of as a foundational and enabling technology. By applying quantum networks, quantum information can be communicated between connected devices enabling large quantum sensing arrays or large quantum computers that are distributed over many devices.

Thus, we argue that quantum networking is essential to scaling quantum information technologies. As quantum technologies scale, so will their technological advantage and value. Therefore, the quantum revolution in science, technology, and society will be achieved through the development of quantum networks. Therefore, we focus our investigation towards quantum networks and the advantage that they can provide.

## *1.2 Problem*

Setting our optimism aside, a more accurate description of our current stage of technological development is, perhaps, the Noisy Intermediate-Scale Quantum (NISQ) era [38]—an era in which *noisy* quantum technologies are being developed but their advantage is fleeting. Although, a wide range of quantum computing architectures are under development [39], their advantage has yet to be demonstrated. For instance, the recent "quantum

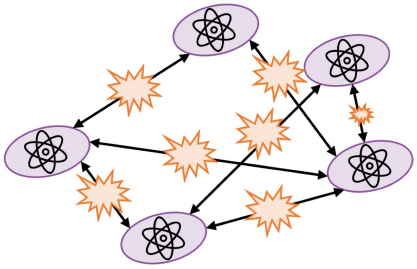


Figure 1.2: A quantum network with noisy communication links between devices.

supremacy" result that Google touted in 2019 [40] was later shown to be reproducible on a laptop computer [41]. Likewise, a wealth of quantum communication networks have been constructed for the purpose of sharing secret keys [42]–[49], yet experts from both academia [50] and the National Security Agency [51] are skeptical of any information security improvements offered by these quantum secret sharing protocols. Despite this skepticism, these early-stage quantum technologies will be iterated upon and scaled as quantum networks [52]. However, as quantum networks scale, noise will remain the key challenge to overcome to realize a quantum advantage [52]–[56].

Noise is a natural process occurring when quantum systems interact with their environments. Through this interaction, the information encoded into a quantum system leaks into the environment and is lost. In existing quantum information technologies, noise limits the ability to store, process, and transmit quantum information [52], [55], [56]. The type of noise that exists in a quantum system depends on its hardware architecture, meaning that different hardware platforms each have their own strengths and weaknesses. Although, error correction and entanglement purification protocols can be used to remove noise in quantum communication networks [57]–[62], near-term NISQ devices are not expected to have enough qubits or entangled states available to be able to perform with fault-tolerance [38], [52].

When quantum networks are constructed out of NISQ devices, their state preparation, evolution, and measurement are all noisy. Therefore, noise must be considered as an inherent property of quantum hardware that leads to technological limitations. As a result of noise, several restrictions on quantum networks follow:

- **Memoryless:** Quantum states cannot be stored for extended periods of time.
- **Low-Depth:** Extensive processing is unavailable.
- **Intermediate-Scale:** The number of qubits held by each device is limited.
- **Inefficient Detectors:** The shot rate may be limited and detectors may have inherent noise.

Overall, noisy quantum networks will operate in a *prepare-and-measure* setting in which quantum states are prepared, communicated, and measured without having extensive processing, long storage times, or high shot counts [52].

Clearly the vision of the quantum revolution is at odds with the reality of the noisy quantum information technologies that can be built. While the vision may have piqued the interest of investors, funding will quickly dissipate if practical quantum advantages are not demonstrated. Nonetheless, noisy quantum networks are being developed, making it a matter of paramount importance quantify their advantages in the noisy setting. To this end, we seek a measurable, operational, and indisputable quantifier of quantum advantage. In doing so, we may assign a practical value to quantum information technologes, justifying their cost and continued development towards the bettering of society through science and technology.

## 1.3 Thesis

### 1.3.1 Main Question

**How does the *operational advantage* of quantum networks deteriorate in the presence of noise?**

### 1.3.2 Approach

To investigate our main question, we must quantify the operational advantage of a quantum network. One way to quantify operational advantage is through *quantum nonclassicality*, a phenomenon in which the measurement data of a quantum system cannot be reproduced by a similar classical system. Such nonclassical correlations are directly related to advantages in quantum information processing systems [63]–[68]. Nonclassicality is a practical quantifier of advantage in noisy quantum networks because it is *device-independent*, meaning that it does not depend upon the physical devices, but only their measured data, the network's causal structure, and the network's communication resources.

To investigate how the operational value of quantum networks deteriorate in the presence of noise, we study the noise robustness of nonclassicality. That is, how much noise can a network tolerate before it can no longer demonstrate nonclassicality. If a network cannot demonstrate nonclassicality, then it provides no information processing advantage over a classical network. To this end, we apply simulation and optimization methods to evaluate the noise robustness of nonclassicality.

### 1.3.3 Technical Challenges

In order to set up applications on a quantum network, the noise in the network's devices and links must be characterized. Furthermore, if noise is dynamic and changes over time, then the characterization process needs to be repeated to maintain robust performance of quantum applications over extended periods of time. This characterization task can be performed using quantum process tomography to learn the noise model [69]–[74]. Quantum process tomography requires many identical copies of the same quantum system, and the number of needed measurements typically grows exponentially with the number of qubits. Regardless, full tomography can be performed in many existing quantum systems because they have only a few qubits, however, current methods fail to scale efficiently to systems having many qubits. Thus, how can we develop robust quantum networking applications if we can't efficiently characterize its noisy or unknown processes in networks?

Suppose now that the network and its noise is characterized and and can be modeled accurately, a second challenge is then to design a networking protocol or algorithm that is robust to the inherent noise. This task will undoubtedly rely upon simulation and optimization software tools, however, quantum simulation and

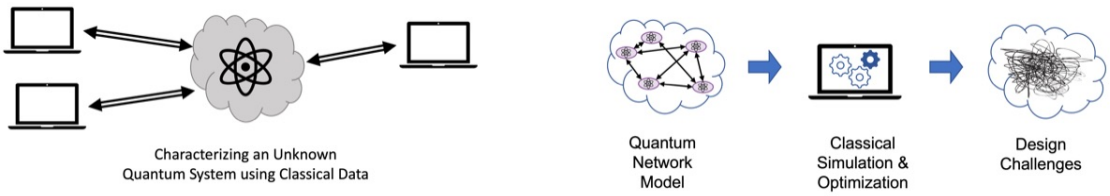



Figure 1.3: Quantum networking characterization and design challenges. (Left) Classical computers try to characterize an unknown quantum system. (Right) Classical computers can

optimization on classical hardware is generally very difficult [75]. A promising approach is to use a tensor network representation for the complex quantum system, which can help reduce complex networks into smaller systems whose dynamics can be simulated and optimized more efficiently [76]. In many cases, tensor networks can lead to efficient classical simulation of quantum systems [77]–[85]. Furthermore, the tensor networks can be optimized via differential programming [86]–[88]. However, these efficiencies typically vanish for the complex, many-bodied entanglement structures, which may be common in quantum networks. Similarly, an efficient classical algorithm is known for simulating certain entangled quantum systems called stabilizer circuits [89], [90]. However, stabilizer circuits cannot perform universal computation nor can they account for noise in a precise manner. Thus, efficient simulation of stabilizer circuits may not necessarily be helpful for simulating noisy quantum systems in the real-world. Moreover, another common tool for optimizing quantum systems is semi-definite programming (SDP)[91]–[94]. Despite the fact that these SDP methods compute in polynomial time, they fail to scale to large quantum systems such as networks because the dimension of quantum systems scales exponentially. Furthermore, SDP methods only solve convex optimization problems, limiting their general applicability. Thus, how can we design future quantum networks and applications if our classical simulation and optimization tools fail to scale?

## *1.4 Summary of Methods*

We thus study the noise robustness of nonclassicality in quantum networks. We take a two-pronged approach in which we develop a theoretical and operational framework for quantifying and certifying nonclassicality in networks. Alongside this framework we also develop hybrid quantum-classical computational methods for performing the requisite derivations, simulations, and optimizations. Our results are both reproducible on a laptop computer, and extensible to real-world quantum network hardware. Furthermore, our hybrid approaches show promise in circumventing technical challenges in quantum networking. Hence, we provide a practical and comprehensive approach to establishing nonclassical correlations for operational tasks.

### *1.4.1 A Black-Box Approach to Communication Networks*

In Chapter 2, we introduce a black-box model for characterizing physical systems in the discrete memoryless setting. Here a black-box can be thought of as a device that maps an input $x$ to an output $y$ with probability $P(y|x)$. The black-boxes behavior is then characterized by its set of conditional probabilities $\mathbf{P} = \sum_{x\in\mathcal{X}} \sum_{y\in\mathcal{Y}} P(y|x) \, |y\rangle\langle x|$, which can be arranged as a column stochastic matrix. We also introduce a formulation of black-box game $\mathbf{G} \in \mathbb{R}^{|\mathcal{Y}|\times|\mathcal{X}|}$ where a reward $G_{y,x} \geq 0$ is given to the black-box each time it outputs $y$ given input $x$. After many independent shots the average score is $\langle \mathbf{G}, \mathbf{P} \rangle = \sum_{x,y} G_{y,x} P(y|x)$. Such linear black-box games can serve as device-independent tests that quantify a systems ability to perform the task encoded by the game.

In Chapter 3, we extend our black-box model to communication where network devices are black-boxes and the network's links are classical communication channels. We show that such networks can be expressed as a directed acyclic graph DAG depicting the networks causal flow of information from input to output. Assuming that each black-box operates in the memoryless setting, then each shot of the network can be considered independent from the other. Furthermore, we introduce the resource of classical communication and how it can be quantified by the signaling dimension of each communication channel. We also introduce shared randomness, as a correlation between black-boxes that can be distributed prior to when shots are taken.

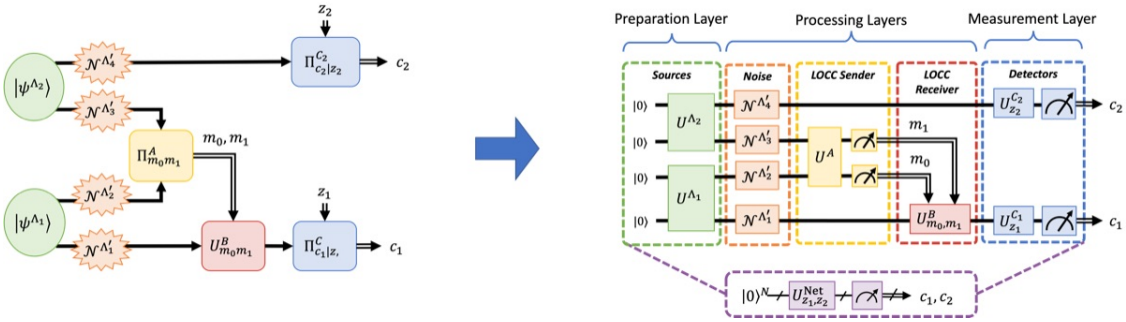



Figure 1.4: (Right) A quantum network's directed acyclic graph (DAG) and (Left) a quantum circuit that simulates the network.

In Chapter 4, we provide a general characterization of set of behaviors $\mathcal{C}^{\text{Net}}$ that classical black-box networks can exhibit. In general $\mathcal{C}^{\text{Net}}$ is nonconvex, however, when all network devices share an unlimited amount of shared randomness the set of classical network behaviors becomes convex as $\mathcal{C}^{\text{Net}} \subseteq \text{Conv}\left(\mathcal{C}^{\text{Net}}\right) = \mathcal{C}^{\text{Net}}_{\Lambda}$ where the $\Lambda$ subscript signifies the global shared randomness. The set of $\mathcal{C}^{\text{Net}}_{\Lambda}$ forms a convex polytope, meaning that it is bound tightly by a finite set of linear inequalities $\mathcal{G}^{\text{Net}} = \{\gamma_k \geq \langle \mathbf{G}_k, \mathbf{P} \rangle\}_{k=1}^{|\mathcal{G}|}$, which are each satisfied for any behavior $\mathbf{P} \in \mathcal{C}^{\text{Net}}_{\Lambda}$. These linear inequalities are black-box games $\mathbf{G}$ where $\gamma$ is an upper bound on the average score that can be achieved by the considered network. These games can then serve as operational tests where a violation to one of these inequalities, $\langle \mathbf{G}, \mathbf{P} \rangle > \gamma$, implies that $\mathbf{P} \notin \mathcal{C}^{\text{Net}}_{\Lambda}$. We then discuss various approaches for deriving linear inequalities that tightly bound a classical network's behaviors.

### *1.4.2 Noisy Quantum Communication Networks*

In Chapter 5 we review the fundamentals of quantum physics, taking an axiomatic approach. Here quantum states are first prepared on a complex-valued Hilbert space $|\psi\rangle \in \mathcal{H}_d$, then quantum state evolve under unitary evolution $|\psi'\rangle = U|\psi\rangle$, and finally the state is measured in an orthonormal basis $\{|y\rangle\}_{y \in \mathcal{Y}}$ where the output $y$ is obtained with probability $P(y) = |\langle y|\psi\rangle|^2$. Furthermore, we discuss how noisy quantum systems are modeled by considering unitary operators that couple the system with its environment as $U(|\psi\rangle \otimes |\phi\rangle^{\text{Env}})$.

In Chapter 6 we discuss how quantum systems can be applied to information processing tasks in sensing, computing, and communications. We argue that quantum networks are a foundational technology that enables quantum applications to scale. Moreover, we discuss how advantages in quantum systems are often quantified in terms of computing or communication resources.

In Chapter 7, we combine the quantum formalism with the classical model of networks. That is, we extend a network DAG with quantum devices that prepare, process, or measure quantum states. Furthermore, these devices are connected by quantum communication. We note that that the amount of quantum communication can also be quantified by the signaling dimension allowing the resources of classical communication to be exchanged for quantum communication in a one-to-one manner. In general the set of quantum network behaviors $\mathcal{Q}^{\text{Net}}$ contains the set of classical network behaviors $\mathcal{C}^{\text{Net}} \subseteq \mathcal{Q}^{\text{Net}}$ where each network is described by the same DAG and the signaling dimension of each communication channel is the same. We then introduce entanglement-assisted communication networks in which quantum entanglement is shared amongst devices in the network. Finally, we discuss how the quantum network model can be encoded into a quantum circuit that simulates the network.

In Chapter 8, we describe how noise can be modeled in quantum communication networks. We apply the

quantum channel formalism in which a completely-positive trace-preserving (CPTP) map models a noisy quantum process as $\widetilde{\rho} = \mathcal{N}(\rho)$ where $\rho$ is a density operator representation of a quantum state. We also describe common noise models that are used throughout the work. Moreover, we consider classical noise models in which stochastic operations are applied as postprocessing to a network's behavior as $\widetilde{\mathbf{P}} = \mathbf{LP}$.

### *1.4.3 An Operational Framework for Certifying Nonclassicality*

In Chapter 9 we develop a device-independent framework for understanding nonclassicality in noisy quantum networks with limited communication resources. In other words, we describe a quantum resource theory of nonclassicality, in which the amount of communication is fixed while network devices are able to apply a generic set of free operations [95]. We characterize both classical and quantum networks by their respective sets of behavior $\mathcal{C}^{\text{Net}}$ and $\mathcal{Q}^{\text{Net}}$ where any quantum behavior $\mathbf{P} \in \mathcal{Q}^{\text{Net}}$ is nonclassical if and only if $\mathbf{P} \notin \mathcal{C}^{\text{Net}}$.

We define two nonclassicality, weak and strong. Weak nonclassicality describes quantum behaviors $\mathbf{P} \in \mathcal{Q}^{\text{Net}}$ that can be simulated when globally across all classical network devices implying that $\mathbf{P} \notin \mathcal{C}^{\text{Net}}$, but $\mathbf{P} \in \mathcal{C}_{\Lambda}^{\text{Net}}$. On the contrary, a quantum behavior $\mathbf{P} \in \mathcal{Q}^{\text{Net}}$ is strongly nonclassical if $\mathbf{P} \notin \mathcal{C}_{\Lambda}^{\text{Net}}$ such that the quantum behavior cannot be simulated when randomness is shared globally to all classical network devices. Indeed, strong nonclassicality *requires* extra classical communication to simulate the quantum correlations, implying a strict resource advantage over classical systems.

Strongly nonclassical behaviors demonstrate a clear operational advantage because such behaviors demonstrate a specific task that cannot be reproduced using classical resources [63]–[68], [96], [97]. Therefore, witnessing nonclassicality can be used verify the presence of quantum resources. The resource advantage can either be quantified as the amount of classical communication needed to reproduce the nonclassical behavior, or by the distance between the behavior and the classical set [98].

The nonclassical correlations can be certified in an operational manner. First, a linear inequality $\gamma \geq \langle \mathbf{G}, \mathbf{P}^C \rangle$ is derived that is satisfied by all classical network behaviors $\mathbf{P}^C \in \mathcal{C}_{\Lambda}^{\text{Net}}$. For a given behavior $\mathbf{P} \in \mathcal{Q}^{\text{Net}}$, the score $\langle \mathbf{G}, \mathbf{P} \rangle > \gamma$ implies that the behavior $\mathbf{P} \notin \mathcal{C}_{\Lambda}^{\text{Net}}$ and is, therefore, nonclassical. Moreover, to exhibit nonclassicality quantum systems must demonstrate quantum properties that have no classical equivalent, such as entanglement or incompatible measurements. Thus, nonclassicality can be used to test the quality of quantum resources.

In Chapter 10, we introduce the noise robustness of nonclassicality as the amount of noise that can be tolerated by the quantum network before it is unable to demonstrate nonclassical behaviors. That is, noise in the quantum networks causes their nonclassicality to deteriorate. As the nonclassicality decreases the the amount of violation $\langle \mathbf{G}, \mathbf{P} \rangle > \gamma$ approaches the classical bound. Therefore, the amount of nonclassicality can be used to measure the quality of the quantum resources as the distance to the set of classical network behaviors.

A challenge that emerges within our framework is finding the maximally nonclassical behavior $\max_{\mathbf{P} \in \mathcal{Q}^{\text{Net}}} \langle \mathbf{G}, \mathbf{P} \rangle$ for a given set of quantum operations and communication resources. Mainly, the state preparations, processing, and measurements must all be optimized causing the optimization to lose its convexity. Therefore, convex optimization techniques such as semi-definite programming [93] fail to derive obtain the maximal violations for a given set of resources. Alternatively, differential programming and machine learning techniques have shown success in obtain maximally nonclassical behaviors [99]–[102]. Throughout this work, we will apply the variational quantum optimization methods for noisy quantum networks developed in our work in reference [102].

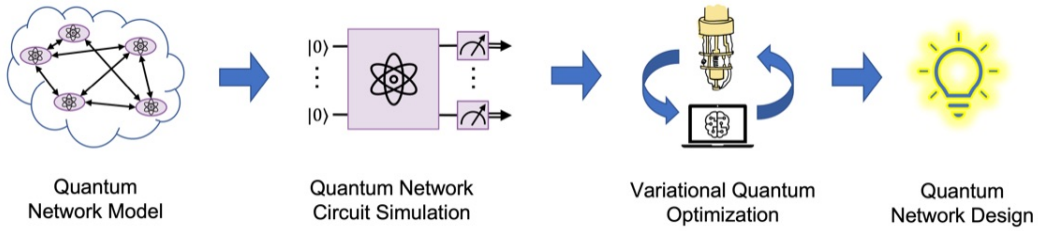



Figure 1.5: A hybrid framework for quantum network design. From left to right, a quantum network is modeled theoretically, the quantum network model is simulated on a quantum computer, hybrid optimization tunes the simulation, the quantum computer advances scientific discovery and engineering design.

### *1.4.4 A Framework for Variational Quantum Optimization of Noisy Quantum Networks*

To address the challenge of simulating and optimizing noisy quantum networks, we apply hybrid quantum-classical computing techniques. Recently, a significant research effort has focused on hybrid quantum-classical computing applications due to their promise of demonstrating practical advantages on NISQ hardware [38]. In these algorithms, classical and quantum computers operate in tandem, each providing its own strengths to the algorithm. In this framework, the classical computer can be viewed as a client that instructs a server-side quantum computer. Generally, the quantum computer is used to efficiently simulate a quantum system or run an algorithm that has a quantum advantage. Thus, hybrid computing algorithms use quantum computers to perform tasks that are classically difficult, but quantumly easy solve.

Variational quantum optimization (VQO) algorithms designate a family of hybrid quantum-classical algorithms that leverage the power of quantum computers to solve optimization problems [103]. The basic goal of VQO is to optimize the settings $\Theta$ of a parameterized quantum circuit $U^{\mathrm{Net}}(\Theta)$ for a particular task by iteratively improving upon a trial solution. The quantum circuit, also known as a variational ansatz, consists of a sequence of quantum operations, each of which is controlled by a parameter that can be adjusted during the optimization process. The task being optimized is then encoded into an objective function referred to as the *cost*, which is minimized during optimization. The goal is to find the settings that minimize the cost.

The optimization objective is to minimize the cost as $\min_{\Theta} \mathrm{Cost}\,(\Theta)$. This optimization problem can be solved using a standard gradient descent algorithm. Namely, the cost is minimized numerically by traversing the path of steepest descent where the path of steepest descent at point $\Theta$ is the negative of the gradient $-\nabla_{\Theta}\mathrm{Cost}\,(\Theta)$. This optimization process is then repeated many times until a satisfactory solution is found. In each iteration a small step is taken in the direction of steepest descent. Remarkably, the gradient can be evaluated on quantum hardware using the *parameter-shift rule* [104]–[107] or a similar finite-differences method, making VQO approaches able to optimize quantum systems themselves.

In Chapter 11 we develop a hybrid quantum-classical framework for quantum network simulation and optimization that we adapt from our work in Ref. [102], [108]. In this framework, a noisy quantum network is expressed as a variational ansatz circuit $U^{\mathrm{Net}}(\Theta)$ that encodes the constraints on the network's communication resources and free operations. The quantum circuit can evaluated on a quantum computer or classical simulator to obtain the network's behavior $\mathbf{P}(\Theta) \in \mathcal{Q}^{\mathrm{Net}}$, thereby simulating the noisy quantum network. Next we apply variational quantum optimization to the noisy network ansatz. Our main focus is nonclassicality, thus the optimization objective is $\max_{\Theta} \langle \mathbf{G}, \mathbf{P}(\Theta) \rangle$ where the inequality $\gamma \geq \langle \mathbf{G}, \mathbf{P}^{C} \rangle$ bounds classical network behavior $\mathbf{P}^{C} \in \mathcal{C}^{\mathrm{Net}}$. However, these variational quantum optimization techniques are extremely general and

can be applied to a range of network optimization problems.

## 1.5 Summary of Results

Our main goal is to investigate how nonclassicality deteriorates in the presence of noise to better understand the noise robustness of the quantum advantage. To this end, we consider a broad range of communication networks and study how incorporating quantum communication resources enables nonclassicality. Hence, for each network, we derive linear inequalities $\gamma \geq \langle \mathbf{G}, \mathbf{P} \rangle$ bounding the classical behaviors $\mathbf{P} \in \mathcal{C}_{\Lambda}^{\text{Net}}$. These linear inequalities serve as operational tests that can quantify the amount of nonclassicality by the amount of violation. Then, using our variational optimization framework, we find examples of violations.

### 1.5.1 Nonclassicality in Nonsignaling networks

In Chapter 12 we consider local nonsignaling networks in which the devices do not communicate, but share global randomness. When quantum entanglement replaces the shared randomness, the nonsignaling network is known to exhibit nonclassical phenomena, typically referred to as *Bell nonlocality* [109]. Within our framework, Bell nonlocality is an example of strong nonclassicality. The maximal violations of the inequalities bounding $\mathcal{C}_{\Lambda}^{\text{Net}}$ are known in many cases allowing us to verify that our VQO framework is able to reproduce known results. Furthermore, we derive the noise robustness of the violations to the CHSH inequality and and verify that the VQO software can obtain the maximal violations of classicality.

In Chapter 13 we consider the $n$-local setting in which $n$ independent shared randomness sources are connect nonsignaling devices into a network. These networks play an important role in entanglement distribution [52], [110]–[113]. In the $n$-local setting, nonlinear $n$-locality inequalities are known that bound the set of classical behaviors. Furthermore, when each source instead emits quantum entanglement, quantum violations of $n$-locality can be found. We find that the quantum violations of $n$-locality are weakly nonclassical because they could be simulated classically if all network devices share randomness. Nevertheless, we verify that our VQO software is able to reproduce known violations in the noiseless case. When noise is introduced we derive maximal $n$-locality violations and verify that the maximal nonclassical behaviors can be obtained using VQO, even in the presence of noise.

### 1.5.2 Nonclassicality in Multipoint Communication Scenarios

In Chapter 14, we consider a point-to-point signaling scenario in which a sender device communicates to a receiver device using a finite amount of communication. Applying our work in Ref. [114], we derive families

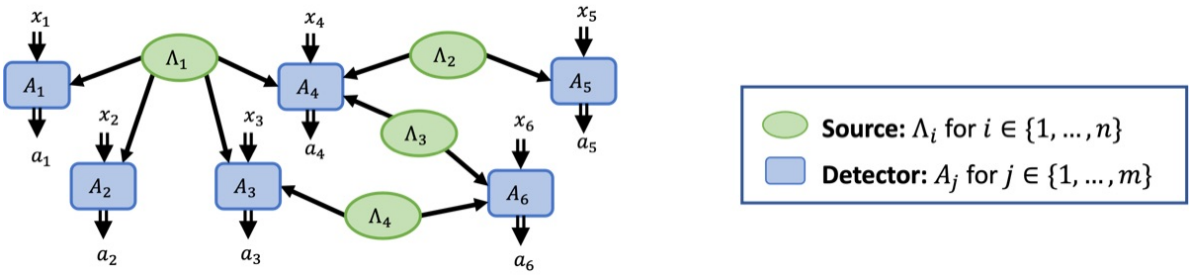



Figure 1.6: An example nonsignaling network DAG in the $n$-local setting.

of linear inequalities that bound the classical signaling polytope $\mathcal{C}_{\Lambda}^{\mathcal{X}\overset{d}{\to}\mathcal{Y}}$. Furthermore, in special cases, we completely characterize the bounds on classical set. In general it is known that quantum signaling does not exhibit strong nonclassicality such that $\mathcal{C}^{\mathcal{X}\overset{d}{\to}\mathcal{Y}} \subseteq \mathcal{Q}^{\mathcal{X}\overset{d}{\to}\mathcal{Y}} \subseteq \mathcal{C}_{\Lambda}^{\mathcal{X}\overset{d}{\to}\mathcal{Y}}$ [115]. Interestingly, we prove that there exist weakly nonclassical quantum signaling systems whose behaviors require an unbounded amount of classical communication to simulate. However, we observe that even small amounts of white noise appear to diminish this significant resource advantage. Furthermore, we observe that when entanglement assists classical communication of one bit that strongly nonclassical behaviors can be produced, one example of which having been recently derived in the literature [116].

In Chapter 15 we consider networks multiple access networks in which multiple senders signaling to a single receiver. We derive novel linear inequalities that bound the set of classical multiple access network behaviors $\mathcal{C}_{\Lambda}^{\mathrm{MAN}}$. Then applying our variational quantum optimization methods, we find that unassisted quantum communication and entanglement-assisted classical communication can produce strongly nonclassical behaviors. As a special case of the multiple access network we also consider the random access code task in which either quantum communication or entanglement assisted classical communication are able to yield an advantage.

In Chapter 16 we considder broadcast networks in which one sender signals to multiple receivers. We derive novel linear inequalities that bounds the set of classical broadcast network behaviors $\mathcal{C}_{\Lambda}^{\mathrm{BCN}}$. Applying our variational optimization methods, we do not find any examples where quantum communication is able to produce strongly nonclassical behaviors. However, we prove by example that quantum broadcasts can require an unbounded amount of classical communication to simulate, hence we argue that unassisted quantum broadcasts are weakly nonclassical such that $\mathcal{C}^{\mathrm{BCN}} \subseteq \mathcal{Q}^{\mathrm{BCN}} \subseteq \mathcal{C}_{\Lambda}^{\mathrm{BCN}}$. However, when we consider entanglement assisted classical or quantum communication resources, strong nonclassiicality can be observed.

In Chapter 17 we consider multipoint communication networks that have multiple senders and receivers. In particular, we consider interference and butterfly network topologies. Applying our framework of nonclassicality, we derive linear inequalities that witness nonclassicality in these networks. Furthermore, we identify examples in which quantum resources lead to strong nonclassicality.

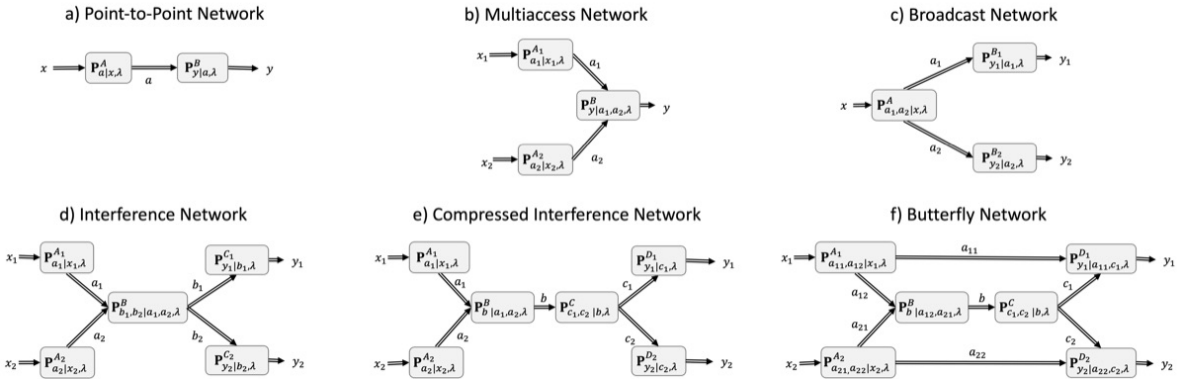



Figure 1.7: A collection DAGs representing the multipoint communication networks that are studied throughout this work.

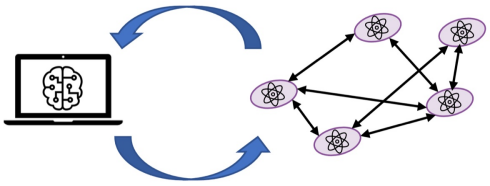


Figure 1.8: Quantum network hardware can also is controlled in a similar manner allowing variational quantum algorithm and other hybrid computing techniques to be applied directly on quantum networks.

### *1.5.3 A Framework for Hardware-Agnostic Quantum Networking*

In Chapter 18 we consider the advantage of applying our variational quantum optimization approaches on quantum networks directly. Since noise in quantum networks can be dynamic and change over time, we argue that the network devices will have to recalibrate their operations over time. We then introduce *variational quantum networking* as a hardware-agnostic approach for establishing and maintaining connections between quantum network devices. In particular we describe a generic protocol in which devices in the network share their tunable parameters while a classical optimizer applies VQO on the networking hardware to minimize the cost function. If successful, the set of optimal network settings $\Theta^\star$ is obtained, as well as a certificate $\text{Cost}\,(\Theta)$ quantifying the performance of the optimization. As examples, we provide examples of variational quantum networking protocols simulated on both classical and quantum hardware.

## *1.6 Outro*

Hence we embark on the journey, so to speak, in an attempt to quantify the quantum advantage in a general, yet practical manner. From a physical perspective, we seek to improve our understanding of the quantum advantage and what it implies about nature. From an engineering perspective, we seek to identify practical applications in which quantum advantages can be realized, even in the presence of noise. Moreover, we focus our theoretical investigation on quantum networks due to their foundational role in scaling quantum information technologies.

# Part II

# Methods

# Chapter 2

# Black-Box Models

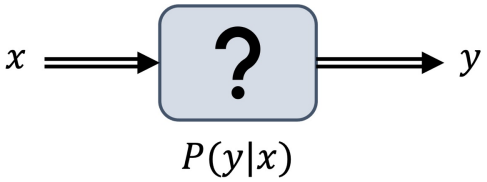



Figure 2.1: A black-box model of a unknown system or process. When value $x \in \mathcal{X}$ is input, the value $y \in \mathcal{Y}$ is output with probability $P(y|x)$.

**Definition 1. Black-Box:** An unknown physical system that accepts as input a classical value $x \in \mathcal{X}$ and as output produces a classical value $y \in \mathcal{Y}$ with some probability $1 \geq P(y|x) \geq 0$ where $\sum_{x \in \mathcal{X}} P(y|x) = 1$ for all $y \in \mathcal{Y}$, and $\mathcal{X} \equiv \{0, \ldots, |\mathcal{X}| - 1\}$ and $\mathcal{Y} \equiv \{0, \ldots, |\mathcal{Y}| - 1\}$ denote finite input and output alphabets respectively.

Black-boxes are extremely prevalent throughout our lives. The inner workings of nearly all electronic and mechanical devices that people use daily are hidden from the people using them. In most cases, a black-box can be opened up and someone with specialized training can determine the precise process mapping an input to an output. For example, a mechanic can open up the hood of a car to diagnose problem, or a programmer can investigate a software's source code to determine how the underlying algorithm works. However, many people do not have the skills to understand how a car or software works, yet this lack of knowledge does not detract from their ability to apply the system. Simply put, you don't need to understand the mechanics of a car to operate it, nor do you need to understand the electronics in a computer to browse the internet.

A layperson's use of black-box technologies can be compared with a physicist's approach to quantum system. That is, not all physical systems in nature are so easy to open up and investigate. For example, a physicist cannot simply peer inside a quantum system to observe what is going on because quantum mechanics cannot be observed directly. All quantum phenomena are observed indirectly by probing the system and measuring the result. That is, the choice of probe used by the experimenter corresponds to a classical input $x$ and the measurement result is a classical output $y$. However, despite our lack of understanding of quantum mechanics, we can still apply quantum systems in technology. Much of this technological success lies on the shoulder of the black-box models that can be used to infer the existence of quantum mechanics, or more

accurately, that nature does not always obey classical mechanics. However, in the spirit of black-boxes, we leave our discussion of quantum mechanics to Chapter 5. In this chapter, we formalize black-boxes whose underlying physical systems, quantum or classical, are hidden from the experimenter.

## 2.1 The Discrete Memoryless Setting

**Definition 2. Discrete Memoryless Setting:** An experimental setting consisting of many independent runs or shots of a black-box device where the prior $q(x)$ and conditional probabilities $P(y|x)$ are independent from the data collected in previous shots.

A black-box's output $y$ bears a conditional dependence on the input $x$, which reflects the causal nature of the black-box. That is, the input $x$ causes the output $y$ while the output $y$ has no affect on the input $x$, or its prior distribution $q(x)$. This type of causal structure is often referred to as the *discrete memoryless setting* because neither the black-box nor the experimenter perform actions based upon the information accumulated during the experiment. In this setting, each run or *shot* of the experiment is completely independent from any other shot. At a high-level, the discrete memoryless setting mirrors a one-shot setting where a black-box is given only a single input. The black-box is then evaluated by its average performance of mapping a uniformly random input $x \in \mathcal{X}$ to the "correct" output $y \in \mathcal{Y}$.

The one-shot setting is interesting from a theoretical and practical perspective because it does not allow for memory effects to take place [117]. However, it doesn't quite describe realistic systems in quantum experiments that often consist of many shots. Such experiments can be considered in the one-shot setting if they are also memoryless, meaning that the black-box and experimenter do not store data for the purpose of conditioning future actions. In practice, it is a key objective to ensure that each run of an experiment is independent from its previous runs. If this is not the case, then the experimental results will likely be very difficult to reproduce. Thus, the discrete memoryless setting serves as a good model of quantum experiments and technology.

The importance of imposing the memoryless constraint can be observed through a simple example. In principle, if a black-box stored its $N$ previous inputs $\vec{x}_0 = (x_i \in \mathcal{X})_{i=1}^N$, then its conditional probabilities $P(y|x, \vec{x}_0)$ could depend on both the current input $x$ and the history of black-box inputs $\vec{x}_0$. Likewise, if each input $x$ depended on the black-box's $N$ previous outputs $\vec{y}_0 \in (y_i \in \mathcal{Y})_{i=1}^N$, then the prior distribution could depend on the black-box's outputs as $q(x|\vec{y}_0)$. Thus, if the black-box and experimenter are not assumed to be memoryless, then the shots cannot be independent. Such a memory-ful scenario could correspond to an adversarial system where all parties can game the system using their local data. Despite being an interesting scenario, such memory-ful models can be quite complicated to understand and will largely be disregarded in this dissertation.

Hence we assert a foundational assumption in this dissertation: all considered black-box systems, either quantum or classical, are assumed to operate in the discrete memoryless settings. That is, all shots in an experiment are assumed to be independent and any deviation from this assumption will be acknowledged.

## 2.2 Black-Box Behaviors

In the discrete memoryless setting, a black-box's *behavior* is characterized by its conditional probabilities $P(y|x)$ [109]. These conditional probabilities are independent from the prior distribution $q(x)$ on its input, which is controlled by the experimenter. Thus, a uniform prior distribution is often assumed where $q(x) = 1/|\mathcal{X}|$,

**Definition 3. Black-Box Behavior:** $\mathbf{P} : \mathcal{X} \to \mathcal{Y}$, a column stochastic matrix that characterizes a black-box

$$\mathbf{P} = \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} P(y|x) \, |y\rangle\langle x| \, , \tag{2.1}$$

where $\{\langle x|\}_{x \in \mathcal{X}}$ and $\{|y\rangle\}_{y \in \mathcal{Y}}$ are orthonormal bases having $0/1$ elements and for $x \in \mathcal{X}$ and $y \in \mathcal{Y}$, the probability $P(y|x)$ is subject to the constraints of

$$\textbf{Non-Negativity:} \quad P(y|x) \geq 0, \quad \textbf{Normalization:} \quad \sum_{y \in \mathcal{Y}} P(y|x) = 1 \tag{2.2}$$

A black-box's behavior is a characteristic of that instance of a black-box, another black-box may output may exhibit a different behavior. In general, a black-box behavior is only constrained by the non-negativity and normalization of its underlying conditional probabilities. Thus, a black-box can produce any probability distribution.

**Definition 4. Full Probability Polytope:** $\mathcal{P}_{\mathcal{Y}|\mathcal{X}}$, A convex polytope that contains all column stochastic matrices where

$$\mathcal{P}_{\mathcal{Y}|\mathcal{X}} = \Big\{ \mathbf{P} \in \mathbb{R}^{|\mathcal{Y}| \times |\mathcal{X}|} \;\Big|\; P(y|x) \geq 0 \quad \text{and} \quad \sum_{y \in \mathcal{Y}} P(y|x) = 1 \Big\}, \tag{2.3}$$

which is equivalent to the set all black-box behavior maps $\mathbf{P} : \mathcal{X} \to \mathcal{Y}$.

**Lemma 1.** The set of black-box behaviors $\mathcal{P}_{\mathcal{Y}|\mathcal{X}}$ is convex.

*Proof.* The convex combination of any two behaviors $\mathbf{P}_1, \mathbf{P}_2 \in \mathcal{P}_{\mathcal{Y}|\mathcal{X}}$ with weights $\mu \in [0, 1]$ is

$$\mu \mathbf{P}_1 + (1 - \mu)\mathbf{P}_2 = \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} \left( \mu P_1(y|x) + (1 - \mu) P_2(y|x) \right) |y\rangle\langle x| = \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} P_3(y|x) \, |y\rangle\langle x| = \mathbf{P}_3. \tag{2.4}$$

The elements of $\mathbf{P}_3$ are non-negative because $P_3(y|x) = \mu P_1(y|x) + (1 - \mu) P_2(y|x) \geq 0$. Additionally, the columns of $\mathbf{P}_3$ are normalized because

$$\sum_{y \in \mathcal{Y}} P_3(y|x) = \mu \sum_{y \in \mathcal{Y}} P_1(y|x) + (1 - \mu) \sum_{y \in \mathcal{Y}} P_2(y|x) = \mu + (1 - \mu) = 1. \tag{2.5}$$

Thus, $\mathbf{P}_3 \in \mathcal{P}_{\mathcal{Y}|\mathcal{X}}$ and the set $\mathcal{P}_{\mathcal{Y}|\mathcal{X}}$ is convex. □

While Lemma 1 proves that the set of behaviors, $\mathcal{P}_{\mathcal{Y}|\mathcal{X}}$ is convex, it is also important to prove that it is a convex polytope, which requires that the set $\mathcal{P}_{\mathcal{Y}|\mathcal{X}}$ has a finite number of extreme points [118]. These extreme points correspond to *deterministic behaviors* that map an input $x$ to the output $y$ with certainty.

**Definition 5. Deterministic Behavior:** A black-box behavior whose matrix elements are $0/1$ values such that

$$P(y|x) \in \mathbb{B} \quad \forall \quad x \in \mathcal{X} \quad \text{and} \quad y \in \mathcal{Y}, \tag{2.6}$$

where $\mathbb{B} \equiv \{0, 1\}$.

**Lemma 2.** The extreme points of the full probability polytope $\mathcal{P}_{\mathcal{Y}|\mathcal{X}}$ are exactly the complete set of deterministic behaviors.

*Proof.* Let $\mathbf{E} \in \mathbb{R}^{|\mathcal{Y}|\times|\mathcal{X}|}$ be a perturbation to a behavior $\mathbf{P} \in \mathcal{P}_{\mathcal{Y}|\mathcal{X}}$ such that

$$\frac{1}{2}(\mathbf{P}+\mathbf{E}) + \frac{1}{2}(\mathbf{P}-\mathbf{E}) = \mathbf{P} \in \mathcal{P}_{\mathcal{Y}|\mathcal{X}} \tag{2.7}$$

and $\mathbf{P}$ is rewritten as the convex combination of two behaviors $(\mathbf{P} \pm \mathbf{E}) \in \mathcal{P}_{\mathcal{Y}|\mathcal{X}}$. A behavior is extreme if it cannot be rewritten in terms of perturbations as in Eq. (2.7). Let $\mathcal{V}_{\mathcal{Y}|\mathcal{X}}$ denote the complete set of deterministic behaviors $\mathbf{D} : \mathcal{X} \to \mathcal{Y}$ where $D(y|x) \in \mathbb{B}$ for all $x \in \mathcal{X}$ and $y \in \mathcal{Y}$. The deterministic behavior is extreme because one perturbation, $(\mathbf{P} \pm \mathbf{E})$, is not a valid behavior. Furthermore, consider a stochastic behavior $\mathbf{P} \notin \mathcal{V}_{\mathcal{Y}|\mathcal{X}}$. $\mathbf{P}$ always admits the decomposition in (2.7) where $\mathbf{E}$ is

$$\mathbf{E} = \sum_{x\in\mathcal{X}} \sum_{y\in\mathcal{Y}} \delta_{y,y_{\max}(x)} P(y_{\min}(x)|x) - \delta_{y,y_{\min}(x)} P(y_{\min}(x)|x)\, |y\rangle\langle x|\,, \tag{2.8}$$

where $y_{\max}(x)$ and $y_{\min}(x)$ are the row ids of largest and smallest values in column $x$. Thus, the extreme points of $\mathcal{P}_{\mathcal{Y}|\mathcal{X}}$ is the complete set of deterministic behaviors $\mathcal{V}_{\mathcal{Y}|\mathcal{X}}$. □

Deterministic behaviors provide an interesting interpretation of the discrete memoryless setting. Given a single sampled input $x$, the output data appears deterministic because only one sample has been taken. More formally, let's refer to a round in the discrete memoryless setting as sampling the black-box's output once for each input $x \in |\mathcal{X}|$. In aggregate, the collection of $|\mathcal{X}|$ samples describes a deterministic behavior because the black-box outputs one value $y \in \mathcal{Y}$ for each input. After sampling the black-box many times in the discrete memoryless setting, the data can be grouped into $N$ complete rounds where $\mathbf{D}_i : \mathcal{X} \to \mathcal{Y}$ denotes the deterministic behavior describing the $i^{th}$ round of the experiment. The average behavior over $N$ samples is then

$$\mathbf{P} = \frac{1}{N}\sum_{i=1}^{N} \mathbf{D}_i, \tag{2.9}$$

which is only deterministic if $\mathbf{D}_i = \mathbf{D}_j$ for all $i, j \in [N]$. An important fact results from Eq. (2.9) where any black-box behavior decomposes as a convex combination of deterministic behaviors

$$\mathbf{P} = \sum_{i=1}^{N} \mu_i \mathbf{D}_i \quad \text{where} \quad \mu_i > 0 \quad \text{and} \quad \sum_{i=1}^{N} \mu_i = 1. \tag{2.10}$$

In Eq. (2.10), the numerical precision of the convex weights $\mu_i$ is dependent upon the number of sampled rounds used to construct the behavior. When the number of shots is large, the convex weights effectively become continuous on the range $\mu_i \in [0, 1]$, meaning that any behavior $\mathbf{P} \in \mathcal{P}_{\mathcal{Y}|\mathcal{X}}$ can be expressed as a convex combination of deterministic behaviors.

One drawback of this sampling approach is that the input is drawn from a uniform distribution, meaning that there is no guarantee that $N$ samples of each input $x \in \mathcal{X}$ will be obtained within a total of $n$ shots. In principle, a black-box that has large input alphabet, may require a very large number of shots to ensure that at least $N$ samples of each input have been obtained. A workaround for this sampling issue can be achieved if the black-box is trusted to be memoryless. In this case, the inputs no longer need be drawn from a uniform distribution because the black-box's behavior is not conditioned upon the data of previous experiments. As a result, $N$ complete rounds of one-shot experiments can be completed in exactly $|\mathcal{X}|N$ shots.

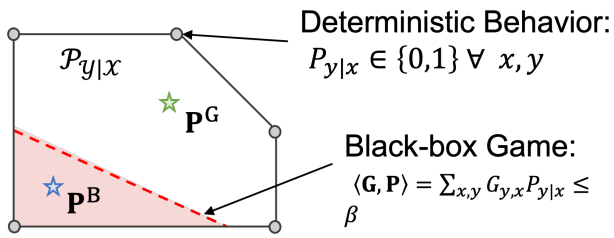



Figure 2.2: A two-dimensional graphical analogy of the probability polytope $\mathcal{P}_{\mathcal{Y}|\mathcal{X}}$ where a particular black-box behavior corresponds to a point within this space. The whole set of black-box behaviors is contained by the convex polygon with blue edges. The vertices of the polygon denote deterministic behaviors that are extreme points of the set of black-box behaviors. The red dashed line denotes a linear black-box game. The shaded red region shows the set of all behaviors that achieve an average score that loses the game, *e.g.*, the blue star $\mathbf{P}^B$. The nonshaded region shows the set of all behaviors that achieve an average score that wins the game, *e.g.*, the green star $\mathbf{P}^G$.

## *2.3 Playing Games with Black-Boxes*

A black-box is a processing device that maps the input alphabet $\mathcal{X}$ to the output alphabet $\mathcal{Y}$. From an operational perspective, a black-box device performs a processing task on its input data. For example, a black-box whose task is to serve as a communication channel between two parties would need its output to match the input $y = x$ with a high-probability. In this case, the optimal black-box behavior would be the identity $\mathbf{P} = \mathbb{I}$ where $P(y = x|x) = 1$ for all $x \in \mathcal{X}$. Different tasks require different types of behaviors, thus a framework is needed to quantify a black-box's performance at a particular task. To this end, we apply a game theoretic approach to black-box models.

In this game theoretic framework, the black-box is viewed as a player in the game and the experimenter is viewed as the referee. The referee samples input $x \in \mathcal{X}$ with uniform probability $P(x) = 1/|\mathcal{X}|$ and passes it to the black-box, which outputs $y \in \mathcal{Y}$ with probability $P(y|x)$. The black-box is assumed to operate in the discrete memoryless setting hence, after many experimental runs, a full behavior $\mathbf{P}$ can be characterized. Each pairing of input and output is then associated with either a reward or a cost, allowing the average score to be evaluated from the probabilities $P(y|x)$.

**Definition 6. Black-Box Game:** A tuple $(\gamma, g)$ that contains a minimum average winning score $\gamma \in \mathbb{R}$ and a predicate function $g : \mathcal{P}_{\mathcal{Y}|\mathcal{X}} \to \mathbb{R}$ that quantifies the average performance of a black-box $\mathbf{P} \in \mathcal{P}_{\mathcal{Y}|\mathcal{X}}$ in the one-shot setting given uniform prior distribution $P(x) = 1/|\mathcal{X}|$. A winning condition for the game is encoded into the inequality

$$\gamma \geq g(\mathbf{P}) \tag{2.11}$$

where $g(\mathbf{P})$ is the average score and $\gamma \in \mathbb{R}$ is the minimal average score needed to "win" the game.

In general, the predicate function $g(\mathbf{P})$ is unrestricted beyond the fact that it must accept a black-box behavior and output a scalar, which leaves quite a bit of flexibility in how a task is quantified. By Lemma 2 any behavior can be expressed as a convex combination of deterministic behaviors $\mathbf{P} = \sum_i \mu_i \mathbf{D}_i$. It follows that a linear predicate function satisfies

$$g(\mathbf{P}) = g\left(\sum_i \mu_i \mathbf{D}_i\right) = \sum_i \mu_i g(\mathbf{D}_i), \tag{2.12}$$

while a nonlinear predicate function does not. When a game is linear the score $g(\mathbf{P})$ is the weighted average of the independent scores for the deterministic behaviors $g(\mathbf{D}_i)$. On the other hand, nonlinear predicate functions do not decompose into a weighted average over $g(\mathbf{D}_i)$.

An example of a nonlinear game that quantifies communication through a black-box is the mutual information between input or output can be considered

$$g(\mathbf{P}) = I(X;Y) \equiv H(X) + H(Y) - H(X,Y) \tag{2.13}$$

where $X$ and $Y$ denote random variables on the input and output and $H(\cdot)$ denotes the Shannon entropy of a marginal or a joint distribution

$$H(X) \equiv -\sum_{x\in\mathcal{X}} P(x)\log_2\left(P(x)\right) \quad \text{and} \quad H(X,Y) \equiv -\sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}} P(x,y)\log_2\left(P(x,y)\right). \tag{2.14}$$

However, a difficulty with nonlinear predicates, such as entropic quantities, is that the score can only be evaluated after many shots are taken. This is because the Shannon entropy calculated from a single shot of the experiment is always 0. Thus, nonlinear predicate functions should only be scored against behaviors, which have been estimated from many shots. We now turn to formalize linear black-box games with linear predicate functions.

**Definition 7. Linear Black-Box Game:** A tuple $(\gamma, \mathbf{G})$ containing the minimum average score $\gamma \in \mathbb{R}$ and predicate matrix $\mathbf{G} \in \mathbb{R}^{|\mathcal{Y}|\times|\mathcal{X}|}$. For a given linear black-box game, the average score achieved by a black-box behavior $\mathbf{P} : \mathcal{X} \to \mathcal{Y}$ with uniform prior distribution is

$$\gamma \geq \langle \mathbf{G}, \mathbf{P}\rangle \equiv \mathrm{Tr}\left[\mathbf{G}^T\mathbf{P}\right] = \sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}} G_{y,x}P(y|x). \tag{2.15}$$

Since the predicate function is linear, we may express it as a weighted average over the deterministic behaviors

$$g(\mathbf{P}) = \langle \mathbf{G}, \mathbf{P}\rangle = \left\langle \mathbf{G}, \sum_i \mu_i \mathbf{D}_i \right\rangle = \sum_i \mu_i \langle \mathbf{G}, \mathbf{D}_i\rangle. \tag{2.16}$$

Furthermore, the inner-product $\langle \mathbf{G}, \mathbf{P}\rangle = \sum_{x,y} G_{y,x}P(y|x)$ shows that each conditional probability $P(y|x)$ is scaled by a distinct game coefficient $G_{y,x}$. Additionally, $G_{y,x}$ can be interpreted as the reward or loss for outputting $y$ given input $x$. When averaged over many shots of the game with $x$ being drawn from a uniform distribution the score $g(x) = \langle \mathbf{G}, \mathbf{P}\rangle$ is recovered.

**Lemma 3.** Given a linear black-box game $(\gamma, \mathbf{G})$ the maximal score that can be achieved for any behavior $\mathbf{P} \in \mathcal{P}_{\mathcal{Y}|\mathcal{X}}$ is

$$\gamma^\star = \max_{\mathbf{P}\in\mathcal{P}_{\mathcal{Y}|\mathcal{X}}} \langle \mathbf{G}, \mathbf{P}\rangle = \sum_{x\in\mathcal{X}} \max_{y\in\mathcal{Y}} G_{y,x} \tag{2.17}$$

and the minimal score that can be achieved is

$$\min_{\mathbf{P}\in\mathcal{P}_{\mathcal{Y}|\mathcal{X}}} \langle \mathbf{G}, \mathbf{P}\rangle = \sum_{x\in\mathcal{X}} \min_{y\in\mathcal{Y}} G_{y,x} \tag{2.18}$$

where in both cases a deterministic behavior $\mathbf{P}^\star$ having 0/1-elements can achieve the optimum.

Although Lemma 3 provides hard bounds on the possible scores that can be achieved in a black-box game,

in practice, the lower bound can be avoided by considering a black-box that outputs white noise.

**Lemma 4.** Given a linear black-box game $(\gamma, \mathbf{G})$, an ignorant black-box with no knowledge of the game being played can always achieve the score

$$\gamma^0 = \sum_{x\in\mathcal{X}} \sum_{y\in\mathcal{Y}} \frac{G_{y,x}}{|\mathcal{Y}|} = \langle \mathbf{G}, \mathbf{P} \rangle \tag{2.19}$$

by exhibiting the white noise behavior $\mathbf{P} = \sum_{x\in\mathcal{X}} \sum_{y\in\mathcal{Y}} \frac{1}{|\mathcal{Y}|} |y\rangle\langle x|$.

As given by Lemma 4 a behavior that exhibits white noise, always outperforms the worst case lower bound in Eq. (2.18) unless, for all $x \in \mathcal{X}$ and $y \in \mathcal{Y}$, $G_{y,x} = c$ for some constant $c$. In this case, however, the black-box game is quite useless because its score is constant for all behaviors $\mathbf{P} \in \mathcal{P}_{\mathcal{Y}|\mathcal{X}}$. As a result, in any game scenario, if a black-box performs worse than the white noise score in Lemma 4 it should be considered as failing to play the game. The reason being that if any amount of correlation exists between the input $\mathcal{X}$ and output $\mathcal{Y}$ alphabets, then the score must be better than the white noise score because the white noise score can be achieved when there exists no correlation between input or output.

## *2.4 Simulating Black-Box Behaviors*

Since a black-box is characterized by its behavior $\mathbf{P} \in \mathcal{P}_{\mathcal{Y}|\mathcal{X}}$, a black-box is simulated if and only if its behavior is reproduced. Following the approach used by Brito *et al.* [98], we quantify the closeness of two black-box behaviors using the following distance measure. Then, we can formally define exact and approximate simulations of black-boxes.

**Definition 8. Behavior Distance:** $D(\mathbf{P}, \mathbf{P}')$, The distance between two behaviors $\mathbf{P}, \mathbf{P}' \in \mathcal{P}_{\mathcal{Y}|\mathcal{X}}$ where

$$D(\mathbf{P}, \mathbf{P}') = \frac{1}{2|\mathcal{X}|} ||\mathbf{P} - \mathbf{P}'||_1 = \frac{1}{2|\mathcal{X}|} \sum_{x\in\mathcal{X}} \sum_{y\in\mathcal{Y}} |P(y|x) - P'(y|x)|, \tag{2.20}$$

and the factor of $1/|\mathcal{X}|$ results from the inputs $x \in \mathcal{X}$ being drawn with uniform probability.

**Definition 9.** Consider two black-box behaviors $\mathbf{P}, \mathbf{P}' \in \mathcal{P}_{\mathcal{Y}|\mathcal{X}}$.

(i) **Exact Simulation:** One behavior *exactly* simulates another if and only if $D(\mathbf{P}, \mathbf{P}') = 0$.

(ii) **Approximate Simulation:** One behavior *approximately* simulates another if and only if $D(\mathbf{P}, \mathbf{P}') \leq \epsilon$ where $0 < \epsilon \ll 1$ is an allowable tolerance for error.

While the exact simulation of a behavior may be theoretically useful, the approximate simulation is more practical. To see this, we must understand that behaviors result from stochastic processes, in which two behaviors are the same only if their respective finite sets of sampled data are identical up to a reordering. Thus, the exact simulation of a behavior exists only in an idealized sense and, as a result, it is more physically relevant to consider simulating a behavior in the approximate sense.

# Chapter 3

# Classical Communication Networks

Classical communication networks are ubiquitous across science, technology, and society. Everyday we use communication networks for making online purchases, making wireless phone calls, and using GPS. In general, classical communication networks consists of many communicating parties, each operating independently from the others based upon their local information. There is also a causal flow of information through the parties in the communication network. At their root, communication networks transmit data between devices enabling distributed applications in computing, sensing, and communications.

In this section, we extend the black-box model to a general model of classical communication networks. In our model, black-boxes serve as network devices and the communication between the black-box devices imposes a causal structure on the network. Our approach builds upon the frameworks of network information theory [119], causal inference [120], and Bayesian network theory [121].

## *3.1 A Communication Network as a Black-Box*

In a classical communication network, there is a causal flow of information between devices. That is, the network accepts input data, processes the information, and outputs some data at a later time. Naturally, a classical communication network can be modeled as a black-box. To extend the black-box model of Chapter 2 to networks we consider multiple inputs represented as sequence of values $(x_1, \ldots, x_n) = \vec{x}^{\text{Net}} \in \mathcal{X}^{\text{Net}}$ where each value is given to a specific device in the network. Likewise the output $(y_1 \ldots, y_m) = \vec{y}^{\text{Net}} \in \mathcal{Y}^{\text{Net}}$ can distributed amongst many devices. As a black-box, a classical communication network is characterized by its behavior $\mathbf{P}^{\text{Net}} : \mathcal{X}^{\text{Net}} \to \mathcal{Y}^{\text{Net}}$ as described in Definition 3.

Caution should be taken when modeling a communication network as a black-box because many classical communication networks do no operate in the discrete memoryless setting. That is, real-world communication networks, such as the internet, often rely upon persistent memory and maintaining a record of transactions. Such data record is essential to the networking applications that society depends on. However, the objective of this dissertation is not to model real-world classical networks, but to compare classical networks to their quantum counterparts. As will be reiterated later, the discrete memoryless setting is practical for many real-world quantum systems. Therefore, it is illustrative to view classical networks as a black-box in the discrete memoryless setting.

The black-box model provides a convenient framework for characterizing the behavior of networks, however,

the black-box model is too general because the black-box can construct any conditional probability distribution that maps the input set $\mathcal{X}^{\text{Net}}$ to the output set $\mathcal{Y}^{\text{Net}}$. Since the network's probabilities are unrestricted, its ability to perform any task encoded as black-box game is also unrestricted. That is, the network can solve any communication or processing problem perfectly. Realistically, a communication network has restrictions on the behaviors it can exhibit. For example, if the amount of communication between parties is restricted, data is lost as it propagates from input to output. This loss of data constrains the network's behavior such that it cannot achieve perfect communication from sender to receiver. Thus, this lost of information must be accounted for in the behaviors produced by the network.

In general, a communication network's causal structure and communication resources impose natural restrictions on the network behaviors $\mathbf{P}^{\text{Net}} : \mathcal{X}^{\text{Net}} \to \mathcal{Y}^{\text{Net}}$. It is therefore important to peer inside the black-box and investigate its causal structure and information flow. In doing so, we can identify important constraints on the network's behaviors given its causal structure and communication resources. As will be discussed later, these classical restrictions are precisely where nonclassicality can be observed in quantum systems.

## *3.2 A Communication Network of Black-Boxes within a Black-Box*

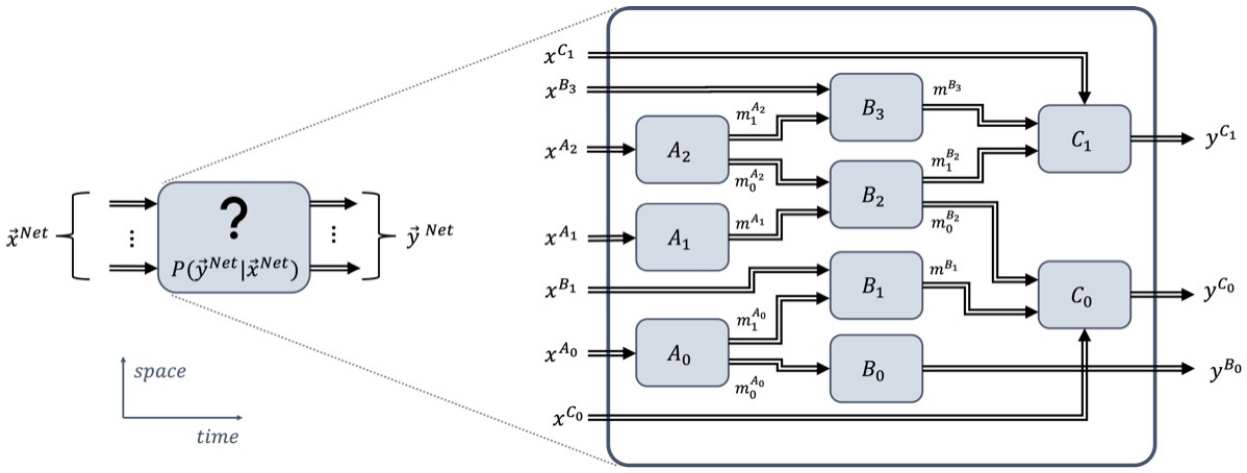



Figure 3.1: A classical communication network of black-box devices within a black-box device. Gray box represents a network device modeled as a black-box. The double-lined arrows denote the classical communication between devices.

**Definition 10. Communication Network Topology:** $\text{Net}(\vec{N}, \vec{E})$, A communication network's causal structure is represented graphically as a directed acyclic graph (DAG) where each node represents a networking device and the double-lined arrows depict classical communication (see Fig. 3.1). Equivalently, the topology can be expressed as a collection of nodes $\vec{N}$ and their connecting edges $\vec{E}$ where

$$\vec{N} \equiv \left(\vec{N}_i \equiv (\text{Dev}_{i,j})_{j=1}^{|\vec{N}_i|}\right)_{i=1}^{|\vec{N}|} \quad \text{and} \quad \vec{E} \equiv \left(\text{id}^{\text{Tx}_i \to \text{Rx}_i}\right)_{i=1}^{|\vec{E}|}. \tag{3.1}$$

In Eq. (3.1), the nodes $\vec{N}$ are ordered as layers of devices $\vec{N}_i$ where $\text{Dev}_{i,j}$ labels the $j^{th}$ device in the $i^{th}$ layer. The device layers occur sequentially in time and designate time steps in the flow of information while devices in the same layer process their local information in parallel. While the device labels $\text{Dev}_{i,j}$ can be

adapted to any network structure, the notation can become quite onerous. Our solution is to label network layers alphabetically, *e.g.*, $\vec{N} = (\vec{N}^A, \vec{N}^B, \vec{N}^C, \dots)$. Devices within each layer share the same alphabetic label, but with a numeric subscript, *e.g.*, $\vec{N}^A = (A_1, A_2, A_3, \dots)$. Throughout this work, the notation $\text{Dev}_{i,j}$ is used when discussing devices in general networks while alphabetic labels are used when referring to devices in a particular network or DAG.

The edges $\vec{E}$ in Eq. (3.1) represent noiseless one-way channels $\text{id}^{\text{Tx}\to\text{Rx}}$ from a sender device ($\text{Tx} \in \vec{N}$) to a receiver device ($\text{Rx} \in \vec{N}$). In each use of the one-way classical channel a single message $m \in \mathcal{M}^{\text{Tx}\to\text{Rx}}$ is transmitted where $\mathcal{M}^{\text{Tx}\to\text{Rx}}$ is the message alphabet used for the communication and $\mathcal{M}^{\text{Net}} \equiv \prod_{i=1}^{|\vec{E}|} \mathcal{M}^{\text{Tx}_i\to\text{Rx}_i}$ denotes the set of all message alphabets. It is important to note that devices within the same layer are assumed to not communicate. Therefore, if the sender is $\text{Dev}_{i,j}$ and receiver is $\text{Dev}_{i',j'}$, then $i < i'$ must hold, meaning that communication occurs only between devices in subsequent layers.

Additionally, the network's devices may have inputs and outputs. If we denote the input and output alphabet of $\text{Dev}_{i,j}$ as $\mathcal{X}_{i,j}$ and $\mathcal{Y}_{i,j}$, then the network's aggregate input and output alphabets are

$$\mathcal{X}^{\text{Net}} \equiv \prod_{i=1}^{|\vec{N}|} \prod_{j=1}^{|\vec{N}_i|} \mathcal{X}_{i,j} \quad \text{and} \quad \mathcal{Y}^{\text{Net}} \equiv \prod_{i=1}^{|\vec{N}|} \prod_{j=1}^{|\vec{N}_i|} \mathcal{Y}_{i,j}. \tag{3.2}$$

Note that in Eq. (3.2) both the number of inputs $|\mathcal{X}^{\text{Net}}|$ and outputs $|\mathcal{Y}^{\text{Net}}|$ for the network scale exponentially with number of nodes $|\vec{N}| = \prod_{i=1}^{|\vec{N}|} |\vec{N}_i|$.

By imposing a strict spatial and causal structure onto the information flow of the black-box, we limit the behaviors $\mathbf{P}^{\text{Net}} : \mathcal{X}^{\text{Net}} \to \mathcal{Y}^{\text{Net}}$ that the network can produce. Indeed, the limitations on the network's behavior correspond to fundamental physical constraints. Namely, devices within the same layer $\vec{N}_j$ are independent from one another. This independence corresponds to a *space-like separation* between devices in the same layer, meaning that they do not communicate or are *nonsignaling*. Likewise, when there is communication between devices, a strict *time-like separation* is imposed. That is, there is a causal link connecting the sender to the receiver, which is represented by an arrow in the network's DAG. A device $\text{Dev}_{i,j}$ is said to have causal influence on $\text{Dev}_{i',j'}$ if a path can be traversed from $\text{Dev}_{i,j}$ to $\text{Dev}_{i',j'}$ while following the flow of the network's DAG.

## 3.3 Classical Networking Devices

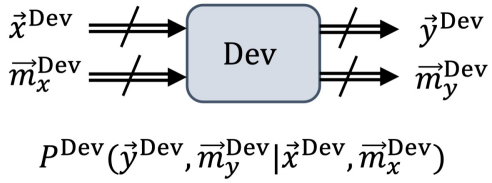



Figure 3.2: A black-box model of a generic classical networking device.

**Definition 11. Classical Networking Device:** $\text{Dev}_{i,j}$, A device that receives, processes, and transmits classical data in a communication network $\text{Net}(\vec{N}, \vec{E})$. Let $\text{Dev}_{i,j}$ denote the networking device represented by the $i^{th}$ node in the $j^{th}$ layer of the network's nodes $\vec{N}$. As input, the device may accept classical value $x_{i,j} \in \mathcal{X}_{i,j}$ and a collection of classical messages $\vec{m}_{i,j}^{\text{Rx}} \in \mathcal{M}_{i,j}^{\text{Rx}}$ that $\text{Dev}_{i,j}$ receives from its parent nodes. As

output, the device may produce a classical output $y_{i,j} \in \mathcal{Y}_{i,j}$ and a collection of classical messages $\vec{m}_{i,j}^{\mathrm{Tx}}$ that $\mathrm{Dev}_{i,j}$ transmits to its children nodes.

Without loss of generality, classical networking devices can be modeled as a black-box (see Def. 1). Just like a black-box, networking devices receive, process, and transmit classical data in a causal manner. Furthermore, if the device is assumed to operate in the discrete memoryless setting (see Def. 2), it can be characterized by its black-box behavior $\mathbf{P}_{i,j} : \mathcal{X}_{i,j} \times \mathcal{M}_{i,j}^{\mathrm{Rx}} \to \mathcal{Y}_{i,j} \times \mathcal{M}_{i,j}^{\mathrm{Tx}}$ where by Definition 3 the device behavior is expressed as

$$\mathbf{P}_{i,j} \equiv \sum_{x \in \mathcal{X}_{i,j}} \sum_{y \in \mathcal{Y}_{i,j}} \sum_{\vec{m}^{\mathrm{Rx}} \in \mathcal{M}_{i,j}^{\mathrm{Rx}}} \sum_{\vec{m}^{\mathrm{Tx}} \in \mathcal{M}_{i,j}^{\mathrm{Tx}}} P(y, \vec{m}^{\mathrm{Tx}} | x, \vec{m}^{\mathrm{Rx}}) \left| y, \vec{m}^{\mathrm{Tx}} \right\rangle \left\langle x, \vec{m}^{\mathrm{Rx}} \right| . \tag{3.3}$$

Modeling classical networking devices as black-boxes is important to this dissertation because no limitations are placed on the device's operation other than the size of of the device's input and output. Therefore, any constraints imposed on the behavior of a network $\mathrm{Net}(\vec{N}, \vec{E})$ must result from the network's causal structure and the amount of communication used between signaling devices.

## 3.4 Classical Networking Resources

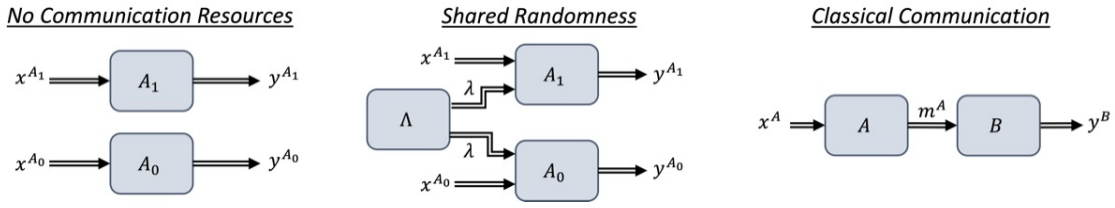



Figure 3.3: **Resources in Classical Communication Scenarios.** From right to left we see a *nonsignaling* scenario where no communication resources are present, a *one-way communication* scenario where device $A$ sends a message to device $B$, and a *shared randomness* scenario where two nonsignaling devices, $A_0$ and $A_1$, each receive a message from a common source device $\Lambda$.

From an operational perspective, the goal of a classical communication network is to perform a task in multiparty communication or distributed information processing. A network's ability to perform its task is limited by the amount of communication between networking devices. Indeed, the communication between devices can be thought of as a fundamental resource that enables a network to perform its task. Classical communication resources are quantified by the number of communicated *bits* where each bit corresponds to a binary value $b \in \mathbb{B} \equiv \{0, 1\}$ that is communicated between two devices. Furthermore, resources can be given value and used to formally establish cost and performance metrics for the considered networking task.

We consider two general types of communication resources, dynamic and static. Dynamic resources, such as classical communication, are transmitted from a sender to a receiver. This communication leads to a causal link from the sender to the receiver. Static resources, such as shared randomness, can be used to correlate the operations of networking devices. However, correlation does not imply causation, meaning that correlated devices do not communicate information.

### 3.4.1 Nonsignaling Devices

Consider the case where devices in the network $\text{Net}(\vec{N}, \vec{E})$ do not communicate such that $|\vec{E}| = 0$. In this case, the network can be expressed as a single layer of devices $\vec{N}^A = (A_j)_{j=1}^n$. The behaviors of each device then combine as a tensor product

$$\mathbf{P}^{\text{Net}} = \bigotimes_{j=1}^{n} \mathbf{P}^{A_j} \quad \text{and} \quad P(\vec{a}|\vec{x}) = \prod_{j=1}^{n} P(a_j|x_j). \tag{3.4}$$

Resulting from Eq. (3.4) are the nonsignaling constraints

$$P(a_j|x_j) = \sum_{\vec{a}_{j'\neq j} \in \mathcal{A}/\mathcal{A}_j} P(\vec{a}_{j'\neq j}, a_j|\vec{x}) \quad \forall \quad a_j \in \mathcal{A}_j, \quad \text{and} \quad \vec{x} \in \mathcal{X} \tag{3.5}$$

where the sum $\vec{a}_{j'\neq j}\mathcal{A}/\mathcal{A}_j$ is over all outputs, excluding the output of the $j^{th}$ device. The nonsignaling constraint in Eq. (3.5) simply states that each device's output is independent from the inputs given to other devices.

### 3.4.2 Classical Communication

**Definition 12. Classical Communication Resource:** A dynamic communication resource used to send a discrete message $m \in \mathcal{M}^{\text{Tx}\to\text{Rx}} \equiv \{0, 1, \ldots |\mathcal{M}^{\text{Tx}\to\text{Rx}}| - 1\}$ from a sender to receiver.

In a network's DAG, classical communication is represented as a double-lined arrow from the sender device (Tx) to the receiver device (Rx). Since the amount of communication between black-boxes restricts the possible behaviors that a network can exhibit, it is important that we quantify precisely the amount of communication between each device. To this end, we quantify the amount of communication by the size of the alphabet used for communication, $\kappa(\mathbf{P}^{\text{Tx}\to\text{Rx}}) = |\mathcal{M}^{\text{Tx}\to\text{Rx}}|$, a quantity referred to as the *signaling dimension*. We adopt the signaling dimension from previous works [114], [122], [123] and provide a formal description in Chapter 14. Generally, the signaling dimension quantifies the amount of classical messages needed to simulate a given behavior, however when the channel is noiseless, the signaling dimension is simply the size of the message alphabet. The total amount of communication used in a network $\text{Net}(\vec{N}, \vec{E})$ is then

$$|\mathcal{M}^{\text{Net}}| = \prod_{i=1}^{|\vec{E}|} |\mathcal{M}^{\text{Tx}_i\to\text{Rx}_i}|, \tag{3.6}$$

where the corresponding number of bits is then $\log_2\left(|\mathcal{M}^{\text{Net}}|\right)$, however, care should be taken because Eq. (3.6) dissociates the communication resources from the network's causal structure.

While we investigate the signaling dimension of noisy behaviors in Chapter 14, here we focus on noiseless communication channels such that the classical network is restricted by the amount of communication, not the quality of communication. Thus, the behavior $\mathbf{P}^{\text{Tx}\to\text{Rx}} = \mathbb{I}_d$ where $d = |\mathcal{M}^{\text{Tx}}| = |\mathcal{M}^{\text{Rx}}|$ is deterministic and communicates the message $m \in \mathcal{M}^{\text{Tx}}$ from the sender to the receiver without error. The signaling dimension is then $\kappa(\mathbf{P}^{\text{Tx}\to\text{Rx}}) = d$, which corresponds to the size of message alphabet used for communication. In this way, one-bit of communication would have $d = 2$ while $n$-bits of communication would have $d = 2^n$. Note that if the communication channel were noisy, then the signaling dimension can decrease, hence, assuming noiseless communication is the least restrictive constraint to place on the classical network's communication.

When there is communication between two devices, their behaviors combine through matrix multiplication. Let $\mathbf{P}^A$ and $\mathbf{P}^B$ be the behaviors of the sender and receiver devices respectively (see Fig. 3.3 center). Their joint behavior is expressed as

$$\mathbf{P}^{A\to B} = \mathbf{P}^B\mathbf{P}^A \quad \text{and} \quad P(y|x) = \sum_{m\in\mathcal{M}^{A\to B}} P(y|m)P(m|x) \tag{3.7}$$

where the resulting behavior $\mathbf{P}^{A\to B}$ is indexed only by the input $x \in \mathcal{X}$ and output $b \in \mathcal{B}$ because we sum over all possible messages $m \in \mathcal{M}^{A\to B}$. When device $A$ and $B$ have more inputs and outputs, the matrix multiplication becomes a tensor contraction, however, we will discuss this case in greater detail in Section 3.5.

### *3.4.3 Shared Randomness*

Shared randomness is produced by source devices $\Lambda_i$ that are represented in a network's DAG as an elliptical node. A source uses classical communication to distributes a shared classical message $\lambda_i \in \mathcal{M}^{\Lambda_i}$ to all linked devices (see Fig. 3.3). The devices can then coordinate their operations using their shared random value.

**Definition 13. Source Device:** $\Lambda$, a device that produces $m$ random variables $\vec{\lambda} = (\lambda_i \in [d])_{i=1}^m$ and distributes each value to its receiving device using classical communication. Since source device's do not have inputs, their behaviors are expressed as vectors on a $d^m$ dimensional vector space.

$$\mathbf{P}^\Lambda = \sum_{\vec{\lambda}\in[d]^m} P(\vec{\lambda}) \left|\vec{\lambda}\right\rangle \tag{3.8}$$

where $\left|\vec{\lambda}\right\rangle = \bigotimes_{i=1}^m |\lambda_i\rangle$ and $\lambda_i \in \vec{\lambda}$.

**Definition 14. Shared Randomness:** A static communication resource used to correlate the behaviors of two or more devices. A source device distributes shared randomness if its behavior decomposes as

$$\mathbf{P}^\Lambda = \sum_{\lambda=1}^{d} P(\lambda) \bigotimes_{i=1}^{m} |\lambda\rangle \tag{3.9}$$

such that each recipient device receives the value $\lambda$ with probability $P(\lambda)$.

It is important to note that not all random distributions output from a source device are considered shared randomness, only those in which each subscribed device receives the same value such that $\lambda$ is shared. The effect of shared randomness on receiving device can be described through example. Consider a layer of $n$ nonsignaling devices that share a common randomness source $\Lambda$. Their behaviors combine as

$$\mathbf{P}^A = \sum_{\lambda\in\mathcal{M}^\Lambda} P(\lambda)\mathbf{P}_\lambda^{A_1} \otimes \cdots \otimes \mathbf{P}_\lambda^{A_n} \tag{3.10}$$

where $\mathbf{P}_\lambda^{A_i} \in \mathcal{P}_{\mathcal{A}_i|\mathcal{X}_i}$, $\sum_{\lambda\in\mathcal{M}^\Lambda} P(\lambda) = 1$, and $P(\lambda) \geq 0$. The conditional probabilities are

$$P(\vec{a}|\vec{x}) = \sum_{\lambda\in\mathcal{M}^\Lambda} \prod_{j=1}^{n} P(a_j|x_j,\lambda)P(\lambda). \tag{3.11}$$

Note that Eq. (3.11) satisfy the nonsignaling constraints of Eq. (3.5), showing that shared randomness does not allow devices to communicate.

To quantify the amount of shared randomness, we can again apply the signaling dimension. If we assume that the shared randomness is communicated without noise, the respective classical communication channels have behaviors that correspond to the identity matrix $\mathbf{P}^{\Lambda\to A_i} = \mathbb{I}_{|\Omega^\Lambda|}$ where the signaling dimension is $\kappa(\mathbf{P}^{\Lambda\to A_i}) = |\Omega^\Lambda|$ because $\Omega^\Lambda$ is the alphabet of discrete shared random values. Thus, the amount of shared randomness is quantified as the minimum number of discrete random values needed to distribute the shared random variable to one device. Note that the number of devices that share randomness could also be considered, however, the amount of random values that each device can apply is still fixed as $|\Omega^\Lambda|$.

In this work we will often consider the case where an unlimited amount of shared randomness is available, in which case $|\Omega^\Lambda| \to \infty$. One reason being that this presents the best case for classical network. Note that the devices could distribute the shared random values prior to use and used at a later time to correlate the devices without needing communication.

## *3.5 Characterizing the Behaviors of Classical Networks*

**Definition 15. Classical Network Behavior: $\mathbf{P}^{\text{Net}}$**, The black-box behavior associated with a classical network $\text{Net}(\vec{N}, \vec{E})$ where

$$\mathbf{P}^{\text{Net}} = \sum_{\vec{x}\in\mathcal{X}^{\text{Net}}} \sum_{\vec{y}\in\mathcal{Y}^{\text{Net}}} P(\vec{y}|\vec{x})\, |\vec{y}\rangle\langle\vec{x}| \tag{3.12}$$

with each conditional probability decomposing as

$$P(\vec{y}|\vec{x}) = \sum_{\vec{m}^{\text{Net}}\in\mathcal{M}^{\text{Net}}} \prod_{i=1}^{|\vec{N}|} \prod_{j=1}^{|\vec{N}_j|} P_{i,j}(y, \vec{m}^{\text{Tx}}|x, \vec{m}^{\text{Rx}}) \tag{3.13}$$

where $x \in \mathcal{X}_{i,j}$ and $y \in \mathcal{Y}_{i,j}$ are the inputs and output of $\text{Dev}_{i,j}$ while $\vec{m}^{\text{Rx}}$ and $\vec{m}^{\text{Tx}}$ are the messages received and transmitted by $\text{Dev}_{i,j}$.

**Definition 16. Set of Classical Network Behaviors: $\mathcal{C}^{\text{Net}}$**, The set of behaviors that can be generated by the network structure $\text{Net}(\vec{N}, \vec{E})$ provided in a given a classical network Net. All behaviors $\mathbf{P}_{\text{Net}} \in \mathcal{C}^{\text{Net}}$ must satisfy Def. 15. Note that the notation $\mathcal{C}^{\text{Net}}$ will typically be used when discussing the sets of classical networks in general.

For any classical network, the set of behaviors is contained by the set of black-box behaviors

$$\mathcal{C}^{\text{Net}} \subseteq \mathcal{P}_{\mathcal{Y}|\mathcal{X}}. \tag{3.14}$$

The subset is strict when communication resource limitations restrict the set of behaviors. Recall that the extreme points of $\mathcal{P}_{\mathcal{Y}|\mathcal{X}}$ are the deterministic behaviors with 0/1 matrix elements. Therefore, any classical network behavior that is deterministic must be extreme. Furthermore, the extreme points of the set of classical network behaviors are also deterministic [109], [124]. These extreme points will be denoted $\mathcal{V}^{\text{Net}}$. Since the extreme points are deterministic, they can be achieved without using any shared randomness because shared randomness would necessarily incorporate stochasticity into the behavior.

When randomness is shared amongst all devices in the network. The resulting set of network behaviors becomes a convex polytope

$$\mathcal{C}_\Lambda^{\text{Net}} \equiv \text{Conv}\left(\mathcal{C}^{\text{Net}}\right) = \text{Conv}\left(\mathcal{V}^{\text{Net}}\right) \tag{3.15}$$

where the deterministic behaviors correspond to vertices of the polytope and the $\Lambda$ subscript denotes the convex hull. On the other hand, when randomness is not shared amongst all devices the classical network set becomes nonconvex

$$\mathcal{C}^{\text{Net}} \subset \mathcal{C}^{\text{Net}}_{\Lambda}. \tag{3.16}$$

In many settings, shared randomness can be considered as a free resource because it could have been distributed prior to the experiment. Therefore, it is often convenient to assume that all devices are correlated and the set of classical network behaviors is convex. We will discuss the geometric structure of this set in Chapter 4.

# Chapter 4

# Classical Network Polytopes

The results in this dissertation build upon the geometric structure of convex polytopes defined on the spaces of black-behaviors. At a high-level, a convex polytope is a closed convex set that has a finite number of extreme points called vertices where flat faces connect the vertices to bound the set.

In general, we refer to convex polytopes on the set of black-box behaviors as *probability polytopes*, however, we make note of two important convex polytopes. In Chapter 2, we introduced the set of black-box behaviors, $\mathcal{P}_{\mathcal{Y}|\mathcal{X}}$ (see Def. 4), which is referred to as the *full probability polytope*. Furthermore, we will refer to as *classical network polytopes* the set $\mathcal{C}_{\Lambda}^{\text{Net}} \equiv \text{Conv}\left(\mathcal{C}^{\text{Net}}\right)$ where $\mathcal{C}^{\text{Net}}$ is the set of behaviors for a general classical network (see Def. 16) and the convex hull $\text{Conv}\left(\cdot\right)$ means that the devices have an unlimited amount of shared randomness at their disposal. Note that $\mathcal{C}_{\Lambda}^{\text{Net}} \subseteq \mathcal{P}_{\mathcal{Y}|\mathcal{X}}$.

In this chapter, we will provide the necessary background for understanding the structure of classical network polytopes. Given their geometry, we point out key features that are important for understanding the methods and results of this work. For an introduction to polyhedral theory, please see Ref. [118].

## *4.1 The Vertex Representation*

A convex polytope $\mathcal{P}$ is expressed in the *vertex representation* as its set of extreme points or vertices $\mathcal{V}$. The polytope is then defined as the convex hull of the set of vertices $\mathcal{P} = \text{Conv}\left(\mathcal{V}\right)$ [118]. We now define classical network polytopes in the vertex representation because it admits a simple representation and the vertices are useful for deriving other useful properties of the geometric structure.

**Definition 17. Vertex Representation of a Classical Network Polytope:** $\mathcal{C}_{\Lambda}^{\text{Net}}$, Let $\mathcal{V}^{\text{Net}}$ denote the set vertices, or deterministic behaviors, that can be achieved by a classical communication network $\mathcal{C}^{\text{Net}}$. The classical network polytope is then defined

$$\mathcal{C}_{\Lambda}^{\text{Net}} \equiv \text{Conv}\left(\mathcal{V}^{\text{Net}}\right). \tag{4.1}$$

The vertices of probability polytopes correspond to deterministic behavior matrices that have 0/1 elements (see Def. 5). Since the full probability polytope's vertices, $\mathcal{V}_{\mathcal{Y}|\mathcal{X}}$, are deterministic behaviors, it follows that

$$\mathcal{V}^{\text{Net}} \subseteq \mathcal{V}_{\mathcal{Y}|\mathcal{X}} \tag{4.2}$$

and $\mathcal{C}_{\Lambda}^{\text{Net}} \subseteq \mathcal{P}_{\mathcal{Y}|\mathcal{X}}$.

Since classical network polytope have deterministic vertices, they belong to a category of integral polytopes known as 0/1 polytopes where the vertices of 0/1 polytopes lie on the unit hypercube [125]. Furthermore, classical network polytopes can be viewed as a generalization of the correlation polytopes introduced by Itamar Pitowsky [126], which characterize the joint probabilities of binary events.

The set of vertices for classical network polytopes can be enumerated, however, the task can be difficult given the large number of vertices. Let's begin with the full probability polytope. The set of vertices $\mathcal{V}_{\mathcal{Y}|\mathcal{X}}$ contains

$$|\mathcal{V}_{\mathcal{Y}|\mathcal{X}}| = |\mathcal{Y}|^{|\mathcal{X}|} \tag{4.3}$$

unique deterministic behaviors where $\mathcal{V}_{\mathcal{Y}|\mathcal{X}}$ enumerates all possible ways to combine $|\mathcal{Y}|$ letters into a word of length $|\mathcal{X}|$. The exponential scaling of the number of vertices foreshadows the difficulties that will arise in performing computations on classical network polytopes.

Since classical networks also have deterministic behaviors as vertices, the set of vertices can be constructed by brute force. That is, deterministic network behaviors result from all devices performing deterministic behaviors. Thus, to construct the set $\mathcal{V}^{\text{Net}}$, the deterministic behaviors $\mathcal{V}_{\mathcal{Y}_{i,j}|\mathcal{X}_{i,j}}$ can first be enumerated for each device $\text{Dev}_{i,j}$. Then, using Def. 15 the individual deterministic device behaviors can be combined into a deterministic network behavior. If all combinations of device behaviors are taken, then $\mathcal{V}^{\text{Net}}$ is enumerated in full.

When the vertices $\mathcal{V}^{\text{Net}}$ are enumerated in this manner, they produce a set that is typically over complete, however, $\mathcal{V}^{\text{Net}}$ should only contain unique vertices. Hence we derive the upper bound on the number of classical network vertices to be

$$|\mathcal{V}^{\text{Net}}| \leq \prod_{i=1}^{|\vec{N}|} \prod_{j=1}^{|\vec{N}_j|} |\mathcal{Y}_{i,j}|^{|\mathcal{X}_{i,j}|} \tag{4.4}$$

where $\vec{N}$ denotes the set of nodes in the classical network's DAG. In general, more efficient algorithms can be used to enumerate the vertices without redundancy. Unfortunately, the number of vertices will usually still scale exponentially with number of inputs $|X_{i,j}|$.

## *4.2 Polytope Dimension*

Convex polytopes in two or three dimension form familiar shapes such as triangles, squares, and cubes. In this dissertation, the considered convex polytopes will require more than just three spatial dimensions to represent. Fortunately, the geometry of convex polytopes generalizes naturally and intuitively to higher dimensions.

**Definition 18. Polytope Dimension:** $\text{Dim}(\mathcal{P})$, The dimension of the Euclidean space required to represent the convex polytope $\mathcal{P}$.

A polytope's dimension $\text{Dim}(\mathcal{P})$ can be evaluated from the set of vertices $\mathcal{V}$ in the following steps.

1. Consider the vectorization $|\mathbf{V}\rangle$ of a deterministic behavior $\mathbf{V} \in \mathcal{V}$ where

$$|\mathbf{V}\rangle = \sum_{x \in \mathcal{X}} \sum_{y=1}^{|\mathcal{Y}|-1} P(y|x)\, |x\rangle \otimes |y\rangle \iff \sum_{x,y} P(y|x)\, |y\rangle\langle x| = \mathbf{V} \tag{4.5}$$

where the last value of each column is removed to account for the $|\mathcal{X}|$ normalization constraints on the behavior $\mathbf{V}$.

2. Construct the matrix $M \in \mathbb{R}^{|\mathcal{Y}\times\mathcal{X}|\times|\mathcal{V}|-1}$ as the concatenation of column vectors,

$$M = \sum_{i=1}^{|\mathcal{Y}\times\mathcal{X}|} \sum_{\mathbf{V}\in\mathcal{V}/\mathbf{V}_0} |\mathbf{V}\rangle\langle i| - |\mathbf{V}_0\rangle\langle i| \tag{4.6}$$

where $\mathbf{V}_0 \in \mathcal{V}$ serves as the coordinates of the origin.

3. The dimension of the convex polytope is $\mathrm{Dim}\,(\mathcal{P}) = \mathrm{Rank}\,(M)$.

**Lemma 5.** The dimension of the full probability polytope is $\mathrm{Dim}\left(\mathcal{P}_{\mathcal{Y}|\mathcal{X}}\right) = |\mathcal{X}|\big(|\mathcal{Y}| - 1\big)$.

*Proof.* A behavior matrix $\mathbf{P} \in \mathcal{P}_{\mathcal{Y}|\mathcal{X}}$ is isomorphic to a vector $|\mathbf{P}\rangle$ on euclidean space of dimension $|\mathcal{Y}\times\mathcal{X}|$. The behavior matrix is subject to a normalization constraint $\sum_{y\in\mathcal{Y}} P(y|x) = 1$ for each $x \in \mathcal{X}$. Since there are $n$ normalization constraints, $\dim \mathcal{P}_{\mathcal{Y}|\mathcal{X}} = |\mathcal{Y}\times\mathcal{X}| - |\mathcal{X}| = |\mathcal{X}|\big(|\mathcal{Y}| - 1\big)$ where $|\mathcal{Y}\times\mathcal{X}| = |\mathcal{Y}||\mathcal{X}|$. □

As another example, consider a network of non-signaling devices $\mathrm{Net}(\vec{N}^A, \vec{E})$ where $|\vec{N}^A| = n$, and $|\vec{E}| = 0$. The corresponding classical network polytope has dimension [127]

$$\mathrm{Dim}\left(\mathcal{C}_\Lambda^{\mathrm{Net}}\right) = \prod_{i=1}^{n} \left(1 + |\mathcal{X}_i|(|\mathcal{Y}_i| - 1)\right) - 1 \tag{4.7}$$

due to normalization and the general nonsignaling constraints in Eq. (3.5).

## *4.3 The Half-Space Representation*

A convex polytope can alternatively be expressed in *half-space representation* as the intersection of a finite set of linear half-space inequalities $\mathcal{G}$, referred to as facet inequalities [118]. Each facet inequality in $\mathcal{G}$ designates a tight bound on the convex polytope $\mathcal{P}$. In the case of the full probability polytope $\mathcal{P}_{\mathcal{Y}|\mathcal{X}}$, the set of linear half-space inequalities $\mathcal{G}_{\mathcal{Y}|\mathcal{X}}$ corresponds to the non-negativity constraints $P(y|x) \geq 0$.

Recall from Def. 7 that a linear black-box game represents a linear inequality on the probability space $\mathcal{P}_{\mathcal{Y}|\mathcal{X}}$. Here a linear inequality is expressed as a tuple $(\gamma, \mathbf{G})$ with $\mathbf{G} \in \mathbb{R}^{|\mathcal{Y}|\times|\mathcal{X}|}$ and $\gamma \in \mathbb{R}$ where the tuple describes the inequality $\gamma \geq \langle \mathbf{G}, \mathbf{P}\rangle = \sum_{x,y} G_{y,x} P(y|x) \leq \gamma$ for arbitrary behavior $\mathbf{P} \in \mathcal{P}_{\mathcal{Y}|\mathcal{X}}$. For convenience, we introduce two important subspaces of probability polytopes.

**Definition 19. Probability Half-Space:** $\mathcal{H}(\gamma, \mathbf{G})$, A subspace of the full probability polytope $\mathcal{P}_{\mathcal{Y}|\mathcal{X}}$ where

$$\mathcal{H}(\gamma, \mathbf{G}) \equiv \left\{ \mathbf{P} \in \mathcal{P}_{\mathcal{Y}|\mathcal{X}} \;\middle|\; \gamma \geq \langle \mathbf{G}, \mathbf{P}\rangle \quad \text{and} \quad \mathrm{Dim}\left(\mathcal{H}(\gamma, \mathbf{G})\right) = \mathrm{Dim}\left(\mathcal{P}_{\mathcal{Y}|\mathcal{X}}\right) \right\}. \tag{4.8}$$

**Definition 20. Probability Hyper-Plane:** $\mathcal{F}(\gamma, \mathbf{G})$, a subspace of the full probability polytope $\mathcal{P}_{\mathcal{Y}|\mathcal{X}}$ where

$$\mathcal{F}(\gamma, \mathbf{G}) \equiv \left\{ \mathbf{P} \in \mathcal{P}_{\mathcal{Y}|\mathcal{X}} \;\middle|\; \gamma = \langle \mathbf{G}, \mathbf{P}\rangle \quad \text{and} \quad \mathrm{Dim}\left(\mathcal{F}(\gamma, \mathbf{G})\right) = \mathrm{Dim}\left(\mathcal{P}_{\mathcal{Y}|\mathcal{X}}\right) - 1 \right\}. \tag{4.9}$$

**Definition 21. Half-Space Representation of a Probability Polytope:** A probability polytope $\mathcal{P} \subset \mathcal{P}_{\mathcal{Y}|\mathcal{X}}$ is expressed in the half-space representation as the intersection of a finite set of linear half-spaces $\mathcal{G}^{\text{Net}}$

$$\mathcal{P} = \bigcap_{i=1}^{|\mathcal{G}^{\text{Net}}|} \mathcal{H}(\gamma_i, \mathbf{G}_i) \tag{4.10}$$

where each inequality $\gamma_i \geq \langle \mathbf{G}_i, \mathbf{P} \rangle$ is a *facet inequality*, meaning that the inequality is both a proper half-space (see Def. 19) and that it tightly bounds the probability polytope $\mathcal{P}$ and belongs to the set $\mathcal{G}^{\text{Net}}$.

**Lemma 6.** A linear inequality $(\gamma, \mathbf{G})$ belongs to the set facet inequalities $\mathcal{G}$ for the probability polytope $\mathcal{P}$ if and only if:

1. $\mathcal{P} \subseteq \mathcal{H}(\gamma, \mathbf{G})$;
2. $\text{Dim}\left(\mathcal{P} \cap \mathcal{F}(\gamma, \mathbf{G})\right) = \text{Dim}\left(\mathcal{P}\right) - 1$.

*Proof.* Condition 1 must hold because the half-space inequality bounds the polytope $\mathcal{P}$. Condition 2 enforces that $\mathcal{F}(\gamma, \mathbf{G})$ is a proper half-space that tighly bounds the probability polytope $\mathcal{P}$. □

An important consequence of Condition 2, Lemma 6, is that a linear inequality can be verified to be a member of $\mathcal{G}$ if the equality holds

$$\text{Dim}\left(\{\mathbf{V} \in \mathcal{V} \,|\gamma = \langle \mathbf{G}, \mathbf{V} \rangle\}\right) = \text{Dim}\left(\mathcal{P}\right) - 1. \tag{4.11}$$

That is, the upper bound for each facet inequality bounding polytope $\mathcal{P}$ is achieved by exactly $\text{Dim}\left(\mathcal{P}\right)$ affinely independent vertices.

A general classical network polytope facet inequality $(\gamma, \mathbf{G}) \in \mathcal{G}^{\text{Net}}$ does not have a unique form. It is therefore convenient to establish a *lexicographic normal form* for a given facet inequality [128]. Using the following observations, a normal form can be established.

1. Multiplying a linear inequality $(\gamma, \mathbf{G})$ by a scalar $a \in \mathbb{R}$ does not alter the half-space contained by the linear inequality, that is, $\mathcal{H}(\gamma, \mathbf{G}) = \mathcal{H}(a(\gamma), a\mathbf{G})$.

2. When the set of vertices $\mathcal{V}$ contains only 0/1 matrices, all coefficients of the game matrix $\mathbf{G}$ are rational [118], [129]. Therefore, there exists a rational scalar $a$ such that $aG_{y,x}$ and $a\gamma$ are integers for all $x \in \mathcal{X}$ and $y \in \mathcal{Y}$.

3. The normalization and non-negativity constraints on black-box behaviors implies equivalence between the following two inequalities

$$\gamma \geq \langle \mathbf{G}, \mathbf{P} \rangle \iff \gamma + 1 \geq \langle \mathbf{G}, \mathbf{P} \rangle + \sum_{y \in \mathcal{Y}} G_{y,x'} P(y|x') \tag{4.12}$$

   for any $x' \in \mathcal{X}$. Therefore, it is always possible to find a form of the linear inequality $(\gamma, \mathbf{G})$ where $\gamma \geq 0$ and $G_{y,x} \geq 0$ for all $y \in \mathcal{Y}$ and $x \in \mathcal{X}$.

**Definition 22. Normal Form:** A facet inequality $(\gamma, \mathbf{G}) \in \mathcal{G}^{\text{Net}}$ is expressed in its unique normal form when the following two conditions are met:

1. The inequality is expressed as $\gamma \geq \langle \mathbf{G}, \mathbf{P} \rangle$, otherwise the inequality is scaled by $-1$ to reverse the inequality.

2. The smallest value in each column of matrix $\mathbf{G}$ is 0, otherwise the smallest value is made to be zero by adding or subtracting normalization constraints from the column as in Eq. (4.12).

3. For inequality $(\gamma, \mathbf{G})$, $\gamma$ and all $G_{y,x}$ are integers with a greatest common factor of 1, otherwise the inequality is scaled by the appropriate rational number.

One advantage of defining the normal form for each facet inequality is that it admits a lexicographic ordering of inequalities. This means that if two game matrices are vectorized as $|\mathbf{G}_1\rangle$ and $|\mathbf{G}_2\rangle$, then they can each be assigned a ranking and sorted. Interpreting the vector $|\mathbf{G}\rangle = G_{y,x} |x\rangle \otimes |y\rangle$ as an integer with $|\mathcal{Y} \times \mathcal{X}|$ digits, then the vector $|\mathbf{G}\rangle$ that corresponds to a larger integer has a larger rank. For example, let

$$|\mathbf{G}_1\rangle = (1, 0, 3, 3)^T \quad \text{and} \quad |\mathbf{G}_2\rangle = (1, 2, 0, 4)^T, \tag{4.13}$$

then $|\mathbf{G}_2\rangle > |\mathbf{G}_1\rangle$ because the first digit is the same in each case, but the second digit is larger for $|\mathbf{G}_2\rangle$. Note that two vectors will only have the same rank if they have the same normal form for their game matrices.

## 4.4 Facet Liftings

The facet inequalities that bound a classical network polytope $\mathcal{C}_\Lambda^{\text{Net}}$ can be lifted to the polytopes of similar networks with more inputs, outputs, or parties [127], [129]. The advantage of facet liftings is that known facet inequalities can be generalized to networks of arbitrary scale.

A facet inequality $\gamma \geq \langle \mathbf{G}, \mathbf{P} \rangle$ is *input lifted* to a new facet $\gamma \geq \langle \mathbf{G}'', \mathbf{P} \rangle$ if $\mathbf{G}'' \in \mathbb{R}^{|\mathcal{Y}| \times |\mathcal{X}'|}$ is obtained from $\mathbf{G} \in \mathbb{R}^{|\mathcal{Y}| \times |\mathcal{X}|}$ by padding it with $(|\mathcal{X}'| - |\mathcal{X}|)$ all-zero columns where $|\mathcal{X}'| > |\mathcal{X}|$. Likewise, a facet inequality is *output lifted* to $\gamma \geq \langle \mathbf{G}', \mathbf{P} \rangle$ if $\mathbf{G}' \in \mathbb{R}^{|\mathcal{Y}'| \times |\mathcal{X}|}$ is obtained from $\mathbf{G} \in \mathbb{R}^{|\mathcal{Y}| \times |\mathcal{X}|}$ by copying rows; *i.e.*, there exists a surjective function $f : \mathcal{Y}' \to \mathcal{Y}$ such that $G'_{y',x} = G_{f(y'),x}$ for all $y' \in \mathcal{Y}'$ and $x \in \mathcal{X}$ where $|\mathcal{Y}'| > |\mathcal{Y}|$. As an example, we demonstrate input and output liftings on the identity matrix $\mathbf{G} = \mathbb{I}_3$,

$$\mathbf{G} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \xrightarrow[\text{Lifting}]{\text{Input}} \mathbf{G}'' = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}, \tag{4.14}$$

$$\mathbf{G} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \xrightarrow[\text{Lifting}]{\text{Output}} \mathbf{G}' = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}. \tag{4.15}$$

It is important to note that lifting a facet inequality does not alter its upper bound.

## 4.5 Symmetries and Generators

As shown in Eq. (4.3) and Eq. (4.4), the number of vertices for the classical network polytopes can grow exponentially with the number of inputs. Generally, the number of facet inequalities $|\mathcal{G}^{\text{Net}}|$ that bound a classical network polytope $\mathcal{C}_\Lambda^{\text{Net}}$ are significantly larger than the set of vertices, $|\mathcal{G}^{\text{Net}}| \gg |\mathcal{V}^{\text{Net}}|$. Therefore, it

can be a significant challenge to perform calculations upon the set of facet inequalities or vertices. Fortunately, probability polytopes offer a great deal of symmetry, which can be exploited to simplify their complex geometric structure.

The input and output values $x\mathcal{X}$ and $y \in \mathcal{Y}$ are merely labels for a black-box's behavior $\mathbf{P} \in \mathcal{P}_{\mathcal{Y}|\mathcal{X}}$. Therefore, the full probability polytope $\mathcal{P}_{\mathcal{Y}|\mathcal{X}}$ is invariant under swapping the labels $x \leftrightarrow x'$ and $y \leftrightarrow y'$ where $x, x' \in \mathcal{X}$ and $y, y' \in \mathcal{Y}$ **Rosset2014**. The relabeling operation corresponds to permutations on the input and output.

**Definition 23. Set of Permutations:** Let $\pi^{\mathcal{X}} \in \mathcal{S}^{\mathcal{X}}$ represent a permutation as a doubly stochastic matrix having elements $\pi^{\mathcal{X}}_{i,j} \in \mathbb{B}$. The set of permutations is then

$$\mathcal{S}^{\mathcal{X}} \equiv \left\{ \pi \in \mathbb{R}^{|\mathcal{X}|\times|\mathcal{X}|} \;\middle|\; \sum_{i\in\mathcal{X}} \pi_{i,j} = \sum_{j\in\mathcal{X}} \pi_{i,j} = 1 \quad \text{and} \quad \pi_{i,j} \in \mathbb{B} \;\forall\; i,j \in \mathcal{X} \right\}. \tag{4.16}$$

A black-box behavior's inputs and outputs can be permuted as

$$\mathbf{P'} = \pi^{\mathcal{Y}}\mathbf{P}\pi^{\mathcal{X}} \quad \text{where} \quad \mathbf{P}, \mathbf{P'} \in \mathcal{P}_{\mathcal{Y}|\mathcal{X}}, \quad \pi^{\mathcal{Y}} \in \mathcal{S}^{\mathcal{Y}}, \quad \text{and} \quad \pi^{\mathcal{X}} \in \mathcal{S}^{\mathcal{X}}. \tag{4.17}$$

Likewise, a facet inequality $(\gamma, \mathbf{G}) \in \mathcal{G}_{\mathcal{Y}|\mathcal{X}}$ can be permuted into a new facet inequality $(\gamma, \mathbf{G}') \in \mathcal{G}_{\mathcal{Y}|\mathcal{X}}$ where $\mathbf{G}' = \pi^{\mathcal{Y}}\mathbf{G}\pi^{\mathcal{X}}$. As we will see later, the probability polytopes $\mathcal{P} \subseteq \mathcal{P}_{\mathcal{Y}|\mathcal{X}}$ that we consider will also be invariant under certain permutations. Given this permutation symmetric structure, we can construct broad classes of vertices and facets that greatly simplify the description and computation of probability polytopes.

**Definition 24. Vertex Class:** A set of probability polytope vertices that can be transformed by permutation into a canonical vertex $\mathbf{V}^\star$, which is referred to as a *generator vertex*. The set corresponding to the vertex class is constructed as

$$\left\{ \mathbf{V} \in \mathcal{V} \;\middle|\; \mathbf{V}^\star\pi^{\mathcal{Y}}\mathbf{V}\pi^{\mathcal{X}} \quad \forall \quad \pi^{\mathcal{X}} \in \mathcal{S}^{\mathcal{X}} \quad \text{and} \quad \pi^{\mathcal{Y}} \in \mathcal{S}^{\mathcal{Y}} \right\}. \tag{4.18}$$

**Definition 25. Facet Class:** A set of probability polytope facets that can be transformed by permutation into a canonical facet $(\gamma, \mathbf{G}^\star)$, which is referred to as a *generator facet*. The set corresponding to the facet class is constructed as

$$\left\{ (\gamma, \mathbf{G}) \in \mathcal{G} \;\middle|\; \mathbf{G}^\star\pi^{\mathcal{Y}}\mathbf{G}\pi^{\mathcal{X}} \quad \forall \quad \pi^{\mathcal{X}} \in \mathcal{S}^{\mathcal{X}} \quad \text{and} \quad \pi^{\mathcal{Y}} \in \mathcal{S}^{\mathcal{Y}} \right\}. \tag{4.19}$$

In both cases, the generator vertex or facet is arbitrary where it may be convenient to use the vertex or facet whose lexicographic rank is either maximum or minimum. Furthermore, when considering general probability polytopes, it may be the case that the sets of permutations may be restricted in some way, however, the facet and vertex classes can each be defined with respect to any permutation operations.

**Definition 26. Set of Generator Vertices:** $\mathcal{V}^\star$, The set of generator vertices $\mathbf{V}^\star \in \mathcal{V}$ that each represent a distinct vertex class.

**Definition 27. Set of Generator Facets:** $\mathcal{G}^\star$, The set of generator facets $(\gamma, \mathbf{G}^\star) \in \mathcal{G}$ that each represent a distinct facet class.

Since the number of input and output permutations scale as factorials of $|\mathcal{X}|$ and $|\mathcal{Y}|$ the sets of generators

dramatically smaller than the full set of vertices or facets. That is,

$$|\mathcal{V}^\star| \ll |\mathcal{V}| \quad \text{and} \quad |\mathcal{G}^\star| \ll |\mathcal{G}| \tag{4.20}$$

where the complete set of vertices or facets can always be recovered by taking permutations of the generator. Thus, it is important to exploit the relabeling symmetries of network polytope [129].

## *4.6 Transforming Polytope Representations*

Convex polytopes can be equivalently expressed in either the vertex or half-space representation, however, the choice of representation largely depends on how the polytope is defined. Since each representation has its own advantages, it is important to be able to transform a convex polytope from one representation to the other

$$\mathcal{V} \Longleftrightarrow \mathcal{G}. \tag{4.21}$$

In practice, it is easy to enumerate the vertices of classical network polytopes, but challenging to determine the polytope's facets. Thus the main challenge is to convert a set of vertices into their bounding facets.

### *4.6.1 Deriving Facet Inequalities via Linear Programming*

Linear programming is perhaps the quickest way to derive a facet inequality $(\gamma, \mathbf{G}) \in \mathcal{G}^{\text{Net}}$ from the set of vertices $\mathcal{V}^{\text{Net}}$ for a classical network polytope $\mathcal{C}_\Lambda^{\text{Net}}$. For a formal introduction to linear programming please refer to Boyd and Vandenberghe's textbook on convex optimization [92]. Linear programs are typically solved using software. In this work, we use the *High Performance Software for Linear Optimization* (HiGHS) optimizer [130] via the Julia Mathematical Programming (JuMP.jl) interface [131] exposed through the Julia Programming Language [132].

In general, a linear program solves an optimization problem with a linear objective function subject to linear constraints. Linear programs can be solved efficiently using algorithms that have a time complexity that scales polynomially with the number of variables. Due to its efficiency linear programming is a powerful tool when it can be applied.

A linear program that derives facets of a classical network polytope is written [109], [133], [134]

$$\max_{(\gamma,\mathbf{G})} \langle \mathbf{G}, \mathbf{P} \rangle - \gamma \tag{4.22}$$

$$\text{s.t.} \ \langle \mathbf{G}, \mathbf{V} \rangle - \gamma \leq 0 \quad \forall \quad \mathbf{V} \in \mathcal{V}^{\text{Net}} \tag{4.23}$$

$$\langle \mathbf{G}, \mathbf{P} \rangle - \gamma \leq 1 \tag{4.24}$$

where the linear program requires the complete set of vertices $\mathcal{V}^{\text{Net}}$ and a black-box behavior $\mathbf{P} \in \mathcal{P}_{\mathcal{Y}|\mathcal{X}}$. If $\mathbf{P} \in \mathcal{C}_\Lambda^{\text{Net}}$, then the objective evaluates to $0 \geq \langle \mathbf{G}, \mathbf{P} \rangle - \gamma$, otherwise, the objective evaluates to $1 = \langle \mathbf{G}, \mathbf{P} \rangle - \gamma$, which implies that $\mathbf{P} \notin \mathcal{C}_\Lambda^{\text{Net}}$. In both cases, the optimal inequality $(\gamma^\star, \mathbf{G}^\star)$ will designate a linear bound on the polytope. When $\mathbf{P} \notin \mathcal{C}_\Lambda^{\text{Net}}$ the bound corresponds to a linear inequality that separates $\mathbf{P}$ from $\mathcal{C}_\Lambda^{\text{Net}}$. Although, care should be taken to ensure that the produced inequality corresponds to a proper facet as described in Lemma 6.

The advantage of this technique for deriving polytope facets is its performance that scales polynomially with the problem size. Furthermore, no knowledge about the facet structure is needed, only the set of vertices

$\mathcal{V}^{\text{Net}}$ and a test point $\mathbf{P}$. However, a drawback of the linear programming approach is that the set vertices grows exponentially with the input, which will eventually render the linear program as inefficient an unable to solve the problem at hand. Additionally, the linear program only derives a single example facet. Other methods are needed to obtain the complete set of facets or generator facets in a guaranteed manner.

### *4.6.2 Full Polytope Transformation via Fourier-Motzkin Elimination*

The complete set of facets can be obtained from the vertices using an algorithm based on Fourier-Motzkin elimination [118]. In this work, we apply the Polyhedron Representation Transformation Algorithm (PORTA) [135]. This polyhedral analysis software is especially convenient for 0/1 polytopes because the algorithm performs computations on rational numbers rather than floats. Hence the exact values for vertices and facets can be computed without significant risk of numerical error.

A drawback of the PORTA software is that it is not parallelized. This means that it cannot be run in a distributed manner. In order to scale to larger polytopes parallelism is needed due to the computational complexity of polytope transformation algorithms. Nevertheless, the algorithm is still useful for rigorously exploring many interesting examples.

A second drawback of the PORTA software is that it is no longer maintained by the authors. As part of this work, we have developed a wrapper *XPORTA.jl* [136] that integrates PORTA software with the Julia programming language [132] through the *Polyhedra.jl* package interface [137]. The XPORTA.jl software significantly improves the ease of using PORTA by making it executable by code written in Julia. The XPORTA.jl wrapper also tests core functionality of the PORTA software to ensure that it still performs as expected on modern operating systems.

### *4.6.3 Deriving Generator Facets via Adjacency Decomposition*

The set of facets $\mathcal{G}^{\text{Net}}$ for a classical network polytope $\mathcal{C}_{\Lambda}^{\text{Net}}$ is completely described by its set of generators $(\mathcal{G}^{\text{Net}})^{\star}$ and its input and output symmetry group $\mathcal{S}^{\mathcal{X}}(\mathcal{C}_{\Lambda}^{\text{Net}})$ and $\mathcal{S}^{\mathcal{Y}}(\mathcal{C}_{\Lambda}^{\text{Net}})$. Therefore it is wasteful to perform a complete polytope representation transformation because only the generators $(\mathcal{G}^{\text{Net}})^{\star} \subseteq \mathcal{G}^{\text{Net}}$ are needed. This task can be achieved due to some important consequences of the polytope's permutation symmetry. First, we define facet adjacency.

**Definition 28. Facet Adjacency:** Two facets $(\gamma_1, \mathbf{G}_1), (\gamma_2, \mathbf{G}_2) \in \mathcal{G}$ are adjacent if and only if they share a *ridge* $\mathcal{R}$ defined as:

1. $\mathcal{R} := \mathcal{F}(\gamma_1, \mathbf{G}_1) \cap \mathcal{F}(\gamma_2, \mathbf{G}_2) \cap \mathcal{P}$,
2. where $\dim(\mathcal{R}) = \dim(\mathcal{P}) - 2$.

A ridge can be understood as a facet of the facet polytope $\mathcal{F}(\gamma, \mathbf{G})$. To compute the ridges of a given facet $(\gamma, \mathbf{G}) \in \mathcal{G}$ we can use the standard Fourier-Motzkin elimination procedure for computing facets [118], [135]. A facet adjacent to $(\gamma, \mathbf{G})$ can then computed from the ridge using a *rotation algorithm* described by Christof and Reinelt (see Algorithm 2 of Ref. [128]). Furthermore, the ridge shared between two facets is unique to those facets. Then, consider two adjacent facet polytopes $\mathcal{F}_1, \mathcal{F}_2 \in \mathcal{G}$. Given the permutation symmetry of the probability polytope, all permutations of $\mathcal{F}_1$ must lie adjacent to a similarly permuted $\mathcal{F}_2$. This means that the complete set of generator facets lie in a connected neighborhood with each other. The neighborhood represents a tiling pattern that covers the probability polytope. By this symmetry, any representative of a

given facet class has the same fixed set of facet classes adjacent to it and there cannot be two disjoint sets of generator facets where the members of one set do not lie adjacent to the members of the other.

**Definition 29. Adjacency Decomposition Algorithm:** [128] An iterative algorithm which requires as input the probability polytope vertices $\mathcal{V}$, a generator facet $(\gamma_{\text{init}}, \mathbf{G}^{\star}_{\text{init}}) \in \mathcal{G}^{\star}$, and the permutation symmetry groups for $\mathcal{S}^{\mathcal{X}}$ and $\mathcal{S}^{\mathcal{Y}}$. During operation, the algorithm maintains a list of generator facets $\mathcal{L}$ where each generator facet $(\gamma, \mathbf{G}^{\star}) \in \mathcal{L}G$ is marked either as *considered* or *unconsidered*. Upon input, the algorithm adds the input facet $(\gamma_{\text{init}}, \mathbf{G}^{\star}_{\text{init}})$ to $\mathcal{L}$ and marks it as *unconsidered*. In each iteration, the algorithm proceeds as follows:

1. An *unconsidered* generator facet $(\gamma, \mathbf{G}^{\star}) \in \mathcal{L}$ is selected.
2. All facets adjacent to $(\gamma, \mathbf{G}^{\star})$ are computed.
3. Each adjacent facet is converted into its lexicographic normal form.
4. Any new generator facets identified are marked as *unconsidered* and added to $\mathcal{L}$.
5. Facet $(\gamma, \mathbf{G}^{\star})$ is marked as *considered*.

The algorithm completes when all generator facets in $\mathcal{L}$ are marked as *considered* upon completion $\mathcal{L} = \mathcal{G}^{\star}$ is guaranteed to hold.

The adjacency decomposition algorithm provides a few key computational advantages over the PORTA software and Fourier-Motzkin elimination techniques:

1. The adjacency decomposition algorithm requires considerably less memory because the algorithm stores only the generator facets $\mathcal{G}^{\star}$ instead of the complete set of facets $\mathcal{G}$.
2. A new generator facet is guaranteed to be output in each iteration. That is, the algorithm computes facets with less redundancy and does not need to run to completion to obtain new generator facets.
3. The algorithm can be widely parallelized because each generator facet can be considered independently and the search for adjacent facets can be parallelized [128].

Our developed adjacency decomposition algorithm is implemented in the *BellScenario.jl* Julia package [138].

# Chapter 5

# Quantum Fundamentals

Quantum physics describes the dynamics of systems that are incredibly small. At its heart, quantum mechanics is a phenomenological theory of the natural world, meaning that our understanding of quantum systems is inferred from the data we measure from them. Although quantum mechanics accurately predicts the behavior of physical processes on average, quantum physics fails to describe the inner workings of quantum systems with certainty. This paradox has confounded physicists and philosophers for over a century. Indeed, it is a complete mystery as to why quantum mechanics behaves in the way it does, and any model that describes its inner workings is metaphysical at best. Nevertheless, quantum physics can be leveraged as an important tool for understanding the natural world.

The study of quantum mechanics dates back over a century to when researchers began to observe that the natural world is not only *quantized* into discrete energy packets, but that physical systems also exhibit a *wave-particle duality* [139]–[142]. These theoretical results were subsequently demonstrated experimentally indicating that both matter and light exhibit a wave-particle duality nature. In 1930, Paul Dirac used these established results to formulate the axiomatic approach to quantum theory that we continue to use nearly a century later [143], while similar results were given by Von Neumann in 1932 [144].

At the highest level, the axiomatic approach to quantum mechanics breaks the dynamics of quantum systems into three stages: preparation, evolution, and measurement. In the preparation stage, a quantum state $|\psi\rangle$ is represented by a complex-valued vector having norm one. In the evolution stage, the prepared state is transformed into a new state vector $|\psi'\rangle = U|\psi\rangle$ where $U$ is a unitary operator. In the measurement stage, the evolved quantum state is projected onto an orthonormal basis $\{|\phi_x\rangle\}_{x=1}^{d}$ to obtain a classical result $x \in \{0, \ldots, d-1\}$ with probability $P(x) = |\langle\phi_x|U|\psi\rangle|^2$. A familiar classical example to which a quantum system can be compared is a coin flip where, with some probability, we expect the coin to show either the heads or tails orientation. However, we will never observe both heads and tails at the same time. Quantum systems are similar in that we will only observe one outcome at a time, but the ways quantum coins flip in nature seems defy our classical understanding of physics.

In the following sections, we provide a more detailed description of the quantum formalism. The curious reader can find a more detailed approach to quantum theory in references [61], [145]. Furthermore, we've implemented the described quantum mechanical framework as an open-source software package "QBase.jl: A base library for quantum information" [146]. This software is written in the Julia programming language [132] and is used in numerical calculation and investigation of quantum systems.

## 5.1 Representing Quantum Systems

To discuss quantum mechanics with any precision, we must first elaborate upon the mathematical objects that describe quantum mechanics. At its root, quantum mechanics applies physical meaning to linear algebra and its familiar matrices and vectors. When representing quantum systems, matrix-like objects are referred to as operators while vector-like objects are referred to as either a *bra* or a *ket* depending upon whether it is a column or row vector. The bras, kets, and operators are then combined algebraically to describe the dynamics of quantum systems. We now define each of these components and the mathematical operations under which they can be combined.

### 5.1.1 Bra-Ket Notation

**Definition 30. Ket:** a column-vector on a complex-valued, $d$-dimension vector space $\mathbb{C}^d$

$$|\psi\rangle \equiv \sum_{i=0}^{d-1} a_i |i\rangle = (a_0, \ldots, a_{d-1})^T. \tag{5.1}$$

**Definition 31. Bra:** a row-vector on a complex-valued, $d$-dimension vector space $(\mathbb{C}^d)^\dagger$

$$\langle\psi| \equiv \sum_{i=0}^{d-1} a_i^* \langle i| = (a_0^*, \ldots, a_{d-1}^*) \tag{5.2}$$

where $a_i^*$ denotes the complex conjugate of $a_i$.

Bras and kets can be expressed with respect to any orthonormal basis $\{|\phi_k\rangle \in \mathbb{C}^d\}_{k=1}^d$ provided that it spans $\mathbb{C}^d$. To express bras and kets, we use the conventional *computational basis*

$$\{|0\rangle, |1\rangle, \ldots, |d-1\rangle\} \equiv \{(1,0,\ldots,0)^T, (0,1,\ldots,0)^T, \ldots, (0,0,\ldots,1)^T\}. \tag{5.3}$$

Generally, the dimension of bras and kets is unrestricted. In many quantum mechanical cases, an infinite dimensional complex-valued vector space is needed, in which the corresponding vector space is regarded as a *Hilbert Space* $\mathcal{H}$. This dissertation will focus mostly on finite dimensional vector spaces.

The vector spaces of bras and kets are dual to each other in a one-to-one correspondence. For every ket $|\psi\rangle \in \mathbb{C}^d$ there exists a bra $\langle\psi| \in (\mathbb{C}^d)^\dagger$, and *vice versa*. Bras and kets can each be transformed into their corresponding dual via the conjugate transpose , $\langle\psi| = |\psi\rangle^\dagger$ and $\langle\psi|^\dagger = |\psi\rangle$.

As dual vector spaces, an inner-product operation exists between bras and kets. For a bra $\langle\phi| = (b_1, \ldots, b_d)$ and a ket $|\psi\rangle = (a_1, \ldots a_d)^T$ the inner product is expressed as

$$\langle\phi|\psi\rangle = \sum_{i=1}^{d} b_i a_i, \tag{5.4}$$

which is a complex-valued scalar. Note that the naming of bras and kets comes from their inner-product because it forms a *bra(c)ket*. In quantum mechanics, bras and kets are typically normalized such that

$$1 = \langle\psi|\psi\rangle = \sum_{i=1}^{d} |a_i|^2 \quad \text{where} \quad |a_i|^2 = a_i^\dagger a_i. \tag{5.5}$$

Likewise, the *outer-product* between $\langle\phi|$ and ket $|\psi\rangle$ to produce a complex-valued $d \times d$ matrix

$$|\psi\rangle\langle\phi| = \sum_{i,j=1}^{d} a_i b_j \, |i\rangle\langle j| \,, \tag{5.6}$$

where $|\psi\rangle\langle\phi| \in \mathbb{C}^{d\times d}$ and the outer-product of computational basis elements indexes row $i$ and column $j$ of the resulting matrix.

### *5.1.2 Operators in Quantum Mechanics*

**Definition 32. Operator:** A linear operator or matrix $M$ defined on a complex-valued matrix space $\mathbb{C}^{d\times d'}$,

$$M = \sum_{i=1}^{d}\sum_{j=1}^{d'} M_{i,j} \, |i\rangle\langle j| = \sum_{i=1}^{d}\sum_{j=1}^{d'} \langle i|M|j\rangle \, |i\rangle\langle j| \tag{5.7}$$

where the subscript $i, j$ indexes the matrix element.

Since operators can be represented as matrices, they admit the same algebraic operations and properties as matrices. For example, an operator $M$ can be multiplied by Bras, kets, or both as

$$M\,|\psi\rangle = |\phi\rangle \,, \quad \langle\psi|\,M = \langle\phi| \,, \quad \text{and} \quad \langle\psi|M|\psi\rangle = \langle\psi|\phi\rangle \,. \tag{5.8}$$

The eigenvectors (*eigenkets*) of an operator then satisfy $M\,|\psi\rangle = \mu\,|\psi\rangle$, in which $\mu$ is a complex-valued scalar corresponding to the eigenvalue. Furthermore, the vector and matrix spaces of multiple bras, kets, or operators can be combined as tensor products. For example, the tensor of bras and kets produces bras and kets in a higher dimension,

$$\langle\phi| \otimes \langle\psi| = \langle\phi, \psi| \quad \text{and} \quad |\phi\rangle\,|\psi\rangle = |\phi, \psi\rangle \tag{5.9}$$

where if $|\phi\rangle\,, |\psi\rangle \in \mathbb{C}^d$, then $|\phi, \psi\rangle \in \mathbb{C}^{d^2}$. Likewise, the tensor product of operators $M, M' \in \mathbb{C}^{d\times d}$ produces the matrix

$$N = M \otimes M' \tag{5.10}$$

where $N \in \mathbb{C}^{d^2\times d^2}$. While the tensor product of matrices or vectors can always be expressed as a matrix or vector respectively, it can be convenient to regard these objects as tensors whose tensor rank increases as more subsystems are combined under the tensor product.

## *5.2 The State Space Axiom*

Quantum systems are understood to exist as a *state* within an abstract metaphysical state space. Whether quantum states correspond to a physical element in reality is a matter of interpretation. Nevertheless, quantum states can serve as an important model in predicting the behaviors of quantum systems.

**Definition 33. Quantum State:** An abstract model of a physical quantum system represented as a normalized $d$-dimension bra or ket

$$|\psi\rangle \in S(\mathbb{C}^d) \equiv \{|\phi\rangle \in \mathbb{C}^d \mid \langle\psi|\psi\rangle = 1\}. \tag{5.11}$$

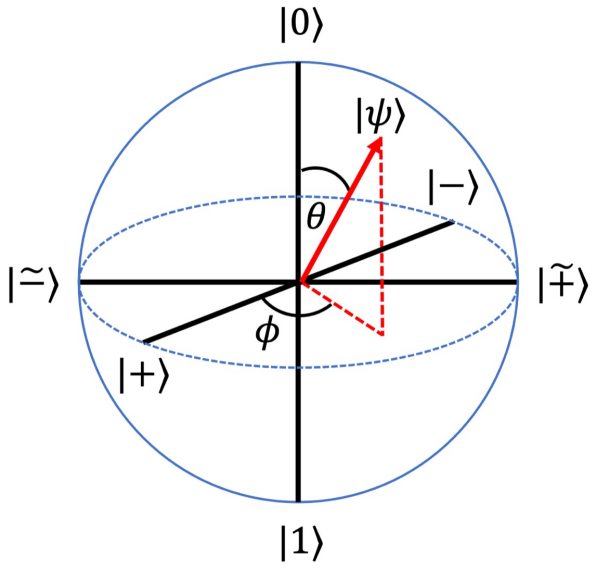



Figure 5.1: Bloch Sphere: The quantum state space of all qubit ($d = 2$) states can be mapped onto a sphere where the $z$-axis corresponds to the computational basis states $|0\rangle$ and $|1\rangle$, the $x$-axis corresponds to the basis states $|\pm\rangle = \frac{1}{\sqrt{2}}(|0\rangle \pm |1\rangle)$, and the $y$-axis corresponds to the basis states $|\widetilde{\pm}\rangle = \frac{1}{\sqrt{2}}(|0\rangle \pm i\,|1\rangle)$. A general qubit state is expressed as $|\psi\rangle = \cos(\frac{\theta}{2})\,|0\rangle + e^{i\phi}\sin(\frac{\theta}{2})\,|1\rangle$. Pure quantum states map to the surface of the sphere while mixed quantum states map to the interior where the maximally mixed state $\rho = \frac{1}{2}\mathbb{I}_2$ corresponds to the origin.

The dimension of the quantum state $|\psi\rangle$ corresponds to the fact that $d$ orthonormal basis vectors $\{|i\rangle\}_{i=0}^{d-1}$ are needed to span $\mathbb{C}^d$. As a result, an orthonormal basis must be complete

$$\mathbb{I}_d = \sum_{i=1}^{d} |i\rangle\langle i|\,. \tag{5.12}$$

We will refer to quantum states that have $d = 2$ as *qubits*, $d = 3$ as *qutrits*, and $d$-level systems as *qudits*. Quantum states can also be infinite dimensional, in which case the complex vector space $\mathbb{C}^d$ becomes a Hilbert space $\mathcal{H}$. This work mainly considers finite dimensional quantum systems hence we apply complex valued vector space $\mathbb{C}^d$.

The computational basis states $\{|i\rangle\}_{i=0}^{d-1}$ may be interpreted as the set of *classical states* for the quantum system. These classical basis states are orthonormal such that $\langle i|j\rangle = \delta_{i,j}$, meaning that they represent mutually exclusive states for the system.

Quantum states are distinct from classical states because they can exist in a superposition of basis states $|\psi\rangle = \sum_{i=0}^{d-1} a_i\,|i\rangle$. Whether a state is in a superposition or not is subjective and depends upon the orthonormal basis that is chosen by the researcher to represent the state. However, superposition allows for two quantum states, $|\psi\rangle = \sum_{i=0}^{d-1} a_i\,|i\rangle$ and $|\phi\rangle = \sum_{j=0}^{d-1} b_j\,|j\rangle$, to be nonorthogonal

$$0 \leq |\,\langle\phi|\psi\rangle\,| \leq 1. \tag{5.13}$$

Quantum states that are not orthogonal are said to overlap and can not be distinguished with certainty.

Quantum states may also be combined by tensor products into larger composite systems. For example,

consider the qubit states $|\psi\rangle, |\phi\rangle \in S(\mathbb{C}^2)$. The joint two-qubit state is then calculated as the tensor product

$$|\psi\rangle \otimes |\phi\rangle \, a_0 b_0 \, |00\rangle + a_0 b_1 \, |01\rangle + a_1 b_0 \, |10\rangle + a_1 b_1 \, |11\rangle \, . \tag{5.14}$$

where $|00\rangle, |01\rangle, |10\rangle, |11\rangle = |0\rangle, |1\rangle, |2\rangle, |3\rangle$ are the computational basis states expressed as binary numbers.

**Definition 34. Density Operator:** $\rho$, An abstract model of a physical quantum system represented as an operator

$$\rho \in D(\mathbb{C}^{d\times d}) \equiv \left\{ M \in \mathbb{C}^{d\times d} \;\middle|\; M = M^\dagger, \; M \geq 0, \; \mathrm{Tr}\left[M\right] = 1 \right\} \tag{5.15}$$

.

Any quantum state can be expressed as a density operator by taking the outer-product $\rho = |\psi\rangle\langle\psi|$. Such a density operator has $\mathrm{Rank}\left(\rho\right) = 1$ and is referred to as a pure state, otherwise the density operator is said to be a *mixed state*. In general, the purity of a quantum state can be quantified as $\mathrm{Tr}\left[\rho^2\right]$. Given a quantum state $\rho$ the purity $\mathrm{Tr}\left[\rho^2\right]$ quantifies how mixed the state is. Note that $\mathrm{Tr}\left[\rho^2\right] \leq \mathrm{Tr}\left[\rho\right] = 1$ and that for pure states $\mathrm{Tr}\left[\rho^2\right] = \mathrm{Tr}\left[\rho\right]$, meaning that their purity is 1. The smallest purity is achieved by the maximally mixed state $\rho_{\mathrm{mix}} = \frac{1}{d}\mathbb{I}_d$ and its purity is $\mathrm{Tr}\left[\rho_{\mathrm{mix}}\right] = \frac{1}{d}$. Thus, the purity of any density operator $\rho \in D(\mathbb{C}^{d\times d})$ is bounded as

$$\frac{1}{d} \leq \mathrm{Tr}\left[\rho^2\right] \leq 1 \tag{5.16}$$

The set of density operators is convex, $D(\mathbb{C}^{d\times d}) = \mathrm{Conv}\left(D(\mathbb{C}^{d\times d})\right)$ with pure state lying on the boundary (see Fig. 5.1). Therefore, any density operator $\rho$ may be expressed as a statistical mixture of pure states

$$\rho' = \sum_{i=1}^{n} \mu_i \, |\psi_i\rangle\langle\psi_i| \quad \text{where} \quad |\psi_i\rangle \in S(\mathbb{C}^d), \quad \mu_i \geq 0, \quad \text{and} \quad \sum_{i=1}^{n} \mu_i = 1. \tag{5.17}$$

These pure states $|\psi_i\rangle$ represent possible orientations of a quantum system and $\mu_i$ their bias. Thus mixed states can be interpreted as an experimenter's uncertainty of a system's quantum state.

## *5.3 The Time Evolution Axiom*

Quantum systems are dynamic and their states are constantly evolving in time. To formalize the time evolution of quantum systems we begin with closed quantum systems and extend the model to open quantum systems. A quantum system is closed when it does not interact with its environment. In this case, a quantum system's evolution is ideal and modeled using unitary operators. On the other hand, an open quantum system interacts with its environment. The coupling to the environment leads to noisy evolution of quantum states because information is lost to the environment. In open quantum systems, the observed evolution of the encoded quantum state is nonunitary and generally represented a completely-positive trace-preserving (CPTP) map.

### *5.3.1 The Unitary Dynamics of Closed Quantum Systems*

The time evolution of a state is defined according to the system's *Hamiltonian*, an Hermitian operator $H \in \mathbb{C}^{d\times d}$ that assigns energy to quantum states $|\psi\rangle \in S(\mathbb{C}^d)$. By solving the *Schrödinger equation* for the respective Hamiltonian, the time evolution of a quantum state $|\psi(t_1)\rangle \to |\psi(t_2)\rangle$ is calculated to be the

unitary operator $U = e^{-iHt}$ where $|\psi(t_2)\rangle = U|\psi(t_1)\rangle$ and $t_2 = t_1 + t$. In general, a unitary operator is defined as follows.

**Definition 35. Unitary Operator:** An operator $U \in \mathbb{C}^{d\times d}$ satisfying $U^\dagger = U^{-1}$ where $U^\dagger U = UU^\dagger = \mathbb{I}_d$.

The set of unitary operators form a group because it contains an identity element $\mathbb{I}_d$, an inverse element $U^\dagger = U^{-1}$, and is associative $(AB)C = A(BC)$ for unitaries $A, B, C$. Furthermore, consider two unitaries $U$ and $V$. Both their matrix product, $UV$, and their tensor product, $U \otimes V$, results in unitaries.

An important property of unitary evolution is that it preserves the normalization and purity of quantum states. This means that when a pure state $|\psi\rangle \in S(\mathbb{C}^d)$ is evolved by unitary $U$, we find the resulting state to be $U|\psi\rangle = |\phi\rangle \in S(\mathbb{C}^d)$ because

$$|\langle\phi|\phi\rangle| = |\langle\psi|U^\dagger U|\psi\rangle| = |\langle\psi|\psi\rangle| = 1. \tag{5.18}$$

Similarly, when a mixed state $\rho \in D(\mathbb{C}^{d\times d})$ is evolved by a unitary, the resulting state is $U\rho U^\dagger = \sigma \in D(\mathbb{C}^{d\times d})$ because

$$\mathrm{Tr}[\sigma] = \mathrm{Tr}\left[U\rho U^\dagger\right] = \mathrm{Tr}\left[\rho U^\dagger U\right] = \mathrm{Tr}[\rho] = 1 \tag{5.19}$$

where the cyclic property of the trace is applied. Under unitary evolution, we see that the purity remains constant because

$$\mathrm{Tr}\left[\sigma^2\right] = \mathrm{Tr}\left[U\rho U^\dagger U\rho U^\dagger\right] = \mathrm{Tr}\left[U\rho^2 U^\dagger\right] = \mathrm{Tr}\left[\rho^2\right] \tag{5.20}$$

where the cyclic property of the trace is used.

Perhaps one of the most interesting features of unitary operators in quantum theory are their time-reversibility. Since $U^\dagger = U^{-1}$, the dynamics of a quantum system evolved by unitary $U$ can be reversed by applying the unitary $U^\dagger$

$$|\psi(t_2)\rangle = U|\psi(t_1)\rangle \quad \text{and} \quad |\psi(t_1)\rangle = U^\dagger|\psi(t_2)\rangle. \tag{5.21}$$

The reversibility of quantum mechanics (and classical mechanics for that matter) introduces an interesting paradox. Simply put, if physics is reversible, then why are certain processes seemingly irreversible? For example, an ice cube will melt into a puddle over time on a hot day, however, there is no sensible way in which an ice cube will spontaneously form from a puddle, even at sub-zero temperatures. Th *reversibility paradox*, as it is known, is one of the more puzzling phenomena to appear as we try to map theory of physics on to the physical universe.

**Definition 36. Quantum Circuit:** A graphical representation of the unitary evolution of a quantum system where lines called *wires* represent qubit subsystems and blocks represent unitary operations on those gates (see Fig. 5.2). In the diagram, time passes from left to right where unitary operators are applied in sequential time steps. Additionally, each qubit is assumed to be independent unless acted upon by a unitary operator coupling it to other qubits.

Quantum circuit diagrams explicitly describe when unitary operations are applied and to qubit subsystem(s). Quantum circuit diagrams are often easier to read than their equivalent algebraic representation, which consists of a series of tensor products, matrix multiplications, and permutations of tensor indices. Fortunately, all of this tedious book keeping is explicitly handled by the wiring and ordering of unitary gates in the circuit diagram.

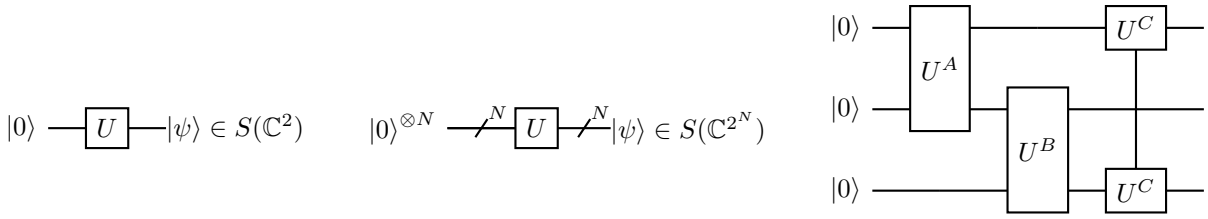



Figure 5.2: (Left) A unitary operator is applied to the zero qubit state as $|\psi\rangle = U|0\rangle$. (Center) An $N$-qubit unitary operator is applied to $N$ zero qubit states as $|\psi\rangle = U|0\rangle^{\otimes N}$. (Right) Two-qubit unitary operators are applied in a cyclic pattern on three qubits where the vertical line connecting two $U^C$ gates denotes that $U^C$ acts on only the first and third qubit wire.

### *5.3.2 Quantum Channels and the Nonunitary Dynamics of Open Quantum Systems*

While the unitary time evolution of a quantum system is a fundamental axiom of quantum mechanics, unitary evolution fails to describe the noisy quantum dynamics observed in both natural and engineered systems. The problem being that unitary operators preserve normalization and purity of a quantum state, but physical systems typically become more mixed over time as noise interferes with the expected quantum process. This disconnect raises the question: how can the unitary model of quantum mechanics describe noisy physical universe?

The natural answer is that quantum systems do not exist in isolation. They are open and constantly interacting with their environment. Through these environment interactions, the quantum states in the investigated system are observed locally to exhibit nonunitary dynamics. This means that quantum states are actually defined on a composite state space $|\psi\rangle^{\text{Sys}} \otimes |\phi\rangle^{\text{Env}} \in S(\mathbb{C}^d \otimes \mathcal{H}^{\text{Env}})$ where $\mathcal{H}^{\text{Env}}$ denotes the Hilbert space of the environment, which may have unbounded dimension. When a unitary operator, $U \in \mathbb{C}^d \otimes \mathcal{H}^{\text{Env}}$, is applied jointly to the system and environment ass $U|\psi\rangle^{\text{Sys}} \otimes |\phi\rangle^{\text{Env}}$, the information encoded into the quantum state $|\psi\rangle^{\text{Sys}}$ is spread out into the environment. Furthermore, the resulting quantum state on Sys is the mixed state obtained via the partial trace

$$\rho^{\text{Sys}} = \text{Tr}_{\text{Env}}\left[U|\psi\rangle\langle\psi|^{\text{Sys}} \otimes |\phi\rangle\langle\phi|^{\text{Env}} U^\dagger\right]. \tag{5.22}$$

In principle, the original quantum state $|\psi\rangle$ can be recovered by reversing the unitary operation that couples the system and environment, however, an experimenter does not always have control over the environment. This general process of noisy evolution of quantum systems is described by the formalism of quantum channels.

**Definition 37. Quantum Channel:** A completely-positive trace-preserving (CPTP) map, $\mathcal{N} : D(\mathcal{H}_d) \to D(\mathcal{H}_{d'})$, that models the nonunitary dynamics of a noisy quantum system. For an input matrix $M \in \mathbb{C}^{d\times d}$, the completely-positive constraint requires that $\mathcal{N}(M) \geq 0$ for $M \geq 0$, while the trace-preserving constraint requires that $\text{Tr}[M] = \text{Tr}[\mathcal{N}(M)]$. The set of all CPTP maps from one density operator space to another is denoted $\text{CPTP}(\mathcal{H}_d, \mathcal{H}_{d'})$.

Quantum channels can be represented in many different ways. The choice of representation largely depends on context. The simplest representation is to express the quantum channel as a function acting upon a quantum state, *e.g.*,

$$\mathcal{N}_\gamma(\rho) = (1-\gamma)\rho + \gamma U\rho U^\dagger \text{Tr}[\rho] \tag{5.23}$$

where $\gamma \in [0,1]$ parameterizes the amount of noise and $U$ is an arbitrary unitary operator. In general, a

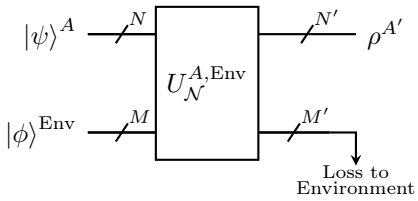



Figure 5.3: A quantum channel is described by a quantum circuit where the system $|\psi\rangle^A$ is modeled on an $N$-qubit register and the environment is modeled on an $M$-qubit register $|\phi\rangle^{\text{Env}}$. The output of the quantum channel is an $N'$-qubit mixed state $\rho^{A'}$ where the remaining $M'$ qubits are lost to the environment. Note that all qubits are conserved, $N + M = N' + M'$, but the input and output registers $A$ and $A'$ do not to have the same dimension.

channel's functional $\mathcal{N}_\gamma$ is arbitrary under the condition that it produces a CPTP map for all $\gamma \in [0, 1]$. While it is often easy to write down a function that describes a channel's action on a quantum state, the detriment of this representation is that the researcher must take care to ensure that the produced channels are CPTP. Alternatively, more direct representations of CPTP maps can be used such as the *systsem-environment representation* and *operator-sum representation* of a quantum channel.

**Definition 38. System-Environment Representation:** [61], [147] A CPTP map is expressed as a unitary operator $U_{\mathcal{N}}^{A,\text{Env}}$ acting upon a system $A$ and its environment as

$$\rho^{A'} = \mathcal{N}(|\psi\rangle\langle\psi|^A) = \text{Tr}_{\text{Env}}\left[U_{\mathcal{N}}^{A,\text{Env}}(|\psi\rangle\langle\psi|^A \otimes |0\ldots\rangle\langle 0\ldots|^{\text{Env}})(U_{\mathcal{N}}^{A,\text{Env}})^\dagger\right], \tag{5.24}$$

where $|\psi\rangle\langle\psi|^A \in S(\bigotimes_{i=1}^N \mathbb{C}^2)$ is an $N$ qubit state, the environment is an $M$-qubit state $|0\ldots\rangle^{\text{Env}} = |0\rangle^{\otimes M}$, and $\text{Tr}_{\text{Env}}[\cdot]$ is the partial trace over the environment. Note that the environment can be assumed to be the zero-state without loss of generality.

In the quantum circuit, the system is modeled using $N$ qubits while the environment is modeled using $M$ ancillary qubits. Generally, the number of ancilla needed to perform any quantum channel on $N$ qubits is bounded as $M \leq 2N$. Therefore, given an $N$-qubit system, its full range of CPTP dynamics can be modeled by a quantum circuit having in total $3N$ qubits where $2N$ of those qubits are reserved as ancilla. However, real-world quantum computers are noisy meaning that additional noise from the quantum computer's environment will be incorporated into the quantum channel in an irreversible manner.

**Definition 39. Opterator-Sum Representation:** [61], [148] A quantum channel $\mathcal{N} \in \text{CPTP}(D^A(\mathbb{C}^{d\times d}), D^{A'}(\mathbb{C}^{d'\times d'}))$ can be expressed as a set of Kraus operators

$$\mathcal{K}_{\mathcal{N}} = \left\{K_i \in \mathbb{C}^{d'\times d} \;\middle|\; \sum_{i=1}^{|\mathcal{K}_{\mathcal{N}}|} K_i^\dagger K_i = \mathbb{I}_d\right\} \tag{5.25}$$

where the Kraus operatore need not be square and the channel's map is applied as

$$\rho^{A'} = \mathcal{N}(|\psi\rangle\langle\psi|^A) = \sum_{i=1}^{|\mathcal{K}_{\mathcal{N}}|} K_i\,|\psi\rangle\langle\psi|^A\,K_i^\dagger. \tag{5.26}$$

## 5.4 The Measurement Axiom

Quantum states are observed through measurement. In this process a researcher applies a measurement $x \in \mathcal{X}$ to a quantum system to obtain a result $y \in \mathcal{Y}$ where $\mathcal{X}$ and $\mathcal{Y}$ are discrete sets. Upon measurement, quantum systems behave stochastically such that given measurement $x$, outcome $y$ is obtained with probability $P(y|x)$. If the experiment can be identically repeated many times the researcher can reconstruct the probability distribution describing the quantum system's behavior. Thus, quantum states are not observed directly, but through their sampled data, therefore, they can be modeled as black-boxes (see Chapter 2).

In the remainder of this section, we discuss in greater detail how the measurement process is incorporated into quantum mechanics. We introduce three types of measurement operators, projector-valued measures (PVMs), positive-operator-valued measures (POVMs), and observables.

### 5.4.1 Projective Measurements

**Definition 40. Projector-Valued Measure (PVM):** A quantum measurement with $|\mathcal{Y}|$ outcomes is modeled as a complete set of projective and orthogonal operators

$$PVM(\mathbb{C}^{d\times d}) \equiv \left\{ \{M_i \in \mathbb{C}^{d\times d}\}_{i=1}^{|\mathcal{Y}|} \;\middle|\; M^2 = M = M^\dagger,\ M \geq 0,\ \sum_{i=1}^{|\mathcal{Y}|} \mathcal{Y}|M_i = \mathbb{I}_d,\ M_i M_j = \delta_{i,j} M_i \right\} \tag{5.27}$$

where each operator $M_i$ corresponds to a distinct measurement outcome and $|\mathcal{Y}| \leq d$

In general a PVM element is an operator $M_y$ that corresponds to the measurement outcome $y \in \mathcal{Y}$. Given a density operate state $\rho \in D(\mathbb{C}^{d\times d})$ the probability of observing measurement outcome $y$ is given by the Born rule

$$P(y) = \mathrm{Tr}\left[M_y \rho\right] \tag{5.28}$$

where $0 \leq P(y) \leq 1$. Note that the researcher has no control over the measurement outcome, but the researcher can control choose the PVM measurement operators $\{M_{y|x}\}_y$ where $x \in \mathcal{X}$ labels the PVM.

When $|\mathcal{Y}| = d$, a PVM forms an orthonormal basis where each PVM element has $\mathrm{Rank}\,(M_y) = 1$. In this case, we refer to the PVM as a measurement basis $\{|\phi_y\rangle\}_{y=1}^{|\mathcal{Y}|}$ where $M_y = |\phi_y\rangle\langle\phi_y|$.

$$P(y) = |\,\langle\phi_y|\psi\rangle\,|^2 \quad \text{where} \quad \rho = |\psi\rangle\langle\psi|\,. \tag{5.29}$$

When a measurement result is obtained, the experimenter gains information about the state of the quantum system. In essence, the quantum state has been projected onto one of the measurement outcomes. Mathematically, the post-measurement state is expressed

$$\rho' = \frac{\Pi_y \rho \Pi_y^\dagger}{\mathrm{Tr}\left[\Pi_y \rho\right]}. \tag{5.30}$$

When $\mathrm{Rank}\,(\Pi_y) = 1$, Eq. (5.30) becomes

$$\rho' = \frac{|\phi_y\rangle\langle\phi_y|\,\rho\,|\phi_y\rangle\langle\phi_y|}{\langle\phi_y|\rho|\phi_y\rangle} = |\phi_y\rangle\langle\phi_y|\,. \tag{5.31}$$

In other words, the quantum state is projected onto one of the measurement basis states.

In a quantum circuit diagram, a projective measurement in the computational basis is denoted using a *meter* symbol and occurs at the end of a wire (see Fig. 5.4). To measure in a basis other than the computational basis, a unitary operator can be used to perform a change of measurement basis operation. Since the quantum measurement process is stochastic, the measurement result can be regarded as a sample from a random distribution as defined for the considered measurement operator and quantum state.

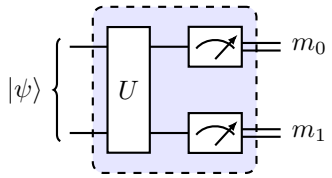



Figure 5.4: A projective two-qubit measurement in an arbitrary measurement basis that yields a two-bit measurement result $m_0, m_1 \in \mathbb{B}$. The meters measure in the computational basis while the preceding unitary operator can arbitrarily transform the computational basis into any measurement basis.

### *5.4.2 General Measurements*

Projective measurements are restrictive because they require orthogonality between measurement operators. In general, quantum measurement operators need not be orthogonal projectors. The advantage of relaxing these constraints is that the measurement operator interacts with the quantum system more weakly, meaning that its effect on the measured state is minimal at the expense of gaining less information about the quantum system.

**Definition 41. Positive Operator-Valued Measure (POVM):** A quantum measurement represented as a set of operators representing distinct measurement outcomes

$$POVM(\mathbb{C}^{d\times d}) \equiv \left\{ \{\Pi_i \in \mathbb{C}^{d\times d}\}_{i=1}^{|\mathcal{Y}|} \;\middle|\; \Pi \geq 0,\ \sum_{i=1}^{|} \mathcal{Y}|\Pi_i = \mathbb{I}_d \right\} \tag{5.32}$$

where $|\mathcal{Y}| > 0$.

Since POVM elements are not required to be orthogonal projectors, the number of measurements operators $n$ can exceed the dimension of the state space $d$. However, the probability of obtaining outcome $y \in \mathcal{Y}$ is determined by the Born rule

$$P(y) = \mathrm{Tr}\left[\Pi_y \rho\right]. \tag{5.33}$$

The post measurement state can be calculated by factoring the POVM elements ass $\Pi_y = K_y^\dagger K_y$ where $K_y$ are Kraus operators. In this way, a POVM resembles a quantum channel in the operator-sum representation of Eq. (5.26). The post-measurement state is then expressed as

$$\rho' = \frac{K_y \rho K_y^\dagger}{\mathrm{Tr}\left[\Pi_y \rho\right]}. \tag{5.34}$$

When a POVM is viewed as a quantum channel, the Stinespring dilation theorem [147] used in system-environment representation of a quantum in Eq. (5.24) can be extended to POVMs. This result, referred to as the Neumark dilation theorem [149], proves that any POVM can be written as a PVM on a larger state space.

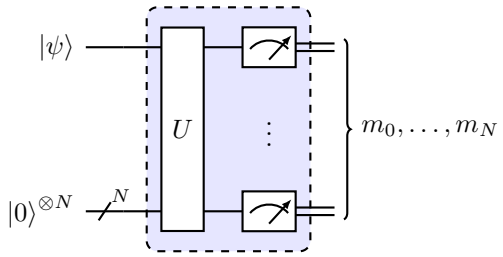



Figure 5.5: A qubit POVM measurement is implemented on a quantum circuit as a dilation on the input state space $|\psi\rangle \in S(\mathbb{C}^{2\times 2})$ extended with an $N$-qubit ancilla. Note that the $N+1$ bit string that is obtained through measurement may require additional post-processing the number of measurement outcomes it not $2^{N+1}$.

### *5.4.3 Observables*

In many cases, quantum systems are characterized by the expectation of their measurement results rather than the probability of seeing different outcomes. The formalism of *observables* simplifies the calculation of expectation values in quantum systems.

**Definition 42. Observable:** A Hermitian operator that describes the expectation of a measurable quantity in a quantum system

$$Obs(\mathbb{C}^{d\times d}) \equiv \left\{ O \in \mathbb{C}^{d\times d} \;\middle|\; O = O^\dagger \right\}. \tag{5.35}$$

An observable is $O \in Obs(\mathbb{C}^{d\times d})$ can be expressed in terms of its eigenvectors $|\phi_y\rangle$ and its eigenvalues $b_y$ as

$$O = \sum_{y=0}^{d-1} b_y \, |\phi_y\rangle\langle\phi_y| \, . \tag{5.36}$$

Note that $\{|\phi_y\rangle\}_y$ form an orthonormal measurement basis, which is a PVM, and the eigenvalues simply apply a physical to each outcome. Since $O$ is Hermitian, its eigenvalues are real $b_y \in \mathbb{R}$, but otherwise unrestricted. For example, the Hamiltonian of a quantum system is an observable that measures the system's energy and its eigenvalues correspond to energy of each possible state the system can be observed in.

The expected measurement outcome for given a density operator $\rho \in D(\mathbb{C}^{d\times d})$ is expressed as

$$\langle O \rangle_\rho = \mathrm{Tr}\left[O\rho\right] = \sum_y b_y \mathrm{Tr}\left[|\phi_y\rangle\langle\phi_y| \, \rho\right] = \sum_y b_y P(y). \tag{5.37}$$

Hence, we explicitly show the relation between an observable and the PVM described by its eigenvectors. Furthermore, from the far right of Eq. (5.37), we recover the standard definition of the expectation value from probability theory.

An important distinction between quantum and classical measurements is compatibility. In brief, two observables $A, B \in Obs(\mathbb{C}^{d\times d})$ are compatible if and only if

$$[A, B] = AB - BA = 0, \tag{5.38}$$

which implies that both observables $A$ and $B$ can be known simultaneously. While Eq. (5.37) must always hold in classical systems, it does not need to hold in quantum systems.

## *5.5 Does Quantum Mechanics Describe Reality?*

When quantum physics is applied, it predicts strange behaviors that defy the familiar classical intuition. We expect that past events events *causally* affect the future but future events do not affect the past. We expect that nature is *realistic* by having properties that are objectively defined, even when they are not being measured. Finally, we expect that nature is *local* such that the dynamics of a system depend only on its objective properties and any forces cannot propagate faster than the speed of light. These notions of causality, realism, and locality our deeply rooted in our human experience. Thus when quantum physics defies these classical ideals, quantum systems appear to be quite strange. This strangeness was even unsettling for the researchers Einstein, Podolsky, and Rosen who argued in their 1935 work that quantum mechanics was not a complete description of nature because it does not describe nature in a local and realistic way [11]. To support this claim, they draw upon two counterintuitive examples where quantum mechanics seemingly defies classical physics.

For instance, when incompatible observables, *e.g.*, $\sigma_x$ and $\sigma_z$, are measured sequentially on same quantum particle, the result of the second measurement precludes the result of the first in a destructive manner [11]. The dynamics of quantum systems appear to be dependent upon their applied measurements, making it impossible to separate the observer from the observed. For this reason, the physical properties measured by the two measurements are not described by an objective reality.

Second, consider two quantum particles that are entangled and separated by a great distance. If one particle is measured, then the other particle is left in a state dependent upon the first particle's measurement. To Einstein, Podolsky, and Rosen, this theoretical nuance cannot describe nature because for the second particle's state to depend upon the first particle's measurement some propagating force is needed to travel from one to the other. If the second particle is not in the first particle's light cone, this force would need to propagate faster than the speed of light. However, a physical mechanism for superluminal signaling is refuted strongly by Einstein's theory of special relativity [150].

Given these peculiarities, the researchers postulated that quantum mechanics is an incomplete theory. As a correction, they argued that local hidden variables must exist that determine the behaviors of quantum systems. If these hidden variables were known, then the outcomes of quantum experiments could be calculated with certainty. As a result, quantum mechanics is an incomplete theory resulting from our incomplete understanding of the subatomic universe. Although, the notion of local hidden variables is enticing, it was later proven wrong [12] and verified experimentally [151] that the entanglement correlations cannot be described by local hidden variables.

Much later, in 1978 when J.A. Wheeler proposed a thought experiment referred to as the *delayed-choice experiment*. The goal of the experiment was to determine when a quantum particle "decided" to exhibit a wave-like or particle-like behavior [152]. In general, the delayed-choice experiment is a scenario in which the wave-particle duality of individual quantum particles is tested as they pass a through double-slit interferometer [153]. That is, if both slits are open, then the particle will exhibit a wave-like interference pattern. However, if one of the two slits is blocked, then the quantum particle exhibits the particle-like behavior of having a fixed trajectory. These two measurements are incompatible because the particle can't be both wave-like and particle-like simultaneously. To explain this phenomenon, it was thought by Wheeler that setting the apparatus up for wave-like measurement or particle-like measurements could affect whether or not the prepared quantum state is a wave or particle. In the delayed-choice experiment, this behavior is ruled out because the quantum particle is first prepared and sent towards quantum double-slit interferometer, but while

the particle is in flight, the choice of measurement is made. Thus, upon initialization, the particle is unable to know which measurement will occur. The delayed-choice experiment can be demonstrated experimentally using a Mach-Zehnder interferometer where it is shown that quantum particle always exhibits the properties of the applied measurement [154].

It is no doubt that incompatible measurements, entanglement, and wave-particle duality allow for counterintuitive quantum phenomena to occur. However, experiments have verified that these phenomena also occur in nature. As a result, quantum mechanics accurately predicts the behavior of the natural world. However, it is somewhat unsatisfying because quantum mechanics fails to make predictions with certainty. Nevertheless, if these strange quantum phenomena exist, then they can be measured and applied.

## *5.6 The Nonclassicality of Quantum Mechanics and Nature*

Nearly three decades passed after Einstein, Podolsky, and Rosen posited their conjecture that local hidden variables determined quantum mechanics with certainty. However it was remarkably rebutted by Bell who showed that two entangled quantum particles can produce correlations that are not reproducible by any physical theory, that is local, realistic, and causal, classical physics being such an example theory [12]. Bell's result is interesting because it implies the existence of an operational test that can verify whether or not the natural world obeys local realism. Experimentally, such Bell tests have been implemented and it has been demonstrated that the predictions of quantum mechanics are correct [14]–[17], [155]–[157]. Therefore, quantum physics is nonclassical and defies our classical understanding of physics.

Nonclassicality can also occur in systems where there is a strict time-like separation such as the delayed-choice experiment. Although, the experimental results of delayed-choice experiments seem to indicate that upon creation, the quantum particle has yet to "decide" whether it is a wave or particle [154], it was recently shown that Wheeler's delayed-choice experiment can be reproduced by a classical model [158]. However, if the state preparation is conditioned upon a random input such that for each input $x$ a different quantum state $|\psi_x\rangle$ is prepared, then quantum systems can exhibit nonclassical behaviors that cannot be reproduced by a classical theory [158]–[160]. Furthermore, these nonclassical behaviors have been verified experimentally using a Mach-Zehnder interferometer [161], [162].

The nonclassical phenomena that occur in both Bell tests and delayed-choice experiments has spurred a broad field of research into the study of various nonclassical phenomena [64], [109], [124], [163]. The concepts of nonclassicality have even been extended to networks using a range of theoretical frameworks **Tava**, [164]–[171]. From a perspective in foundational physics, these examples of nonclassicality demonstrate how locality, causality, and realism cannot all hold simultaneously in quantum systems. However, exactly how these classical assumptions break down is still a mystery and remains open to interpretation.

# Chapter 6

# Quantum Information Technology

From an engineering standpoint, quantum technologies apply quantum physics to solve problems. Quantum technologies can be categorized into three groups: computing, sensing, and communications. A wide range of quantum computing architectures are being developed [39], however, their advantage has yet to be demonstrated. Quantum computers are theorized to provide significant advantages in simulation tasks [18]–[22], which may lead to improved algorithms for designing drugs [23] and materials [24]. Similarly, quantum algorithms have been identified that provide exponential or quadratic improvements over the best existing algorithms [25]. Examples include searching over big data sets [26], or factoring large numbers [27]. Alternatively, quantum systems can be used to sense their environment with greater precision than existing techniques [28]. Quantum sensing techniques are already being used in practical applications such as improving sensitivity gravitational wave detectors [29]. More generally, quantum sensing techniques are expected to improve telescopes [30], [31], clock synchronization [32], and biomedical imaging applications [33], [34]. Finally, quantum communications technologies can help improve the efficiency of data transmission, improve information security, or be used transmit quantum systems [35]–[37].

## *6.1 Quantum Information*

The quantization of quantum systems permits information to be encoded into its discrete quantum states. From this perspective, a *bit* of information (a 0 or 1 value) can be encoded into a two-level quantum particle called a *qubit*. However, a qubit is distinct from a bit because the qubit can be in a *superposition* of the 0 and 1 state simultaneously. The qubit superposition state is represented by the complex-valued vector $|\psi\rangle = \cos(\theta)\,|0\rangle + \sin(\theta)e^{i\phi}\,|1\rangle$, where $|0\rangle = (1,0)^T$ and $|1\rangle = (0,1)^T$ correspond to the two possible classical values. When measured, a qubit is projected into a definite outcome, 0 or 1, which is a classical bit. The probability of obtaining each measurement outcome is calculated from the qubit's quantum state as $P(x=0) = |\,\langle 0|\psi\rangle\,|^2$ and $P(x=1) = |\,\langle 1|\psi\rangle\,|^2$. The process of measurement alters the quantum state such that the post-measurement qubit matches the classical bit value observed as output. There is also a contextuality to measurements in quantum systems. That is, two measurements that could be applied to a qubit may be incompatible such that their outcome cannot be simultaneously known. This property formally known as the Heisenberg uncertainty principle [172], which means that when one quantity of a quantum system is measured with certainty, any information associated with a second incompatible measurement is

lost, hence the certainty of which both quantities can be predicted is bounded.

In quantum communication a quantum state, such as a qubit, is sent from one party to another raising the question of whether quantum physics allows for more efficient communication. After all, an enormous amount of information could be encoded into the continuum of possible qubit superpositions. However, there is a fundamental limit such that only one bit of classical information can be decoded from the qubit. This result was first formulated by Holevo stating that the classical communication capacity of a quantum system does not exceed that of its corresponding classical system of equal dimension [173]. These results were later strengthened by Frenkel and Weiner who showed that the classical data of a quantum communication channel can always be simulated exactly using a dimension no greater than that of the quantum channel [115].

The prepare-evolve-measure model of quantum mechanics can be recast as an information processing scheme: encode, process, and decode. First, information is encoded into quantum state through state preparation. Next, the information is processed as the state undergoes unitary time evolution. Finally, information is decoded from the state through measurement.

The encoding, processing, and decoding are controlled by classical data, meaning that they can each be given an input $x_1 \in \mathcal{X}_1$, $x_2 \in \mathcal{X}2$, and $x_3 \in \mathcal{X}_3$ respectively. In this manner, $x_1$ corresponds to the applied quantum state encoding $\rho_{x_1} \in D(\mathcal{H}^A)$, $x_2$ corresponds to the applied channel processing $\mathcal{E}_{x_2} \in \text{CPTP}(A \to A')$, and $x_3$ corresponds to the applied measurement decoding $\{\Pi_{y|x_3}\}_{y\in\mathcal{Y}} \in \text{POVM}(\mathcal{H}^{A'})$ where mixed states, CPTP maps, and POVMs are considered for full generality. Upon measurement, the system outputs the value $y \in \mathcal{Y}$ with probability

$$P(y|x_1, x_2, x_3) = \text{Tr}\left[\Pi_{y|x_3}\mathcal{E}_{x_2}(\rho_{x_1})\right]. \tag{6.1}$$

The quantum information processing system can be modeled as a black-box because classical data is input to and output from the system, however, the evolution of the quantum state space is hidden from the researcher. Like a black-box, the probabilities in Eq. (6.1) can be used to construct a black-box behavior $\mathbf{P} : \mathcal{X}_1 \times \mathcal{X}_2 \times \mathcal{X}_3 \to \mathcal{Y}$ for the quantum system. Through the formalism of black-box games, the quantum system can be given task and it performance evaluated.

## *6.2 The Quantum Internet Vision and Road Map*

Central to the development and scaling of quantum technologies is connecting many quantum devices together to construct a quantum network. In a quantum communication network, entanglement and quantum communication are fundamental resources that enable the network to perform its functions. The advantages of quantum information technologies are then derived from operating upon these resources. The vision is to build a quantum network at global scale referred to as the quantum internet [52]–[54], [174], [175].

In the long-term, quantum communication networks are predicted to provide a slew of advantages in applied communications. A core function of quantum networks is to use quantum communication approaches [35]–[37] to distribute quantum information and entanglement to many network users. These networked resources can be used to bolster many existing technologies. For instance, quantum networks, enable distributed quantum sensing [176], [177], in which quantum communication and entanglement assist in applications such as synchronizing atomic clocks [32] or building larger telescopes arrays [178], [179]. Similarly, quantum communications can help improve information security in applications such as secret key sharing [180]–[182] and evaluating algorithms on a quantum computer privately [183], among many other applications [184]. Furthermore, quantum networks can be used to implement distributed quantum computing [55], [56], [64], in which large quantum computers are developed by distributing their operations across many small quantum

computers. While the prospect of these advantages may be exciting, we must first overcome the challenge posed by noise in quantum networking hardware.

We are far from building the quantum internet, however, it is predicted that a quantum internet can be developed and scaled incrementally as new functionality and novel quantum capabilities emerge [52]. Currently, quantum networks allow for *prepare-and-measure* communication where sender devices prepare quantum states and transmit them to receiver devices. The receiver device measures the incoming state to produce a classical result. Prepare-and-measure networks lack a quantum memory, meaning that quantum states are short-lived. As a result, the inputs and outputs of a prepare-and-measure network are classical and extensive processing cannot be applied to quantum states. Despite these limitations, such networks can be used for applications in key distribution [181], [182], [185] and randomness generation [186]–[192]. In practice, quantum prepare-and-measure networks have been deployed with applications in quantum key distribution [42]–[49]. These networks are generally constructed using linear optics, leading to measurement limitations. Furthermore, the nonclassical behaviors that occur in prepare-and-measure networks have been both studied extensively [124] and demonstrated experimentally [193]–[197].

A more advanced quantum networking functionality incorporates the use of classical communication to assist quantum communication resources. This functionality is often referred to as local operations and classical communication (LOCC) where network devices operate on their local quantum states while coordinating their operations using classical communication. LOCC networks enables important protocols such as teleportation [198], entanglement distribution [110], [111], and quantum network coding [199], which can be used to prepare complex multipartite quantum states or send quantum states over long-distances using quantum repeaters [112], [113]. While LOCC networks still operate in a prepare-and-measure format, they do require general improvements to control, short-term quantum memory, and noise robustness to accommodate the complications of LOCC. A further challenge is that entangled measurements are needed for entanglement distribution protocols, however, such measurements are not available using the linear optics typically sufficient for prepare-and-measure networks. As a result, LOCC networks apply matter-based qubit architectures [200]–[202], although, network teleportation protocols have been demonstrated using photonic qubits [35].

An interesting extension of LOCC networks is measurement-based quantum computation [203]–[207], in which computation is performed on a large entangled quantum state and measurement results condition downstream operations. Such measurement-based quantum computers are actively being developed by companies such as PsiQuantum [208] and Xanadu [209]. Measurement-based quantum computation is attractive for distributed quantum computing because once the entanglement is distributed, only local operations and classical communication are needed to run the computation.

Indeed it is an exciting time in the development of quantum networking. We are at the cusp of being able to scale the first real-world applications of prepare-and-measure networks while LOCC networks are actively under development. Despite having noisy hardware, these simple quantum networking technologies are still predicted to provide value in communications and sensing technologies. As quantum networks become more robust to noise, their processing and storage capabilities will improve, leading to the quantum internet [52].

## *6.3 Quantifying the Quantum Advantage*

From a physical standpoint, it is interesting to ask why quantum physics seems to offer an advantage in information processing tasks. Indeed, if one technology provides an advantage over another, there ought to be a way to quantify that advantage through measurement and observation. Demonstrating such *operational*

advantages is necessary to justify the cost of continued research into quantum information technology.

The operational advantage of information processing tasks can be quantified in terms of the fundamental resources consumed during the information processing task. In classical information processing system, the resource correspond to the bits that encode information and the operations applied to those bits. Similarly, in quantum information systems, the resources are the quantum states that encode information and the operations applied to those states. In either setting, the number of resources can be counted simply by inspecting the system and its applied algorithm or protocol.

In quantum communication networks, a resource advantage can be shown in two ways. A communication complexity advantage is shown when fewer quantum resources are needed to perform an information processing task than analogous classical resources. Alternatively, a nonclassicality advantage is demonstrated when a fixed set of quantum communication resources produces statistics that cannot be reproduced using a similar classical network using an equal amount of classical communication resources.

Resource complexity advantages are a standard approach taken in quantum computing where a quantum advantage is demonstrated by showing that fewer quantum resources are needed than classical [64], however, such an approach presents a few issues. First, it is easy to place an upper bound on the number of resources required to solve a task because the resource complexity of any known algorithm can be counted. However, it is generally difficult to place a lower bound on the number of resources because this requires knowledge of the best algorithm to apply, which is a nontrivial task even in the classical case. As a result, many of the resource complexity advantages that have been identified are compared to the classical upper bound, meaning that the possibility of a more efficient classical algorithm has not been ruled out. Furthermore, complexity advantages often require a large amount of quantum resources to show their advantage, but NISQ devices of may have too few qubits or too much noise to realize the advantage. Third, measuring the resource complexity requires the researcher to have knowledge of the applied algorithm and the hardware that runs the algorithm. Indeed, if device performs an information processing task without error, then it is impossible to assert whether the device used quantum or classical resources and how many resources were used. Thus, resource complexity advantages are not very operational because they cannot be observed from the measured data easily.

Alternatively, nonclassicality advantages provide a more operational way in which the quantum advantage can be quantified. More precisely, a quantum network exhibits nonclassicality if its produced probability distribution cannot be reproduced using similar resources in a classical network. In order to reproduce the quantum behavior, the classical network would need to use more classical resources. which implies that the quantum resources are more valuable. Nonclassicality is a practical quantifier of advantage because it requires minimal quantum resources to demonstrate and it quantifies quantum resources precisely in terms of classical resources.

# Chapter 7

# Quantum Communication Networks

Equipped with our black-box model for classical networks, we are now ready to take the quantum leap. Similar to a classical communication network, a quantum network is a collection of communicating devices that process information in a distributed manner. However, the network devices are able to perform quantum operations and utilize quantum communication resources. As a classical communication network enables classical information technologies to be distributed, a quantum network enables quantum information technologies to be distributed. As discussed in Chapter 6, many applications can be developed using quantum networks. Indeed, this foundational framework for quantum information technologies can be used broadly throughout quantum information science and technology.

## *7.1 Graphic Representation of Quantum Communication Networks*

When quantum networking devices are connected together using quantum communication resources, they form a quantum communication network. As in the classical network case, there is a well-defined causal flow and ordering of the network's events and operations. This causal structure imposes clear time-like and space-like separations between networking devices, as well as the amount of communication between devices.. The quantum network DAG is distinct from the classical network DAG, as shown in Fig. 3.1, because quantum networks incorporate quantum communication resources. Furthermore, devices are no longer treated as black-boxes, but are assumed to have a set of free operations that they can apply.

**Definition 43. Quantum Communication Network Topology:** The structure of a quantum communication network is analogous to the communication network topology in definition 10. A quantum network's topology can be represented graphically by a directed acyclic graph (DAG) (see Fig. 7.1). In the DAG, each node represents a quantum networking device and the single-lined arrows depict quantum communication between devices while the double-lined arrow depicts classical communication. The network's topology can be equivalently expressed as a collection of nodes $\vec{N}$ that represent networking devices and their connecting edges $\vec{E}_Q$ and $\vec{E}_C$ that represent quantum and classical communication respectively. The network is then described as

$$\text{Net} \equiv \left\{ \vec{N} \equiv \left( \vec{N}_i \equiv (\text{Dev}_{i,j})_{j=1}^{|\vec{N}_i|} \right)_{i=1}^{|\vec{N}|}, \quad \vec{E}_Q \equiv \left( \text{id}^{\text{Tx}_i \to \text{Rx}_i} \right)_{i=1}^{|\vec{E}_Q|}, \quad \vec{E}_C \equiv \left( \text{id}^{\text{Tx}_i \to \text{Rx}_i} \right)_{i=1}^{|\vec{E}_C|} \right\}. \tag{7.1}$$

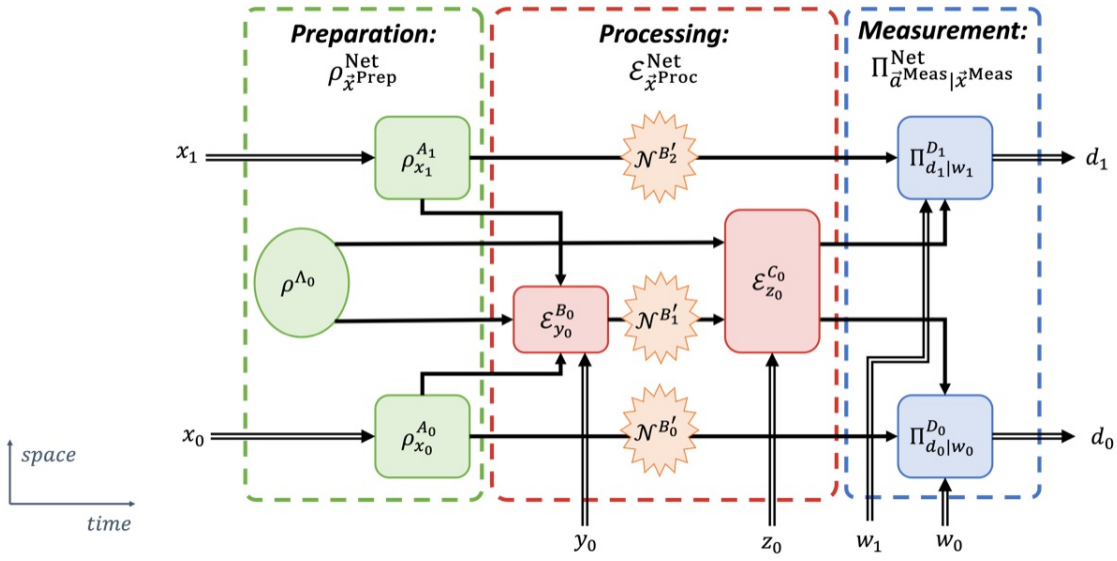



Figure 7.1: Example DAG Representing a Quantum Communication Network. A quantum network can be decomposed into three operational layers: preparation, processing, and measurement. Network layers occur sequentially in time. Devices that are linked directly or indirectly are time-like separated, whereas devices with no causal link between each other are space-like separated.

where Eq. (7.1) is analogous to the definition for classical communication network topology (see Def. 10). However, the classical case is extended with quantum communication $\vec{E}_Q$ where $\mathrm{id}^{\mathrm{Tx}\to\mathrm{Rx}}$ denotes the ideal transmission of a quantum state from a sender device (Tx) to a receiver device (Rx).

Each node in a network's DAG corresponds to a quantum networking device. The superscripts index the device's location in the network where the layer's are labeled alphabetically and device's in the same layer are labeled numerically. The links between nodes correspond to dynamic quantum or classical communication resources being used to communicate. We refer to the quantum communication as wires because they will ultimately correspond to wires in a quantum circuit when we introduce our network simulation framework.

Fundamentally, quantum networks prepare, process, and measure quantum states. The classical data that is input to the system is encoded into the network's operations and measurements. When the quantum state is measured, a classical value is output from the network. Hence, a quantum network has a classical input and output and the quantum system and communication resources are responsible for mapping the in input alphabet $\vec{x} \in \mathcal{X}^{\mathrm{Net}}$ in the the output alphabet $\vec{y} \in \mathcal{Y}^{\mathrm{Net}}$.

At a high-level, a quantum network decomposes into three layers: preparation, processing, and measurement. In the preparation layer, the network state $\rho^{\mathrm{Net}}_{\vec{x}^{\vec{A}}}$ is initialized. In the processing layer, the quantum channel operation $\mathcal{E}^{\mathrm{Net}}_{\vec{x}^{\vec{B}}}$ is applied. Finally, in the measurement layer the network measurement $\Pi^{\mathrm{Net}}_{\vec{y}^{\vec{C}}|\vec{x}^{\vec{C}}}$ is applied. In this simplified networking model the three layers can be combined into a concise expression using the Born rule of Eq. (5.29) and Eq. (5.28). For pure states, unitary processing, and projective measurements, we find

$$P^{\mathrm{Net}}(\vec{y}|\vec{x}) = \left| \left\langle \phi^{\mathrm{Net}}_{\vec{y}^C|\vec{x}^C} \middle| U^{\mathrm{Net}}_{\vec{x}^B} \middle| \psi^{\mathrm{Net}}_{\vec{x}^A} \right\rangle \right|^2 . \tag{7.2}$$

Similarly for mixed states, quantum channel processing, and POVM measurements, we find

$$P^{\mathrm{Net}}(\vec{y}|\vec{x}) = \mathrm{Tr}\left[ \Pi^{\mathrm{Net}}_{\vec{y}^C|\vec{x}^C} \mathcal{E}^{\mathrm{Net}}_{\vec{x}^B}(\rho^{\mathrm{Net}}_{\vec{x}^A}) \right] . \tag{7.3}$$

In both cases, the classical input is defined $\vec{x} = \{\vec{x}^A, \vec{x}^B, \vec{x}^C\}$ and the the output is $\vec{y} = \vec{y}^C$. Furthermore, each layer is assumed to implicitly permute its subsystems such that each layer is properly wired together as expressed explicitly by the network's DAG. That is the identity channels $\mathrm{id}^{\vec{A}\to\vec{B}}$ and $\mathrm{id}^{\vec{B}\to\vec{C}}$ are implicitly assumed to be in Eq. (7.2) and Eq. (7.3), mapping each layer to the next as directed by the DAG.

It is important to remark that our quantum network model has a classical input and output, while many quantum networking applications, such as teleportation, rely consider quantum inputs and/or quantum outputs. However, we argue that quantum states themselves are not operational, hence, such applications having quantum inputs and/or outputs are not operational on their own. More simply put, no advantage can be realized in the applications until a measurement is made, producing a classical outcome. Thus, a device that has quantum inputs and quantum outputs serves as a convenient component of a quantum network, but shouldn't be regarded as being operational on its own. After all, what is the use of a teleportation device if its input and output are quantum states that are hidden from our objective reality? Indeed, the teleportation device might as well have not existed in the first place—Just like the wavefunction! This is not to say that protocols like teleportation do not provide value, it's just that the value is not realized until measurements are made yielding an operational advantage. Therefore, we ground our approach to quantum networking in reality, not in the metaphysical quantum state space.

Since we consider quantum networks that have classical inputs and outputs, we can characterize the network by its behavior. In this way, a quantum network can be regarded similarly to a black-box in the discrete memoryless setting. This setting is especially practical for near-term quantum networks, which are unlikely to have robust quantum memories and each shot of the network can be treated as being independent.

**Definition 44. Quantum Network Behavior:** $\mathbf{P}^{\mathrm{Net}} \in \mathcal{P}_{\mathcal{Y}|\mathcal{X}}$, A quantum network is characterized by its behavior $\mathbf{P}^{\mathrm{Net}} = \sum_{\vec{y}\in\mathcal{Y}} \sum_{\vec{x}\in\mathcal{X}} P^{\mathrm{Net}}(\vec{y}|\vec{x}) \, |\vec{y}\rangle\langle\vec{x}|$. For a layer of pure state preparations $\left|\psi^{\mathrm{Net}}_{\vec{x}\vec{A}}\right\rangle \in S(\mathcal{H}^{\vec{A}})$, a layer of unitary network processing $U^{\mathrm{Net}}_{\vec{x}\vec{B}} \in U(\mathcal{H}^{\vec{B}})$, and a layer of rank-one projective measurements $\left|\phi^{\mathrm{Net}}_{\vec{x}\vec{C}}\right\rangle\!\left\langle\phi^{\mathrm{Net}}_{\vec{x}\vec{C}}\right|$ the network's transition probabilities are given by Eq. (7.2) or Eq. (7.3).

**Definition 45. The Set of Quantum Network Behaviors:** $\mathcal{Q}^{\mathrm{Net}}$, the set of quantum network behaviors that satisfy Def. 44.

## *7.2 Quantum Networking Devices*

The devices in a quantum network are similar to the network of black-boxes used to model classical systems. To extend the black-box model to quantum systems, we need to incorporate quantum inputs and outputs into the devices. Note that each device has a discrete classical input that controls the operation, however a classical output is only provided if the device performs measurement.

**Definition 46. Quantum Networking Device:** A device that can receive, processes, and transmit both quantum and classical information. A quantum networking device may have the same classical inputs and outputs of as black-box networking devices (see Def. 11), however, quantum devices may also have a quantum input $\rho^{\mathrm{Dev}}_{in} \in S(\mathbb{C}^d)$ and quantum output $\rho^{\mathrm{Dev}}_{out} \in S(\mathbb{C}^{d'})$.

While Def. 46 is quite generic, it is convenient to give the quantum networking devices greater specificity (see Fig. 7.2). That is, as quantum information flows through the network it must first be encoded, then processed, and finally measured, where each step in the network may implemented by an independent device in the network. Indeed, devices can fulfill three key roles that align with the fundamental axioms of quantum

mechanics: preparation, evolution, and measurement. To help distinguish the roles and operations applied at each device, we use a color-coding scheme where green nodes correspond to preparation devices, red nodes correspond to processing devices, blue nodes correspond to measurement devices, and orange nodes correspond to noise.

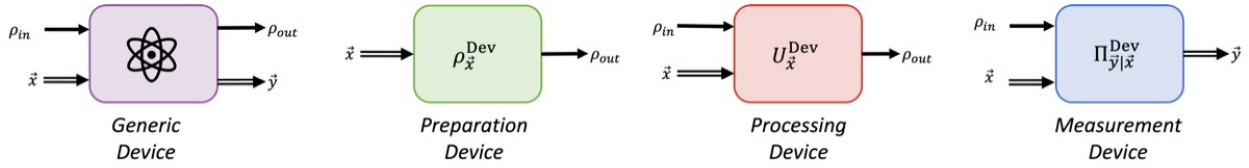



Figure 7.2: The basic types of quantum networking devices. From left to right, a generic quantum device (purple), a preparation device (green), a processing device (red), a measurement device (blue).

**Definition 47. Preparation Device:** A quantum networking device that prepares a quantum state $\rho_{\vec{x}}^{\text{Dev}} \in D(\mathcal{H}_d)$ conditioned upon the classical input $\vec{x}$. An $N$-qubit pure state $|\psi\rangle_{\vec{x}}^{\text{Dev}} \in S(\mathcal{H}_2^{\otimes N})$ is prepared by unitary evolution on the zero state, $|\psi\rangle_{\vec{x}}^{\text{Dev}} = U_{\vec{x}}^{\text{Dev}} |0\rangle\langle 0|^{\otimes N} (U_{\vec{x}}^{\text{Dev}})^\dagger$. Mixed state preparations can be achieved by a unitary evolution of the zero state on a dilated state space

$$\rho_{\vec{x}}^{\text{Dev}} = \text{Tr}_{\text{Env}} \left[ U_{\vec{x}} |0 \dots 0\rangle\langle 0 \dots 0| (U_{\vec{x}})^\dagger \right] \tag{7.4}$$

**Definition 48. Processing Device:** A quantum networking device that applies an operation to the input state $\rho_{in}$. The operation is conditioned on a classical input $\vec{x}$ and may either be unitary or a CPTP map. For unitary operations, the output is $\rho_{out} = U_{\vec{x}}^{\text{Dev}} \rho_{in} (U_{\vec{x}}^{\text{Dev}})^\dagger$. For CPTP operations, the output is $\rho_{out} = \mathcal{N}_{\vec{x}}(\rho_{in})$, which may be implemented unitarily using the system-environment representation of a quantum channel.

**Definition 49. Measurement Device:** A quantum networking device that measures the input state $\rho_{in}^{\text{Dev}}$ and outputs the classical result $\vec{y}$. The measurement is conditioned on a classical input $\vec{x}$ and the measurement operator may either be a PVM or POVM $\Pi_{\vec{y}|\vec{x}}^{\text{Dev}}$. The probability that the device outputs $\vec{y}$ is given by the Born rule, $P(\vec{y}|\vec{x}) = \text{Tr}\left[ \Pi_{\vec{y}|\vec{x}}^{\text{Dev}} \rho_{in}^{\text{Dev}} \right]$. Furthermore, measurement devices may measure an observable $O_{\vec{x}}^{\text{Dev}}$ in place of PVM or POVM operators, in which case the device outputs the expectation $\langle O_{\vec{x}}^{\text{Dev}} \rangle_{\rho_{in}^{\text{Dev}}}$.

**Definition 50. Noise Node:** Noise is represented in a quantum network DAG as a jagged starburst node in orange. A noise node applies the noisy quantum channel $\mathcal{N} \in \text{CPTP}(\mathcal{H}_{in}, \mathcal{H}_{out})$ taking the input state to the output state as $\rho_{out} = \mathcal{N}(\rho_{in})$. Note that noise nodes are not conditioned upon a classical input because the noise is not controllable. Furthermore, by the system-environment representation of a quantum channel, noise can be modeled using unitary operators as $\rho_{out} = \text{Tr}_B \left[ U^{AB} (\rho_{in}^A \otimes |0 \dots 0\rangle\langle 0 \dots 0|^B) (U^{AB})^\dagger \right]$.

## *7.3 Quantum Network Communication Resources*

Quantum systems can be utilized as communication resources that link the devices of the network. In classical networks, classical communication and shared randomness were considered as resources. In quantum networks, similar resources of quantum communication and entanglement can be considered in addition to classical communication.

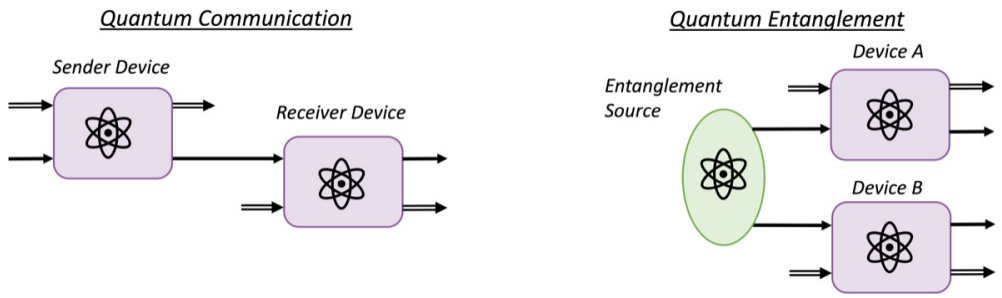



Figure 7.3: Depictions of quantum communication and entanglement resources. (Left) A sender device communicates a quantum state to the receiver device. (Right) Devices $A$ and $B$ share an entangled quantum state.

### *7.3.1 Quantum Communication*

**Definition 51. Quantum Communication:** A dynamic resource corresponding to a single use of a quantum channel in which a quantum state $\rho \in D(\mathcal{H}_d^{\text{Tx}})$ is sent to a receiver using a CPTP map $\mathcal{N} \in \text{CPTP}(\mathcal{H}_d^{\text{Tx}}, \mathcal{H}_{d'}^{\text{Rx}})$ a $\rho' = \mathcal{N}(\rho)$ where $\rho' \in D(\mathcal{H}_{d'}^{\text{Rx}})$.

Quantum communication is a dynamic resource because it involves the transmission of a physical quantum message from one device to another. This quantum system can either be a pure state $|\psi\rangle \in S(\mathcal{H}_d)$ or a density operator $\rho \in D(\mathcal{H}_d)$. The medium over which the message is sent is a quantum channel modeled by a CPTP map as in Definition 37. In general the quantum channel is noisy, however, noiseless quantum communication is important to consider when comparing quantum and classical communication resources.

To quantify the amount of communication, we consider the noiseless case in which the quantum channel is the identity $\text{id} : D(\mathcal{H}_d^{\text{Tx}}) \to D(\mathcal{H}_d^{\text{Rx}})$. Then the amount of communication is quantified by the signaling dimension of the quantum channel $\kappa(\text{id}) = d$, which is the dimension of the Hilbert space used for communication. We adopt this terminology from our work in Ref. [114] and elaborate upon the signaling dimension in Chapter 14. The advantage of the signaling dimension is its correspondence with classical signaling where the classical communication capacity of a quantum channel cannot exceed its signaling dimension [115], [173]. That is, a qubit of communication cannot communicate any more information than a bit of communication. Thus, the signaling measures these to resources' ability to signal classical information, which is a convenient device-independent way to approach an unknown communication resource.

### *7.3.2 Quantum Entanglement*

In the quantum network DAG, quantum entanglement resources are depicted as elliptical nodes that have no classical or quantum inputs. These entanglement producing devices are commonly referred to as entanglement sources or, more simply, sources.

Quantum entanglement is a static resource similar to shared randomness. An entangled system can be considered as two or more quantum particles that originate at the same source and are distributed to different network devices. The network devices may use the entanglement to correlate their information, teleport quantum states, or boost the amount of communication. In the following, we define entanglement for mixed and pure state systems. Each networking device has its own subsystem of quantum state space $S(\mathcal{H}_{d_k}^{A_k})$ where the entangled state is defined on the composite system of quantum networking devices $S(\bigotimes_k \mathcal{H}_{d_k}^{A_k})$.

**Definition 52. Pure State Entanglement:** A composite quantum pure state $|\psi\rangle \in S(\bigotimes_k \mathcal{H}_{d_k})$ is *entangled* if and only if $|\psi\rangle$ cannot be written as a tensor product state of its subsystems. An entangled state $|\psi\rangle$ must

satisfy

$$|\psi\rangle \notin \left\{ |\phi\rangle \in S(\bigotimes_k \mathcal{H}_{d_k}) \;\middle|\; |\phi\rangle = \bigotimes_k |\phi_k\rangle \quad \text{where} \quad |\phi_k\rangle \in S(\mathcal{H}_{d_k}) \;\forall\; k \right\}. \tag{7.5}$$

**Definition 53. Mixed State Entanglement:** A composite quantum mixed state $\rho \in D(\bigotimes_k \mathcal{H}_{d_k})$ is *entangled* if and only if $\rho$ cannot be written as a convex combination of tensor products of mixed states on its subsystems, that is, $\rho$ must satisfy

$$\rho \notin \text{Conv}\left( \left\{ \sigma \in D(\bigotimes_k \mathcal{H}_{d_k}) \;\middle|\; \sigma = \bigotimes_k \sigma_k \quad \text{where} \quad \sigma_k \in D(\mathcal{H}_{d_k}) \;\forall\; k \right\} \right). \tag{7.6}$$

Quantum entanglement is a stronger resource than shared randomness because entanglement can simulate shared randomness. For example, consider entangled state $|\psi\rangle = \cos(\theta)\,|00\rangle + \sin(\theta)\,|11\rangle$ shared between device $A$ and $B$. If each device measures its qubit in then two outcomes 00 and 11 occur with probability $P(00) = \cos^2(\theta) = |\,\langle 00|\psi\rangle\,|^2$ and $P(11) = \sin^2(\theta) = |\,\langle 11|\psi\rangle\,|^2$, which is equivalent to the two devices sharing the result of a biased coin flip. However, a shared coin flip between two devices is actually represented by the mixed state

$$\rho = \cos^2(\theta)\,|00\rangle\langle 00| + \sin^2(\theta)\,|11\rangle\langle 11|\,, \tag{7.7}$$

which is not entangled because it lies in the convex hull of separable mixed states in Def. 53. When measured in the computational basis, the devices obtain the either 00 or 11 with probability $\cos^2(\theta) = P(00)$ and $\sin^2(\theta) = P(11)$. The fact that shared coin flip is represented by a mixed state in quantum systems where the EPR pair is pure indicates that entanglement may bear some operational advantage over shared randomness.

Entanglement can be quantified similarly to shared randomness where the signaling dimension is used to quantify the minimum amount of quantum communication needed to send a network device its portion of the entangled state. For instance, the $N$-qubit entangled state $|\psi\rangle \in \bigotimes_{k=1}^{N} \mathcal{H}_{d_k}$ requires a channel of signaling dimension $\kappa = \min_k d_k$, where it is assumed that each device gets a single particle.

### *7.3.3 Local Operations and Classical Communication*

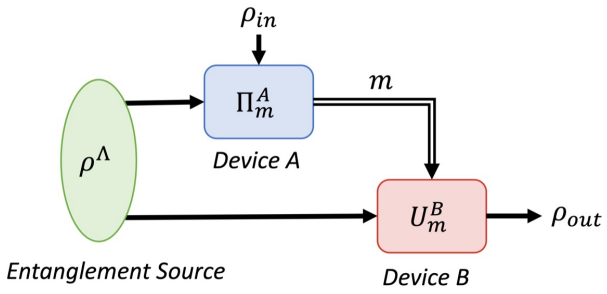



Figure 7.4: A quantum teleportation protocol. Devices $A$ and $B$ share a maximally entangled two-qubit state $|\Phi^+\rangle$. An unknown input qubit state $\rho_{in}$ can be communicated from device $A$ to device $B$. In the protocol device $A$ applies a two-qubit Bell measurement on the input state and their entangled qubit and sends the two-bit measurement result $m$ to device $B$ using classical communication. Device $B$ can then apply to their entangled qubit a local unitary operation $U_m^B$ that is conditioned upon $m$. If performed without error, the input state is "teleported" to the output such that $\rho_{out} = \rho_{in}$.

Local operations and classical communication (LOCC) is an important quantum networking resource in which classical communication assists the quantum network devices. Within the LOCC framework, a sender device $A$ measures its local quantum state $\rho^A$ to obtain a classical value $m$ with probability

$P(m|x) = \mathrm{Tr}\left[\Pi^A_{m|x}\rho^A\right]$. The measurement result is then transmitted to a receiver device $B$ that conditions its local operation $\mathcal{E}^B_m(\rho^B)$ on the received message $m$. Thus, LOCC allows devices to coordinate their operations.

LOCC is a natural framework for quantum networks because classical communications are already necessary for controlling quantum devices. Although classical communication is a trivial task, incorporating classical communication into a quantum network's causal flow can be a challenging problem in precision and control, especially if there are quantum memory limitations. One important LOCC protocol is quantum teleportation because it shows how two-bits of classical communication and entanglement between sender and receiver is sufficient to transmit an unknown quantum state. That is, quantum teleportation establishes a qubit channel from the sender to the receiver. Teleportation is interesting because it exemplifies how two communication resources, entanglement and classical communication, combine to simulate the stronger resource of quantum communication.

LOCC operations can be modeled as a quantum channel $\mathcal{E}^{\mathrm{LOCC}} \in \mathrm{CPTP}(\mathcal{H}^A \otimes \mathcal{H}^B \to \mathcal{H}^{B'})$ where

$$\mathcal{E}^{\mathrm{LOCC}}(\rho^{AB}) = \sum_m \mathrm{Tr}_A\left[\Pi^A_m \otimes \mathcal{E}^{B\to B'}_m(\rho^{AB})\right] = \sum_m \mathcal{E}^{B\to B'}_m(\rho^B_m)P(m) \tag{7.8}$$

and

$$\rho^B_m P(m) = \mathrm{Tr}_A\left[\Pi^A_m \otimes \mathbb{I}^B \rho^{AB}\right]\frac{P(m)}{P(m)}, \quad \text{and} \quad P(m) = \mathrm{Tr}\left[\Pi^A_m \otimes \mathbb{I}^B \rho^{AB}\right]. \tag{7.9}$$

In this setting, the measurement $\Pi^A_m$ is assumed to be destructive such that the quantum state on system $A$ is lost upon measurement.

It is important to note that our channel model of LOCC resources is very general. There are two edge cases that may seem to differ, but these edge cases can still be accommodated by the model in Eq. (7.8). First, suppose that a quantum state is prepared based upon the communicated measurement result $m$. Then, $\rho^B$ would simply correspond to $|0\rangle\langle 0|$ and $\mathcal{E}_m(|0\rangle\langle 0|) = \rho_m$ would initialize the desired state preparation from the zero state. Second, it is convenient within our quantum network DAG for the measurements to be performed in the final layer. Although it seems that the measurement $\Pi^A$ must be applied to obtain the classical result, this is not actually the case. Indeed, the *deferred measurement principle* [61] can be applied where the classical communication is replaced with a controlled quantum operation and the measurement is deferred until the last layer.

## 7.4 Noisy Quantum Network Behaviors

The DAG representation of a quantum network is a mathematically explicit description of a tensor network describing the behavior of the quantum network. When contracted, the tensor network computes conditional probabilities and expectations. In Eq. (7.3), we give a general three-layer decomposition of the quantum network, however in this section, we give a more complete model how the operations in each layer are constructed.

First, the preparation layer initializes the state

$$\rho^{\mathrm{Net}}_{\vec{x}^A} \equiv \bigotimes_{j=1}^{|\Lambda|} \rho^{\Lambda_i} \otimes \bigotimes_{i=1}^{|A|} \rho^{A_i}_{x^{A_i}} \quad \text{where} \quad \vec{x}^{\mathrm{Prep}} = (x^{A_i})_{i=1}^{|A|}, \quad \rho^{\Lambda_i} \in D(\mathcal{H}^{\Lambda_i}), \quad \text{and} \quad \rho^{A_i}_{x^{A_i}} \in D(\mathcal{H}^{A_i}). \tag{7.10}$$

Since each preparation device initializes an independent quantum state, the whole preparation layer belongs

to preparation layer Hilbert space $\mathcal{H}^{\text{Prep}} = \mathcal{H}^{\Lambda} \otimes \mathcal{H}^{A}$, which is the tensor product of the entanglement source layer $\Lambda$ and preparation device layer $\mathcal{H}^{A}$. Note that a similar definition can be made for the case of pure states

$$\left|\psi^{\text{Net}}_{\vec{x}^{A}}\right\rangle \equiv \bigotimes_{j=1}^{|\Lambda|} \left|\psi^{\Lambda_i}\right\rangle \otimes \bigotimes_{i=1}^{|\vec{A}|} \left|\psi^{A_i}_{x^{A_i}}\right\rangle. \tag{7.11}$$

The processing layer(s) sit between preparation and measurement and can be modeled generally as a quantum channel $\mathcal{E}^{\text{Net}}_{\vec{x}^{\vec{B}}}$ or as a unitary $U^{\text{Net}}_{\vec{x}^{\vec{B}}}$. A given processing layer labeled using index $B$ is then

$$\mathcal{E}^{\vec{B}}_{\vec{x}^{\vec{B}}} = \bigotimes_{j=1}^{|\vec{B}|} \mathcal{E}^{B_j}_{x^{B_j}} \tag{7.12}$$

where $\mathcal{E}^{B_j}_{x^{B_j}} \in \text{CPTP}(\mathcal{H}^{B_j}_{in}, \mathcal{H}^{B_j}_{out})$ is the operation applied by each device in the layer. Note that the quantum channel $\mathcal{E}^{B_j}_{x^{B_j}}$ could be replaced with a unitary $U^{B_j}_{x^{B_j}}$. Furthermore, it might be the case that a wire passes through a layer without being operated upon. In this case, the identity map $\rho = \text{id}(\rho)$ is applied to preserve the quantum state. Suppose now that second processing layer $\vec{B}'$ is considered. Then the two layers combine as

$$\mathcal{E}^{\text{Net}}_{\vec{x}^{\text{Proc}}} = \mathcal{E}^{B'}_{\vec{x}^{\vec{B}'}} \circ \text{id}^{\vec{B} \to \vec{B}'} \circ \mathcal{E}^{B}_{\vec{x}^{\vec{B}}} \tag{7.13}$$

where the identity map $\text{id}^{\vec{B} \to \vec{B}'}$ is needed to ensure that the quantum states are wired correctly between the two layers.

At the end of the network DAG is a measurement layer, which we label as $C$. The measurement layer decomposes as a collection of independent measurement devices

$$\Pi^{\text{Net}}_{\vec{y}^{\vec{C}}|\vec{x}^{\vec{C}}} \equiv \bigotimes_{i=1}^{|\vec{C}|} \Pi^{C_i}_{y^{C_i}|x^{C_i}} \quad \text{where} \quad \Pi^{\text{Net}}_{\vec{y}^{\vec{C}}|\vec{x}^{\vec{C}}} \in POVM(\bigotimes_{i=1}^{|\vec{C}|} \mathcal{H}^{C_i}), \quad \text{and} \quad \Pi^{C_i}_{y^{C_i}|x^{C_i}} \in POVM(\mathcal{H}^{C_i}). \tag{7.14}$$

Note that $\vec{x^{C}} = (x^{C_i})_{i=1}^{|C|}$ and $\vec{y^{C}} = (y^{C_i})_{i=1}^{|C|}$ are the inputs and outputs for the measurement devices. in the case where rank-one projective measurements are applied, then the POVM operators $\Pi^{C_i}_{y^{C_i}|x^{C_i}}$ become $\left|\phi_{y^{C_i}|x^{C_i}}\right\rangle\!\left\langle\phi_{y^{C_i}|x^{C_i}}\right|^{C_i}$.

If the network does not contain any noise in its communication, then

$$P(\vec{y}^{C}|\vec{x}^{A}, \vec{x}^{B}, \vec{x}^{C}) = \text{Tr}\left[\Pi^{\text{Net}}_{\vec{y}^{C}|\vec{x}^{C}} \text{id}^{B \to C} \circ \mathcal{E}^{\text{Net}}_{\vec{x}^{\text{Proc}}} \circ \text{id}^{A \to B}(\rho^{\text{Net}}_{\vec{x}^{A}})\right] \tag{7.15}$$

where the identity channels $\text{id}^{A \to B}$ and $\text{id}^{B \to C}$ are needed to correctly map each communicated state to its measurement device.

When a noise node is placed on a quantum communication link, it implies that the communication is noisy such that the input state $\rho$ is taken to a noisy state $\widetilde{\mathcal{N}}$ as $\widetilde{\rho} = \mathcal{N}(\rho)$. Noise can also be incorporated into the model for a device's operations. During a shot of the network, we assume that the noise is static and unchanging. Furthermore, to maintain node labeling consistency between the noisy and noiseless DAGs, noise layers can be place between device layers, where noise layers can be labeled with a $\prime$ symbol on the alphabetic

superscript that indexes the layer. Thus, a layer of noise after the preparation layer $\vec{A}$ would decompose as

$$\mathcal{N}^{\vec{A}'} = \bigotimes_{i=1}^{|\vec{A}'|} \mathcal{N}^{A_i} \tag{7.16}$$

where $\mathcal{N}^{\vec{A}'}$ should be thought of as a processing layer in the network. As such the identity map $\mathrm{id}^{\vec{A}\to\vec{A}'}$ is needed to map the preparation layer $\vec{A}$ to the noise layer $\vec{A}'$ and the identity map $\mathrm{id}^{\vec{A}'\to\vec{B}}$ is needed to map the output of the noise layer to the input of layer $\vec{B}$.

Overall, a given quantum network DAG, noise model, and set of free operations at each device dictate the precise decomposition for the network behavior $\mathbf{P}^{\mathrm{Net}}$ from Def. 44 and the corresponding set of quantum network behaviors $\mathcal{Q}^{\mathrm{Net}}$ from Def. 45. While an explicit mathematical description can always be obtained from the DAG, the simplest representation for the network is by its DAG.

## *7.5 Simulating Noisy Quantum Communication Networks*

It is important that noisy quantum networks can be simulated so that they can be studied and designed prior to their implementation In the previous sections, we introduce quantum networks and characterize them by their black-box behaviors. Recall from Section 2.4 that a black-box behavior is simulated if it is reproduced within a tolerable amount of error. Thus, a quantum network is simulated if its behavior $\mathbf{P}^{\mathrm{Net}}$ can be produced by another system, such as a classical or quantum computer. Indeed, the quantum network model developed so far can easily be extended to a quantum circuit model for simulation on a quantum computer.

In this work, we develop a quantum simulation framework for noisy quantum networks that can be run on either classical or quantum hardware. We implement our approach in the Python package qNetVO: the Quantum Network Variational Optimizer [108]. In this package, we extend the PennyLane software [210] with tools for constructing quantum circuits that simulate a given noisy quantum network's DAG. Since PennyLane provides a hardware-agnostic circuit model, our software can easily be applied on a range of classical simulators and existing quantum computing platforms. In the rest of this section, we elaborate upon our framework for simulating quantum networks.

### *7.5.1 A Quantum Network as a Quantum Circuit*

As discussed in Chapter 5, a quantum circuit is a DAG that explicitly describes a sequence of quantum operations applied to an array of qubit subsystems on a quantum computer. Likewise, the DAG representation of a quantum network introduced in Chapter 7 describes a quantum network in an analogous manner. Throughout this section, we will show that there is a direct correspondence between the two DAGs, meaning that any quantum network DAG corresponds to an explicit quantum circuit DAG (see Fig. 7.5). Thus, we construct a procedure for transforming a quantum network DAG into a quantum circuit that simulates the network on any quantum computing hardware or classical simulator.

The quantum circuit simulation describes network using a realistic model where physical communication resources and device operations correspond directly to quantum circuit components that simulate those resources and operations. Hence the constraints on locality and communication resources are embedded implicitly into the structure of the quantum circuit. Furthermore, if the dynamics of a quantum network are encoded properly into a quantum circuit and that circuit is executed faithfully without noise, then an exact

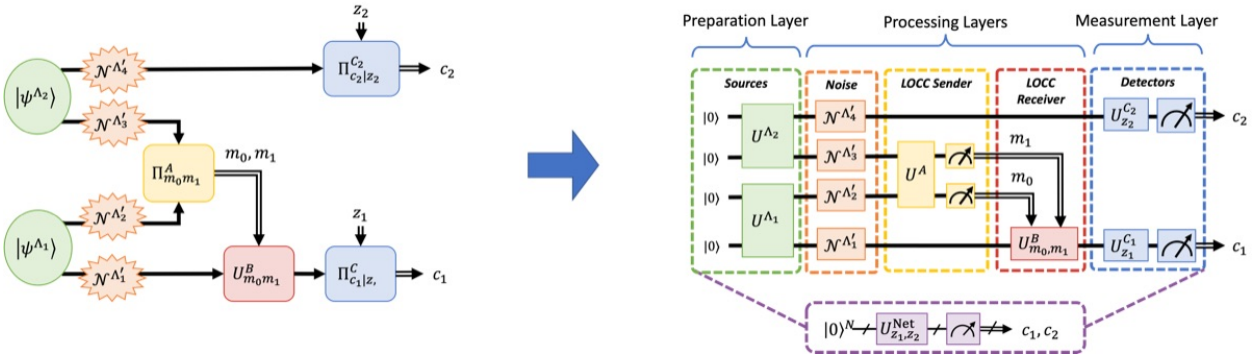



Figure 7.5: Quantum circuit that simulates an entanglement swapping scenario. (Left) A DAG depicting the quantum communication network with two entanglement sources (green), a qubit noise model (orange), an LOCC sender device (yellow), an LOCC receiver device (red), and a measurement device (Blue). (Right) The same network is depicted by quantum circuit.

simulation of the quantum network is achieved, meaning that the resulting network $\mathbf{P}^{\text{Net}}$ is indistinguishable from the real quantum network behavior.

Since a quantum network is constructed modularily from quantum network devices, the network's simulations circuit simulation also admits a modular construction. That is, the processing applied by each device is described by a unitary operation $U_x^{\text{Dev}} \in U(\mathcal{H}^{\text{Dev}})$ that is conditioned upon the device's classical input $x \in \mathcal{X}^{\text{Dev}}$. Note that $\mathcal{H}^{\text{Dev}}$ corresponds to the device's local Hilbert space where the number of qubits operated upon by Dev is $N^{\text{Dev}} = \log_2 \left[\text{Dim}\left(\mathcal{H}^{\text{Dev}}\right)\right]$. The unitary operation then serves as a gate in the quantum circuit model and must be expressed in terms of fundamental quantum circuit gates.

Each wire in the quantum circuit model corresponds to a qubit of communication linking two unitaries that model network devices. The $N$-qubits in the network are labeled using the natural numbers $q \in [N]$. The qubits local to a networking device Dev are notated $\vec{q}^{\text{Dev}} \subseteq [N]$ and these qubit wires are operated upon by the unitary operator $U_x^{\text{Dev}}$.

A quantum network's simulation circuit can then be constructed modularily its constituent quantum networking devices. The construction will mirror the one used in Chapter 7. The quantum networking devices are grouped into sequential layers where no devices in a given layer operate upon the same qubits

$$\bigcap_{\text{Dev} \in \text{Lay}} \vec{q}^{\text{Dev}} = \emptyset, \tag{7.17}$$

implying independence between the networking device. Each layer describes a time-step in the network simulation and is ordered and indexed using capitalized alphabetic characters. The events in two sequential layers operating on a given qubit are said to be time-like separated. Furthermore, devices within the same layer are labeled using the alphabetic character with a numeric subscript. Devices within the same layer are independent from each other and are said to be space-like separated. Finally, any qubits that are not operated upon in a given layer can be understood as the identity operation.

In the sections that follow, we elucidate the details of the quantum circuits that simulate quantum networking devices.

### 7.5.2 Simulating Network Dynamics

Formally, the quantum network circuit is represented by the $N$-qubit unitary operator $U_{\vec{x}}^{\text{Net}} \in \mathbb{C}_2^{\otimes N}$ where $\vec{x} \in \mathcal{X}$ is the set of classical inputs that condition the network's dynamics. Since the network decomposes into sequential layers of quantum networking devices, the unitary operator admits the decomposition

$$U_{\vec{x}}^{\text{Net}} = \prod_{\text{Lay} \in \text{Net}} U_{\vec{x}^{\text{Lay}}}^{\text{Lay}} = \ldots U_{\vec{x}^C}^C U_{\vec{x}^B}^B U_{\vec{x}^A}^A U^\Lambda = U_{\vec{x}^{\text{Meas}}}^{\text{Meas}} U_{\vec{x}^{\text{Proc}}}^{\text{Proc}} U_{\vec{x}^{\text{Prep}}}^{\text{Prep}} \tag{7.18}$$

where each iterated layer's unitary evolves the quantum network's state by one time step and the operations applied in each layer may be conditioned on its own classical inputs $\vec{x}^{\text{Lay}}$.

The network's dynamics are simulated by evolving the zero state as $U_{\vec{x}}^{\text{Net}} |0\rangle^{\otimes N}$. The resulting quantum state is then measured in the computational basis $\{|\vec{z}\rangle\}_{\vec{z} \in \mathbb{B}^M}$ where $M \leq N$ are the measured qubits. The measured quantum state is then the reduced density matrix

$$\rho_{\vec{x}}^{\text{Meas}} = \text{Tr}_{\text{Env}} \left[ U_{\vec{x}}^{\text{Net}} |0\rangle\langle 0|^{\otimes N} \left( U_{\vec{x}}^{\text{Net}} \right)^\dagger \right] \tag{7.19}$$

where the traced out subsystem Env denotes the unmeasured $(N - M)$ qubit subspace. The resulting measurement probabilities are then

$$P(\vec{z}|\vec{x}) = \text{Tr} \left[ |\vec{z}\rangle\langle\vec{z}| \, \rho_{\vec{x}}^{\text{Meas}} \right]. \tag{7.20}$$

The black-box behavior characterizing the quantum circuit is then given as

$$\mathbf{P}^{\text{QC}} = \sum_{\vec{z} \in \mathbb{B}^M} \sum_{\vec{x} \in \mathcal{X}} \text{Tr} \left[ |\vec{z}\rangle\langle\vec{z}| \, \rho_{\vec{x}}^{\text{Meas}} \right]. \tag{7.21}$$

It is often the case that a quantum network does not output an $M$-bit string. In this case, a classical postprocessing map $\mathbf{L} : \mathbb{B}^M \to \mathcal{A}$ is needed to map the $M$ bit output of the quantum circuit to the $|\mathcal{A}|$ outputs of the quantum network. That is, consider the quantum network behavior $\mathbf{P}^{\text{Net}} : \mathcal{X} \to \mathcal{A}$ and the simulation circuit behavior $\mathbf{P}^{\text{QC}} : \mathcal{X} \to \mathbb{B}^M$, then

$$\mathbf{P}^{\text{Net}} = \mathbf{L}\mathbf{P}^{\text{QC}} \tag{7.22}$$

where $\mathbf{P}^{\text{QC}}$ is defined in Eq. (7.21). The post-processing map $\mathbf{L}$ is a column stochastic matrix

$$\mathbf{L} = \sum_{\vec{a} \in \mathcal{A}} \sum_{\vec{z} \in \mathbb{B}^M} P(\vec{a}|\vec{z}) \, |\vec{a}\rangle\langle\vec{z}|, \tag{7.23}$$

and is therefore a black-box behavior. A quantum computer can be used to construct the network behavior $\mathbf{P}^{\text{Net}}$ by repeatedly running the quantum circuit to estimate the conditional probabilities $P(\vec{a}|\vec{x})$ across all inputs $\vec{x} \in \mathcal{X}$. We now discuss the layered decomposition of $U_{\vec{x}}^{\text{Net}}$ in more detail.

### 7.5.3 Evaluating Quantum Circuits

To this point, we have only described the quantum network model. To simulate the quantum network's behavior, we quantum network circuit must be evaluated. A quantum circuit can either be run on quantum hardware such as a quantum computer or a classical simulation software the reproduces the behavior of the quantum computer.

On a classical simulator, the quantum circuit's probabilities can be simulated in the absence of noise, however, simulations of quantum systems cannot scale to large systems with many qubits. Many simulation software are available and can be run on a laptop computer, or on remote classical simulation environments. Through PennyLane, the qNetVO software can simulate quantum networks in any simulator supported by PennyLane. In this work, we use PennyLane's `default.qubit` state vector simulator and the `default.mixed` mixed state simulator. Both simulations rely on a qubit based computing architecture.

The main advantage of simulating a quantum network on quantum hardware is the promise of efficient computation. In principle, quantum computers ought to be able to simulate quantum systems with greater efficiency than classical computers [18], [19]. This improved efficieny is important because at some point, quantum systems become so complex that they can no longer be efficiently simulated or stored in memory on a classical computer. On a quantum computer however, the quantum system is simulated by the natural evolution of the wavefunction.

To demonstrate this proof of concept we provide a few examples of evaluating simulating quantum networks on IBM quantum hardware. However, it is important to acknowledge that existing noisy intermediate-scale quantum (NISQ) computers may not provide significant simulation advantages [38]. In many cases, it is beneficial to perform simulations on noiseless qubits that are simulated classically than it is to use noisy qubits, even if it means that fewer qubits can be considered. The qNetVO framework offers the flexibility to simulate the quantum system classically, hence it bridges the gap between classical and quantum simulation tools for quantum networks. This feature is convenient because the quantum networks being developed are still small enough to be simulated and optimized on a classical computer. Therefore, the approaches we introduce in this work can be applied in scientific discovery and engineering design.

In general, when simulating quantum networks on NISQ computers, the noise present on the computer hardware may not resemble the noise in the quantum network hardware. When simulating the network, our objective is to precisely reproduce a noisy quantum network's behavior. However, evaluating a simulation circuit on noisy quantum hardware adds the quantum computer's noise into the simulation, which may not match the network's noise model. Thus, to achieve precise noise modeling, we must rely upon classical simulators or fault-tolerant quantum computers. Additionally, we accurate simulation requires accurate characterization of the noise model of the physical quantum system that is under investigation.

# Chapter 8

# Noise in Quantum Communication Networks

## *8.1 Limitations of Noisy Quantum Networks*

In general, quantum systems are noisy and quantum networks are no exception. Noise in quantum networks can be found in all processes including preparation, processing, communication, and measurement. The noise is very detrimental to the implementation of quantum networking applications leading to key limitations and restrictions on the capabilities of quantum networks in the foreseeable future.

From a quantum theory perspective, noise is modeled as a quantum channel $\mathcal{N} \in \text{CPTP}$ as discussed in Chapter 5. Thus, the behavior of quantum network systems can be evaluated explicitly for any noise model considered in the network. However, simulating noise can be challenging because CPTP maps require either ancillary qubits, or a mixed state simulation of the system. In either case, simulating noisy quantum processes adds a significant overhead to the already difficult task of simulating quantum systems.

From an experimental perspective, noise in quantum hardware limits its performance and efficiency, leading to fundamental restrictions on the quantum systems that can be constructed in practice. Namely, these restrictions lead to a plethora of features that are characteristic of near-term quantum networks. We summarize these key limitations below.

- **Small Hilbert Space Dimension:** In the presence of noise, it is difficult to control large arrays of qubits or high-dimensional quantum states. As a result, noisy quantum networks will be limited in the number of qubits they can utilize.
- **Low Circuit Depth:** In the presence of noise, it is difficult to perform extensive processing on quantum systems because the error grows exponentially with the number of operations. As a result, noisy quantum networks will be limited in the number of operations that can be applied to quantum states between preparation and measurement.
- **Short Quantum Memory Times:** In the presence of noise, it is difficult to preserve a quantum states in memory. As a result, noisy quantum networks will either be memoryless or have short-term storage capabilities.

- **Detection Efficiency:** In the presence of noise, the rate at which quantum states are lost is large leading to a decline in the rate of detection of quantum states. Therefore the shot rate of noisy networks will be low. As a result, the efficiency of quantum networks will be diminished, limiting their ability to scale.

Given the restrictions of noisy quantum networks, their functionality will be limited to the prepare-and-measure setting, in which quantum communication is available between devices, but extensive processing and memory are not. The prepare-and-measure setting is a type of discrete memoryless setting, allowing our general black-box framework to accurately describe the behavior of these networks. Furthermore, given the limitations on detection efficiency and Hilbert space dimension, near-term quantum networks will also be relatively small in scale consisting of only a handful of devices that each contain only a few qubits.

While these limitations may be disheartening, they also make studying such systems more tractable from a theoretical perspective. Recall that large-scale, complex quantum systems are generally very difficult to simulate and optimize. However, if we restrict the Hilbert space dimension and the circuit depth of these networks, the complexity of these systems can be decreased considerably, leading tractable simulation and optimization on a classical computer. Although, if quantum networks can be simulated classically, it is very difficult to argue that they can provide much advantage. Nevertheless, understanding the theory of these rudimentary noisy quantum networks will lead to important insights that will help these systems scale.

## *8.2 Loss in Quantum Systems*

In quantum communication systems, loss is one of the most prevalent types of noise. In general, loss describes a noisy process where a quantum state may have been input into a system, but never makes it to the output. A common example can be found for photonic qubits in which a single photon encodes quantum information. As the photon propagates through a medium, such as a fiber optic cable, the photon has a certain probability of getting absorbed by a particle in the medium. Hence the photon and its encoded quantum information never make it to the output of the fiber. Indeed, as the length of the fiber, or optical medium, increases, the loss scales as $I(L) = I_0 e^{-\alpha L}$ where $\alpha \geq 0$ is a constant that depends on the medium and frequency of light, $L$ is the length, and $I_0$ is the intensity of light at the input. This exponential scaling in loss is one of the key factors limiting the development of long-distance quantum communications.

While it is crucial to understand how to build quantum systems robust to loss, the way that loss is modeled depends on the physical implementation of the quantum system. Within our operational framework loss can take on a few different forms. Namely, the way loss is treated is different if the errors are flagged or unflagged. If an error is flagged, given one shot of the experiment, it can be determined whether the sampled output has an error. The flag can be thought of a dedicated output value or set of values indicating that an error occurred. On the other hand, it cannot be determined whether or not an unflagged error occurred in an experiment given a single shot of data.

In most settings, flagged errors are preferable because they allow for either error correction to be applied or for the erroneous data samples to be thrown away. Quantum error correction, unfortunately, is generally resource intensive meaning that it is unlikely to be available in the noise limited quantum networks being discussed in this work [52]. On the contrary, classical error correction techniques are feasible where postprocessing can be combined with knowledge of the physical system in attempt to make an educated guess to what the result would have been without error, but such techniques are limited in their ability to function in a black-box setting. Finally, the simplest approach is to ignore the flagged errors altogether. In

doing so, the unflagged sampled data will perform its task more precisely, however, the detriment is that high error-rates will only serve to decrease the efficiency proportionally.

In practice, it might not always be possible to flag errors due to physical constraints, or the sheer complexity of the network. If this is the case, then error mitigation strategies could be applied in which knowledge of the physical quantum system and its noise are used to make the system more robust or apply improved postprocessing. However, without having direct access to the quantum devices and their underlying physical systems, error mitigation reduces to the task of finding the device settings that perform best in the given noisy network.

## *8.3 The Noise of Relative Unitary Rotations*

A very important type of noise in a network can be compared to a calibration task. In particular consider a quantum device $A$ that wants to send a qubit to a quantum device $B$, however, to do so they must agree upon their bases used to encode and decode information. The problem is that the two devices $A$ and $B$ may not agree upon the orientation of the zero state, and the communication medium may apply some unknown unitary transformation to the communicated states as $P(y|x) = \text{Tr}\left[|0\rangle\langle 0| U_{\mathcal{N}} |0\rangle\langle 0| U_{\mathcal{N}}^{\dagger}\right]$. To calibrate their communication channel, the devices need to perform some optimization over their quantum channel to ensure that it is set up appropriately to remove the unwanted transformation $U_{\mathcal{N}}$. While this example of noise is very simple, failure to perform a proper calibration between all the networking devices will be extremely detrimental to a protocol. Furthermore, the fact that transformation applied by the communication medium may not be static and that each device's bases might have a relative rotation with respect to the other devices implies that calibration is will be a constant task in quantum networks. As a result, automated methods will be needed to calibrate quantum networking devices.

## *8.4 Common Quantum Noise Models*

It is not feasible to consider all types of noise that could affect a quantum network's communication and processing resources. However, there are some common types of error that can be modeled in a network to gauge the general affect of noise on in a network. We break noise models into two categories, unital and nonunital, in which the preserves the identity, but the latter does not. In each case, we list a handful of important noise models cover a broad of common types of noise. Furthermore, we place an emphasis on qubit noise models because of their simplicity and prevalence in qubit communication scenarios.

### *8.4.1 Unital Noise Models*

Unital channels model common types of hardware noise such as depolarizing and dephasing. A quantum channel $\mathcal{U}$ is unital if and only if it satisfies $\mathcal{U}(\mathbb{I}) = \mathbb{I}$. Consequently, unital noise cannot improve the purity of a general quantum state $\rho$, that is, $\text{Tr}\left[\mathcal{U}(\rho)^2\right] \leq \text{Tr}\left[\rho^2\right]$.

In the qubit case, the theory of unital channels is well known [211]. A unital qubit channel $\mathcal{U}$ can be expressed in Pauli basis as a $3 \times 3$ matrix

$$M_{\mathcal{U}} = \sum_{i,j \in \{x,y,z\}} \frac{1}{2} \text{Tr}\left[\sigma_j \mathcal{U}(\sigma_i)\right] |j\rangle\langle i| \,, \tag{8.1}$$

where the operator $M_{\mathcal{U}}$ is applied to the Bloch vector $\vec{s}$ of the input state $\rho$ as $\vec{r} = M_{\mathcal{U}}\vec{s}$ to produce $\vec{r}$ the Bloch vector of the output states $\mathcal{U}(\rho)$. An important property of unital qubit channels is that the matrix $M_{\mathcal{U}}$ can always be diagonalized using rotations $R_{\text{in}}$, $R_{\text{out}} \in SO(3)$ such that

$$R_{\text{out}} M_{\mathcal{U}} R_{\text{in}} = \text{diag}(u_0, u_1, u_2) = M_{\mathcal{U}^\star}, \tag{8.2}$$

where, without loss of generality, the singular values are bound as $1 \geq u_x, u_y, u_z \geq 0$ and can be permuted along the diagonal of $M_{\mathcal{U}^\star}$.

By the isometry between $SO(3)$ and $SU(2)$ [212], the rotation operators $R_{\text{in}}$ and $R_{\text{out}}$ correspond directly to unitaries $V_{\text{in}}$ and $V_{\text{out}} \in SU(2)$ applied to $\mathcal{U}$ as

$$\mathcal{U}^\star(\rho) = V_{\text{out}}\mathcal{U}(V_{\text{in}}\rho V_{\text{in}}^\dagger)V_{\text{out}}^\dagger, \tag{8.3}$$

where $u_i\sigma_i = \mathcal{U}(\sigma_i)$ is an *eigenoperator* of the unital channel $\mathcal{U}$. We now describe some common examples of unitary noise.

**Definition 54. Depolarizing Channel:** $\mathcal{W}_v \in \text{CPTP}(\mathcal{H}_d, \mathcal{H}_d)$, A noise model that mixes the maximally mixed state $\frac{1}{2}\mathbb{I}_d$ with an input state $\rho$ as

$$\mathcal{W}_v(\rho) = v\rho + \frac{(1-v)}{d}\mathbb{I}_d \text{Tr}\left[\rho\right], \tag{8.4}$$

where $v$ is a parameter commonly referred to as the *visibility*. The visibility can be expressed as a noise parameter

$$\gamma = \frac{d^2-1}{d^2}(1-v) \tag{8.5}$$

where if $v \in [0,1]$, then $\gamma \in [0, \frac{d^2-1}{d^2}]$, however, $\mathcal{W}_\gamma^{A\to A'} \in \text{CPTP}(\mathcal{H}^A, \mathcal{H}^{A'})$ for all $\gamma \in [0,1]$.

In the operator-sum representation of a quantum channel, an $N$-qubit depolarizing channel's Kraus operators are expressed as

$$\mathcal{K}(\mathcal{W}_\gamma^{A\to A'}) = \left\{ k_{\vec{x}} \bigotimes_{i=0}^{N-1} \sigma_{x_i} \right\}_{\vec{x}\in\{0,1,2,3\}^N} \quad \text{where} \quad k_{\vec{x}} = \begin{cases} \sqrt{1-\gamma} & \text{if } \vec{x} = (0,\ldots,0) \\ \sqrt{\frac{\gamma}{d^2-1}} & \text{otherwise} \end{cases} \tag{8.6}$$

and $\sigma_{x_i}$ denotes on of the four qubit Pauli operators. Furthermore, in the case of qubit depolarizing noise, the Pauli decomposition of the channel is $M_{\mathcal{W}_v^{A\to A'}} = \text{diag}\left(v, v, v\right)$ where $v$ is the visibility.

**Definition 55. Qubit Pauli Channel:** A family of unital qubit channels having the form

$$\mathcal{U}(\rho) = p_0\rho + p_1\sigma_x\rho\sigma_x + p_2\sigma_y\rho\sigma_y + p_3\sigma_z\rho\sigma_z \quad \text{where} \quad \sum_{i=0}^{3} p_i = 1 \tag{8.7}$$

where $\sum_i p_i = 1$ and $p_i \geq 0$. The vector $\vec{p} = (p_i)_{i=0}^3$ corresponds to a distinct channel and can be interpreted as a point confined to the convex hull of a tetrahedron The Pauli decomposition of the Pauli channel is then

$$M_{\mathcal{U}} = \text{diag}\left(p_0 + p_1 - p_2 - p_3, p_0 - p_1 + p_2 - p_3, p_0 - p_1 - p_2 + p_3\right). \tag{8.8}$$

The family of qubit Pauli channels is incredibly general. Looking at the extreme points of the tetrahedron,

we find that when $p_0 = 1$, the channel is ideal. When $p_1 = 1$, the channel corresponds to a bit-flip. When $p_2 = 1$, the channel corresponds to a phase-flip. When $p_3 = 1$, the channel corresponds to a dephasing channel. Furthermore, the center of the tetrahedron, when $p_i = \frac{1}{4}$ for all $i \in \{0, 1, 2, 3\}$, the channel becomes the completely depolarizing channel.

As an example we will consider a partially dephasing channel that describes the decoherence process as

$$\mathcal{D}_\gamma^{A\to A'}(\rho) = \frac{1+\sqrt{1-\gamma}}{2}\rho + \frac{1-\sqrt{1-\gamma}}{2}\sigma_z\rho\sigma_z, \quad \text{where} \quad \gamma \in [0,1], \tag{8.9}$$

where the channel causes the off-diagonal terms of $\rho$ to go to zero as the noise parameter $\gamma$ increases. In the Pauli decomposition, the partially dephasing channel is then expressed as $M_{\mathcal{D}_\gamma^{A\to A'}} = \text{diag}\left(\sqrt{1-\gamma}, \sqrt{1-\gamma}, 1\right)$. The Kraus operators for the dephasing channel are

$$K_0 = \begin{pmatrix} 1 & 0 \\ 0 & \sqrt{1-\gamma} \end{pmatrix} \text{ and } \quad K_1 = \begin{pmatrix} 0 & 0 \\ 0 & \sqrt{\gamma} \end{pmatrix}. \tag{8.10}$$

In general, a partial bit-flip or partial phase-flip channel can be described similarly to Eq. (13.87) by swapping out $\sigma_z$ for $\sigma_x$ or $\sigma_y$ respectively. However, by Eq. (8.3), local qubit unitary operators are sufficient to convert dephasing channel $\mathcal{D}_\gamma^{A\to A'}$ into a phase-flip or bit-flip channel. Without loss of generality, the results of the dephasing channel can be extended to bit-flip or phase-flip channels.

### *8.4.2 Nonunital Noise Models*

A quantum channel is nonunital if and only if it does not preserve the identity, $\mathcal{N}(\mathbb{I}) \neq \mathbb{I}$, where important examples include amplitude damping channels and replacer channels. It is useful to note that upon local unitary rotations as in Eq. (8.3), a qubit nonunital channel can be represented in the Pauli basis as where

$$\mathbb{T}_{\mathcal{N}_\gamma} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ t_x & u_x & 0 & 0 \\ t_y & 0 & u_y & 0 \\ t_z & 0 & 0 & u_z \end{pmatrix} \tag{8.11}$$

where we write the $3 \times 3$ matrix block on the bottom right as $M_{\mathcal{N}_\gamma} = \text{diag}(u_x, u_y, u_z)$. Furthermore, the matrix elements are evaluated similarly to the case of unital channels in Eq. (8.1), but we now include the qubit identity matrix $\mathbb{I}_2$ in addition to the Pauli operators.

**Definition 56. Qubit Amplitude Damping Channel:** $\mathcal{A}_\gamma \in \text{CPTP}(\mathcal{H}_2 \to \mathcal{H}_2)$, A channel modeling an energy dissipation process where a high energy state $|1\rangle\langle 1|$ transitions into a low energy state $|0\rangle\langle 0|$.The Kraus operators are defined as

$$K_0 = \begin{pmatrix} 1 & 0 \\ 0 & \sqrt{1-\gamma} \end{pmatrix}, \quad K_1 = \begin{pmatrix} 0 & \sqrt{\gamma} \\ 0 & 0 \end{pmatrix} \tag{8.12}$$

where the effect on a qubit density matrix is

$$\mathcal{A}_\gamma(\rho) = \begin{pmatrix} \gamma + (1-\gamma)\rho_{00} & \sqrt{1-\gamma}\rho_{01} \\ \sqrt{1-\gamma}\rho_{10} & (1-\gamma)\rho_{11} \end{pmatrix}. \tag{8.13}$$

In the Pauli basis, the amplitude damping channel is expressed as

$$\mathbb{T}_{\mathcal{A}_\gamma} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & \sqrt{1-\gamma} & 0 & 0 \\ 0 & 0 & \sqrt{1-\gamma} & 0 \\ \gamma & 0 & 0 & (1-\gamma) \end{pmatrix}. \tag{8.14}$$

**Definition 57. Replacer Channel:** $\mathcal{R}_\gamma \in \mathrm{CPTP}(\mathcal{H}_d \to \mathcal{H}_{d'})$, Replaces the input state $\rho \in D(\mathcal{H}_d)$ with the state $\sigma \in D(\mathcal{H}_{d'})$ with probability $\gamma$ such that

$$\mathcal{R}_\gamma(\rho) = (1-\gamma)\rho + \gamma\sigma \mathrm{Tr}\left[\rho\right]. \tag{8.15}$$

Indeed replacer channels represent a broad family of noise models because each distinct replacer state corresponds to a distinct channel. In general, a replacer channel is nonunital if $\gamma > 0$ and $\sigma \neq \frac{1}{d}\mathbb{I}_d$. Thus, the depolarizing channel is a special case of replacer channel because its replacer state $\sigma = \frac{1}{d}\mathbb{I}_d$ allows the channel to be unital.

To express the replacer channel in its operator-sum representation (see Def 39), we first consider the case where the replacer state is pure, $\sigma = |\psi\rangle\langle\psi| \in \mathcal{H}_{d'}$. As a result, the Kraus operators for the replacer channel $\mathcal{R}_\gamma$ are

$$K_i = \sqrt{\gamma}\,|\psi\rangle\langle i| \;\; \forall\, i \in [0,d),\;\; K_d = \sqrt{1-\gamma}\mathbb{I}_{d'\times d}, \tag{8.16}$$

where $\{|i\rangle\}_{i=0}^{d-1}$ form an orthonormal basis and $\mathbb{I}_{d'\times d} = \sum_{j=1}^{d'}\sum_{k=1}^{d}\delta_{j,k}\,|j\rangle\langle k|$ is the identity map from the input to the output. The replacer state can be expressed as a mixed state using the fact that $\sigma = \sum_j c_j\,|\psi_j\rangle\langle\psi_j|$ where $j$ indexes each pure state in the mixture and $c_j \in \mathbb{R}$ is its weight. Then, let $K_{i,j} = \sqrt{\gamma}\sqrt{c_j}\,|\psi_i\rangle\langle i|$ for such that the operator-sum representation becomes

$$\mathcal{R}_\gamma(\rho) = K_d\rho K_d^\dagger + \sum_{i=0}^{d-1}\sum_j K_{i,j}\rho K_{i,j}^\dagger = (1-\gamma)\rho + \gamma\sum_j c_j\,|\psi_i\rangle\langle\psi_i|\,\mathrm{Tr}\left[\rho\right] = (1-\gamma)\rho + \gamma\sigma\mathrm{Tr}\left[\rho\right]. \tag{8.17}$$

Hence we give the construction of a channel whose replacer state is an arbitrary mixed state.

**Definition 58. Colored Noise:** $\mathcal{C}_\gamma \in \mathrm{CPTP}(\mathcal{H}_4 \to \mathcal{H}_4)$, A nonunital two-qubit channel modeling depolarization on a preferred, entangled axis of the state,

$$\mathcal{C}_\gamma(\rho) = (1-\gamma)\rho + \frac{\gamma}{2}(\left|\Psi^+\right\rangle\left\langle\Psi^+\right| + \left|\Psi^-\right\rangle\left\langle\Psi^-\right|), \tag{8.18}$$

where $|\Psi^\pm\rangle = (|01\rangle \pm |10\rangle)/\sqrt{2}$.

Colored noise is typically found on the singlet states $|\Psi^-\rangle = \frac{1}{\sqrt{2}}(|01\rangle - |10\rangle)$ produced by parametric down conversion [213]. Note that colored noise is an example of a replacer channel where the replacer state is $\sigma = \frac{1}{2}(|01\rangle\langle 01| + |10\rangle\langle 10|)$ The Kraus operators for colored noise are

$$K_0 = \sqrt{1-\gamma}\mathbb{I}, \quad K_{\Psi^\pm,\Phi^\pm} = \sqrt{\frac{\gamma}{2}}\left|\Psi^\pm\right\rangle\left\langle\Phi^\pm\right|, \quad \text{and} \quad K_{\Psi^\pm,\Psi^\pm} = \sqrt{\frac{\gamma}{2}}\left|\Psi^\pm\right\rangle\left\langle\Psi^\pm\right|, \tag{8.19}$$

where $|\Phi^\pm\rangle$ and $|\Psi^\pm\rangle$ constitute the Bell basis and all combinations of $\pm$ are considered leading to 9 Kraus operators in total.

**Definition 59. Loss Channel:** $\mathcal{L}_\gamma \in \text{CPTP}(\mathcal{H}_d \to \mathcal{H}_d)$, Models loss of quantum information in an unflagged manner. This noisy process is a special case of replacer channel in which the replacer state is $\sigma = |0\rangle\langle 0|$, therefore, its Kraus operators are

$$K_0 = \sqrt{1-\gamma}\mathbb{I}_d, \quad K_i = \sqrt{\gamma}\,|0\rangle\langle i| \;\; \forall\, i \in [1, d-1]. \tag{8.20}$$

The error is said to be unflagged, because the error state $\sigma$ is used to encode quantum information, meaning that the error's occurrence in one shot of data cannot be determined with certainty. In the Pauli basis the loss channel is

$$\mathbb{T}_{\mathcal{L}_\gamma} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1-\gamma & 0 & 0 \\ 0 & 0 & 1-\gamma & 0 \\ \gamma & 0 & 0 & 1-\gamma \end{pmatrix} \tag{8.21}$$

**Definition 60. Erasure Channel:** $\mathcal{E}_\gamma \in \text{CPTP}(\mathcal{H}_d \to \mathcal{H}_{d+1})$ models a loss process in which the error is flagged. The erasure channel is distinct from the loss channel where, with probability $\gamma$, the output is an error flag $\sigma = |d\rangle\langle d|$ that is orthogonal to the input Hilbert space. The erasure channel is expressed as

$$\mathcal{E}_\gamma(\rho) = (1-\gamma)\mathbb{I}_{d'\times d}\rho\mathbb{I}^T_{d'\times d} + \gamma\,|d\rangle\langle d|\,\text{Tr}\,[\rho] \tag{8.22}$$

where we note that the input $\rho$ is defined on the Hilbert space spanned by $\{|i\rangle\}_{i=0}^{d-1}$ such that $\text{Tr}\,[\sigma\rho] = 0$ and $\mathbb{I}_{d'\times d} = \sum_{j=0}^{d}\sum_{k=0}^{d-1}\delta_{j,k}\,|j\rangle\langle k|$. The error is said to be flagged because the error state $\sigma$ is orthogonal from the rest of the Hilbert space, meaning that the output data from each shot indicates whether or not the state was lost or not.

## *8.5 Detector Errors as Classical Post-Processing*

Noise in quantum systems does not always have to be modeled as CPTP map acting upon a Hilbert space. In principle, classical postprocessing can be applied to the data measured from the quantum system to model certain noisy processes.

**Definition 61. Classical Error Map:** $\mathbf{E}_\gamma : \mathcal{Y} \to \mathcal{Y}'$ where $\mathbf{E}_\gamma \in \mathcal{P}_{\mathcal{Y}'|\mathcal{Y}}$. The classical error map acts upon a network's behavior $\mathbf{P}^{\text{Net}} \in \mathcal{P}_{\mathcal{Y}|\mathcal{X}}$ via matrix multiplication to produce a noisy quantum network behavior

$$\widetilde{\mathbf{P}}^{\text{Net}} = \mathbf{E}\mathbf{P}^{\text{Net}} \quad \text{where} \quad \widetilde{\mathbf{P}}^{\text{Net}} : \mathcal{X} \to \mathcal{Y}'. \tag{8.23}$$

Modeling noise in quantum systems using classical error maps is advantageous because of its computational efficiency. That is, it considerably more efficient to multiply two matrices $\mathbf{E}_\gamma$ and $\mathbf{P}^{\text{Net}}$ than it is to model noise in quantum systems, which requires either an enlarged Hilbert space, or a mixed state simulator to be used. Both expanding the Hilbert space or using a mixed state simulator in place of a state vector simulator adds significant computational overhead to the simulation of a quantum system. However, only certain noise models are able to be represented using classical error maps, limiting the cases in which they can be applied.

**Definition 62. Classical Replacer Error:** $\mathbf{R}_\gamma \in \mathcal{P}_{\mathcal{Y}|\mathcal{Y}}$, a classical error map where each output $y \in \mathcal{Y}$ is replaced by a new output $y'$ with fixed probability $P(y'|y)$ for all $y \in \mathcal{Y}$. The error map is

$$\mathbf{R}_\gamma = (1-\gamma)\mathbb{I}_{|\mathcal{Y}|} + \gamma\mathbf{R} \tag{8.24}$$

where $\mathbf{R} \in \mathcal{P}_{\mathcal{Y}|\mathcal{X}}$ is the replacer behavior that takes the special form $\mathbf{R} = \vec{\mu} \cdot \vec{1}^T$ where $\vec{\mu} = (\mu_1, \ldots, \mu_{|\mathcal{Y}|})^T$, $\sum_y \mu_y = 1$ and $\mu_y \geq 0$. The behavior with replacer noise is then

$$\widetilde{\mathbf{P}} = \mathbf{R}_\gamma \mathbf{P} = (1-\gamma)\mathbf{P} + \gamma \mathbf{R} \tag{8.25}$$

where $\widetilde{\mathbf{P}}, \mathbf{P}, \mathbf{R} \in \mathcal{P}_{\mathcal{Y}|\mathcal{X}}$.

It is important to note that the classical replacer error is defined for a measurement that has no input, however, measurements can be conditioned upon an classical input $x' \in \mathcal{X}'$. In this case, the classical error map can be conditioned upon input as $\mathbf{R}_{\gamma,x'}$ such that replaced behavior is distinct for input $x$. We now show that the classical error map can simulate a family of POVM measurements that be can derived from a PVM measurement preceded by a quantum replacer channel.

**Proposition 1.** A classical replacer map $\mathbf{R}_\gamma \in \mathcal{P}_{\mathcal{Y}|\mathcal{Y}}$ is equivalent to the POVM $R_\gamma \in \text{POVM}(\mathcal{H}_d)$ whose $y^{th}$ is

$$R_{y,\gamma} = (1-\gamma)\Pi_y + \gamma R_y \quad \text{where} \quad R_y = \mu_y \mathbb{I}_d \tag{8.26}$$

where $\{\Pi_y\}_{y\in\mathcal{Y}} \in \text{POVM}(\mathcal{H}_d)$ is the measurement whose probabilities $P(y|x) = \text{Tr}\left[\Pi_y \rho_x\right]$ achieve the noiseless behavior $\mathbf{P}$ for a given state ensemble $\{\rho_x\}$.

*Proof.* To prove equivalence, we show that the behavior $\widetilde{\mathbf{P}} = \mathbf{R}_\gamma \mathbf{P}$ is achieved by the quantum probabilities $P'(y|x) = \text{Tr}\left[R_{y,\gamma}\rho_x\right]$. First, from Eq. (8.25) note that $\widetilde{P}(y|x) = (1-\gamma)P(y|x) + \gamma\mu_y$ where $\sum_y \mu_y = 1$ and $\mu_y \geq 0$. Then, using Eq. (8.26) we find that, for any quantum state $\rho_x \in D(\mathcal{H}_d)$,

$$P'(y|x) = \text{Tr}\left[R_{y,\gamma}\rho_x\right] = (1-\gamma)\text{Tr}\left[\Pi_y\rho_x\right] + \gamma\mu_y \text{Tr}\left[\mathbb{I}_d\rho_x\right] = (1-\gamma)P(y|x) + \gamma\mu_y = \widetilde{P}(y|x). \tag{8.27}$$

□

**Proposition 2.** A classical replacer error map $\mathbf{R}_\gamma$ is equivalent to a quantum measurement device that implements a projective measurement $\{\Pi_y\}_{y\in\mathcal{Y}} \in \text{PVM}(\mathcal{H}_d)$ and has replacer channel $\mathcal{R}_\gamma(\rho) = (1-\gamma)\rho + \gamma\sigma$ where $\text{Tr}\left[\Pi_y\sigma\right] = \mu_y$ are the probabilities of the replacer behavior $\mathbf{R} = \sum_{y,x} \mu_y |y\rangle\langle x|$.

*Proof.* Consider a projective measurement $\{\Pi_y\}_{y\in\mathcal{Y}} \in \text{PVM}(\mathcal{H}_d)$ that achieves the noiseless probabilities $P(y|x) = \text{Tr}\left[\Pi_y\rho_x\right]$ for a state ensemble $\{\rho_x\}_{x\in\mathcal{X}}$. Then, incorporate the replacer channel between the state and measurement as

$$P'(y|x) = \text{Tr}\left[\Pi_y \mathcal{R}_\gamma(\rho_x)\right] = \text{Tr}\left[\mathcal{R}_\gamma^\dagger(\Pi_y)\rho_x\right]. \tag{8.28}$$

Noting that the replacer channel Kraus operators are $K_0 = \sqrt{1-\gamma}\mathbb{I}_d$ and $K_{i,j} = \sqrt{\gamma}\sqrt{\mu_j}\,|\psi_j\rangle\langle i|$ where $\sum_j \mu_j = 1$ $\mu_j \geq 0$, we can write

$$\mathcal{R}_\gamma^\dagger(\Pi_y) = (1-\gamma)\Pi_y + \gamma \sum_{i,j=1}^{d} \mu_j |i\rangle\langle\psi_j|\,\Pi_y\,|\psi_j\rangle\langle i|. \tag{8.29}$$

Since $\Pi_y$ is a PVM element, it is orthogonal to the rest of measurement operators, therefore, each $|\psi_j\rangle$ can be assigned to the subspace of $\mathcal{H}_d$ spanned by each distinct measurement operator $\Pi_y$. Taking $j \in \mathcal{Y}$, such

that $\delta_{j,y} = \langle\psi_j|\Pi_y|\psi_j\rangle$, Eq. (8.29) becomes

$$\mathcal{R}^\dagger_\gamma(\Pi_y) = (1-\gamma)\Pi_y + \gamma\sum_{i=1}^{d}\mu_y\,|i\rangle\langle i| = (1-\gamma)\Pi_y + \mu_y\mathbb{I}_d = R_{y,\gamma} \tag{8.30}$$

Note that the replacer state $\sigma$ must be selected carefully such that weights $\mu_j$ and states $\{|\psi_j\rangle\}_j$ are set appropriately to ensure equivalence between the POVM $R_\gamma$ and the replacer channel $\mathcal{R}_\gamma$. Finally, using Proposition 1 we find equivalence between channel $\mathcal{R}_\gamma$ and the classical replacer error map $\mathbf{R}_\gamma$. □

Remarkably, Proposition 1 and Proposition 2 show that a classical postprocessing map can suffice for implementing either POVM measurements or a nonunitial quantum channel such the replacer channel. In practice, if a POVM or quantum channel can be subsituted for classical postprocessing than this substitution is advantageous because it reduces the overhead of implementing more complicated processing and measurements. As a special case of the classical replacer map we consider the classical white noise map.

**Definition 63. Classical White Noise Map:** $\mathbf{W}_\gamma \in \mathcal{P}_{\mathcal{Y}|\mathcal{Y}}$, a classical replacer error map that adds white noise to a behavior. The white noise map is represented as a column stochastic matrix

$$\mathbf{W}_\gamma = (1-\gamma)\mathbb{I}_{|\mathcal{Y}|} + \frac{\gamma}{|\mathcal{Y}|}\mathbf{1}_{|\mathcal{Y}|\times|\mathcal{Y}|} \tag{8.31}$$

where $\mathbf{1}_{|\mathcal{Y}|\times|\mathcal{Y}|}$ is the $|\mathcal{Y}|\times|\mathcal{Y}|$ matrix of all ones. The resulting noisy behavior is

$$\widetilde{\mathbf{P}} = \mathbf{W}_\gamma\mathbf{P} = (1-\gamma)\mathbf{P} + \frac{\gamma}{|\mathcal{Y}|}\mathbf{1}_{|\mathcal{Y}|\times|\mathcal{X}|} \tag{8.32}$$

Using Proposition 1, we can express the classical white noise map in Eq. (8.31) as the POVM $\{W_{y|x,\gamma}\}_y \in \mathcal{Y} \in \text{POVM}(\mathcal{H}_d)$

$$W_{y|x,\gamma} = (1-\gamma)\Pi_{y|x} + \gamma W_{y|x}, \tag{8.33}$$

where $W_{y|x} = \frac{1}{|\mathcal{Y}|}\mathbb{I}_d$ are white noise POVM elements and $\Pi_{y|x}$ are the noiseless POVM elements. Note that $\mu_y = \frac{1}{|\mathcal{Y}|}$ because all outputs have equal probability.

**Proposition 3.** For any device whose measurement $\{\Pi_y\}_{y\in\mathcal{Y}} \in \text{POVM}(\mathcal{H}_d)$ satisfies $\text{Tr}\left[\Pi_y\right] = \text{Tr}\left[\Pi'_y\right]$ for all $y, y' \in \mathcal{Y}$, the classical white noise error map is equivalent to a depolarizing channel $\mathcal{W}_v$ being applied to the measured state $\rho_x$ where $\mathcal{W}_v$ is defined in Eq. (13.69) and $v = (1-\gamma)$.

*Proof.* From Eq. (13.82) and Proposition 1, the classical white noise map is $W_{y,\gamma} = (1-\gamma)\Pi_y + \gamma/|\mathcal{Y}|\mathbb{I}_d$ where $\Pi_y$ is a POVM that achieves the noiseless probabilities $P(y|x) = \text{Tr}\left[\Pi_y\rho_x\right]$. Applying the depolarizing channel before measurement, we find

$$P'(y|x) = \text{Tr}\left[\Pi_y\mathcal{W}_v(\rho_x)\right] = \text{Tr}\left[\mathcal{W}^\dagger_v(\Pi_y)\rho_x\right]. \tag{8.34}$$

Since the Kraus operators of the depolarizing channel self-adjoint, the depolarizing channel is also self-adjoint as $\mathcal{W}^\dagger_v = \mathcal{W}_v$. Therefore, the POVM in Eq. (8.34) becomes

$$\mathcal{W}^\dagger_v = v\Pi_y + \frac{(1-v)}{d}\text{Tr}\left[\Pi_y\right]\mathbb{I}_d. \tag{8.35}$$

Finally, to recover Eq. (13.82) we substitute $v = (1-\gamma)$ and require that $\text{Tr}\left[\Pi_y\right]/d = 1/|\mathcal{Y}|$ holds for all $y$.

It follows that the classical white noise map simulates the depolarized measurement for any POVM satisfying $\mathrm{Tr}\left[\Pi_y\right] = \mathrm{Tr}\left[\Pi_{y'}\right]$ for all $y, y' \in \mathcal{Y}$. □

As a result of Proposition 3, classical white noise can simulate depolarizing noise in certain measurement devices. Thus, we show that these classical postprocessing maps can accommodate both unital and nonunital noise.

**Definition 64. Classical Loss Error:** $\mathbf{L}_\gamma \in \mathcal{P}_{\mathcal{Y}|\mathcal{Y}}$, consider the special case of classical replacer error where the output $y$ is mapped to a fixed output $y_0 = 0$ with probability $P(y_0|y) = 1$. The classical loss error is expressed as

$$\mathbf{L}_\gamma = (1-\gamma)\mathbb{I}_{|\mathcal{Y}|} + \gamma \sum_{y,y' \in \mathcal{Y}} \delta_{y_0,y'} \, |y'\rangle\langle y| \tag{8.36}$$

and its effect on a behavior is

$$\widetilde{\mathbf{P}} = \mathbf{L}_\gamma \mathbf{P} = (1-\gamma)\mathbf{P} + \gamma \sum_{y \in \mathcal{Y}} \sum_{x \in \mathcal{X}} \delta_{y_0,y} \, |y\rangle\langle x| \, . \tag{8.37}$$

Using Proposition 1, the behavior of the classical loss error is equivalent to the behavior constructed using the POVM $\{L_{y,\gamma}\}_{y \in \mathcal{Y}} \in \mathrm{POVM}(\mathcal{H}_d)$ where

$$L_{y,\gamma} = (1-\gamma)L_y + \gamma \delta_{y_0,y} \mathbb{I}_d. \tag{8.38}$$

Moreover, if $\{\Pi_y\}_{y \in \mathcal{Y}} \in \mathrm{PVM}(\mathcal{H}_d)$ then by Proposition 2, we find that the classical loss error can simulate the replacer channel $\mathcal{R}_\gamma$ whose replacer state is $\sigma = |\psi\rangle\langle\psi|$ where $\delta_{y,0} = \mathrm{Tr}\left[\Pi_y \sigma\right]$. For this reason, we find that the classical loss error is analogous to the loss channel.

# Chapter 9

# Nonclassicality in Quantum Networks

We are unable to peer inside a quantum system to observe its inner workings. Thus, to prove that quantum physics or that its operational advantage exists, we must observe in the classical data the input to and output from the quantum system. To this end, it is constructive to take on a device-independent perspective, in which we try infer physical properties of a system using only its measured input-output statistics. This approach is typically referred to as being *device-independent* [214] because no assumptions are made on the physical workings of the devices involved, only their causal structure and communication resources.

**Definition 65. Device-Independent:** A theory, protocol, or test that depends only on a black-box system's classical data, which is referred to as its behavior $\mathbf{P}$.

From a device-independent perspective, constraints may be placed on the communication resources used to link the devices, and remarkably, a clear separation can be found between the set of classical network behaviors $\mathcal{C}^{\text{Net}}$ and the set of quantum network behaviors $\mathcal{Q}^{\text{Net}}$. That is, there exists a nonclassicality in the behaviors of quantum networks where a given behavior $\mathbf{P} \in \mathcal{Q}^{\text{Net}}$ may not be reproducible by a classical network implying that $\mathbf{P} \notin \mathcal{C}^{\text{Net}}$.

## 9.1 Quantum Nonclassicality

Consider a given network topology $\text{Net}(\vec{N}, \vec{E})$ where $\mathcal{C}^{\text{Net}}$ and $\mathcal{Q}^{\text{Net}}$ denote the sets of classical and quantum network behaviors for a given network topology. We consider the black-box model in the discrete memoryless setting to place quantum and classical networks on equal footing such that both networks produce behaviors belonging to the same full probability $\mathcal{P}^{\text{Net}}_{\mathcal{Y}|\mathcal{X}}$. The distinction between the two networks is that the classical communication channels $\text{id}^{\text{Tx}\to\text{Rx}} \in \vec{E}$ are exchanged for noiseless quantum communication channels acting upon a $d$ dimensional quantum state space. Generally,

$$\mathcal{C}^{\text{Net}} \subseteq \mathcal{Q}^{\text{Net}} \quad \text{and} \quad \text{Conv}\left(\mathcal{C}^{\text{Net}}\right) \subseteq \text{Conv}\left(\mathcal{Q}^{\text{Net}}\right) \tag{9.1}$$

where ideal quantum communication networks can simulate the behaviors of the analogous classical network. However, quantum resources can produce correlations that cannot be reproduced using classical communication

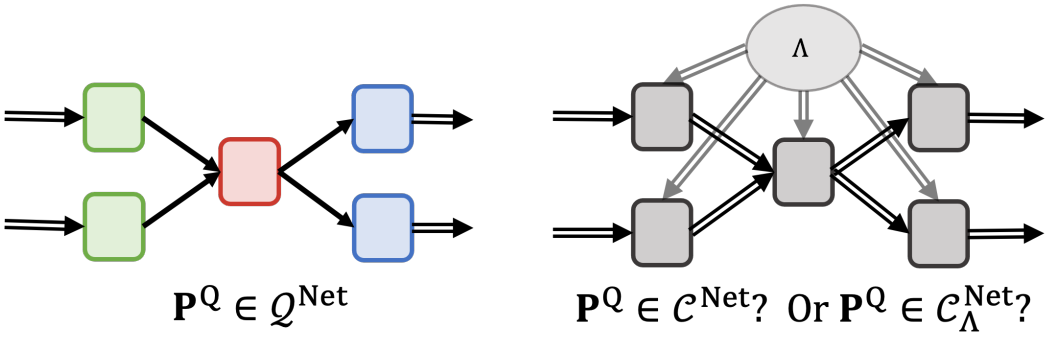



Figure 9.1: Nonclassicality can be understood in terms of classical simulability. The main question is whether the quantum network on the left can be simulated by the classical network on the right. If the classical network can simulate the quantum network behavior without using shared randomness, the quantum behavior is classically simulable. If the classical network needs shared randomness to simulate, the quantum behavior is weakly nonclassical. If the classical networks cannot simulate the quantum network behavior when an unlimited amount of global shared randomness is present, then the quantum network behavior is strongly nonclassical.

resources and black-box devices.

**Definition 66. Nonclassical Behavior:** For a given network topology $\text{Net}(\vec{N}, \vec{E})$, a quantum network behavior $\mathbf{P}^{\text{Net}} \in \mathcal{Q}^{\text{Net}}$ is nonclassical if and only $\mathbf{P} \notin \mathcal{C}^{\text{Net}}$.

**Definition 67. Weak Nonclassicality:** For a given network topology $\text{Net}(\vec{N}, \vec{E})$, a nonclassical quantum network behavior $\mathbf{P} \in \mathcal{Q}^{\text{Net}}$ where $\mathbf{P} \notin \mathcal{C}^{\text{Net}}$ is *weakly nonclassical* if and only $\mathbf{P} \in \mathcal{C}_{\Lambda}^{\text{Net}} = \text{Conv}\left(\mathcal{C}^{\text{Net}}\right)$.

**Definition 68. Strong Nonclassicality:** For a given network topology $\text{Net}(\vec{N}, \vec{E})$, a nonclassical quantum network behavior $\mathbf{P} \in \mathcal{Q}^{\text{Net}}$ is *strongly nonclassical* if and only if $\mathbf{P} \notin \mathcal{C}_{\Lambda}^{\text{Net}} = \text{Conv}\left(\mathcal{C}^{\text{Net}}\right)$.

In these definitions, any behavior that cannot be reproduced by its equivalent classical resources is nonclassical (see 9.2). We categorize nonclassical behaviors as being weak or strong. The distinction being that weak nonclassicality can be simulated using classical communication and an unlimited amount of shared randomness is available between all devices while strong nonclassicality cannot. To classically simulate strong nonclassicality, additional communication resources are required. Both types of nonclassicality demonstrate a quantum advantage, however, strong nonclassicality provides a more convincing demonstration of quantum advantage because the quantum behaviors can only be reproduced using additional communication. Nonclassicality is a device-independent property because a behavior being nonclassical depends only on the produced statistics. However, the assumed causal structure $\text{Net}(\vec{N}, \vec{E})$ must hold, otherwise the bounds of the classical correlations $\mathcal{C}^{\text{Net}}$ no longer apply. Thus, observing nonclassical behaviors in this classical data is interesting from both a physical and technological perspective.

*Nonclassicality* describes a general phenomena that occurs in quantum systems that have locality constraints. Nonclassicality in the form described by J.S. Bell is, in which entanglement links two quantum devices, is historically referred to as Bell nonlocality [109]. Through a range of theoretical frameworks, Bell nonlocality can be extended to quantum networks [124]. Bell nonlocality can be viewed as a special example of a broad class of nonclassical phenomena referred to as quantum contextuality [163]. Here, incompatible measurements referred to as *contexts* and are viewed as being foundational to realizing quantum computing advantage [66]–[68], [96]. Another example of nonclassicality is found the field of quantum causal modeling, the

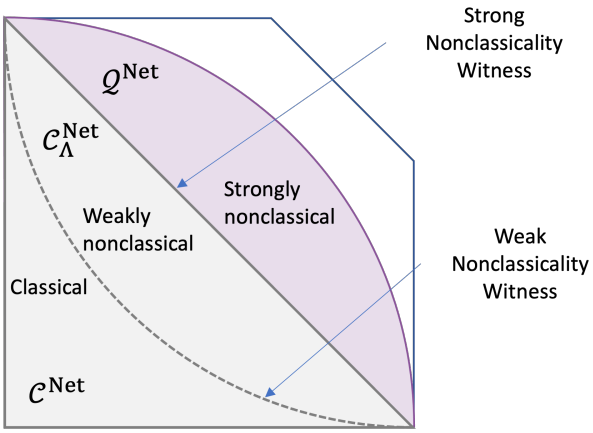



Figure 9.2: An abstraction of the probability polytope and relative sets of quantum and classical network behaviors. The convex polytope of the border denotes the full probability polytope $\mathcal{P}_{\mathcal{Y}|\mathcal{X}}$. All quantum and classical network behaviors are subsets of the full probability polytope. The classical set of network behaviors $\mathcal{C}^{\text{Net}}$ is typically nonconvex, the triangular region $\mathcal{C}_{\Lambda}^{\text{Net}}$ is convex due to the shared randomness. A set of strongly nonclassical quantum behaviors are shown in purple. It is assumed that shared randomness is present to make the quantum set convex and that it can extend beyond $\mathcal{C}_{\Lambda}^{\text{Net}}$. To witness behaviors as being outside of either $\mathcal{C}^{\text{Net}}$ or $\mathcal{C}_{\Lambda}^{\text{Net}}$, violations of the respective boundaries can be used. These witnesses take the form of black-box games and are discussed in more detail later.

fundamental restrictions on information flow imposed by a causal structure is examined for both quantum and classical communication networks [215]–[217]. Interestingly, quantum causal models can exhibit nonclassicality in the delayed-choice settings with a single particle and no entanglement [158], [161]. Alternatively, quantum nonclassicality can be understood in terms of its communication complexity, in which classical and quantum resources are compared in their ability to solve a problem [64]. Furthermore, entropic methods have been applied to distinguish the correlations of quantum and classical data [218]–[222].

### *9.1.1 The Operational Advantage of Quantum Nonclassicality*

From an engineering perspective, it is natural consider how these nonclassical correlations can be applied. Indeed, the nonclassicality of quantum systems can lead to operational advantages in distributed information processing, certifying quantum resources, and information security.

The information processing advantage of nonclassicality can be observed in distributed computing settings in which devices operate independently and share limited communication resources. In such settings, the devices can play games where the maximal probability of winning the game is directly dependent upon the communication resources used. Indeed, when entanglement is shared between devices, nonclassical winning probabilities can be obtained [223]–[225]. In a similar approach, the communication complexity of distributed information processing tasks can be dramatically reduced when quantum communication resources are used [64], showing that the amount of quantum communication needed to solve a problem can be less than the required amount of classical communication. In general, any nonclassical behavior must perform some task at an advantage [63], [65], although the value of this task may be in question. As a simple example, consider a nonclassical behavior $\mathbf{P} \in \mathcal{Q}^{\text{Net}}$. If the task is to use classical communication resources to reproduce $\mathbf{P}$, then there is a quantum advantage because additional classical communication is required to simulate the quantum communication. Similarly, in quantum computing measurement contextuality has been observed to

be a source of information processing advantage [66]–[68], [96], [97].

Quantum nonclassicality can also provide advantages certifying quantum resources and functionality. If a black-box communication system with a known causal structure is observed to exhibit nonclassical correlations, then it can be asserted that quantum resources were applied, however, the assumed causal structure and amount of signaling must hold. These concepts are the foundation of device self-testing in which known quantum devices are used to characterize the functionality of another [226]. Self-tests can certify a range of properties such as dimensionality [227]–[231], state preparations [232]–[235], and measurements [236]–[239]. These methods can then be extended into network settings [164], [240]–[244].

Fundamental cryptographic protocols such as key distribution or randomness generation can be made more secure by using nonclassicality to assert that quantum resources are being used. Such protocols are based upon the principles of self-testing procedures. In these device-independent protocols the nonclassical behavior asserts the presence of quantum resources, which are often needed to assume security. For example, nonclassicality can be used to certify that randomness was produced by quantum systems [186]–[192], [245], [246]. Furthermore, the presence of nonclassicality can help ensure that eavesdroppers are not present in key distribution [182], [247]–[255]. Such device-independent protocols can even be scaled to networking scenarios [256]–[260].

## *9.2 The Hierarchy of Network Communication Resources*

For a given network topology, the set of classical behaviors is contained by the set of behaviors for the quantum network having the same topology, but upgrading some or all of the communication resources to quantum. That is, the relation $\mathcal{C}^{\text{Net}} \subseteq \mathcal{Q}^{\text{Net}}$ must hold for any two networks that have the same causal structure and communication resources. As a result, the set of deterministic behaviors $\mathcal{V}^{\text{Net}}$ that are extreme points of $\mathcal{C}^{\text{Net}}$ are also extreme points of $\mathcal{Q}^{\text{Net}}$, however, it is not always clear whether

$$\text{Conv}\left(\mathcal{C}^{\text{Net}}\right) \subseteq \text{Conv}\left(\mathcal{Q}^{\text{Net}}\right) \tag{9.2}$$

holds with equality, or if the quantum set is greater because both cases can occur in networking systems.

The quantum and classical communication resources can be ranked based on their operational value [64]. From strongest to weakest the hierarchy of communication resources is:

1. Quantum Communication
2. Classical Communication
3. Quantum Entanglement
4. Classical Shared Randomness

The relative *strength* of two communication resource compares their operational value. In general, this comparison can be quantified in the communication complexity of using a set of resources to simulate another. To better understand the resource hierarchy, we show in example how the various communication resources can be simulated.

Classical shared randomness is the weakest of all resources. As described in Definition 14, distributing a shared random value $\lambda \in [d]$ to $m$ subscribing devices requires $m \log_2(d)$ bits of classical communication. Furthermore, the communication cost of distributing shared to solve any two-party computation $f : \{0,1\}^N \times$

$\{0,1\}^M \to \{0,1\}$ is negligible for large $N$ and $M$ [261]. Therefore, shared randomness can not only be reproduced using classical communication, but in some cases the added cost of distributing shared randomness can be insignificant.

Generally, randomness shared amongst $m$ devices can be modeled as a single source $\Lambda$ that emits a message $\lambda_i \in [d]$ to party $A_i$ with probability $P(\lambda_i)$. The messages are independent, but can be chosen arbitrarily such that the behavior decomposes as

$$\mathbf{P}^{\Lambda} = \bigotimes_{i=1}^{m} \sum_{\lambda_i=1}^{d} P(\lambda_i) \left|\lambda_i\right\rangle = \sum_{\vec{\lambda}\in[d]^m} P(\vec{\lambda}) \left|\vec{\lambda}\right\rangle. \tag{9.3}$$

We find that Eq. (9.3) can be reproduced exactly by measuring the state

$$\rho^{\Lambda} = \mathrm{diag}\left(P(\vec{\lambda})\right) \vec{\lambda} \in [d]^m \in D(\mathcal{H}_d^{\otimes m}) \tag{9.4}$$

in the computational basis such that $P(\vec{\lambda}) = \left\langle \vec{\lambda}|\rho^{\Lambda}|\vec{\lambda}\right\rangle$. Thus, any shared randomness is simulated using a quantum mixed state distribution. In a more restricted example, let $\lambda_i) = \lambda_{i'}$ where $P(\lambda_i) = P(\lambda_{i'})$ for all $i, i' \in [m]$ and $\lambda_i\lambda_{i'} \in [d]$ such that $\mathbf{P}^{\Lambda} = \bigotimes_{i=1}^{m} \sum_{\lambda=1}^{d} P(\lambda) \left|\lambda\right\rangle$. This behavior can be reproduced by measuring the entangled state

$$\left|\psi^{\Lambda}\right\rangle = \sum_{\lambda=1}^{d} \sqrt{P(\lambda)} \bigotimes_{j=1}^{m} \left|\lambda\right\rangle \tag{9.5}$$

in the computational basis such that $P(\lambda) = |\left\langle\lambda|\psi^{\Lambda}\right\rangle|^2$. Interestingly, shared randomness cannot be used to simulate the correlations produced by quantum entanglement [12], nor can it be used to communicate information, meaning that it is the weakest of the communication resources.

While quantum entanglement is certainly a stronger communication resource than classical shared randomness, it is still weaker than classical communication. It was shown by Bacon and Toner that a single bit of classical communication and and unlimited amount of shared randomness are sufficient to simulate Bell correlations [262], [263]. Furthermore, non-signaling theories such as classical and quantum physics obey the *principle of information causality*, meaning that quantum entanglement cannot be used to communicate when assisted by classical communication [264].

Moving up the hierarchy to the dynamic resources, quantum communication can be found to be stronger than classical communication. Quantum communication can exactly reproduce classical communication by encoding signals into the computational basis states, however, classical communication cannot reproduce quantum communication without using extra resources. For example, quantum communication of an unknown quantum state can be simulated using a teleportation protocol, which requires only 2 bits of classical communication and an entangled state shared between the sender and the receiver [198]. However, by Holeveo's theorem the classical communication capacity of a quantum channel cannot exceed its dimension [173]. This statement is made even stronger by Frenkel and Weiner's result [115], in which it was proven that one *dit* of classical communication and an unlimited amount of shared randomness are sufficient to simulate one *qudit* of quantum communication. However, additional classical communication is needed to simulate the nonclassical correlations of quantum random access code in the delayed-choice setting.

Similar to quantum teleportation, entanglement-assisted classical communication can demonstrate communication advantages over shared randomness-assisted classical communication. Throughout this work, we show different configurations of entanglement-assistance in communication networks to produce strongly

nonclassical correlations. However, the principle of information causality restricts the amount of information that can be communicated [264]. An interesting example occurs when quantum communication is assisted by entanglement. In this case, dense coding protocols are known where two-bits of classical communication can be encoded into a shared entangled state and quantum communication [265]. It has recently been observed that there is a duality between dense coding protocols and teleportation protocols [266].

The fact that different configurations of classical resources can simulate the correlations values the correlations in terms of classical resources. If the task requires the processing of classical information, the quantum and classical resources will be indistinguishable in their application, although more classical resources may be needed to simulate the quantum system, if classical resources are cheap, then this cost may not be very significant.

## *9.3 Witnessing Nonclassicality*

Nonclassicality is a device-independent quantity that is observed in the network's classical data where the network's causal structure and communcation resources are known. This task is the foundation of device-independent certification and is even an important task to observing and proving quantum physics in nature. Since the classical set $\mathcal{C}^{\text{Net}}$ is bounded, there exist inequalities that separate the behaviors achievable by the network $\mathbf{P} \in \mathcal{C}^{\text{Net}}$ from those that are not $\mathbf{P} \notin \mathcal{C}^{\text{Net}}$. Thus, nonclassicality can be witnessed in classical data by demonstrating a violation of the classical boundary.

**Lemma 7.** Consider a nonclassical behavior $\mathbf{P}^{NC} \notin \mathcal{C}^{\text{Net}}$. There exists an inequality $\gamma \geq g(\mathbf{P}^{C})$ that is satisfied for all $\mathbf{P}^{C} \in \mathcal{C}^{\text{Net}}$, but violated by the nonclassical behavior as $\gamma < g(\mathbf{P}^{NC})$.

Applying Lemma 7, the task of witnessing nonclassicality reduces to identifying a function $g : \mathbf{P} \to \mathbb{R}$ that bounds the set of classical behaviors, but is violated by the nonclassical behavior. Inequalities bounding the set of classical behaviors can be derived using the fact that that $\mathcal{C}^{\text{Net}}$ is closed, connected, and bound by a set of inequalities referred to as nonclassicality witnesses [109], [124]. We express nonclassicality witnesses as the game $(\gamma, S_{\text{Net}}(\mathbf{P}))$ where $S_{\text{Net}} : \mathbf{P} \to \mathbb{R}$ is referred to as the nonclassicality score and $\gamma \geq S_{\text{Net}}(\mathbf{P})$ for all $\mathbf{P} \in \mathcal{C}^{\text{Net}}$. Unless randomness is shared amongst all device such that $\mathcal{C}^{\text{Net}}$ is convex, the nonclassicalilty score $S_{\text{Net}}(\mathbf{P})$ is a nonlinear function of the probabilities $P(\vec{a}|\vec{x})$. Fortunately, nonclassicality witnesses have been derived for a large number of communication network scenarios [124], [267]–[275].

We mainly focus on the case when the nonclassicality witness linear and an unlimited amount of shared randomness is held by all device in the network. In this setting $\mathcal{C}^{\text{Net}} = \mathcal{C}^{\text{Net}}_{\Lambda}$ becomes convex and corresponds to a classical network polytope, which is bound by linear inequalities as discussed in Chapter 4. In this case, a given behavior $\mathbf{P}$ can be tested to belong to $\mathcal{C}^{\text{Net}}$ by using the linear program in Eq. (4.22). However, the linear program will only obtain an example facet inequality that is violated by the behavior $\mathbf{P}$. On the other hand, the classical network polytope can be completely characterized using Fourier-Motzkin elimination or the adjacency decomposition algorithm discussed in Section 4.6. However, these algorithms can only scale due to their computational complexity. Fortunately, the tight classical bounds have been computed for many network scenarios. Therefore, a breadth of nonclassicality can be studied without needing to derive new bounds.

We can also consider the case where the set of classical network behaviors is nonconvex $\mathcal{C}^{\text{Net}} \subset \text{Conv}\left(\mathcal{C}^{\text{Net}}\right)$, the bounds become more difficult to characterize. For many network topologies, these nonlinear bounds have been derived and general techniques for constructing new bounding inequalities have been developed [276]–[279].

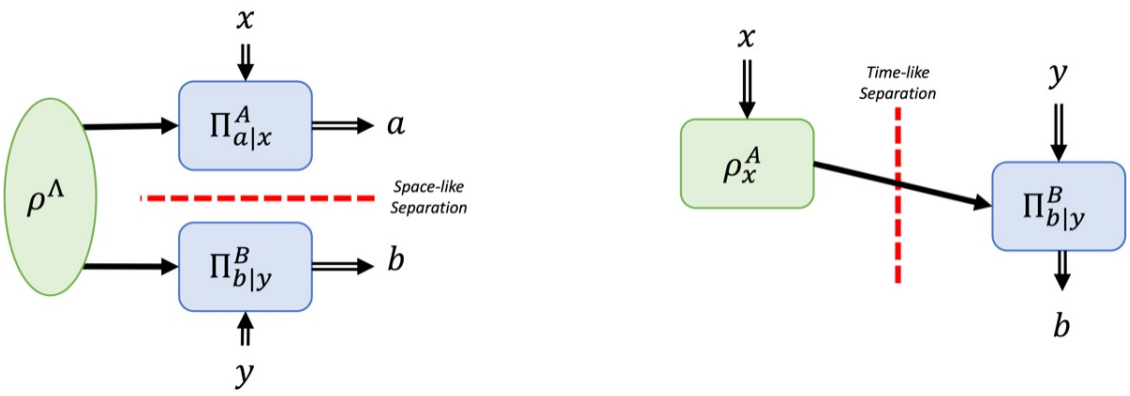



Figure 9.3: (Left) The CHSH scenario in which two space-like separated measurement devices, $A$ and $B$ share an entangled state $\rho^\Lambda$. (Right) The delayed-choice setting where device $A$ sends a quantum state to device $B$.

Nonclassical network behaviors can also be witnessed using a game theoretic approach in which quantum communication resources increase the probability of winning a game [64], [223]. These games are typically played under the assumption that the players are not communicating and the advantage comes from the use of entanglement resources. These types of games are referred to as *nonlocal games*. Nonlocal games are linear black-box game $(\gamma, \mathbf{G})$ (see Def. 7) where there is an associated reward $G_{y,x}$ for outputting $y \in \mathcal{Y}$ given the input $x \in \mathcal{X}$ and output $y \in \mathcal{Y}$. The upper bound $\gamma$ corresponds to the maximal average score that the classical network can achieve given that $x$ is drawn uniformly from $\mathcal{X}$ and sufficiently many data points have been taken to dampen statistical fluctuations. Games do not necessarily represent tight bounds on the set of classical network behaviors, but they do associate with an explicit task that demonstrates an advantage using quantum resources.

## 9.4 Examples of Nonclassicality

### 9.4.1 Delayed-Choice Signaling Scenario

Consider the delayed-choice scenario in which a sender encodes a two-bit input $x \in \mathcal{X} = \mathbb{B}^2$ into a qubit state $\rho_x$ and sends it to a receiver who decodes the quantum state using a quantum measurement $\Pi^B_{b|y}$ conditioned on an input $y \in \mathcal{Y} = \mathbb{B}$ to produce an output $b \in \mathcal{B}\mathbb{B}$. In this scenario, $x$ and $y$ are locally independent variables, and a time-like separation exists between the sender and receiver.

An interesting quantum advantage emerges when a random access coding (RAC) game is considered where the objective is for the receiver to guess the $y^{th}$ bit of the sender's input $x$, however, only one bit of communication is allowed. That is, the gam is won if $b = x_y$ where $x_y$ is the $y^{th}$ bit of the senders input $x$. This game can be encoded into the game matrix that rewards the correct output for each input

$$\mathbf{G} = \sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}}\sum_{b\in\mathcal{B}} G_{b,x,y}\,|b\rangle\langle x,y| = \begin{bmatrix} 1 & 1 & 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 & 1 & 1 \end{bmatrix}. \tag{9.6}$$

The classical bound is $\gamma = 6 \geq \langle \mathbf{G}, \mathbf{P} \rangle$ for all $\mathbf{P} \in \mathcal{C}^{\text{Net}}_\Lambda$ [280]. However, the maximal quantum violation is $\gamma = 4 + 2\sqrt{2}$. The RAC game can be extended to any system where $x \in \mathbb{B}^n$, $|\mathcal{Y}| = n$, and $\mathcal{B} = \mathbb{B}$.

In the classical setting, the probability that the receiver outputs the right answer is $P_{Win} \leq \frac{3}{4}$ where a

perfect strategy doesn't exists because the senders input cannot be compressed into one-bit without loss. Interestingly, when one-bit of quantum communication is used instead, the probability of winning is bound as $P^{Q}_{Win} \leq \frac{2+\sqrt{2}}{4} > \frac{3}{4}$ marking a quantum advantage in this task. The advantage in quantum random access code has been generalized and extensively studied [187], [280]–[284]. Note however, that quantum communication does not allow for a perfect quantum winning strategy nor does is allow for more than one-bit of information to be communicated. Nevertheless, it does improve the probability of winning the random access code game.

The quantum random access code advantage has been studied [282], [283]. This example of strong nonclassicality have been studied as dimensionality witnesses [227], [228] and semi-device independent entanglement certification [234]. These violations of classicality also semi-device independent quantum key distribution [252], and semi-device independent random number generation [188], [192], [246]. Furthermore, the nonclassicality violations in these settings have been experimentally verified [161], [229].

### *9.4.2 Nonsignaling Bipartite Nonclassicality*

Consider the scenario in which one source $\Lambda$ links two devices, $A$ and $B$ where the inputs and outputs of each device are binary such that $x, y, a, b \in \mathbb{B}$. This scenario is a fundamental example in nonclassicality and is commonly referred to as the CHSH scenario after Clauser Horne Shimony and Holt [109]. These researchers showed that the classical behaviors $\mathbf{P} \in \mathcal{C}^{\text{CHSH}}_{\Lambda}$ are bound by the CHSH inequality [13]

$$S_{\text{CHSH}}(\mathbf{P}^{\text{Net}}) \equiv \left\langle O^A_0 \otimes O^B_0 \right\rangle + \left\langle O^A_1 \otimes O^B_0 \right\rangle + \left\langle O^A_0 \otimes O^B_1 \right\rangle - \left\langle O^A_1 \otimes O^B_1 \right\rangle \tag{9.7}$$

$$S_{\text{CHSH}}(\mathbf{P}^{\text{Net}}) \equiv \sum_{y \in \mathbb{B}} \left\langle O^{AB}_y \right\rangle \leq 2 \quad \text{where} \quad O^{AB}_y = (O^A_0 + (-1)^y O^A_1) \otimes O^B_y. \tag{9.8}$$

The CHSH inequality is a linear inequality whose game matrix is

$$\mathbf{G}^{\text{CHSH}} = \begin{pmatrix} 1 & 1 & 1 & -1 \\ -1 & -1 & -1 & 1 \\ -1 & -1 & -1 & 1 \\ 1 & 1 & 1 & -1 \end{pmatrix} \tag{9.9}$$

where $S_{\text{CHSH}}(\mathbf{P}^{AB}) = \left\langle \mathbf{G}^{\text{CHSH}}, \mathbf{P}^{AB} \right\rangle$ and behaviors decompose as

$$\mathbf{P}^{AB}_{a,b|x,y} = \sum_{\lambda} \mathbf{P}^{A}_{a|x,\lambda} \mathbf{P}^{B}_{b|y,\lambda} \mathbf{P}^{\Lambda}_{\lambda}. \tag{9.10}$$

The CHSH score quantifies the performance in a game where the goal is to maximize the likelihood that the binary inputs and outputs satisfy $a \oplus b = x \wedge y$ [109].

The maximal quantum violation of the CHSH inequality is $S^{\star}_{\text{CHSH}}(\mathbf{P}^{\star}) = 2\sqrt{2} > 2$ where $\mathbf{P}^{\star} \in \mathcal{Q}^{\text{CHSH}}$ [285]. It follows that all quantum behaviors $\mathbf{P} \in \mathcal{Q}^{\text{CHSH}}$ satisfy $S_{\text{CHSH}}(\mathbf{P}) \leq 2\sqrt{2}$. To achieve the optimal behavior $\mathbf{P}^{\star}$, the two devices measure the respective observables

$$O^A_x = (1-x)\sigma_3 + x\sigma_1 \quad \text{and} \quad O^B_y = \frac{1}{\sqrt{2}}(\sigma_3 + (-1)^y \sigma_1) \tag{9.11}$$

where $\sigma_3 = \sigma_z$ and $\sigma_1 = \sigma_x$ are Pauli matrices. If the two parties are correlated using shared randomness, their winning probability is bound as $P_{Win} \leq \frac{3}{4}$, however, when quantum entanglement is shared between

the two parties, their winning probability is bound as $P^{Q}_{Win} \leq \cos^2(\frac{\pi}{3}) > \frac{3}{4}$. Since there is no communication, the quantum entanglement is shown to be a stronger resource than classical shared randomness, hence there exists operational advantage when entanglement is used.

## *9.5 Nonclassical Hidden Variables in Quantum Mechanics*

There is still much to understand about quantum physics, and as postulated by Einstein, Podolsky, and Rosen, it may not be complete. The fact that operational quantum advantages seem to exist implies that quantum physics is somehow extends beyond classical limits.

**Definition 69. Realism:** A model is *realistic* if its elements correspond to physical elements of nature. That is, there exist hidden variables or information, that if known, allow the system's dynamics to be predicted with certainty.

**Definition 70. Locality:** A model is *local* if each network device operates only upon the information present.

**Definition 71. Causality:** A model is *causal* if the transmission of information is the forward lightcone and it operates only on information contained within its past lightcone.

However, demonstrating nonclassicality alone does not assert that one of these classical assumptions has failed. Indeed, the causal structure of the communication network and its communication resources must be known. Knowledge which cannot always be taken as fact. Otherwise, the nonclassical quantum behaviors can be simulated using classical communication. Consider the randomn access code game where the sender is allowed to send two rather than one bit. The receiver would always be able to output the right answer because the sender can send their entire input. Similarly, the CHSH violations can be simulated if classical communication is used [262]. The fact that nonclassicality can be simulated using classical resources allows it to be used to quantify the amount of operational advantage provided in the quantum task.

## *9.6 Simulating Nonclassicality with Classical Resources*

One of the main applications of nonclassicality is to certify the presence of quantum resources. Since the security of device-independent protocols and self-testing certification relies on nonclassical behaviors, it is important to understand how these protocols could be attacked using classical resources. In particular, consider a simulation-based attack in which classical resources are able to reproduce nonclassical statistics, a device-independent protocol would be unable to discern whether or not the underlying resources are quantum. To this end, we consider two types of simulation-based attacks in which the network is malicious and attempts to simulate quantum resources, while the researcher supplies the network with classical data attempting to discern correctly whether quantum resources are being used.

The first simulation-based attack we consider is a side-channel attack in which network devices communicate extra classical communication via a side-channel that is hidden from the researcher. The extra communication expands the classical network polytope $\mathcal{C}^{\text{Net}}_{\Lambda}$ such that it includes behaviors that would otherwise have been strongly nonclassical. If the researcher is not aware of the side-channel, then the network would appear to exhibit a nonclassical behavior. To mitigate the risk of this simulation-based attack, the researcher needs to characterize the network's causal structure and the amount of communication between devices. In other words, the research must confirm that no side-channels exist.

In practice, asserting that no classical side-channels exist is difficult because quantum networks will require classical communications between devices to coordinate their actions. Furthermore, since classical communication is easy to implement, there will be few limitations on the amount of classical communication between devices. Thus, it is important that network devices are trusted to not implement simulation attacks via a side-channel.

A second approach to simulating nonclassicality is a memory-based attack. Recall that we have so far only discussed black-boxes in the discrete memoryless setting. However, since classical data can be stored, network devices could use their classical memory to store their history of classical data. Consider the memoryful black-box with probabilities $P^{\text{Mem}}(y|x, \vec{x}_t, \vec{y}_t)$ where $\vec{x}_t$ and $\vec{y}_t$ are the history of inputs and outputs received by the memoryful black-box. Now, if the researcher draws the input $x$ from a uniform prior distribution independent of $\vec{x}_t$ and $\vec{y}_t$, then $\vec{x}_t$ and $\vec{y}_t$ can be understood as random variables local to the black-box. Therefore, a black-box that remembers its inputs and outputs can use that history to simulate stochasticity. However, as a single black-box this is not interesting because even in the memoryless setting, the black-box can be stochastic.

When considering memoryful black-boxes in a quantum network, the worst-case is that all black-boxes know the history of classical inputs and outputs for all devices. That is, the behavior of one memoryful network device is $P(\vec{y}|\vec{x}, \vec{x}_t^{\text{Net}}, \vec{y}_t^{\text{Net}})$ where $\vec{x}_t^{\text{Net}}$ and $\vec{y}_t^{\text{Net}}$ contain the history of inputs and outputs given to each device. Once again, if the researcher draws the input $\vec{x}$ from a uniform distribution independent of the device's history, then $\vec{x}_t^{\text{Net}}$ and $\vec{y}_t^{\text{Net}}$ can be understood as global hidden variables shared amongst all devices. As a result, the set of memoryful classical network behaviors $\mathcal{C}_{\text{Mem}}^{\text{Net}}$ must obey the hierarchy,

$$\mathcal{C}^{\text{Net}} \subseteq \mathcal{C}_{\text{Mem}}^{\text{Net}} \subseteq \mathcal{C}_{\Lambda}^{\text{Net}} \tag{9.12}$$

such that the memoryful network can simulate weakly nonclassical behaviors but not strongly nonclassical behaviors. Hence the assumption that network devices are memoryless is not necessary when strongly nonclassical behaviors are being used, but it is necessary when weak nonclassicality is considered. Furthermore, it is important that the researcher's input be uniformly random and independent from the black-box's history because otherwise a malicious network could use its history to correlate its behavior with the researcher's prior distribution, leading possibly to nonclassicality.

# Chapter 10

# The Noise Robustness of Nonclassical Behaviors

Quantum communication networks are naturally noisy. Within our model of quantum communication networks, noise can be modeled as a collection of CPTP maps $\mathcal{N}^{\text{Net}}$ that are applied between the ideal operations of the quantum networking devices. As discussed in Chapter 5, noisy quantum channels can expressed using unitary operators in the system-environment representation (see Def. 38).

We can interpret the noise in quantum networks as being caused by unwanted interactions between the encoded quantum information and its physical surrounding environment. The environment is typically not controlled by researchers, meaning that the environment cannot be measured, or intentionally operated upon. This means that the noise in a quantum network cannot be recovered when lost to the environment. However, the effects of noise can be mitigated by using extra quantum resources to perform error correction protocols. For instance, in entanglement distillation protocols, many copies of noisy entangled states are used to produce an entangled state with less noise [62]. Although, the extra quantum resources needed to mitigate noise are not always available.

## *10.1 Nonclassicality Breaking Noise*

Noise causes the set of quantum network behaviors $\mathcal{Q}^{\text{Net}}$ to deform. That is, considering an arbitrary quantum network topology (Net) and its noise model $\mathcal{N}^{\text{Net}}$, the set of noisy quantum network behaviors becomes constrained as $\mathcal{Q}^{\text{Net}}(\mathcal{N}^{\text{Net}}) \subset \mathcal{Q}^{\text{Net}}(\text{id}^{\text{Net}})$ where $\mathcal{Q}^{\text{Net}}(\text{id}^{\text{Net}})$ and $\mathcal{Q}^{\text{Net}}(\mathcal{N}^{\text{Net}})$ denote the sets of noiseless and noisy quantum network behaviors respectively. If $\mathcal{Q}^{\text{Net}}$ contains strongly nonclassical behaviors, then it the hierarchy $\text{Conv}\left(\mathcal{C}^{\text{Net}}\right) = \mathcal{C}_{\Lambda}^{\text{Net}} \subset \mathcal{Q}^{\text{Net}}$ must hold. Therefore, if a sufficient amount of noise is present, it is possible that the set of noisy quantum behaviors can deform such that it is completely contained by the set of classical network behaviors and its nonclassicality is said to be broken.

**Definition 72. Nonclassicality Breaking Noise:** A network's noise model $\mathcal{N}^{\text{Net}}$ is said to be strong nonclassicality breaking if and only if

$$\mathcal{Q}^{\text{Net}}(\mathcal{N}^{\text{Net}}) \subseteq \mathcal{C}_{\Lambda}^{\text{Net}} \subset \mathcal{Q}^{\text{Net}}(\text{id}^{\text{Net}}), \tag{10.1}$$

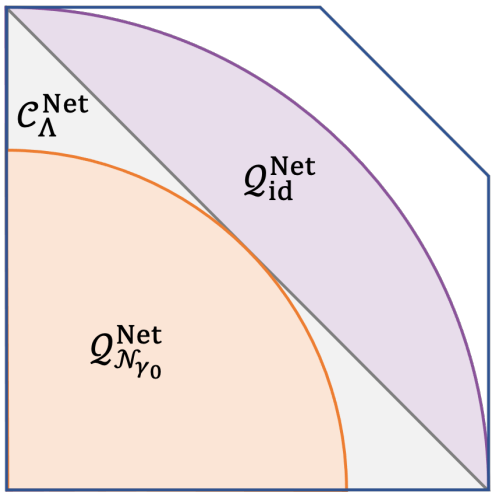



Figure 10.1: A depiction of the full probability polytope $\mathcal{P}_{\mathcal{Y}|\mathcal{X}}$ (outside blue polygon) that contains all quantum and classical network behaviors. The set of noiseless quantum behaviors $\mathcal{Q}^{\text{Net}}$ is purple and the set of noise quantum behaviors $\mathcal{Q}^{\text{Net}}(\mathcal{N}_{\vec{\gamma}}^{\text{Net}})$ is given in orange. The nonclassicality of the network is broken when the noise is sufficiently large such that the set of noisy quantum behaviors is contained by the set of classical network behaviors $\mathcal{C}_{\Lambda}^{\text{Net}}$.

and is said to be weak nonclassicality breaking if and only if

$$\mathcal{Q}^{\text{Net}}(\mathcal{N}^{\text{Net}}) \subseteq \mathcal{C}^{\text{Net}} \subset \mathcal{Q}^{\text{Net}}(\text{id}^{\text{Net}}), \tag{10.2}$$

where $\text{Conv}\left(\mathcal{C}^{\text{Net}}\right) = \mathcal{C}_{\Lambda}^{\text{Net}}$.

The concept of strong nonclassicality breaking was originally introduced in the context of nonlocality breaking channels [286], [287] where in these works, the authors considered the amount of noise at which the nonlocality with respect to the CHSH inequality is broken. We extended these results in reference [102], in which we considered the amount of noise needed to break the non-$n$-locality of star and chain quantum networks. The latter work is an example of *weak nonclassicality breaking* while the former works are examples of *strong nonclassicality breaking*.

In general, it is a difficult task to assert whether or not the set of quantum network behaviors is contained by the classical set $\mathcal{Q}^{\text{Net}}(\mathcal{N}^{\text{Net}}) \subseteq \mathcal{C}^{\text{Net}}$ (See Fig. 10.1). In practice, it is easier to assert whether or not nonclassicality broken with respect to a tight nonclassicality witness $(\beta, S_{\text{Net}})$. That is, the nonclassicality of a network is broken if $\beta \geq S_{\text{Net}}(\mathbf{P})$ is satisfied for all $\mathbf{P} \in \mathcal{Q}^{\text{Net}}(\mathcal{N}^{\text{Net}})$, or equivalently, if

$$\beta \geq S_{\text{Net}}(\mathbf{P}^{\star}) \equiv \max_{\mathbf{P} \in \mathcal{Q}^{\text{Net}}(\mathcal{N}^{\text{Net}})} S_{\text{Net}}(\mathbf{P}). \tag{10.3}$$

## *10.2 Quantifying Noise Robustness*

Since quantum nonclassicality deteriorates in the presence of noise, it is important to characterize the noise robustness of nonclassical quantum behaviors. To define the noise robustness, we first quantify the amount of noise present in a network. This can be done by parameterizing a quantum channel $\mathcal{N}_{\vec{\gamma}}^{\text{Net}}$ or classical postprocessing map $\mathbf{E}_{\vec{\gamma}}$ by a *noise parameter* $\vec{\gamma} = (\gamma_i \in [0,1])_{i=1}^{|\vec{\gamma}|}$ where $\gamma_i = 0$ corresponds to the noiseless case and $\mathcal{N}_{\vec{\gamma}}^{\text{Net}} \in \text{CPTP}$ for all $\gamma_i \in [0,1]$.

**Definition 73. The Noise Robustness of Nonclassicality:** The amount of noise that can be tolerated by a quantum network before nonclassicality is broken. The noise robustness of quantum nonclassicality can be quantified by the critical noise parameters $\gamma_0$ at which nonclassicality is broken, that is,

$$\gamma_0 = \sup\left\{\vec{\gamma} \in [0,1] \middle| \; \mathcal{Q}^{\text{Net}}(\mathcal{N}_\gamma^{\text{Net}}) \subseteq \mathcal{C}^{\text{Net}}\right\}. \tag{10.4}$$

where $\mathcal{Q}^{\text{Net}}(\mathcal{N}_\gamma^{\text{Net}})$ is the set of quantum network behaviors having noise model $\mathcal{N}_\gamma^{\text{Net}}$. The network's noise robustness is defined with respect to a nonclassicality witness $(\beta, S_{\text{Net}})$ as

$$\gamma_0 = \sup\left\{\gamma \in [0,1] \;\middle|\; \beta \geq S_{\text{Net}}(\mathbf{P}^\star) = \max_{\mathbf{P} \in \mathcal{Q}^{\text{Net}}(\mathcal{N}_\gamma^{\text{Net}})} S_{\text{Net}}(\mathbf{P})\right\}. \tag{10.5}$$

However, the challenge still remains of finding the maximal Bell score for a given noisy network.

Note that the critical noise parameter $\gamma_0 \in [0,1]$ corresponds to a single point in the domain of noise parameters. We can also consider cases where there are many noise parameters such that $|\vec{\gamma}| > 1$. In this case, the set of critical noise parameters $\vec{\gamma}_0 \in [0,1]^{|\vec{\gamma}|}$ forms a surface. The surface corresponds to the set of all $\vec{\gamma}_0$ for which $\beta = S_{\text{Net}}(\mathbf{P}^\star)$.

**Theorem 1.** Consider a linear nonclassicality witness $(0, \mathbf{G})$ that has been scaled such that the classical upper bound is zero and the maximal possible score is one where $1 = \max_{\mathbf{P} \in \mathcal{P}_{\mathcal{Y}|\mathcal{X}}} \langle \mathbf{G}, \mathbf{P} \rangle$. If white noise is applied to a behavior as

$$\widetilde{\mathbf{P}}_\gamma = \mathbf{W}_\gamma \mathbf{P} = (1-\gamma)\mathbf{P} + \gamma \mathbf{W} \quad \text{where} \quad \mathbf{W} = \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} \frac{1}{|\mathcal{Y}|} |y\rangle\langle x| \,, \tag{10.6}$$

then the nonclassicality is broken at the critical noise parameter

$$\gamma_0 = \frac{\langle \mathbf{G}, \mathbf{P}^Q \rangle}{\langle \mathbf{G}, \mathbf{P}^Q \rangle - \langle \mathbf{G}, \mathbf{W} \rangle} \tag{10.7}$$

where $\mathbf{P}^Q$ is the optimal quantum behavior for the considered system.

*Proof.* The optimal quantum strategy is unchanged in the presence of white noise, therefore the game score is expressed for the noisy behavior as

$$\left\langle \mathbf{G}, \widetilde{\mathbf{P}}_\gamma^Q \right\rangle = (1-\gamma) \left\langle \mathbf{G}, \mathbf{P}^Q \right\rangle + \gamma \left\langle \mathbf{G}, \mathbf{W} \right\rangle. \tag{10.8}$$

We seek the critical parameter at which the nonclassity witness is no longer violated, therefore, we set Eq. (10.8) equal to zero and solve for $\gamma_0$

$$0 = (1-\gamma_0) \left\langle \mathbf{G}, \mathbf{P}^Q \right\rangle + \gamma_0 \left\langle \mathbf{G}, \mathbf{W} \right\rangle \tag{10.9}$$

$$\gamma_0(\left\langle \mathbf{G}, \mathbf{P}^Q \right\rangle - \left\langle \mathbf{G}, \mathbf{W} \right\rangle) = \left\langle \mathbf{G}, \mathbf{P}^Q \right\rangle \tag{10.10}$$

$$\gamma_0 = \frac{\left\langle \mathbf{G}, \mathbf{P}^Q \right\rangle}{\left\langle \mathbf{G}, \mathbf{P}^Q \right\rangle - \left\langle \mathbf{G}, \mathbf{W} \right\rangle}. \tag{10.11}$$

□

The noise robustness of nonclassicality can characterize the value of quantum communication resources.

If a quantum network does not exhibit nonclassical behaviors, then the operational advantages derived from quantum nonclassicality must also vanish because there exists a classical network that can simulate exactly the quantum network's behavior. If a quantum technology is to provide advantage it ought to be able to exhibit nonclassical behaviors. Thus, the noise robustness of nonclassicality can quantify a network's ability to demonstrate quantum advantages in the presence of noise.

Since the noise parameter is specific to a noise model, the noise robustness of a quantum network is defined with respect to a particular noise model. For this reason, the robustness should not be quantitatively compared between different noise models, even when the same network topology is considered in both cases. On the other hand, the noise robustness of two different networks with similar noise models can be compared. As a result, the relative noise robustness of different quantum network topologies can be compared.

## *10.3 Optimizing Nonclassicality in Noisy Quantum Networks*

Evaluating a given network's robustness to a noise model $\mathcal{N}_\gamma^{\text{Net}}$ requires following optimization problem to be solved

$$S_{\text{Net}}(\mathbf{P}^\star) = \max_{\mathbf{P}\in\mathcal{Q}^{\text{Net}}(\mathcal{N}^{\text{Net}})} S_{\text{Net}}(\mathbf{P}). \tag{10.12}$$

Since the network's probabilities decompose as

$$P(\vec{b}|\vec{x},\vec{y}) = \text{Tr}\left[\Pi_{\vec{b}|\vec{y}}^{\text{Net}} \mathcal{N}^{\text{Net}}(\rho_{\vec{x}}^{\text{Net}})\right], \tag{10.13}$$

the maximization over $\mathbf{P} \in \mathcal{Q}^{\text{Net}}(\mathcal{N}^{\text{Net}})$ is recast as a maximization over all network state preparations $\rho_{\vec{x}}^{\text{Net}} \in D(\mathcal{H}^{\text{Prep}})$ and all network measurements $\Pi^{\text{Net}} \in \text{POVM}(\mathcal{H}^{\text{Meas}})$. For example, let $S^{\text{Net}}$ be a linear nonclassicality witness, then Eq. (10.12) becomes

$$S_{\text{Net}}(\mathbf{P}^\star) = \max_{\substack{\{\rho_{\vec{x}}^{\text{Net}}\in D(\mathcal{H}^{\text{Prep}})\}_{\vec{x}\in\mathcal{X}} \\ \{\{\Pi_{\vec{b}|\vec{y}}^{\text{Net}}\}_{\vec{b}\in\mathcal{B}}\in\text{POVM}(\mathcal{H}^{\text{Meas}})\}_{\vec{y}\in\mathcal{Y}}}} \sum_{\vec{b}}\sum_{\vec{x}}\sum_{\vec{y}} S_{\vec{b},\vec{x},\vec{y}}^{\text{Net}} \text{Tr}\left[\Pi_{\vec{b}|\vec{y}}^{\text{Net}} \mathcal{N}^{\text{Net}}(\rho_{\vec{x}}^{\text{Net}})\right]. \tag{10.14}$$

Generally, Eq. (10.14) is a bilinear optimization because both states and measurements must be optimized. If either the set of state preparations $\{\rho_{\vec{x}}^{\text{Net}}\}_{\vec{x}}$ or measurements $\{\{\Pi_{\vec{b}|\vec{y}}^{\text{Net}}\}_{\vec{b}}\}_{\vec{y}}$ is fixed, then Eq. (10.14) can be solved as a semidefinite program [92]. While there exist numerical methods for solving such bilinear programs, *e.g.* the see-saw method, they are difficult to solve in general. The main drawbacks of using numerical convex optimization techniques include there being no guarantee of convergence to a global optima, a priori requirement of characterizing the noise model $\mathcal{N}^{\text{Net}}$, and the need to perform convex optimization on a Hilbert space that grows exponentially with the number of qubits. Furthermore, when an optimization is run, the optimal state preparations and measurements are output as matrices where it is not necessarily clear how the state preparations or measurements are achieved on a quantum networking device.

To resolve some of these issues, we use variational quantum optimization to maximize the nonclassicality in Eq. (10.12). Although VQO methods are still subject to finding local optima, they provide interesting advantages over the convex optimization techniques. These advantages include not needing to characterize the noise $\mathcal{N}^{\text{Net}}$ and obtaining as output from the optimization the device settings for which the optimum is achieved. Furthermore, when implemented on quantum hardware, there is no need to store the exponentially large Hilbert space into memory.

The maximization of nonclassicality in noisy quantum networks in Eq. (10.12) is well suited for our VQO

framework. The noisy quantum network ansatz circuit can be constructed as described in Chapter 7. The cost function that is minimized during variational optimization is then expressed as

$$\text{Cost}\left(\mathbf{P}_{\text{Net}}(\Theta)\right) = -S_{\text{Bell}}(\mathbf{P}_{\text{Net}}(\Theta)), \tag{10.15}$$

where the minus sign transforms the minimization of the cost into the maximization of the Bell score. It is then a simple matter of using software such as qNetVO [108] to perform the VQO algorithm shown in Fig. 11.3. Furthermore, the network settings $\Theta$ parameterize the network's state preparation $\rho_{\vec{x}}^{\text{Net}}$ and measurements $\Pi_{\vec{b}|\vec{y}}^{\text{Net}}$.

We now formalize some notation used throughout the work. First, let $S_{\text{Bell}}(\Pi_{\vec{b}|\vec{y}}^{\text{Net}}, \mathcal{N}^{\text{Net}}, \rho_{\vec{x}}^{\text{Net}})$ be the Bell score for a fixed network state preparation $\rho_{\vec{x}}^{\text{Net}}$, noise model $\mathcal{N}^{\text{Net}}$, and measurements $\Pi_{\vec{b}|\vec{y}}^{\text{Net}}$. We define the maximal Bell score for a set of fixed state preparations as

$$S_{\text{Net}}^{\star}(\rho_{\vec{x}}^{\text{Net}}) = \max_{\Pi_{\vec{b}|\vec{y}}^{\text{Net}}} S_{\text{Net}}(\Pi_{\vec{b}|\vec{y}}^{\text{Net}}, \text{id}^{\text{Net}}, \rho^{\text{Net}}) \tag{10.16}$$

where the optimization is over all network measurements and $\text{id}^{\text{Net}}$ denotes a noiseless channel on all qubits. Furthermore, for a static noise model $\mathcal{N}^{\text{Net}}$, we define

$$\widetilde{S}_{\text{Net}}^{\star}(\mathcal{N}^{\text{Net}}) = \max_{\rho_{\vec{x}}^{\text{Net}} \in D(\mathcal{H}^{\text{Net}})} S_{\text{Net}}^{\star}\left(\mathcal{N}^{\text{Net}}(\rho_{\vec{x}}^{\text{Net}})\right), \tag{10.17}$$

where the optimization is over the set of all state preparations.

# Chapter 11

# Optimizing Noisy Quantum Communication Networks

This section introduces a general framework for simulating and optimizing noisy quantum networks. At a high-level, a quantum network's DAG and free operations is transformed into an equivalent quantum circuit representation. The quantum circuit can then be run by either a quantum computer, or on a classical computer that simulates the circuit execution. Through the use of differential programming, the simulations can be optimized in a similar manner to training a neural network.

The simulation and optimization framework described here was introduced in *Variational Quantum Optimization of Nonlocality in Noisy Quantum Networks* [102], and is implemented as a Python package called *qNetVO: the Quantum Network Variational Optimizer* [108]. The qNetVO software extends the PennyLane framework for cross-platform differential programming of quantum computers [210]. That is, using PennyLane the qNetVO simulation framework can run its simulation on a wide range of quantum hardware platforms and simulators. Furthermore, PennyLane enables our quantum simulations to be trained like neural networks by integrating with differential programming software packages such as Tensor Flow [288] or PyTorch [289], however, we mainly apply Autograd, PennyLane's default automatic differentiation software.

The variational optimization techniques applied by the qNetVO framework are compatible with both classical and quantum simulations of the network. The advantage of having an optimizer compatible with quantum hardware is two-fold. First, if the quantum system is too complex to simulate classically, the network cannot be numerically optimized on a classical computer with efficiency. Second, if the quantum circuit is evaluated on an actual quantum network, then the circuit's functionality can be optimized with respect to the noise model inherent to the network. This last point is quite significant because removes the necessity of characterizing the noise in the network when running quantum networking protocols.

## *11.1 Variational Optimization and Differential Programming*

In our optimization framework, the objective is to minimize a cost function over its variable parameters $\Theta \in \mathbb{R}^n$ to obtain the optimal settings

$$\Theta^\star = \arg\min_{\Theta} \text{Cost}\,(\Theta) \quad \text{where} \quad \text{Cost} : \mathbb{R}^n \to \mathbb{R}. \tag{11.1}$$

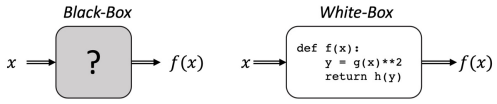



Figure 11.1: A function $f(x)$ is shown as a black-box and a white-box. The contents of the black-box are unknown but the contents of the white-box function is shown and expressed as code.

To solve the optimization problem in Eq. (11.1) we apply gradient descent [290], [291] to find local optima. This algorithm optimizes a function by incrementally stepping in the direction of steepest descent toward the minimum. In each step of this iterative procedure the settings are updated as $\Theta \rightarrow \Theta'$ where

$$\Theta' = \Theta - \eta \nabla_{\Theta} \text{Cost}\left(() \, \mathbf{P}_{\text{Net}}(\Theta)\right), \tag{11.2}$$

in which $\eta \in \mathbb{R}$ is the step size and $\nabla_{\Theta} \text{Cost}\left(\Theta\right) = (\frac{d\text{Cost}}{d\Theta_1}, \ldots, \frac{d\text{Cost}}{d\Theta_n})$ is the gradient of the cost function evaluated at $\Theta$. The gradient $\nabla_{\Theta} \text{Cost}\left(() \, \mathbf{P}_{\text{Net}}(\Theta)\right)$ is evaluated numerically using automatic differentiation [292], [293]. There are three main approach in which gradients can be evaluated computationally, numerical differentiation, automatic differentiation, and symbolic differentiation. Although we describe each method in terms of obtaining the first order derivative, each of these approaches can be used tto obtain higher-order derivatives as well.

In numerical differentiation the derivative of a function $f : \mathbb{R}^n \rightarrow \mathbb{R}$ is evaluated at point $\vec{x} \in \mathbb{R}^n$ using a finite differences rule

$$\frac{d}{dx_i} f(\vec{x}) = \frac{1}{2\varepsilon}(f(\vec{x} + \varepsilon \vec{e}_i) - f(\vec{x} - \varepsilon \vec{e}_i)) + O(\varepsilon^2) \tag{11.3}$$

where $\vec{e}_i \in \mathbb{R}^n$ is the $i^{th}$ basis vector. Note that to evaluate all $n$ derivatives, $2n$ total function evaluations are needed such that the computational complexity of finite difference gradient is linear complexity $O(n)$. Importantly, numerical differentiation approximates to first order the gradient at point $\Theta$. However, its advantage is that numerical differentiation can be applied to a black-box function with no knowledge about the function other than its input and output.

In symbolic differentiation, a software interprets mathematical expressions symbolically using mathematical rules to construct a new function [294]. The main advantage of symbolic differentiation is that it is analytic, meaning that the gradient is evaluated exact. In general, the main drawbacks of symbolic differentiation are that it requires a closed form expression of the function $f(\vec{x})$, which can grow exponentially larger than the original expression when differentiated [293]. These long expressions can be costly to evaluate and compute. It is important to note that symbolic differentiation is a white-box approach because the code of the differentiated function must be known.

Automatic differentiation is similar to symbolic differentiation in that it constructs the exact derivative function of $f(x)$ analytically. However, automatic differentiation supports programmatic control flow within the function it differentiates and improved efficiency over symbolic differentiation methods [293]. In this method a DAG modeling the function $f : \mathbb{R}^n \rightarrow \mathbb{R}$ where the inputs $\vec{x}$ flow to the output $f(\vec{x})$ with each node corresponding to a function $g_i$, and each link an intermediate variable $v_j$. To compute the gradient two passes are taken through the function $f(x)$. In the first pass, the computational DAG is constructed and th intermediate variables $v_i$ and functions $g_i$ are evaluated for the specified input $\vec{x}_0$. Then in a single backward pass, the gradient is evaluated by accumulating all of the intermediate derivatives into a linear combination of products using the chain to differentiate the nested functions, *e.g.*, $f(g(x))$ are differentiated

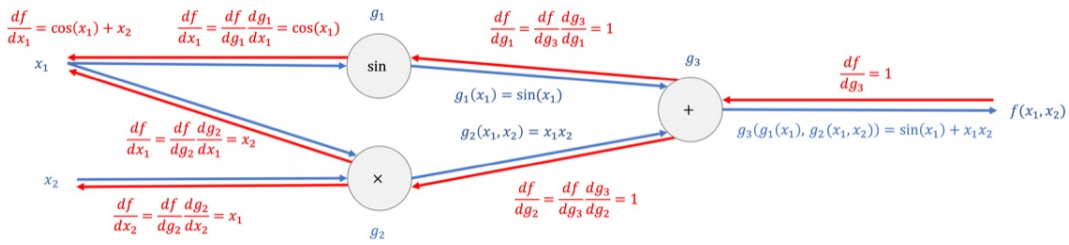



Figure 11.2: The computational DAG representing the function $f(x_1, x_2) = \sin(x_1) + x_1 x_2$ is constructed and differentiated via backpropagation. In the graph the The circular gray nodes correspond to subfunctions of $f$. The forward pass is shown using blue arrows and text. The backward pass is shown using blue arrows and text. When traversed from output to each input, a complete derivative is calculated.

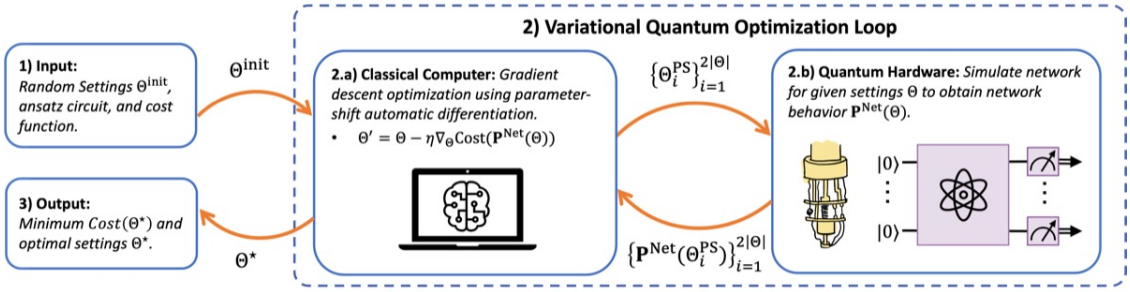



Figure 11.3: **Variational Quantum Optimization Algorithm:** As input the algorithm takes randomized settings $\Theta^{\text{init}}$, and ansatz circuit $U^{\text{Net}}(\Theta)$, and a cost function Cost $\left(\mathbf{P}^{\text{Net}}(\Theta)\right)$ that operates upon the a behavior $\mathbf{P}^{\text{Net}}$. The algorithm then begins the variational quantum optimization loop in which the classical computer uses the quantum hardware as an oracle to produce the probabilities for the ansatz circuit evaluated at the specified settings. Using the queried data, the classical computer evaluates gradients of the cost and performs a gradient descent optimization. Iterating the optimization optimizes the cost. At the end of the optimization the optimal settings $\Theta^\star$ and minimal cost Cost $(\Theta^\star)$ are output.

as $\frac{df}{dx} = \frac{df}{dg}\frac{dg}{dx}$. The algorithm for this process is referred to as backpropagation and can be evaluated in order $O(1)$ complexity where the function $f(\vec{x})$ is essentially evaluated twice, once for the forward pass and once for the backward pass (see Fig. 11.2) [295]. The advantage of backpropagation is its $O(1)$ efficiency, even when there are millions or billions of parameters. Indeed, the main drawback of backpropagation occurs when $f(x)$ cannot be evaluated efficiently, although, such scaling challenges can be offset by wide parallelization and the use of GPUs [296], [297].

## *11.2 Variational Quantum Optimization Framework*

This section outlines the hybrid variational quantum optimization procedure applied in this work. The described algorithm is implemented by the qNetVO software and can be run using remote quantum hardware or a classical simulator [108].

The main difference between variational quantum optimization and variational optimization is that the former evaluates the cost and gradients by treating a quantum device as an oracle. That is, a classical optimizer queries the oracle asking it to run a quantum circuit and the oracle returns the results for the specified quantum circuit evaluation. In this way, the quantum system can be treated as a black-box and black-box

methods for evaluating gradients can be applied. However, if the applied unitary operators associated with given optimization parameters are known, then alternative methods can be applied to evaluate the gradient.

In particular our software computes the gradient of a quantum circuit on quantum hardware using the parameter-shift rule [104], [107]. In practice, the parameter-shift rule first runs the network simulation over a collection of "shifted" settings $\{\Theta^g\}_g = \hat{\Theta}_{\text{PS}}$ and then, the classical optimizer uses the resulting circuit behaviors $\{\mathbf{P}_{\text{QC}}(\Theta^g)\}_{\Theta^g \in \hat{\Theta}_{\text{PS}}}$ to construct the gradient. Furthermore, the parameter-shift rule can be naturally extended to evaluate higher-order derivatives quantum hardware [105], however, this takes on a significant overhead in the number of circuits that must be evaluated. Similarly, the quantum natural gradient can be used to improve the gradient at the cost of additional circuit evaluations [298].

In practice, we apply the basic parameter-shift rule that is exposed by the PennyLane software [210]. The major advantage of using the parameter-shift rule is that it is more robust to the stochastic noise on a quantum computer than the finite-differences rule [299]. While the parameter-shift rule and the finite-difference rule each rely upon shifting the value of a parameter, the parameter-shift rule applies a macroscopic shift whereas the finite-differences rule applies a microscopic shift. These shifts respectively lead to large and small differences in the cost landscape. Indeed, if we allow for stochastic fluctuations in the cost evaluation, then the finite differences can lead to large errors in the calculated gradient whereas the parameter-shift rule is less susceptible to such fluctuations.

**Algorithm 1. Variational Quantum Optimization:** As input, the VQO algorithm requires a cost function $\text{Cost}\left(\mathbf{P}^{\text{Net}}(\Theta)\right)$ and a parameterized variational circuit ansatz $U^{\text{Net}}(\Theta_{\vec{x}})$. As hyperparameters, the algorithm accepts the step size $\eta$, the maximum number of gradient descent iterations `max_iter`. The VQO algorithm then proceeds as follows (see Fig. 11.3):

1. The network settings $\Theta_{\text{init}}$ are randomly initialized.
2. set `iter` $= 0$ and begin the hybrid VQO loop.
3. While `iter` $<$ `max_iter` do:
   (a) The classical optimizer constructs the collection of parameter-shift settings $\{\Theta^g\}_g = \hat{\Theta}_{\text{PS}}$.
   (b) The quantum computer evaluates $\mathbf{P}^{\text{QC}}(\Theta^g)$ for all settings in $\Theta^g \in \hat{\Theta}_{\text{PS}}$.
   (c) The classical optimizer evaluates the gradient $\nabla_\Theta \text{Cost}\left(\mathbf{P}^{\text{Net}}(\Theta)\right)$ using the parameter-shift rule and updates the network settings as $\Theta = \Theta - \eta \nabla_\Theta \text{Cost}\left(\mathbf{P}^{\text{Net}}(\Theta)\right)$.
   (d) Increment the iteration `iter` $= 1+$ `iter`.
4. Evaluate $\text{Cost}\left(\mathbf{P}^{\text{Net}}(\Theta)\right)$ on the quantum computer
5. Returns the optimal settings $\Theta$ and the minimal cost $\text{Cost}\left(\mathbf{P}^{\text{Net}}(\Theta)\right)$.

Our VQO algorithm is not guaranteed to find $\Theta^\star$, the global minimum of the cost function. The algorithm instead outputs settings $\Theta'$ that upper bounds the global minimum as $\text{Cost}(\Theta') \geq \text{Cost}(\Theta^\star)$. The bound can be improved by repeating the optimization many times with randomly initialized settings. Additionally, hyperparameters such as the step size $\eta$ can often be adjusted to improve the tightness of the bound $\text{Cost}(\Theta') \geq \text{Cost}(\Theta^\star)$.

It may also be constructive to evaluate the cost $\text{Cost}\left(\mathbf{P}^{\text{Net}}(\Theta)\right)$ each time after each step of gradient descent. In doing so, the performance of the cost minimization can be logged or tracked throughout the

optimization. This information can be used to exit the optimization loop early if a satisfactory cost is found, or the loop may also be exited if the optimization appears to be stuck in a local minimum by not making sufficient progress towards an expected minimum. However, incorporating an extra cost evaluation require at least one additional circuit to be evaluated on quantum hardware, which could be costly in time or money. Thus, it may be preferred to evaluate only once per some number of steps. On the other hand, Algorithm 1 does not evaluate the cost until after the optimization. Although this approach is most efficient, the settings obtained in the last set of gradient descent are not necessarily the most optimal because a step size that is too large can lead to oscillations in the optimized cost. That is, when the gradient descent algorithm ends up overshooting the minimum such that the cost increases rather than decreases.

## *11.3 Variational Quantum Optimization in Practice*

In this section we discuss how variational quantum optimization is applied as a hybrid quantum-classical optimization algorithm in practice. We consider VQO from two general perspectives *design* and *automation*. From a design perspective, VQO is used to optimize a simulation of a quantum network in order to gain insights. Here a quantum computer or a classical simulator is used to model a quantum network. The results the simulation and optimization can be used to assist the design of the network. On the other hand, VQO can be used to help automate a task on quantum hardware. Here a cost function encodes the desired performance of quantum hardware, hence by minimizing the cost, the quantum system is inadvertently trained to perform the desired task. Thus, VQO can serve as a mechanism for establishing and maintaining the performance of quantum protocols between networking devices.

In design tasks, the use of quantum hardware can be important because it may be able to simulate the quantum network system more efficiently. That is, a quantum computer may be able to efficiently evaluate a cost function in cases where a classical computers cannot. If there is a big enough separation between the evaluation times for classical and quantum systems, then a quantum computer may be more efficient to run the $O(N)$ number of circuits needed to evaluate the quantum gradients using the parameter shift rule for $N$ parameters, than a classical computer can perform the $O(1)$ complexity of back propagation.

To scale the performance of the quantum computations it is important to note that all parameter shifts can be run in parallel. That is, the parameter-shift settings for each parameter are evaluated independently from the remainder of the settings, meaning that they can be run in parallel on different quantum computers. Thus, as long as there are many quantum computers available, the overhead of the parameter-shift rule can be mitigated. Furthermore, as quantum computing becomes more accessible, parallelizing computation across many independent quantum computers will be a straightforward task.

### *11.3.1 The Quantum Cost of Evaluating the Cost*

In a standard variational quantum optimization task, a quantum computer or device is used as an oracle that accepts a question $\Theta$ and provides an answer $\mathbf{P}(\Theta)$. However, what is the complexity of evaluating the cost $\text{Cost}\left(\mathbf{P}(\theta)\right)$ or its gradient $\nabla_\Theta \text{Cost}\left(\mathbf{P}(\Theta)\right)$?

We first consider the number of circuit evaluations needed to obtain $\text{Cost}\left(\mathbf{P}(\Theta)\right)$. At the minimum the behavior $\mathbf{P}(\Theta)$ can be evaluated from a single circuit evaluation, However, if inputs are encoded into the settings as $\Theta = \{\Theta_x\}_{x\in\mathcal{X}}$, each $\mathbf{P}_x(\Theta_x)$ must be evaluated independently and combined as $\mathbf{P}(\Theta) = (\mathbf{P}_x(\Theta_x))_{x\in\mathcal{X}}$. Therefore a single evaluation of the cost function may require up to $|\mathcal{X}|$ circuit evaluations, hence the evaluating the cost scales as order $O(|\mathcal{X}|)$. The number of circuit evaluations is a meaningful quantifier of

complexity because there can be a nontrivial cost in setting up the quantum computer or device to run the circuit for the given settings.

We can then apply a similar analysis to the parameter-shift rule where we note that $\nabla_{\theta_i}\text{Cost}\left(\mathbf{P}(\Theta)\right) = \beta(\text{Cost}\left(\mathbf{P}(\Theta + \alpha\hat{\theta}_i)\right) - \text{Cost}\left(\mathbf{P}(\Theta - \alpha\hat{\theta}_i)\right))$ where $\hat{\theta}_i$ is a unit vector on the $i^{th}$ setting and $\alpha, \beta \in \mathbb{R}$ are determined by the unitary operator and cost function being differentiated. Indeed, the gradient for each parameter scales linearly with the number of settings $|\Theta|$. Since each cost function requires $O(|\mathcal{X}|)$ circuit evaluations, the gradient evaluation then requires $O(|\mathcal{X}| \times |\Theta|)$ circuit evaluations. Indeed, for certain variational ansatzes, or for higher-order gradients, the number of parameter-shift terms can grow polynomially, however, we will ignore that in our analysis.

The number of shots can also be considered when quantifying the cost of running a quantum circuit. That is, each shot takes a well-defined amount of time to take and is usually associated a monetary cost in practice. While the number of shots is arbitrary with respect to the variational optimization protocol, it does affect the performance of the protocol. Namely, a small number of shots leads to large fluctuations about the mean while a large number of shots leads to small fluctuations about the mean. Furthermore, the number of shots also corresponds to the numeric precision in a behavior $\mathbf{P}$. Let $N$ be the number of shots, then $P(x) \in \{\frac{i}{N}\}_{i=0}^{N}$ where the probability $P(x)$ can only take on a finite set of values. Therefore as $N \to \infty$, the probability approximately lies on the continuum $P(x) \in [0, 1]$.

Other costs can be associated with the integration between classical and quantum computing devices. In the worst case, the quantum computer is a remote device that is accessed via the cloud and the classical optimizer is a laptop computer running from a remote location on Earth. When the classical optimizer wishes to query the quantum oracle, the optimizer must first make a web request to the quantum oracle delivering the set of quantum circuits $\{U^{\text{Net}}(\Theta_i)\}_{i=1}^{M}$ that are to be run on the remote quantum computer. These circuits then wait in a queue for their chance to run on the quantum computer. The queue wait times are typically the bottleneck in the hybrid algorithm because they can be anywhere between a minute to several hours on open-access quantum computing platforms such as IBMs quantum experience. The latency of the web requests to the cloud device could also add a nontrivial cost if they amount of data to communicate is large.

While network latency and queuing costs are bottlenecks for realizing practical VQO, these bottlenecks are minor because they can be mitigated or circumvented. Namely, suppose that the collection of circuits $\{U^{\text{Net}}(\Theta_i)\}_{i=1}^{M}$ can be batched into a single request to the quantum computer where the batch queues together. By implementing batching, the number of queue wait times is reduced by a factor of $M$. On the other hand, if the quantum computer is reserved such that it cannot be used by others, and the classical optimizer is located in close proximity to the quantum device, then the queue wait times and network latency are no longer relevant to the evaluation of the quantum circuit.

### *11.3.2 Scaling Variational Quantum Optimization with Parallelization*

Scaling our VQO framework is largely limited by available quantum technology. As more qubits and larger circuit depths become available, the number of parameters $|\Theta|$ in variational ansatz can become quite large, growing exponentially with the number of qubits. As discussed in the previous section the cost of evaluating a gradient on a quantum computer scales as $O(|\mathcal{X}| \times |\Theta|)$ where $|\mathcal{X}|$ is the number of circuits needed to evaluate the cost. As we will find in our quantum networking applications, the number of network inputs $|\mathcal{X}|$ can grow exponentially with number of devices in the network. Since both $|\Theta|$ and $|\mathcal{X}|$ can become large, it is important to discuss how these algorithms can be scaled. In general, the main approaches to scaling are

parallelization of existing quantum computers, development of larger quantum computers, and improvement of fidelity and gate speed of quantum computers.

The first, most important scaling technique is parallelization across many quantum computers. In principle, all $O(|\mathcal{X}|)$ circuit evaluations needed to evaluate the cost $\text{Cost}\left(\mathbf{P}(\Theta)\right)$ or the $O(|\mathcal{X}| \times |\Theta|)$ circuit evaluations need to evaluate the gradient are independent. That is, each circuit evaluation can be run on a separate quantum computer at the same time. Hence the evaluation of the cost and its gradient is a widely parallelizable task. For this reason, we argue that the most effective to scale variational quantum algorithm is through parallelization. Since a wide range of noisy few-qubit quantum computers have already been developed, it is only a matter of manufacturing to construct more.

Scaling VQO to larger quantum systems with more qubits presents a separate challenge. Indeed, techniques are known by which a quantum circuit with many qubits is simulated using a small quantum computer [300], [301]. However, this approach increases the number of circuits to be run exponentially. While parallelization can mitigate this scaling somewhat, this technique itself is not scalable. Thus, the best way to scale to larger systems is to use a larger quantum computer. In this regime, we may find practical simulation advantages where a quantum circuit can solve computational problems more efficiently than classical hardware. However, NISQ devices have yet to show a simulation advantage [302]–[307]. Nevertheless, NISQ devices are predicted to provide simulation advantages in the near-term [38], [308] As an aside, we note that larger networks can be simulated using smaller quantum devices where an exponential increase in the number of circuit evaluations is accrued [300], [301]. Such methods are only feasible if wide parallelization across NISQ devices can offset the exponential increase in circuit evaluations.

Finally, if large quantum computers are available in bulk, the remaining strategy to improve the performance of quantum circuit evaluation is to improve the performance of the quantum computing devices. First, the noise in the device should be mitigated or corrected to the fidelity of quantum gate operations leading to more accurate gradient calculations and improved convergence of the optimization. Second, the speed at which the quantum computer initializes circuits, applies gates, and performs measurements can be improved decreasing the overall run time. In this way, the depth of quantum circuits can be increased without adding significant error or time to the circuit execution.

Overall, we argue that variational quantum algorithms are largely limited by the hardware's capability. The most import way to mitigate the cost of evaluating quantum circuits in VQO is to rely upon massive parallelization because the $O(|\mathcal{X}| \times |\Theta|)$ circuit evaluations to calculate the gradient becomes one circuit evaluation, nullifying the cost altogether. Parallelization is the most important scaling strategy because it can be implement right now through manufacturing existing quantum devices. Second, increasing the number of qubits will allow variational quantum algorithms scale to the regime where practical quantum advantages may emerge. However, since scaling the number of qubits often leads to the number of parameters $|\Theta|$ and inputs $|\mathcal{X}|$ increasing as well. Thus, scaling to more qubits will require parallelization across quantum computers. Similarly, increasing the fidelity and speed of quantum operations will lead to greater circuit depths, causing $|\Theta|$ to increase. Hence parallelization will be needed once again to offset the cost of increasing $|\Theta|$.

### *11.3.3 Challenges in Scaling Variational Quantum Optimization*

As discussed in the previous section, scaling VQO techniques to more qubits and larger circuit depths having more parameters causes the number of circuit evaluations to increase. However, the added cost of these circuit evaluations is offset by the fact that the circuits are independent and therefore, parallelizable. In this section we highlight other challenges that emerge as variational quantum algorithms scale.

A widely reported issued and major challenge in scaling variational quantum algorithms is that they can exhibit barren plateaus, a phenomenon in variational quantum algorithms where the gradient $\nabla_\Theta \text{Cost}\left(\mathbf{P}(\Theta)\right) \to 0$ approaches zero [309]. As the gradient becomes small the optimization fails to make any progress towards its optimum. The term barren plateau likens the cost landscape to a plateau, which is flat and any major change in the cost landscape occurs abruptly as a spike. Barren plateaus can have may causes including random initialization of settings [309], costs that measure nonlocal observables [310], or noise in the quantum circuit gates [311]. Furthermore, it has been observed that higher order gradients cannot completely resolve the challenge presented by barren plateaus [312].

Another major blocker to scaling variational quantum algorithms is that the computational complexity of the classical optimizer can be NP-hard [313]. This observation is quite unfortunate as it suggests that the optimizations themselves become difficult to solve as they scale in size. Furthermore, it is observed in Ref. [314] that the variational quantum eigensolver, a widely studied variational quantum algorithm, may not provide any practical advantage even on noiseless quantum devices.

Indeed, there is some healthy skepticism doubting the ability of variational quantum algorithms to scale and realize practical quantum advantage. While this may be disheartening, it also allows to expand the application of variational quantum algorithms to novel problems in which the hybrid algorithms can use their quantum resources to the fullest extent. Mainly, we argue that variational quantum algorithms could provide value in ways that do not rely upon a classical advantage. For example, preliminary results suggest that VQO can be used to optimize quantum systems against their inherent, uncharacterized noise models [102], [299].

## *11.4 Variational Optimization Applied to Quantum Networks*

Our VQO framework finds the optimal settings $\Theta^\star$ that produce a network behavior $\mathbf{P}^{\text{Net}}(\Theta^\star)$ that is optimal for a particular task, *e.g.*, violating a nonclassicality witness. In our optimization framework, the task is encoded as a cost function $\text{Cost}\left(\mathbf{P}^{\text{Net}}(\Theta)\right)$, and must be tailored to the problem at hand. The optimization objective is then expressed as a minimization of the cost function,

$$\Theta^\star = \arg\min_\Theta \text{Cost}\left(\mathbf{P}^{\text{Net}}(\Theta)\right). \tag{11.4}$$

The cost function can quantify a wide range of network properties such as entropic quantities, the distance to a desired network behavior, or the winning probability of a multipartite game.

A quantum network corresponds to a quantum circuit, and through parameterization, the quantum circuit can be transformed into a variational ansatz for the quantum network. Suppose the unitary operations $U_{x_i}^{\text{Dev}_i}$ applied by the device $\text{Dev}_i$ are parameterized by a set of real-valued settings $\vec{\theta}_{x_i}$ such that $U_{x_i}^{\text{Dev}_i} = U^{\text{Dev}_i}(\vec{\theta}_{x_i})$. The scalar values $\theta \in \theta_{x_i}$ are continuous, differentiable, and tune a unitary gate within the device's circuit. In aggregate the network is parameterized by $\Theta_{\vec{x}} = (\theta_{x_i})_i$ where $\vec{x} = (x_i)_i \in \mathcal{X}^{\text{Net}}$ is the network's classical input. Then the network's unitary $U_{\vec{x}}^{\text{Net}} = U^{\text{Net}}(\Theta_{\vec{x}})$ leads to the behavior

$$\mathbf{P}^{\text{QC}}(\Theta) = \sum_{\vec{z}\in\mathcal{Z}} \sum_{\vec{x}\in\mathcal{X}} \left| \langle \vec{z}| \, U^{\text{Net}}(\Theta_{\vec{x}}) \, |0\rangle^N \right|^2 |\vec{z}\rangle\langle\vec{x}| \,. \tag{11.5}$$

The network behavior is then constructed by applying a suitable classical postprocessing map $\mathbf{L} : \mathcal{Z} \to \mathcal{Y}$ such that $\mathbf{P}^{\text{Net}}(\Theta) = \mathbf{L}\mathbf{P}^{QC}$.

In the qNetVO software [108], we use PennyLane [210] to automatically differentiate quantum network

circuit ansatzes. In this way, qNetVO is an extension of the PennyLane framework, giving it the ability to run on a wide range of hardware platforms and simulations thereof. Furthermore, PennyLane can be integrated with new hardware and software simulators through its plugin interface.

When simulated classically, the gradients of the variational circuit ansatz can be evaluated using back propagation [293]. On quantum hardware, our software computes the gradient of a quantum circuit using the parameter-shift rule [104], [107]. In practice, the parameter-shift rule first runs the network simulation over a collection of "shifted" settings $\{\Theta^g\}_g = \hat{\Theta}_{\text{PS}}$ and then, the classical optimizer uses the resulting circuit behaviors $\{\mathbf{P}_{\text{QC}}(\Theta^g)\}_{\Theta^g \in \hat{\Theta}_{\text{PS}}}$ to construct the gradient.

### *11.4.1 Variational Quantum Optimization on Quantum Network Hardware*

In principle, the variational quantum optimization framework developed in this chapter can implemented directly on a quantum network. The key requirement is that $\Theta$, the set of free parameters of the quantum network devices, is differentiable. Furthermore, if the parameter-shift rule is used, the unitary decomposition of the applied operators must also be known. In fact, hybrid black-box optimization techniques have previously been used to maximize the violation of the CHSH inequality in a photonic system [315], [316]. Hence, our VQO framework could be extended to similar photonic implementations of quantum networks [193]–[197].

A key advantage of extending our VQO framework to quantum network hardware is that network protocols can be optimized against the noise inherent to the quantum network hardware, leading to network protocols with greater robustness to noise [102], [299]. That is, the noise and biases on quantum computers may not accurately reflect the noise and biases on quantum network hardware. Hence, communications protocols optimized on quantum network hardware will be more robust because they are tailored specially for the hardware they run on. Furthermore, the noise model does not need to be characterized for VQO methods to optimize a quantum network. Therefore, the cost of noise tomography [69], [71], [72] can be avoided altogether.

VQO on quantum networks can also be used to establish and maintain networking protocols on noisy quantum networks. Using VQO, quantum network devices may be able to automatically align photon polarization bases, maintain the communication capacity of channels, or set up device-independent protocols. Such self-organization amongst network devices may significantly reduce the manual work and expertise needed to build, scale, and maintain quantum networks. We refer to this network automation paradigm as variational quantum networking and we formally introduce its methodology in Chapter 18

## *11.5 Table of Variational Circuit Ansatzes*

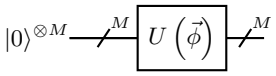
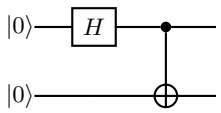
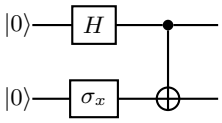
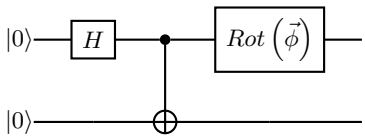
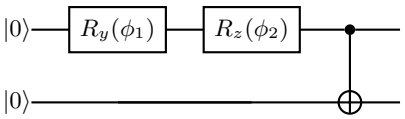
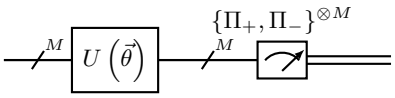
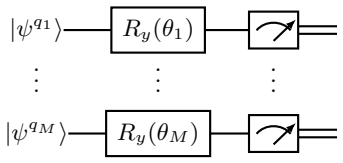
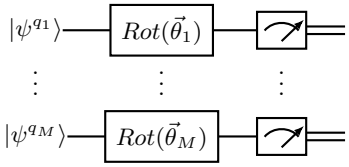
| Ansatz Name | Ansatz Circuit | Number of Parameters | Ansatz Description |
|---|---|---|---|
| Arbitrary $M$-Qubit Pure State Preparation |  | $\left\vec{\phi}\right\vert = 2^{M+1} - 2$ | Parameterizes all $M$-qubit pure state preparations. See the `ArbitraryStatePreparation` method in PennyLane for implementation details. |
| $\left\vert\Phi^+\right\rangle$ State Preparation |  | None | Prepares the state $\left\vert\Phi^+\right\rangle = (\vert 00\rangle + \vert 11\rangle)/\sqrt{2}$. |
| $\left\vert\Psi^+\right\rangle$ State Preparation |  | None | Prepares the state $\left\vert\Psi^+\right\rangle = (\vert 01\rangle + \vert 10\rangle)/\sqrt{2}$. |
| Maximally Entangled State Preparation |  | $\left\vert\vec{\phi}\right\vert = 3$ | Parameterizes all 2-qubit maximally entangled state preparations using an arbitrary qubit rotation $Rot(\vec{\phi}) = R_z(\phi_1)R_y(\phi_2)R_z(\phi_3)$. |
| Nonmaximally Entangled State Preparation |  | $\left\vert\vec{\phi}\right\vert = 2$ | Parameterizes a family of separable through maximally entangled states as $\vert\psi\rangle = \cos(\phi_1/2)\vert 00\rangle + \sin(\phi_1/2)e^{i\phi_2}\vert 11\rangle$. |
| Arbitrary $M$-Qubit Projective Measurement |  | $\left\vert\vec{\theta}\right\vert = 2^{2M} - 1$ | Parameterizes all $M$-qubit projective measurements. See the `ArbitraryUnitary` method in PennyLane for implementation details. |
| $M$-Qubit Local $R_y$ Measurement |  | $\left\vert\vec{\theta}\right\vert = M$ | Parameterizes all $M$-qubit projective measurements that decompose as $\Pi^A_{\vec{z}\vert\vec{x}} = \bigotimes_{i=1}^{M} R_y(\phi_i)\vert z^{q_i}\rangle\langle z^{q_i}\vert$. |
| Arbitrary Local Qubit Measurement |  | $\left\vert\vec{\theta}\right\vert = 3M$ | Parameterizes all $M$-qubit projective measurements that decompose as $\Pi^A_{\vec{z}\vert\vec{x}} = \bigotimes_{i=1}^{M} Rot(\vec{\phi}_i)\vert z^{q_i}\rangle\langle z^{q_i}\vert$ where $Rot(\vec{\phi}) = R_z(\phi_1)R_y(\phi_2)R_z(\phi_3)$. |

Table 11.1: Variational Circuit Ansatz Glossary: We list ansatzes by name, circuit, number of parameters, and description.

# Part III

# Results

# Chapter 12

# Nonclassicality in Local Setting of Nonsignaling Quantum Networks

**Definition 74. Local Nonsignaling Scenario:** A network scenario in which a source $\Lambda$ signals to $m$ detectors $A_i \in \vec{A}$, each having inputs $|\mathcal{X}_i|$ and outputs $|\mathcal{A}_i|$. The DAG is expressed as

$$\mathrm{LNS}(m) \equiv \mathrm{Net}(\vec{N}, \vec{E}) \quad \text{where} \quad \vec{N} = (\Lambda, \vec{A}), \quad \vec{E} = \{\Lambda \prec (\vec{A})\} \tag{12.1}$$

where $\vec{A} = (\vec{A})_{i=1}^{m}$ and the source $\Lambda$ signals to all detectors $A_i \in \vec{A}$, but the detectors do not communicate.

## *12.1 Classical Bounds on Nonsignaling Networks in the Local Setting*

Since the source $\Lambda$ shared randomness with all devices in the network, if we allow $\Lambda$ to emit an unlimited amount of randomness, the classical set forms a convex polytope $\mathcal{C}_{\Lambda}^{\mathrm{LNS}(m)}$ (see Chapter 4). Therefore, classical bounds correspond to linear inequalities that describe facets of the polytope $\mathcal{C}_{\Lambda}^{\mathrm{LNS}(m)}$. Therefore, any bound on the classical set has the form [317]

$$S_{\mathrm{LNS}(m)}(\mathbf{P}) = \langle \mathbf{G}, \mathbf{P} \rangle = \sum_{x_1 \in \mathcal{X}_1} \cdots \sum_{x_n \in \mathcal{X}_n} \sum_{a_1 \in \mathcal{A}_1} \cdots \sum_{a_n \in \mathcal{A}_n} G_{\vec{a},\vec{x}} P_{\vec{a}|\vec{x}} \leq \beta \tag{12.2}$$

where the network's behavior is expressed

$$\mathbf{P}^{\mathrm{LNS}(m)} = \sum_{\lambda} P_{\lambda}^{\Lambda} \bigotimes_{i=1}^{m} \mathbf{P}_{\lambda}^{A_i} \tag{12.3}$$

with $\mathbf{P}_{\lambda}^{A_i} : \mathcal{X}_i \to \mathcal{A}_i$ being the local behavior of device $A_i$ conditioned upon the shared random value $\lambda$. The classical bound on the inequality is found by maximizing the extreme points of the convex polytope $\mathcal{C}_{\Lambda}^{\mathrm{LNS}(m)}$

$$\beta = \max_{\mathbf{P} \in \mathcal{V}^{\mathrm{LNS}(m)}} S_{\mathrm{LNS}(m)}(\mathbf{P}). \tag{12.4}$$

### 12.1.1 Classical Bounds in Bipartite Nonsignaling Networks

In this section we consider the local nonsignaling scenario LNS(2) in which two measurement devices, $A$ and $B$, exhibit the behavior

$$\mathbf{P}^{AB} : \mathcal{X} \times \mathcal{Y} \to \mathcal{A} \times \mathcal{B} \quad \text{where} \quad \mathbf{P}^{AB} = \sum_{\lambda} P_{\lambda}^{\Lambda} \mathbf{P}_{\lambda}^{A} \otimes \mathbf{P}_{\lambda}^{B} \tag{12.5}$$

where $\mathbf{P}_{\lambda}^{A} : \mathcal{X} \to \mathcal{A}$ and $\mathbf{P}_{\lambda}^{B} : \mathcal{Y} \to \mathcal{B}$ are the local behaviors of each device conditioned upon the shared random value $\lambda$, which occurs with probability $P_{\lambda}^{\Lambda}$.

**Definition 75. Full-Correlation Inequality:** [318] Consider the local nonsignaling scenario LNS(2) in which each device's outcome is binary, $|\mathcal{A}| = |\mathcal{B}| = 2$ such that $a, b \in \mathbb{B}$. We refer to as a full-correlation nonclassicality witness any linear inequality taking the form

$$S(\mathbf{P}^{AB}) = \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} G_{x,y} \left\langle O_x^A O_y^B \right\rangle_{\Lambda} \quad \text{where} \quad \left\langle O_x^A O_y^B \right\rangle = \sum_{a,b \in \mathbb{B}} (-1)^{a \oplus b} P_{a,b|x,y}^{AB} \leq \beta. \tag{12.6}$$

Any full-correlation nonclassicality witness can be expressed as

$$S(\mathbf{P}^{AB}) = \sum_{y \in \mathcal{Y}} \left\langle O_y^{AB} \right\rangle \quad \text{where} \quad \left\langle O_y^{AB} \right\rangle = \sum_{x \in \mathcal{X}} G_{x,y} \left\langle O_x^A O_y^B \right\rangle. \tag{12.7}$$

The classical bound $\beta$ is obtained by optimizing over deterministic strategies having $O_x^A, O_y^B \in \{\pm 1\}$ to find

$$\beta = \max_{\{O_x^A \in \{\pm 1\}\}_{x \in \mathcal{X}}} \sum_{y \in \mathcal{Y}} \sum_{x \in \mathcal{X}} G_{x,y} O_x^A O_y^B = \max_{\{O_x^A \in \{\pm 1\}\}_{x \in \mathcal{X}}} \sum_{y \in \mathcal{Y}} \left| \sum_{x \in \mathcal{X}} G_{x,y} O_x^A \right| \tag{12.8}$$

where it is assumed that $O_y^B = \text{sign}(\sum_x G_{x,y} O_x^A)$.

**Definition 76. CHSH Inequality:** [13] $(S_{\text{CHSH}}, 2)$, For binary inputs $|\mathcal{X}| = |\mathcal{Y}| = 2$, the CHSH inequality is the full-correlation witness

$$S_{\text{CHSH}}(\mathbf{P}^{AB}) = \sum_{y \in \mathbb{B}} \left\langle O_y^{\text{CHSH}} \right\rangle \leq 2 \quad \text{where} \quad \left\langle O_y^{\text{CHSH}} \right\rangle = \sum_{x \in \mathbb{B}} (-1)^{x \wedge y} \left\langle O_x^A O_y^B \right\rangle \tag{12.9}$$

where $G_{x,y}^{\text{CHSH}} = (-1)^{x \wedge y}$.

### 12.1.2 Maximal CHSH Violations for Arbitrary Mixed State Preparations

Not all quantum states can be used to generate nonclassical behaviors. After all, a quantum state with a density matrix $\rho = \sum_i \mu_i |i\rangle\langle i|$ that is diagonal in the $z$-basis is a classical state, which cannot exhibit nonclassical behaviors. From a theoretical perspective, it is useful to quantify the amount of nonclassicality achievable by a given density matrix $\rho$ with respect to some nonclassicality witness $S_{\text{Net}}(\rho)$. Indeed, the Horodeckis' derived the maximal violation of the CHSH inequality for all two-qubit mixed states [319]. The proof of this result provides useful mathematical machinery for obtaining the optimal measurements for maximal CHSH violation. We now elucidate the details of this important method.

We begin by characterizing the two-qubit correlation matrix because it contains a quantum state's nonclassical content with respect to the CHSH inequality $S_{\text{CHSH}}(\mathbf{P}) \leq 2$ in Eq. (12.9).

**Definition 77. Two-Qubit Correlation Matrix:** $T_\rho \in \mathbb{R}^{3\times 3}$, The correlation matrix of a two-qubit mixed state $\rho \in D(\mathcal{H}_2^A \otimes \mathcal{H}_2^B)$ is

$$T_\rho = \sum_{i,j=1}^{3} \text{Tr}\left[\sigma_i^A \otimes \sigma_j^B \rho^{AB}\right] |i\rangle\langle j| \tag{12.10}$$

where $\sigma_1$, $\sigma_2$, and $\sigma_3$ are the Pauli matrices $\sigma_x$, $\sigma_y$, and $\sigma_z$ respectively.

Due to the homomorphism mapping $SU(2)$ to $SO(3)$ [212], local qubit unitaries $U^A, U^B \in SU(2)$ correspond to rotations $R^A, R^B \in SO(3)$ such that unitary evolution of a state $\tilde{\rho}$ corresponds to a rotation of its correlation matrix $T_{\tilde{\rho}}$ such that

$$T_\rho = R^A T_{\tilde{\rho}} (R^B)^T \iff U^A \otimes U^B \tilde{\rho}^{AB} (U^A \otimes U^B)^\dagger = \rho^A B. \tag{12.11}$$

Therefore, there exists local qubit unitaries $U^A$ and $U^B$ such that the correlation matrix is diagonal [320]

$$T_\rho = \text{diag}\,(\vec{\tau}) = R^A T_{\tilde{\rho}} (R^B)^T \tag{12.12}$$

where $\vec{\tau} \equiv (\tau_1, \tau_2, \tau_3)$ and $1 \geq \tau_1 \geq \tau_2 \geq |\tau_3| \geq 0$ are the singular values of $T_{\tilde{\rho}}$. If the qubit unitaries $U^A$ and $U^B$ can be chose freely, then the rotations $R^A$ and $R^B$ allow for the the singular values $\tau_j \in \vec{\tau}$ to be permuted and/or sign-flipped provided that $\det(T_{\tilde{\rho}}) = \det(T_\rho) = \prod_{j=1}^3 \tau_j$. Therefore, as we optimize over local qubit measurements it can be assumed without loss of generality that the correlation matrix of $\rho$ is diagonal.

Using the correlation matrix, we can now describe the Horodeckis' main result. Given a two-qubit mixed state $\rho$, the maximal CHSH score is [319]

$$S^\star_{\text{CHSH}}(\rho) = 2\sqrt{\tau_1^2 + \tau_2^2} \leq 2\sqrt{2} \tag{12.13}$$

where $\tau_j$ are the two largest singular values of $T_\rho$. Remarkably, $\rho$ can be used to generate nonlocal correlations if and only if $\tau_1^2 + \tau_2^2 > 1$. Otherwise, a state is classical with respect to the CHSH inequality. Examples of classical states include the product of two pure qubit states, $\gamma = |\psi\rangle\langle\psi| \otimes |\phi\rangle\langle\phi|$, or a shared coin flip, $\gamma = \frac{1}{2}(|00\rangle\langle 00| + |11\rangle\langle 11|)$.

To obtain the maximal CHSH violation, $S^\star_{\text{CHSH}}(\rho)$, there are two choices of optimal qubit observables,

$$\{O_x^A = (1-x)\sigma_1 + (x)\sigma_x,\ O_y^B = \tfrac{\tau_1\sigma_1 + \tau_2(-1)^y\sigma_x}{\sqrt{\tau_1^2+\tau_2^2}}\} \quad \text{or} \quad \{\widehat{O}_x^A = \tfrac{\tau_1\sigma_1 + \tau_2(-1)^x\sigma_2}{\sqrt{\tau_1^2+\tau_2^2}},\ \widehat{O}_y^B = (1-y)\sigma_1 + y\sigma_2\} \tag{12.14}$$

where in the two cases, the observables are swapped as $O_x^A = \widehat{O}_y^B$ and $O_y^B = \widehat{O}_x^A$. This fact follows from $T_\rho$ being being symmetric and the two devices being indistinguishable. However note that one party always measures using mutually unbiased bases $\sigma_1$ and $\sigma_2$ while the other device measures observables that can be nonorthogonal such that $\text{Tr}\,[O_0 O_1] > 0$. From a self-testing perspective, a maximal CHSH score, $S^\star_{\text{CHSH}}(\rho)$, is insufficient on its own to determine with certainty which of the two parties measured in mutually unbiased bases.

Although the two optimal measurement strategies in Eq. (12.14) yield the same maximal CHSH score $S^\star_{\text{CHSH}}(\rho)$, they yield distinct expectations for the CHSH bipartite correlator

$$\langle O_y^{AB}\rangle_\rho = \sqrt{\tau_1^2+\tau_2^2} = \frac{1}{2}S^\star_{\text{CHSH}}(\rho) \quad \text{and} \quad \langle \widehat{O}_y^{AB}\rangle_\rho = \frac{2}{\sqrt{\tau_1^2+\tau_2^2}}\left(\tau_1^2(1-y) + \tau_2^2 y\right). \tag{12.15}$$

where $O_y^{AB} = (O_0^A + (-1)^y O_1^A) \otimes O_y^B$ and $\widehat{O}_y^{AB} = (\widehat{O}_0^A + (-1)^y \widehat{O}_1^A) \otimes \widehat{O}_y^B$. As we will see later in star and chain networks, the distinct optimal CHSH observables in Eq. (12.15) will play an important role in their nonclassicality.

### *12.1.3 Unital Noise Robustness of CHSH Nonlocality*

In this section, we provide a theoretical analysis of the maximal CHSH violations of locality in the presence of unital noise. For various qubit unital CPTP maps $\mathcal{U} : \mathcal{H}_2 \to \mathcal{H}_2$ where $\mathcal{U}(\mathbb{I}) = \mathbb{I}$, we derive the maximal CHSH score $\widetilde{S}^{\star}_{\text{CHSH}}(\mathcal{U}^A \otimes \mathcal{U}^B)$ that can be attained for any choice of quantum state and local observables.

**Theorem 2.** Consider the CHSH scenario in which the state preparation has a unital noise model that is separable across qubits $\mathcal{U}^A \otimes \mathcal{U}^B$. The maximal noisy CHSH score is achieved using a maximally entangled state preparation $|\psi\rangle^{\Lambda} = V^A \otimes V^B |\Phi^+\rangle$, that is,

$$\widetilde{S}^{\star}_{\text{CHSH}}(\mathcal{U}^A \otimes \mathcal{U}^B) = S^{\star}_{\text{CHSH}} \left( \mathcal{U}^A \otimes \mathcal{U}^B (|\psi\rangle\langle\psi|^{\Lambda}) \right). \tag{12.16}$$

*Proof.* Consider a two-qubit mixed state $\rho \in D(\mathcal{H}^A \otimes \mathcal{H}^B)$ and local qubit PVM observables, $O_x^A = \vec{a}_x \cdot \vec{\sigma}$ and $O_y^B = \vec{b}_y \cdot \vec{\sigma}$, where $\vec{a}_x$ are $\vec{b}_y$ Bloch vectors of unit length $|\vec{a}_x| = |\vec{b}_y| = 1$. The CHSH score is $S_{\text{CHSH}} = \sum_{y \in \mathbb{B}} \langle O_y^{\text{CHSH}} \rangle_{\rho}$ where

$$O_y^{\text{CHSH}} = (O_0^A + (-1)^y O_1^A) \otimes O_y^B = (\vec{a}_0 + (-1)^y \vec{a}_1) \cdot \vec{\sigma} \otimes \vec{b}_y \cdot \vec{\sigma}. \tag{12.17}$$

Let $\tilde{\rho}^{\Lambda} = \mathcal{U}^A \otimes \mathcal{U}^B(\rho^{\Lambda})$, then

$$\langle O_y^{\text{CHSH}} \rangle_{\tilde{\rho}^{\Lambda}} = \text{Tr}\left[ O_y^{\text{CHSH}} \mathcal{U}^A \otimes \mathcal{U}^B(\rho^{\Lambda}) \right] = \text{Tr}\left[ (\mathcal{U}^A \otimes \mathcal{U}^B)^{\dagger}(O_y^{\text{CHSH}}) \rho^{\Lambda} \right] = \text{Tr}\left[ \widetilde{O}_y^{\text{CHSH}} \rho^{\Lambda} \right] \tag{12.18}$$

where we use the fact that the adjoint of a unital channel is unital. Given local qubit unitary freedom on the unital channels' input and output, the channels $\mathcal{U}^A$ and $\mathcal{U}^B$ can be diagonalized such that their Bloch matrix representation is $M_{\mathcal{U}^{A\star}} = \text{diag}(u_x^A, u_y^A, u_z^A)$ where $u_x^A \geq u_y^A \geq u_z^A$ and similarly for $M_{\mathcal{U}^{B\star}}$. Then, the correlation operator for observable $O_y^{\text{CHSH}}$ is expressed in terms of Pauli vectors as $T_{O_y^{\text{CHSH}}} = |\vec{a}_0 + (-1)^y \vec{a}_1\rangle\langle\vec{b}_y|$ and the correlation operator of the noisy observable $\widetilde{O}_y^{\text{CHSH}}$ is

$$T_{\widetilde{O}_y^{\text{CHSH}}} = M_{\mathcal{U}^{A\star}} |\vec{a}_0 + (-1)^y \vec{a}_1\rangle\langle\vec{b}_y| M_{\mathcal{U}^{B\star}} = |\vec{\alpha}_0^i + (-1)^y \vec{\alpha}_1\rangle\langle\vec{\beta}_y|, \tag{12.19}$$

where $\vec{\alpha}_x = (u_x^A a_x,\, u_y^A a_y,\, u_z^A a_z)$ and $\vec{\beta}_y = (u_x^B b_x,\, u_y^B b_y,\, u_z^B b_z)$. We make the important observation that $\text{Rank}(T_{\widetilde{O}_y^{\text{CHSH}}}) \leq 1$ where the rank is 0 in only the case where the unital channel is fully depolarizing $u_j = 0$ for all $j \in \{x, y, z\}$. It follows that $\text{Rank}\left( T_{\widetilde{O}_{y=0}^{\text{CHSH}}} + T_{\widetilde{O}_{y=1}^{\text{CHSH}}} \right) \leq 2$. Furthermore, observables $O_x^A$ and $O_y^B$ can always be chosen such that $\text{Tr}\left[ T^T_{\widetilde{O}_{y=0}^{\text{CHSH}}} T_{\widetilde{O}_{y=1}^{\text{CHSH}}} \right] = 0$ are orthogonal. Indeed, orthogonality is satisfied provided that either of the following equalities hold

$$\langle \vec{\alpha}_0 + \vec{\alpha}_1 | \vec{\alpha}_0 - \vec{\alpha}_1 \rangle = (u_x^A)^2 (a_{0,x}^2 - a_{1,x}^2) + (u_y^A)^2 (a_{0,y}^2 - a_{1,y}^2) + (u_z^A)^2 (a_{0,z}^2 - a_{1,z}^2) = 0, \tag{12.20}$$

$$\left\langle \vec{\beta}_0 | \vec{\beta}_1 \right\rangle = (u_x^B)^2 b_{0,x} b_{1,x} + (u_y^B)^2 b_{0,y} b_{1,y} + (u_z^B)^2 b_{0,z} b_{1,z} = 0. \tag{12.21}$$

Next, Proposition 1 from Zhang *et al.* [287] states that for any $4 \times 4$ matrix $M = \sum_{i,j=1}^{3} t_{i,j} \sigma_i \otimes \sigma_j$

whose correlation operator $T_M = \sum_{i,j=1}^{3} t_{i,j} |i\rangle\langle j|$ satisfies $\text{Rank}(T_M) \leq 2$, the quantity $\text{Tr}[M\rho]$ is maximized by a maximally entangled state $\rho = |\psi\rangle\langle\psi|$ where $|\psi\rangle = V^A \otimes V^B |\Phi^+\rangle$. It follows that there exists a maximally entangled state $\rho = |\psi\rangle\langle\psi|$ that maximizes simultaneously $\text{Tr}\left[\widetilde{O}^{\text{CHSH}}_{y=0}\rho\right]$, $\text{Tr}\left[\widetilde{O}^{\text{CHSH}}_{y=1}\rho\right]$, and $\text{Tr}\left[\left(O^{\text{CHSH}}_{y=0} + O^{\text{CHSH}}_{y=1}\right)\rho\right]$ because the correlation operators $T_{\widetilde{O}^{\text{CHSH}}_{y=0}}$ and $T_{\widetilde{O}^{\text{CHSH}}_{y=1}}$ are each rank-one and can be chosen to be orthogonal. That is, the correlation operator for a maximally entangled state is diagonalized as $T_\rho = \text{diag}(1, -1, 1)$, hence $\rho$ can simultaneously maximize two observables if their correlation operators are orthogonal as $\text{Tr}\left[T_{\widetilde{O}^{\text{CHSH}}_{y=0}} T_{\widetilde{O}^{\text{CHSH}}_{y=1}}\right]$. Since there exists a maximally entangled state, $|\psi\rangle\langle\psi|^\Lambda$, that maximizes the expectation of $\left\langle O^{\text{CHSH}}_y \right\rangle_{\rho^\Lambda}$ for both $y = 0, 1$, maximally entangled states are optimal for achieving the maximal CHSH score in the presence of unital qubit noise. □

Theorem 2 is quite remarkable because it indicates that the presence of unital qubit noise does not alter the optimal state preparations for CHSH violation. However, multi-qubit unital noise models do not necessarily preserve the maximal state preparation. Nevertheless, Theorem 2 allows to derive the maximal CHSH violation in a straight forward manner because the optimal state preparation is known.

**Theorem 3.** Consider a CHSH scenario that has unital qubit noise model $\mathcal{U}^A \otimes \mathcal{U}^B$. The maximal CHSH score is

$$\widetilde{S}^\star_{\text{CHSH}}(\mathcal{U}^A \otimes \mathcal{U}^B) = 2\sqrt{(u^A_0 u^B_0)^2 + (u^A_1 u^B_1)^2}, \tag{12.22}$$

where $u_0 \geq u_1$ are the two largest singular values of matrix $M_\mathcal{U}$.

*Proof.* By Theorem 2, the optimal state preparation for CHSH violation is a maximally entangled state $|\psi\rangle = V^A \otimes V^B |\Phi^+\rangle$ where the Bell state's correlation operator is $T_{|\Phi^+\rangle\langle\Phi^+|} = \text{diag}(1, -1, 1)$. Since we allow qubit unitary freedom on the state preparation and measurement, we can diagonalize the unital qubit channels in the Pauli basis as $M_{\mathcal{U}^A} = \text{diag}(u^A_1, u^A_2, u^A_0)$ and $M_{\mathcal{U}^B} = \text{diag}(u^B_1, u^B_2, u^B_0)$ where $u_0 \geq u_1 \geq u_2$. Then, $\rho^\Lambda = \mathcal{U}^A \otimes \mathcal{U}^B(|\Phi^+\rangle\langle\Phi^+|)$ can be expressed as

$$T_{\rho^\Lambda} = M_{\mathcal{U}^A} T_{|\Phi^+\rangle\langle\Phi^+|} M^T_{\mathcal{U}^B} = \text{diag}(u^A_1 u^B_1, -u^A_2 u^B_2, u^A_0 u^B_0). \tag{12.23}$$

Noting that $\tau_j = u^A_j u^B_j$, we use Eq. (12.13) to find

$$S^\star_{\text{CHSH}}(\mathcal{U}^A \otimes \mathcal{U}^B) = 2\sqrt{(u^A_0 u^B_0)^2 + (u^A_1 u^B_1)^2}. \tag{12.24}$$

Finally, since $T_{|\Phi^+\rangle\langle\Phi^+|}$, $M^A_\mathcal{U}$, and $M^B_\mathcal{U}$ are all diagonal, the singular values of $T_{\rho^\Lambda}$ and the resulting Bell scores are maximal. □

We can now apply Theorem 3 to explicitly evaluate the noise robustness of the CHSH violation with respect to different types of unital qubit channels. Each derivation requires only the calculation of each unital channel's singular values $M_{\mathcal{U}^\star} = \text{diag}(u_0, u_1, u_2)$, then Theorem 3 can be applied directly.

**Proposition 4.** Consider a CHSH scenario having a qubit depolarizing noise model, $\mathcal{W}^A_{v^A} \otimes \mathcal{W}^B_{v^B}$. The maximal $n$-local star and chain scores are

$$\widetilde{S}^\star_{\text{CHSH}}(\mathcal{W}^A_{v^A} \otimes \mathcal{W}^B_{v^B}) = 2\sqrt{2}\left(v^A v^B\right), \tag{12.25}$$

*Proof.* A qubit depolarizing channel is diagonalized in the Pauli basis as $M_{\mathcal{D}_v} = \text{diag}(v, v, v)$. Eq. (12.25) follows directly from Theorem 3. □

**Proposition 5.** Consider a star or chain quantum network having a qubit dephasing noise model, $\mathcal{D}^A_{\gamma^A} \otimes \mathcal{D}^B_{\gamma^B}$. The maximal $n$-local star and chain scores are

$$\widetilde{S}^{\star}_{\text{CHSH}}(\mathcal{D}^A_{v^A} \otimes \mathcal{D}^B_{v^B}) = \prod_{i=1}^{n} \left(1 + (1-\gamma^A)(1-\gamma^B)\right)^{\frac{1}{2n}}, \tag{12.26}$$

*Proof.* A qubit dephasing channel is diagonalized in the Pauli basis as $M_{\mathcal{D}_\gamma} = \text{diag}(\sqrt{1-\gamma}, \sqrt{1-\gamma}, 1)$. Eq. (12.26) follows directly from Theorem 3. □

Up to this point, we have only considered the noise robustness of the CHSH scenario with respect to single qubit unital noise models. We now derive the robustness for two-qubit depolarizing and classical white noise models.

**Proposition 6.** Consider the CHSH scenario that has a two-qubit depolarizing channel $\mathcal{W}^{AB}_v \in \text{CPTP}(\mathcal{H}_4 \to \mathcal{H}_4)$ on its source. The maximal noisy CHSH score is

$$\widetilde{S}^{\star}_{\text{CHSH}}(\mathcal{W}^{AB}_v) = (1 - \frac{16}{15}\gamma)2\sqrt{2} \tag{12.27}$$

where $\gamma = 15/16(1-v)$ and $\gamma \in [0, 1]$.

*Proof.* Applying the two-qubit depolarizing channel to a two qubit state $\rho^{AB} \in D(\mathcal{H}^A_2 \otimes \mathcal{H}^B_2)$ yields $\tilde{\rho}^{AB} = \mathcal{W}_v(\rho^{AB}) = v\rho^{AB} + (1-v)\frac{1}{4}\mathbb{I}_4$. Noting that the correlation matrix $T_{\frac{1}{4}\mathbb{I}_4}$ is all zeros shows that the term $(1-v)\frac{1}{4}\mathbb{I}_4$ contributes 0 to the CHSH score. Then, note that the correlation matrix $T_{\rho^{AB}} = \text{diag}\,(\tau_0, \tau_1, \tau_2)$ where $1 \geq \tau_0 \geq \tau_1 \geq |\tau_2| \geq 0$. Thus, the noisy state $\tilde{\rho}^{AB}$ has the correlation matrix $T_{\tilde{\rho}^{AB}} = v\text{diag}\,(\tau_0, \tau_1, \tau_2)$. It follows from Eq. (12.13) that the maximal CHSH score is $\widetilde{S}^{\star}_{\text{CHSH}}(\mathcal{W}_v) = 2\sqrt{v^2(\tau_0^2 + \tau_1^2)} = v2\sqrt{\tau_0^2 + \tau_1^2}$. If the maximally entangled state $\rho^{AB} = |\Phi^+\rangle\langle\Phi^+|$ is used, then $\tau_0 = \tau_1 = 1$ and $\widetilde{S}^{\star}_{\text{CHSH}}(\mathcal{W}_v) = 2v\sqrt{2}$. Finally, substituting the noise parameter for the visibility as $\gamma = 15/16(1-v)$ where $\gamma \in [0, 1]$ recovers Eq. (12.27). □

**Proposition 7.** Consider the CHSH scenario in which each detector has a classical white noise model $\mathbf{W}^A_{\gamma^A} \otimes \mathbf{W}^B_{\gamma^B} \in \mathcal{P}_{\mathcal{A}\times\mathcal{B}|\mathcal{A}\times\mathcal{B}}$ acting on each detector's output. The maximal noisy CHSH score is

$$\widetilde{S}^{\star}_{\text{CHSH}}(\mathbf{W}^A_{\gamma^A} \otimes \mathbf{W}^B_{\gamma^B}) = (1-\gamma^A)(1-\gamma^B)2\sqrt{2}. \tag{12.28}$$

*Proof.* As given by Eq. (13.82), white noise on each dichotomic detector can be modeled by a POVM with elements $W^A_{\pm|x,\gamma^A} = (1-\gamma^A)P^A_{+|x} + \gamma W^A_{\pm|x}$ where $P^A_{\pm|x} \in \text{PVM}(\mathcal{H}^A_2)$ are projectors onto even and odd parity subspaces and $W^A_{\pm|x} = \frac{1}{2}\mathbb{I}_2$. The corresponding noisy qubit dichotomic observable is constructed

$$\widetilde{O}^A_{x,\gamma^A} = W_{+|x,\gamma^A} - W_{-|x,\gamma^A} = (1-\gamma^A)(P^A_{+|x} - P^A_{-|x}) = (1-\gamma^A)O^A_x, \tag{12.29}$$

where $O^A_x$ is the observable used in the noiseless case whereas $\widetilde{O}^B_{y,\gamma^B}$ is similarly constructed. The noisy joint observable is then $\widetilde{O}^{\text{CHSH}}_y = (1-\gamma^A)(1-\gamma^B)O^{\text{CHSH}}_y$. Finally, if $O^A_x$ and $O^B_y$ attain the maximal CHSH score of $2\sqrt{2}$ in the noiseless case, then maximal CHSH score in the noisy case is $(1-\gamma^A)(1-\gamma^B)2\sqrt{2}$. □

### *12.1.4 Nonunital Noise Robustness of CHSH Nonlocality*

In this section, we derive the maximal violations of the CHSH inequality in the presence of nonunital noise models. For an introduction to nonunital noise, please refer to Chapter 8.

**Theorem 4.** Consider the family of noisy two-qubit state $\tilde{\rho}_\lambda = \mathcal{N}^A \otimes \mathcal{N}^B(\rho_\lambda)$ where $\rho_\lambda = |\psi_\lambda\rangle\langle\psi_\lambda|$ and $\sqrt{\lambda}\,|00\rangle + \sqrt{1-\lambda}\,|11\rangle$ is a nonmaximally entangled state with $\lambda \in [0,1]$ and $\mathcal{N}^A, \mathcal{N}^B \in \mathrm{CPTP}(\mathcal{H}_2 \to \mathcal{H}_2)$ are nonunital qubit channels that can be diagonalized in the Pauli basis as

$$\mathbb{T}_{\mathcal{N}_\gamma} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & u_x & 0 & 0 \\ 0 & 0 & u_y & 0 \\ t_z & 0 & 0 & u_z \end{pmatrix}. \tag{12.30}$$

If the amount entanglement specified by $\lambda$ can be varied freely, the maximal CHSH score for a given channel two-qubit nonunital channel is

$$S^\star_{\mathrm{CHSH}}(\tilde{\rho}_{\lambda^\star}) = \max \begin{cases} 2\sqrt{(u_x^A u_x^B)^2 + (u_y^A u_y^B)^2}, \\ 2\sqrt{(u_x^A u_x^B)^2 + (t_z^A t_z^B + u_z^A u_z^B)^2 - \frac{(t_z^A t_z^B + u_z^A u_z^B)^2 (t_z^A u_z^B + u_z^A t_z^B)^2}{(t_z^A u_z^B + u_z^A t_z^B)^2 - (u_x^A u_x^B)^2}} \end{cases} \tag{12.31}$$

where $\lambda^\star = \arg\max_{\lambda\in[0,1]} S^\star_{\mathrm{CHSH}}(\tilde{\rho}_\lambda)$ is the optimal entanglement parameter for maximal CHSH violation and, $\lambda^\star = \frac{1}{2}$ if $S^\star_{\mathrm{CHSH}}(\tilde{\rho}_{\lambda^\star}) = 2\sqrt{(u_x^A u_x^B)^2 + (u_y^A u_y^B)^2}$ and otherwise,

$$\lambda^\star = \frac{1}{2}\Big(1 - \frac{(t_z^A t_z^B + u_z^A u_z^B)(t_z^A u_z^B + u_z^A t_z^B)}{(t_z^A u_z^B + u_z^A t_z^B)^2 - (u_x^A u_x^B)^2}\Big). \tag{12.32}$$

*Proof.* The correlation matrix for $\rho_\lambda = |\psi_\lambda\rangle\langle\psi_\lambda|$ is expressed

$$\mathbb{T}_{\rho_\lambda} = \begin{pmatrix} 1 & 0 & 0 & 2\lambda - 1 \\ 0 & 2\sqrt{\lambda}\sqrt{1-\lambda} & 0 & 0 \\ 0 & 0 & -2\sqrt{\lambda}\sqrt{1-\lambda} & \\ 2\lambda - 1 & 0 & 0 & 1 \end{pmatrix} \tag{12.33}$$

and the qubit nonunital channels, $\mathcal{N}^A$ and $\mathcal{N}^B$, are applied as $\mathbb{T}_{\tilde{\rho}_\lambda} = \mathbb{T}_{\mathcal{N}^A}\mathbb{T}_{\rho_\lambda}\mathbb{T}^T_{\mathcal{N}^B}$, leading to the singular values of $\mathbb{M}_{\tilde{\rho}_\lambda}$, the 3x3 correlation matrix subblock, to be

$$\tau_x = 2u_x^A u_x^B\sqrt{\lambda(1-\lambda)}, \quad \tau_y = 2u_y^A u_y^B\sqrt{\lambda(1-\lambda)}, \quad \text{and} \quad \tau_z = t_z^A t_z^B + u_z^A u_z^B + (2\lambda-1)(t_z^A u_z^B + u_z^A t_z^B). \tag{12.34}$$

Using Eq. (12.13), we find $\max_{\lambda\in[0,1]} S^\star_{\mathrm{CHSH}}(\tilde{\rho}_\lambda)$ by checking the singular value pairs $(\tau_x, \tau_y)$ and $(\tau_x, \tau_z)$ where it is assumed without loss of generality that $|\tau_x| \geq |\tau_y|$. For the pair $(\tau_x, \tau_y)$, we find

$$\max_{\lambda\in[0,1]} \tau_x^2 + \tau_y^2 = \max_{\lambda\in[0,1]} 4\lambda(1-\lambda)\Big((u_x^A u_x^B)^2 + (u_y^A u_y^B)^2\Big) = (u_x^A u_x^B)^2 + (u_y^A u_y^B)^2, \tag{12.35}$$

where $\lambda^\star = \frac{1}{2}$ is the optimal entanglement parameter. Inserting Eq. (12.35) into the maximal CHSH score in Eq. (12.13) achieves the maximal violation in the upper case of Eq. (12.31). Next, we consider the singular

value pair $(\tau_x, \tau_z)$ where

$$\max_{\lambda\in[0,1]} \tau_x^2 + \tau_z^2 = \max_{\lambda\in[0,1]} 4\lambda(1-\lambda)(u_x^A u_x^B)^2 + \Big(t_z^A t_z^B + u_z^A u_z^B + (2\lambda-1)(t_z^A u_z^B + u_z^A t_z^B)\Big)^2. \tag{12.36}$$

To maximize Eq. (12.36), we take the derivative of the RHS with respect to $\lambda$ and set the result equal to 0. Solving for the critical point, we find $\lambda^\star$ as given by Eq. (12.32). Next, inserting $\lambda^\star$ into Eq. (12.36) and rearranging the terms recovers the lower case of Eq. (12.31). □

Theorem 4 is significant because for a broad family of qubit nonunital channels, the maximal CHSH violation can be computed. The nonunital channel in Eq. (12.30) can describe both loss channels and amplitude damping channels. Furthermore, the maximal violation is found to not always be by a maximally entangled state. In fact, based on the parameters in the matrices $\mathbb{T}_{\mathcal{N}^A}$ and $\mathbb{T}_{\mathcal{N}^B}$ it can be asserted whether a maximally or nonmaximally entangled state achieves the maximum. Thus, the CHSH violation of nonunital qubit channels is distinct from unital qubit channels because the noise model optimal state is no longer the maximally entangled state as shown in Theorem 3. As an example, we now derive the maximal violation of both maximally and nonmaximally entangled states with amplitude damping on each qubit. We find a region of noise parameter for which nonmaximally entangled states achieve larger CHSH violations than maximally entangled states.

**Proposition 8.** Consider the CHSH scenario in which a maximally entangled state $\rho = |\psi\rangle\langle\psi|$ is prepared where $|\psi\rangle = U^A \otimes U^B |\Phi^+\rangle$. When an amplitude damping channel is applied to each qubit as $\tilde{\rho} = \mathcal{A}_{\gamma^A} \otimes \mathcal{A}_{\gamma^B}(\rho)$, the CHSH inequality is violated only if

$$(1-\gamma^A)(1-\gamma^B) > \frac{1}{2} \tag{12.37}$$

and for $\gamma^A$ and $\gamma^B \in [0, \frac{1}{2}]$, the maximal CHSH score is

$$S^\star_{\text{CHSH}}(\tilde{\rho}) = \sqrt{2(1-\gamma^A)(1-\gamma^B)}. \tag{12.38}$$

*Proof.* Let $\rho = |\Phi^+\rangle\langle\Phi^+|$ be a maximally entangled state preparation and $u_x = u_y = \sqrt{1-\gamma}$, $u_z = (1-\gamma)$, and $t_z = \gamma$ be the nonunital channel parameters for the amplitude damping channel. Substituting these values into the singular value expressions listed in the proof of Theorem 4, we find

$$\tau_x = \tau_y = \sqrt{(1-\gamma^A)(1-\gamma^B)} \tag{12.39}$$

$$\tau_z = \gamma^A\gamma^B + (1-\gamma^A)(1-\gamma^B) \tag{12.40}$$

In the domain $\gamma^A, \gamma^B \in [0, \frac{1}{2}]$, the maximal singular values are $(\tau_x, \tau_y)$. Therefore, Eq. (12.13) calculates the maximal CHSH score to be

$$S^\star_{\text{CHSH}}(\tilde{\rho}) = 2\sqrt{2(1-\gamma^A)(1-\gamma^B)}, \tag{12.41}$$

for which the CHSH inequality is violated if and only if $(1-\gamma^A)(1-\gamma^B) > \frac{1}{2}$. For all $\gamma^A \in [\frac{1}{2}, 1]$ and $\gamma^B \in [0,1]$ no violations occur because $\tau_x^2 + \tau_z^2 \leq 1$ and $\tau_x^2 + \tau_y^2 \leq 1$ (and similarly for the values $\gamma^A \in [0,1]$ and $\gamma^B \in [\frac{1}{2}, 1]$). Thus, the CHSH inequality is violated only when $(1-\gamma^A)(1-\gamma^B) > \frac{1}{2}$. Since $\tau_x = \tau_y$ are the two maximal eigenvalues of $R_\rho$ when $\gamma^A, \gamma^B \in [0, \frac{1}{2}]$, Eq. (12.13) can be applied to recover the $n$-local star score in Eq. (12.38). Finally, the Bell state $|\Phi^+\rangle\langle\Phi^+|$ and amplitude damping channel are both diagonal

in the Pauli basis, therefore the evaluated score is maximal. □

**Proposition 9.** Consider the CHSH scenario with single qubit amplitude damping noise $\mathcal{A}_\gamma^A \otimes \mathrm{id}^B$ where $\gamma^A = \gamma$. For a noisy nonmaximally entangled state preparation $\tilde{\rho}_{\lambda,\gamma} = \mathcal{A}_\gamma^A \otimes \mathrm{id}^B(|\psi_\lambda\rangle\langle\psi_\lambda|)$ where $|\psi_\lambda\rangle = \sqrt{\lambda}\,|00\rangle + \sqrt{1-\lambda}\,|11\rangle$, the maximal CHSH score is

$$S^\star_{\mathrm{CHSH}}(\tilde{\rho}_{\lambda^\star,\gamma}) = \begin{cases} 2\sqrt{2(1-\gamma)}, & \gamma \in [0, \frac{1}{2}] \\ 2, & \gamma \in [\frac{1}{2}, 1] \end{cases} \tag{12.42}$$

where $\lambda^\star = \frac{1}{2}$ when $\gamma \in [0, \frac{1}{2}]$ and $\lambda^\star = 1$ when $\gamma \in [\frac{1}{2}, 1]$.

*Proof.* Considering nonmaximally entangled states, we use Theorem 4 to find that

$$\lambda^\star = \frac{1}{2}\left(1 - \frac{(1-\gamma)\gamma}{\gamma^2 - (1-\gamma)}\right), \tag{12.43}$$

which yields

$$S^\star_{\mathrm{CHSH}}(\tilde{\rho}_\lambda) = 2\sqrt{(1-\gamma) + (1-\gamma)^2 - \frac{(1-\gamma)^2\gamma^2}{\gamma^2 - (1-\gamma)}}. \tag{12.44}$$

However, $\lambda^\star = 1$ when $\gamma = \frac{1}{2}$ and within the domain $\gamma \in [0, \frac{1}{2}]$ it can be verified that

$$(1-\gamma) \geq (1-\gamma)^2 - \frac{(1-\gamma)^2\gamma^2}{\gamma^2 - (1-\gamma)}. \tag{12.45}$$

Thus, when amplitude damping is applied to a single qubit, maximally entangled state preparations are optimal on the range $\gamma \in [0, \frac{1}{2}]$ and the classical state preparation is optimal for $\gamma \in [\frac{1}{2}, 1]$. □

**Proposition 10.** Consider the CHSH scenario with uniform qubit amplitude damping noise $\mathcal{A}_\gamma^A \otimes \mathcal{A}_\gamma^B$ where $\gamma^A = \gamma^B = \gamma$. For a noisy nonmaximally entangled state preparation $\tilde{\rho}_{\lambda,\gamma} = \mathcal{A}_\gamma^A \otimes \mathcal{A}_\gamma^B(|\psi_\lambda\rangle\langle\psi_\lambda|)$ where $|\psi_\lambda\rangle = \sqrt{\lambda}\,|00\rangle + \sqrt{1-\lambda}\,|11\rangle$, the maximal CHSH score is

$$S^\star_{\mathrm{CHSH}}(\tilde{\rho}_{\lambda^\star,\gamma}) = \begin{cases} 2\sqrt{2(1-\gamma)^2}, & \gamma \in [0, \gamma_c] \\ 2\sqrt{(1-\gamma)^2 - \frac{(2\gamma^2-2\gamma+1)^2}{4\gamma^2-1}}, & \gamma \in [\gamma_c, \frac{1}{3}] \\ 2 & \gamma \in [\frac{1}{3}, 1] \end{cases} \tag{12.46}$$

where the optimal entanglement parameter is

$$\lambda^\star = \begin{cases} \frac{1}{2}, & \gamma \in [0, \gamma_c] \\ \frac{1}{2}\left(1 - \frac{(\gamma^2+(1-\gamma)^2)(2\gamma(1-\gamma))}{(2\gamma(1-\gamma))^2-(1-\gamma)^2}\right) & \gamma \in [\gamma_c, \frac{1}{3}] \\ 1, & \gamma \in [\frac{1}{3}, 1] \end{cases}, \tag{12.47}$$

the crossover noise parameter is

$$\gamma_c = \frac{1}{6}\left(4 + \frac{(3\sqrt{114} - 32)^{\frac{1}{3}}}{2^{\frac{2}{3}}} - \frac{1}{(6\sqrt{114} - 64)^{\frac{1}{3}}}\right), \tag{12.48}$$

and the nonlocality is broken with respect to the CHSH inequality when $\gamma = \gamma_0 = \frac{1}{3}$.

*Proof.* Using Theorem 4, we substitute the values $u_x = u_y = \sqrt{1-\gamma}$, $u_z = (1-\gamma)$, and $t_z = \gamma$ into Eq. (12.31) and Eq. (12.32) to obtain Eq. (12.46) and Eq. (12.47). Note that when $\gamma \in [\frac{1}{3}, 1]$, then $\lambda^\star = 1$ and the optimal state preparation corresponds to the classical state $|00\rangle\langle 00|$. Setting $\lambda = 1$ in Eq. (12.36) yields $S^\star_{\text{CHSH}}(\tilde{\rho}_{\lambda=1}) = 2$, implying that the CHSH nonlocality is broken for all nonmaximally entangled states when $\gamma = \frac{1}{3}$. For $\gamma \geq \frac{1}{3}$, the classical state preparation $|00\rangle\langle 00|$ is optimal and achieves the classical bound. To obtain the crossover noise parameter $\gamma_c$ we set the two cases in Eq. (12.50) to be equal and solve for $\gamma$ to find Eq. (12.48). □

## *12.2 Variational Optimization of Nonlocality in Noisy Networks*

### *12.2.1 Maximizing CHSH Violation in the Presence of Nonunital Noise*

In this section we demonstrate that variational optimization methods can find the maximal violations in the presence of amplitude damping noise. Hence, we reproduce the theoretical results obtained in Section 12.1.4, in which we show that when uniform qubit amplitude damping noise is present, the maximal CHSH violation is not obtained by an entangled state.

Consider the CHSH scenario in which uniform qubit amplitude damping noise is applied to a maximally entangled $\rho_{\lambda=\frac{1}{2}}$ state as $\tilde{\rho}_{\lambda=\frac{1}{2}} = \mathcal{A}_\gamma \otimes \mathcal{A}_\gamma(\rho_{\lambda=\frac{1}{2}})$ where the maximally entangled state. For maximally entangled states, we use Proposition 15 to find that the maximal CHSH score is

$$S^\star_{\text{CHSH}}(\tilde{\rho}_{\frac{1}{2},\gamma}) = \begin{cases} 2\sqrt{2(1-\gamma)^2}, & \gamma \in [0, \frac{1}{2}] \\ 2\sqrt{(1-\gamma)^2 + (2\gamma^2 - 2\gamma + 1)^2}, & \gamma \in [\frac{1}{2}, 1] \end{cases} \tag{12.49}$$

where $\gamma_{c,\frac{1}{2}} = \frac{1}{2}$ is the noise parameter at which the crossover occurs between the two curves in Eq. (12.49) Furthermore, we see from Eq. (12.49) that the CHSH nonlocality is broken for all maximally entangled states $\rho_{\lambda=\frac{1}{2}}$ at the critical noise parameter $\gamma_{0,\frac{1}{2}} = (1 - \frac{1}{\sqrt{2}})$.

Next, we show in Proposition 10 that for noisy nonmaximally entangled states, the maximal CHSH score is

$$S^\star_{\text{CHSH}}(\tilde{\rho}_{\lambda^\star,\gamma}) = \begin{cases} 2\sqrt{2(1-\gamma)^2}, & \gamma \in [0, \gamma_c] \\ 2\sqrt{(1-\gamma)^2 - \frac{(2\gamma^2-2\gamma+1)^2}{4\gamma^2-1}}, & \gamma \in [\gamma_c, \frac{1}{3}] \\ 2 & \gamma \in [\frac{1}{3}, 1] \end{cases} \tag{12.50}$$

where the optimal entanglement parameter is

$$\lambda^\star = \begin{cases} \frac{1}{2}, & \gamma \in [0, \gamma_c] \\ \frac{1}{2}\left(1 - \frac{(\gamma^2+(1-\gamma)^2)(2\gamma(1-\gamma))}{(2\gamma(1-\gamma))^2-(1-\gamma)^2}\right) & \gamma \in [\gamma_c, \frac{1}{3}] \\ 1, & \gamma \in [\frac{1}{3}, 1] \end{cases}, \tag{12.51}$$

the crossover noise parameter is

$$\gamma_c = \frac{1}{6}\left(4 + \frac{(3\sqrt{114} - 32)^{\frac{1}{3}}}{2^{\frac{2}{3}}} - \frac{1}{(6\sqrt{114} - 64)^{\frac{1}{3}}}\right), \tag{12.52}$$

and the nonlocality is broken with respect to the CHSH inequality when $\gamma = \gamma_0 = \frac{1}{3}$. Therefore, in the

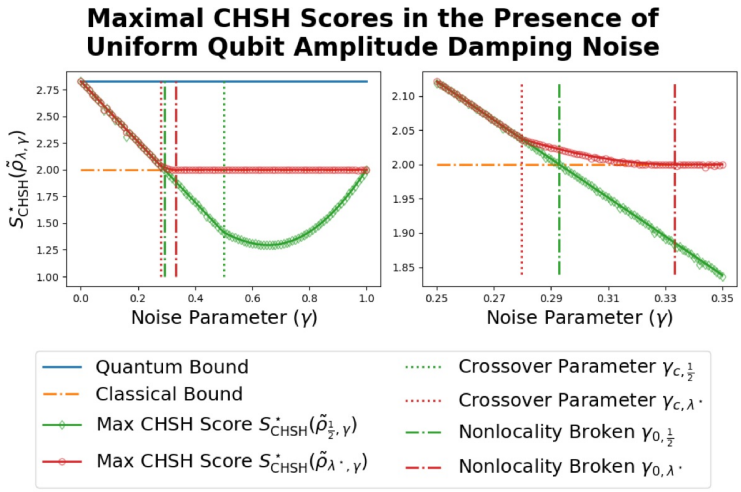



Figure 12.1: **Uniform qubit amplitude damping in the CHSH scenario.** We plot the maximal CHSH scores obtained when uniform qubit amplitude damping noise is applied to maximally entangled states (green) and nonmaximally entangled states (red). (Left) The noise parameter $\gamma \in [0, 1]$ is scanned over an interval of 0.01. (Right) The noise parameter $\gamma \in [0.25, 0.35]$ is scanned over an interval of 0.001. The solid green line plots the maximal CHSH score for maximally entangled states as given by Eq. (12.49) while the diamond markers plot numerical results obtained using VQO. The solid red line plots the maximal CHSH score for nonmaximally entangled states as given by Eq. (12.50). The green and red dotted vertical lines plot the crossover parameters at which the cases switch in Eq. (12.49) and Eq. (12.50) respectively. Likewise, the green and red dash-dotted vertical lines plot the critical parameters at which nonlocality is broken for maximally entangled states and nonmaximally entangled states respectively.

presence of uniform qubit amplitude damping noise, maximally entangled states maximize the CHSH score in the range $\gamma \in [0, \gamma_c]$, nonmaximally entangled states maximize the CHSH score in the range $\gamma \in (\gamma_c, \frac{1}{3})$, and the classical state $|00\rangle\langle 00|$ maximizes the CHSH score in the range $\gamma \in [\frac{1}{3}, 1]$. The maximal separation between the CHSH scores of maximally and nonmaximally entangled states occurs at $\gamma = (1 - \frac{1}{\sqrt{2}})$ where the nonmaximally entangled state achieves $S^{\star}_{\mathrm{CHSH}}(\tilde{\rho}_{\lambda^{\star}}) = 2\sqrt{\frac{1}{14}(3 + 8\sqrt{2})} \approx 2.0222836 > 2$. We plot this data in Fig. 12.1 in which we show that VQO successfully finds the maximal CHSH score for both maximally entangled and nonmaximally entangled state preparations.

# Chapter 13

# Nonclassicality in the $n$-Local Setting of Nonsignaling Quantum Networks

The $n$-local setting describes a class of prepare-and-measure networks where $n$ independent entanglement sources link together $m$ nonsignaling measurement devices. The $n$-local setting extends the local case by incorporating multiple sources into the nonsignaling scenario. In this setting, a network's DAG is expressed as

$$n\text{-Loc} = \text{Net}\left(\vec{N} = \left(\vec{\Lambda}, \vec{A}\right), \quad \vec{E} = \{\Lambda_i \prec \Lambda_i^{\text{Tx}}\}_{i=1}^{n}\right) \tag{13.1}$$

where $\vec{\Lambda} = (\Lambda_i)_{i=1}^{n}$ denotes a layer of source devices, $\vec{A} = (A_j)_{j=1}^{m}$ denotes a layer of nonsignaling measurement devices, and $\Lambda_i^{\text{Tx}}$ denotes the set measurement devices that are children of source $\Lambda_i$. As input each detector device $A_j$ accepts a classical value $x_j \in \mathcal{X}_j$ and outputs a classical value $a_j \in \mathcal{A}_j$. Since the network is nonsignaling, the network's behavior must satisfy the nonsignaling constraints (see Subsection 3.4.1)

$$P(a_j|x_j) = \sum_{\vec{a}_{j'\neq j} \in \mathcal{A}/\mathcal{A}_j} P(\vec{a}_{j'\neq j}, a_j|\vec{x}) \quad \forall \quad a_j \in \mathcal{A}_j, \quad \text{and} \quad \vec{x} \in \mathcal{X} = \mathcal{X}_1 \times \cdots \times \mathcal{X}_m. \tag{13.2}$$

When classical resources are used, the sources emit shared randomness. Here the $i^{th}$ source outputs the

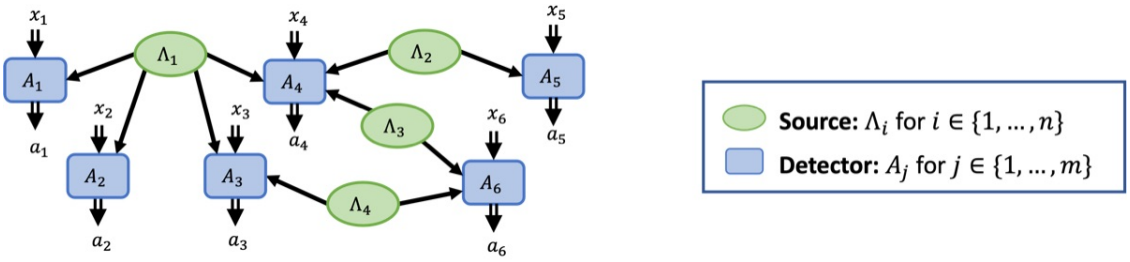



Figure 13.1: **Example Quantum Network in the $n$-Local Setting:** The network consists of $n$ entanglement sources (green ellipses) and $m$ detectors (blue rectangles).

classical value $\lambda_i$ drawn randomly from the distribution $\Omega^{\Lambda_i}$ with probability $P(\lambda_i)$ and transmits the value to its linked children devices $A_j \in \Lambda_i^{\text{Tx}}$. All sources are assumed to be independent such that $P(\vec{\lambda}) = \prod_{i=1}^n P(\lambda_i)$, where $\vec{\lambda} = (\lambda_i)_{i=1}^n$ contains the random value output from each source. A given detector $A_j$ may receive random values from multiple sources, thus $\vec{\lambda}^{A_j} \subseteq \vec{\lambda}$ denotes the set of random values received from its parent sources. The transition probabilities in the classical setting then decomposes as

$$P(\vec{a}|\vec{x}) = \sum_{\lambda_1 \in \Omega^{\Lambda_1}} \cdots \sum_{\lambda_n \in \Omega^{\Lambda_n}} \prod_{i=1}^{n} P(\lambda_i) \prod_{j=1}^{m} P(a_j|x_j, \vec{\lambda}^{A_j}). \tag{13.3}$$

**Definition 78.** ***n*-Local Set of Behaviors:** The set of behaviors $\mathcal{C}^{n\text{-Loc}} \subseteq \mathcal{P}^{\text{Net}}_{\mathcal{Y}|\mathcal{X}}$ such that Eq. (13.3) is satisfied. Note that when $n = 1$, the $n$-local set corresponds to the local set.

In the quantum setting, each source prepares a multi-qubit state $\rho^{\Lambda}$ that is distributed to the detector devices that are its children. The detectors then each apply a measurement $\Pi^{A_j}\{\Pi^{A_j}_{a_j|x_j}\}_{x_j \in \mathcal{X}_j}$ to their local quantum states $\rho^{A_j^{\text{Rx}}}$. The transition probabilities are then

$$P(\vec{a}|\vec{x}) = \text{Tr}\left[\Pi^{\vec{A}}_{\vec{a}|\vec{x}} \text{id}^{\vec{\Lambda}\to\vec{A}}(\rho^{\vec{\Lambda}})\right] = \text{Tr}\left[\left(\bigotimes_{j=1}^m \Pi^{A_j}_{a_j|x_j}\right) \text{id}^{\vec{\Lambda}\to\vec{A}} \left(\bigotimes_{i=1}^n \rho^{\Lambda_i}\right)\right] \tag{13.4}$$

where $\text{id}^{\vec{\Lambda}\to\vec{A}}$ is the identity channel that transmits the quantum states prepared at each source to its child detectors.

**Definition 79. Set of Quantum Behaviors in the *n*-Local Setting:** The set of behaviors $\mathcal{Q}^{n\text{-Loc}} \subseteq \mathcal{P}^{\text{Net}}_{\mathcal{Y}|\mathcal{X}}$ that satisfy Eq. (13.4).

Having defined the set of quantum and classical behaviors for nonsignaling networks in the $n$-local setting, we can formaly describe their nonclassicality, which is typically referred to as non-$n$-locality [124].

**Definition 80. Non-*n*-Locality:** A quantum behavior $\mathbf{P} \in \mathcal{Q}^{n\text{-Loc}}$ is said to be non-$n$-local if and only if $\mathbf{P} \notin \mathcal{C}^{n\text{-Loc}}$ where $n$-Loc denotes the DAG corresponding to a nonsignaling network in the $n$-local setting.

The set of $n$-local correlations is bound by $n$-locality inequalities, whose violation witnesses correlations as non-$n$-local [124]. These inequalities are expressed in terms of the $n$-body correlator

$$\left\langle O^{\vec{A}}_{\vec{x}} \right\rangle_{\rho^{\vec{\Lambda}}} = \text{Tr}\left[O^{\vec{A}}_{\vec{x}} \text{id}^{\vec{\Lambda}\to\vec{A}}(\rho^{\vec{\Lambda}})\right] \tag{13.5}$$

where $O^{\vec{A}}_{\vec{x}} = \bigotimes_{j=1}^m O^{A_j}_{x_j}$ is the joint network Hermitian observable. When the outputs are binary strings such that $\vec{a} \in \mathcal{A} = \mathbb{B}^m$, the observable $O^{\vec{A}}_{\vec{x}}$ has eigenvalues $\pm 1$ that correspond to the parity of the output bit string $\vec{a}$ such that

$$\left\langle O^{\vec{A}}_{\vec{x}} \right\rangle_{\rho^{\vec{\Lambda}}} = \sum_{\vec{a} \in \mathcal{A}} (-1)^{\oplus_{j=1}^m a_j} P(\vec{a}|\vec{x}). \tag{13.6}$$

Quantum violations of $n$-locality inequalities are known for many networks [267]–[273], [275], [276], [318], [321]–[325], and have been demonstrated experimentally [193]–[197], [326], [327].

Like all quantum nonclassicality, non-$n$-local correlations deteriorate in the presence of noise. The noise robustness has been investigated with respect to various noise models on the sources, communication, or measurement [102], [269]–[271], [328], [329]. However, noise robustness can more generally be characterized in terms of the maximal $n$-local violation attainable for a given mixed state. Previously, we have discussed how

the maximal violation of the CHSH inequality [13] is known for any two-qubit mixed state [319]. This result has been extended to star and chain networks in the $n$-local setting [330]–[332]. In these works, multiqubit observables are assumed to be obtained using general GHZ measurements where the qubits are projected onto the GHZ basis. It was observed in Ref. [331] that the GHZ measurement results can be coarse-grained into two observables in mutually unbiased bases, namely $\bigotimes_{i=1}^{m} X$ and $\bigotimes_{i=1}^{m} Z$ where the two observables correspond to outputting either the phase or the parity of the measured $m$-qubit GHZ state.

Recently, however, we applied variational optimization methods to obtained larger $n$-local violations using arbitrary local qubit observables [102]. Hence we were motivated to generalize the $n$-local violations for local qubit observables in Ref. [333], and compare them with the maximal violations obtained when the central parties measures their qubits in mutually unbiased bases [330]–[332]. We find that reference [318] misses an equality condition between the star network's $n$-local violation and its upper bound, the geometric mean of each source's CHSH violation. Exploiting this condition, we achieve the upper bound for all two-qubit mixed states where the external parties must measure observables in mutually unbiased bases, *e.g.*, $X$ and $Z$. These $n$-local violations are larger and more robust to noise than those obtained previously under the assumption that the central parties, instead, measure observables in mutually unbiased bases.

## *13.1 Local Qubit Measurements in the n-Local Setting*

When investigating the nonclassicality of quantum networks in the $n$-local setting, interesting theoretical results can readily be obtained when measurements are assumed to be separable across qubits, or even general qudit systems. As such, consider an $M$-qubit measurement performed by device $A_j$ where $\vec{q}^{\,A_j} = (q_1, \ldots, q_M)$. The corresponding observable is

$$O_{x_j}^{A_j} = \bigotimes_{i=1}^{M} O_{x_j}^{q_i} \tag{13.7}$$

where each local qubit observable is independent, but conditioned upon the same input $x_j$. Furthermore, if all qubit observables $O_{x_j}^{q_i}$ have eigenvalues $\pm 1$, then they take the form

$$O_{x_j}^{q_i} = \vec{a} \cdot \vec{\sigma} \quad \text{where} \quad \vec{\sigma} = (\sigma_x, \sigma_y, \sigma_z), \quad \vec{a} \in \mathbb{R}^3 \quad \text{and} \quad |\vec{a}| = 1. \tag{13.8}$$

It follows that $O_{x_j}^{A_j}$ also has eigenvalues $\pm 1$ such that $\left\langle O_{x_j}^{A_j} \right\rangle_{\rho}$ is the expected parity of the $M$-bit string containing each individual qubit's measurement result. That is, $a_j = \bigoplus_{i=1}^{M} a_{j,i}$ where $a_j, a_{j,i} \in \mathbb{B}$ and $a_{j,i}$ is the value output from the $i^{th}$ qubit at the $j^{th}$ measurement device. Alternatively, the qubit observables could be $O_{x_j}^{q_i} = \pm \mathbb{I}_2$, which corresponds to a static output of $\pm 1$ respectively. When the qubit observable is the identity, the qubit's measurement result is disregarded and replaced with a fixed classical value.

When local qubit measurements are used such that Eq. (13.7) holds for all measurement devices $A_j \in \vec{A}$, then the network correlator factors across the sources as

$$\left\langle \bigotimes_{j=1}^{m} O_{x_j}^{A_j} \right\rangle_{\rho^{\vec{\Lambda}}} = \prod_{i=1}^{n} \left\langle \bigotimes_{q \in \vec{q}^{\,\Lambda_i}} O_{x_q}^{q} \right\rangle_{\rho^{\Lambda_i}} \tag{13.9}$$

where $x_q$ is the classical input to the device that measures qubit $q$ as determined by the network's DAG.

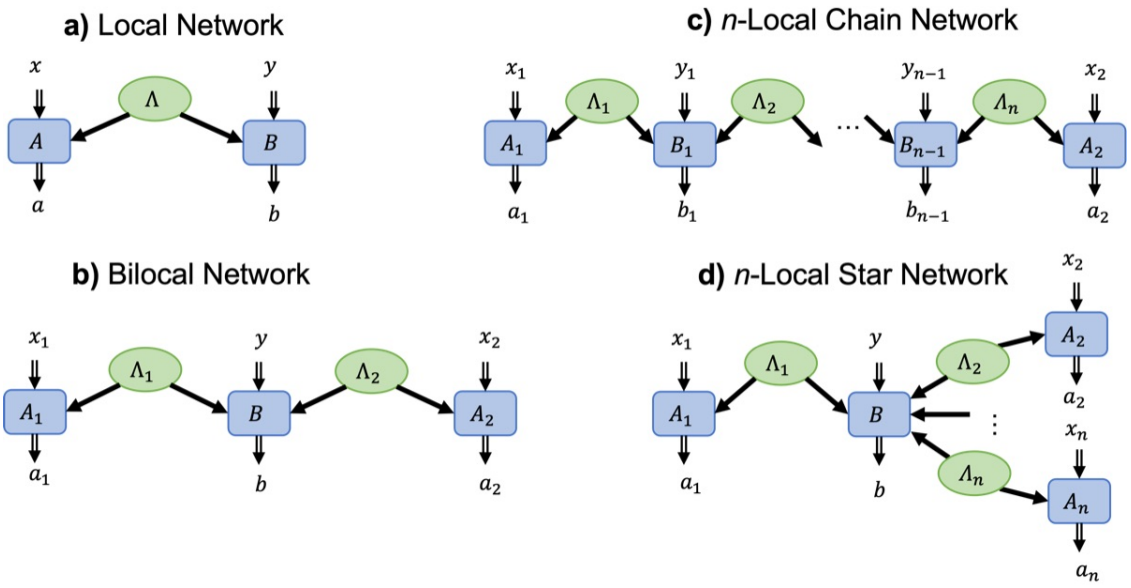



Figure 13.2: DAGs for $n$-local star and chain networks where green ellipses are sources and blue rectangles are detectors. a) The local network that corresponds to the CHSH scenario. b) The bilocal network corresponding to the simplest $n$-local network. c) The $n$-local Chain network. d) The $n$-local star network.

## *13.2 Nonclasicality in Star and Chain Networks*

In this work we consider star [270] and chain [271] networks in the $n$-local setting (see Fig. 13.1). Star networks are known to have important applications in entanglement distribution protocols [110], [111], [198], [270], while chain networks can enable long-distance quantum communication via quantum repeater relay nodes [112], [113]. These applications, however, require signaling between measurement devices and are not representative of a nonsignaling scenario. Nevertheless, investigating the nonclassicality of the nonsignaling case is important because the space-like separation between measurement devices can often be enforced experimentally [193]–[197]. Under these space-like separations, the $n$-local nonsignaling setting can be used in a wide range of device-independent certification tasks [164], [240]–[244] and security protocols [182], [256]–[260].

### *13.2.1 Nonclassicality in Bilocal Networks*

Consider the bilocal $n = 2$ network depicted in Fig. 13.2.b having two sources and three measurement devices. In the classical setting an unlimited amount of shared randomness is produced as each source, while in the quantum setting each source emits a two-qubit entangled state. This scenario was among the first $n$-local cases to be investigated due to its relative simplicity and its application in entanglement swapping [268], [269].

In the quantum case, the network consists of four qubits total where the two sources each emit two-qubits, $\vec{q}^{\Lambda_1} = (q_1, q_2)$ and $\vec{q}^{\Lambda_2} = (q_3, q_4)$. The two exterior measurement nodes measure the first and last qubits as $\vec{q}^{A_1} = (q_1)$ and $\vec{q}^{A_2} = (q_4)$, and the central node jointly measures the qubits $\vec{q}^{B} = (q_2, q_3)$ where $\Lambda_1$ and $\Lambda_2$.

The bilocal network lays the theoretical foundation for its generalizations to star and chain networks. For this reason, we refrain from any serious analysis and reserve the technicalities to our discussion of star and chain networks. Nevertheless, we provide an example bilocality inequality that can be used to witness

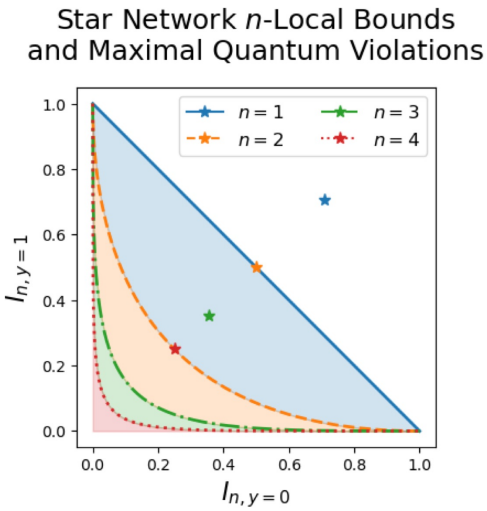



Figure 13.3: The $n$-local bounds for the star network network are plotted in terms of the $I_{n,y}$ network correlators where the shaded region denote the $n$-local regions. The maximal quantum violation is also plotted by the star marker. Note that for $n \geq 2$, the local bound is not violated by the maximally nonclassical strategy, however, for $n > 2$ the $I_{n,y}$ network correlators can also violate the bound for the $n' = n - 1$ case.

nonclassical quantum behaviors in the bilocal setting.

Consider the case where the inputs and outputs of all measurement devices are binary such that $\mathcal{X}_j = \mathcal{A}_j = \mathbb{B}$ for all $j \in [m]$. The set of bilocal behaviors $\mathcal{C}^{2\text{-Loc}}$ is bounded by the nonlinear inequality [268], [269]

$$S_{2\text{-Loc}}(\mathbf{P}) = \sqrt{|I_{2,y=0}(\mathbf{P})|} + \sqrt{|I_{2,y=1}(\mathbf{P})|} \leq 1, \tag{13.10}$$

where the quantity

$$I_{2,y}(\mathbf{P}) = \frac{1}{4} \sum_{x_1,x_2} (-1)^{y \wedge (x_1 \oplus x_2)} \langle O_{x_1}^{A_1} O_y^B O_{x_2}^{A_2} \rangle_{\rho^{\Lambda_1} \otimes \rho^{\Lambda_2}} \tag{13.11}$$

is a linear combination of tripartite correlators on the state $\rho^{\Lambda_1} \otimes \rho^{\Lambda_2}$ with the correlators being expressed as

$$\langle O_{x_1}^{A_1} O_y^B O_{x_2}^{A_2} \rangle_{\rho^{\Lambda_1} \otimes \rho^{\Lambda_2}} = \sum_{a_1,a_2,b \in \mathbb{B}} (-1)^{a_1 \oplus a_2 \oplus b} P(a_1, a_2, b | x_1, x_2, y). \tag{13.12}$$

The maximal quantum violation is $S_{2\text{-Loc}}(\mathbf{P}^\star) = \sqrt{2} > 1$ where $\mathbf{P}^\star \in \mathcal{Q}^{2\text{-Loc}}$ [269].

## *13.3 Nonclassicality in Noisy Quantum Star Networks*

**Definition 81. Star Network:** $n$-Star $= \text{Net}(\vec{N}, \vec{E})$, A network consisting of $n$, independent sources $\{\Lambda_i\}_{i=1}^n$ and $n$ external nonsignaling measurement devices $\{A_j\}_{j=1}^n$ that each share a source with a central measurement device $B$ (see Fig. 13.2.d). The network nodes and edges are expressed

$$\vec{N} = (\{\Lambda_i\}_{i=1}^n, \{A_1, \ldots, A_n, B\}), \quad \vec{E} = \{\Lambda_i \prec (A_i, B)\}_{i=1}^n. \tag{13.13}$$

### 13.3.1 Nonclassicality Witnesses in the Star Network

A family of nonclassicality witnesses for star networks can be constructed from a collection of $n$ bipartite full-correlation inequalities applied to each independent source [318] (see Def. 75). Here the full-correlation inequality is described by the tuple $(\gamma^i, \mathbf{G}^i)$

$$S_{n\text{-Star}\{\mathbf{G}^i\}_i}(\mathbf{P}) = \sum_{y\in\mathcal{Y}} |I_{n,y}(\mathbf{P})|^{\frac{1}{n}} \leq \left(\prod_{i=1}^{n} \gamma^i\right)^{\frac{1}{n}} = \beta \tag{13.14}$$

where

$$I_{n,y} = \sum_{x_1\in\mathcal{X}_1} \cdots \sum_{x_n\in\mathcal{X}_n} G^1_{x_1,y} \ldots G^n_{x_n,y} \left\langle O^{A_1}_{x_1} \ldots O^{A_n}_{x_n} O^B_y \right\rangle_{\rho^{\vec{\Lambda}}} \tag{13.15}$$

and $G^i_{x,y} \in \mathbb{R}$ is a matrix element of $\mathbf{G}^i$. Furthermore, the correlator is

$$\left\langle O^{A_1}_{x_1} \ldots O^{A_n}_{x_n} O^B_y \right\rangle_{\rho^{\vec{\Lambda}}} = \sum_{a_1\in\mathcal{A}_i} \cdots \sum_{a_n\in\mathcal{A}_n} \sum_{b\in\mathcal{B}} (-1)^{\oplus_{i=1}^n a_i} (-1)^{f_y(\vec{b})} P(\vec{a},\vec{b}|\vec{x},y) \tag{13.16}$$

where $\vec{a},\vec{x},\vec{b} \in \mathbb{B}^n$ are $n$-bit strings and the central node outputs $b = f_y(\vec{b})$ for some function $f_y : \mathbb{B}^n \to \mathbb{B}$ mapping the central node's $n$-bit measurement result $\vec{b}$ into a binary value $b$. Each set of full-correlator inequalities $\{(\gamma^i, \mathbf{G}^i)\}_{i=1}^n$ and binary functions $\{f_y\}_{y\in\mathcal{Y}}$ correspond to a distinct classical bound on the star network in the $n$-local setting.

An important case occurs when local qubit (or qudit) observables are considered on the central node such that $O^B_y = \bigotimes_{j=n+1}^{2n} O^{q_j}_y$ and $b = f_y(\vec{b}) = \bigoplus_{i=1}^n b_i$ for all $y$. In this case, the correlator factors as in Eq. (13.9) such that

$$I_{n,y} = \sum_{x_1\in\mathcal{X}_1} \cdots \sum_{x_n\in\mathcal{X}_n} \prod_{i=1}^{n} G^i_{x_i,y} \left\langle O^{A_i}_{x_i} O^{B_i}_y \right\rangle_{\rho^{\Lambda_i}} \tag{13.17}$$

where $O^{B_i}_y$ is the qubit observable applied on the central node's side of source $\Lambda_i$. Let

$$O^{A_iB_i}_y = \sum_{x_i\in\mathcal{X}_i} G^i_{x_i,y} O^{A_i}_{x_i} \otimes O^{B_i}_y, \tag{13.18}$$

then Eq. (13.17) becomes

$$I_{n,y} = \prod_{i=1}^{n} \left\langle O^{A_iB_i}_y \right\rangle_{\rho^{\Lambda_i}}. \tag{13.19}$$

It follows that the star network inequality for local qubit observables is

$$S_{n\text{-Star}\{\mathbf{G}^i\}_i}(\mathbf{P}) = \sum_{y\in\mathcal{Y}} \prod_{i=1}^{n} \left| \left\langle O^{A_iB_i}_y \right\rangle_{\rho^{\Lambda_i}} \right|^{\frac{1}{n}} \tag{13.20}$$

where the absolute value results from the fact that the observables $O^{B_i}_y$ can always be selected to enforce that $\left\langle O^{A_iB_i}_y \right\rangle_{\rho_i} \geq 0$. Then the following inequality must hold for any state preparation $\rho^{\vec{\Lambda}}$ and measurement

$\langle O_y^{A_iB_i}\rangle_{\rho^{\Lambda_i}}$ [318]

$$\sum_{y\in\mathcal{Y}}\prod_{i=1}^{n}\left|\left\langle O_y^{A_iB_i}\right\rangle_{\rho^{\Lambda_i}}\right|^{\frac{1}{n}}\leq\prod_{i=1}^{n}\left(\sum_{y\in\mathcal{Y}}\left|\left\langle O_y^{A_iB_i}\right\rangle_{\rho^{\Lambda_i}}\right|\right)^{\frac{1}{n}} \tag{13.21}$$

$$S_{n\text{-Star}\{\mathbf{G}^i\}_i}(\mathbf{C}^{\text{Net}})\leq\prod_{i=1}^{n}S^i_{\text{LNS}(2)}\left(\mathbf{C}^{AB_i}\right)^{\frac{1}{n}} \tag{13.22}$$

where, for local qubit observables, the star inequality score is upper bounded by the geometric mean of bipartite full-correlator inequalities. The conditions for equality are given in the following lemma.

**Lemma 8.** [333] Equality is obtained in Eq. (13.21) if and only if the matrix with elements $M_{y,i}=|\left\langle O_y^{A_iB_i}\right\rangle_{\rho^{\Lambda_i}}|$ has Rank $(M)=1$, or for some $i\in[n]$, $\left\langle O_y^{A_iB_i}\right\rangle_{\rho^{\Lambda_i}}=0$ for all $y\in\mathcal{Y}$.

*Proof.* Let $M_{y,i}=a_yb_i$, then the RHS and LHS of Eq. (13.21) are equal because

$$\sum_{y\in\mathcal{Y}}\left(\prod_{i=1}^{n}M_{y,i}\right)^{\frac{1}{n}}=\sum_{y\in\mathcal{Y}}\left(\prod_{i=1}^{n}a_yb_i\right)^{\frac{1}{n}}=\left(\sum_{y\in\mathcal{Y}}a_y\right)\left(\prod_{i=1}^{n}b_i\right)^{\frac{1}{n}}=\prod_{i=1}^{n}\left(\sum_{y\in\mathcal{Y}}a_yb_i\right)^{\frac{1}{n}}=\prod_{i=1}^{n}\left(\sum_{y\in\mathcal{Y}}M_{y,i}\right)^{\frac{1}{n}}. \tag{13.23}$$

Since equality holds when $M=\vec{a}\cdot\vec{b}^T$ for positive vectors $\vec{a}$ and $\vec{b}$ any matrix that has Rank $(M)=1$ must yield equality. As an edge case, since both RHS and LHS of Eq. (13.21) are positive, equality is also obtained when, for some $i\in[n]$, $M_{y,i}=0$ for all $y\in\mathcal{Y}$ because the upper bound is zero $\sum_{y\in\mathcal{Y}}M_{y,i}=0$. Finally, if $M_{y,i}\neq a_yb_i$ for all $y\in\mathcal{Y}$ and $i\in[n]$, then the inequality in Eq. (13.21) is strict because the sum and product cannot be interchanged. □

Note that reference [318] presents a similar result to Lemma 8, however, the authors incorrectly state that equality is obtained *if and only if* $M_{z,i}=M_{z,i'}$ for all $i,i'\in[n]$. While the example matrix $M$ is rank-one, the authors' equality condition neglects a large number of rank-one matrices. Indeed, these matrices correspond to correlations that lead to greater violations of star $n$-locality than previously known.

### *13.3.2 Maximal n-Locality Violations in Noisy Star Networks*

We now consider the maximal violations of the star $n$-locality inequalities when there is noise present in a quantum star network. To this end, we first derive general bounds on the maximal $n$-locality violations given arbitrary mixed state preparations and local observables that are separable across qudit subsystems. Then, we compare our results with the maximal star $n$-locality violations discussed in Ref. [331].

**Theorem 5.** For any ensemble of two-qudit mixed states $\rho_{[n]}=\bigotimes_{i=1}^{n}\rho_i$ where $\rho_i\in D(\mathcal{H}_d^{A_i}\otimes\mathcal{H}_d^{B_i})$, the maximal star $n$-locality score obtained using local qudit observables satisfies

$$S^{\star}_{n\text{-Star}\{\mathbf{G}^i\}_i}(\rho_{[n]})\leq\prod_{i=1}^{n}G^{i\star}(\rho_i)^{\frac{1}{n}} \tag{13.24}$$

where equality occurs if and only if the conditions in Lemma **??** are satisfied by the matrix $M\in\mathbb{R}^{|\mathcal{Z}|\times n}$ where $M_{z,i}=\left\langle O_z^{A_iB_i}\right\rangle^{\star}_{\rho_i}\geq 0$ are the correlators that maximize $G^{i\star}(\rho_i)$.

*Proof.* For any set of correlators $\{\langle O_z^{A_iB_i}\rangle_{\rho_i}\}_{i\in[n],z\in\mathcal{Z}}$ where $\langle O_z^{A_iB_i}\rangle_{\rho_i}$ is defined in Eq. (12.7), the inequality in Eq. (**??**) must hold. Thus, if the set of correlators $\{\langle O_z^{A_iB_i}\rangle^\star_{\rho_i}\}_{i\in[n],z\in\mathcal{Z}}$ maximize the star $n$-locality score, then

$$S^\star_{n\text{-Star}\{\mathbf{G}^i\}_i}(\rho_{[n]}) \leq \prod_{i=1}^n \left(\sum_{z\in\mathcal{Z}} \langle O_z^{A_iB_i}\rangle^\star_{\rho_i}\right)^{\frac{1}{n}}. \tag{13.25}$$

where we may assume that $\langle O_z^{A_iB_i}\rangle^\star_{\rho_i} \geq 0$ because the qubit observable $B_z^i$ may be selected to ensure nonnegativity. However, this set of observables is not guaranteed to maximize each respective locality inequality, implying that $S^\star_{n\text{-Star}\{\mathbf{G}^i\}_i}(\rho_{[n]}) \leq \prod_{i=1}^n G^{i\star}(\rho_i)^{\frac{1}{n}}$. Now let $M$ be a matrix with elements $M_{z,i} = \langle O_z^{A_iB_i}\rangle^\star_{\rho_i} \geq 0$. Then, using Lemma 8, $S^\star_{n\text{-Star}\{\mathbf{G}^i\}_i}(\rho_{[n]}) = \prod_{i=1}^n G^{i\star}(\rho_i)^{\frac{1}{n}}$ if and only if either $\text{Rank}\,(M) = 1$ or, for some $i \in [n]$, $\langle O_z^{A_iB_i}\rangle^\star_{\rho_i} = 0$ for all $z \in \mathcal{Z}$. □

Remarkably, Theorem 5 shows that any full-correlation star inequality as defined in Eq. (13.14) can achieve its maximal score $\beta = \prod_{i=1}^n (\gamma^i)^{1/n}$ using local qubit observables and a coarse-graining function $f_y(\vec{b}) = \bigoplus_{i=1}^n b_i$.

Although Theorem 5 applies to a broad family of star $n$-locality inequalities, it fails to dictate whether or not there exist local qudit observables that achieve the upper bound. However, we make an interesting observation when all parties are given binary inputs where $\mathcal{X}_i = \mathcal{Z} = \mathbb{B}$ for all $i \in [n]$. In this setting, the CHSH inequality in Eq. (12.9) is considered for each source such that $\prod_{i=1}^n G^i_{x_i,z} = \prod_{i=1}^n \frac{1}{2}(-1)^{x_i\wedge z}$ and $\beta = 1$.

**Theorem 6.** For any ensemble of two-qubit mixed states $\rho_{[n]} = \bigotimes_{i=1}^n$ where $\rho_i \in D(\mathcal{H}_2^{A_i} \otimes \mathcal{H}_2^{B_i})$, the maximal CHSH $n$-locality star score obtained using local qubit observables is

$$S^\star_{n\text{-Star}\{\mathbf{G}^{\text{CHSH}_i}\}_i}(\rho_{[n]}) = \prod_{i=1}^n S^\star_{\text{CHSH}}(\rho_i)^{\frac{1}{n}} \tag{13.26}$$

where $S^\star_{\text{CHSH}}(\rho_i) = \sqrt{\tau_{i,0}^2 + \tau_{i,1}^2}$ as in Eq. (12.13).

*Proof.* By Theorem 5, $S^\star_{n\text{-Star}\{\mathbf{G}^{\text{CHSH}_i}\}_i}(\rho_{[n]}) \leq \prod_{i=1}^n S^\star_{\text{CHSH}}(\rho_i)$. To prove equality, consider the optimal CHSH observables $A^{i\star}_{x_i}$ and $B^{i\star}_z$ from Eq. (12.14). Since the expectations in Eq. (**??**) satisfy $\left\langle O_0^{\text{CHSH}_i}\right\rangle^\star_{\rho_i} = \left\langle O_1^{\text{CHSH}_i}\right\rangle^\star_{\rho_i}$ for all $i \in [n]$, the matrix $M$ with elements $M_{z,i} = \langle O_z^{\text{CHSH}_i}\rangle^\star_{\rho_i} \geq 0$ has $\text{Rank}\,(M) = 1$. Thus, Lemma **??** is satisfied, implying that $S^\star_{n\text{-Star}\{\mathbf{G}^{\text{CHSH}_i}\}_i}(\rho_{[n]}) = \prod_{i=1}^n S^\star_{\text{CHSH}}(\rho_i)^{\frac{1}{n}}$. □

Theorem 6 contrasts with previous results, which assume that the multi-qubit observables are in the mutually unbiased bases

$$\widehat{C}_z^{[n]} = (1-z)\bigotimes_{i=1}^n Z^{B_i} + z\bigotimes_{i=1}^n X^{B_i}. \tag{13.27}$$

This choice of observables has physical significance because they can be obtained by coarse-graining the outcomes of the GHZ measurement, which is necessary for generalized entanglement swapping the star network [111]. The observables $\widehat{C}_z^{[n]}$ lead to the maximal $n$-local star score [330]–[332]

$$\widehat{S}^\star_{n\text{-Star}\{\mathbf{G}^i\}_i}(\rho^{\text{Net}}) = \sqrt{\prod_{i=1}^n \tau_{i,0}^{2/n} + \prod_{i=1}^n \tau_{i,1}^{2/n}} \tag{13.28}$$

where each external party measures the qubit observables

$$\hat{A}^i_{x_i} = \frac{\prod_{i=1}^n \tau_{i,0}^{1/n} Z + (-1)^{x_i} \prod_{i=1}^n \tau_{i,1}^{1/n} X}{\sqrt{\prod_{i=1}^n \tau_{i,0}^{2/n} + \prod_{i=1}^n \tau_{i,1}^{2/n}}}. \tag{13.29}$$

In general, $\widehat{S}^\star_{n\text{-Star}\{\mathbf{G}^i\}_i}(\rho^{\text{Net}}) \leq S^\star_{n\text{-Star}\{\mathbf{G}^{\text{CHSH}_i}\}_i}(\rho^{\text{Net}})$ where taking the square of both sides and substituting Eq. (13.28) and Eq. (13.26) yields the inequality in Eq. (13.21)

$$\prod_{i=1}^n \tau_{i,0}^{2/n} + \prod_{i=1}^n \tau_{i,1}^{2/n} \leq \prod_{i=1}^n (\tau_{i,0}^2 + \tau_{i,1}^2)^{\frac{1}{n}}. \tag{13.30}$$

Applying Lemma 8, we derive the condition for equality between these two contrasting $n$-local star scores.

**Corollary 1.** $S^\star_{n\text{-Star}\{\mathbf{G}^{\text{CHSH}_i}\}_i}(\rho^{\text{Net}}) = \widehat{S}^\star_{n\text{-Star}\{\mathbf{G}^i\}_i}(\rho^{\text{Net}})$ if and only if either $\frac{\tau_{i,0}}{\tau_{i,1}} = \frac{\tau_{j,0}}{\tau_{j,1}}$ for all $i, j \in [n]$, or $\tau_{i,0} = 0$ for some $i \in [n]$.

*Proof.* Let $M_{j,i} = \tau^2_{i,j}$, then $\text{Rank}\,(M) = 1$ when the two rows of $M$ are scalar multiples of each other. Hence equality requires the ratio $\frac{\tau_{i,0}}{\tau_{i,1}}$ to be constant for all $i \in [n]$. Otherwise, equality is obtained when $\tau_{i,0} = 0$ because $\tau_{i,0} \geq \tau_{i,1} = 0$. □

In theoretical works, the conditions in Corollary 1 are often assumed to hold. For instance, equality holds when all sources emit the same mixed state, or when white noise is modeled on sources. However, these examples of equality are exceptions that would rarely occur in general ensembles of two-qubit mixed states.

While $\widehat{S}^\star_{n\text{-Star}\{\mathbf{G}^i\}_i}(\rho^{\text{Net}})$ is not globally maximal, it is maximal under the assumption that the central party measures local qubit observables in Eq. (13.27), which are in mutually unbiased bases. On the contrary, $S^\star_{n\text{-Star}\{\mathbf{G}^{\text{CHSH}_i}\}_i}(\rho^{\text{Net}})$ is globally maximal for all mixed state ensemble where $S^\star_{n\text{-Star}\{\mathbf{G}^{\text{CHSH}_i}\}_i}(\rho^{\text{Net}})$ is achieved when the *external* parties each measure their qubit observables in mutually unbiased bases, *e.g.*, $A^i_{x_i} \in \{X, Z\}$. As noted earlier, a CHSH violation $S^\star_{\text{CHSH}}(\rho) > 1$ is alone insufficient to determine which qubit was measured in mutually unbiased bases. However, this is not the case in the star network because the maximal violation $S^\star_{n\text{-Star}\{\mathbf{G}^{\text{CHSH}_i}\}_i}(\rho^{\text{Net}}) > \widehat{S}^\star_{n\text{-Star}\{\mathbf{G}^i\}_i}(\rho^{\text{Net}})$ for a known state $\rho^{\text{Net}}$ requires the external parties to measure observables in mutually unbiased bases. This fact could be used to self-test whether the external parties are measuring in mutually unbiased bases.

In reference [102], an extreme example is noted where $S^\star_{n\text{-Star}\{\mathbf{G}^{\text{CHSH}_i}\}_i}(\rho^{\text{Net}}) > 1 \geq \widehat{S}^\star_{n\text{-Star}\{\mathbf{G}^i\}_i}(\rho^{\text{Net}})$ in which $k \in [1, n)$ sources each emit the classical state $\sigma_i = |00\rangle\langle 00|$ while the remaining sources each emit maximally entangled states. In this case,

$$S^\star_{n\text{-Star}\{\mathbf{G}^{\text{CHSH}_i}\}_i}(\rho^{\text{Net}}) = 2^{\frac{n-k}{2n}} > 1 = \widehat{S}^\star_{n\text{-Star}\{\mathbf{G}^i\}_i}(\rho^{\text{Net}}), \tag{13.31}$$

thus showing an example where significant $n$-local violations occurs where no violation is predicted by previous results [331]. We now use Theorem 6 to generalize this example.

**Corollary 2.** For two-qubit states $\sigma_i, \rho_i \in D(\mathcal{H}_2^{A_i} \otimes \mathcal{H}_2^{B_i})$, consider the ensemble of $k$ classical states $\sigma_{[k]} = \bigotimes_{i=1}^k \sigma_i$ where $T_{\sigma_i} = \text{diag}\,(0, 0, 1)$, and the ensemble of general mixed states $\rho_{[k+1,n]} = \bigotimes_{i=k+1}^n \rho_i$. Then,

$$\widehat{S}^\star_{n\text{-Star}\{\mathbf{G}^i\}_i}(\sigma_{[k]} \otimes \rho_{[k+1,n]}) \leq 1 \leq S^\star_{n\text{-Star}\{\mathbf{G}^{\text{CHSH}_i}\}_i}(\sigma_{[k]} \otimes \rho_{[k+1,n]}) = S^\star_{(n-k)\text{-CHSH}}(\rho_{[k+1,n]})^{\frac{n-k}{n}}. \tag{13.32}$$

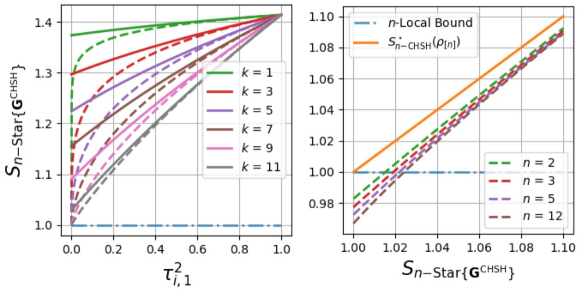



Figure 13.4: **Comparing the Star $n$-Locality Violations: of $S^{\star}_{n\text{-Star}\{\mathbf{G}^{\text{CHSH}_i}\}_i}(\rho^{\text{Net}})$ (solid) and $\widehat{S}^{\star}_{n\text{-Star}\{\mathbf{G}^i\}_i}(\rho^{\text{Net}})$ (dashed):** (Left) We consider $k < n = 12$ noisy sources that have $\tau_{i,0} = 1$ and $\tau_{i,1} \in [0, 1]$. The remaining $(n-k)$ sources have $\tau_{i,0} = \tau_{i,1} = 1$. (Right) For all $i \in [n]$, we set $\frac{1}{2} S_{\text{CHSH}}\star(\rho_i) = \beta^{\star} \in [1.0, 1.1]$ while $\tau^2_{i,1} = (\beta^{\star})^2 - \tau^2_{i,0}$ where $\tau^2_{i,0}$ are evenly spaced in $\left[\frac{1}{2}(\tau^2_{i,0} + \tau^2_{i,1}), \min\{1, \tau^2_{i,0} + \tau^2_{i,1}\}\right]$.

Corollary 2 leads to bounds whose violation witnesses *full quantum network nonlocality*, in which all sources are verified to be nonclassical [170]. Namely, if $k = 1$, then the maximal $n$-local violation is bounded as $S^{\star}_{n\text{-Star}\{\mathbf{G}^{\text{CHSH}_i}\}_i}(\rho^{\text{Net}}) \leq 2^{\frac{n-1}{2n}}$. Thus, if $S_{n\text{-Star}\{\mathbf{G}^{\text{CHSH}_i}\}_i}(\mathbf{C}^{\text{Net}}) > 2^{\frac{n-1}{2n}}$, then all sources are nonclassical. Similarly, if the central party measures the observable $\widehat{C}^{[n]}_z$ such that $\widehat{S}^{\star}_{n\text{-Star}\{\mathbf{G}^i\}_i}(\rho^{\text{Net}})$ is maximal, then the violation $\widehat{S}^{\star}_{n\text{-Star}\{\mathbf{G}^i\}_i}(\rho^{\text{Net}}) > 1$, witnesses all sources to be nonclassical. Thus for either choice of observables, a sufficiently large $n$-local violation asserts that no classical sources are present.

Overall, the $n$-locality violations of $S^{\star}_{n\text{-Star}\{\mathbf{G}^{\text{CHSH}_i}\}_i}(\rho^{\text{Net}})$ are more robust to noise than $\widehat{S}^{\star}_{n\text{-Star}\{\mathbf{G}^i\}_i}(\rho^{\text{Net}})$. In Fig. 13.4, we illustrate the separation between $S^{\star}_{n\text{-Star}\{\mathbf{G}^{\text{CHSH}_i}\}_i}(\rho^{\text{Net}})$ and $\widehat{S}^{\star}_{n\text{-Star}\{\mathbf{G}^i\}_i}(\rho^{\text{Net}})$. In the left plot, we consider $k$ sources to be affected by noise that dampens $\tau_{i,1}$ but preserves $\tau_{i,0}$, while the remaining sources are noiseless. As $\tau_{i,1}$ becomes small, a large separation exists. In the right plot, we consider a case where Corollary 1 does not hold. That is, for all $i \in [n]$, $S^{\star}_{\text{CHSH}}(\rho_i)$ is constant, but the pair $\tau_{i,0}$ and $\tau_{i,1}$ are unique. We thus construct examples where $S^{\star}_{n\text{-Star}\{\mathbf{G}^{\text{CHSH}_i}\}_i}(\rho^{\text{Net}}) > 1 > \widehat{S}^{\star}_{n\text{-Star}\{\mathbf{G}^i\}_i}(\rho^{\text{Net}})$.

## *13.4 Nonclassicality in Noisy n-Local Chain Networks*

**Definition 82. Chain Network:** $n$-Chain $= \text{Net}(\vec{N}, \vec{E})$, A network consisting of $n$, independent sources $\{\Lambda_i\}_{i=1}^n$, two external nonsignaling measurement devices $A_1$ and $B_n$, and $n-1$ central nonsignaling measurement devices devices $C_{i,i+1}$ where the subscript indexes the linked sources (see Fig. 13.2.c). The layers of nodes are then

$$\vec{N} = (\{\Lambda_i\}_{i=1}^n, \{A_1, C_{1,2}, \ldots, C_{n-1,n}, B_n\}) \tag{13.33}$$

where the collection of edges are

$$\vec{E} = \{\Lambda_1 \prec (A_1, C_{1,2}), \ldots, \Lambda_i \prec (C_{i-1,i}, C_{i,i+1}), \ldots, \Lambda_n \prec (C_{n-1,n}, B_n)\}\,. \tag{13.34}$$

### 13.4.1 Nonclassicality Witnesses in the Chain Network

Consider the $n$-local chain network in Def. 82. When each party outputs a binary value $a, b, c \in \{\pm 1\}$, we introduce a family of chain network $n$-locality inequalities

$$S_{n\text{-Chain}\{\mathbf{G}^1,\mathbf{G}^n\}}(\mathbf{C}^{\text{Net}}) = \sum_{z\in\mathcal{Z}} \left|J_{n,z}\left(\mathbf{C}^{\text{Net}}\right)\right|^{\frac{1}{2}} \leq \beta, \tag{13.35}$$

$$J_{n,z}\left(\mathbf{C}^{\text{Net}}\right) = \sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}} G^1_{x,z} G^n_{y,z} \left\langle O^{\text{Chain}}_{x,y,z}\right\rangle, \tag{13.36}$$

where $O^{\text{Chain}}_{x,y,z} = A^1_x \otimes (\bigotimes_{i=1}^{n-1} C^{i,i+1}_z) \otimes B^n_y$ with $C^{i,i+1}_z = B^i_z \otimes A^{i+1}_z$ and the coefficients for the bipartite full-correlation locality inequalities are $G^1_{x,z}, G^n_{y,z} \in \mathbb{R}$. The $n$-local bound of $\beta = \sqrt{\gamma^1\gamma^n}$ is derived by assuming that each party's correlator satisfies $A^1_x, B^n_y, C^{i,i+1}_z \in \{\pm 1\}$ and is separable from the other parties [271], [318], [334]. When considering local qudit observables, we use Eq. (13.9) to rewrite Eq. (13.36) as

$$J_{n,z}(\mathbf{C}^{\text{Net}}) = \prod_{i\in\{1,n\}} \langle O^{A_iB_i}_z\rangle_{\rho_i} \prod_{i=2}^{n-1} \left\langle A^i_z \otimes B^i_z\right\rangle_{\rho_i} \tag{13.37}$$

where $\left\langle O^{A_1B_1}_z\right\rangle_{\rho_1} = \sum_{x\in\mathcal{X}} G^1_{x,z} \left\langle A^1_x \otimes B^1_z\right\rangle_{\rho_1}$ (similarly for $n$) is the full-correlation observable from Eq. (12.7). Then, inserting Eq. (13.37) into Eq. (13.35) and noting that $\prod_{i\in\{1,n\}}\langle O^{A_iB_i}_z\rangle_{\rho_i} = I_{2,z}(\mathbf{C}^{1,n})$, we find

$$S_{n\text{-Chain}\{\mathbf{G}^1,\mathbf{G}^n\}}(\mathbf{C}^{\text{Net}}) = \sum_{z\in\mathcal{Z}} \left| I_{2,z}(\mathbf{C}^{1,n}) \prod_{i=2}^{n-1} \left\langle A^i_z \otimes B^i_z\right\rangle_{\rho_i} \right|^{\frac{1}{2}} \tag{13.38}$$

where $\mathbf{C}^{1,n}$ denotes the correlations of sources 1 and $n$.

### 13.4.2 Maximal n-Locality Violations in Noisy Chain

We now consider the maximal violations of the chain $n$-locality inequalities when there is noise present in a quantum chain network. To this end, we first derive general bounds on the maximal $n$-locality violations given arbitrary mixed state preparations and local observables that are separable across qudit subsystems. Then, we compare our results with the maximal chain $n$-locality violations discussed in Ref. [332].

First, we derive an upper bound on the maximal chain $n$-locality score in Eq. (13.38) for any quantum correlations $\mathbf{C}^{\text{Net}} \in \mathcal{Q}^{\text{Net}}$. First, since $\langle A^i_z \otimes B^i_z\rangle_{\rho_i} \leq \tau_{i,0}$ for all $z \in \mathcal{Z}$, it follows that

$$S_{n\text{-Chain}\{\mathbf{G}^1,\mathbf{G}^n\}}(\mathbf{C}^{\text{Net}}) \leq \sum_{z\in\mathcal{Z}} \left|I_{2,z}(\mathbf{C}^{1,n})\right|^{\frac{1}{2}} \prod_{i=2}^{n-1} \sqrt{\tau_{i,0}} \leq S^{\star}_{2\text{-Star}\{\mathbf{G}^1,\mathbf{G}^2\}}(\rho_1\otimes\rho_n) \prod_{i=2}^{n-1}\sqrt{\tau_{i,0}} \tag{13.39}$$

where in the second line $\sum_{z\in\mathbb{B}} |I_{2,z}(\mathbf{C}^{1,n})|^{\frac{1}{2}} = S_{2\text{-Star}\{\mathbf{G}^1,\mathbf{G}^2\}}(\mathbf{C}^{1,n}) \leq S^{\star}_{2\text{-Star}\{\mathbf{G}^1,\mathbf{G}^2\}}(\rho_1\otimes\rho_n)$. When the number of inputs for a party exceeds 3, sources $i \in \{1, n\}$ may need to prepare bipartite entangled states of dimension larger than two to obtain the maximal quantum violation [334]. However, as shown in the following theorem, the maximal violation of the chain $n$-locality inequalities can be obtained when the central sources $i \in [2, n-1]$ prepare classical two-qubit states.

**Theorem 7.** For any ensemble of two-qudit mixed states $\rho^{\text{Net}} = \bigotimes_{i=1}^{n} \rho_i$ where $\rho_i \in D(\mathcal{H}_d^{A_i} \otimes \mathcal{H}_d^{B_i})$, the

maximal $n$-local chain score obtained by measuring local dichotomic observables is

$$S^{\star}_{n\text{-Chain}\{\mathbf{G}^1,\mathbf{G}^n\}}(\rho^{\text{Net}}) = S^{\star}_{2\text{-Star}\{\mathbf{G}^1,\mathbf{G}^2\}}(\rho_1 \otimes \rho_n) \prod_{i=2}^{n-1} \sqrt{\tau_{i,0}} \tag{13.40}$$

where $S^{\star}_{2\text{-Star}\{\mathbf{G}^1,\mathbf{G}^2\}}(\rho_1 \otimes \rho_n)$ is the maximal bilocal star score using local qudit observables.

*Proof.* For all sources $i \in [2, n-1]$, let $\rho_i \in D(\mathcal{H}_2^{A_i} \otimes \mathcal{H}_2^{B_i})$ be measured by the observable $A_z^i \otimes B_z^i = Z \otimes Z$ such that $\langle A_z^i \otimes B_z^i \rangle_{\rho_i} = \tau_{i,0}$ for all $z \in \mathcal{Z}$. Then, for the two sources $i \in \{1, n\}$, consider the optimal full-correlation observables $\langle O_z^{A_i B_i} \rangle^{\star}_{\rho_i}$ that maximize the bilocal star score to obtain $S^{\star}_{n\text{-Chain}\{\mathbf{G}^1,\mathbf{G}^n\}}(\rho^{\text{Net}}) = S^{\star}_{2\text{-Star}\{\mathbf{G}^1,\mathbf{G}^2\}}(\rho_1 \otimes \rho_n)$. Inserting these observables into Eq. (13.38) yields the upper bound on the $n$-local chain score in Eq. (13.39). □

Theorem 7 shows that to obtain the maximal $n$-local chain score, it is sufficient to measure all central sources $i \in [2, n-1]$ using the fixed local qubit observable $Z \otimes Z$. Only the measurements that share a source with an external party depend on $z$. Thus, the maximal $n$-local chain score can be achieved even when all sources $i \in [2, n-1]$ emit a classical state such as $\gamma$ where $T_\gamma = \text{diag}\,(0, 0, 1)$. Similarly, consider the single-outcome observables $A_z^i \otimes B_z^i = \mathbb{I}_d \otimes \mathbb{I}_d$ measured on sources $i \in [2, n-1]$, then for any mixed state $\rho_i$, $\langle A_z^i \otimes B_z^i \rangle_{\rho_i} = 1$ for all $z \in \mathcal{Z}$ and

$$S^{\star}_{n\text{-Chain}\{\mathbf{G}^1,\mathbf{G}^n\}}(\rho^{\text{Net}}) = S^{\star}_{2\text{-Star}\{\mathbf{G}^1,\mathbf{G}^2\}}(\rho_1 \otimes \rho_n). \tag{13.41}$$

That is, the maximal violation of chain $n$-locality is achieved by disregarding all sources $i \in [2, n-1]$ and replacing their measurements with a constant output.

Consider now the special case where all inputs are binary $\mathcal{X} = \mathcal{Y} = \mathcal{Z} = \mathbb{B}$. The $n$-local chain inequality then becomes a generalization of the CHSH inequality in Eq. (12.9). Since all inputs are binary, the maximal $n$-local chain score can be obtained using maximally entangled two-qubit states [334]. In this setting, $G^1_{x,z} G^n_{y,z} = \frac{1}{4}(-1)^{x \wedge z}(-1)^{y \wedge z}$ and $\beta = 1$. Inserting these values into Eq. (13.36) and considering local qubit observables we find

$$J_{n,z}(\mathbf{C}^{\text{Net}}) = \prod_{i \in \{1,n\}} \langle O_z^{\text{CHSH}_i} \rangle_{\rho_i} \prod_{i=2}^{n-1} \langle A_z^i \otimes B_z^i \rangle_{\rho_i} \tag{13.42}$$

where $O_z^{\text{CHSH}_i} = \frac{1}{2}(A_0^i + (-1)^z A_1^i) \otimes B_z^i$ as in Eq. (12.9). It follows that

$$S_{n\text{-Chain}\{\mathbf{G}^{\text{CHSH}}\}}(\mathbf{C}^{\text{Net}}) \equiv \sum_{z \in \mathbb{B}} \prod_{i \in \{1,n\}} \left| \langle O_z^{\text{CHSH}_i} \rangle_{\rho_i} \right|^{\frac{1}{2}} \prod_{i=2}^{n-1} \left| \langle A_z^i \otimes B_z^i \rangle_{\rho_i} \right|^{\frac{1}{2}}. \tag{13.43}$$

**Theorem 8.** For any ensemble of two-qubit mixed states $\rho^{\text{Net}} = \bigotimes_{i=1}^n \rho_i$ where $\rho_i \in D(\mathcal{H}_2^{A_i} \otimes \mathcal{H}_2^{B_i})$, the maximal $n$-local chain score obtained by local qubit observables is

$$S^{\star}_{n\text{-Chain}\{\mathbf{G}^{\text{CHSH}}\}}(\rho^{\text{Net}}) = \prod_{i \in \{1,n\}} \sqrt{S^{\star}_{\text{CHSH}}(\rho_i)} \prod_{i=2}^{n-1} \sqrt{\tau_{i,0}}. \tag{13.44}$$

*Proof.* By direct application of Theorem 7, $S^{\star}_{n\text{-Chain}\{\mathbf{G}^{\text{CHSH}}\}}(\rho^{\text{Net}}) = S^{\star}_{2\text{-Star}\{\mathbf{G}^{\text{CHSH}}\}}(\rho_1 \otimes \rho_n) \prod_{i=2}^{n-1} \sqrt{\tau_{i,0}}$, then using Theorem 6 we recover Eq. (13.44) because $S^{\star}_{2\text{-Star}\{\mathbf{G}^{\text{CHSH}}\}}(\rho_1 \otimes \rho_n) = \prod_{i \in \{1,n\}} \sqrt{\mathbf{G}^{\text{CHSH}}(\rho_i)}$. To achieve the maximal score, the central sources $i \in [2, n-1]$ are measured by the observables $A_z^i \otimes B_z^i = Z \otimes Z$ for all $z \in \mathbb{B}$ while sources $i \in \{1, n\}$ are measured by the observables described in the proof of Theorem 6. That is,

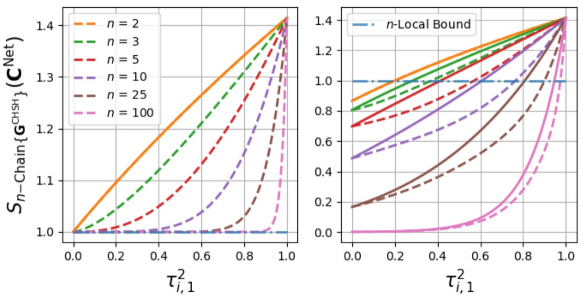



Figure 13.5: **Comparing the chain $n$-locality scores $S^{\star}_{n\text{-Chain}\{\mathbf{G}^{\text{CHSH}}\}}(\rho^{\text{Net}})$ (solid) is compared with $\widehat{S}^{\star}_{n\text{-Chain}\{\mathbf{G}^{\text{CHSH}}\}}(\rho^{\text{Net}})$ (dashed):** (Left) For various $n$, We set $\tau^2_{i,0} = 1$ and vary $\tau^2_{i,1} \in [0,1]$. (Right) We set $\tau^2_{i,0} = \frac{3}{4} + \frac{1}{4}\tau^2_{i,1}$ and vary $\tau^2_{i,1} \in [0,1]$.

the external parties measure $A^1_0 = B^n_0 = Z$ and $A^1_1 = B^n_1 = X$, while $B^1_z = (\tau_{1,0}Z + \tau_{1,1}(-1)^z X)/(\tau^2_{1,0} + \tau^2_{1,1})^{1/2}$ and $A^n_z = (\tau_{n,0}Z + \tau_{n,1}(-1)^z X)/(\tau^2_{n,0} + \tau^2_{n,1})^{1/2}$. □

The maximal $n$-local chain score $S^{\star}_{n\text{-Chain}\{\mathbf{G}^{\text{CHSH}}\}}(\rho^{\text{Net}})$ derived in Theorem 8 is distinct from previous results that assume the central parties to measure the observables [332]

$$\widehat{C}^{i,i+1}_z = (1-z)Z \otimes Z + zX \otimes X. \tag{13.45}$$

This choice of observable is physically motivated because it can be implemented as a coarse-graining on the outputs of a Bell state measurement, which is important for entanglement swapping [110]. However, these observables lead to a suboptimal $n$-local chain score

$$\widehat{S}^{\star}_{n\text{-Chain}\{\mathbf{G}^{\text{CHSH}}\}}(\rho^{\text{Net}}) \equiv \sqrt{\textstyle\prod_{i=1}^n \tau_{i,0} + \prod_{i=1}^n \tau_{i,1}} \tag{13.46}$$

where the two external parties measure the observables

$$\hat{A}^1_x = \frac{\prod_{i=1}^n \sqrt{\tau_{i,0}}Z + (-1)^x \prod_{i=1}^n \sqrt{\tau_{i,1}}X}{\sqrt{\prod_{i=1}^n \tau_{i,0} + \prod_{i=1}^n \tau_{i,1}}}, \tag{13.47}$$

and similarly for $\widehat{B}^n_y$. In general, $S^{\star}_{n\text{-Chain}\{\mathbf{G}^{\text{CHSH}}\}}(\rho^{\text{Net}}) \geq \widehat{S}^{\star}_{n\text{-Chain}\{\mathbf{G}^{\text{CHSH}}\}}(\rho^{\text{Net}})$ where equality occurs only in special cases.

**Corollary 3.** $S^{\star}_{n\text{-Chain}\{\mathbf{G}^{\text{CHSH}}\}}(\rho^{\text{Net}}) = \widehat{S}^{\star}_{n\text{-Chain}\{\mathbf{G}^{\text{CHSH}}\}}(\rho^{\text{Net}})$ if $\tau_{i,0} = \tau_{i,1}$ for all $i \in [2, n-1]$ and $S^{\star}_{2\text{-Star}\{\mathbf{G}^1,\mathbf{G}^2\}}(\rho_1 \otimes \rho_n) = \widehat{S}^{\star}_{2\text{-Star}\{\mathbf{G}^{\text{CHSH}}\}}(\rho_1 \otimes \rho_n)$.

The result Corollary 3 was observed in reference [102] where significant chain $n$-locality observations were obtained when $\widehat{S}_{n\text{-Chain}\{\mathbf{G}^{\text{CHSH}}\}}(\mathbf{C}^{\text{Net}}) \leq 1$. For instance, let $\rho^{\text{Net}} = \rho_1 \otimes (\bigotimes_{i=2}^{n-1} \gamma_i) \otimes \rho_n$ where $T_{\gamma_i} = \text{diag}\,(0,0,1)$, then $S^{\star}_{n\text{-Chain}\{\mathbf{G}^{\text{CHSH}}\}}(\rho^{\text{Net}}) = S^{\star}_{2\text{-Star}\{\mathbf{G}^{\text{CHSH}}\}}(\rho_1 \otimes \rho_n)$ and $\widehat{S}^{\star}_{n\text{-Chain}\{\mathbf{G}^{\text{CHSH}}\}}(\rho^{\text{Net}}) \leq 1$. However, if the central parties measure the observables $\widehat{C}^{(i,i+1)}_z$ as in Eq. (13.45), then an $n$-local violation $\widehat{S}_{n\text{-Chain}\{\mathbf{G}^{\text{CHSH}}\}}(\mathbf{C}^{\text{Net}}) > 1$ asserts that each source is nonclassical. As a consequence, full quantum network nonlocality [170] cannot be witnessed in the chain network with respect to local qubit observables, but can if central parties measure their qubits in mutually unbiased bases.

Overall, the $n$-local violations of $S^{\star}_{n\text{-Chain}\{\mathbf{G}^{\text{CHSH}}\}}(\rho^{\text{Net}})$ are more robust to noise than the $n$-local violations of $\widehat{S}^{\star}_{n\text{-Chain}\{\mathbf{G}^{\text{CHSH}}\}}(\rho^{\text{Net}})$. In Fig. 13.5, we illustrate cases where $S^{\star}_{n\text{-Chain}\{\mathbf{G}^{\text{CHSH}}\}}(\rho^{\text{Net}}) > \widehat{S}^{\star}_{n\text{-Chain}\{\mathbf{G}^{\text{CHSH}}\}}(\rho^{\text{Net}})$. In both plots, we consider uniform noise on all sources such that Corollary 1 is satisfied such that $S^{\star}_{2\text{-Star}\{\mathbf{G}^{\text{CHSH}}\}}(\rho_1 \otimes \rho_n) = S^{\star}_{2\text{-Star}\{\mathbf{G}^{\text{CHSH}}\}}(\rho_1 \otimes \rho_n)$. Thus, we ensure that the separation between $S^{\star}_{n\text{-Chain}\{\mathbf{G}^{\text{CHSH}}\}}(\rho^{\text{Net}})$ and $\widehat{S}^{\star}_{n\text{-Chain}\{\mathbf{G}^{\text{CHSH}}\}}(\rho^{\text{Net}})$ is inherent to the chain network. In the left plot, we consider noise where $\tau_{i,0}$ is preserved but $\tau_{i,1}$ is damped. For all $n > 2$, we find $S^{\star}_{n\text{-Chain}\{\mathbf{G}^{\text{CHSH}}\}}(\rho^{\text{Net}}) = S^{\star}_{2\text{-Star}\{\mathbf{G}^{\text{CHSH}}\}}(\rho_1 \otimes \rho_n) \geq \widehat{S}^{\star}_{n\text{-Chain}\{\mathbf{G}^{\text{CHSH}}\}}(\rho^{\text{Net}}) \geq 1$ where the separation increases with $n$. In the right plot, we consider noise such that $\tau_{i,0}$ is also damped, but the bias $\tau_{i,0} > \tau_{i,1}$ is preserved. For all $n > 2$, we find that

$$S^{\star}_{2\text{-Star}\{\mathbf{G}^{\text{CHSH}}\}}(\rho_1 \otimes \rho_n) > S^{\star}_{n\text{-Chain}\{\mathbf{G}^{\text{CHSH}}\}}(\rho^{\text{Net}}) \geq \widehat{S}^{\star}_{n\text{-Chain}\{\mathbf{G}^{\text{CHSH}}\}}(\rho^{\text{Net}}) \tag{13.48}$$

and we find examples where $S^{\star}_{n\text{-Chain}\{\mathbf{G}^{\text{CHSH}}\}}(\rho^{\text{Net}}) > 1 > \widehat{S}^{\star}_{n\text{-Chain}\{\mathbf{G}^{\text{CHSH}}\}}(\rho^{\text{Net}})$. Hence the local qubit measurement strategy discussed in Theorem 8 lead to $n$-locality violations in noisy settings when the results of previous works predict the correlations to be local [332].

## *13.5 The Noise Robustness of Star and Chain Non-n-Locality*

In the previous sections we derived general results for the maximal $n$-locality violations in the star and chain networks. Here we extend these results to the noise robustness of star and chain non-$n$-locality. Since we derive explicit maximal violations for $S_{n\text{-Star}\{\mathbf{G}^{\text{CHSH}_i}\}_i}$ and $S_{n\text{-Chain}\{\mathbf{G}^{\text{CHSH}}\}}$ in terms of the CHSH violation on each source, the following noise robustness results are generalizations of the CHSH noise robustness results in Chapter 12.

### *13.5.1 Unital Noise Robustness of Star and Chain Non-n-Locality*

**Theorem 9.** Consider an $n$-local star or chain network that has qubit-separable PVM measurements and a unital qubit noise model on each source such that $\mathcal{N}^{\text{Net}} = \bigotimes_{i=1}^{n} \mathcal{U}^{A_i} \otimes \mathcal{U}^{B_i}$. The maximal noisy $n$-local star and chain scores are

$$\widetilde{S}^{\star}_{n\text{-Star}\{\mathbf{G}^{\text{CHSH}_i}\}_i}(\mathcal{N}^{\text{Net}}) = \frac{1}{2}\prod_{i=1}^{n} \widetilde{S}^{\star}_{\text{CHSH}}(\mathcal{U}^{A_i} \otimes \mathcal{U}^{B_i})^{\frac{1}{n}}, \tag{13.49}$$

$$\widetilde{S}^{\star}_{n\text{-Chain}\{\mathbf{G}^{\text{CHSH}}\}}(\mathcal{N}^{\text{Net}}) = \widetilde{S}_{2\text{-Star}\{\mathbf{G}^{\text{CHSH}}\}}(\mathcal{U}^{A_1} \otimes \mathcal{U}^{B_1} \otimes \mathcal{U}^{A_n} \otimes \mathcal{U}^{B_n}) \prod_{i=2}^{n-1} \sqrt{u_0^{A_i} u_0^{B_i}} \tag{13.50}$$

where $\widetilde{S}^{\star}_{\text{CHSH}}(\mathcal{U}^{A_i} \otimes \mathcal{U}^{B_i}) = 2\sqrt{(u_0^{A_i} u_0^{B_i})^2 + (u_1^{A_i} u_1^{B_i})^2}$ as given in Theorem 3 and $u_0 \geq u_1$ are the two largest singular values of the respective channel's Pauli matrix representation $M_{\mathcal{U}}$.

*Proof.* In Theorem 3 we showed that $\widetilde{S}^{\star}_{\text{CHSH}}(\mathcal{U}^{A_i} \otimes \mathcal{U}^{B_i}) = 2\sqrt{(u_0^{A_i} u_0^{B_i})^2 + (u_1^{A_i} u_1^{B_i})^2}$. Then, since Theorem 6 and Theorem 8 give the maximal $n$-locality star and chain scores for arbitrary mixed states in terms of their CHSH violation, we simply insert the maximal CHSH violation into these results to recover Eq. (13.49) and Eq. (13.50). Note that in Eq. (13.50), we use $\tau_{i,0} = u_0^{A_i} u_0^{B_i}$. □

**Proposition 11.** Consider a star or chain quantum network having a qubit depolarizing noise model,

$\mathcal{W}_{\vec{v}}^{\text{Net}} = \bigotimes_{i=1}^{n} \mathcal{D}_{v^{A_i}}^{A_i} \otimes \mathcal{D}_{v^{B_i}}^{B_i}$. The maximal $n$-local star and chain scores are

$$\widetilde{S}^{\star}_{n\text{-Star}\{\mathbf{G}^{\text{CHSH}_i}\}_i}(\mathcal{D}_{\vec{v}}^{\text{Net}}) = \sqrt{2}\left(\prod_{i=1}^{n} v^{A_i} v^{B_i}\right)^{\frac{1}{n}}, \tag{13.51}$$

$$\widetilde{S}^{\star}_{n\text{-Chain}\{\mathbf{G}^{\text{CHSH}}\}}(\mathcal{D}_{\vec{v}}^{\text{Net}}) = \sqrt{2}\left(\prod_{i=1}^{n} v^{A_i} v^{B_i}\right)^{\frac{1}{2}}. \tag{13.52}$$

*Proof.* A qubit depolarizing channel is diagonalized in the Pauli basis as $M_{\mathcal{D}_v} = \text{diag}(v, v, v)$. Eq. (13.51) and Eq. (13.52) follow directly from Theorem 9. □

**Proposition 12.** Consider a star or chain quantum network having a qubit dephasing noise model, $\mathcal{D}_{\vec{\gamma}}^{\text{Net}} = \bigotimes_{i=1}^{n} \mathcal{D}_{\gamma^{A_i}}^{A_i} \otimes \mathcal{D}_{\gamma^{B_i}}^{B_i}$. The maximal $n$-local star and chain scores are

$$\widetilde{S}^{\star}_{n\text{-Star}\{\mathbf{G}^{\text{CHSH}_i}\}_i}(\mathcal{D}_{\vec{\gamma}}^{\text{Net}}) = \prod_{i=1}^{n} \left(1 + (1-\gamma^{A_i})(1-\gamma^{B_i})\right)^{\frac{1}{2n}}, \tag{13.53}$$

$$\widetilde{S}^{\star}_{n\text{-Chain}\{\mathbf{G}^{\text{CHSH}}\}}(\mathcal{D}_{\vec{\gamma}}^{\text{Net}}) = \widetilde{S}^{\star}_{2\text{-Star}\{\mathbf{G}^{\text{CHSH}}\}}(\mathcal{D}_{\gamma^{A_1}}^{A_1} \otimes \mathcal{D}_{\gamma^{B_1}}^{B_1} \otimes \mathcal{D}_{\gamma^{A_n}}^{A_n} \otimes \mathcal{D}_{\gamma^{B_n}}^{B_n}). \tag{13.54}$$

*Proof.* A qubit dephasing channel is diagonalized in the Pauli basis as $M_{\mathcal{D}_\gamma} = \text{diag}(\sqrt{1-\gamma}, \sqrt{1-\gamma}, 1)$. Eq. (13.53) and Eq. (13.54) follow directly from Theorem 9. □

**Proposition 13.** Consider a star or chain quantum network having a two-qubit depolarizing noise model on each source $\mathcal{W}_{\vec{\gamma}}^{\text{Net}} = \bigotimes_{i=1}^{n} \mathcal{W}_{v^i}^{i}$ where $\mathcal{W}_{v^i}^{i} \in \text{CPTP}(\mathcal{H}_4^{A_iB_i} \to \mathcal{H}_4^{A_iB_i})$ and $\gamma_i = \frac{15}{16}(1 - v_i)$. The maximal noisy $n$-local star and chain scores are

$$\widetilde{S}^{\star}_{n\text{-Star}\{\mathbf{G}^{\text{CHSH}_i}\}_i}(\mathcal{W}_{\vec{\gamma}}^{\text{Net}}) = \sqrt{2}\left(\prod_{i=1}^{n} \left|1 - \frac{16}{15}\gamma_i\right|\right)^{\frac{1}{n}} \tag{13.55}$$

$$\widetilde{S}^{\star}_{n\text{-Chain}\{\mathbf{G}^{\text{CHSH}}\}}(\mathcal{W}_{\vec{\gamma}}^{\text{Net}}) = \sqrt{2}\sqrt{\prod_{i=1}^{n} \left|1 - \frac{16}{15}\gamma_i\right|}. \tag{13.56}$$

*Proof.* In references [269], [270] it was derived that the maximal $n$-local star score is $\widetilde{S}^{\star}_{n\text{-Star}\{\mathbf{G}^{\text{CHSH}_i}\}_i}(\mathcal{W}_{\vec{\gamma}}^{\text{Net}}) = \sqrt{2}\left(\prod_{i=1}^{n} v_i\right)^{\frac{1}{n}}$ and in reference [271] it was derived that the maximal $n$-local chain score is $\widetilde{S}^{\star}_{n\text{-Chain}\{\mathbf{G}^{\text{CHSH}}\}}(\mathcal{W}_{\vec{\gamma}}^{\text{Net}}) = \sqrt{2}\sqrt{\prod_{i=1}^{n} v_i}$. Substituting the noise parameter $\gamma_i$ for the visibility $v_i$ as $\gamma_i = \frac{15}{16}(1 - v_i)$ recovers Eq. (13.55) and Eq. (13.56). □

**Proposition 14.** Consider a quantum network with a classical white noise model on each dichotomic detector $\mathbf{W}_{\vec{\gamma}}^{\text{Net}} = \bigotimes_{j=1}^{m} \mathbf{W}_{\gamma_i}^{i}$. The maximal noisy $n$-local scores for the star and chain networks are

$$\widetilde{S}^{\star}_{n\text{-Star}\{\mathbf{G}^{\text{CHSH}_i}\}_i}(\mathbf{W}_{\vec{\gamma}}^{\text{Net}}) = \sqrt{2}\left(\prod_{j=1}^{n+1} (1 - \gamma_j)\right)^{\frac{1}{n}}, \tag{13.57}$$

$$\widetilde{S}^{\star}_{n\text{-Chain}\{\mathbf{G}^{\text{CHSH}}\}}(\mathbf{W}_{\vec{\gamma}}^{\text{Net}}) = \sqrt{2}\sqrt{\prod_{j=1}^{n+1} (1 - \gamma_j)}. \tag{13.58}$$

*Proof.* As given by Eq. (13.82), white noise on each detector is modeled by a POVM with elements $W_{\pm|x,\gamma} = (1-\gamma)P_{+|x} + \gamma W_{\pm|x}$ where $P_{\pm|x}$ are projectors onto even and odd parity subspaces and $W_{\pm|x} = \frac{1}{2}\mathbb{I}_{2^M}$ is a white noise POVM. For each external node $P^{A_j}_{\pm|x_j} \in \text{PVM}(\mathcal{H}_2)$ while for the $M$-qubit central nodes, $P^{C_k}_{\pm|z_k} \in \text{PVM}(\mathcal{H}_{2^M})$ where $\text{Tr}\left[P^{C_k}_{+|z_k}\right] = \text{Tr}\left[P^{C_k}_{-|z_k}\right]$. In both cases, the corresponding dichotomic observables having $\pm 1$ eigenvalues are constructed as

$$\widetilde{O}^{A_j}_{x_j,\gamma_j} = W_{+|x_j,\gamma_j} - W_{-|x_j,\gamma_j} = (1-\gamma_j)(P_{+|x} - P_{-|x}) = (1-\gamma_j)O^{A_j}_{x_j}, \tag{13.59}$$

where $O^{A_j}_{x_j}$ is the device's observable in the noiseless case. Note here that the classical white noise has the same effect on the observables of both single and multi-qubit detectors. Then, consider the noisy $m$-partite correlator

$$\left\langle \widetilde{O}^{\text{Net}}_{\vec{x},\vec{\gamma}} \right\rangle_{\rho^{\text{Net}}} = \text{Tr}\left[\left(\bigotimes_{j=1}^{m} \widetilde{O}^{A_j}_{x_j}\right)\rho^{\text{Net}}\right] = \text{Tr}\left[\left(\bigotimes_{j=1}^{m}(1-\gamma_j)O^{A_j}_{x_j}\right)\rho^{\text{Net}}\right] = \prod_{j=1}^{m}(1-\gamma_j)\left\langle O^{\text{Net}}_{\vec{x}}\right\rangle_{\rho^{\text{Net}}}, \tag{13.60}$$

where the $m$ detectors can be in any entangled topology and $O^{\text{Net}}_{\vec{x}}$ is the product of all $m$ detecotors' noiseless observables. Since $|\langle O^{\text{Net}}_{\vec{x}}\rangle| \leq 1$, it must hold that $|\left\langle \widetilde{O}^{\text{Net}}_{\vec{x},\vec{\gamma}}\right\rangle| \leq \prod_{j=1}^{m}(1-\gamma_j)$ where the upper bound is achieved only if the optimal observables are chosen for maximal non-$n$-locality. For the noisy quantum network the quantity $I_{n,y}$ defined in Eq. (13.15) is bound as

$$|I_{n,y}| \leq \frac{1}{\sqrt{2^n}}\prod_{j=1}^{n+1}(1-\gamma_j) \tag{13.61}$$

hence the maximal noisy $n$-local star score is

$$\widetilde{S}^{\star}_{n\text{-Star}\{\mathbf{G}^{\text{CHSH}_i}\}_i}(\mathbf{W}^{\text{Net}}_{\vec{\gamma}}) = \sqrt{2}\left(\prod_{j=1}^{n+1}(1-\gamma_j)\right)^{\frac{1}{n}} \tag{13.62}$$

and achieved using the optimal noiseless strategy for the star network [270]. Likewise, if the optimal strategy for the chain inequality is used [271], then

$$|J_{2,y}| = \frac{1}{2}\prod_{j=1}^{n+1}(1-\gamma_j), \tag{13.63}$$

and the maximal noisy $n$-local chain score is

$$\widetilde{S}^{\star}_{n\text{-Chain}\{\mathbf{G}^{\text{CHSH}}\}}(\mathbf{W}^{\text{Net}}_{\vec{\gamma}}) = \sqrt{2}\sqrt{\prod_{j=1}^{n+1}(1-\gamma_j)}. \tag{13.64}$$

□

### *13.5.2 Nonunital Noise Robustness of Star and Chain Non-n-Locality*

In this section we take a theoretical approach to the noise robustness of $n$-locality violations in star and chain networks. Using Theorem 6 and Theorem 8, we can extend the nonunital noise robustness results obtained

for the CHSH scenario in Section 12.1.4 to star and chain networks having local qubit measurements.

Consider the case where a nonunital channel is applied to each qubit such that

$$\widetilde{\rho}^{\text{Net}} = \mathcal{N}^{\text{Net}}(\rho^{\text{Net}}) = \bigotimes_{i=1}^{n} \mathcal{N}^{A_i} \otimes \mathcal{N}^{B_i}(\rho^{A_iB_i}), \quad \text{where} \quad \mathcal{N}^{A_i}, \mathcal{N}^{B_i} \in \text{CPTP}(\mathcal{H}_2 \to \mathcal{H}_2). \tag{13.65}$$

when local qubit observables are measured, then Theorem 6 and Theorem 8 can be used to obtain the maximal star and chain $n$-locality violations $\widetilde{S}^{\star}_{n\text{-Star}\{\mathbf{G}^{\text{CHSH}_i}\}_i}(\mathcal{N}^{\text{Net}})$ and $\widetilde{S}^{\star}_{n\text{-Chain}\{\mathbf{G}^{\text{CHSH}}\}}(\mathcal{N}^{\text{Net}})$.

In general, in the presence of nonunital noise, the CHSH violation is found to be greater for nonmaximally entangled state preparations than maximally entangled state preparations. Here a state is maximally entangled if $|\psi\rangle = U^A \otimes U^B |\Phi^+\rangle$ while a state is nonmaximally entangled if $|\psi_\lambda\rangle = \sqrt{\lambda}\,|00\rangle + \sqrt{1-\lambda}\,|11\rangle$ and local unitary rotations thereof.

**Proposition 15.** Consider an $n$-local star or chain network that prepares maximally entangled states $\rho^{\text{Net}} = \bigotimes_{i=1}^{n} |\psi\rangle\langle\psi|$ where $|\psi\rangle = U^A \otimes U^B |\Phi^+\rangle$, applies local qubit PVM measurements, and has an amplitude damping channel applied to each qubit as $\mathcal{N}^{\text{Net}}_{\vec{\gamma}} = \bigotimes_{i=1}^{n} \mathcal{A}_{\gamma^{A_i}} \otimes \mathcal{A}_{\gamma^{B_i}}$. The maximal $n$-local scores for the star and chain networks are

$$S^{\star}_{n\text{-Star}\{\mathbf{G}^{\text{CHSH}_i}\}_i}(\tilde{\rho}^{\text{Net}}) = \prod_{i=1}^{n} \left( \sqrt{2(1-\gamma^{A_i})(1-\gamma^{B_i})} \right)^{\frac{1}{n}}, \tag{13.66}$$

$$S^{\star}_{n\text{-Chain}\{\mathbf{G}^{\text{CHSH}}\}}(\tilde{\rho}^{\text{Net}}) = \sqrt{\sqrt{2(1-\gamma^{A_1})(1-\gamma^{B_1})}\sqrt{2(1-\gamma^{A_n})(1-\gamma^{B_n})}} \prod_{i=2}^{n-1} \sqrt{\tau_{i,0}} \tag{13.67}$$

where $\tau_{i,0} = \min\{\sqrt{(1-\gamma^{A_i})(1-\gamma^{B_i})}, \gamma^{A_i}\gamma^{B_i} + (1-\gamma^{A_i})(1-\gamma^{B_i})\}$

*Proof.* Using Proposition 15 we find that $S^{\star}_{\text{CHSH}}(\tilde{\rho}^{A_iB_i}) = 2\sqrt{2(1-\gamma^{A_i})(1-\gamma^{B_i})}$. Then using Theorem 6, we find the maximal $n$-local star score to be as shown in Eq. (13.66). Finally, using Theorem 8, we find the maximal $n$-local chain score to be as shown in Eq. (13.67) where we note that the maximal singular value $\tau_0 = \max\{\sqrt{(1-\gamma^{A_i})(1-\gamma^{B_i})}, \gamma^{A_i}\gamma^{B_i} + (1-\gamma^{A_i})(1-\gamma^{B_i})\}$ where the two options are taken from Eq. (12.39) and Eq. (12.40). □

## *13.6 Variational Optimization of Non-n-locality in Noisy Star and Chain Networks*

In this section, we demonstrate on a classical simulator that VQO can reproduce the maximal $n$-local scores derived Section 13.5. We begin with an overview of our use of VQO to investigate the noise robustness of quantum non-$n$-locality. Then, for unital and nonunital channel examples, we show that our VQO methods can obtain the theoretical $n$-local scores. We find that network noise models consisting of general unital qubit channels have their non-$n$-locality maximized by maximally entangled states are optimal. On the contrary, we show that when nonunital qubit amplitude damping channels are considered, there exists nonmaximally entangled state preparations that outperform maximally entangled preparations.

Our objectives are to verify that our VQO framework can reproduce known noise robustness results. Our investigative approach to noise robustness is distinct from previous works [286], [287] that evaluate the precise noise parameters at which nonlocality is broken. Instead, we use VQO to find maximal violations of a Bell inequality given a static noise model $\mathcal{N}^{\text{Net}}_{\vec{\gamma}}$. By scanning through the noise parameters, we create a picture

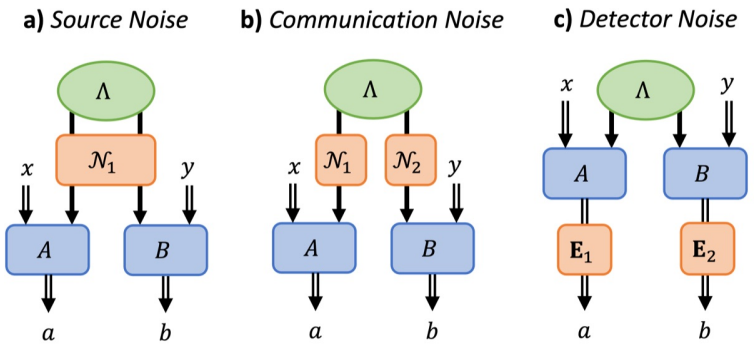



Figure 13.6: **Noise Models in Quantum Networks.** a) Source Noise: A quantum channel is applied to all qubits at a given source. b) Communication Noise: A quantum channel is applied independently to each qubit. c) Detector Noise: Classical post-processing is applied to the classical data output from the network.

of how the non-$n$-locality deteriorates as the amount of noise increases. Hence we are able to compare the relative noise robustness across different quantum network topologies.

We consider noise applied during the preparation, communication, and measurement stages of a quantum network in the $n$-local setting (see Fig. 13.6). Source noise occurs during the state preparation at each source and is modeled as $\mathcal{N}_{\vec{\gamma}}^{\text{Net}} = \bigotimes_{i=1}^{n} \mathcal{N}_{\gamma_i}^{\Lambda_i}$. Communication noise occurs during transmission of quantum states and is modeled as $\mathcal{N}_{\vec{\gamma}}^{\text{Net}} = \bigotimes_{k=1}^{l} \mathcal{N}_{\gamma_k}^{L_k}$. Detector noise occurs during measurement and is modeled as $\mathbf{E}_{\vec{\gamma}}^{\text{Net}} = \bigotimes_{j=1}^{m} \mathbf{E}_{\gamma_j}^{A_j}$. Alternatively, detector noise can be modeled as an adjoint channel applied to the measurement $\mathcal{N}_{\gamma}^{\text{Net}\dagger}(\Pi_{\vec{a}|\vec{x}}^{\text{Net}})$. For more details about dector noise please refer to Section 8.5.

To simplify our investigation, we characterize the network noise using one parameter. However, our VQO approach can easily be extended to nonuniform noise models as it introduces no additional computational overhead in comparison to the uniform noise model. First, we consider an ideal quantum network that has a single faulty component such as a noisy source, link, or measurement device. In this case, the noise parameters take the form $\gamma_v = (\gamma, 0, \ldots, 0)$. Second, we consider quantum networks having noise applied uniformly to all sources, links, or measurement devices. In this case, the noise parameters are $\gamma_v = (\gamma, \ldots, \gamma)$.

To evaluate the noise robustness we begin with a static noise model $\mathcal{N}_{\gamma}^{\text{Net}}$. To create a high-level overview of the noise robustness, we scan through the noise parameter $\gamma \in [0, 1]$ using an interval of 0.05. For each $\gamma$, we use VQO to find the optimal state preparations and measurements that maximize non-$n$-locality with respect to the Bell inequality $S_{\text{Bell}}$. We repeat this procedure for all considered network topologies depicted in Fig. 13.2 and compare their relative noise robustness. Furthermore, we compare the optimized results with theoretical bounds on the max violation.

We organize our investigation into two broad classes of noise, unital and nonunital. In each case, we consider noise applied to sources, qubit communication, and detectors. For each noise model we compare the $n \leq 4$ cases for the chain and star $n$-local inequalities $S_{n\text{-Star}\{\mathbf{G}^{\text{CHSH}_i}\}_i}$ and $S_{n\text{-Chain}\{\mathbf{G}^{\text{CHSH}}\}}$. Finally, we explore a wide range of state preparation and measurement ansatzes to broadly investigate the relation between entanglement and noise robustness. For details, please refer to Table 11.1.

### *13.6.1 Unital Noise Robustness of Star and Chain non-n-locality*

In this section, we demonstrate on a classical simulator that our VQO framework finds the theoretically maximal $n$-local violations in the presence of unital channels such as qubit depolarizing noise, source

depolarizing noise, detector white noise, and qubit dephasing noise. We show that network ansatzes allowing maximally entangled state preparation and local qubit measurements can achieve the theoretical maximum. Furthermore, we find no example of entangled measurements or nonmaximally entangled states that achieve a larger Bell score. These numerical results provide evidence that maximally entangled state preparations and local qubit measurements are optimal for general unital noise models on sources and measurements. Indeed, the numerical VQO results in this section led to the derivation of Theorem 6 and Theorem 8 in reference [333], thus our VQO framework demonstrates the ability to obtain novel theoretical insights.

We first consider qubit depolarizing noise on each qubit as

$$\mathcal{W}_{\vec{\gamma}}^{\text{Net}} = \bigotimes \mathcal{W}_{\gamma^{A_i}}^{A_i} \otimes \mathcal{W}_{\gamma^{B_i}}^{B_i} \tag{13.68}$$

where we consider either uniform noise or noise on just qubit $A_1$. A qubit depolarizing channel mixes white noise with the input qubit state as

$$\mathcal{D}_v(\rho) = v\rho + \frac{(1-v)}{2}\mathbb{I}_2 \text{Tr}\left[\rho\right], \tag{13.69}$$

where $v$ is a parameter commonly referred to as the *visibility*. The visibility relates to the noise parameter as

$$\gamma = \frac{3}{4}(1-v). \tag{13.70}$$

The qubit depolarizing channel's Kraus operators are

$$\begin{aligned} K_0 &= \sqrt{1-\gamma}\,\mathbb{I}_2, & K_1 &= \sqrt{\frac{\gamma}{3}}\,\sigma_x, \\ K_2 &= \sqrt{\frac{\gamma}{3}}\,\sigma_y, & K_3 &= \sqrt{\frac{\gamma}{3}}\,\sigma_z. \end{aligned} \tag{13.71}$$

In Proposition 11, we show that in the presence of qubit depolarizing noise, $\mathcal{D}_{\vec{v}}^{\text{Net}} = \bigotimes_{i=1}^{n} \mathcal{D}_{v^{A_i}}^{A_i} \otimes \mathcal{D}_{v^{B_i}}^{B_i}$ the maximal $n$-local star and chain scores are

$$S^{\star}_{n\text{-Star}\{\mathbf{G}^{\text{CHSH}_i}\}_i}(\mathcal{D}_{\vec{v}}^{\text{Net}}) = \sqrt{2}\left(\prod_{i=1}^{n} v^{A_i} v^{B_i}\right)^{\frac{1}{n}}, \tag{13.72}$$

$$S^{\star}_{n\text{-Chain}\{\mathbf{G}^{\text{CHSH}}\}}(\mathcal{D}_{\vec{v}}^{\text{Net}}) = \sqrt{2}\left(\prod_{i=1}^{n} v^{A_i} v^{B_i}\right)^{\frac{1}{2}}. \tag{13.73}$$

In Fig. 13.7, we find a close correspondence between the theoretical maximal Bell scores given by Eq. (13.72) and Eq. (13.73) and the maxima obtained using our VQO framework. In our optimizations, we consider arbitrary preparation and measurement ansatzes (see Table 11.1). We find that maximal violations can be achieved using the Bell state $|\Phi^+\rangle$ and local qubit measurements.

We now consider depolarizing noise on each source where the network noise model is

$$\mathcal{W}_{\vec{v}}^{\text{Net}} = \bigotimes_{i=1}^{n} \mathcal{W}_{v^i}^{\Lambda_i} \tag{13.74}$$

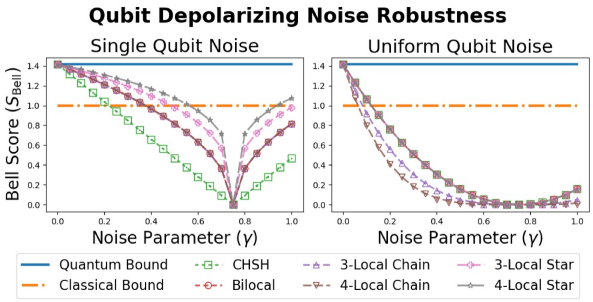



Figure 13.7: **Qubit depolarizing noise robustness of non-$n$-locality.** (Left) Qubit depolarizing noise is applied to qubit $q_1$. (Right) Qubit depolarizing noise is applied uniformly to all qubits. The markers show the maximal Bell score achieved using VQO. The dashed lines show the theoretical maximal given by Eq. (13.72) and Eq. (13.73) where the relation between the visibility $v$ and the noise parameter $\gamma$ is $\gamma = \frac{3}{4}(1-v)$.

where $\mathcal{W}_v^{\Lambda_i} \in \text{CPTP}(\mathcal{H}_4 \to \mathcal{H}_4)$. Depolarizing noise on two-qubit sources is expressed as

$$\mathcal{W}_v(\rho) = v\rho + \frac{(1-v)}{4}\mathbb{I}_4 \text{Tr}\left[\rho\right] \tag{13.75}$$

where the visibility $v$ relates to the noise parameter as

$$\gamma = \frac{15}{16}(1-v). \tag{13.76}$$

The Kraus operators for a two-qubit depolarizing channel are expressed as

$$K_{0,0} = \sqrt{1-\gamma}\mathbb{I}_4, \quad \left\{K_{i,j} = \sqrt{\frac{\gamma}{15}}\sigma_i \otimes \sigma_j\right\}_{i,j=0}^{3}, \tag{13.77}$$

where all $i$ and $j$ are considered except $i = j = 0$. The maximal Bell scores have previously been derived for $n$-local star and chain networks having source depolarizing noise. Namely, the maximal noisy Bell score for the star network is $\widetilde{S}^{\star}_{n\text{-Star}\{\mathbf{G}^{\text{CHSH}_i}\}_i}(\bigotimes_{i=1}^n \mathcal{W}_{v_i}^{\Lambda_i}) = \sqrt{2}\left(\prod_{i=1}^n v_i\right)^{1/n}$ [269], [270] and the maximal noisy Bell score for the chain network is $\widetilde{S}^{\star}_{n\text{-Chain}\{\mathbf{G}^{\text{CHSH}}\}}(\bigotimes_{i=1}^n \mathcal{W}_{v_i}^{\Lambda_i}) = \sqrt{2}\sqrt{\prod_{i=1}^n v_i}$ [271]. We use Eq. (13.76) to redefine the maximal violations in terms of the noise parameter as

$$\widetilde{S}_{n\text{-Star}\{\mathbf{G}^{\text{CHSH}_i}\}_i}(\bigotimes_{i=1}^n \mathcal{W}_{v_i}^{\Lambda_i}) = \sqrt{2}\left(\prod_{i=1}^n \left|1 - \frac{16}{15}\gamma_i\right|\right)^{1/n}, \tag{13.78}$$

$$\widetilde{S}^{\star}_{n\text{-Chain}}(\bigotimes_{i=1}^n \mathcal{D}_{v_i}^{\Lambda_i}) = \sqrt{2}\sqrt{\prod_{i=1}^n \left|1 - \frac{16}{15}\gamma_i\right|}. \tag{13.79}$$

In Fig. 13.8, we show that there is a close correspondence between the theoretical maximal Bell scores given by Eq. (13.78) and Eq. (13.79) and the maxima obtained using VQO optimization. In our optimization we consider source depolarizing noise an a broad range of preparation and measurement ansatzes (see Table 11.1). We find that the state preparation $|\Phi^+\rangle$ and local qubit measurements optimized over rotations about the $y$-axis are sufficient for maximal violations in all cases.

In the case when all detectors have binary outputs, consider a classical white noise error on all detectors

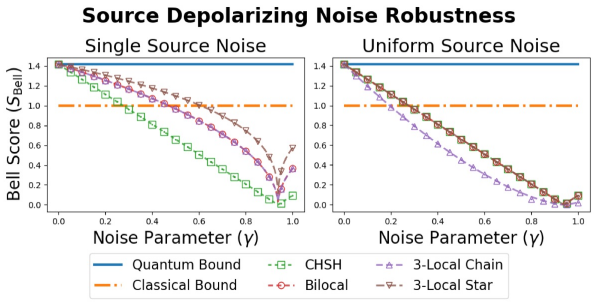



Figure 13.8: **Source depolarizing noise robustness of non-$n$-locality.** (Left) Depolarizing noise is applied to source $\Lambda_1$. (Right) Depolarizing noise is applied uniformly to all sources. The markers show the maximal Bell score achieved using VQO. The dashed lines show the theoretical maximal score from Eqs. (13.78) and (13.79).

in the network $\vec{A}$ where

$$\mathbf{W}^{\text{Net}}_{\vec{\gamma}} = \bigotimes_{A_i \in \vec{A}} \mathbf{W}^{A_i}_{\gamma_i^A} \tag{13.80}$$

and we define a white noise error as the classical postprocessing map

$$\mathbf{W}_\gamma = (1-\gamma)\mathbb{I}_2 + \frac{\gamma}{2}\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}, \tag{13.81}$$

where with probability $\gamma$, the detector outputs a binary value drawn from a uniform random distribution. Noting that Eq. (13.81) is a convex combination of noiseless and white noise measurements, we can equivalently express $\mathbf{W}_\gamma$ as the POVM

$$W_{\pm|x,\gamma} = (1-\gamma)\Pi_{\pm|x} + W_{\pm|x}, \tag{13.82}$$

where $W_{\pm|x} = \frac{1}{2}\mathbb{I}_M$ are white noise POVM elements and $\Pi_{\pm|x}$ are projectors onto even $(+)$ and odd $(-)$ parity subspaces that satisfy $\text{Tr}\left[\Pi_{\pm|x}\right] = 2^{M-1}$. The projector onto even and odd parity subspaces corresponds to the fact that we use parity to coarse grain a multi-bit output into a single bit.

In Proposition 3, we show that detector white noise post-processing map $\mathbf{W}_\gamma$ is a unital process equivalent to a depolarizing channel on the detector's $M$-qubits provided that $\text{Tr}\left[\Pi_{+|x}\right] = \text{Tr}\left[\Pi_{-|x}\right]$ for all $x$,

$$\mathcal{W}^{A_j}_{(1-\gamma)}(X) = (1-\gamma)X + \frac{\gamma}{2^M}\mathbb{I}_{2^M}\text{Tr}\left[X\right]. \tag{13.83}$$

Then in Proposition 14, we find that given the network noise model $\mathcal{N}^{\text{Net}}_{\vec{\gamma}} = \bigotimes_{j=1}^m \mathcal{D}^{A_j}_{(1-\gamma_j)}$, the maximal $n$-local star and chain scores are

$$\widetilde{S}^\star_{n\text{-Star}}(\bigotimes_{j=1}^m \mathcal{D}^{A_j}_{(1-\gamma_j)}) = \sqrt{2}\left(\prod_{j=1}^{n+1}(1-\gamma_j)\right)^{1/n}, \tag{13.84}$$

$$\widetilde{S}^\star_{n\text{-Chain}}(\bigotimes_{j=1}^m \mathcal{D}^{A_j}_{(1-\gamma_j)}) = \sqrt{2}\sqrt{\prod_{j=1}^{n+1}(1-\gamma_j)}. \tag{13.85}$$

In Fig. 13.9, we show that there is a close correspondence between the theoretical maximal $n$-local violations given by Eq. (13.84) and Eq. (13.85) and the maxima obtained using VQO. In our optimizations we consider a broad range of preparation and measurement ansatzes (see Table 11.1). For all networks, we find that the state preparation $|\Phi^+\rangle$ and local qubit measurements optimized over rotations about the $y$-axis are sufficient to achieve the theoretical maximum.

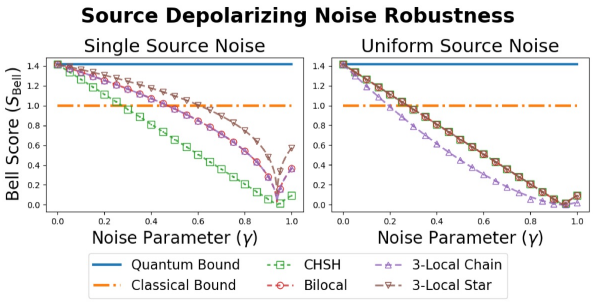



Figure 13.9: **Detector white noise robustness of non-$n$-locality.** (Left) Detector white noise is applied to measurement device $A_1$. (Right) Detector white noise is applied uniformly to all measurement devices. The markers show the maximal Bell score achieved using VQO. The dashed lines show the theoretical maximal score from Eqs. (13.84) and (13.85).

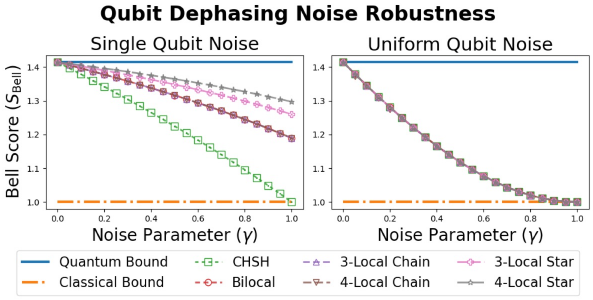



Figure 13.10: **Qubit dephasing noise robustness of non-$n$-locality.** (Left) Dephasing noise is applied to qubit $q_1$. (Right) Dephasing noise is applied uniformly to all qubits. The markers show the maximal Bell score achieved using VQO. The dashed lines show the maximal scores predicted by Eq. (13.89).

We now consider the case where dephasing noise is applied to each qubit such that we consider the noise model

$$\mathcal{D}^{\text{Net}}_{\vec{\gamma}} = \bigotimes_{i=1}^{n} \mathcal{D}^{A_i}_{\gamma^A} \otimes \mathcal{D}^{B_i}_{\gamma^B}. \tag{13.86}$$

Qubit dephasing noise is a unital channel that describes the decoherence process as

$$\mathcal{D}_\gamma(\rho) = \frac{1+\sqrt{1-\gamma}}{2}\rho + \frac{1-\sqrt{1-\gamma}}{2}\sigma_z\rho\sigma_z, \tag{13.87}$$

where the off diagonal terms go to zero as the noise parameter $\gamma$ increases. The Kraus operators for the dephasing channel are

$$K_0 = \begin{pmatrix} 1 & 0 \\ 0 & \sqrt{1-\gamma} \end{pmatrix} \text{ and } \quad K_1 = \begin{pmatrix} 0 & 0 \\ 0 & \sqrt{\gamma} \end{pmatrix}. \tag{13.88}$$

In Proposition 12 we show that, given a network noise model $\mathcal{D}^{\text{Net}}_{\vec{\gamma}} = \bigotimes_{i=1}^{n} \mathcal{D}^{A_i}_{\gamma^{A_i}} \otimes \mathcal{D}^{B_i}_{\gamma^{B_i}}$, the maximal $n$-local star and chain scores are

$$\widetilde{S}^{\star}_{n\text{-Star}\{\mathbf{G}^{\text{CHSH}_i}\}_i}(\mathcal{D}^{\text{Net}}_{\vec{\gamma}}) = \prod_{i=1}^{n} \left(1 + (1-\gamma^{A_i})(1-\gamma^{B_i})\right)^{\frac{1}{2n}}, \tag{13.89}$$

$$\widetilde{S}^{\star}_{n\text{-Chain}\{\mathbf{G}^{\text{CHSH}}\}}(\mathcal{D}^{\text{Net}}_{\vec{\gamma}}) = \widetilde{S}^{\star}_{2\text{-Star}\{\mathbf{G}^{\text{CHSH}}\}}(\mathcal{D}^{A_1}_{\gamma^{A_1}} \otimes \mathcal{D}^{B_1}_{\gamma^{B_1}} \otimes \mathcal{D}^{A_n}_{\gamma^{A_n}} \otimes \mathcal{D}^{B_n}_{\gamma^{B_n}}). \tag{13.90}$$

In Fig. 13.11, we show that there is a close correspondence between the theoretical maximal $n$-local violations given by Eq. (13.89) and Eq. (13.90) and the maxima obtained using VQO. In our optimizations we consider a broad range of preparation and measurement ansatzes (see Table 11.1). We find that maximally entangled states and local qubit measurements are sufficient for maximal $n$-local violation.

### *13.6.2 Nonunital Noise Robustness of Star and Chain non-n-locality*

In this section, we investigate the noise robustness of quantum non-$n$-locality with respect to nonunital noise. A quantum channel is nonunital if and only if it does not preserve the identity, $\mathcal{N}(\mathbb{I}) \neq \mathbb{I}$, where important examples include qubit amplitude damping, fixed errors on measurements, and two-qubit colored noise. As opposed to unital channels, we show that maximally entangled state preparations are not necessarily optimal in the presence of nonunital noise. We verify that VQO can be used to find the optimal state preparations and measurements for maximal $n$-local violation results by comparing our numerical results with the theoretical maximal noisy $n$-local scores derived in Section 13.5.2.

We first consider the case where an amplitude damping channel is applied to each qubit such that the network's noise model is

$$\mathcal{A}^{\text{Net}}_{\vec{\gamma}} = \bigotimes_{i=1}^{n} \mathcal{A}^{A_i}_{\gamma^{A_i}} \otimes \mathcal{A}^{B_i}_{\gamma^{B_i}}. \tag{13.91}$$

The qubit amplitude damping channel $\mathcal{A}_\gamma$ has Kraus operators

$$K_0 = \begin{pmatrix} 1 & 0 \\ 0 & \sqrt{1-\gamma} \end{pmatrix}, \quad K_1 = \begin{pmatrix} 0 & \sqrt{\gamma} \\ 0 & 0 \end{pmatrix} \tag{13.92}$$

where the effect on a qubit density matrix is

$$\mathcal{A}_\gamma(\rho) = \begin{pmatrix} \gamma + (1-\gamma)\rho_{00} & \sqrt{1-\gamma}\rho_{01} \\ \sqrt{1-\gamma}\rho_{10} & (1-\gamma)\rho_{11} \end{pmatrix}. \tag{13.93}$$

In star and chain networks, we compare the noise robustness of $n$-local violations for nonmaximally entangled states, $|\psi_\lambda\rangle = \sqrt{\lambda}\,|00\rangle + \sqrt{1-\lambda}\,|11\rangle$ for $\lambda \in [0,1]$, and maximally entangled states $\left|\psi_{\lambda=1/2}\right\rangle$. We express the noisy maximally entangled states as $\tilde{\rho}_{\frac{1}{2},\gamma} = \mathcal{A}^A_\gamma \otimes \mathcal{A}^B_\gamma(\rho_{\lambda=\frac{1}{2}})$, and noisy nonmaximally entangled states as $\tilde{\rho}_{\lambda,\gamma} = \mathcal{A}^A_\gamma \otimes \mathcal{A}^B_\gamma(\rho_\lambda)$. We now prove through example that nonmaximally entangled states achieve larger Bell scores than maximally entangled states. Then, we verify that our VQO software can reproduce our theoretical results.

In the CHSH scenario it was observed that, in the presence of uniform qubit amplitude damping noise nonmaximally entangled states lead to larger violation than maximally entangled states (see Fig. 12.1). This improvement in CHSH noise robustness can be generalized to our uniform qubit amplitude damping results in $n$-local star and chain networks. Using Proposition 10 and Theorem 6, we can write

$$S^\star_{n\text{-Star}\{\mathbf{G}^{\text{CHSH}_i}\}_i}(\textstyle\bigotimes_{i=1}^n \tilde{\rho}_{\lambda=\frac{1}{2},\gamma}) = \frac{1}{2} S^\star_{\text{CHSH}}(\tilde{\rho}_{\lambda^\star,\gamma}) \tag{13.94}$$

. For the chain network, we note that $\mathcal{A}_\gamma \otimes \mathcal{A}_\gamma(|00\rangle\langle 00|) = |00\rangle\langle 00|$, that is, the $|00\rangle\langle 00|$ state experiences no noise from amplitude damping. Since for $\rho = |00\rangle\langle 00|$, the largest singular value of $T_\rho$ is $\tau_0 = 1$, then

Theorem 8 can be used to obtain

$$S^{\star}_{n\text{-Chain}}(\bigotimes_{i=1}^{n}\tilde{\rho}_{\lambda=\frac{1}{2},\gamma}) = \tfrac{1}{2}S^{\star}_{\text{CHSH}}(\tilde{\rho}_{\lambda^{\star},\gamma}). \tag{13.95}$$

However, in the case of maximally entangled state preparations, the maximal singular value is $\tau_{0,\gamma} = \max\{1-\gamma, \gamma^2 + (1-\gamma)^2\}$, thus by Eq. (**??**)

$$S^{\star}_{n\text{-Chain}}\left(\bigotimes_{i=1}^{n}\tilde{\rho}_{\lambda=\frac{1}{2},\gamma}\right) = \frac{1}{2}S^{\star}_{\text{CHSH}}(\tilde{\rho}_{\lambda=\frac{1}{2},\gamma})\prod_{i=2}^{n-1}(\tau_{0,\gamma})^{\frac{1}{2}}. \tag{13.96}$$

We now consider the case where amplitude damping noise is applied to a single qubit in the network. Like in the uniform noise case, we start with the CHSH scenario and maximally entangled states. Using Proposition 15 and setting $\gamma^B = 0$ and $\gamma^A = \gamma \in [0,1]$, we find that $\tau_0 = \tau_1 = \sqrt{1-\gamma} \geq \tau_2 = 1-\gamma$. Thus,

$$S^{\star}_{n\text{-Star}\{\mathbf{G}^{\text{CHSH}_i}\}_i}(\tilde{\rho}_{\lambda=\frac{1}{2},\gamma}) = 2\sqrt{2(1-\gamma)} \tag{13.97}$$

and by Theorem 6 we see that

$$S^{\star}_{n\text{-Star}\{\mathbf{G}^{\text{CHSH}_i}\}_i}(\tilde{\rho}_{\lambda=\frac{1}{2},\gamma}\bigotimes_{i=2}^{n}|\Phi^+\rangle\langle\Phi^+|) = \left((\sqrt{2})^{n-1}\sqrt{2(1-\gamma)}\right)^{\frac{1}{n}}. \tag{13.98}$$

Note that the $n$-local chain score is equivalent to Eq. (13.98) because the interior chain nodes are not affected by the amplitude damping noise.

When nonmaximally entangled states are considered we use Proposition 9 to find that

$$S^{\star}_{\text{CHSH}}(\tilde{\rho}_{\lambda^{\star},\gamma}) = \begin{cases} 2\sqrt{2(1-\gamma)}, & \gamma \in [0,\frac{1}{2}] \\ 2, & \gamma \in [\frac{1}{2},1] \end{cases} \tag{13.99}$$

where $\lambda^{\star} = \frac{1}{2}$ when $\gamma \in [0,\frac{1}{2}]$ and $\lambda^{\star} = 1$ when $\gamma \in [\frac{1}{2},1]$. Thus, when amplitude damping is applied to a single qubit, maximally entangled state preparations are optimal on the range $\gamma \in [0,\frac{1}{2}]$ and the classical state preparation is optimal for $\gamma \in [\frac{1}{2},1]$. Then by Theorem 6, we find

$$\begin{aligned} &S^{\star}_{n\text{-Star}\{\mathbf{G}^{\text{CHSH}_i}\}_i}(\tilde{\rho}_{\lambda=\frac{1}{2},\gamma}\bigotimes_{i=2}^{n}|\Phi^+\rangle\langle\Phi^+|) = \\ &\qquad = \max\left\{\sqrt{2}^{\frac{n-1}{n}}, \left(\sqrt{2}^{n-1}\sqrt{2(1-\gamma)}\right)^{\frac{1}{n}}\right\}. \end{aligned} \tag{13.100}$$

In Fig. 12.1 and Fig. 13.11 we see a close correspondence between theoretical results and those found using VQO. In our optimizations, we consider a range of preparation and measurement ansatzes (see Table 11.1). We find that nonmaximally entangled state preparations and local qubit measurements are sufficient for maximal $n$-local violation

Now consider the case in which colored noise is applied to each source in the network as

$$\tilde{\rho}^{\text{Net}} = \bigotimes_{i=1}^{n}\mathcal{C}_{\gamma^{\Lambda_i}}(\rho^{\Lambda_i}). \tag{13.101}$$

Colored noise is typically found on the singlet state $|\Psi^-\rangle = \frac{1}{\sqrt{2}}(|01\rangle - |10\rangle)$ produced by parametric down conversion [213]. Colored noise is a two-qubit noise model representing depolarization on a preferred axis of

the state,

$$\mathcal{C}_\gamma(\rho) = (1-\gamma)\rho + \frac{\gamma}{2}(\left|\Psi^+\right\rangle\left\langle\Psi^+\right| + \left|\Psi^-\right\rangle\left\langle\Psi^-\right|), \tag{13.102}$$

where $|\Psi^\pm\rangle = (|01\rangle \pm |10\rangle)/\sqrt{2}$. The Kraus operators for colored noise are

$$\begin{aligned} K_0 &= \sqrt{1-\gamma}\mathbb{I}, \\ K_{\Psi^\pm,\Phi^\pm} &= \sqrt{\gamma/2}\left|\Psi^\pm\right\rangle\left\langle\Phi^\pm\right|, \\ K_{\Psi^\pm,\Psi^\pm} &= \sqrt{\gamma/2}\left|\Psi^\pm\right\rangle\left\langle\Psi^\pm\right|, \end{aligned} \tag{13.103}$$

where $|\Phi^\pm\rangle$ and $|\Psi^\pm\rangle$ constitute the Bell basis. Without loss of generality, we focus colored noise on $|\Psi^+\rangle$ and $|\Phi^+\rangle$ Bell states.

We begin by evaluating the CHSH score of the state $\tilde{\rho}_\gamma^\Psi = \mathcal{C}_\gamma(|\Psi^+\rangle\langle\Psi^+|)$. It follows that that the correlation matrix is $T_{\tilde{\rho}_\gamma^\Psi} = \text{diag}(1-\gamma, 1-\gamma, -1)$. Then, we use Eq. (12.13) to obtain $S^\star_{\text{CHSH}}(\tilde{\rho}_\gamma^\Psi) = 2\sqrt{1+(1-\gamma)^2}$. We note that this CHSH score is equivalent to Eq. (13.89), the score obtain for uniform qubit dephasing noise, hence, there is a direct correspondence between uniform qubit dephasing and colored noise applied to the $|\Psi^+\rangle\langle\Psi^+|$ state.

For uniform source colored noise, the maximal $n$-local star score is given by Theorem 6 as

$$S^\star_{n\text{-Star}\{\mathbf{G}^{\text{CHSH}_i}\}_i}(\bigotimes_{i=1}^n \tilde{\rho}_\gamma^\Psi) = \sqrt{1+(1-\gamma)^2}. \tag{13.104}$$

Since the maximal singular value of $T_{\tilde{\rho}_\gamma^\Psi}$ is $\tau_0 = 1$, $S^\star_{n\text{-Chain}\{\mathbf{G}^{\text{CHSH}}\}}(\bigotimes_{i=1}^n \tilde{\rho}_\gamma^\Psi) = S^\star_{2\text{-Star}\{\mathbf{G}^{\text{CHSH}}\}}(\bigotimes_{i=1}^2 \tilde{\rho}_\gamma^\Psi)$.

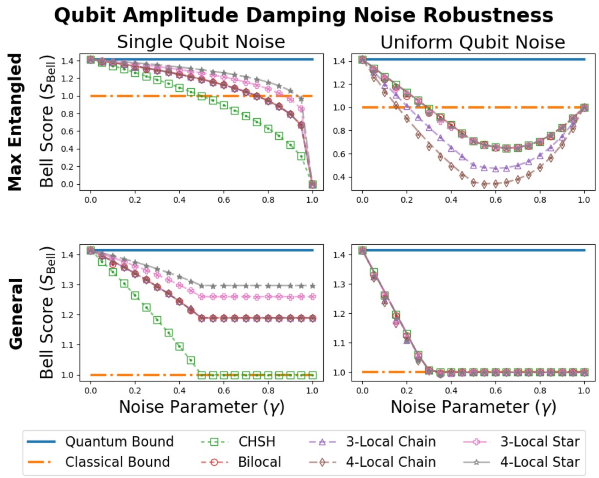



Figure 13.11: **Qubit amplitude damping noise robustness** Amplitude damping noise is applied to a single qubit (left column) and uniformly to all qubits (right column). The markers show the maximal Bell score achieved using VQO while the dashed lines show the theoretical maximum. The top row plots the maximal $n$-local scores for maximally entangled state preparations and local qubit measurements while the bottom row plots the maximal $n$-local scores over arbitrary state preparations and measurements. The theoretical scores are given by Eq. (13.98) (top right), Eq. (13.94) and Eq. (13.96) for the $n$-local star and chain networks respectively (top left), Eq. (13.100) (bottom left), and Eq. (12.50) (bottom right).

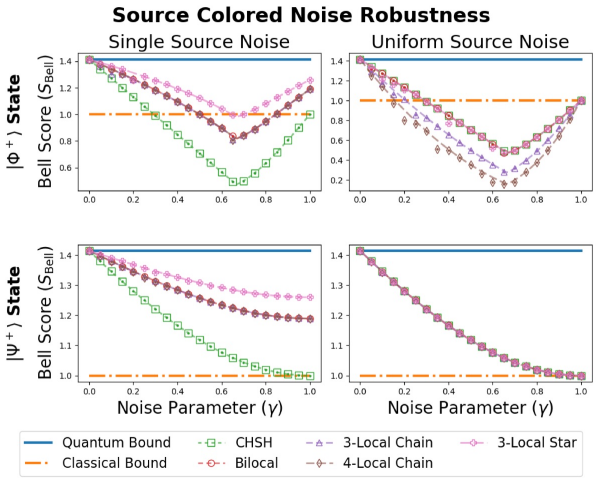



Figure 13.12: Colored noise is applied to a single source (left column) and uniformly to all sources (right column). The markers show the maximal Bell score achieved using VQO over $|\Phi^+\rangle$ state preparations and local qubit measurements (top row) and over $|\Psi^+\rangle$ state preparations and arbitrary measurements (bottom row). The dashed lines show the theoretically maximal score. (Top row) The theoretical lines are given by Eq. (13.109) for single source noise and by Eq. (13.107) and Eq. (13.108) for uniform source noise. (Bottom row) The theoretical lines are given by Eq. (13.105) and (13.104) for single source and uniform source noise.

When colored noise is applied to a single source we find that

$$S^{\star}_{n\text{-Star}\{\mathbf{G}^{\text{CHSH}_i}\}_i}(\tilde{\rho}^{\Psi}_{\gamma} \bigotimes_{i=2}^{n} |\Psi^+\rangle\langle\Psi^+|) = \left(\sqrt{2}^{n-1}\sqrt{1+(1-\gamma)^2}\right)^{\frac{1}{n}}. \tag{13.105}$$

Next we evaluate the CHSH score of the state $\tilde{\rho}^{\Phi}_{\gamma} = \mathcal{C}_{\gamma}(|\Phi^+\rangle\langle\Phi^+|)$. It follows that that the correlation matrix is $T_{\tilde{\rho}^{\Phi}_{\gamma}} = \text{diag}(1-\gamma, 1-2\gamma, -1)$. Then, we use Eq. (12.13) to obtain

$$S^{\star}_{\text{CHSH}}(\tilde{\rho}^{\Phi}_{\gamma}) = \begin{cases} 2\sqrt{2(1-\gamma)^2}, & \gamma\in[0,\frac{2}{3}] \\ 2\sqrt{(1-\gamma)^2+(1-2\gamma)^2}, & \gamma\in[\frac{2}{3},1] \end{cases}. \tag{13.106}$$

When colored noise is applied uniformly to $|\Phi^+\rangle\langle\Phi^+|$ sources, it follows from Theorem 6 that

$$S^{\star}_{n\text{-Star}\{\mathbf{G}^{\text{CHSH}_i}\}_i}(\bigotimes_{i=1}^{n} \tilde{\rho}^{\Phi}_{\gamma}) = \tfrac{1}{2} S^{\star}_{\text{CHSH}}(\tilde{\rho}^{\Phi}_{\gamma}). \tag{13.107}$$

Furthermore, using Theorem 8 in the presence of uniform colored noise, we find that the $n$-local chain score is

$$S^{\star}_{n\text{-Chain}\{\mathbf{G}^{\text{CHSH}}\}}(\bigotimes_{i=1}^{n} \tilde{\rho}^{\Phi}_{\gamma}) = \tfrac{1}{2} S^{\star}_{\text{CHSH}}(\tilde{\rho}^{\Phi}_{\gamma}) \prod_{i=2}^{n-1} \sqrt{\tau_0} \tag{13.108}$$

where $\tau_0 = 1-\gamma$. When colored noise is applied to a single source in a star we find that

$$S^{\star}_{n\text{-Star}\{\mathbf{G}^{\text{CHSH}_i}\}_i}(\tilde{\rho}^{\Phi}_{\gamma} \bigotimes_{i=2}^{n} |\Psi^+\rangle\langle\Psi^+|) = \left(\sqrt{2}^{n-1} \tfrac{1}{2} S^{\star}_{\text{CHSH}}(\tilde{\rho}_{\gamma})\right)^{\frac{1}{n}} \tag{13.109}$$

where the $n$-local chain score is given by the bilocal score.

In Fig. 13.12, we show that there is a close correspondence between the theoretical maximal $n$-local violations in the presence of colored noise and the maxima obtained using VQO. We consider a range of preparation and measurement ansatzes (see Table 11.1). We find that $|\Psi^+\rangle$ is the optimal state preparation

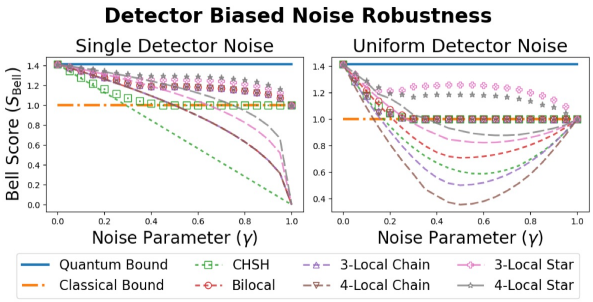



Figure 13.13: **Biased detector noise robustness.** (Left) Biased detector noise is applied to a single detector. (Right) Biased detector noise is applied uniformly to all detectors. The markers show the maximal Bell score achieved using VQO when using arbitrary state preparations. The dashed lines show the maximal Bell score achieved using VQO with respect to maximally entangled states

and that local qubit measurements are sufficient to achieve the maximal $n$-local violation in all cases.

Now consider the case in which each dichotomic detector has a bias in their output. We define a biased detector error as the classical post-processing map,

$$\mathbf{R}_\gamma = (1-\gamma)\mathbb{I}_2 + \gamma \begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix}, \tag{13.110}$$

where the fixed classical value of $+1$ is output whenever an error occurs. The biased detector error can be described by the POVM with elements

$$\widehat{\Pi}_{+|x} = (1-\gamma)\Pi_{+|x} + \gamma \mathbb{I}_{2^M}, \text{ and} \tag{13.111}$$

$$\widehat{\Pi}_{-|x} = (1-\gamma)\Pi_{-|x}, \tag{13.112}$$

where $\Pi_{+|x}$ and $\Pi_{-|x}$ constitute an $M$-qubit PVM. In Proposition 2, we show that the post-processing map $\mathbf{R}_\gamma$ in Eq. (13.110) is nonunital and equivalent to a partial replacer channel

$$\mathcal{R}_{\gamma,x}(\rho^{A_j}) = (1-\gamma)\rho^{A_j} + \gamma \rho'_x \text{Tr}\left[\rho^{A_j}\right] \tag{13.113}$$

applied to the local quantum state $\rho^{A_j}$ where the replacer state, $\rho'_x$, is a density operator contained by the projective subspace of $\Pi^{A_j}_{+|x}$.

In Fig. 13.13 we compare the VQO results of maximally entangled state preparations and arbitrary state preparations. We find significant improvements when arbitrary state preparations are considered, implying that nonmaximally entangled states are optimal for $n$-local violations in the presence of biased detector noise.

# Chapter 14

# Nonclassicality in Point-to-Point Signaling Systems

**Definition 83. Signaling Scenario DAG:** Consider the signaling scenarios as depicted in Fig. 14.1 in which a sender device $A$ encodes a classical input $x \in \mathcal{X}$ into a physical signal that is communicated to a receiver device $B$ that decodes the signal into a classical output $y \in \mathcal{Y}$. The physical signal may be either a classical message $m \in \mathcal{M}$ where $|\mathcal{M}| = d$ or quantum state $\rho \in D(\mathcal{H}_d)$ (see Fig. 14.1.a) and b)) where the corresponding DAG is expressed

$$A \to B = \text{Net}(\vec{N} = (A, B), \vec{E} = \{A \to B\}). \tag{14.1}$$

The communication can be assisted by a static nonsignaling resource shared by the sender and receiver such as shared randomness or entanglement (see. Fig 14.1.c-e)) where the corresponding DAG is expressed

$$\Lambda \prec A \to B = \text{Net}(\vec{N} = (\Lambda, A, B), \vec{E} = \{\Lambda \prec (A, B), A \to B\}). \tag{14.2}$$

Both DAGs have a strict ordering of events in which a source $\Lambda$ shares the static resource, the sender $A$ encodes its message, and the receiver $B$ decodes the message where the source and input $x$ are independent.

No matter which resources, quantum or classical, are considered in the signaling scenario, a behavior $\mathbf{P} : \mathcal{X} \to \mathcal{Y}$ is produced. Thus, we can investigate the nonclassicality of the quantum resource configurations depicted in the signaling scenarios in Fig. 14.1.b,d,e) by comparing them with the similar classical resource configurations in Fig. 14.1.a,c). While the lens of operational nonclassicality allows us to view signaling scenarios with an interesting new perspective, their behaviors have been extensively studied over the past 5 decades. Nevertheless, our theoretical and computational frameworks are able to push our understanding of signaling systems further.

A famous result by Holevo implies that when quantum signals are used, the classical communication capacity is limited as $\log_2(d)$, where $d$ is the input Hilbert space dimension of the quantum signaling state $\rho_x \in D(\mathcal{H}_d)$ [173]; hence a noiseless classical channel transmitting $d$ messages has a capacity no less than the quantum signaling system. However, channel capacity is just one figure of merit, and there may be other features of the quantum system's behavior that do not readily admit a classical simulation. The strongest form of simulation being an exact replication of the transition probabilities $P(y|x) = \text{Tr}\left[\Pi_y \rho_x\right]$ for any set of

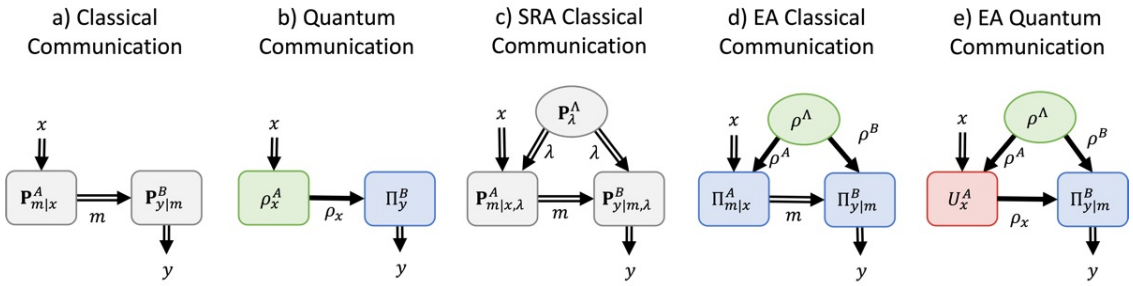



Figure 14.1: **Signaling Systems:** From left to right signaling resources increase in their operational value when considering the behavior of a signaling systems $\mathbf{P} : \mathcal{X} \to \mathcal{Y}$ where $x \in \mathcal{X}$ and $y \in \mathcal{Y}$. Double-lined arrows depict classical communication while single-lined arrows depict quantum communication. Classical devices are shown in gray while quantum devices are colored by their function where green represents preparation devices, blue represents measurement devices, and red represents processing devices. a) Classical communication, device $A$ sends a classical message $m$ to device $B$. b) Quantum communication, device $A$ sends a quantum state $\rho_x$ to device $B$. c) Shared randomness-assisted classical communication where devices $A$ and $B$ share a randomness source. d) Entanglement-assisted classical communication where devices $A$ and $B$ share an entanglement source. e) Entanglement-assisted quantum communication where devices $A$ and $B$ share entanglement and perform quantum communication.

states $\{\rho_x\}_{x\in\mathcal{X}}$ and POVM $\{\Pi_y\}_{y\in\mathcal{Y}}$. This problem falls in the domain of zero-error quantum information theory [335]–[339], which considers the classical and quantum resources needed to perfectly simulate a given quantum channel $\mathcal{N}$. Unlike the communication capacity, a zero-error simulation of $\mathbf{P}_\mathcal{N}$ typically requires additional communication beyond the input dimension of $\mathcal{N}$. For example, a noiseless qubit channel $\mathrm{id}_2$ can generate channels $\mathbf{P}_{\mathrm{id}_2}$ that cannot be faithfully simulated using a one bit of classical communication [231].

The simulation question becomes more interesting if "static" nonsignaling resources are used for the channel simulation [340], [341], in addition to the "dynamic" resource of noiseless classical communication. For example, shared randomness is a relatively inexpensive classical resource that Alice and Bob can use to coordinate their encoding and decoding maps used in the simulation protocol shown in Fig. **??**.c). Using shared randomness, a channel can be exactly simulated with a forward noiseless communication rate that asymptotically approaches the channel capacity; a fact known as the Classical Reverse Shannon Theorem [342]. More powerful static resources such as shared entanglement or non-signaling correlations could also be considered [338], [343], [344].

While the Classical Reverse Shannon Theorem describes many-copy channel simulation, we focus on zero-error channel simulation in the single-copy case. That is, we define the signaling dimension classical and quantum channels to be the minimum amount of classical communication (with unlimited shared randomness) needed to perfectly simulate every behavior that can be produced by a channel. The concept of signaling dimension was first introduced in Reference **??** in which it was shown that both quantum and classical signaling systems obey a *no hyper-signaling principle* that prevents physical systems from communicating too much. Significant progress in understanding the signaling dimension was made by Frenkel and Weiner who showed that every $d$-dimensional quantum channel requires no more than $d$ classical messages to perfectly simulate [115]. This result is a "fine-grained" version of Holevo's Theorem for channel capacity mentioned above. However, the Frenkel-Weiner bound is not tight in general when noisy quantum communication channels are considered. For example, suppose the completely depolarizing channel acts upon a $d$ dimensional quantum state as $\mathcal{D}(\rho) = \mathbb{I}/d$. For any choice of inputs $\{\rho_x\}_x$ and POVM $\{\Pi_y\}_y$, the Frenkel-Weiner protocol yields a simulation of $\mathbf{P}_\mathcal{D}$ that uses a forward transmission of $d$ messages. However, this is clearly not

optimal since $\mathbf{P}_{\mathcal{D}}$ can be reproduced with no forward communication whatsoever; Bob just samples from the distribution $P(y) = \mathrm{Tr}\left[[]\,\Pi_y\right]/d$. A fundamental problem is then to understand when a noisy classical channel sending $d$ messages from Alice to Bob actually requires $d$ noiseless classical messages for zero-error simulation. As a key result of this chapter, we provide a family of simple tests that determine when this amount of communication is needed. In other words, we characterize the conditions in which the simulation protocol of Frenkel and Weiner is optimal for the purposes of sending $d$ messages over a $d$-dimensional quantum channel.

To this end, we develop a framework for device-independent certification and estimation of signaling dimension similar to previous approaches used for the device-independent dimensionality testing of classical and quantum devices [227], [228], [345]–[347]. Specifically, we obtain signaling dimension witnesses that stipulate necessary conditions on the signaling dimension of $\mathcal{N}$ in terms of the probabilities $P_{\mathcal{N}}(y|x)$, with no assumptions made about the quantum states $\{\rho_x\}_x$, POVM $\{\Pi_y\}_y$, or channel $\mathcal{N}$ **DallArno-2017a**. Complementary results have been obtained by Dall'Arno *et al.* who approached the simulation problem from the quantum side and characterized the set of channels $\mathbf{P}_{\mathcal{N}}$ that can be obtained using binary encodings for special types of quantum channels $\mathcal{N}$ **DallArno-2017a**. In this paper, we compute a wide range of signaling dimension witnesses using the adjacency decomposition technique [128], recovering prior results of Frenkel and Weiner [115] and generalizing work by Heinosaari and Kerppo [348]. For certain cases we prove that these inequalities are complete, i.e. providing both necessary and sufficient conditions for signaling dimension. As a further application, we compute bounds for the signaling dimension of partial replacer channels. The supporting software and numerics for this work is found on Github [349].

## *14.1 The Behaviors of Signaling Systems*

In this section, we discuss the general structure of the behaviors that can be produced by signaling scenarios and compare broadly, the behaviors that can be produced when classical and quantum resources are present. We consider each of the signaling scenarios depicted by the DAGs in Fig. 14.1 when a limited amount of communication is allowed between the sender and receiver. Such scenrios describe a lossy communication process where the sender must compress the input, making it impossible for the receiver to decode the message without error. Overall, we build upon results in the literature to establish a general hierarchy based upon the operational value of each signaling resource configuration.

### *14.1.1 Unassisted Classical Communication*

Consider the signaling scenario $A \to B$ in the discrete memoryless setting where for each input $x \in \mathcal{X}$ given to the sender $A$, the receiver outputs $y \in \mathcal{Y}$ with probability $P(y|x)$ (see Fig. 14.1.a). The associated behavior $\mathbf{P}^{A\to B}$ then belongs to the full set of probabilities such that $\mathbf{P}^{A\to B} \in \mathcal{P}_{\mathcal{Y}|\mathcal{X}}$. However, if the amount of communication is restricted such that in each shot, the sender can only send one classical message $m \in \mathcal{M}$ where $|\mathcal{M}| = d$.

**Definition 84. The Set of Classical Signaling Behaviors:** $\mathcal{C}^{\mathcal{X}\xrightarrow{d}\mathcal{Y}} \subseteq \mathcal{P}_{\mathcal{Y}|\mathcal{X}}$, the set of behaviors $\mathbf{P}^{A\xrightarrow{d}B} : \mathcal{X} \to \mathcal{Y}$ that decompose as

$$\mathbf{P}^{A\xrightarrow{d}B} = \mathbf{P}^B\mathbf{P}^A = \sum_{m\in\mathcal{M}}\sum_{m\in\mathcal{M}} \left|\mathbf{P}^B_m\right\rangle\!\left\langle\mathbf{P}^A_m\right| = \sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}}\sum_{m=\in\mathcal{M}} P^B_{y|m}P^A_{m|x}\,|y\rangle\langle x| \tag{14.3}$$

where $|\mathcal{M}| = d$ and $\mathbf{P}^A : \mathcal{X} \to \mathcal{M}$ and $\mathbf{P}^B : \mathcal{M} \to \mathcal{Y}$ correspond to the sender's and receiver's respective

encoding and decoding behaviors. Furthermore, $\langle \mathbf{P}^A_m | = \sum_{x\in\mathcal{X}} P^A_{m|x} |m\rangle\langle x|$ is the $m^{th}$ row of behavior $\mathbf{P}^A$, and $|\mathbf{P}^B_m\rangle = \sum_{y\in\mathcal{Y}} P(y|m) |y\rangle\langle m|$ is the $m^{th}$ column of behavior $\mathbf{P}^B$.

**Lemma 9.** [123], [350] A behavior $\mathbf{P} \in \mathcal{P}_{\mathcal{Y}|\mathcal{X}}$ belongs to $\mathcal{C}^{\mathcal{X}\xrightarrow{d}\mathcal{Y}} \subseteq \mathcal{P}_{\mathcal{Y}|\mathcal{X}}$ if and only if the nonnegative rank of $\mathbf{P}$ satisfies

$$\text{Rank}_+ (\mathbf{P}) \leq d \tag{14.4}$$

where the nonnegative rank of a matrix $M \in \mathbb{R}^{n'\times n}$ where $M_{i,j} \geq 0$ is defined [351]

$$\text{Rank}_+ (M) = \min \left\{ k \mid \sum_{i=1}^{k} |a_i\rangle\langle b_i| = M \quad \text{where} \quad |a_i\rangle, |b_i\rangle \in \mathbb{R}^n_+ \right\}, \tag{14.5}$$

or alternatively,

$$\text{Rank}_+ (M) = \min \left\{ k \mid M_{y,x} = \langle a_y | b_x \rangle \quad \text{where} \quad |a_y\rangle, |b_x\rangle \in \mathbb{R}^k_+ \right\} \tag{14.6}$$

where the two nonnegative rank definitions are equivalent, but differ in how the matrix $M$ is decomposed.

Since the set of classical signaling behaviors is characterized by the nonnegative, it is important to note that for a nonnegative matrix $M \in \mathbb{R}^{n'\times n}$ that

$$\text{Rank} (M) \leq \text{Rank}_+ (M) \leq \min\{n, n'\}. \tag{14.7}$$

**Lemma 10.** The extreme points of the set $\mathcal{C}^{\mathcal{X}\xrightarrow{d}\mathcal{Y}}$ are

$$\mathcal{V}^{A\xrightarrow{d}B} = \left\{ \mathbf{D}^{A\xrightarrow{d}B} \in \mathbb{B}^{|\mathcal{Y}|\times|\mathcal{X}|} \mid \text{Rank} \left( \mathbf{D}^{A\xrightarrow{d}B} \right) \leq d \right\} \tag{14.8}$$

where $\mathbf{D}^{A\xrightarrow{d}B}$ are deterministic behaviors having binary elements.

*Proof.* The extreme points of a set of classical behaviors correspond to deterministic behaviors. Then, behavior $\mathbf{P}^B$ and $\mathbf{P}^A$ have ranks $\text{Rank} \left( \mathbf{P}^B \right) \leq d$ and $\mathbf{P}^A \leq d$ then $\text{Rank} \left( \mathbf{P}^B \mathbf{P}^A \right) \leq \min\{\text{Rank} \left( \mathbf{P}^B \right), \text{Rank} \left( \mathbf{P}^A \right)$. □

### *14.1.2 Unassisted Quantum Signaling*

Consider the quantum signaling scenario $A \to B$ in the discrete memoryless setting where for each input $x \in \mathcal{X}$ given to the sender $A$, the receiver outputs $y \in \mathcal{Y}$ with probability $P(y|x)$ (see Fig. 14.1.b). The associated behavior $\mathbf{P}^{A\to B}$ then belongs to the full set of probabilities such that $\mathbf{P}^{A\to B} \in \mathcal{P}_{\mathcal{Y}|\mathcal{X}}$. However, the amount of quantum communication is restricted such that, in each shot, the sender can transmit only one quantum state $\rho \in \mathcal{H}_d$ .

**Definition 85. The Set of Quantum Signaling Behaviors:** $\mathcal{Q}^{\mathcal{X}\xrightarrow{d}\mathcal{Y}} \subseteq \mathcal{P}_{\mathcal{Y}|\mathcal{X}}$, the set of behaviors $\mathbf{P}^{A\xrightarrow{d}B} : \mathcal{X} \to \mathcal{Y}$ that decompose as

$$\mathbf{P}^{A\xrightarrow{d}B} = \sum_{x\in\mathcal{X}} \sum_{y\in\mathcal{Y}} \text{Tr} \left[ \Pi^B_y \text{id}^{A\to B}(\rho^A_x) \right] |y\rangle\langle x| \tag{14.9}$$

where $\rho_x \in D(\mathcal{H}^A_d)$ is the encoded quantum state, $\text{id}^{A\to B} \in \text{CPTP}(\mathcal{H}^A_d \to \mathcal{H}^B_{d'})$ is a noiseless quantum communication channel, and $\{\Pi^B_y\}_{y\in\mathcal{Y}} \in \text{POVM}(\mathcal{H}^B_{d'})$ is the receiver's measurement. Note that the dimension of the quantum state and measurement need not be equal, that is, $d \neq d'$ is permitted as long as the channel satisfies $\mathcal{N}^{A\to B} : D(\mathcal{H}^A_d) \to D(\mathcal{H}^B_{d'})$.

**Definition 86. The Set of Noisy Quantum Signaling Behaviors:** $\mathcal{Q}^{\mathcal{X}\xrightarrow{d}\mathcal{Y}}(\mathcal{N}) \subseteq \mathcal{Q}^{\mathcal{X}\xrightarrow{d}\mathcal{Y}}$, the set of behaviors $\mathbf{P}^{A\xrightarrow{d}B} : \mathcal{X} \to \mathcal{Y}$ that decompose as

$$\mathbf{P}^{A\xrightarrow{d}B} = \sum_{x\in\mathcal{X}} \sum_{y\in\mathcal{Y}} \text{Tr}\left[\Pi_y^B \mathcal{N}^{A\to B}(\rho_x^A)\right] |y\rangle\langle x| \tag{14.10}$$

where $\rho_x \in D(\mathcal{H}_d^A)$ is the encoded quantum state, $\mathcal{N} \in \text{CPTP}(\mathcal{H}_d^A \to \mathcal{H}_{d'}^B)$ is a noisy quantum communication channel, and $\{\Pi_y^B\}_{y\in\mathcal{Y}} \in \text{POVM}(\mathcal{H}_{d'}^B)$ is the receiver's measurement. Note that the dimension of the quantum state and measurement need not be equal, that is, $d \neq d'$ is permitted as long as the channel satisfies $\mathcal{N}^{A\to B} : D(\mathcal{H}_d^A) \to D(\mathcal{H}_{d'}^B)$.

**Lemma 11.** (Ref. [352] Lemma 5 and Ref. [123] Proposition 6) A behavior $\mathbf{P} \in \mathcal{P}_{\mathcal{Y}|\mathcal{X}}$ belongs to $\mathcal{Q}^{\mathcal{X}\xrightarrow{d}\mathcal{Y}} \subseteq \mathcal{P}_{\mathcal{Y}|\mathcal{X}}$ if and only if the positive semidefinite rank of $\mathbf{P}$ satisfies

$$\text{Rank}_{\text{PSD}}(\mathbf{P}) \leq d \tag{14.11}$$

where the positive semidefinite rank of a matrix $M \in \mathbb{R}^{n'\times n}$ where $M_{i,j} \geq 0$ is defined [353]

$$\text{Rank}_{\text{PSD}}(M) = \min\left\{k \in \mathbb{Z}_{\geq 0} \mid M_{i,j} = \text{Tr}[A_i B_j] \ \forall\, i \in [n'],\ j \in [n] \quad \text{where} \quad 0 \leq A_i, B_j \in \mathbb{C}^{k\times k}\right\}. \tag{14.12}$$

For a nonnegative matrix $M \in \mathbb{R}^{n'\times n}$, the positive semidefinite rank and nonnegative rank are related as [353]

$$\frac{1}{2}\sqrt{1 + 8\text{Rank}(M)} - \frac{1}{2} \leq \text{Rank}_{\text{PSD}}(M) \leq \text{Rank}_{+}(M) \leq \min\{n, n'\}. \tag{14.13}$$

In Eq. (14.13), the inequality $\text{Rank}_{\text{PSD}}(M) \leq \text{Rank}_{+}(M)$ asserts that any behavior $\mathbf{P}\mathcal{C}^{\mathcal{X}\xrightarrow{d}\mathcal{Y}}$ satisfies $\text{Rank}_{\text{PSD}}(\mathbf{P}) \leq \text{Rank}_{+}(\mathbf{P}) \leq d$, implying that quantum signaling resources can simulate the classical behavior using equal, or less communication. As a result, the relation $\mathcal{C}^{\mathcal{X}\xrightarrow{d}\mathcal{Y}} \subseteq \mathcal{Q}^{\mathcal{X}\xrightarrow{d}\mathcal{Y}}$ must hold in general, which is not surprising as a qudit of quantum communication ought to be able to simulate a dit of classical communication. Nevertheless, the inequality in Eq. (14.13) provides sufficient grounds to suggest the existence of weakly nonclassical quantum signaling correlations.

### *14.1.3 Shared Randomness-Assisted Classical Signaling*

Consider the classical signaling scenario $\Lambda \overset{\infty}{\prec} (A\xrightarrow{d}B)$ in which $d$ classical messages are used, but the sender and receiver share an unlimited amount of randomness (see Fig. 14.1.c). The set of classical signaling behaviors then forms a convex polytope that we refer to as the *signaling polytope*.

**Definition 87. Signaling Polytope:** $\mathcal{C}_\Lambda^{\mathcal{X}\xrightarrow{d}\mathcal{Y}} = \text{Conv}\left(\mathcal{C}^{\mathcal{X}\xrightarrow{d}\mathcal{Y}}\right)$, The set of all classical signaling behaviors $\mathbf{P} \in \mathcal{P}_{\mathcal{Y}|\mathcal{X}}$ whose probabilities decompose as

$$\mathbf{P} = \sum_{\lambda\in\Omega^\Lambda} P_\lambda^\Lambda \mathbf{P}_\lambda^B \mathbf{P}_\lambda^A = \sum_{\lambda\in\Omega^\Lambda} \sum_{m\in\mathcal{M}} \sum_{y\in\mathcal{Y}} \sum_{x\in\mathcal{X}} P_\lambda^\Lambda P_{y|m}^B P_{m|x}^A |y\rangle\langle x| \tag{14.14}$$

where $\mathbf{P}_\lambda^A \in \mathcal{P}_{\mathcal{X}|\mathcal{M}}$ and $\mathbf{P}_\lambda^B \in \mathcal{P}_{\mathcal{Y}|\mathcal{M}}$, $|\mathcal{M}| = d$, and $\Omega^\Lambda$ is a countably infinite set from which $\lambda$ is drawn with probability $P_\lambda^\Lambda$.

It is important to note that the signalling polytope has been described previously in references [114], [115], [347]. The signaling polytope $\mathcal{C}_\Lambda^{\mathcal{X}\xrightarrow{d}\mathcal{Y}}$ is convex due to the fact that any randomness in the encoding or

decoding maps of Eq. (14.14) can be absorbed into the shared randomness distribution $P_\lambda^\Lambda$. The vertices of the signaling polytope $\mathcal{C}_\Lambda^{\mathcal{X}\overset{d}{\to}\mathcal{Y}}$ are then $\mathcal{V}^{A\overset{d}{\to}B}$ where $\mathcal{C}_\Lambda^{\mathcal{X}\overset{d}{\to}\mathcal{Y}} = \text{Conv}\left(\mathcal{V}^{A\overset{d}{\to}B}\right)$ where $\mathcal{V}^{A\overset{d}{\to}B}$ is the set of deterministic behaviors having linear rank $d$ or less as defined in Lemma 14.8. When $d \geq 2$, we find that $\text{Dim}\left(\mathcal{C}_\Lambda^{\mathcal{X}\overset{d}{\to}\mathcal{Y}}\right) = |\mathcal{X}|(|\mathcal{Y}|-1)$ when $d \geq 2$ as a result of the $|\mathcal{X}|$ normalization constraints on the full probability polytope $\mathcal{P}_{\mathcal{Y}|\mathcal{X}}$. When $d = 1$, $\text{Dim}\left(\mathcal{C}_\Lambda^{\mathcal{X}\overset{d=1}{\to}\mathcal{Y}}\right) = |\mathcal{Y}|$ because the input and and output are independent such that $P(y|x) = P(y|x')$ for all $x, x' \in \mathcal{X}$ and $y \in \mathcal{Y}$. By the Weyl-Minkowski Theorem [118], the signaling polytope can be described as the intersection of a finite set of linear half-spaces

$$\mathcal{C}_\Lambda^{\mathcal{X}\overset{d}{\to}\mathcal{Y}} = \bigcap_{(\gamma,\mathbf{G})\in\mathcal{G}^{A\overset{d}{\to}B}} \mathcal{H}(\gamma,\mathbf{G}) \tag{14.15}$$

where $\mathcal{G}$ is the set signaling polytope facet inequalities $\langle \mathbf{G}, \mathbf{P}\rangle \leq \gamma$ and a half-space is explicitly defined as

$$\mathcal{H}(\gamma,\mathbf{G}) \equiv \left\{ \mathbf{P} \in \mathcal{P}_{\mathcal{Y}|\mathcal{X}} \;\middle|\; \gamma \geq \langle \mathbf{G}, \mathbf{P}\rangle \quad \text{and} \quad \text{Dim}\left(\mathcal{H}(\gamma,\mathbf{G})\right) = \text{Dim}\left(\mathcal{P}_{\mathcal{Y}|\mathcal{X}}\right) \right\}. \tag{14.16}$$

**Lemma 12.** A behavior $\mathbf{P} \in \mathcal{P}_{\mathcal{Y}|\mathcal{X}}$ belongs to the signaling polytope such that $\mathbf{P} \in \mathcal{C}^{\mathcal{X}\overset{d}{\to}\mathcal{Y}}$ if and only if it satisfies all facet inequalities $(\gamma, \mathbf{G}) \in \mathcal{G}^{A\overset{d}{\to}B}$ where $\langle \mathbf{G}, \mathbf{P}\rangle \leq \gamma$.

**Lemma 13.** [115] The hierarchy $\mathcal{Q}^{\mathcal{X}\overset{d}{\to}\mathcal{Y}} \subset \mathcal{C}_\Lambda^{\mathcal{X}\overset{d}{\to}\mathcal{Y}}$ holds such that the behavior of any quantum signaling system that transmits $d$-dimensional states $\rho_x \in D(\mathcal{H}_d)$ can be simulated exactly using noiseless $|\mathcal{M}| = d$-level classical communication.

### *14.1.4 Entanglement-Assisted Classical Signaling*

Consider the classical signaling scenario $\Lambda \overset{\infty}{\prec} (A\overset{d}{\to}B)$ in which $d$ classical messages are used, but the sender and receiver share an entangled state $\rho^{AB} \in D(\mathcal{H}_{d'}^A \otimes \mathcal{H}_{d'}^B$ (see Fig. 14.1.d). Note that the dimension of the entangled state is independent from the amount of communication. Moreover, the the entanglement is a correlating resource such that it does not require communication to use and can be distributed prior to the experiment being run.

**Definition 88. The Set of Entanglement-Assisted Classical Signaling Behaviors:** $\mathcal{C}_{\text{EA}}^{\mathcal{X}\overset{d}{\to}\mathcal{Y}} \subseteq \mathcal{P}_{\mathcal{Y}|\mathcal{X}}$, the set of behaviors $\mathbf{P}^{A\overset{d}{\to}B} : \mathcal{X} \to \mathcal{Y}$ that decompose as

$$\mathbf{P}^{A\overset{d}{\to}B} = \sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}}\sum_{m\in\mathcal{M}} \text{Tr}\left[\Pi_{m|x}^A \otimes \Pi_{y|m}^B \rho^{AB}\right] |y\rangle\langle x| = \sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}}\sum_{m\in\mathcal{M}} \text{Tr}\left[\Pi_{y|m}^B \rho_{m|x}^B\right] P_{m|x}^A\, |y\rangle\langle x| \tag{14.17}$$

where $|\mathcal{M}| = d$ is the number of distinct classical messages $m$, $\{\Pi_{m|x}^A\}_{m\in\mathcal{M}}, \{\Pi_{y|m}^B\}_{y\in\mathcal{Y}} \in \text{POVM}(\mathcal{H}_{d'}^A)$ measure the entangled state,

$$P_{m|x}^A = \text{Tr}\left[\Pi_{m|x}^A \otimes \mathbb{I}_{d'}^B \rho^{AB}\right], \quad \rho_{m|x}^B = \text{Tr}_A\left[\Pi_{m|x}^A \otimes \mathbb{I}_{d'}^B \rho^{AB}\right] \frac{1}{P_{m|x}^A} \tag{14.18}$$

.

Interestingly, it has been observed that entanglement-assisted classical signaling can lead to behaviors that cannot be reproduced by shared randomness-assisted classical signaling [116], [350]. Furthermore, if the dimension of entanglement is unrestricted, then it must hold that $\mathcal{C}_\Lambda^{\mathcal{X}\overset{d}{\to}\mathcal{Y}} \subseteq \mathcal{C}_{\text{EA}}^{\mathcal{X}\overset{d}{\to}\mathcal{Y}}$, indicating that entanglement assisted classical signaling can produce strongly nonclassical behaviors. Indeed, we will show

examples for which a classical bit of communication assisted by a two-qubit entangled state can achieve strongly nonclassical behaviors that cannot be reproduced, even with an unlimited amount of shared randomness.

### *14.1.5 Entanglement-Assisted Quantum Signaling*

Consider the quantum signaling scenario $\Lambda \overset{\infty}{\prec} (A \overset{d}{\to} B)$ in which $d$-dimensional quantum states $\rho_x \in D(\mathcal{H}_d)$ are used as messages and the sender and receiver share an entangled state $\rho^{AB} \in D(\mathcal{H}_d^A \otimes \mathcal{H}_d^B)$ (see Fig. 14.1.e). Note that we set the subsystems of the entangled state to have dimension $d$ for convenience. Moreover, the entanglement is a correlating resource such that it does not require communication to use and can be distributed prior to the experiment being run.

**Definition 89. The Set of Entanglement-Assisted Quantum Signaling Behaviors:** $\mathcal{Q}_{\text{EA}}^{\mathcal{X} \overset{d}{\to} \mathcal{Y}} \subseteq \mathcal{P}_{\mathcal{Y}|\mathcal{X}}$, the set of behaviors $\mathbf{P}^{A \overset{d}{\to} B} : \mathcal{X} \to \mathcal{Y}$ that decompose as

$$\mathbf{P}^{A \overset{d}{\to} B} = \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} \text{Tr}\left[\Pi_y^{BB'} \mathcal{E}_x^{A \to B'} \otimes \mathbb{I}_d^B(\rho^{AB})\right] |y\rangle\langle x| \tag{14.19}$$

where $\mathcal{E}_x^{A \to B'} \in \text{CPTP}(\mathcal{H}_d^A \to \mathcal{H}_d^{B'})$ is the quantum encoding map, $\{\Pi_{m|x}^A\}_{m \in \mathcal{M}}, \{\Pi_y^{BB'}\}_{y \in \mathcal{Y}} \in \text{POVM}(\mathcal{H}_d^B \otimes \mathcal{H}_d^{B'})$ is the receiver's joint measurement on the the entangled state $\mathcal{E}_x^{A \to B'} \otimes \mathbb{I}_d^B(\rho^A B)$.

The entanglement-assisted quantum signaling scenario encompasses the standard dense-coding protocol [37], [265] in which the classical communication capacity of a quantum channel is doubled through the use of entanglement-assistance. In this way, entanglement-assisted quantum signaling behaviors demonstrate a significant operational advantage over other resource configurations in signaling scenarios, leading to behaviors that are more strongly nonclassical than those produced in entanglement-assisted classical signaling scenarios.

## *14.2 Signaling Dimension*

In this section, we describe the classical bounds on signaling scenarios and how it relates to the classical simulation cost of the signaling system. Indeed, we introduce the *signaling dimension* as the amount classical communication needed to exactly reproduce a given signaling behavior $\mathbf{P}$, which may apply classical or quantum resources. Furthermore, in characterizing the signaling dimension we obtain witnesses of both weak and strong nonclassicality.

**Definition 90.** Consider the behavior $\mathbf{P} \in \mathcal{P}_{\mathcal{Y}|\mathcal{X}}$.

(i) **Classical Signaling Dimension:** $\kappa(\mathbf{P}) = \min\{d \in \mathbb{N} \mid \mathbf{P} \in \mathcal{C}^{\mathcal{X} \overset{d}{\to} \mathcal{Y}}\}$

(ii) **Signaling Dimension:** $\kappa_\Lambda(\mathbf{P}) = \min\{d \in \mathbb{N} \mid \mathbf{P} \in \mathcal{C}_\Lambda^{\mathcal{X} \overset{d}{\to} \mathcal{Y}}\}$

where $\mathcal{C}^{\mathcal{X} \overset{d}{\to} \mathcal{Y}}$ is the set of classical signaling behaviors (see Def. 84 and Fig. 14.1.a), and $\mathcal{C}_\Lambda^{\mathcal{X} \overset{d}{\to} \mathcal{Y}}$ is the signaling polytope (see Def. 87 and Fig. 14.1.c).

Remarkably, these device-independent quantities describe the minimal classical communication resources needed to simulate a given behavior. The key difference between the two is that the *classical signaling dimension* does not permit shared randomness while the *signaling dimension* allows for an unlimited amount of shared randomness between sender and receiver. A similar quantity to the classical signaling dimension is

given by Heinosaari *et al.* [123] while the signaling dimension was introduced by Dall'Arno *et al.* [122]. In general, the two signaling dimensions relate as

$$1 \leq \kappa_\Lambda(\mathbf{P}) \leq \kappa(\mathbf{P}) \leq \min\{|\mathcal{X}|, |\mathcal{Y}|\}, \tag{14.20}$$

where the inclusion of shared randomness can reduce the cost of simulation. Indeed, when the upper bound on the RHS is achieved, the sender and receiver can simulate any $\mathbf{P} \in \mathcal{P}_{\mathcal{Y}|\mathcal{X}}$: either the sender applies channel $\mathbf{P}$ on the input $x \in \mathcal{X}$ and sends the output $y \in \mathcal{Y}$ to the receiver, or the sender transmits the input $x$ to the receiver who locally applies the behavior $\mathbf{P}$. In Theorem 10 we provide necessary and sufficient conditions for when this trivial upper bound is attained.

In our work "Certifying the Classical Simulation Cost of a Quantum Channel" [114], we extend the concept of signaling dimension to quantum channels.

**Definition 91.** For a quantum channel $\mathcal{N} : \mathrm{CPTP}(\mathcal{H}^A_{d_A} \to \mathcal{H}^B_{d_B})$, let $\mathcal{Q}^{\mathcal{X}\to\mathcal{Y}}(\mathcal{N})$ denote the set of all quantum signaling behaviors whose probabilities decompose as $P(y|x) = \mathrm{Tr}\left[\Pi^B_y \mathcal{N}(\rho^A_x))\right]$ for any choice of state preparations $\{\rho^A_x \in D(\mathcal{H}^A_{d_A})\}_{x\in\mathcal{X}}$ and measurements $\{\Pi^B_y\}_{y\in\mathcal{Y}} \in \mathrm{POVM}(\mathcal{H}^B_{d_B})$.

(i) **The $\mathcal{X} \to \mathcal{Y}$ Signaling Dimension of a Quantum Channel:** $\kappa_\Lambda^{\mathcal{X}\to\mathcal{Y}}(\mathcal{N}) = \min\{d \in \mathbb{N} \mid \mathcal{Q}^{\mathcal{X}\to\mathcal{Y}}(\mathcal{N}) \subseteq \mathcal{C}_\Lambda^{\mathcal{X}\xrightarrow{d}\mathcal{Y}}\}$ where $\mathcal{X}$ and $\mathcal{Y}$ are fixed.

(ii) **The Signaling Dimension of a Quantum Channel:** $\kappa_\Lambda(\mathcal{N}) = \min\{d \in \mathbb{N} \mid \mathcal{Q}^{\mathcal{X}\to\mathcal{Y}}(\mathcal{N}) \subseteq \mathcal{C}_\Lambda^{\mathcal{X}\xrightarrow{d}\mathcal{Y}} \; \forall \; \mathcal{X}, \mathcal{Y}\}$ where the optimization is over all $\mathcal{X}$ and $\mathcal{Y}$.

The signaling dimension of a quantum channel quantifies the amount of classical communication needed to simulate any behavior that can be generated using a particular channel and any choice of state preparation and measurement. The two quantum channel signaling dimensions relate as

$$1 \leq \kappa_\Lambda^{\mathcal{X}\to\mathcal{Y}}(\mathcal{N}) \leq \kappa_\Lambda(\mathcal{N}) \leq \min\{d_A, d_B\}. \tag{14.21}$$

where $d_A$ and $d_B$ are the respective Hilbert space dimensions of the sender and receiver's systems. This bound is a direct consequence of Frenkel and Weiner's result [115], summarized in Lemma 13, which can be restated in our terminology as $\kappa_\Lambda(\mathrm{id}_d) = d$, where $\mathrm{id}_d \in \mathrm{CPTP}(\mathcal{H}_d \to \mathcal{H}_d)$ is the noiseless channel on a $d$-dimensional quantum system. To prove Eq. (14.21), either the sender prepares the states $\{\rho_x\}_{x\mathcal{X}}$ to the receiver who measures the POVM $\{\mathcal{N}^\dagger(\Pi_y)\}_y$, or the sender transmits the states $\{\mathcal{N}(\rho_x)\}_{x\in\mathcal{X}}$ to the receiver who measures the POVM $\{\Pi_y\}_{y\in\mathcal{Y}}$. Here $\mathcal{N}^\dagger$ denotes the adjoint map of $\mathcal{N}$. Furthermore, we consider only classical simulations assisted by unlimited shared randomness because otherwise an unbounded amount of classical communication may be needed to simulate a channel. Moreover, since quantum signaling systems have signaling dimension no larger than their Hilbert space, allowing shared randomness allows the minimal Hilbert space dimension to be estimated. Hence a device-independent certification of signaling dimension leads to a device-independent certification of the physical input/output Hilbert spaces of the channel connecting the sender and receiver.

Another relationship we observe is

$$\kappa_\Lambda^{\mathcal{X}\to\mathcal{Y}}(\mathcal{N}) = \kappa_\Lambda^{\mathcal{X}\to[d_B^2]}(\mathcal{N}) \qquad \forall \quad |\mathcal{Y}| \geq d_B^2. \tag{14.22}$$

This follows from Carathéodory's Theorem [354], which implies that every POVM on a $d_B$-dimensional system can be expressed as a convex combination of POVMs with no more than $d_B^2$ outcomes [355]. Since unlimited

shared randomness sis considered, the sender and receiver can always restrict their attention to POVMs with no more than $d_B^2$ outcomes for the purposes of simulating any behavior in $\mathcal{Q}^{\mathcal{X}\xrightarrow{d}\mathcal{Y}}(\mathcal{N})$ when $|\mathcal{Y}| \geq d_B^2$.

The signaling dimension of quantum and classical channels can be unified through the following observation. A classical channel from set $\mathcal{X}$ to $\mathcal{Y}$ can be represented by a CPTP map $\mathcal{N} \in \text{CPTP}(\mathcal{H}_d^A \to \mathcal{H}_d^B)$ that completely dephases its input and output in fixed orthonormal bases $\{|x\rangle\}_{x\in\mathcal{X}}$ and $\{|y\rangle\}_{y\in\mathcal{Y}}$, respectively. The transition probabilities of $\mathbf{P}_{\mathcal{N}} \in \mathcal{Q}^{\mathcal{X}\xrightarrow{d}\mathcal{Y}}(\mathcal{N})$ are then given by Eq. (14.10) as $P_{\mathcal{N}}(y|x) = \text{Tr}\left[|y\rangle\langle y|\mathcal{N}\big(|x\rangle\langle x|\big)\right]$. The channel $\mathcal{N}$ can be used to generate another channel $\overline{\mathcal{N}}$ with input and output alphabets $\overline{\mathcal{X}}$ and $\overline{\mathcal{Y}}$ by performing a preprocessing column stochastic map $\mathbf{T} : \overline{\mathcal{X}} \to \mathcal{X}$ and postprocessing column stochastic map $\mathbf{R} : \mathcal{Y} \to \overline{\mathcal{Y}}$, thereby yielding the channel $\mathbf{P}_{\overline{\mathcal{N}}} = \mathbf{R}\mathbf{P}_{\mathcal{N}}\mathbf{T}$. When this relationship holds, $\mathbf{P}_{\overline{\mathcal{N}}}$ is said to be *ultraweakly majorized* by $\mathbf{P}_{\mathcal{N}}$ [123], [348], and the signaling dimension of $\kappa_\Lambda(\mathbf{P}_{\overline{\mathcal{N}}}) \leq \kappa_\Lambda(\mathbf{P}_{\mathcal{N}})$ [338]. Thus, $\kappa_\Lambda(\mathbf{P}_{\mathcal{N}})$ can always be obtained from the transition probabilities $P_{\mathcal{N}}(y|x)$ directly with no need for pre/postprocessing.

It is important to note that the signaling dimension of the quantum channel as in Definition 91 assumes that there is no entanglement assistance between the sender and receiver. In principle, however, these quantities can be generalized to account for entanglement assistance.

**Definition 92.** Let $\mathcal{N}^{\Lambda\to AB} \in \text{CPTP}(\mathcal{H}^\Lambda \to \mathcal{H}^{AB})$ denote a noisy channel applied to bipartite entangled state $\rho^\Lambda \in D(\mathcal{H}_{d_\Lambda^2})$ and $\mathcal{N}^{A\to B'} \in \text{CPTP}(\mathcal{H}^A \to \mathcal{H}^{B'})$ denote a noisy quantum channel between sender and receiver where $\mathcal{C}_{\text{EA}}^{\mathcal{X}\to\mathcal{Y}}(\mathcal{N}^{\Lambda\to AB})$ and $\mathcal{Q}^{\mathcal{X}\to\mathcal{Y}}(\mathcal{N}^{\Lambda\to AB},\mathcal{N}^{A\to B'})$ denote the sets of entanglement-assisted classical and quantum behaviors respectively.

(i) **Signaling Dimension of an Entanglement-Assisted Classical Channel:**

$$\kappa_\Lambda(\mathcal{N}^{\Lambda\to AB}) = \min\{d \in \mathbb{N} \mid \mathcal{C}_{\text{EA}}^{\mathcal{X}\xrightarrow{d'}\mathcal{Y}}(\mathcal{N}^{\Lambda\to AB}) \subseteq \mathcal{C}_\Lambda^{\mathcal{X}\xrightarrow{d}\mathcal{Y}}\} \tag{14.23}$$

(ii) **Signaling Dimension of an Entanglement-Assisted Quantum Channel:**

$$\kappa_\Lambda(\mathcal{N}^{\Lambda\to AB},\mathcal{N}^{A\to B'}) = \min\{d \in \mathbb{N} \mid \mathcal{Q}_{\text{EA}}^{\mathcal{X}\to\mathcal{Y}}(\mathcal{N}^{\Lambda\to AB},\mathcal{N}^{A\to B'}) \subseteq \mathcal{C}_\Lambda^{\mathcal{X}\xrightarrow{d}\mathcal{Y}}\} \tag{14.24}$$

### *14.2.1 Nonnegative Rank as a Signaling Dimension Witness*

As discussed in Lemma 9, the set of classical signaling behaviors $\mathcal{C}^{\mathcal{X}\xrightarrow{d}\mathcal{Y}}$ is exactly the subset of behaviors $\mathbf{P} \in \mathcal{P}_{\mathcal{Y}|\mathcal{X}}$ that have nonnegative rank satisfying $\text{Rank}_+(\mathbf{P}) \leq d$. Therefore,

$$\kappa(\mathbf{P}) = \text{Rank}_+(\mathbf{P}), \tag{14.25}$$

meaning that the behavior $\mathbf{P}$ can be simulated using only $|\mathcal{M}| = \text{Rank}_+(\mathbf{P})$ classical messages. In this way, a behavior's nonnegative rank places a lower bound on the amount of unassisted classical communication present in the signaling system.

On the contrary, recall from Lemma 11 that the set of quantum signaling behaviors $\mathcal{Q}^{\mathcal{X}\xrightarrow{d}\mathcal{Y}}$ is exactly the subset of behaviors $\mathbf{P} \in \mathcal{P}_{\mathcal{Y}|\mathcal{X}}$ that have positive semidefinite rank satisfying $\text{Rank}_{\text{PSD}}(\mathbf{P}) \leq d$. Therefore, the positive semidefinite rank of a behavior quantifies the amount of unassisted *quantum* communication needed to simulate the behavior. Thus if a quantum signaling system exhibits a behavior $\mathbf{P}$ that has $\text{Rank}_{\text{PSD}}(\mathbf{P}) = k$, then $k$ lower bounds the dimension of the Hilbert space of the communicated quantum state.

As a result, to witness nonclassicality in unassisted quantum signaling behavior $\mathbf{P} \in \mathcal{Q}^{\mathcal{X}\xrightarrow{d}\mathcal{Y}}$, it is sufficient

to show that classical signaling dimension is bounded as $\kappa(\mathbf{P}) > d$ because

$$\kappa(\mathbf{P}) = \text{Rank}_{+}(\mathbf{P}) > \text{Rank}_{\text{PSD}}(\mathbf{P}) = d \tag{14.26}$$

where $\text{Rank}_{\text{PSD}}(\mathbf{P}) = d$ follows from the fact that the Hilbert space $\mathcal{H}_d$ is used for signaling. The inequality in Eq. (14.26) follows directly from Eq. (14.13), in which it is stated that the nonnegative rank is no less than the positive semidefinite rank. In the literature, examples have been identified where strict separations between these two ranks exist [123], indicating the presence of operational nonclassicality in unassisted quantum signaling scenarios. In general, such quantum signaling behaviors are regarded as being weakly nonclassical because Lemma 13 shows that the set of quantum signaling behaviors is contained by the set of shared randomness-assisted classical behaviors such that $\mathcal{Q}^{\mathcal{X}\overset{d}{\to}\mathcal{Y}} \subseteq \mathcal{C}_{\Lambda}^{\mathcal{X}\overset{d}{\to}\mathcal{Y}}$.

### *14.2.2 Linear Black-Box Games as Signaling Dimension Witnesses*

The signaling dimension $\kappa_{\Lambda}(\mathbf{P})$ is a device-independent quantity that can be ascertained from any black-box behavior. The goal is to find the smallest positive integer $d$ that satisfies $\mathbf{P} \in \mathcal{C}_{\Lambda}^{\mathcal{X}\overset{d}{\to}\mathcal{Y}}$. To develop a device-independent test for signaling dimension, we can exploit the structure of general signaling polytopes $\mathcal{C}_{\Lambda}^{\mathcal{X}\overset{d}{\to}\mathcal{Y}}$. The polytope is convex and exhibits a permutation symmetry on its input and output alphabets $\mathcal{X}$ and $\mathcal{Y}$. Therefore, a signaling polytope $\mathcal{C}_{\Lambda}^{\mathcal{X}\overset{d}{\to}\mathcal{Y}}$ is bound by a set of linear facet inequalities $\mathcal{G}^{\mathcal{X}\overset{d}{\to}\mathcal{Y}}$ that tightly bound $\mathcal{C}_{\Lambda}^{\mathcal{X}\overset{d}{\to}\mathcal{Y}}$. These linear facet inequalities can then test whether or not a behavior $\mathbf{P}$ belongs to $\mathcal{C}_{\Lambda}^{\mathcal{X}\overset{d}{\to}\mathcal{Y}}$. This approach is taken by Dall'Arno *et al.* to demonstrate using a toy physical model to violate the No-Hypersignaling Principle [122]. Moreover, these linear witnesses are similar to those applied in classical and quantum dimensionality witnessing [227], [228], [345]–[347].

**Definition 93. Signaling Dimension Witness:** $(\gamma, \mathbf{G}) \in \mathcal{G}^{\mathcal{X}\overset{d}{\to}\mathcal{Y}}$, A linear inequality $\langle \mathbf{G}, \mathbf{P} \rangle \leq \gamma$ that satisfies $\mathcal{C}_{\Lambda}^{\mathcal{X}\overset{d}{\to}\mathcal{Y}} \subset \{\mathbf{P} \in \mathcal{P}_{\mathcal{Y}|\mathcal{X}} \mid \langle \mathbf{G}, \mathbf{P} \rangle \leq \gamma\}$. A signaling dimension witness is "tight" if the equality $\langle \mathbf{G}, \mathbf{P} \rangle = \gamma$ is satisfied by $\text{Dim}\left(\mathcal{C}_{\Lambda}^{\mathcal{X}\overset{d}{\to}\mathcal{Y}}\right) = |\mathcal{X}|(|\mathcal{Y}| - 1)$ affinely independent vertices. In other words, a tight signaling dimension witness describes a facet of $\mathcal{C}_{\Lambda}^{\mathcal{X}\overset{d}{\to}\mathcal{Y}}$.

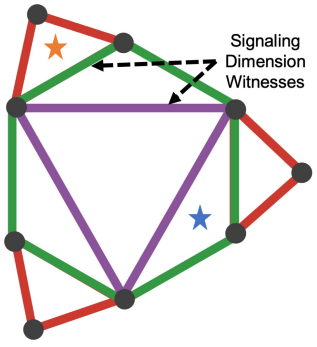



Figure 14.2: A graphical analogy illustrating key features of signaling polytopes. Note that the geometry has been simplified by removing vertices and and projecting the polytope onto a plane. The example shows a hierarchy of three signaling polytopes $\mathcal{C}^{\mathcal{X}\overset{d}{\to}\mathcal{Y}}$ where $d \in [1, 3]$ and $|\mathcal{X}| = |\mathcal{Y}| = 3$. Each signaling polytope is a convex set where the hierarchy $\mathcal{C}_{\Lambda}^{\mathcal{X}\overset{d=1}{\to}\mathcal{Y}} \subset \mathcal{C}_{\Lambda}^{\mathcal{X}\overset{d=2}{\to}\mathcal{Y}} \subset \mathcal{C}_{\Lambda}^{\mathcal{X}\overset{d=3}{\to}\mathcal{Y}}$ holds for any $|\mathcal{X}|$ and $|\mathcal{Y}|$. The vertices of each polytope are depicted as circles while the facets are lines. The $\mathcal{C}_{\Lambda}^{\mathcal{X}\overset{d=1}{\to}\mathcal{Y}}$ polytope is blue, the $\mathcal{C}_{\Lambda}^{\mathcal{X}\overset{d=2}{\to}\mathcal{Y}}$ polytope is green, and the $\mathcal{C}_{\Lambda}^{\mathcal{X}\overset{d=3}{\to}\mathcal{Y}} = \mathcal{P}_{\mathcal{Y}|\mathcal{X}}$ polytope is red. A facet of the signaling polytope constitutes a signaling dimension witness. The behavior $\mathbf{P}_{Orange}$ (orange star) has $\kappa_{\Lambda}(\mathbf{P}_{Orange}) = 3$ while the behavior $\mathbf{P}_{Blue}$ (blue star) has $\kappa_{\Lambda}(\mathbf{P}_{Blue}) = 2$.

If a behavior $\mathbf{P} \in \mathcal{P}_{\mathcal{Y}|\mathcal{X}}$ violates a signaling dimension witness as $\langle \mathbf{G}, \mathbf{P} \rangle > \gamma$, then $\mathbf{P} \notin \mathcal{C}_{\Lambda}^{\mathcal{X} \overset{d}{\to} \mathcal{Y}}$ and a lower bound $\kappa_{\Lambda}(\mathbf{P}) > d$ is placed on the signaling dimension. Alternatively, if $\mathbf{P}$ satisfies all signaling dimension witnesses belonging to the complete set of signaling polytope facets $\mathcal{G}^{\mathcal{X} \overset{d}{\to} \mathcal{Y}}$, then $\mathbf{P} \in \mathcal{C}_{\Lambda}^{\mathcal{X} \overset{d}{\to} \mathcal{Y}}$ and the upper bound $\kappa_{\Lambda}(\mathbf{P}) \leq d$ is placed on the signaling dimension. It then becomes our objective to procure signaling dimension witnesses that can be applied to certify the signaling dimension in practice.

In addition to certifying the classical simulation cost, signaling dimension witnesses can also be used to witness strong nonclasasicality in signaling scenario. That is, if the amount of classical or quantum communication is known for a given signaling system, a violation of a signaling dimension witness would observe a behavior that cannot be reproduced classically. Since signaling dimension witnesses allow for an unlimited amount of shared randomness between devices, such violations would observe strong nonclassicality in signaling scenarios. Although it is known that $\mathcal{Q}^{\mathcal{X} \overset{d}{\to} \mathcal{Y}} \subseteq \mathcal{C}_{\Lambda}^{\mathcal{X} \overset{d}{\to} \mathcal{Y}}$, holds generally as given by Lemma 13, violations to signaling dimension witnesses have been observed when the sender and receiver share entanglement [116]. These violations correspond an increase in the classical simulation cost of entanglement-assisted signaling scenarios, an observation made previously by Cubitt *et al.* [350], in which the zero-error simulation using entanglement-assisted communication is considered. However, the extent to which such violations occur in quantum systems has not been investigated thoroughly.

## *14.3 Signaling Dimension Witnesses*

In this section we derive the signaling dimension witnesses that bound signaling polytopes. The signaling dimension is introduced in Section 14.2 as a device-independent quantifier of the amount communication. Since signaling polytopes are invariant under the relabelling of inputs and outputs, all discussed inequalities describe a family of inequalities where each member of the family is obtained by a distinct relabelling of the inputs and/or outputs, *i.e.*, the rows and columns of the behavior matrix are permuted. To obtain signaling polytope facets, we follow the standard procedure where we first enumerate the vertices and then, use a polytope transformation technique such as Fourier-Motzkin elimination to derive the facets [118]. Software such as PORTA [135], [136] assists in this computation, but the large number of vertices leads to impractical run times. To improve efficiency, we utilize the adjacency decomposition technique which heavily exploits the permutation symmetry of signaling polytopes [128] (see Section 4.6 of Chapter 4). Our software and computed facets are publicly available on GitHub [349] while a catalog of general tight signaling dimension witnesses are provided in this section.

In general, each signaling dimension witness corresponds to a linear black-box game $(\gamma, \mathbf{G})$ played by the signaling system. For an overview of black-box games, please refer to Section 2.3 of Chapter 2. In signaling scenarios, the black-box game is viewed as a cooperative game played by the sender and receiver. Here the referee, or researcher, gives the sender an input $x \in \mathcal{X}$ drawn with uniform probability $1/|\mathcal{X}|$. The sender encodes $x$ into a message that is transmitted to the receiver and then decoded to produce the output $y \in \mathcal{Y}$ with probability $P(y|x)$. The reward $G_{y,x} \in \mathbb{R}$ then scales the behavior to achieve the score

$$S(\mathbf{P}) = \langle \mathbf{G}, \mathbf{P} \rangle = \mathrm{Tr}\left[\mathbf{G}^T \mathbf{P}\right] = \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} G_{y,x} P(y|x). \tag{14.27}$$

Hence a violation of a signaling dimension witness corresponds precisely to achieving a score $S(\mathbf{P}) > \gamma$.

Not all linear inequalities that bound a signaling polytope are facet inequalities. To prove that an inequality $(\gamma, \mathbf{G})$ is facet inequality of $\mathcal{C}_{\Lambda}^{\mathcal{X} \overset{d}{\to} \mathcal{Y}}$, the signaling dimension must be tight as described in Def. 91.

Furthermore, by the permutation symmetry of the signaling polytope, it suffices to use describe a family of signaling dimension witnesses by generator facet. The other members of the family are produced from the generator by taking row and column permutations of the game matrix $\mathbf{G}$. Similarly, facet inequalities can be lifted to signaling polytopes with more inputs and/or outputs (for details, see Section 4.4 of Chapter 4).

### *14.3.1 Ambiguous Guessing Games*

**Definition 94. Ambiguous Guessing Game:** $(d, \mathbf{G}_k^{\mathcal{X}\xrightarrow{d}\mathcal{Y}})$, A signaling dimension witness. For $k \in \{0, \dots |\mathcal{Y}|\}$ and $d \leq \min\{|\mathcal{X}|, |\mathcal{Y}|\}$, let $\mathbf{G}_k^{\mathcal{X}\xrightarrow{d}\mathcal{Y}} \in \mathbb{R}^{|\mathcal{Y}|\times|\mathcal{X}|}$ such that (i) $k$ rows are stochastic with $0/1$ elements, and (ii) the remaining $(|\mathcal{Y}| - k)$ rows have the value $1/(|\mathcal{X}| - d + 1)$ in each column.

It will be helpful to refer to rows of type (i) as "guessing rows" and rows of type (ii) as "ambiguous rows." For example, if $|\mathcal{X}| = |\mathcal{Y}| = 6$, $k = 5$, and $d = 2$, then up to a permutation of rows and columns we have

$$\mathbf{G}_k^{\mathcal{X}\xrightarrow{d}\mathcal{Y}} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ \frac{1}{5} & \frac{1}{5} & \frac{1}{5} & \frac{1}{5} & \frac{1}{5} & \frac{1}{5} \end{bmatrix}. \tag{14.28}$$

Then, for any channel $\mathbf{P} \in \mathcal{C}_\Lambda^{\mathcal{X}\xrightarrow{d}\mathcal{Y}}$, the inequality

$$\langle \mathbf{G}_k^{\mathcal{X}\xrightarrow{d}\mathcal{Y}}, \mathbf{P} \rangle \leq d \tag{14.29}$$

is satisfied. To prove this bound holds in general, suppose without loss of generality that the first $k$ rows of $\mathbf{G}_k^{\mathcal{X}\xrightarrow{d}\mathcal{Y}}$ are guessing rows. Then, let $\mathbf{V}$ be any vertex of $\mathcal{C}_\Lambda^{\mathcal{X}\xrightarrow{d}\mathcal{Y}}$ where $t$ of its first $k$ rows are nonzero. If $t = d$, then clearly Eq. (14.29) holds. Otherwise, if $t < d$, then $\langle \mathbf{G}_k^{\mathcal{X}\xrightarrow{d}\mathcal{Y}}, \mathbf{V} \rangle \leq t + (|\mathcal{X}| - t)/(|\mathcal{X}| - d + 1) \leq d$, where the last inequality follows after some algebraic manipulation.

Equation (14.29) can be interpreted as the score of cooperative game played between the sender and receiver. In this game, the $y^{th}$ guessing row corresponds to an answer $y$ that is correct for only a single input $x$. If the receiver correctly outputs $y$ when the sender is given the input $x$ awards 1 point, whereas an incorrect guess awards 0 points. On the other hand, if the $y^{th}$ row is an ambiguous row, then it corresponds to an answer $y$ that is correct for all inputs $x \in \mathcal{X}$. That is, if the receiver outputs a value $y$ that indexes an ambiguous row, then the associated reward is $1/(|\mathcal{X}| - d + 1)$ for any $x \in \mathcal{X}$. Note that the action of guessing an ambiguous row can be interpreted as the receiver asserting that the message is ambiguous or indistinguishable, meaning that the receiver is trying to mitigate risk by taking a small reward with certainty instead guessing the correct answer with small probability.

**Definition 95. ($k$,$d$)-Ambiguous Polytope:** $\mathcal{A}_k^{\mathcal{X}\xrightarrow{d}\mathcal{Y}} \subset \mathcal{P}_{\mathcal{Y}|\mathcal{X}}$, The collection of all behaviors $\mathbf{P} \in \mathcal{P}_{\mathcal{Y}|\mathcal{X}}$ that satisfy Eq. (14.29) for all $\mathbf{G}_k^{\mathcal{X}\xrightarrow{d}\mathcal{Y}}$ where $k$, $d$, $\mathcal{X}$, and $\mathcal{Y}$ are fixed.

Naturally, $\mathcal{C}_\Lambda^{\mathcal{X}\xrightarrow{d}\mathcal{Y}} \subseteq \mathcal{A}_k^{\mathcal{X}\xrightarrow{d}\mathcal{Y}}$ for all $k \in \{0, \dots, |\mathcal{Y}|\}$, therefore, if $\mathbf{P} \notin \mathcal{A}_k^{\mathcal{X}\xrightarrow{d}\mathcal{Y}}$ for some $k$, then $\mathbf{P} \notin \mathcal{C}_\Lambda^{\mathcal{X}\xrightarrow{d}\mathcal{Y}}$. This observation is convenient when it comes to deciding whether or not $\mathbf{P} \in \mathcal{A}_k^{\mathcal{X}\xrightarrow{d}\mathcal{Y}}$.

**Proposition 16.** A behavior $\mathbf{P} \in \mathcal{P}_{\mathcal{Y}|\mathcal{X}}$ belongs to $\mathcal{A}_k^{\mathcal{X}\overset{d}{\to}\mathcal{Y}}$ if and only if

$$\max_{\pi\in\mathcal{S}^{\mathcal{Y}}} \sum_{i=1}^{k} \|\mathbf{r}_{\pi(i)}\|_\infty + \frac{1}{|\mathcal{X}|-d+1} \sum_{i=k+1}^{|\mathcal{Y}|} \|\mathbf{r}_{\pi(i)}\|_1 \leq d, \tag{14.30}$$

where the maximization is taken over $\mathcal{S}^{\mathcal{Y}}$,the set of all permutations on $\mathcal{Y}$, $\mathbf{r}_i$ denotes the $i^{th}$ row of $\mathbf{P}$, $\|\mathbf{r}_i\|_\infty$ is the largest element in the row $\mathbf{r}_i$, and $\|\mathbf{r}_i\|_1$ is the row sum of $\mathbf{r}_i$.

The maximization on the LHS of Eq. (14.30) is performed efficiently using the following procedure. For each row $\mathbf{r}_i$ we assign a pair $(a_i, b_i)$ where $a_i = \|\mathbf{r}_i\|_\infty$ and $b_i = \frac{1}{|\mathcal{X}|-d+1}\|\mathbf{r}_i\|_1$. Define $\delta i = a_i - b_i$, and relabel the rows of $\mathbf{P}$ in non-increasing order of the $\delta_i$. Then according to this sorting, we have an ambiguous guessing game score of $\sum_{i=1}^{k} a_i + \sum_{i=k+1}^{|\mathcal{Y}|} b_i$, which we claim attains the maximum on the LHS of Eq. (14.30). Indeed, for any other row permutation $\pi$, the guessing game score is given by

$$\sum_{\substack{i\in\{1,\cdots,k\}\\ \pi(i)\in\{1,\cdots,k\}}} a_i + \sum_{\substack{i\in\{1,\cdots,k\}\\ \pi(i)\in\{k+1,\cdots,|\mathcal{Y}|\}}} b_i + \sum_{\substack{i\in\{k+1,\cdots,|\mathcal{Y}|\}\\ \pi(i)\in\{1,\cdots,k\}}} a_i + \sum_{\substack{i\in\{k+1,\cdots,|\mathcal{Y}|\}\\ \pi(i)\in\{k+1,\cdots,|\mathcal{Y}|\}}} b_i. \tag{14.31}$$

Hence the difference in these two scores is

$$\sum_{\substack{i\in\{1,\cdots,k\}\\ \pi(i)\in\{k+1,\cdots,|\mathcal{Y}|\}}} (a_i - b_i) - \sum_{\substack{i\in\{k+1,\cdots,|\mathcal{Y}|\}\\ \pi(i)\in\{1,\cdots,k\}}} (a_i - b_i) \geq 0, \tag{14.32}$$

where the inequality follows from the fact that we have ordered the indices in non-increasing order of $(a_i - b_i)$, and the number of terms in each summation is the same since $\pi$ is a bijection.

A special case of the ambiguous guessing games arises when $k = |\mathcal{Y}|$ such that all rows are guessing rows. Then up to a normalization factor $\frac{1}{n}$, we interpret the LHS of Eq. (14.30) as the success probability when the receiver performs maximum likelihood estimation of the sender's input value $x$ based on the message $m$ (i.e. the receiver guesses the value $y$ that maximizes $P(y = x|x)$).

**Definition 96. Maximum Likelihood (ML) Estimation Polytope:** $\mathcal{M}^{\mathcal{X}\overset{d}{\to}\mathcal{Y}} = \mathcal{A}_{|\mathcal{Y}|}^{\mathcal{X}\overset{d}{\to}\mathcal{Y}}$.

Using Proposition 16 we find that

$$\mathbf{P} \in \mathcal{M}^{\mathcal{X}\overset{d}{\to}\mathcal{Y}} \iff \sum_{y=\in\mathcal{Y}} \max_{x\in\mathcal{X}} P(y|x) \leq d. \tag{14.33}$$

In general, the ambiguous guessing signaling dimension witnesses of Eq. (14.29) are not tight for a signaling polytope $\mathcal{C}_\Lambda^{\mathcal{X}\overset{d}{\to}\mathcal{Y}}$. For instance, $\langle \mathbf{G}_k^{\mathcal{X}\overset{d}{\to}\mathcal{Y}}, \mathbf{P}\rangle \leq d$ is trivially satisfied whenever $k = 0$. Nevertheless, in many cases we can establish tightness of these inequalities.

### *14.3.2 Fully Characterized Signaling Polytopes*

In general, it is not feasible to identify the complete set of tight signaling dimension witnesses that bound each signaling polytope $\mathcal{C}_\Lambda^{\mathcal{X}\overset{d}{\to}\mathcal{Y}}$. However, we analytically solve the problem in special cases.

**Theorem 10.** Consider the signaling polytope $\mathcal{C}_\Lambda^{\mathcal{X}\overset{d}{\to}\mathcal{Y}}$:

(i) If $d = |\mathcal{Y}| - 1$, then $\mathcal{C}_\Lambda^{\mathcal{X}\overset{d}{\to}\mathcal{Y}} = \mathcal{M}^{\mathcal{X}\overset{d}{\to}\mathcal{Y}}$.

(a) $2 \geq \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \end{bmatrix}$ (b) $2 \geq \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \end{bmatrix}$

(c) $3 \geq \begin{bmatrix} 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \end{bmatrix}$ (d) $4 \geq \begin{bmatrix} 2 & 0 & 0 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 2 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 & 0 \end{bmatrix}$

(e) $4 \geq \begin{bmatrix} 2 & 0 & 0 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 & 0 \end{bmatrix}$ (f) $4 \geq \begin{bmatrix} 2 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 0 \\ 1 & 1 & 1 & 0 & 0 & 0 \end{bmatrix}$

(g) $4 \geq \begin{bmatrix} 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 \\ 1 & 1 & 1 & 0 & 0 & 0 \end{bmatrix}$ (h) $5 \geq \begin{bmatrix} 1 & 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 & 1 & 1 \end{bmatrix}$

Table 14.1: Generator facets for the signaling polytope $\mathcal{C}_{\Lambda}^{\mathcal{X} \xrightarrow{d} \mathcal{Y}}$ where $d = 2$, $|\mathcal{X}| = 6$, and $|\mathcal{Y}| = 4$. Each inequality is expressed as $\gamma \geq \mathbf{G}$ where the inner product $\langle \mathbf{G}, \mathbf{P} \rangle$ is implied. (a) ML facet input/output lifted from the $d = 2$ and $|\mathcal{X}| = |\mathcal{Y}| = 3$ signaling polytope. (b) ML facet output lifted from the $d = 2$ and $|\mathcal{X}| = |\mathcal{Y}| = 4$ signaling polytope. (c) Anti-guessing facet output lifted from the $d = 2$ and $|\mathcal{X}| = |\mathcal{Y}| = 4$ signaling polytope. (d) Ambiguous guessing facet output lifted from the $d = 2$, $|\mathcal{X}| = 3$, and $|\mathcal{Y}| = 4$ signaling polytope. (e-g) Rescalings of the $d = 2$, $|\mathcal{X}| = 3$, and $|\mathcal{Y}| = 4$ambiguous guessing facet output lifted to $\mathcal{C}_2^{6 \to 4}$. General forms of these tight signaling dimension witnesses are derived in Appendix 14.3. (h) $k$-guessing facet of the $d = 2$, $|\mathcal{X}| = 6$, and $|\mathcal{Y}| = 4$ signaling polytope.

(ii) If $d = |\mathcal{X}| - 1$, then $\mathcal{C}_{\Lambda}^{\mathcal{X} \xrightarrow{d} \mathcal{Y}} = \bigcap_{k=|\mathcal{X}|}^{|\mathcal{Y}|} \mathcal{A}_k^{\mathcal{X} \xrightarrow{d} \mathcal{Y}}$.

In other words, to decide whether or not a channel can be simulated by an amount of classical messages strictly less than the input/output alphabets, it suffices to consider the ambiguous guessing games. Moreover, by Eq. (14.30) it is simple to check if these conditions are satisfied for a given channel $\mathbf{P}$.

*Proof.* For a complete proof, refer to Appendix A.1 or our work in Ref. [114]. □

We also fully characterize the family of signaling polytopes where $d = 2$, $|\mathcal{Y}| = 4$, and $|\mathcal{X}|$ is arbitrary. This result can be used to understand the classical simulation cost of performing Bell measurements on a two-qubit system, since this process induces a classical channel with four outputs. Similarly, this scenario can be used to witness strong nonclassicality in entanglement-assisted signaling scenarios using either a bit or qubit of communication.

**Theorem 11.** For any integer $|\mathcal{X}|$, $d = 2$, and $|\mathcal{Y}| = 4$ a behavior $\mathbf{P} \in \mathcal{P}_{\mathcal{Y}|\mathcal{X}}$ belongs to $\mathcal{C}_{\Lambda}^{\mathcal{X} \xrightarrow{d} \mathcal{Y}}$ if and only if it satisfies the eight signaling dimension witnesses depicted in Fig. 14.1 and all their input/output liftings and permutations.

*Proof.* For a complete proof, please see Appendix A.2.3 or our work in Ref. [114]. □

Remarkably, the result in Theorem 11 shows that when $|\mathcal{Y}| = 4$ and $d = 2$, no new facet classes for $\mathcal{C}_{\Lambda}^{\mathcal{X} \xrightarrow{d} \mathcal{Y}}$ are found when $|\mathcal{X}| > 6$. Consequently, to demonstrate that a behavior $\mathbf{P} \in \mathcal{P}^{n \to 4}$ requires more than one bit

for simulation, it suffices to consider input sets of size no greater than six. For $|\mathcal{X}| < 6$, the signaling polytope facet classes are given by the facets in Table 14.1 having $(6 - |\mathcal{X}|)$ all-zero columns. We conjecture that in general, no more than $|\mathcal{X}| = \binom{|\mathcal{Y}|}{d}$ inputs are needed to certify that a behavior $\mathbf{P} \in \mathcal{P}_{\mathcal{Y}|\mathcal{X}}$ has a signaling dimension larger than $d$.

### *14.3.3 Tight Signaling Dimension Witnesses for General Signaling Polytopes*

In this section, we generalize the facets in Table 14.1 to broad families of signaling polytopes. For each inequality, we describe the black-box game and the conditions for which the inequality is tight. Proof that each inequality is tight can be found in Appendix B or in our work [114].

It is important to note that we describe the generator form for each inequality $(\gamma, \mathbf{G})$. By the permutation symmetry of the signaling polytope, each input and output permutation of $\mathbf{G}$ must also be a tight signaling dimension witness. Likewise, the tight signaling dimension witnesses can be lifted to signaling scenarios having more inputs and outputs. Therefore, the example generator facet inequalities that we list below can be extended to nearly all signaling polytopes. This fact is important because it enables the use of adjacency decomposition for any signaling polytope because a seed facet can always be obtained.

### *14.3.4 Maximum Likelihood Signaling Polytope Facets*

**Definition 97. Maximum Likelihood Facet:** $(d, \mathbf{G}_{\text{ML}}) \in \mathcal{G}^{\mathcal{X} \xrightarrow{d} \mathcal{Y}}$, Let $|\mathcal{Y}| = |\mathcal{X}|$ and consider the game $\mathbf{G}_{\text{ML}}^{|\mathcal{Y}|} = \mathbb{I}_{|\mathcal{Y}|}$, the $|\mathcal{Y}| \times |\mathcal{Y}|$ identity matrix. For all vertices $\mathbf{V} \in \mathcal{V}^{\mathcal{X} \xrightarrow{d} \mathcal{Y}}$ the inequality

$$\langle \mathbf{G}_{\text{ML}}^{|\mathcal{Y}|}, \mathbf{V} \rangle \leq d, \tag{14.34}$$

is satisfied because $\text{Rank}\,(\mathbf{V}) \leq d$ and both $\mathbf{V}$ and $\mathbf{G}_{\text{ML}}^{|\mathcal{Y}|}$ are doubly stochastic. The maximum likelihood game is a tight signaling dimension witness whenever and $|\mathcal{Y}| > d > 1$. For proof please refer to Proposition 23 in Appendix B, or our work [114].

### *14.3.5 Ambiguous Guessing Signaling Polytope Facets*

**Definition 98. Ambiguous Guessing Facet:** $(\gamma_?^{|\mathcal{Y}|,d}, \mathbf{G}_?^{|\mathcal{Y}|,d}) \in \mathcal{G}^{\mathcal{X} \xrightarrow{d} \mathcal{Y}}$, Let $|\mathcal{X}| = |\mathcal{Y}| - 1$ and consider the game where

$$\mathbf{G}_?^{|\mathcal{Y}|,d} = \sum_{x \in \mathcal{X}} (|\mathcal{Y}| - d)\, |x\rangle\langle x| + ||\mathcal{Y}|\rangle\langle x|\,, \quad \text{and} \quad \gamma_?^{|\mathcal{Y}|,d} = d(|\mathcal{Y}| - d). \tag{14.35}$$

The ambiguous guessing game is tight signaling dimension witness whenever $|\mathcal{Y}| - 1 > d > 1$. For proof see Proposition 24 in Appendix B.

This signaling dimension witness can be interpreted as a combination between a maximum likelihood game for which a correct answer provides $(|\mathcal{Y}| - d)$ points extended by one ambiguous row for which 1 point

is scored for any input $x \in \mathcal{X}$. For example, when $|\mathcal{Y}| = 6$ and $d = 2$ we have

$$\mathbf{G}_{?}^{|\mathcal{Y}|,d} = \begin{bmatrix} (|\mathcal{Y}| - d)\mathbf{G}_{\text{ML}}^{(|\mathcal{Y}|-1)} \\ \vec{1} \end{bmatrix}, \quad \text{e.g.} \quad \mathbf{G}_{?}^{6,2} = \left[\begin{array}{ccccc} 4 & 0 & 0 & 0 & 0 \\ 0 & 4 & 0 & 0 & 0 \\ 0 & 0 & 4 & 0 & 0 \\ 0 & 0 & 0 & 4 & 0 \\ 0 & 0 & 0 & 0 & 4 \\ \hline 1 & 1 & 1 & 1 & 1 \end{array}\right], \tag{14.36}$$

where we refer to rows of the $\mathbf{G}_{\text{ML}}^{|\mathcal{Y}|-1)}$ block as *guessing rows* and $\vec{1}$ is a row vector of ones which we refer to as the *ambiguous row*. Note that $\mathbf{G}_{?}^{|\mathcal{Y}|,d}$ is a special case of the ambiguous guessing game $\mathbf{G}_{k}^{\mathcal{X}\xrightarrow{d}\mathcal{Y}}$, and without loss of generality, we express $\mathbf{G}_{?}^{|\mathcal{Y}|,d}$ in a normal form where all elements $G_{y,x}$ are non-negative integers. For any vertex $\mathbf{V} \in \mathcal{V}^{\mathcal{X}\xrightarrow{d}\mathcal{Y}}$, the inequality

$$\langle \mathbf{G}_{?}^{|\mathcal{Y}|,d}, \mathbf{V} \rangle \leq d(|\mathcal{Y}| - d) \tag{14.37}$$

is satisfied

An ambiguous guessing facet $(\gamma_{?}^{|\mathcal{Y}|,d}, \mathbf{G}_{?}^{|\mathcal{Y}|,d})$ exhibits nontrivial input lifting to $\mathbf{G}_{?}^{\prime|\mathcal{Y}|,2} \in \mathbb{R}^{|\mathcal{Y}|\times(|\mathcal{X}|+1)}$ by taking a guessing row $y$ where $G_{y,x} = (|\mathcal{Y}| - d)$ distributing the value between between two columns such that $G'_{y,x} = 1$ and $G'_{y,x'} = (|\mathcal{Y}| - d) - 1$ where $x' = |\mathcal{X}| + 1$ is a new column (see Table 14.1.f-h). The bound of the input-lifted facet is then the same as the unlifted version. For example, when $|\mathcal{Y}| = 5$ and $d = 2$, the $\mathbf{G}_{?}^{5,2}$ is rescaled along the 4th row as,

$$\mathbf{G}_{?}^{5,2} = \begin{bmatrix} 3 & 0 & 0 & 0 \\ 0 & 3 & 0 & 0 \\ 0 & 0 & 3 & 0 \\ 0 & 0 & 0 & 3 \\ 1 & 1 & 1 & 1 \end{bmatrix} \rightarrow \mathbf{G}_{?}^{\prime 5,2} = \begin{bmatrix} 3 & 0 & 0 & 0 & 0 \\ 0 & 3 & 0 & 0 & 0 \\ 0 & 0 & 3 & 0 & 0 \\ 0 & 0 & 0 & 1 & 2 \\ 1 & 1 & 1 & 1 & 0 \end{bmatrix}. \tag{14.38}$$

We observe this type of input lifting to be a general trend observed in our computed signaling polytope facets [349], however, it is not quite clear how this lifting rule generalizes.

### *14.3.6 k-Guessing Signaling Polytope Facets*

Consider a guessing game where for each input $x \in \mathcal{X}$ there are $k$ correct answers out of $|\mathcal{Y}|$ possible answers. In this game, the sender has $|\mathcal{X}| = \binom{|\mathcal{Y}|}{k}$ inputs where each value $x$ corresponds to a unique set of $k$ correct answers. Given an input $x \in \mathcal{X}$, the sender signals to the receiver using a message $m \in \mathcal{M}$ where $|\mathcal{M}| = d$ and Bob makes a guess $y \in \mathcal{Y}$. A correct guess scores 1 point while an incorrect guess scores 0 points. This type of guessing game is described by Heinosaari *et al.* [123], [348] and used to test the communication performance of a particular theory, as well as by Frenkel and Weiner [116] who used it to demonstrate a violation of signaling dimension when a bit of communication is assisted by entanglement and $|\mathcal{Y}| = 4$.

**Definition 99.** *k*-**Guessing Game:** $(\gamma_{\text{K}}^{|\mathcal{Y}|,k,d}, \mathbf{G}_{\text{K}}^{|\mathcal{Y}|,k}, ) \in \mathcal{G}^{\mathcal{X}\xrightarrow{d}\mathcal{Y}}$, A linear signaling dimension witness where

$$\gamma_{\text{K}}^{|\mathcal{Y}|,k,d} = \binom{|\mathcal{Y}|}{k} - \binom{|\mathcal{Y}| - d}{k} \tag{14.39}$$

and $\mathbf{G}_{\mathrm{K}}^{|\mathcal{Y}|,k} \in \mathbb{R}^{|\mathcal{Y}| \times \binom{|\mathcal{Y}|}{k}}$ is a matrix with each column containing a unique arrangement of $k$ unit elements and $(|\mathcal{Y}| - k)$ null elements. For example,

$$\mathbf{G}_{\mathrm{K}}^{6,2} = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 1 & 1 \end{bmatrix}. \tag{14.40}$$

The $k$-guessing game is tight signaling dimension witness whenever $|\mathcal{X}| = \binom{|\mathcal{Y}|}{k}$, $|\mathcal{Y}| - 2 \geq k \geq 1$, and $d = |\mathcal{Y}| - k$. For proof see Proposition 25 in Appendix B.

Note that the maximum likelihood signaling dimension witness $(d, \mathbf{G}_{\mathrm{ML}}^{|\mathcal{Y}|})$ is a $(k = 1)$-guessing game where $\mathbf{G}_{\mathrm{ML}}^{|\mathcal{Y}|} = \mathbf{G}_{\mathrm{K}}^{|\mathcal{Y}|,1}$. This general signaling dimension witness was identified by Frenkel and Weiner [115], who showed that given a channel $\mathbf{P} \in \mathcal{C}_{\Lambda}^{\mathcal{X} \xrightarrow{d} \mathcal{Y}}$, the $k$-guessing game score is bounded as

$$\gamma_{\mathrm{K}}^{|\mathcal{Y}|,k,d} = \binom{|\mathcal{Y}|}{k} - \binom{|\mathcal{Y}| - d}{k} \geq \langle \mathbf{G}_{\mathrm{K}}^{|\mathcal{Y}|,k}, \mathbf{P} \rangle \geq \binom{|\mathcal{Y}| - d}{|\mathcal{Y}| - k}. \tag{14.41}$$

### *14.3.7 Anti-Guessing Facets*

A special case of the $k$-guessing game is the *anti-guessing game*, $(|\mathcal{Y}|, \mathbf{G}_{\mathrm{A}}^{|\mathcal{Y}|})$, where $\mathbf{G}_{\mathrm{A}}^{|\mathcal{Y}|} = \mathbf{G}_{\mathrm{K}}^{|\mathcal{Y}|,(|\mathcal{Y}|-1)}$. However, for any behavior $\mathbf{P} \in \mathcal{P}_{\mathcal{Y}|\mathcal{X}}$ that has $|\mathcal{X}| = |\mathcal{Y}|$, the anti-guessing game inequality $\langle \mathbf{G}_{\mathrm{A}}^{|\mathcal{Y}|}, \mathbf{P} \rangle \leq |\mathcal{Y}|$ is satisfied. Therefore, anti-guessing games are not very useful for witnessing signaling dimension. That said, the anti-guessing game is significant because it can be combined with a maximum likelihood game in block form to construct a facet of the $d = (|\mathcal{Y}| - 2)$ signaling polytope $\mathcal{C}_{\Lambda}^{\mathcal{X} \xrightarrow{d} \mathcal{Y}}$.

**Definition 100. Anti-Guessing Facet:** $(\gamma^{\varepsilon,d}, \mathbf{G}_{\mathrm{A}}^{\varepsilon,m'}) \in \mathcal{G}^{\mathcal{X} \xrightarrow{d} \mathcal{Y}}$, A signaling dimension witness where $|\mathcal{X}| = |\mathcal{Y}| = \varepsilon + m'$ and the game matrix is constructed as

$$\mathbf{G}_{\mathrm{A}}^{\varepsilon,m'} = \left[\begin{array}{c|c} \mathbf{G}_{\mathrm{A}}^{\varepsilon} & \hat{0} \\ \hline \hat{0} & \mathbf{G}_{\mathrm{ML}}^{m'} \end{array}\right], \quad e.g. \quad \mathbf{G}_{\mathrm{A}}^{4,2} = \left[\begin{array}{cccc|cc} 1 & 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 1 & 0 & 0 \\ \hline 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{array}\right] \tag{14.42}$$

where $\hat{0}$ is a matrix block of zeros. For any behavior $\mathbf{P} \in \mathcal{C}_{\Lambda}^{\mathcal{X} \xrightarrow{d} \mathcal{Y}}$, the upper bound for the anti-guessing game is

$$\gamma_{\mathrm{A}}^{\varepsilon,d} = \varepsilon + d - 2 \geq \langle \mathbf{G}_{\mathrm{A}}^{\varepsilon,m'}, \mathbf{P} \rangle, \tag{14.43}$$

which follows from the fact that no more than two rows are required to score $\varepsilon$ in the $\mathbf{G}_{\mathrm{A}}^{\varepsilon}$ block and the remaining $d - 2$ rows score one point each against the $\mathbf{G}_{\mathrm{ML}}^{m'}$ block. The anti-guessing inequality is a tight signaling dimension witness whenever $|\mathcal{Y}| - 2 \geq d \geq 2$ and $|\mathcal{Y}| - d + 1 \geq \varepsilon \geq 3$. For proof see Proposition 27 in Appendix B or our work [114].

## *14.4 Certifying Signaling Dimension*

In this section we discuss the how the signaling dimension of classical and quantum channels can be certified. Our approach is practical and can always obtain lower bound on the signaling dimension. Furthermore, when trust is placed on the various devices we find applications in self-testing. We then extend this procedure to quantum channels for which our variational methods and convex optimization techniques can be applied. Remarkably, we find that in the case of general partial replacer channels, the signaling dimension can be certified with a very high-degree of accuracy.

In practice, certification and self-testing tasks reduce to the trust that a researcher has in a device and its resources. That is, the classical or quantum channel from a sender to a receiver may be unknown or not fully characterized. Indeed, the same can be said for sender and receiver devices. In a more extreme case, the devices and channels may be malicious and trying to trick the researcher into believing that they utilize quantum resources when in fact, the devices are classical.

In such scenarios where there are noisy, uncharacterized, or untrustworthy devices, it will still be necessary to ascertain certain properties of the quantum resources at play without having direct access to the physical devices or pre-established trust in their hardware. To perform such tests, a device-independent approach that infers properties of the quantum by analyzing the observed behavior $\mathbf{P}_{\mathcal{Y}|\mathcal{X}}$ obtained as sample averages over many uses of the black-box system, which we assume to be memoryless and uncorrelated with the research who is certifying the system.

To this end, we develop device-independent procedures that certify and bound the signaling dimension of classical and quantum channels. The utility of our tests improves as more trust is placed on the tested devices. Consider first the untrusted black-box signaling system in Fig. 14.3.a). Here the researcher only has access to the behavior $\mathbf{P} \in \mathcal{P}_{\mathcal{Y}|\mathcal{X}}$, however, the researcher can use signaling dimension witnesses to obtain a lower bound on the signaling dimension of the quantum system. Note that without more trust in the devices, the upper bound of $\kappa_\Lambda(\mathbf{P}) \leq \kappa(\mathbf{P}) \leq \min\{|\mathcal{X}|, |\mathcal{Y}|\}$ must hold, but an upper limit on the amount of communication that the devices can perform remains unknown. Furthermore, it cannot be asserted from the behavior alone whether or not quantum resources are used within the black-box signaling system.

In the remaining cases of Fig. 14.3.b-d), one or more of the quantum devices are trusted, meaning that their operations and resources are characterized completely. In these cases, the Hilbert space dimension of either the sender or receiver is known placing an immediate upper bound on the signaling dimension of any produced behavior as $\kappa_\Lambda(\mathbf{P}) \leq \min\{d_A, d_B\}$ as given in Eq. (14.21). Settings in which the dimension of the signaling state is known are typically referred to as semi-device-independent. Such settings are important in practice because they facilitate the use of trusted quantum devices to certify and test the capabilities of untrusted quantum devices.

### *14.4.1 Certifying the Unassisted Signaling Dimension of a Behavior*

To calculate the classical signaling dimension as given in Definition 90.(i) it is sufficient calculate the nonnegative rank of the behavior as $\kappa(\mathbf{P}) = \text{Rank}_+(()\,\mathbf{P})$. However, this task is NP-Hard in general [351], [356]. Nonetheless, we provide an algorithm that can evaluate the nonnegative rank.

In practice, the nonnegative rank can be evaluated by numerically in an iterative algorithm. In each step,

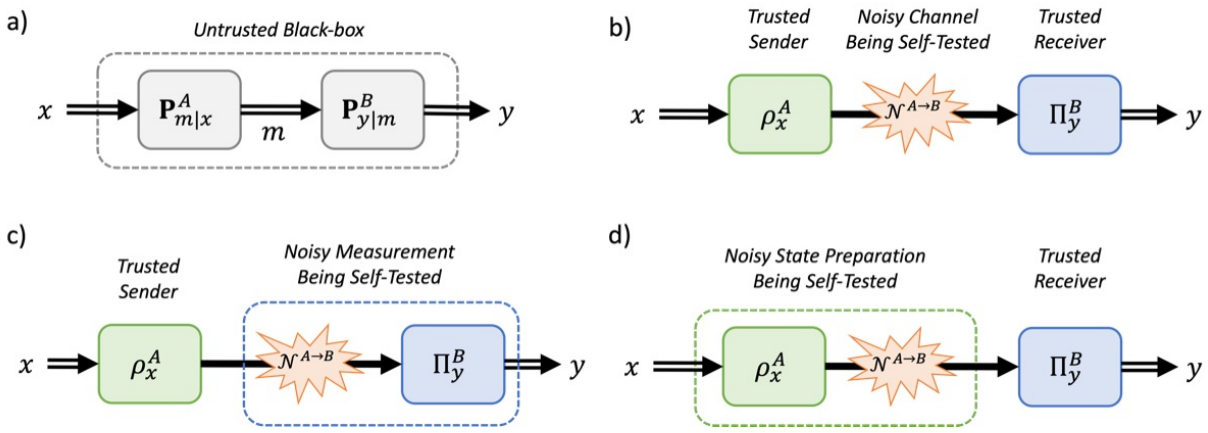



Figure 14.3: A depiction of the required device trust to implement the device-independent certification of the signaling dimension for classical and quantum channels. a) The signaling dimension of an untrusted classical behavior $\mathbf{P}$ is certified using only the probabilities $P(y|x)$ for all $x \in \mathcal{X}$ and $y \in \mathcal{Y}$. b) A trusted sender and receiver device certify the signaling dimension of a noisy quantum channel $\mathcal{N}$. c) A trusted sender device self-tests a noisy measurement applied by a receiver device. d) A trusted receiver device self-tests a noisy state preparation of a sender devices.

a nonnegative matrix factorization $C' = AB$ where $\text{Rank}_+(C') = k$ is optimized as

$$\ell^\star = \min_{A\in\mathbb{R}^{n'\times k}, B\in\mathbb{R}^{k\times n}} \frac{1}{2n}\sum_{i=1}^{n'}\sum_{j=1}^{n}|M_{i,j}| \quad \text{where} \quad M = C - AB \tag{14.44}$$

where $\ell^\star$ is the minimal trace distance between $C$ and $C'$. Note that in nonnegative matrix factorization algorithms more complicated optimization objectives are considered [357]–[359], however, we use the trace distance in Eq. (14.44) to quantify the closeness of the nonnegative matrix factorization to the original matrix. Using the trace distance in Eq. (14.44), if $\ell^\star = 0$, then $C = C'$ and $\text{Rank}_+(C) \leq k$, otherwise $\text{Rank}_+(C) > k$. In practice, the optimization in Eq. (14.44) will have some numerical error, thus a small tolerance $0 < \varepsilon \ll 1$ should be allowed when checking equality, the condition that $\ell^\star \approx 0$ can then be understood as $\ell^\star - \varepsilon \leq 0$ such that $\ell^\star$ falls within the specified tolerance. Thus, a certification procedure of nonnegative rank follows, and by extension, a certification procedure for the unassisted classical signaling dimension $\kappa(\mathbf{P})$.

**Algorithm 2. Certification of $\kappa(\mathbf{P})$ via Nonnegative Rank:** Given a behavior $\mathbf{P} \in \mathcal{P}_{\mathcal{Y}|\mathcal{X}}$, evaluate its nonnegative rank $\text{Rank}_+(\mathbf{P}) = \kappa(\mathbf{P})$. In this algorithm, the nonnegative rank is evaluated iteratively using a binary search over nonnegative matrix factorization $\mathbf{P}' = AB$ where $\text{Rank}_+(\mathbf{P}') = k$. Note that the optimization over nonnegative matrix factorization in Eq. (14.44) can be performed using software tools such as `scikit-learn` [359]. The algorithm proceeds as follows:

1. Use Eq. (14.7) to set lower and upper bounds on the nonnegative rank as the integers $\texttt{lb} := \text{Rank}(\mathbf{P})$ and $\texttt{ub} := \min\{|\mathcal{X}|, |\mathcal{Y}|\}$ and set the integer $k = \texttt{lb} + \lfloor \frac{1}{2}(\texttt{ub} - \texttt{lb}) \rfloor$ such that $\texttt{lb} \leq k \leq \texttt{ub}$.

2. While $\texttt{lb} < \texttt{ub}$ do:

   (a) Use Eq. (14.44) to find an optimal nonnegative matrix factorization $\mathbf{P}' = AB$ that obtains the minimal distance $\ell^\star$ between $\mathbf{P}'$ and $\mathbf{P}$.

   (b) If $\ell^\star - \varepsilon \leq 0$, set $\texttt{ub} := k$ and $k := \texttt{lb} + \lfloor \frac{1}{2}(k - \texttt{lb}) \rfloor$, else, set $\texttt{lb} := k + 1$ and $k := k + \lceil \frac{1}{2}(\texttt{ub} - k) \rceil$.

3. Return $k$ where $\text{Rank}_+(\mathbf{P}) = k$ because $\texttt{lb} = \texttt{ub} = k$.

**Remark.** Since $\text{Rank}(M) = \text{Rank}_+(M)$ when $\text{Rank}(M) = 1$ or $\text{Rank}(M) = 2$ [351], if $\texttt{lb} \leq 2$, then $\text{Rank}_+(\mathbf{P}) = \texttt{lb}$. In this way, the algorithm can short-circuit to improve the performance significantly in special cases.

Since we perform a binary search over $\text{Rank}(\mathbf{P}) \leq \text{Rank}_+(\mathbf{P}') = k \leq \min\{|\mathcal{X}|, |\mathcal{Y}|\}$, the nonnegative rank of the matrix factorization $\mathbf{P}' = AB$, the number of $k$ values to test scales as $O(\log_2(\min\{|\mathcal{X}|, |\mathcal{Y}|\} - \text{Rank}(\mathbf{P})))$. The main challenge in our algorithm is optimizing the nonnegative matrix factorization. While available software tools can compute the nonnegative rank efficiently for many of the examples considered in this work, eventually, large matrix dimensions for nonnegative matrices $A$ and $B$ will limit the performance of these algorithms considerably. Furthermore, our nonnegative matrix factorization algorithm is approximate, susceptible to finding local optima. That is, it may be observed that $\mathbf{P} \not\approx \mathbf{P}' = AB$, although, there exist nonnegative matrices $A'$ and $B'$ having same dimension as $A$ and $B$ such that $\mathbf{P} \approx A'B'$. To mitigate this sort of error, the nonnegative rank factorization can be repeated with randomized initial settings to increase the probability of converging to a global optimum. Finally, the set of matrices having a fixed nonnegative rank is unstable to perturbations. That is, consider a nonnegative matrix $C \in \mathbb{R}^{n' \times n}$ and a small perturbation $C'$ having elements $0 \leq |C'_{i,j}| \ll 1$ such that $C - C'$ is nonnegative. If $\text{Rank}_+(C) = k$, then it is possible that either $\text{Rank}_+(C - C') = k \pm k'$ for some $k' \geq 0$. To accommodate such errors in a practical manner, we apply a numerical tolerance $\ell^\star - \varepsilon \leq 0$, then we can only estimate that $\text{Rank}_+(C) \geq k$. Although, the tolerance $\varepsilon$ should be tuned to the particular task and physical system.

Generally, we must be cautious with the fact that the nonnegative rank computed in Algorithm 2 is approximate. That is, the behavior $\mathbf{P}$ is only approximately equivalent to its computed nonnegative matrix $\mathbf{P}' \approx \mathbf{P}$, meaning that it can only be certified that $\text{Rank}_+(\mathbf{P}') \leq \text{Rank}_+(\mathbf{P})$. Likewise, if a local optimum is found, then Given the limitations of approximate nonnegative rank factorizations, it is important to acknowledge their practical implications. The main advantage being that experimental data from quantum systems is stochastic and fluctuates, meaning that the expected behavior $\mathbf{P}$ and its construction from sampled data $\widetilde{\mathbf{P}}$ are only approximately equal $\mathbf{P} \approx \widetilde{\mathbf{P}}$. In practice, the nonnegative matrix factorization $\mathbf{P}' = AB$ may lie within the expected deviations of the data. Hence, the behavior $\mathbf{P}'$ approximately simulates $\mathbf{P}$, while its nonnegative rank is smaller as $\text{Rank}_+(\mathbf{P}') < \text{Rank}_+(\mathbf{P})$. Indeed, an approximate simulation that resembles experimental data may better characterize the classical simulation cost of the behavior $\mathbf{P}$.

### *14.4.2 Certifying the Signaling Dimension of a Classical Channel*

In this section we outline a certification procedure for the signaling dimension when an unlimited amount of shared randomness is permitted in the corresponding simulation (see Def. 90.(ii). In principle, given a classical behavior $\mathbf{P} \in \mathcal{P}_{\mathcal{Y}|\mathcal{X}}$ the signaling dimension $\kappa_\Lambda(\mathbf{P})$ can be certified incrementally in a similar manner to the nonnegative rank (see Algorithm 2). However, instead of computing the nonnegative rank of the behavior $\mathbf{P}$, it must be verified whether or not the behavior belongs to a given signaling polytope $\mathcal{C}_\Lambda^{\mathcal{X} \xrightarrow{d} \mathcal{Y}}$. This task can be achieved by using signaling dimension witnesses to test for inclusion in or exclusion from a given signaling polytope.

**Algorithm 3. Certification of $\kappa_\Lambda(\mathbf{P})$ via Linear Signaling Dimension Witnesses:** Given a behavior $\mathbf{P} \in \mathcal{P}_{\mathcal{Y}|\mathcal{X}}$, evaluate its signaling dimension $\kappa_\Lambda(\mathbf{P})$. In this algorithm, the signaling dimension is evaluated iteratively using a binary search over the possible signaling dimension $d \in [1, \min\{|\mathcal{X}|, |\mathcal{Y}|\}$. The agorithm proceeds as follows:

1. Use Proposition 16, set the lower bound on the signaling dimension to be

$$\texttt{lb} := \max_{0\leq k\leq |\mathcal{Y}|} \left\lceil \max_{\pi\in\mathcal{S}^{\mathcal{Y}}} \sum_{i=1}^{k} \|\mathbf{r}_{\pi(i)}\|_\infty + \frac{1}{|\mathcal{X}|-d+1} \sum_{i=k+1}^{|\mathcal{Y}|} \|\mathbf{r}_{\pi(i)}\|_1 \right\rceil, \tag{14.45}$$

   which can efficiently be evaluated for each $k$ by first sorting the rows $\mathbf{r}_i$ of $\mathbf{P}$ in non-increasing order by the value $\delta_i = (\|\mathbf{r}_i\|_\infty - \|\mathbf{r}_i\|_1)$ where $\|\mathbf{r}_i\|_\infty$ is the maximum value in each row and $Vert\mathbf{r}_i\|_1$ is the sum of each row.

2. By Theorem 10, if $\texttt{lb} \geq \min\{|\mathcal{X}|, |\mathcal{Y}|\} - 1$, set the upper bound on the signaling dimension as $\texttt{ub} := \texttt{lb}$, else, set $\texttt{ub} := \min\{|\mathcal{X}|, |\mathcal{Y}|\} - 1$.

3. Set the tested signaling dimension as $d := \texttt{lb} + \lfloor \frac{1}{2}(\texttt{ub} - \texttt{lb}) \rfloor$.

4. While $\texttt{lb} < \texttt{ub}$ do:

   (a) For each signaling dimension witness $(\gamma, \mathbf{G}) \in \mathcal{G}^{\mathcal{X}\xrightarrow{d}\mathcal{Y}}$, if $\gamma < \langle \mathbf{G}, \mathbf{P} \rangle$, then set $\texttt{lb} := d + 1$ and $d := d + \lceil \frac{1}{2}(\texttt{ub} - d) \rceil$ and continue to the next iteration of the loop.

   (b) If no signaling dimension witnesses are violated, then set $\texttt{ub} := d$ and $d := \texttt{lb} + \lfloor \frac{1}{2}(d - \texttt{lb}) \rfloor$.

5. Return $d$ where $\kappa_\Lambda(\mathbf{P}) = d = \texttt{lb} = \texttt{ub}$.

The initial upper and lower bounds can be computed with relative efficiency given an behavior matrix $\mathbf{P} \in \mathcal{P}_{\mathcal{Y}|\mathcal{X}}$. Indeed, if $\kappa_\Lambda(\mathbf{P}) \geq \min\{|\mathcal{X}|, |\mathcal{Y}|\} - 1$, then the signaling dimension is evaluated exactly. Unfortunately, computing the complete set of signaling dimension witnesses $\mathcal{G}^{\mathcal{X}\xrightarrow{d}\mathcal{Y}}$ is not generally feasible. In practice, the generalized tight signaling dimension witnesses discussed in Section 14.3 can represent a subset of the complete set of facet inequalities, $\overline{\mathcal{G}}^{\mathcal{X}\xrightarrow{d}\mathcal{Y}} \subset \mathcal{G}^{\mathcal{X}\xrightarrow{d}\mathcal{Y}}$. In this case, the upper bound cannot be improved because not all facet inequalities of the signaling polytope are tested. However, if an inequality $(\overline{\gamma}, \overline{\mathbf{G}}) \in \overline{\mathcal{G}}^{\mathcal{X}\xrightarrow{d}\mathcal{Y}}$ is violated as $\overline{\gamma} < \langle \overline{\mathbf{G}}, \mathbf{P} \rangle$, then the lower bound can be improved beyond the initial bound given by the ambiguous guessing games in Proposition 16.

As a special case, consider the scenario in which $|\mathcal{Y}| = 4$, $1 \leq d \leq 4$, and $|\mathcal{X}| \in \mathbb{N}$. In this setting, Theorem 10 and Theorem 11 give the complete set of facet inequalities for all signaling polytopes $\mathcal{C}_\Lambda^{\mathcal{X}\xrightarrow{d}\mathcal{Y}}$. That is when $d = 2$, the tight signaling dimension witnesses are listed in Table 14.1 and when $d = 3$, the tight signaling dimension witnesses correspond to maximum likelihood games as described in Section 14.3.4.

### *14.4.3 Certifying the Signaling Dimension of a Quantum Channel*

In this section, we extend Algorithm 3 to certifying the signaling dimension of a given quantum channel $\kappa_\Lambda^{\mathcal{X}\to\mathcal{Y}}(\mathcal{N})$ and $\kappa_\Lambda(\mathcal{N})$ (see Def. 91). Although the quantum channel $\mathcal{N} \in \text{CPTP}(\mathcal{H}_{d_A}^A \to \mathcal{H}_{d_B}^B)$ may be untrusted or uncharacterized, the signaling dimension of a quantum channel can be certified provided that both the quantum sender and receiver are be trusted devices (see Fig. 14.3.b). If neither the sender or receiver are trusted to use quantum resources, the device statistics are alone insufficient to decide whether a signaling system is quantum or classical. This restriction results directly from Lemma 13, in which $\mathcal{Q}^{\mathcal{X}\xrightarrow{d}\mathcal{Y}} \subseteq \mathcal{C}_\Lambda^{\mathcal{X}\xrightarrow{d}\mathcal{Y}}$ shows that a classical signaling system can simulate exactly a quantum signaling system [115]. Before describing our certification procedure, we first introduce an important quantity referred to as the *communication value of a quantum channel*.

**Definition 101. The $\mathcal{X} \to \mathcal{Y}$ Communication Value of a Quantum Channel:** [360] For a quantum channel $\mathcal{N} \in \text{CPTP}(\mathcal{H}_{d_A}^A \to \mathcal{H}_{d_B}^B)$ its communication value is

$$\text{cv}^{\mathcal{X}\to\mathcal{Y}}(\mathcal{N}) = \max_{\substack{\{\rho_x^A \in D(\mathcal{H}_{d_A}^A)\}_{x\in\mathcal{X}}, \\ \{\Pi_y^B\}_{y\in\mathcal{Y}} \in \text{POVM}(\mathcal{H}_{d_B}^B)}} \sum_{x=1}^{\min\{|\mathcal{X}|,|\mathcal{Y}|\}} \text{Tr}\left[\Pi_{y=x}^B \mathcal{N}(\rho_x^A)\right] \tag{14.46}$$

where the communication value corresponds to the maximal score for the maximal likelihood game, *e.g.*, $\max_{\mathbf{P}_\mathcal{N} \in \mathcal{Q}^{\mathcal{X}\to\mathcal{Y}}(\mathcal{N})} \left\langle \mathbf{G}_{\text{ML}}^{\min\{|\mathcal{X}|,|\mathcal{Y}|\}}, \mathbf{P}_\mathcal{N} \right\rangle$ (see Section 14.3.4). Since the maximum likelihood facet is a tight signaling polytope facet, $\lceil \text{cv}^{\mathcal{X}\to\mathcal{Y}}(\mathcal{N}) \rceil \leq \kappa_\Lambda^{\mathcal{X}\to\mathcal{Y}}(\mathcal{N})$.

**Algorithm 4. Certification of the Signaling Dimension of a Quantum Channel $\kappa_\Lambda^{\mathcal{X}\to\mathcal{Y}}(\mathcal{N})$:** Given a quantum channel $\mathcal{N} \in \text{CPTP}(\mathcal{H}_{d_A}^A \to \mathcal{H}_{d_B}^B)$, evaluate its signaling dimension for fixed inputs $\mathcal{X}$ and $\mathcal{Y}$.

1. Use Eq. (14.21) to set the upper bound on the signaling dimension to be $\texttt{ub} := \min\{d_A, d_B\}$.
2. Use Definition 101 to set the lower bound on tthe signaling dimension to be $\texttt{lb} := \lceil \text{cv}^{\mathcal{X}\to\mathcal{Y}}(\mathcal{N}) \rceil$.
3. Set the tested signaling dimension as $d := \texttt{lb} + \lfloor \frac{1}{2}(\texttt{ub} - \texttt{lb}) \rfloor$.
4. While $\texttt{lb} < \texttt{ub}$ do:
   (a) For each signaling dimension witness $(\gamma, \mathbf{G}) \in \mathcal{G}^{\mathcal{X}\overset{d}{\to}\mathcal{Y}}$, evaluate the conic optimization program

$$\widetilde{S}_\mathbf{G}^\star(\mathcal{N}) = \max_{\mathbf{P}_\mathcal{N} \in \mathcal{Q}^{\mathcal{X}\to\mathcal{Y}}(\mathcal{N})} \langle \mathbf{G}, \mathbf{P}_\mathcal{N} \rangle = \max_{\substack{\{\rho_x^A \in D(\mathcal{H}_{d_A}^A)\}_{x\in\mathcal{X}}, \\ \{\Pi_y^B\}_{y\in\mathcal{Y}} \in \text{POVM}(\mathcal{H}_{d_B}^B)}} \sum_{x\in\mathcal{X}} \sum_{y\in\mathcal{Y}} G_{y,x} \text{Tr}\left[\Pi_y^B \mathcal{N}\big(\rho_x^A\big)\right], \tag{14.47}$$

   if $\widetilde{S}_\mathbf{G}^\star(\mathcal{N}) > \gamma$, then set $\texttt{lb} := d+1$ and $d := d + \lceil \frac{1}{2}(\texttt{ub} - d) \rceil$, and continue to the next iteration of the while loop.
   (b) If no signaling dimension witnesses are violated, then set $\texttt{ub} := d$ and $d := \texttt{lb} + \lfloor \frac{1}{2}(d - \texttt{lb}) \rfloor$.
5. Return $d$ where $\kappa_\Lambda^{\mathcal{X}\to\mathcal{Y}}(\mathcal{N}) = d = \texttt{lb} = \texttt{ub}$.

**Remark.** The $\mathcal{X} \to \mathcal{Y}$ signaling dimension of a quantum channel can be extended to the full signaling dimension $\kappa_\Lambda(\mathcal{N})$ as in Definition 91.(ii) by repeating the algorithm for various $\mathcal{X}$ and $\mathcal{Y}$. Although, bounds on the signaling dimension should only be expected in practice due to the difficulty of evaluating the conic program in Eq. (14.47) as well as deriving signaling dimension witnesses for general $d$, $\mathcal{X}$, and $\mathcal{Y}$.

The certification in Algorithm 4 presents some practical challenges. First, the conic optimization programs in Eq. (14.46) and Eq. (14.47) can be analytically solved only in special cases [360]. Hence deciding whether a given quantum channel can violate a particular signaling dimension witness is typically quite challenging. However, even if the maximum is not obtained in Eq. (14.46), it can still serve as a lower bound. Furthermore, the challenge of identifying linear signaling dimension witnesses discussed in Algorithm 3 also persists in the certification of the signaling dimension of a quantum channel. In spite of these challenges, one convenience is that the permutations of the induced classical channel $\mathbf{P}_\mathcal{N}$ need not be considered because these permutations are implicit in the optimized quantum states $\{\rho_x\}_{x\mathcal{X}}$ and measurements $\{\Pi_y\}_{y\in\mathcal{Y}}$.

In practice, the optimization problem in Eq. 14.47 is solved numerically using either variational quantum optimization techniques or semidefinite programming. When applying variational optimization, the cost function is

$$\text{Cost}\left(\mathbf{P}_{\mathcal{N}}\right) := -S_{\mathbf{G}}(\mathbf{P}_{\mathcal{N}}) = -\left\langle \mathbf{G}, \mathbf{P}_{\mathcal{N}} \right\rangle \tag{14.48}$$

where the minus sign results because when the cost is minimized, the violation with respect to the signaling dimension should be maximized. On the contrary, when semidefinite programming optimization methods are applied, a see-saw algorithm must be implemented in which the measurements and states are optimized such that either the measurements or states are held constant and the other is optimized as a semidefinite program (see Eq. (14.49) and Eq. (14.50)). Alternating between state and measurement optimization incrementally improves the score $S(\mathbf{P}_{\mathcal{N}})$. For an implementation please refer to our CVChannel.jl julia package [361], which supports our work in reference [360]. In both cases, the optimization can be run on classical computing hardware for a known quantum channel expressed in the operator-sum representation where both methods are subject to finding local optima. The variational quantum optimization methods have a key advantage over convex optimization techniques because variational quantum optimization allows the signaling dimension of an unknown channel to be certified (see Fig. 14.3.a). On the contrary, convex optimization techniques can be applied only when the channel is completely characterized such that it can be expressed in the operator-sum representation.

It is important to point out that the certification procedure in Algorithm 4 requires the sender and receiver to both be trusted and characterized. This trust or characterization may not exist in practice. Consider first the case Fig. 14.3.c) where the measurement device is untrusted. The noise can be incorporated into the measurement as $\widetilde{\Pi}_y^B = \mathcal{N}^\dagger(\Pi_y^B)$ and the conic optimization program reduces to the semidefinite program

$$\max_{\{\rho_x \in D(\mathcal{H}_{d_A}^A)\}_{x\in\mathcal{X}}} \sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}} G_{y,x} \text{Tr}\left[\widetilde{\Pi}_y^B \rho_x^A\right]. \tag{14.49}$$

Similarly, if the sender is not trusted as in Fig. 14.3.d), then the noise is incorporated into the states as $\widetilde{\rho}_x^A = \mathcal{N}(\rho_x^A)$ and the folowing semidefinite program results,

$$\max_{\{\Pi_y^B\}_{y\in\mathcal{Y}} \in \text{POVM}(\mathcal{H}_{d_B}^B)} \sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}} G_{y,x} \text{Tr}\left[\Pi_y^B \widetilde{\rho}_x^A\right]. \tag{14.50}$$

In both cases, if $\widetilde{\rho}_x^A$ or $\widetilde{\Pi}_y^B$ can be expressed in matrix form, then semidefinite programming can be used to optimize the problem with relative efficiency on classical hardware. If these quantum operators are not known, then variational quantum optimization can be applied on the quantum hardware directly. Such scenarios correspond to a self-testing procedure where a trusted device is used to characterize the other device combined with the noisy channel.

In the worst case, neither the sender device or the receiver device are trusted then the upper bound $\kappa_\Lambda(\mathcal{N}) \leq \min\{d_A, d_B\}$ in Eq. (14.21) no longer holds. However, this lack of trust is not entirely detrimental as lower and upper bounds can still be found using the certification procedure for signaling dimension of a classical behavior $\kappa_\Lambda(\mathbf{P}_{\mathcal{N}})$ where $\mathbf{P}_{\mathcal{N}} \in \mathcal{Q}^{\mathcal{X}\xrightarrow{d}\mathcal{Y}}(\mathcal{N})$ is the observed classical channel generated using a noisy quantum channel. Alternatively, if the sender and receiver are untrusted, but their devices can be tuned over a range of state preparations and measurements (without having knowledge of what these states or measurements are), then the conic program in Eq. (14.47) can be optimized using a black-box optimization technique (see Section 11.1 in Chapter 11) to obtain violations of signaling dimension witnesses and improve

the lower bound on $\kappa_\Lambda(\mathcal{N})$. However, with no knowledge of the Hilbert space dimension, the upper bounds are given as $\kappa_\Lambda^{\mathcal{X}\to\mathcal{Y}}(\mathcal{N}) \leq \min\{|\mathcal{X}|, |\mathcal{Y}|\}$ and $\kappa_\Lambda(\mathcal{N}) \leq \infty$.

As a final remark, we can extend Algorithm 4 to scenarios in which entanglement assists the sender and receiver. Here the structure of the quantum system changes to fit the communication resources at play, changing the form of the conic program in Eq. (14.47). Although the optimization program changes, the approach stays remains the same. Indeed, we simply apply our variational optimization methods to the entanglement-assisted resources to find violations of signaling dimension witnesses.

### *14.4.4 The Signaling Dimension of Replacer Channels*

Despite the general difficulty of certifying $\kappa_\Lambda(\mathcal{N})$, we nevertheless establish bounds for the signaling dimension of partial replacer channels. Remarkably, this example demonstrates a given case where the signaling dimension is certified using analytic methods. Consider first a $d$-dimensional partial replacer channel, which has the form

$$\mathcal{R}_\mu(X) = \mu X + (1-\mu)\mathrm{Tr}\left[X\right]\sigma \in \mathrm{CPTP}(\mathcal{H}_d \to \mathcal{H}_{d'}), \tag{14.51}$$

where $1 \geq \mu \geq 0$ and $\sigma$ is some fixed density matrix. The partial depolarizing channel $\mathcal{D}_\mu$ corresponds to $\sigma$ being the maximally mixed state whereas the partial erasure channel $\mathcal{E}_\mu$ corresponds to $\sigma$ being an erasure flag $|E\rangle\langle E|$ with $|E\rangle$ being orthogonal to $\{|1\rangle, \cdots, |d\rangle\}$.

**Theorem 12.** [114] The signaling dimension of a $d$-dimensional partial replacer channel is bounded by

$$\lceil \mu d + 1 - \mu \rceil \leq \kappa_\Lambda(\mathcal{R}_\mu) \leq \min\{d, \lceil \mu d + 1 \rceil\}. \tag{14.52}$$

Moreover, for the partial erasure channel, the upper bound is tight for all $\mu \in [0, 1]$.

*Proof.* We first prove the upper bound in Eq. (14.52). The trivial bound $\kappa_\Lambda(\mathcal{R}_\mu) \leq d$ was already observed in Eq. (14.21). To show that $\kappa_\Lambda(\mathcal{R}_\mu) \leq \lceil \mu d + 1 \rceil$, let $\{\rho_x\}_x$ be any collection of inputs and $\{\Pi_y\}_{y\in\mathcal{Y}} \in \mathrm{POVM}(\mathcal{H}_d)$. Then

$$P_{\mathcal{R}_\mu}(y|x) = \mu P(y|x) + (1-\mu)S(y), \tag{14.53}$$

where $P(y|x) = \mathrm{Tr}\left[\Pi_y \rho_x\right]$ and $S(y) = \mathrm{Tr}\left[\Pi_y \sigma\right]$. From Lemma 13, we know that $P(y|x)$ can be simulated classically using $d$ signals and shared randomness. Substituting this into Eq. (14.53) yields

$$P_{\mathcal{R}_\mu}(u|x) = \sum_\lambda q(\lambda) \sum_{m=1}^{d} R_\lambda(m|x)[\mu T_\lambda(y|m) + (1-\mu)S(y)]. \tag{14.54}$$

For $r = \lceil \mu d + 1 \rceil$, let $\nu$ be a random variable uniformly distributed over $\{\binom{d}{r-1}\}$, which is the collection of all subsets of $[d]$ having size $r-1$. For a given $\lambda$, $\nu$, and input $x$, Alice performs the channel $\mathbf{T}_\lambda$. If $m \in \nu$, Alice sends message $m' = m$; otherwise, Alice sends message $m' = 0$. Upon receiving $m'$, Bob does the following: if $m' \neq 0$ he performs channel $\mathbf{R}_\lambda$ with probability $\frac{\mu d}{r-1}$ and samples from distribution $S(y)$ with probability $1 - \frac{\mu d}{r-1}$; if $m' \neq 0$ he samples from $S(y)$ with probability one. Since $Pr\{m \in \nu\} = \frac{r-1}{d}$, this protocol faithfully simulates $P_{\mathcal{R}_\mu}$. To establish the lower bound in Eq. (14.53), suppose that Alice sends orthogonal states $\{|1\rangle, \cdots, |d\rangle\}$ and Bob measures in the same basis. Then

$$\sum_{i=1}^{d} \langle i| \mathcal{R}_\mu(|i\rangle\langle i|) |i\rangle = d\mu + (1-\mu), \tag{14.55}$$

which will violate Eq. (14.29) for the ML polytope $\mathcal{M}_r^{d\to d}$ whenever $r < \mu d + (1-\mu)$. Hence any zero-error simulation will require at least $\lceil \mu d + 1 - \mu \rceil$ classical messages. For the erasure channel, this lower bound can be tightened by considering the score for other ambiguous games, as detailed in Appendix B.5. □

## *14.5 Nonclassicality in Signaling Scenarios*

The signaling dimension witnesses and certification procedures described earlier can serve as witness of nonclassicality in signaling scenarios. By Lemma 13 we know that $\mathcal{Q}^{\mathcal{X}\xrightarrow{d}\mathcal{Y}} \subseteq \mathcal{C}_\Lambda^{\mathcal{X}\xrightarrow{d}\mathcal{Y}}$, indicating that quantum signaling is not stronger of a resource than classical signaling assisted with an unlimited amount of shared randomness. However, there are also clear examples showing that $\mathcal{C}^{\mathcal{X}\xrightarrow{d}\mathcal{Y}} \subset \mathcal{Q}^{\mathcal{X}\xrightarrow{d}\mathcal{Y}}$ [123]. As a result, the foundation is laid for weak nonclassicality in quantum signaling scenarios. Furthermore, it was recently observed that the $k$-guessing game in Definition 99 can be violated when a bit of communication is assisted by entanglement [116], hence $\mathcal{C}_\Lambda^{\mathcal{X}\xrightarrow{d}\mathcal{Y}} \subset \mathcal{C}_{\text{EA}}^{\mathcal{X}\xrightarrow{d}\mathcal{Y}}$ and strong nonclassicality is expressed. It naturally follows that $\mathcal{C}_{\text{EA}}^{\mathcal{X}\xrightarrow{d}\mathcal{Y}} \subseteq \mathcal{Q}_{\text{EA}}^{\mathcal{X}\xrightarrow{d}\mathcal{Y}}$ because a dense coding protocol can be used to communicate 2-bits using one qubit of communication and entanglement between the sender and receiver [37], [265]. As a result, both classical and quantum entanglement-assisted scenarios can exhibit strong nonclassicality. We now turn to elaborate upon these nonclassical behaviors that can be generated in signaling systems.

### *14.5.1 Weak Nonclassicality in Unassisted Quantum Signaling Scenarios*

In this section we show that unassisted quantum signaling scenarios can exhibit weak nonclassicality. Furthermore, we provide an example where an unbounded amount of classical communication is needed to simulate a quantum signaling behavior. Hence we prove that an unlimited amount of shared randomness can be needed to simulate weakly nonclassical behaviors. Indeed, this observation indicates a resource advantage in quantum signaling systems.

**Theorem 13.** Consider a quantum signaling system where $|\mathcal{X}| = |\mathcal{Y}| = N$ and the noiseless channel $\text{id}_{d=2} \in \text{CPTP}(\mathcal{H}_2^A \to \mathcal{H}_2^B)$ is used to communicate the qubit signals. As $N \to \infty$ there exists behaviors $\mathbf{P} \in \mathcal{Q}^{\mathcal{X}\xrightarrow{d}\mathcal{Y}}$ such that the unassisted signaling dimension is unbounded as $\kappa(\mathbf{P}) \to \infty$.

*Proof.* We prove this theorem by showing an example unassisted signaling scales with the number of inputs and outputs $N$. Consider the planar symmetric $N$-element qubit POVM $\{\Pi_x = \frac{2}{N}|\psi_y\rangle\langle\psi_y|\}_{y=0}^{N-1}$ where $N$ is even and $|\psi_y\rangle = \cos(\frac{\pi y}{N})|0\rangle + \sin(\frac{\pi y}{N})|1\rangle$. The maximum likelihood facet inequality is saturated only if

$$2 = \langle \mathbf{G}, \mathbf{P}\rangle = \sum_{x=0}^{N-1} \text{Tr}\left[\Pi_{y=x}^B \rho_x^A\right] = \frac{2}{N}\sum_{x=0}^{N-1} \text{Tr}\left[|\psi_{y=x}\rangle\langle\psi_{y=x}|\psi_x\rangle\langle\psi_x|\right] = \frac{2}{N}\sum_{x=0}^{N-1} |\langle\psi_x|\psi_x\rangle|^2. \tag{14.56}$$

The resulting conditional probabilities are then

$$P(y|x) = \left(\cos^2(\frac{\pi y}{N})\cos^2(\frac{\pi x}{N}) + 2\cos(\frac{\pi y}{N})\cos(\frac{\pi x}{N})\sin(\frac{\pi y}{N})\sin(\frac{\pi x}{N}) + \sin^2(\frac{\pi y}{N})\sin^2(\frac{\pi x}{N})\right) \tag{14.57}$$

and the quantum signaling behavior takes the form

$$\mathbf{P} = \begin{pmatrix} \frac{2}{N} & \downarrow & \downarrow & 0 & \uparrow & \uparrow \\ \uparrow & \ddots & \downarrow & \downarrow & \ddots & \uparrow \\ \uparrow & \uparrow & \frac{2}{N} & \downarrow & \downarrow & 0 \\ 0 & \uparrow & \uparrow & \frac{2}{N} & \downarrow & \downarrow \\ \downarrow & \ddots & \uparrow & \uparrow & \ddots & \downarrow \\ \downarrow & \downarrow & 0 & \uparrow & \uparrow & \frac{2}{N} \end{pmatrix} \tag{14.58}$$

where indicate the direction of increasing values in each column, the largest value in each column is $P(y = x|x) = \frac{2}{N}$, the smallest value is in each column is $P(y = f(x)|x) = 0$ where $f(x) = x + N/2 \mod N$, and $P(y|x) > 0$ when $y \neq f(x)$. Upon making these observations, we can derive two lower bounds on the $\text{Rank}_+(\mathbf{P})$. The first lower bound is quite naive, but clearly proves the statement that $\text{Rank}_+(\mathbf{P}) \to \infty$ as $N \to \infty$. The second lower bound improves the first, but is less straightforward. The remainder of our proof is adapted from Fawzi *et al.* [353] while similar results are presented by Beasley *et al.* [362].

Using the definition of nonnegative rank in Eq. (14.6), we seek the smallest integer $k$ such that the $N$ nonnegative vectors $|a_0\rangle, \ldots, |a_{N-1}\rangle, |b_0\rangle, \ldots, |b_{N-1}\rangle \in \mathbb{R}_+^k$ satisfy $P(y|x) = \langle a_y|b_x\rangle$ for all $x$ and $y$. Observe that the constraint $P(f(x)|x) = \langle a_{f(x)}|b_x\rangle = 0$ requires that $\text{Supp}(|a_{f(x)}\rangle) \cap \text{Supp}(|b_x\rangle) = \emptyset$ and, for all $y \neq f(x)$, the constraint $P(y|x) = \langle a_y|b_x\rangle > 0$ requires that $\text{Supp}(|a_y\rangle) \cap \text{Supp}(|b_x\rangle) \neq \emptyset$. Suppose now that $\text{Supp}(|a_y\rangle) = \text{Supp}(|a_{f(x)}\rangle)$ for $y \neq f(x)$, then $\langle a_y|b_x\rangle = \langle a_{f(x)}|b_x\rangle = 0$, which is a contradiction because $\langle a_y|b_x\rangle > 0$. Therefore, there cannot be two vectors $|a_y\rangle$ and $|a_{y'}\rangle$ with matching supports for $y \neq y'$ where a similar constraint also holds for the vectors $\{|b_x\rangle\}_{x\in\mathcal{X}}$. Since there are $2^k$ ways to choose the support of $|a_y\rangle$, $k$ must be sufficiently large such that $2^k \geq N$. Taking the logarithm of both sides yields $k \geq \log_2(N)$, therefore it must hold that

$$\text{Rank}_+(\mathbf{P}) \geq \lceil \log_2(N) \rceil. \tag{14.59}$$

It follows that $\text{Rank}_+(\mathbf{P}) \to \infty$ as $N \to \infty$, which proves that an unbounded amount of classical communication is needed to simulate one qubit of quantum communication.

The lower bound in Eq. (14.59) is quite naive, however, because it assumes that all $2^k$ supports of $|a_y\rangle$ are relevant. To be more restrictive, consider the case where $\text{Supp}(|a_y\rangle) \subseteq \text{Supp}(|a_{f(x)}\rangle)$ for $y \neq f(x)$. As a result, we again obtain the contradiction $\langle a_y|b_x\rangle = \langle a_{f(x)}|b_x\rangle = 0$. Therefore, for any $y \neq y'$, $\text{Supp}(|a_y\rangle)$ cannot contain $\text{Supp}(|a_{y'}\rangle)$ and *vice versa*. Obeying this constraint yields $\binom{k}{\lfloor k/2 \rfloor}$ vectors whose supports mutually exclude the other vectors, hence

$$\text{Rank}_+(\mathbf{P}) = \min\left\{k \in \mathbb{N} \mid N \leq \binom{k}{\lfloor \frac{k}{2} \rfloor}\right\} \geq \lceil \log_2(N) \rceil. \tag{14.60}$$

Note that Eq. (14.60) improves upon the naive lower bound in Eq. (14.59). □

To verify the results of Theorem 13, we apply our nonnegative rank certification procedure in Algorithm 2 to the planar symmetric qubit signaling behaviors having the form in Eq. (14.57) and plot the results in Fig. 14.4. While the numerical data confirms that the nonnegative rank, and equivalently the unassisted signaling dimension, is bounded as

$$|\mathcal{X}| \geq \text{Rank}_+(\mathbf{P}) = \kappa(\mathbf{P}) \geq \min\{k \in \mathbb{N} \mid |\mathcal{X}| \leq \binom{k}{\lfloor k/2 \rfloor}\}, \tag{14.61}$$

the numerical data also suggests that the lower bound could be quite loose in general. Indeed, our numerical results imply that the lower bound scales linearly with $|\mathcal{X}|$ rather than logarithmically as proven in Theorem 13.

In Fig. 14.4, we also consider the case when white noise is mixed into the behavior as $\widetilde{\mathbf{P}}_\gamma = (1-\gamma)\mathbf{P} + \gamma\mathbf{W}$ where $\mathbf{W} = \sum_{x\in\mathcal{X}} \sum_{y\in\mathcal{Y}} \frac{1}{|\mathcal{Y}|} |y\rangle\langle x|$. This noise model corresponds to a depolarizing channel being placed on the the state prior to measurement. Even when $\gamma = 0.01$ (purple crosses), we find that the approximate nonnegative rank is below the theoretical lower bound given by the dash-dotted blue steps. Furthermore as the noise parameter $\gamma$ increases, the nonnegative rank $\text{Rank}_+\left(\widetilde{\mathbf{P}}_\gamma\right)$ decreases to its minimal value of 3. Since matrices $M$ having $\text{Rank}\,(M) = 2$ also have $\text{Rank}_+\,(M) = 2$, if $\text{Rank}\,(M) > 2$, then it follows that $\text{Rank}_+\,(M) \geq 3$. Overall, we find that white noise can decrease considerably the unassisted classical simulation cost $\kappa(\widetilde{\mathbf{P}}_\gamma)$.

We must be cautious in asserting these numerical results as fact because Algorithm 2 certifies the approximate nonnegative rank whereas Theorem 13 considers exact nonnegative rank. The difference being that two behaviors $\mathbf{P}, \mathbf{P}' \in \mathcal{P}_{\mathcal{Y}|\mathcal{X}}$ are approximately equal if $D(\mathbf{P}, \mathbf{P}') = \frac{1}{2|\mathcal{X}|} \sum_{x\in\mathcal{X}} \sum_{y\in\mathcal{Y}} |P(y|x) - P'(y|x)| \leq \varepsilon$, while exact equivalence requires $\varepsilon = 0$. An approximate nonnegative rank should always lower bound the true nonnegative rank, implying that the cases in Fig. 14.4 where $|\mathcal{X}| = n = \text{Rank}_+\,(\mathbf{P})$ are asserting the

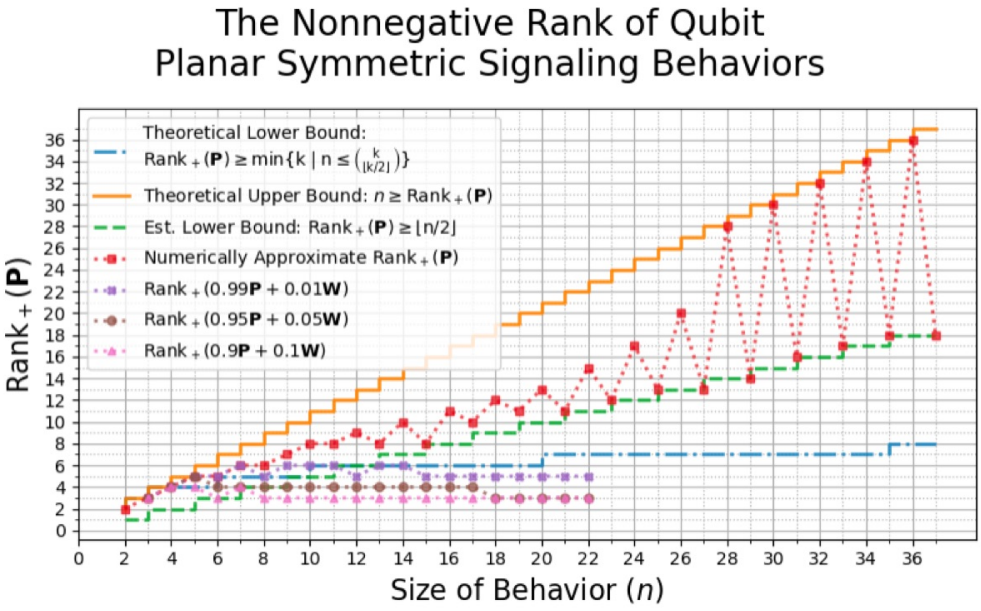



Figure 14.4: We use Algorithm 2 to certify the nonnegative rank of qubit signaling behaviors $\mathbf{P} \in \mathcal{Q}^{\mathcal{X} \overset{d=2}{\to} \mathcal{Y}}$ that have the planar symmetric form in Eq. (14.57) where $|\mathcal{X}| = |\mathcal{Y}| = n \in \{2, \ldots, 37\}$. Furthermore, we also certify the nonnegative rank when white noise $\mathbf{W} = \sum_{x,y} \frac{1}{|\mathcal{Y}|} |y\rangle\langle x|$ is mixed into the planar symmetric behavior as $\widetilde{\mathbf{P}}_\gamma = (1-\gamma)\mathbf{P} + \gamma\mathbf{W}$. To mitigate the risk of finding local optima during the nonnegative matrix factorization, we repeat each factorization 20 times with randomized initial parameters and consider only the factorization having smallest error. The solid orange line plots the trivial upper bound $n \geq \text{Rank}_+\,(\mathbf{P})$, the blue dash-dotted line plots the theoretical lower bound in Eq. (14.60), and the green dashed line plots the lower bound $\text{Rank}_+\,(\mathbf{P}) \geq \lfloor \frac{n}{2} \rfloor$ estimated from the numerical data. The markers plot the $\text{Rank}_+\left(\widetilde{\mathbf{P}}_\gamma\right)$ as certified using Algorithm 2 where red squares mark the noiseless $\gamma = 0$ case, purple crosses mark the $\gamma = 0.01$ case, brown circles mark the $\gamma = 0.05$ case, and pink stars mark the $\gamma = 0.1$ case. The algorithm outputs an approximate classical signaling behavior $\mathbf{P}' \in \mathcal{C}^{\mathcal{X} \overset{d=k}{\to} \mathcal{Y}}$ that has distance $D(\widetilde{\mathbf{P}}_\gamma, \mathbf{P}') = \frac{1}{2n} \sum_{x,y=1}^{n} |P(y|x) - P'(y|x)| \leq \varepsilon = 10^{-9}$.

exact nonnegative rank. However, this is still not necessarily fact because the certification algorithm can still converge to local optima, meaning that, perhaps, the nonnegative rank is much lower in these cases where the algorithm simply failed to find the optimal nonnegative matrix factorization.

To improve upon these numerical results, the most conservative option would be to improve the numerical precision of the nonnegative matrix factorization, however, this could incorporate a significant computational overhead as the computational problem is NP-hard. Alternatively, more trials of nonnegative rank certifications could be run in the hopes that we win the lottery so-to-speak where the ideal initial parameters are chosen at random to allow the optimal matrix factorization to be found. Lastly, perhaps the tolerance for error $\varepsilon = 10^{-9}$ is too strict and should be increased, however, doing so could be detrimental because the resulting behaviors would include more error.

Theorem 13 implies the behaviors of noiseless qubit signaling systems can require an unbounded amount of classical communication to simulate. However, since the hierarchy holds that $\mathcal{C}^{\mathcal{X}\xrightarrow{d}\mathcal{Y}} \subseteq \mathcal{Q}^{\mathcal{X}\xrightarrow{d}\mathcal{Y}} \subseteq \mathcal{C}_{\Lambda}^{\mathcal{X}\xrightarrow{d}\mathcal{Y}}$, the extra communication needed to simulate the qubit serves only to distribute an unbounded amount of shared randomness. Thus, the example of planar symmetric qubit encodings and decodings described in the proof of Theorem 13 describe an instance of weak nonclassicality, in which the weak nonclassicality of quantum signaling scenarios is no more than an exhibition of the randomness intrinsic to quantum systems [214]. This is because the planar symmetric encodings and decodings are analog in nature, meaning that there is no limit to the number of states or measurement operators that can be oriented symmetrically along a plane of the Bloch sphere. From this perspective, the observed weak nonclassicality is likened to a digital classical system attempting to simulate an analog quantum system. At some point, the numerical precision is insufficient to replicate the behavior exactly. However, depending on the context, there may be little value to improving numerical precision of a classical simulation by using additional classical communication.

We show that when white noise is present on the planar symmetric quantum behavior $\widetilde{\mathbf{P}}_\gamma$, the unassisted simulation cost $\kappa(\widetilde{\mathbf{P}})$ decreases considerably. Thus, even the amount of noise is relatively small, the logarithmic scaling in the nonnegative rank of the noiseless behavior vanishes. Although it is not clear from Fig. 14.4 whether the nonnegative rank of the noisy behaviors will continue to increase with $n$ or not, there is still a clear separation from the noisy and noiseless cases. Remarkably, we find that there will always be a slight quantum advantage because $\text{Rank}\left(\widetilde{\mathbf{P}}_\gamma\right) > 2$ for all $n \geq 3$ and $\gamma \in [0, 1)$. Hence, the nonnegative rank of the noisy behavior is bounded below as $\text{Rank}_+\left(\widetilde{\mathbf{P}}_\gamma\right) \geq 3$. However, the signaling dimension of this scenario is $\kappa_\Lambda(\widetilde{\mathbf{P}}_\gamma) \leq 2$, meaning that a quantum advantage persists in the unassisted cases for all noise models except $\gamma = 1$.

### *14.5.2 Strong Nonclassicality in Entanglement-Assisted Signaling Scenarios*

Violations of signaling dimension witnesses can observe strong nonclassicality in entanglement assisted signaling scenarios (see Fig. 14.1.d,e). Consider the certification scenario in which that the amount of communication between sender and receiver is known as $d$ and the amount of entanglement between the sender and receiver is to be tested.

In Theorem 11, we derived the complete set of signaling dimension witnesses that bound the signaling polytopes of any qubit signaling system. These signaling dimension witnesses are listed in Table 14.1 and are labeled similarly using alphabetic subscripts, *e.g*, $(\gamma_a, \mathbf{G}_a)$. These signaling dimension witnesses can test for strong nonclassicality in entanglement-assisted signaling scenarios. Furthermore, these signaling dimension violations can be used to self-test LOCC resources such as entanglement-assisted classical and quantum

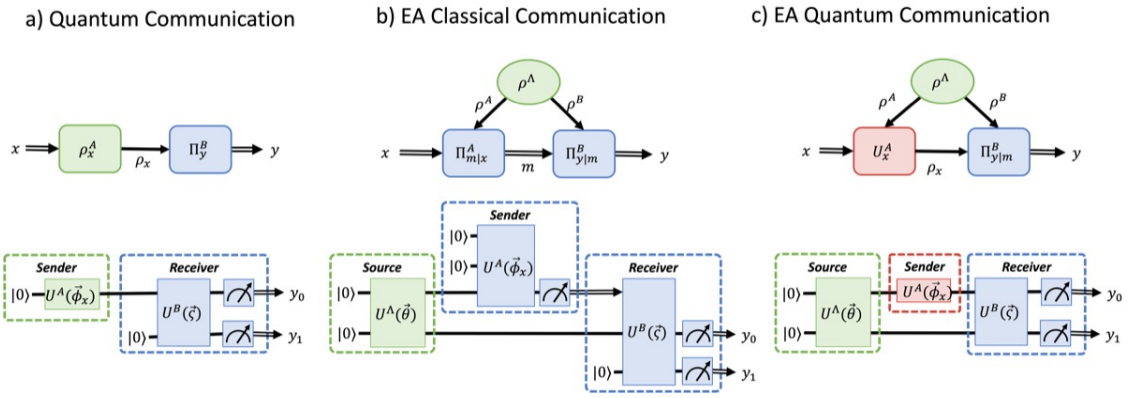



Figure 14.5: Variational ansatzes for entanglement-assisted signaling scenarios. a) Quantum communication with no entanglement assistance where the receiver applies a POVM qubit measurement with four outcomes. b) Entanglement-assisted classical communication where the sender's ansatz is given in Fig. 14.6. Note that the receiver performs a qubit POVM measurement conditioned upon the sender's classical measurement outcome $m$. c) Entanglement-assisted quantum communication where the sender applies an arbitrary qubit unitary to their portion of the entangled state. The receiver jointly measures the two qubits using a projective measurement.

communication.

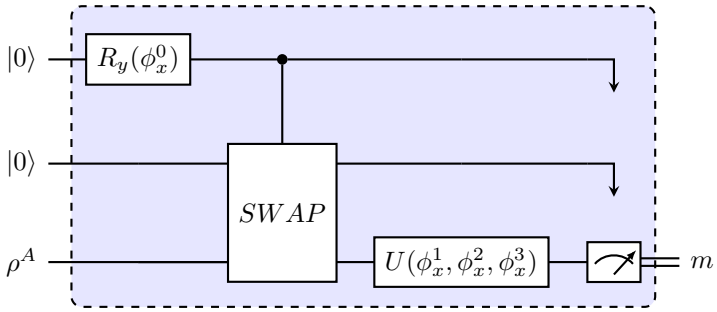



Figure 14.6: The expanded entanglement-assisted classical signaling encoding ansatz $U^A(\vec{\phi}_x)$ shown in Fig. 14.5.b). The $R_y(\phi_x^0)$ gate controls whether or not the $SWAP$ gate exchanges the $|0\rangle$ state with the $\rho^A$ part of the entangled state. The output bit $m$ is obtained from arbitrary qubit measurement applied. Note that either the state $|0\rangle$ or $\rho^A$ is measured depending on the controlled $SWAP$. All remaining qubits are ignored.

To obtain these nonclassical violations we apply our variational optimization methods. The optimization objective being

$$\min_{\Theta} \text{Cost}\left(\mathbf{P}(\Theta)\right) = \min_{\Theta} -\langle \mathbf{G}, \mathbf{P}(\Theta) \rangle \,. \tag{14.62}$$

where $\mathbf{G}$ is one of the signaling dimension witness in Table 14.1. To construct the behavior $\mathbf{P}(\Theta)$, we consider distinct variational ansatzes for classical and quantum communication scenarios as shown in Fig. 14.5. It is important to note that the signaling dimension witness in Table 14.1 have $|\mathcal{X}| \in \{3, 4, 5, 6\}$ and $|\mathcal{Y}| = 4$, meaning that it is sufficient to consider only four outputs.

We then optimize the variational ansatzes in Fig. 14.5 with respect to the signaling dimension witnesses in Table 14.1 to obtain the results in Fig. 14.7. As expected from Lemma 13, we find no violations for the unassisted quantum signaling case in Fig. 14.5.a) because $\mathcal{Q}^{\mathcal{X} \xrightarrow{d} \mathcal{Y}} \subseteq \mathcal{C}_{\Lambda}^{\mathcal{X} \xrightarrow{d} \mathcal{Y}}$. Hence, we do not plot this data.

In the entanglement-assisted classical signaling cases shown by the blue bars we find that the maximum likelihood games $\mathbf{G}_a$ and $\mathbf{G}_b$ do not have violations. This result indicates that entanglement cannot improve

communication value or capacity of a signaling system. Nonetheless there still exists an operational advantage because all remaining signaling dimension witnesses are violated by entanglement-assisted classical signaling. It is important to note that these violations could only be found when the sender uses the encoding ansatz in Fig. 14.6, that is, the violations seem to require the sender to abstain from measuring the entangled state when given certain inputs.

It is very interesting the maximum likelihood games show no advantage whereas the anti-guessing game $\mathbf{G}_c$ and ambiguous guessing game $\mathbf{G}_d$, $\mathbf{G}_e$, $\mathbf{G}_f$, $\mathbf{G}_g$ signaling dimension witnesses do show advantages. We find this interesting because these games bear many similarities to the maximum likelihood game, for which we do not find a violation. Moreover, the signaling dimension witnesses $\mathbf{G}_e$, $\mathbf{G}_f$, and $\mathbf{G}_g$ that are nontrivial input liftings of the ambiguous guessing game yield larger violations than the ambiguous guessing game. Thus, these facet inequalities appear to be distinct from the ambiguous guessing game in $\mathbf{G}_d$.

For the $k$-guessing game $\mathbf{G}_h$, our variational optimization successfully reproduces the results given by Frenkel *et al.* [116]. Indeed, we can reiterate their optimal encoding and decoding strategy as follows. Let entanglement source $\Lambda$ prepare the two-qubit state $|\psi\rangle = \frac{1}{\sqrt{2}}(|01\rangle - |10\rangle)$. Let the sender perform the following measurements

$$\Pi^A_{\pm|0} \in \{\mathbb{I}_2, 0\}, \quad \Pi^A_{\pm|1} \in \{\frac{1}{2}(\mathbb{I}_2 \pm \sigma_z)\}, \qquad \Pi^A_{\pm|2} \in \{\frac{1}{2}(\mathbb{I}_2 \pm \sigma_x)\}, \tag{14.63}$$

$$\Pi^A_{\pm|3} \in \{\frac{1}{2}(\mathbb{I}_2 \mp \sigma_z)\} \quad \Pi^A_{\pm|4} \in \{\frac{1}{2}(\mathbb{I}_2 \mp \sigma_x)\} \qquad \Pi^A_{\pm|5} \in \{0, \mathbb{I}_2\}. \tag{14.64}$$

Then, let the receiver perform the following POVM measurements conditioned upon $m = \pm$

$$\Pi^B_{0|+} = \frac{1}{2}(\mathbb{I}_2 - \frac{1}{\sqrt{2}}(\sigma_z + \sigma_x)) \quad \Pi^B_{1|+}, \quad \Pi^B_{1|+} = \mathbb{I}_2 - \Pi^B_{0|+}, \quad \Pi^B_{2|+} = \Pi^B_{3|+} = 0 \tag{14.65}$$

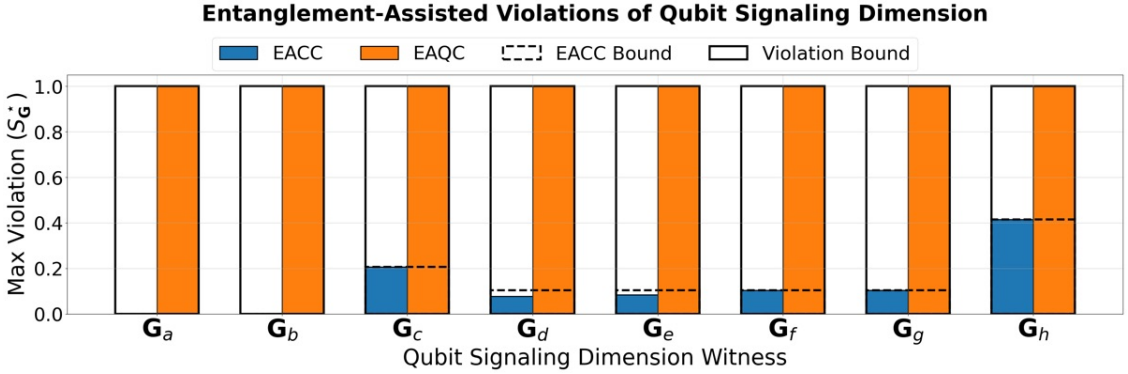



Figure 14.7: For each of the qubit signaling dimension witnesses listed in Table 14.1 we use variational optimization to find maximal violations for entanglement-assisted classical communication (EACC) and entanglement-assisted quantum communication (EAQC). The $y$-axis plots the maximal violation rescaled as $S^\star_{\mathbf{G}_i} = (S^\star_{\mathbf{G}_i}(\mathbf{P}) - \gamma_i)/(\gamma^\star_i - \gamma_i)$ where $\gamma^\star_i > \gamma_i$ is the largest possible violation. This rescaling ensures that, for all signaling dimension witnesses, the maximal violation possible is 1 and classical bound is 0. The orange bar denotes the EAQC violation and obtains the max violation $\gamma^\star$ in all cases. The blue bar denotes the EACC violation that violates all inequalities that are not maximum likelihood facets. The dashed bar is a predicted upper bound for all EACC violatons. For games $\mathbf{G}_d$ and $\mathbf{G}_e$ we predict that this upper bound is loose, while in all other cases, we predict the bound to be tight.

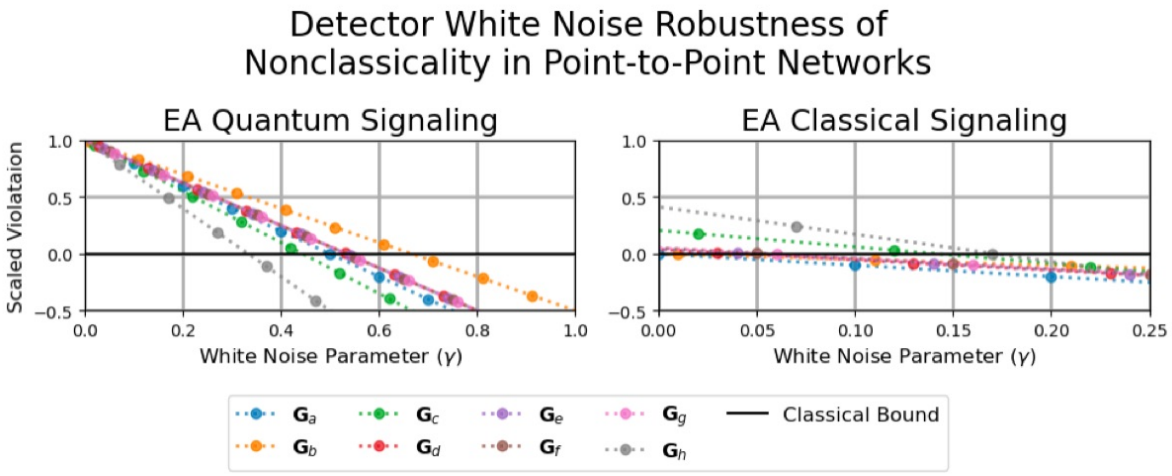



Figure 14.8: The robustness of nonclassicality is plotted for each of the entanglement-assisted violations of signaling dimension shown in Fig. 14.7. For each signaling dimension witness, we plot the curve $\widetilde{\mathbf{P}}_\gamma = (1-\gamma)\mathbf{P} + \gamma\mathbf{W}$ where $\mathbf{W} = \sum_{x,y} \frac{1}{|\mathcal{Y}|} |y\rangle\langle x|$ is a white noise behavior. Each colored line corresponds to a different nonclassicality witness where the markers are aesthetic for distinguishing overlapping lines. The classical violations are scaled such that the maximum possible violation is one where zero is the classical bound. (Left) The noise robustness of entanglement-assisted quantum signaling behaviors. (Right) The noise robustness of entanglement-assisted classical signaling behaviors.

and

$$\Pi^B_{0|-} = \Pi^B_{1|-} = 0 \quad \Pi^B_{2|-} = \frac{1}{2}(\mathbb{I}_2 + \frac{1}{\sqrt{2}}(\sigma_z - \sigma_x)), \quad \Pi^B_{3|-} = \mathbb{I}_2 = \Pi^B_{2|-}. \tag{14.66}$$

Finally, the probabilities are calculated as

$$P(y|x) = \mathrm{Tr}\left[\Pi^A_{+|x} \otimes \Pi^B_{y|+} |\psi\rangle\langle\psi|\right] + \mathrm{Tr}\left[\Pi^A_{-|x} \otimes \Pi^B_{y|-} |\psi\rangle\langle\psi|\right]. \tag{14.67}$$

The maximal score is then $S^\star_{\mathbf{G}_h} = 4 + \sqrt{2} \approx 5.414$.

When quantum signaling is assisted by entanglement as in Fig. 14.5.c) the violations correspond to the orange bars in Fig. 14.7. Naturally, a dense coding protocol can be implemented using these quantum resources such that 2-bits of classical communication can be achieved between the sender and receiver. Since $|\mathcal{Y}| = 4$, we find from Eq. (14.20) that $\kappa_\Lambda(\mathbf{P}) \leq 4$, hence entanglement-assisted qubit communication can produce any behavior in $\mathcal{P}_{\mathcal{Y}|\mathcal{X}}$, the full probability polytope. As a result, the perfect winning score can be achieved for all considered signaling dimension witnesses. Our variational optimizations are able to obtain the perfect winning scores as shown in Fig. 14.7 by the orange bars achieving the violation bound.

### *14.5.3 The Noise Robustness of Strong Nonclassicality in Signaling Systems*

To gain a high-level perspective of the relative noise robustness of each qubit signaling dimension witness, we consider depolarizing noise $\mathcal{W}_{(1-\gamma)}(\Pi^B_y)$ where $\Pi^B_y = (U^B)^\dagger |y\rangle\langle y| U^B$ is a the two-qubit projective measurement given in Fig. 14.5. Note that the action of the depolarizing channel on each measurement operator is

$$\mathcal{W}_{(1-\gamma)}(\Pi^B_y) = (1-\gamma)\Pi^B_y + \gamma\frac{1}{4}\mathbb{I}_4, \tag{14.68}$$

therefore, the noisy probabilities are

$$P(y|x) = (1-\gamma)\mathrm{Tr}\left[(U^B)^\dagger |y\rangle\langle y| U^B \rho_x^B\right] + \frac{\gamma}{4}\mathrm{Tr}\left[\rho_x^B\right] = (1-\gamma)\mathrm{Tr}\left[(U^B)^\dagger |y\rangle\langle y| U^B \rho_x^B\right] + \frac{\gamma}{4}. \tag{14.69}$$

In the entanglement-assisted quantum signaling case, the measured state is $\rho_x^B = (U_x^A \otimes \mathbb{I})\rho^\Lambda (U_x^A \otimes \mathbb{I})^\dagger$ where applying a depolarizing channel to the entangled state produced at the source yields the probalities

$$P(y|x) = \mathrm{Tr}\left[\Pi_y^B (U_x^A \otimes \mathbb{I})\mathcal{W}_{(1-\gamma)}(\rho^\Lambda)(U_x^A \otimes \mathbb{I})^\dagger\right] = (1-\gamma)\mathrm{Tr}\left[\Pi_y^B (U_x^A \otimes \mathbb{I})\mathcal{W}_{(1-\gamma)}(\rho^\Lambda)(U_x^A \otimes \mathbb{I})^\dagger\right] + \frac{\gamma}{4}\mathrm{Tr}\left[\Pi_y^B\right]. \tag{14.70}$$

If $\Pi_y^B$ is a projective rank-one measurement, then $\mathrm{Tr}\left[\Pi_y^B\right] = 1$ and Eq. (14.70) is equivalent to Eq. (14.69), showing that within our entanglement-assisted quantum communication ansatz, a depolarized measurement is equivalent to a depolarized state preparation at the source.

Using Proposition 3 we can conveniently replace the depolarizing noise on the measurement with the postprocessing map $\mathbf{W}_\gamma = (1-\gamma)\mathbb{I}_{|\mathcal{Y}|} + \gamma\mathbf{W}$ where $\mathbf{W} = \sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}} \frac{1}{|\mathcal{Y}|}|y\rangle\langle x|$. It then becomes a simple matter of taking the optimal behavior $\mathbf{P}$ for violating a particular signaling dimension witness $\mathbf{G}$ and calculating the noisy behavior as

$$\widetilde{\mathbf{P}}_\gamma = (1-\gamma)\mathbf{P} + \gamma\mathbf{W} = \mathbf{W}_\gamma\mathbf{P}. \tag{14.71}$$

It then follows that

$$\left\langle \mathbf{G}, \widetilde{\mathbf{P}}_\gamma \right\rangle = (1-\gamma)\left\langle \mathbf{G}, \mathbf{P}\right\rangle + \gamma\left\langle \mathbf{G}, \mathbf{W}\right\rangle. \tag{14.72}$$

where the first term is the maximal quantum violation and the second is the white noise score $\langle \mathbf{G}, \mathbf{W}\rangle = \sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}} \frac{G_{y,x}}{|\mathcal{Y}|}$ from Lemma 4. Then, if the classical bound and game have been rescaled such that for any $\mathbf{P} \in \mathcal{C}_\Lambda^{\mathcal{X}\xrightarrow{d}\mathcal{Y}}$

$$\beta = 0 \geq \langle \mathbf{G}, \mathbf{P}\rangle \tag{14.73}$$

where the maximal possible score is one, then the critical noise parameter $\gamma_0$ is given by Theorem 1 as

$$\gamma_0 = \frac{\langle \mathbf{G}, \mathbf{P}\rangle}{\langle \mathbf{G}, \mathbf{P}\rangle - \langle \mathbf{G}, \mathbf{W}\rangle}. \tag{14.74}$$

In fig. 14.8, we plot the maximal nonclassicality for entanglement-assisted violations of qubit signaling dimension. Furthermore, in Fig. 14.9 we plot the critical parameters at which nonclassicality is broken. It is important to note that the noise robustness of entanglement-assisted quantum signaling is significantly greater than the entanglement-assisted classical signaling case. Although the white noise scores $\langle \mathbf{G}, \mathbf{W}\rangle$ are the same in both cases, the quantum signaling case saturates the maximal possible score giving it a greater robustness to noise across all signaling dimension witnesses. This improved robustness is reflected by the fact that the Fig. 14.8 shows the EA classical signaling plot on the right to have its nonclassicality broken by $\gamma = 0.2$ while the quantum signaling case shows nonclassicality at significantly greater values of $\gamma$. Other key distinctions show that the most robust signaling dimension violation for the EA classical signaling case is $\mathbf{G}_h$, while the EA quantum signaling case shows this signaling dimension witness to be the least robust to noise. Furthermore, each witness has its nonclassicality broken at a different critical noise parameter, showing that not all tests for nonclassicality have the same noise robustness, even when considering the same quantum resources.

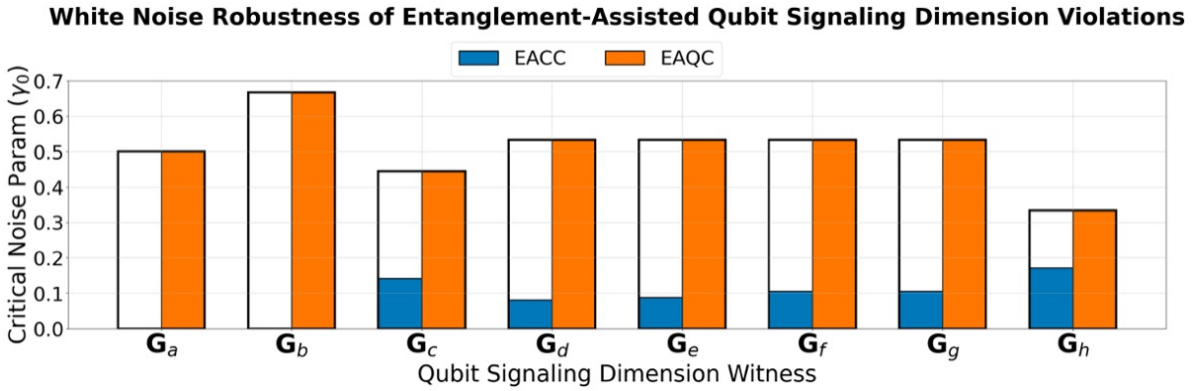



Figure 14.9: The $y$-axis shows the critical noise parameter $\gamma_0$ derived from Eq. (14.74) while the $x$-axis plots each qubit signaling dimension witness. The orange bars show the entanglement-assisted quantum communication (EAQC) critical parameters and the blue bars show the entanglement-assisted classical communication (EACC) critical parameters.

# Chapter 15

# Nonclassicality in Multiple Access Networks

In this chapter we consider nonclassicality in signaling scenarios where multiple senders $A_1, \ldots, A_n$ each send a message to a single receiver $B$. These multi-sender communication networks can be considered as multiple access channel (MAC), hence we refer to these signaling systems as multiple access networks (MANs).

**Definition 102. Multiple Access Network:** $\text{MAN}(\vec{A}, B) = \text{Net}(\vec{N}, \vec{E})$, A DAG describing a communication network that has $n$ independent senders $\vec{A} = \{A_i\}_{i=1}^n$ that signal to a single receiver where

$$\vec{N} = \left(\vec{A}, B\right), \quad \vec{E} = \{A_i \to B\}_{i=1}^n . \tag{15.1}$$

Each sender is given a classical input $x_i \in \mathcal{X}_i$ while the receiver outputs the classical value $y \in \mathcal{Y}$.

In the scientific literature, the communication capacity of multiple access channels has been extensively studied in the classical [363] and quantum settings [364]–[367]. Furthermore, nonclassical behaviors have recently been identified in MAC systems that use either qubit communication between sender and receiver [368], [369] or entanglement between the senders [370]. Hence there is motivation to explore further the nonclassical behaviors in multiple access networks that use quantum resources.

To investigate the nonclassicality in multiple access networks we follow the standard prespriction and derive linear witnesses that bound the classical set of MAN behaviors when the amount of communication between a sender and receiver is limited. Then, we show by examples obtained via variational optimization that application of equivalent quantum resources leads to strongly nonclassical behaviors in MANs. Thus, we derive a collection of operational tests that can be used to certify the quantum resources present in multi-sender communication scenarios.

## *15.1 Multiple Access Network Behaviors*

For simplicity, we focus on a two-sender multiple access networks in which the senders $A_1$ and $A_2$ each accept independent inputs $x_1 \in \mathcal{X}_1$ and $x_2 \in \mathcal{X}_2$, respectively. The receiver $B$ then outputs a classical value $y \in \mathcal{Y}$.

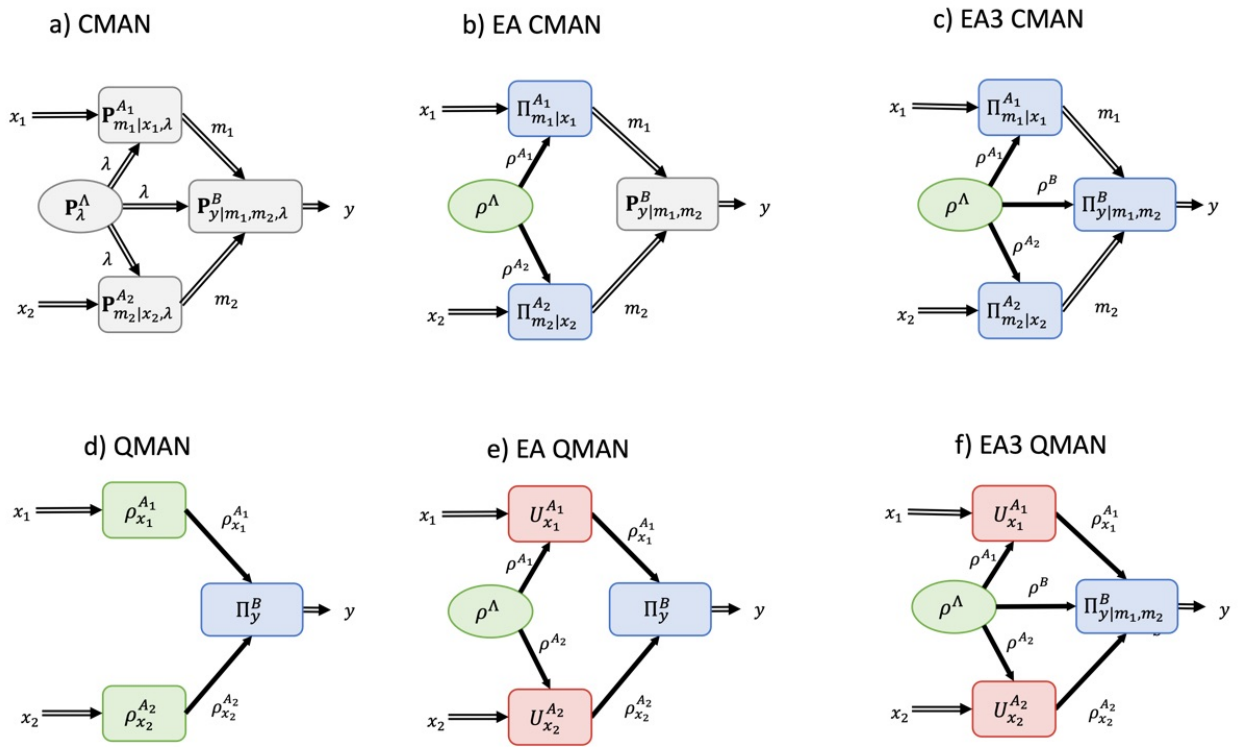



Figure 15.1: **DAGs for MANs with two senders.** Each diagram depicts the causal flow of classical information from inputs $x_1$ and $x_2$ to output $y$. The nodes in the diagram represent devices in the multiple access network and the arrows linking them describe classical communication (double line) or quantum communication (single line). The gray nodes represent classical network devices whereas the colored nodes represent quantum network devices where green nodes are preparation devices, blue nodes are measurement devices, and red nodes are processing devices. Elliptical nodes represent static resources such as classical shared randomness or quantum entanglement. Rectangular nodes represent information processing devices. a) CMAN: a classical MAN where each sender transmits a classical message to the receiver and all devices share randomness. b) EA CMAN: an entanglement-assisted classical MAN where the two senders share entanglement and each send a classical message to the receiver. c) EA3 CMAN: a classical MAN assisted by tripartite entanglement between all devices. d) QMAN: a quantum MAN where each sender transmits a quantum state to the receiver. e) EA QMAN: an entanglement-assisted quantum MAN where the two senders share entanglement. f EA3 QMAN: a quantum MAN where all devices share tripartite entanglement.

Taking a black-box approach, we characterize each multiaccess network by its behavior

$$\mathbf{P}^{\text{MAN}} = \sum_{x_1 \in \mathcal{X}_1} \sum_{x_2 \in \mathcal{X}_2} \sum_{y \in \mathcal{Y}} P^{\text{MAN}}_{y|x_1,x_2}. \tag{15.2}$$

The set of behaviors that can be produced by each respective multiaccess network depend upon the communication resources. Therefore, we describe the behavior decomposition for each respective resource configuration.

**Definition 103.** We define the sets of multiaccess network behaviors $\mathbf{P} \in \mathcal{P}_{\mathcal{Y}|\mathcal{X}_1 \times \mathcal{X}_2}$ for the six resource configurations shown in Fig. 15.1.

a) **Set of Classical MAN Behaviors:** $\mathcal{C}_{\Lambda}^{\text{MAN}}$, The set of all behaviors $\mathbf{P} \in \mathcal{P}_{\mathcal{Y}|\mathcal{X}_1\times\mathcal{X}_2}$ that decompose as

$$\mathbf{P}^{\text{MAN}} = \sum_{\lambda\in\Omega^{\Lambda}} \mathbf{P}_{\lambda}^{B}(\mathbf{P}_{\lambda}^{A_1}\otimes\mathbf{P}_{\lambda}^{A_2})P_{\lambda}^{\Lambda} \tag{15.3}$$

where $P_{\lambda}^{\Lambda} \in [0,1]$, $\mathbf{P}_{\lambda}^{A_1} \in \mathcal{P}_{\mathcal{M}_1|\mathcal{X}_1}$, $\mathbf{P}_{\lambda}^{A_2} \in \mathcal{P}_{\mathcal{M}_2|\mathcal{X}_2}$, and $\mathbf{P}_{\lambda}^{B} \in \mathcal{P}_{\mathcal{Y}|\mathcal{M}_1\times\mathcal{M}_2}$ are the behaviors of the respective network devices.

b) **Set of Entanglement-Assisted Classical MAN Behaviors:** $\mathcal{C}_{\text{EA}}^{\text{MAN}}$, The set of all behaviors $\mathbf{P} \in \mathcal{P}_{\mathcal{Y}|\mathcal{X}_1\times\mathcal{X}_2}$ that decompose as $\mathbf{P}^{\text{MAN}} = \mathbf{P}^{B}\mathbf{P}^{A_1A_2}$ where

$$\mathbf{P}^{A_1A_2} = \sum_{x_1\in\mathcal{X}_1}\sum_{x_2\in\mathcal{X}_2}\sum_{m_1\in\mathcal{M}_1}\sum_{m_2\in\mathcal{M}_2} \text{Tr}\left[(\Pi_{m_1|x_1}^{A_1}\otimes\Pi_{m_2|x_2}^{A_2})\rho^{\Lambda}\right]|m_1,m_2\rangle\langle x_1,x_2| \tag{15.4}$$

and $\rho^{\Lambda} \in D(\mathcal{H}^{A_1}\otimes\mathcal{H}^{A_2})$.

c) **Set of Tripartite Entanglement-Assisted Classical MAN Behavior:** $\mathcal{C}_{\text{EA3}}^{\text{MAN}}$, The set of all behaviors $\mathbf{P} \in \mathcal{P}_{\mathcal{Y}|\mathcal{X}_1\times\mathcal{X}_2}$ whose probabilities decompose as

$$P(y|x_1,x_2) = \sum_{m_1\in\mathcal{M}_1}\sum_{m_2\in\mathcal{M}_2} \text{Tr}\left[(\Pi_{m_1|x_1}^{A_1}\otimes\Pi_{m_2|x_2}^{A_2}\otimes\Pi_{y|m_1,m_2}^{B})\rho^{\Lambda}\right] \tag{15.5}$$

where $\rho^{\Lambda} \in D(\mathcal{H}^{A_1}\otimes\mathcal{H}^{A_2}\otimes\mathcal{H}^{B})$.

d) **Set of Quantum MAN Behaviors:** $\mathcal{Q}^{\text{MAN}}$, The set of all behaviors $\mathbf{P} \in \mathcal{P}_{\mathcal{Y}|\mathcal{X}_1\times\mathcal{X}_2}$ whose probabilities decompose as

$$P(y|x_1,x_2) = \text{Tr}\left[\Pi_{y}^{B}(\rho_{x_1}^{A_1}\otimes\rho_{x_2}^{A_2})\right]. \tag{15.6}$$

e) **Set of Entanglement-Assisted Quantum MAN Behaviors:** $\mathcal{Q}_{\text{EA}}^{\text{MAN}}$, The set of all behaviors $\mathbf{P} \in \mathcal{P}_{\mathcal{Y}|\mathcal{X}_1\times\mathcal{X}_2}$ whose probabilities decompose as

$$P(y|x_1,x_2) = \text{Tr}\left[\Pi_{y}^{B}U_{x_1}^{A_1}\otimes U_{x_2}^{A_2}\rho^{\Lambda}(U_{x_1}^{A_1}\otimes U_{x_2}^{A_2})^{\dagger}\right]. \tag{15.7}$$

f) **Set of Tripartite Entanglement-Assisted Quantum MAN Behaviors:** $\mathcal{Q}_{\text{EA3}}^{\text{MAN}}$, The set of all $\mathbf{P} \in \mathcal{P}_{\mathcal{Y}|\mathcal{X}_1\times\mathcal{X}_2}$ whose probabilities decompose as

$$P(y|x_1,x_2) = \text{Tr}\left[\Pi_{y}^{B}(U_{x_1}^{A_1}\otimes U_{x_2}^{A_2}\otimes\mathbb{I})\rho^{\Lambda}(U_{x_1}^{A_1}\otimes U_{x_2}^{A_2}\otimes\mathbb{I})^{\dagger}\right]. \tag{15.8}$$

In the six resource configuration shown in Fig. 15.1, quantum resources are added incrementallly to the class multiaccess network configuration in Fig. 15.1.a), in which all network devices share aa common randomness source $\Lambda$. Going along the top row, classical communication is used, but the shared randomness is replaced by bipartite entanglement between the senders (Fig. 15.1.b), then by tripartite entanglement between all network devices (Fig. 15.1.c). In the bottom row, the classical communication is replaced with quantum communication in Fig. 15.1.d). Then, the quantum communication is assisted by bipartite and tripartite entanglement in Fig. 15.1.e,f).

## *15.2 Witnessing Nonclassicality in Multiple Access Networks*

Our objective is to quantify and compare the amount of nonclassicality produced in multiaccess networks using quantum resources. To quantify nonclassicality, we apply the standard device-independent approach discussed in Chapter 9. Namely, we follow the approach given in Chapter 4 to enumerate the complete set of deterministic network behaviors $\mathbf{V} \in \mathcal{C}_\Lambda^{\text{MAN}}$ that have 0/1 elements and then we use convex optimization and polyhedral transformations to obtain linear inequalities that bound the multiaccess network polytope $\mathcal{C}_\Lambda^{\text{MAN}}$. Using the variational optimization techniques discussed in Chapter 11 we then find state encodings and measurements that produce strongly nonclassical behaviors $\mathbf{P} \notin \mathcal{C}_\Lambda^{\text{MAN}}$.

### *15.2.1 A Complete Multiaccess Network Polytope*

Consider the multiaccess network where each sender has three inputs $|\mathcal{X}_1| = |\mathcal{X}_2| = 3$, signals to the receiver using a bit or qubit, $|\mathcal{M}_1| = |\mathcal{M}_2| = 2$, while the receiver outputs a binary value $|\mathcal{Y}| = 2$. This example is the minimal example in which strong nonclassicality can be observed.

The set of vertices for the multiaccess network polytope are $\mathcal{V}^{\text{MAN}}$ where $\text{Conv}\left(\mathcal{V}^{\text{MAN}}\right) = \mathcal{C}_\Lambda^{\text{MAN}}$. If each sender uses a bit a communication, then the number of vertices are counted to be

$$
\begin{aligned}
|\mathcal{V}^{\text{MAN}}| = |\mathcal{Y}| + & && \text{(non-communication case)} \\
& + \left\{ {|\mathcal{X}_1| \atop 2} \right\} \binom{|\mathcal{Y}|}{2} 2! + && (|\mathcal{M}_2| = 1,\ |\mathcal{M}_1| = 2 \text{ case}) \\
& + \left\{ {|\mathcal{X}_2| \atop 2} \right\} \binom{|\mathcal{Y}|}{2} 2! + && (|\mathcal{M}_2| = 2,\ |\mathcal{M}_1| = 1 \text{ case}) \\
& + \sum_{m=2}^{\min\{4,|\mathcal{Y}|\}} \left\{ {|\mathcal{X}_1| \atop 2} \right\} \left\{ {|\mathcal{X}_2| \atop 2} \right\} \left( \left\{ {4 \atop m} \right\} - 2 \left\{ {2 \atop m} \right\} \right) \binom{|\mathcal{Y}|}{m} m! && (|\mathcal{M}_1| = 2,\ |\mathcal{M}_2| = 2 \text{ case}).
\end{aligned} \tag{15.9}
$$

where $\left\{ {n \atop k} \right\}$ is Striling's number of the 2nd kind, *i.e.*, the number of ways to partition $n$ elements into $k$ non-empty sets. Note that if $k > n$, then this number is 0 because the partitioning does not exist. Note that the case where senders $A_1$ and $A_2$ each send a bit also contains the cases where $A_1$ or $A_2$ or both do not communicate. Thus, our expression in Eq. (15.9) accounts for each communication individually.

In our example, when $|\mathcal{X}_1| = |\mathcal{X}_2| = 3$ and $|\mathcal{Y}| = 2$, we find 104 unique vertices. For this example, the set of half-space inequalities $\mathcal{G}^{\text{MAN}}$ bounding $\mathcal{C}_\Lambda^{\text{MAN}}$ can be computed efficiently in full using PORTA **XPORTA.jl**, [135]. We find a total of 1230 unique inequalities, however, when relabeling symmetry is taken into account, we find 20 canonical half-space inequalities (See Table 15.1). We will refer to the half-space inequalities by number as $G_j$ where $j$ corresponds to the number in Table 15.1.

Note that inequality $\mathbf{G}_0$ is a trivial bound $1 \geq P(y|x_1, x_2)$, hence, we will not disregard it for the remainder of this work. Furthermore, inequalities $\mathbf{G}_1$ and $\mathbf{G}_2$ are equivalent to the qubit dimensionality witnesses derived in reference [227]. To see equivalence, we note that these inequalities are simply lifted [127], [129] from the scenario where $|\mathcal{X}| = 2$ or $|\mathcal{Y}| = 2$ by adding 3 columns of all zeros to the matrix $\mathbf{G}$. The remainder of the inequalities are novel to the considered multiaccess network scenario.

$$(0)\quad 1 \geq \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \qquad (1)\quad 4 \geq \begin{bmatrix} 1 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 \end{bmatrix}$$

$$(2)\quad 4 \geq \begin{bmatrix} 1 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 \end{bmatrix} \qquad (3)\quad 5 \geq \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 & 0 & 0 & 1 & 0 \end{bmatrix}$$

$$(4)\quad 5 \geq \begin{bmatrix} 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 \\ 1 & 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 \end{bmatrix} \qquad (5)\quad 7 \geq \begin{bmatrix} 0 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 0 \\ 2 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 \end{bmatrix}$$

$$(6)\quad 7 \geq \begin{bmatrix} 0 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 0 \\ 2 & 2 & 0 & 1 & 0 & 0 & 0 & 1 & 0 \end{bmatrix} \qquad (7)\quad 7 \geq \begin{bmatrix} 0 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 0 \\ 2 & 1 & 0 & 2 & 0 & 1 & 0 & 0 & 0 \end{bmatrix}$$

$$(8)\quad 6 \geq \begin{bmatrix} 0 & 0 & 0 & 0 & 1 & 1 & 1 & 0 & 1 \\ 1 & 1 & 0 & 1 & 0 & 0 & 0 & 1 & 0 \end{bmatrix} \qquad (9)\quad 8 \geq \begin{bmatrix} 0 & 0 & 0 & 0 & 1 & 1 & 2 & 0 & 0 \\ 2 & 2 & 0 & 1 & 0 & 0 & 0 & 1 & 0 \end{bmatrix}$$

$$(10)\quad 8 \geq \begin{bmatrix} 0 & 0 & 0 & 0 & 2 & 0 & 1 & 0 & 1 \\ 2 & 2 & 0 & 1 & 0 & 1 & 0 & 0 & 0 \end{bmatrix} \qquad (11)\quad 8 \geq \begin{bmatrix} 0 & 0 & 1 & 0 & 2 & 0 & 0 & 0 & 1 \\ 2 & 1 & 0 & 2 & 0 & 1 & 0 & 0 & 0 \end{bmatrix}$$

$$(12)\quad 8 \geq \begin{bmatrix} 0 & 0 & 1 & 0 & 2 & 0 & 0 & 0 & 1 \\ 2 & 1 & 0 & 2 & 0 & 0 & 0 & 1 & 0 \end{bmatrix} \qquad (13)\quad 8 \geq \begin{bmatrix} 0 & 0 & 1 & 0 & 1 & 0 & 2 & 0 & 0 \\ 2 & 1 & 0 & 0 & 0 & 1 & 0 & 2 & 0 \end{bmatrix}$$

$$(14)\quad 10 \geq \begin{bmatrix} 0 & 0 & 1 & 0 & 2 & 0 & 1 & 0 & 1 \\ 3 & 2 & 0 & 2 & 0 & 1 & 0 & 0 & 0 \end{bmatrix} \qquad (15)\quad 10 \geq \begin{bmatrix} 0 & 0 & 1 & 0 & 2 & 0 & 1 & 0 & 1 \\ 3 & 2 & 0 & 2 & 0 & 0 & 0 & 1 & 0 \end{bmatrix}$$

$$(16)\quad 11 \geq \begin{bmatrix} 0 & 0 & 2 & 0 & 1 & 0 & 2 & 0 & 0 \\ 3 & 1 & 0 & 1 & 0 & 2 & 0 & 2 & 0 \end{bmatrix} \qquad (17)\quad 11 \geq \begin{bmatrix} 0 & 0 & 2 & 0 & 1 & 0 & 2 & 0 & 0 \\ 3 & 1 & 0 & 1 & 0 & 2 & 0 & 2 & 0 \end{bmatrix}$$

$$(18)\quad 16 \geq \begin{bmatrix} 0 & 0 & 2 & 0 & 3 & 0 & 2 & 0 & 1 \\ 5 & 3 & 0 & 3 & 0 & 1 & 0 & 1 & 0 \end{bmatrix} \qquad (19)\quad 17 \geq \begin{bmatrix} 0 & 0 & 2 & 1 & 2 & 0 & 5 & 0 & 1 \\ 4 & 2 & 0 & 0 & 0 & 1 & 0 & 4 & 0 \end{bmatrix}$$

Table 15.1: **Canonical Multiaccess Network Nonclassicality Witnesses.** The 20 canonical witness of nonclassicality are listed in the format $\beta \geq \mathbf{G}$. Note that $\mathbf{G}$ is a $2 \times 9$ matrix reflecting the 2 classical outputs and $3 \times 3$ classical inputs.

### *15.2.2 The Finger Printing Inequality*

The inequalities in Table 15.1 do not admit a simple operational interpretation. Hence we introduce a finger printing game in which the objective is for the multiaccess network to output $y = 0$ when $x_1 = x_2$ and $y = 1$ otherwise. We encode this task into the following linear inequality which we call the *finger-printing inequality*, $\mathbf{G}_{FP}$:

$$\mathbf{G}_{FP} := 7 \geq \begin{bmatrix} 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 \\ 0 & 1 & 1 & 1 & 0 & 1 & 1 & 1 & 0 \end{bmatrix}. \tag{15.10}$$

In the finger-printing game a score of 1 is rewarded for a correct answer and 0 otherwise. The classical bound is 7 and can be achieved by two distinct deterministic classical strategies

$$\mathbf{P}_{\text{MAN}} \in \left\{ \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \end{bmatrix}, \quad \begin{bmatrix} 1 & 0 & 0 & 0 & 1 & 1 & 0 & 1 & 1 \\ 0 & 1 & 1 & 1 & 0 & 0 & 1 & 0 & 0 \end{bmatrix} \right\}. \tag{15.11}$$

In each case, the sender $A_1$ and $A_2$ each send $m_1 = m_2 = 0$ for the inputs $x_1 = 0$ and $x_2 = 0$ respectively. Hence the receiver knows with certainty if $x_1 = x_2 = 0$. The rest of the behavior is implemented either by fixing the output to be $y = 1$ for all $m_1 \neq 0$ or $m_2 \neq 0$ or by inverting sender $A_2$'s bit if sender $A_1$ sends a

1. In total, there are 6 such classical strategies which can achieve a score of 7 against the finger-printing inequality. Since the Dim $\left(\mathcal{C}_{\Lambda}^{\text{MAN}}\right) = 9$ when $|\mathcal{X}_1| = |\mathcal{X}_2| = 3$ and $|\mathcal{Y}| = 2$, the finger printing inequality does not designate a tight nonclassicality witness. In other words, $(7, \mathbf{G}_{FP})$ is not a facet inequality of $\mathcal{C}_{\Lambda}^{\text{MAN}}$. Nevertheless, a violation $\langle \mathbf{G}_{FP}, \mathbf{P} \rangle > 7$ witnesses strong nonclassicality.

## *15.3 Strong Nonclassicality in Multiaccess Networks*

Our main result is that all quantum resource configurations in Fig. 15.1 can produce strongly nonclassical behaviors. We prove this by example for the complete set of nonclassicality witnesses derived for the multiaccess network that has $|\mathcal{X}_1| = |\mathcal{X}_2| = 3$ and $|\mathcal{Y}| = 2$. Then, for more general multiaccess network scenarios we apply derived nonclassicality witnesses.

### *15.3.1 Variational Optimization of Multiaccess Networks*

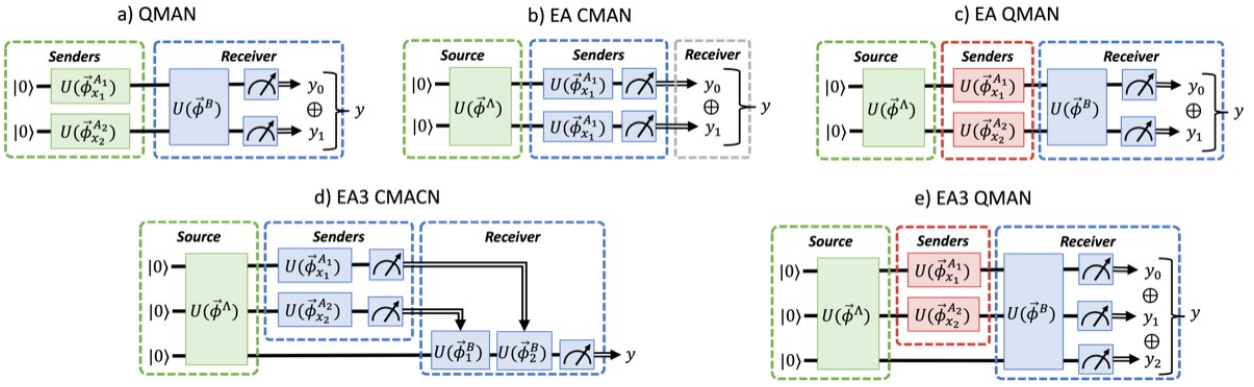



Figure 15.2: **Variational Ansatz Circuits for Multiaccess Networks.** Each diagram depicts a variational ansatz circuit for the MAN DAGs in Fig. 15.1. All gates are fully parameterized arbitrary unitary operators acting on the shown qubits. The operations of preparation devices are green, processing devices are red, measurement devices are blue, and classical devices are gray.

To find strongly nonclassical behaviors, we optimize general variational ansatzes that correspond to each of the quantum resource configurations in Fig. 15.1. These ansatzes are shown in Fig. 15.2. The optimization objective is expressed as

$$\min_{\Theta} \text{Cost}\,(\theta) = -\max_{\Theta} \langle \mathbf{G}, \mathbf{P}(\Theta) \rangle \tag{15.12}$$

where $\mathbf{G}$ is a linear nonclassicality witness, $\Theta \in \mathbb{R}^m$ is a collection of settings for the multiaccess network variational ansatz circuits, and $\mathbf{P}(\Theta)$ is the evaluated behavior for the given settings.

The general variational optimization procedure is given in Chapter 11. Namely, we use a standard gradient descent algorithm to minimize the cost function Eq. (15.12). We evaluate the quantum circuits in Fig. 15.2 on a classical simulator allowing us to use backpropagation to calculate the gradients of the cost function with efficiency. In general, however, our approach could be applied on quantum hardware.

It is important to remarks that there is no guarantee that the global optimum will be found. For this reason, it is best practice to repeat the optimization with randomized initial settings. We argue that the optimized nonclassical behaviors are likely approximately optimal because in all instances of known maximal violations, our variational optimizations have found the global optimum. Nevertheless, when the variational

optimization algorithm obtains a violation of the inequality $\langle \mathbf{G}, \mathbf{P} \rangle \leq \gamma$, the violation confirms the presence of strongly nonclassical behaviors. Thus, we identify a broad range of novel nonclassical behaviors for multiaccess networks using quantum resources.

### *15.3.2 Quantum Nonclassicality in the* $(3,3) \overset{(2,2)}{\succ} 2$ *Multiaccess Network*

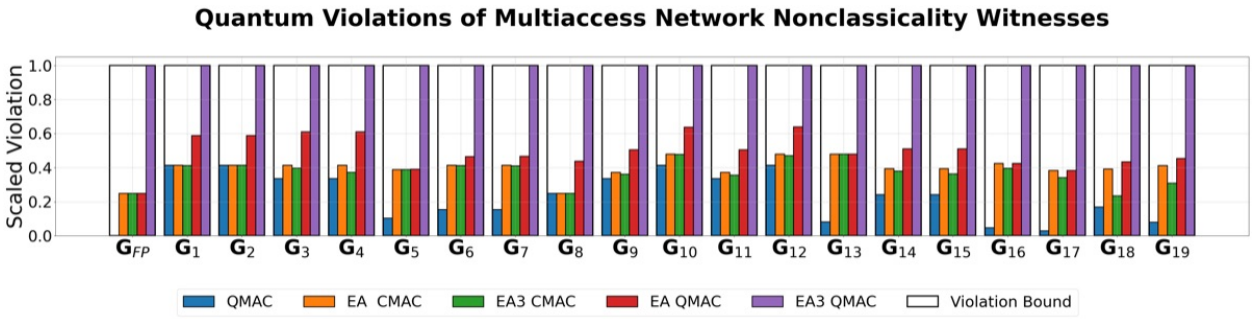



Figure 15.3: The five quantum ansatzes of Fig. 15.2 are optimized against the finger-printing inequality $I_{FP}$ and the 19 nontrivial inequalities bounding the the multiaccess network in Table. 15.1. The amount of violation is plotted on the $y$-axis as $(\langle \mathbf{G}|\mathbf{P}^\star\rangle - \gamma)/(\gamma^\star - \gamma)$ where $\gamma^\star$ is the maximal possible score for the black-box game $(\gamma, \mathbf{G})$ and $\langle \mathbf{G}|\mathbf{P}^\star\rangle$ is the maximal score obtained via optimization.

When the variational ansatzes in Fig. 15.2 are optimized we find that all quantum resource configurations violate nonclassicality witnesses in Table 15.1. The maximal violations obtained are plotted in Fig. 15.3. In all cases, the entanglement-assisted multiaccess networks achieved larger violations than the unassisted quantum multiaccess network. Furthermore, we find that the tripartite entanglement-assisted quantum multiaccess network is able to $\gamma^\star$ the maximal possible violation for each respective black-box game $\mathbf{G}_j$. As a result, we assert that the following general hierarchy likely holds,

$$\mathcal{C}_\Lambda^{\text{MAN}} \subset \mathcal{Q}^{\text{MAN}} \subset \mathcal{C}_{\text{EA}}^{\text{MAN}} \subseteq \mathcal{C}_{\text{EA3}}^{\text{MAN}} \subset \mathcal{Q}_{\text{EA}}^{\text{MAN}} \subseteq \mathcal{Q}_{\text{EA3}}^{\text{MAN}} \tag{15.13}$$

where shared randomness is implicitly assumed in each set to maintain convexity. Indeed, it is possible that the set inclusions are not so simple such that the set is not contained in its entirety, however, the violations demonstrate that each respective resource configuration is insufficient for simulating all behaviors in its containing set.

Additionally, unassisted quantum multiaccess network are found to be insufficient to violate the finger printing inequality $\mathbf{G}_{FP}$. Unfortunately, our numerics only lead to the conjecture that entanglement is required between the senders to violate the finger printing inequality. However, if this conjecture holds, then the finger-printing inequality can serve as a test for entanglement between the two senders.

In the case of entnaglement-assisted classical multiaccess networks, we can describe an encoding and decoding strategy that achieves the maximal violation. Indeed, we find that $\langle \mathbf{G}_{FP}|\mathbf{P}^\star\rangle \leq 7.5$ for $\mathbf{P}^\star \in \mathcal{C}_{\text{EA}}^{\text{MAN}}$ where

$$\mathbf{P}^\star = \begin{bmatrix} 1 & 0.25 & 0.25 & 0.25 & 1 & 0.25 & 0.25 & 0.25 & 1 \\ 0 & 0.75 & 0.75 & 0.75 & 0 & 0.75 & 0.75 & 0.75 & 0 \end{bmatrix}. \tag{15.14}$$

This behavior is achieved using the maximally entangled state preparation $|\Phi^+\rangle = \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle)$ followed

by projective measurements in the following local qubit bases,

$$\left\{ \left|\phi_{0|0}\right\rangle = \left|0\right\rangle, \quad \left|\phi_{1|0}\right\rangle = \left|1\right\rangle \right\}, \tag{15.15}$$

$$\left\{ \left|\phi_{0|1}\right\rangle = \frac{1}{2}\left|0\right\rangle + \frac{\sqrt{3}}{2}\left|1\right\rangle, \quad \left|\phi_{1|1}\right\rangle = \frac{\sqrt{3}}{2}\left|0\right\rangle - \frac{1}{2}\left|1\right\rangle \right\}, \quad \text{and} \tag{15.16}$$

$$\left\{ \left|\phi_{0|2}\right\rangle = \frac{1}{2}\left|0\right\rangle - \frac{\sqrt{3}}{2}\left|1\right\rangle, \quad \left|\phi_{1|2}\right\rangle = \frac{\sqrt{3}}{2}\left|0\right\rangle + \frac{1}{2}\left|1\right\rangle \right\}. \tag{15.17}$$

Note that if $x_1 = x_2$, then senders $A_1$ and $A_2$ measure $|\Phi^+\rangle$ in the same basis resulting in a bitstring $m_1 m_2 \in \{00, 11\}$ with certainty. Otherwise, if $x_1 \neq x_2$, then,

$$P(m_1, m_2 \in \{00, 11\} | x_1, x_2) = \left| \langle \phi^{A_1}_{m_1|x_1}, \phi^{A_2}_{m_2|x_2} | \Phi^+ \rangle \right|^2 = 1/4, \quad \text{and} \tag{15.18}$$

$$P(m_1 m_2 \in \{01, 10\} | x_1, x_2) = \left| \langle \phi^{A_1}_{m_1|x_1}, \phi^{A_2}_{m_2|x_2} | \Phi^+ \rangle \right|^2 = 3/4. \tag{15.19}$$

Finally, the receiver outputs $y = 0$ when the received classical bits $m_1 m_2 \in \{00, 11\}$ and otherwise, outputs $y = 1$. With some calculation it can be verified that the Bell state preparation and the bases in Eqs. (15.15) through (15.17) produce the behavior matrix in Eq. (15.14).

### *15.3.3 Noisy Multiple Access Networks*

In this section, we investigate the noise robustness of the nonclassicality violations obtained in the previous section 15.3.2. As a worst-case example, we consider depolarizing white noise applied to receiver's measurement as $\mathcal{W}_{(1-\gamma)}(\Pi^B_y) = (1-\gamma)\Pi^B_y + \frac{\gamma}{2}\mathbb{I}_2$, where $\Pi^B_y = (U^B)^\dagger |y\rangle\langle y| U^B$ where $\Pi^B_y$ is the receiver's two-outcome measurement, which we assume to measure the parity of each measured qubit as denoted in Fig. 15.2. Using Proposition 3 we can model the depolarizing noise as a classical postprocessing operation on the multiaccess network behavior where the noisy behavior is $\widetilde{\mathbf{P}}_\gamma = (1-\gamma)\mathbf{P} + \gamma\mathbf{W}$ and $\mathbf{W} = \sum_{x_1,x_2 \in \mathcal{X}} \sum_{\vec{y} \in \mathcal{Y}} \frac{1}{|\mathcal{Y}|} |y\rangle\langle x_1, x_2|$. Then we apply Theorem 1 to derive the critical noise parameter at which nonclassicality is broken to be

$$\gamma_0 = \frac{\langle \mathbf{G}, \mathbf{P} \rangle}{\langle \mathbf{G}, \mathbf{P} \rangle - \langle \mathbf{G}, \mathbf{W} \rangle} \tag{15.20}$$

where $\mathbf{P}$ is the maximum violation of the classical bound obtained through variational optimization.

We find that our noise model for depolarizing noise on the measurement also describes depolarizing noise on the entanglement source in in scenarios Fig. 15.2.b,c,e). That is, since all qubits in the ansatz circuit are prepared at the same source and measured at the same measurement device, it doesn't matter whether the depolarizing noise is applied to the produced entangled state $\mathcal{W}_{(1-\gamma)}(\rho^\Lambda)$ or to the joint qubit measurement $\mathcal{W}_{(1-\gamma)}(\Pi^B_y)$. The main condition being that the measurement is projective and $\mathrm{Tr}\left[\Pi^B_y\right] = 1$ for all $y$.

We plot our results in Fig. 15.4 showing that noise robustness is unique for each nonclassicality witness. Furthermore, resources that achieve a larger nonclassicality score with respect to other resource configurations also have a larger critical parameter, indicating greater robustness. As shown in Fig. 15.5, we plot the linear curve that the noisy behavior $\widetilde{\mathbf{P}}_\gamma$ follows as $\gamma$ increases. The $y$-intercept of the line depends on the maximal score for the given nonclassicality witness, while the rate of decrease depends on the white noise score $\langle \mathbf{G}, \mathbf{W} \rangle$.

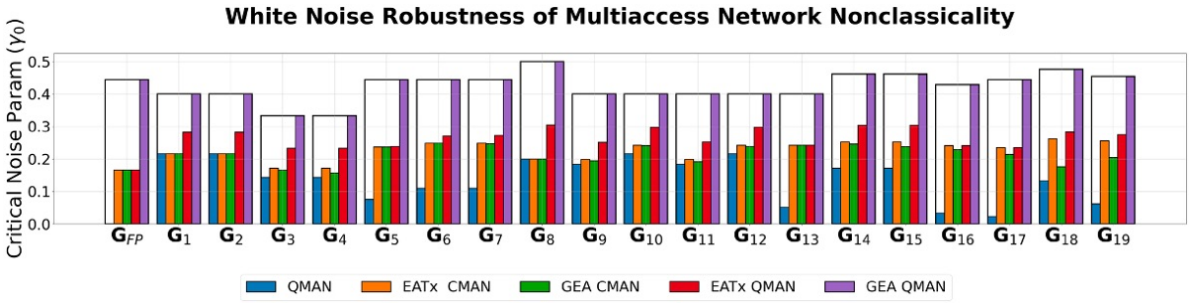



Figure 15.4: For each qubit signaling dimension witness, we plot the critical noise parameter $\gamma_0$ as calculated from Eq. (15.20). From left to right the colored bars show the multiaccess network quantum resource configurations: unassisted quantum signaling (blue), entanglement-assisted classical senders (orange), global entanglement-assisted classical signaling (green), entanglement-assisted quantum senders (red), and global entanglement-assisted quantum senders (purple).

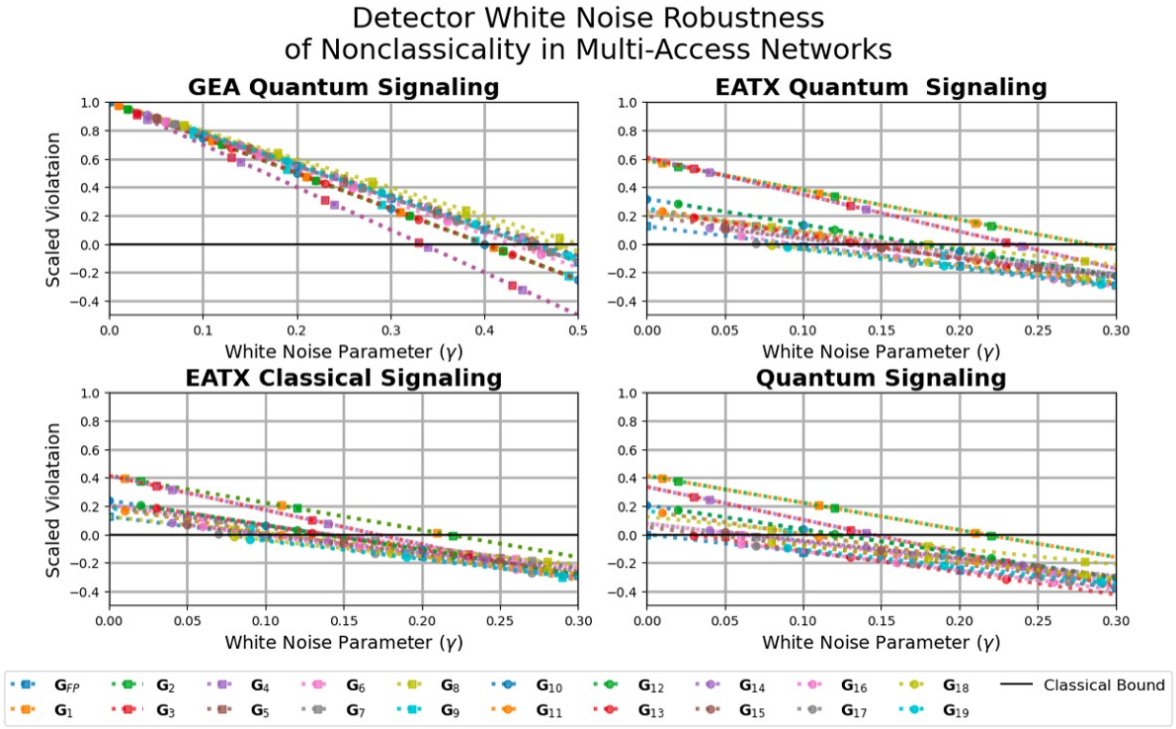



Figure 15.5: The detector white noise robustness of nonclassicality is plotted for the multiaccess network having $|\mathcal{X}_1| = |\mathcal{X}_2| = 3$, $|\mathcal{Y}| = 2$, and $|\mathcal{M}_1| = |\mathcal{M}_2| = 2$. For each signaling dimension witness, we plot the curve $\widetilde{\mathbf{P}}_\gamma = (1-\gamma)\mathbf{P} + \gamma\mathbf{W}$ where $\mathbf{W} = \sum_{x,y} \frac{1}{|\mathcal{Y}|} |y\rangle\langle x|$ is a white noise behavior and $\mathbf{P}$ is the optimal behavior for the specified resources. Each colored line and marker combination corresponds to a different nonclassicality witness where the markers are aesthetic for distinguishing overlapping lines. The classical violations are scaled such that the maximum possible violation is one where zero is the classical bound. (Top Left) The noise robustness of global entanglement-assisted quantum signaling. (Top Right) The noise robustness of entanglement-assisted senders using quantum signaling. (Bottom Left) The noise robustness of entanglement-assisted senders using classical signaling. (Bottom Right) The noise robustness of quantum signaling.

# Chapter 16

# Nonclassicality in Broadcast Networks

In this section we consider another important networking architecture known as broadcast networks. In this network, a single sender signals to many independent receiver devices.

**Definition 104. Broadcast Network:** $\text{BCN}(A, \vec{B}) = \text{Net}((A, \vec{B}), \{A \to B_i\}_{i=1})^{|\vec{B}|}$, A DAG consisting of a sender device $A$ and a collection of receivers $\vec{B}$. The receiver is given a classical input $x \in \mathcal{X}$ and the $i^{th}$ receiver outputs a classical value $y_i \in \mathcal{Y}_i$. The signaling dimension of the channel connecting the sender to the $i^{th}$ receiver is bounded as $|\mathcal{M}_i|$.

## *16.1 Broadcast Network Behaviors*

In general, a broadcast network behavior $\mathbf{P}^{\text{BCN}}$ with $n$ receivers belongs to the full probability polytope $\mathcal{P}_{\mathcal{Y}_1 \times \cdots \times \mathcal{Y}_n | \mathcal{X}}$ where additional constraints on the behaviors are imposed by the amount of communication between the sender and each receiver. For simplicity, we consider broadcast networks with two receivers and consider a range of quantum resource configurations that can be used to assist the broadcast (see Fig. 16.1).

**Definition 105.** We define the sets of broadcast network behaviors $\mathbf{P} \in \mathcal{P}_{\mathcal{Y}_1 \times \mathcal{Y}_2 | \mathcal{X}}$ for the six resource configurations in Fig. 16.1).

a) **Set of Classical Broadcast Behaviors:** $\mathcal{C}^{\text{BCN}}$, The set of all behaviors that decompose as $\mathbf{P} = (\mathbf{P}^{B_1} \otimes \mathbf{P}^{B_2})\mathbf{P}^{A}$ where $\mathbf{P}^{A} \in \mathcal{P}_{\mathcal{M}_1 \times \mathcal{M}_2 | \mathcal{X}}$, $\mathbf{P}^{B_1} \in \mathcal{P}_{\mathcal{Y}_1 | \mathcal{M}_1}$, and $\mathbf{P}^{B_2} \in \mathcal{P}_{\mathcal{Y}_2 | \mathcal{M}_2}$ are the respective behaviors for each device.

b) **Set of Global Shared Randomness Assisted-Classical Broadcast Behaviors:** $\mathcal{C}^{\text{BCN}}_{\Lambda}$, The set of all behaviors that decompose as

$$\mathbf{P} = \sum_{\lambda \in \Omega^{\Lambda}} (\mathbf{P}^{B_1}_{\lambda]} \otimes \mathbf{P}^{B_2}_{\lambda}) \mathbf{P}^{A}_{\lambda} P^{\Lambda}_{\lambda} \tag{16.1}$$

where $P^{\Lambda}_{\lambda} \in [0, 1]$ and $\sum_{\lambda} P^{\Lambda}_{\lambda} = 1$, $\mathbf{P}^{A}_{\lambda} \in \mathcal{P}_{\mathcal{M}_1 \times \mathcal{M}_2 | \mathcal{X}}$, $\mathbf{P}^{B_1}_{\lambda} \in \mathcal{P}_{\mathcal{Y}_1 | \mathcal{M}_1}$, and $\mathbf{P}^{B_2}_{\lambda} \in \mathcal{P}_{\mathcal{Y}_2 | \mathcal{M}_2}$ are the respective behaviors for each device.

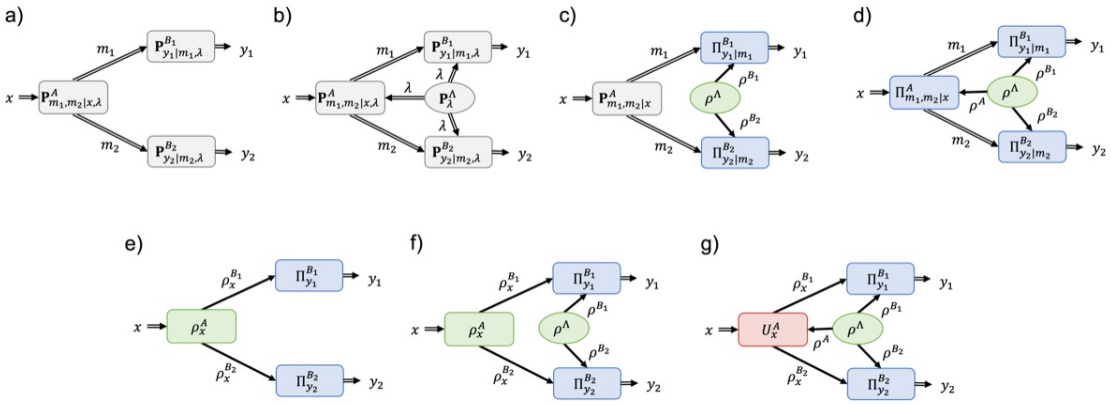



Figure 16.1: **Broadcast Network Resource Configurations:** Directed acyclic graphs depicting different resource configurations for a broadcast network with two receivers. Classical communication is depicted by double-lined arrows and quantum communication is depicted by single-lined arrows. Elliptical nodes show static resources while rectangles show network devices. In the top row classical broadcast networks are shown and in the bottom row, quantum broadcast networks are shown. a) Unassisted classical broadcast. b) Classical broadcast assisted by shared randomness. c) Classical broadcast with entanglement-assisted receivers. d) Classical broadcast assisted by entanglement between all devices. e) Quantum broadcast. f) Quantum broadcast with entanglement-assisted receivers. g) Quantum broadcast assisted by entanglement between all devices.

c) **Set of Entanglement-Assisted Receiver Classical Broadcast Behaviors:** $\mathcal{C}_{\text{EA}}^{\text{BCN}}$, The set of all behaviors that decompose as $\mathbf{P} = \mathbf{P}^{B_1 B_2} \mathbf{P}^{A}$ where

$$\mathbf{P}^{B_1 B_2}_{y_1,y_2|m_1,m_2} = \text{Tr}\left[(\Pi^{B_1}_{y_1|m_1} \otimes \Pi^{B_2}_{y_2|m_2})\rho^{\Lambda}\right]. \tag{16.2}$$

d) **Set of Global Entanglement-Assisted Classical Broadcast Behaviors:** $\mathcal{C}_{\text{EA3}}^{\text{BCN}}$, The set of all behaviors that decompose as

$$\mathbf{P}_{y_1,y_2|x} = \sum_{m_1 \in \mathcal{M}_1} \sum_{m_2 \in \mathcal{M}_2} \text{Tr}\left[(\Pi^{B_1}_{y_1|m_1} \otimes \Pi^{B_2}_{y_2|m_2} \otimes \Pi^{A}_{m_1,m_2|x})\rho^{\Lambda}\right]. \tag{16.3}$$

e) **Set of Unassisted Quantum Broadcast Behaviors:** $\mathcal{Q}^{\text{BCN}}$, The set of all behaviors that decompose as

$$\mathbf{P}_{y_1,y_2|x} = \text{Tr}\left[(\Pi^{B_1}_{y_1} \otimes \Pi^{B_2}_{y_2})\rho^{A}_{x}\right]. \tag{16.4}$$

f) **Set of Entanglement-Assisted Receiver Quantum Broadcast Behaviors:** $\mathcal{Q}_{\text{EA}}^{\text{BCN}}$, The set of all behaviors that decompose as

$$\mathbf{P}_{y_1,y_2|x} = \text{Tr}\left[(\Pi^{B_1}_{y_1} \otimes \Pi^{B_2}_{y_2})(\rho^{A}_{x} \otimes \rho^{\Lambda})\right]. \tag{16.5}$$

g) **Set of Global Entanglement-Assisted Quantum Broadcast Behaviors:** $\mathcal{Q}_{\text{EA3}}^{\text{BCN}}$, The set of all behaviors that decompose as

$$\mathbf{P}_{y_1,y_2|x} = \text{Tr}\left[(\Pi^{B_1}_{y_1} \otimes \Pi^{B_2}_{y_2})(U^{A}_{x} \otimes \mathbb{I}^{B_1 B_2})\rho^{\Lambda}(U^{A}_{x} \otimes \mathbb{I}^{B_1 B_2})^{\dagger}\right] \tag{16.6}$$

To show how these behavior decompositions generalize to arbitrary numbers of receivers, we provide as an example the construction for unassisted classical and quantum broadcasts. For a general classical broadcast network with $n$ receivers, the behavior is constructed as

$$\mathbf{P}^{\mathrm{BCN}} = \left( \bigotimes_{i=1}^{n} \mathbf{P}^{B_i} \right) \mathbf{P}^{A} \tag{16.7}$$

where $\mathbf{P}^{B_i} \in \mathcal{P}_{\mathcal{Y}_i|\mathcal{M}_i}$ and $\mathbf{P}^{A} \in \mathcal{P}_{\mathcal{M}_1 \times \cdots \times \mathcal{M}_n|\mathcal{X}}$. Similarly the quantum broadcast behavior with $n$ receivers decomposes as

$$\mathbf{P}^{\mathrm{BCN}}_{\vec{y}|x} = \mathrm{Tr}\left[ \left( \bigotimes_{i=1}^{n} \Pi^{B_i}_{y_i} \right) \rho^{A}_{x} \right] \tag{16.8}$$

where $\Pi^{B_i}_{y_i} \in \mathrm{POVM}(\mathcal{H}^{B_i}_{|\mathcal{M}_i|})$ and $\rho^{A} \in D(\bigotimes_{i=1}^{n} \mathcal{H}_{|\mathcal{M}_i|})$. Entanglement and shared randomness-assisted broadcasts can be generalized accordingly using the DAG structure the tensor calculi for constructing the behaviors of classical, quantum, and LOCC networks. However, the complexity of different resource configurations is too large and should be handled only as needed.

## *16.2 Bounds on the Behaviors of Classical Broadcasts*

In this section we derive bounds on the shared random-assisted classical broadcast behaviors in both general and specific cases. We begin with some trivial bounds describing when broadcast networks can achieve the full probability polytope $\mathcal{P}_{\mathcal{Y}_1 \times \ldots \mathcal{Y}_n|\mathcal{X}}$, then we explore examples in which more restrictive bounds can be derived. Since we permit an unlimited amount of global shared randomness in our classical model, the classical bounds take the form of linear black-box games that can serve as witnesses of strong nonclassicality.

**Lemma 14.** Consider a broadcast network with $n$ receivers. If $|\mathcal{Y}_i| = |\mathcal{M}_i|$ for all $i \in \{1, \ldots, n\}$, then $\mathcal{C}^{\mathrm{BCN}} = \mathcal{C}^{\mathrm{BCN}}_{\Lambda} = \mathcal{P}_{\mathcal{Y}_1 \times \cdots \times \mathcal{Y}_n|\mathcal{X}}$ for any input alphabet $\mathcal{X}$. Hence all broadcasts using quantum resources can be simulated exactly using unassisted classical communication.

*Proof.* For all $i \in \{1, \ldots, n\}$, let $|\mathcal{Y}_i| = |\mathcal{M}_i|$ and let the device $B_i$ output the received message such that $y_i = m_i \in \mathcal{M}_i$. That is, the receivers do not operate upon their received data and the broadcast network reduces to a single black box $\mathbf{P}^{\mathrm{BCN}} : \mathcal{X} \to \mathcal{Y}_1 \times \cdots \times \mathcal{Y}_n$. The behavior $\mathbf{P}^{\mathrm{BCN}} \in \mathcal{P}_{\mathcal{Y}_1 \times \cdots \times \mathcal{Y}_n}$ is unrestricted and exhibit any behavior in the full probability polytope. Therefore, $\mathcal{C}^{\mathrm{BCN}} = \mathcal{P}_{\mathcal{Y}_1 \times \cdots \times \mathcal{Y}_n}$ and since all quantum resource assisted broadcasts are contained by the probability polytope, the unassisted classical broadcast $\mathcal{C}^{\mathrm{BCN}}$ can simulate all possible broadcast behaviors with zero error. □

**Lemma 15.** Consider a broadcast network with $n$ receivers. If $|\mathcal{X}| \leq \min\{|\mathcal{M}_i|\}_{i=1}^{n}$, then $\mathcal{C}^{\mathrm{BCN}} \subseteq \mathcal{Q}^{\mathrm{BCN}} \subseteq \mathcal{C}^{\mathrm{BCN}}_{\Lambda} = \mathcal{P}_{\mathcal{Y}_1 \times \cdots \times \mathcal{Y}_n}$ for output alphabets sizes of arbitrary size. Hence any entanglement-assisted quantum or classical broadcast behavior can be simulated exactly using classical communication assisted by an unlimited amount of shared randomness.

*Proof.* If $|\mathcal{X}| \leq \min\{|\mathcal{M}_i|\}_{i=1}^{n}$, then the entire input can be communicated to each receiver. Since each receiver knows the entire input $x \in \mathcal{X}$ the collection of receivers $\vec{B}$ can produce any deterministic behavior $\mathbf{V} \in \mathcal{P}_{\mathcal{Y}_1 \times \cdots \times \mathcal{Y}_n|\mathcal{X}}$. Therefore, if the receivers share an unlimited amount of randomness, then they can construct any behavior $\mathbf{P} \in \mathcal{P}_{\mathcal{Y}_1 \times \cdots \times \mathcal{Y}_n|\mathcal{X}}$. Thus, it must hold that $\mathcal{C}^{\mathrm{BCN}}_{\Lambda} = \mathcal{P}_{\mathcal{Y}_1 \times \cdots \times \mathcal{Y}_n|\mathcal{X}}$. Now suppose that the receivers do not share randomness and $|\mathcal{Y}_i| > |\mathcal{M}_i|$ for one or more receivers. It follows that set of behaviors

becomes nonconvex such that $\mathcal{C}^{\text{BCN}} \subset \mathcal{C}_{\Lambda}^{\text{BCN}}$. For example, consider the case where $|\mathcal{M}_1| = |\mathcal{M}_2| = |\mathcal{X}|$, $|\mathcal{Y}_1| = 3$, and $|\mathcal{Y}_2| = 2$ and the receivers are tasked with outputting $y_1 = y_2 = 0$ when $x = 0$ and when $x = 1$, $y_1 = y_2 = 1$ with probability $1/2$ while also $y_1 = 2$ and $y_1 = 0$ with probability $1/2$ such that $P(1,1|1) = P(2,0|1) = \frac{1}{2}$. Since the receivers do not share randomness, they must independently guess the output $y_1 \in \{1,2\}$ and $y_2 \in \{0,1\}$ when the observe $x = 1$. However, they are unable to coordinate such that $P(1,1|1) = P(2,0|1) = 1/2$ with zero error because receiver $B_2$ does not know $B_1$'s output. Therefore, $\mathcal{C}^{\text{BCN}}$ is a strict subset of $\mathcal{C}_{\Lambda}^{\text{BCN}}$. It follows directly that the hierarchy $\mathcal{C}^{\text{BCN}} \subseteq \mathcal{Q}^{\text{BCN}} \subseteq \mathcal{C}_{\Lambda}^{\text{BCN}} = \mathcal{P}_{\mathcal{Y}_1 \times \cdots \times \mathcal{Y}_n}$ holds generally where $\mathcal{C}^{\text{BCN}} = \mathcal{C}_{\Lambda}^{\text{BCN}}$ only occurs in special cases when no receivers have $|\mathcal{Y}_i| > |\mathcal{M}_i|$. □

### *16.2.1 Complete Broadcast Network Polytopes*

In this section, we discuss general and specific cases where we can define complete broadcast network polytopes $\mathcal{C}_{\Lambda}^{\text{BCN}}$ that are not equivalent to the full probability polytope. Hence we obtain complete sets of facet inequalities that bound the set of shared randomness-assisted classical broadcast behaviors. To begin we show that all bounds for the signaling polytopes discussed in Chapter 14 also bound broadcast networks.

**Theorem 14.** Consider the $n$-receiver broadcast network where $\min\{|\mathcal{X}|, |\mathcal{Y}_1|\} \geq |\mathcal{M}_1|$ and for $i > 1$, $|\mathcal{M}_i| = |\mathcal{Y}_i|$. It follows that the broadcast network polytope $\mathcal{C}_{\Lambda}^{\text{BCN}}$ is lifted from the signaling polytope $\mathcal{C}_{\Lambda}^{\mathcal{X} \xrightarrow{d} \mathcal{Y}}$ with $\mathcal{Y} = \mathcal{Y}_1$ where the set of generator facet inequalities $\mathcal{G}^{\text{BCN}}$ bounding $\mathcal{C}_{\Lambda}^{\text{BCN}}$ are constructed from the inequalities $\mathcal{G}^{\mathcal{X} \xrightarrow{d} \mathcal{Y}}$ bounding $\mathcal{C}_{\Lambda}^{\mathcal{X} \xrightarrow{d} \mathcal{Y}}$ as

$$\mathbf{G}^{\text{BCN}} = \mathbf{G}^{\mathcal{X} \xrightarrow{d} \mathcal{Y}} \otimes \left( \bigotimes_{i=2}^{n} \left| \vec{1}_i \right\rangle \right) \tag{16.9}$$

where $\left| \vec{1}_i \right\rangle = (1, \ldots, 1)^T$ and $\left| \vec{1}_i \right\rangle$ has length $|\mathcal{M}_i|$.

*Proof.* Note that for $i > 1$ $|\mathcal{Y}_i| = |\mathcal{M}_i|$, thus following the arguments in Lemma 14, the partial broadcast network $\text{BCN}(A, \{B_i\}_{i=2}^n)$ can exhibit any behavior in the probability polytope $\mathcal{P}_{\mathcal{Y}_2 \times \cdots \times \mathcal{Y}_n | \mathcal{X}}$, which excludes the receiver $B_1$. Therefore, any restrictions imposed on the full broadcast network must be induced by the fact that $\min\{|\mathcal{X}|, |\mathcal{Y}_1|\} > |\mathcal{M}_1|$. This fact is further supported by the vertex structure of the broadcast polytope where the outcomes $(y_2, \ldots, y_n) \in \mathcal{Y}_2 \times \cdots \times \mathcal{Y}_n$ are unrestricted for each input $x \in \mathcal{X}$, hence any restrictions on the vertices are due to the limited amount of communication to the first receiver $B_1$. The limitations on the vertices. That is, let $\mathbf{P}_x^{AB_1} = \sum_{y_1 \in \mathcal{Y}_1} \sum_{m_1 \in \mathcal{M}_1} P^{B_1}(y_1|m_1) P^A(m_1|x) \, |y_1\rangle\langle x|$ and $\mathbf{P}_x^{AB_{2\ldots n}} = \sum_{y_2, \ldots, y_n \in \mathcal{Y}_2 \times \cdots \times \mathcal{Y}_n} P(y_2, \ldots, y_n) \, |y_2, \ldots, y_n\rangle\langle x|$ be column vectors, then $\mathbf{P}_x^{\text{BCN}} = \mathbf{P}_x^{AB_1} \otimes \mathbf{P}_x^{AB_{2\ldots n}}$ where $\mathbf{P}_x^{AB_{2\ldots n}}$ are totally unrestricted whereas $\mathbf{P}_x^{AB_1}$ has restrictions such that $\sum_{x \in \mathcal{X}} \mathbf{P}_x^{AB_1} = \mathbf{P}^{AB_1} \in \mathcal{C}_{\Lambda}^{\mathcal{X} \xrightarrow{d} \mathcal{Y}}$ the signaling polytope where $d = |\mathcal{M}_1|$. Hence $\mathbf{P}^{AB_1}$ satisfies all facet inequalities $\mathcal{G}^{\mathcal{X} \xrightarrow{d} \mathcal{Y}}$ bounding the signaling polytope $\mathcal{C}_{\Lambda}^{\mathcal{X} \xrightarrow{d} \mathcal{Y}}$.

We then note that any given signaling polytope inequality $(\gamma, \mathbf{G}^{\mathcal{X} \xrightarrow{d} \mathcal{Y}})$ can be output lifted to the dimensions of the broadcast network as $\mathbf{G}^{\text{BCN}} = \mathbf{G}^{\mathcal{X} \xrightarrow{d} \mathcal{Y}} \otimes \left( \bigotimes_{i=2}^{n} \left| \vec{1}_i \right\rangle \right)$ where $\left| \vec{1}_i \right\rangle = (1, \ldots, 1)^T$ and $\left| \vec{1}_i \right\rangle$ has length $|\mathcal{M}_i|$. Since $(\gamma, Gbf^{\mathcal{X} \xrightarrow{d} \mathcal{Y}})$ is a signaling polytope facet, it is satisfied with equality by exactly $|\mathcal{X}|(|\mathcal{Y}_1| - 1)$ signaling polytope vertices using only $|\mathcal{Y}_1|$ of the $|\mathcal{Y}^{\text{BCN}}| = \prod_{i=1}^{n} |\mathcal{Y}_i|$ rows of the broadcast behavior. Then since the remaining broadcast receivers are unrestricted, $|\mathcal{X}|$ affinely independent vertices can be enumerate that satisfy the inequality for all remaining $|\mathcal{Y}^{\text{BCN}}| - |\mathcal{Y}_1|$ rows. This enumeration follows precisely from the enumeration used when output lifting a signaling polytope facet. Hence in total we obtain $|\mathcal{X}|(|\mathcal{Y}^{\text{BCN}}| - 1) = \text{Dim}\left(\mathcal{C}_{\Lambda}^{\text{BCN}}\right) - 1$. Hence all signaling polytope facets can be lifted to the broadcast network.

Furthermore, there can be no additional facet inequalities, because the lifted signaling polytope inequalities are all tight, and any other output lifting on the columns of the behavior $\mathbf{P}^{AB_1}$ would not allow a sufficient number of affinely independent vertices to be enumerated. □

### *16.2.2 The 3-22-33 Broadcast Polytope*

Consider the broadcast scenario having two receivers where $|\mathcal{X}| = 3$, $|\mathcal{M}_1| = |\mathcal{M}_2| = 2$ and $|\mathcal{Y}_1| = |\mathcal{Y}_2| = 3$. The classical broadcast network polytope can be computed in full where there are a total of 441 vertices and 417 facets. Upon calculating the generator facets, we find 7 facet classes where one is the trivial non-negativity facet, while the remaining six are nontrivial bounds on the classical 3-22-33 broadcast.

$$(a)\ 2 \geq \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix}, \quad (b)\ 2 \geq \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix}, \quad (c)\ 2 \geq \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix}, \tag{16.10}$$

$$(d)\ 2 \geq \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix}, \quad (e)\ 4 \geq \begin{bmatrix} 0 & 0 & 2 \\ 0 & 0 & 1 \\ 0 & 1 & 1 \\ 0 & 2 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 1 \\ 1 & 0 & 0 \\ 2 & 0 & 0 \\ 1 & 1 & 1 \end{bmatrix}, \quad (f)\ 4 \geq \begin{bmatrix} 0 & 0 & 2 \\ 0 & 2 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \\ 2 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}, \tag{16.11}$$

## *16.3 The Weak Nonclassicality of Unassisted Bipartite Quantum Broadcasts*

In this section, we prove for a wide range of examples, that bipartite quantum broadcast networks having two receiver exhibit weak nonclassicality, but not strong nonclassicality. Although our proofs cover only specific edge cases, we find no examples where unassisted quantum communication is able to violate a linear nonclassicality witnesss bounding a broadcast network. Thus, we conjecture that unassisted quantum broadcasts are classically simulable in general.

In our first main result for the weak nonclassicality of broadcast networks, we show by example that an unbounded amount of classical communication is required to simulate an entanglement source $\Lambda$ emitting a two-qubit state. Thus, even in the simplest case of distributing randomness, a quantum advantage is identified. This result is a natural extension of Theorem 13, in which a similar result was proven for point-to-point

signaling using one qubit. This example is relevant to broadcast networks because it is equivalent to the scenario in which $|\mathcal{X}| = 1$

**Theorem 15.** An entanglement source $\Lambda$ that emits a maximally entangled state $|\Phi^+\rangle$ measured using local qubit POVMs $\Pi_{y_1}^{B_1} \otimes \Pi_{y_2}^{B_2}$ requires and unbounded amount of classical communication to simulate a local qubit POVM measurements on that state.

*Proof.* Consider the planar symmetric $N$-element qubit POVM $\{\Pi_x = \frac{2}{N} |\psi_x\rangle\langle\psi_x|\}_{x=0}^{N-1}$ where $N$ is even and $|\psi_x\rangle = \cos(\frac{\pi x}{N}) |0\rangle + \sin(\frac{\pi x}{N}) |1\rangle$. It follows that

$$P(y_1, y_2) = \mathrm{Tr}\left[\Pi_{y_1}^{B_1} \otimes \Pi_{y_2}^{B_2} \left|\Phi^+\right\rangle\!\left\langle\Phi^+\right|\right] = \frac{1}{N}\mathrm{Tr}\left[\Pi_{y_1}^{B_1} \rho_{y_2}^T\right] = \frac{1}{N} \langle\psi_{y_2}|\Pi_{y_1}]|\psi_{y_2}\rangle = P(y_1|y_2)P(y_2) \tag{16.12}$$

where $P(y_2) = \mathrm{Tr}\left[\mathbb{I}^{B_1} \otimes \Pi_{y_2}^{B_2}\right] = \frac{1}{2}\mathrm{Tr}\left[\Pi_{y_2}^{B_2}\right] = \frac{1}{N}$ and $\rho_{y_2}$ is the post-measurement state after being measured by receiver $B_2$

$$\rho_{y_2} = \frac{1}{P(y_2)}\mathrm{Tr}_{B_1}\left[(\mathbb{I}^{B_1} \otimes \Pi_{y_2}^{B_2}) \left|\Phi^+\right\rangle\!\left\langle\Phi^+\right|\right] = \frac{\Pi_{y_2}^{B_2}}{\mathrm{Tr}\left[\Pi_{y_1}^{B_2}\right]} = |\psi_{y_2}\rangle\langle\psi_{y_2}| \tag{16.13}$$

and $|\psi_{y_2}\rangle$ belongs to the set of a planar symmetric states. Since $|\psi_x\rangle$ is real, $|\psi_x\rangle\langle\psi_x|$ is symmetric for all $x$ hence $\rho_{y_2} = \rho_{y_2}^T$. Expanding Eq. (16.12) we find

$$P(y_1, y_2) = \frac{1}{N}\left(\cos^2(\theta_1)\cos^2(\theta_2) + 2\cos(\theta_1)\cos(\theta_2)\sin(\theta_1)\sin(\theta_2) + \sin^2(\theta_1)\sin^2(\theta_2)\right) \tag{16.14}$$

where $\theta_1 = \frac{\pi y_1}{N}$ and $\theta_2 = \frac{\pi y_2}{N}$. Then, the quantum behavior can be matricized as

$$\mathbf{P}^Q = \mathbf{P}^{B_1}(\mathbf{P}^{B_2})^T = \frac{1}{N}\begin{pmatrix} \frac{2}{N} & \downarrow & \downarrow & 0 & \uparrow & \uparrow \\ \uparrow & \ddots & \downarrow & \downarrow & \ddots & \uparrow \\ \uparrow & \uparrow & \frac{2}{N} & \downarrow & \downarrow & 0 \\ 0 & \uparrow & \uparrow & \frac{2}{N} & \downarrow & \downarrow \\ \downarrow & \ddots & \uparrow & \uparrow & \ddots & \downarrow \\ \downarrow & \downarrow & 0 & \uparrow & \uparrow & \frac{2}{N} \end{pmatrix} \tag{16.15}$$

where the matrix elements are bound as $\frac{2}{N} \geq \mathbf{P}_{b|c}^Q \geq 0$, the scalar factor corresponds to $P(c) = \frac{1}{N}$, and the arrows depict the fact that the matrix elements are increasing as you approach the diagonal of $\frac{2}{N}$ and depart from the shifted diagonal of zeros.

In the classical setting where shared randomness is held between two receiver $B_1$ and $B_2$ the joint behavior can be viewed i the matrix form

$$\mathbf{P}^{B_1B_2} = \sum_{\lambda} \mathbf{P}_{\lambda\in\Omega^\Lambda}^{B_1}(\mathbf{P}_{\lambda}^{B_2})^T, \tag{16.16}$$

and since $\mathbf{P}^{B_1}$ and $\mathbf{P}^{B_2}$ are non-negative,

$$\mathrm{Rank}_+\left(\mathbf{P}^{B_1B_2}\right) = |\Omega_{\mathrm{Min}}^\Lambda| = \min\{k \mid \sum_{\lambda=1}^{k} \mathbf{P}_{\lambda\in\Omega^\Lambda}^{B_1}(\mathbf{P}_{\lambda}^{B_2})^T\} \tag{16.17}$$

is precisely the minimum number of shared random values needed to simulate the joint broadcast behavior. Since we consider a qubit broadcast having $|\mathcal{M}_1| = |\mathcal{M}_2| = 2$, weak nonclassicality is shown to exist if

$\text{Rank}_+\left(\mathbf{P}^{B_1B_2}\right) > 2$. Since the matrix in Eq. (16.15) is identical to the planar symmetric behavior in Theorem 13, it follows that the number of classical message needed to simulate exactly the behavior $\mathbf{P}^Q$ is

$$\text{Rank}_+\left(\mathbf{P}^Q\right) \geq \min\{k \in \mathbb{N} \mid N \leq \binom{k}{\lfloor \frac{k}{2} \rfloor}\} \geq \lceil \log_2(N) \rceil. \tag{16.18}$$

Therefore, as $N \to \infty$ an unbounded amount of classical communication is need to exactly simulate the observed behavior of a maximally entangled two-qubit state. □

The result in Theorem 15 proves that entanglement is weakly nonclassical even in the simplest case where a source distributes a static entangled states to receivers $B_1$ and $B_2$. This example demonstrates the inclusion hierarchy $\mathcal{C}^{\text{BCN}} \subset \mathcal{Q}^{\text{BCN}} \subset \mathcal{C}_\Lambda^{\text{BCN}}$ where an unlimited amount of shared randomness is required to simulate planar symmetric POVM measurements on each qubit.

We now consider the weak nonclassicality of quantum broadcasts when $|\mathcal{X}| > 1$ and more than one input is given to the sender.

**Proposition 17.** Consider the broadcast scenario with $n$ receivers. Suppose that for all $x \in \mathcal{X}$ the sender prepares a separable state $\rho_x^A = \bigotimes_{i=1}^n \rho_x^{B_i}$ where $\rho_x^{B_i} \in D(\mathcal{H}_{d_i}^{B_i})$. It follows that $\mathcal{C}^{\text{BCN}} \subseteq \mathcal{Q}^{\text{BCN}} \subseteq \text{Conv}\left(\mathcal{C}^{\text{BCN}}\right)$, hence all quantum broadcasts using separable states are classically simulable.

*Proof.* Let the state $\rho_x^A = \bigotimes_{i=1}^n \rho_x^{B_i}$ be measured by the $n$ receivers as

$$P(\vec{y}|x) = \text{Tr}\left[\left(\bigotimes_{j=1}^n \Pi_{y_i}^{B_i}\right)\rho_x^A\right] = \prod_{i=1}^n \text{Tr}\left[\Pi_{y_i}^{B_i}\rho_x^{B_i}\right] = \prod_{i=1}^n P(y_i|x). \tag{16.19}$$

Since a classical broadcast's behavior decomposes as

$$\mathbf{P}^{\text{BCN}} = \left(\bigotimes_{j=1}^n \mathbf{P}^{B_j}\right)\mathbf{P}_x^A = \left(\bigotimes_{j=1}^n \mathbf{P}^{B_i}\right)\left(\bigotimes_{i=1}^n \mathbf{P}_x^{A\to B_i}\right) = \bigotimes_{i=1}^n \mathbf{P}^{B_i}\mathbf{P}_x^{A\to B_i}, \tag{16.20}$$

the classical probabilities factor as $P(\vec{y}|x) = \prod_{i=1}^n P(y_i|x)$ where $\mathbf{P}_x^A = \sum_{\vec{m}\in\vec{\mathcal{M}}} P(\vec{m}|x)\,|\vec{m}\rangle\langle x|$. From Eq. (16.19), we find that the broadcast network's behavior also factors as the product of each receiver's local behavior. Hence the quantum broadcast system reduces to $n$ parallel signaling scenarios. If the signaling dimension of each independent broadcast is constant across the quantum and classical versions of the network, then by Lemma 13, the classical signaling scenario can always simulate each individual channel when an unlimited amount of shared randomness is present. That is $\mathcal{Q}^{\text{BCN}} \subseteq \mathcal{C}_\Lambda^{\text{BCN}}$. Similarly using Theorem 13 we know that each independent point-to-point signaling scenario can require an unlimited amount of shared randomness to simulate. Therefore we show the complete hierarchy $\mathcal{C}^{\text{BCN}} \subseteq \mathcal{Q}^{\text{BCN}} \subseteq \mathcal{C}_\Lambda^{\text{BCN}}$. □

**Theorem 16.** Consider the broadcast scenario with two receivers in which the sender prepares the maximally entangled state $\rho_x^A = \left|\psi_x^A\right\rangle\left\langle\psi_x^A\right|$ where $\left|\psi_x^A\right\rangle = U_x^{B_1} \otimes \mathbb{I}^{B_2}\left|\Phi^+\right\rangle$ and $\left|\Phi^+\right\rangle = \sum_{i=1}^d \frac{1}{\sqrt{d}}\left|i\right\rangle\left|i\right\rangle$. All bipartite maximally entangled states are classically simulable in a two-receiver broadcast such that $\mathcal{C}^{\text{BCN}} \subseteq \mathcal{Q}^{\text{BCN}} \subseteq \mathcal{C}_\Lambda^{\text{BCN}}$.

*Proof.* Suppose $\rho_x^A = \left|\psi_x^A\right\rangle\!\left\langle\psi_x^A\right|$ where $\left|\psi_x^A\right\rangle = U_x^{B_1} \otimes \mathbb{I}^{B_2} \left|\Phi^+\right\rangle$ and $\left|\Phi^+\right\rangle = \sum_{i=1}^d \frac{1}{\sqrt{d}} \left|i\right\rangle \left|i\right\rangle$, then

$$P(y_1, y_2|x) = \mathrm{Tr}\left[\Pi_{y_1}^{B_1} \otimes \Pi_{y_2}^{B_2} \rho^A\right] = \mathrm{Tr}\left[(\Pi_{y_1}^{B_1} \otimes \Pi_{y_2}^{B_2})(U_x^{B_1} \otimes \mathbb{I}^{B_2}) \left|\Phi^+\right\rangle\!\left\langle\Phi^+\right| (U_x^{B_1} \otimes \mathbb{I}^{B_2})^\dagger\right] \tag{16.21}$$

$$= \mathrm{Tr}\left[(\Pi_{y_1}^{B_1} U_x^{B_1} \otimes \mathbb{I}^{B_2})(\mathbb{I}^{B_1} \otimes \sqrt{\Pi_{y_2}^{B_2}}) \left|\Phi^+\right\rangle\!\left\langle\Phi^+\right| (\mathbb{I}^{B_1} \otimes \sqrt{\Pi_{y_2}^{B_2}}^\dagger)(U_x^{B_1} \otimes \mathbb{I}^{B_2})^\dagger\right] \tag{16.22}$$

$$= \mathrm{Tr}\left[(\Pi_{y_1}^{B_1} U_x^{B_1} \otimes \mathbb{I}^{B_2})(\sqrt{\Pi_{y_2}^{B_2}}^T \otimes \mathbb{I}^{B_2}) \left|\Phi^+\right\rangle\!\left\langle\Phi^+\right| ((\sqrt{\Pi_{y_2}^{B_2}}^\dagger)^T \otimes \mathbb{I}^{B_2})(U_x^{B_1 \otimes \mathbb{I}^{B_2}})^\dagger\right] \tag{16.23}$$

$$= \mathrm{Tr}\left[\Pi_{y_1}^{B_1} U_x^{B_1} \sqrt{\Pi_{y_2}^{B_2}}^T \frac{1}{d}\mathbb{I} \left(\sqrt{\Pi_{y_2}^{B_2}}^\dagger\right)^T \left(U_x^{B_1}\right)^\dagger\right] \tag{16.24}$$

where in the second to last line, the ricochet property was used to carry the $\sqrt{\Pi_{y_2}^{B_2}}$ terms into the $B_1$ subsystem, and in last line the partial trace is taken over subsystem $B_2$. Furthermore the post-measurement state of a POVM measurement is

$$\rho_{y_2} = \frac{\mathrm{Tr}_{B_1}\left[(\mathbb{I}^{B_1} \otimes \Pi_{y_2}^{B_2}) \left|\Phi^+\right\rangle\!\left\langle\Phi^+\right|\right]}{P(y_2)} = \frac{\sqrt{\Pi_{y_2}^{B_2}} \frac{1}{d}\mathbb{I} \sqrt{\Pi_{y_2}^{B_2}}^\dagger}{\mathrm{Tr}\left[\Pi_{y_2}^{B_2} \frac{1}{d}\mathbb{I}\right]} = \frac{\Pi_{y_2}^{B_2}}{\mathrm{Tr}\left[\Pi_{y_2}^{B_2}\right]}. \tag{16.25}$$

It follows by the ricochet property that when the $B_2$ subsystem is traced out instead we find

$$\rho_{y_2}^T = \frac{\mathrm{Tr}_{B_2}\left[(\Pi_{y_2}^{B_2})^T \otimes \mathbb{I}^{B_2} \left|\Phi^+\right\rangle\!\left\langle\Phi^+\right|\right]}{P(y_2)} = \frac{(\Pi_{y_2}^{B_2})^T}{\mathrm{Tr}\left[\Pi_{y_2}^{B_2}\right]}. \tag{16.26}$$

Then, if we multiply Eq. (16.24) by $P(y_2)/P(y_2)$ we find

$$P(y_1, y_2|x) = \mathrm{Tr}\left[\Pi_{y_1}^{B_1} U_x^{B_1} \sqrt{\Pi_{y_2}^{B_2}}^T \frac{1}{d}\mathbb{I} \left(\sqrt{\Pi_{y_2}^{B_2}}^\dagger\right)^T \left(U_x^{B_1}\right)^\dagger\right] \frac{P(y_2)}{Py_2} \tag{16.27}$$

$$= \mathrm{Tr}\left[\Pi_{y_1}^{B_1} U_x^{B_1} \rho_{y_2}^T \left(U_x^{B_1}\right)^\dagger\right] P(y_2) \tag{16.28}$$

$$= P(y_1|x, y_2) P(y_2) \tag{16.29}$$

where in the last line it is shown $P(y_2)$ is independent from the input $x$. Thus, $y_2$ can be interpreted as a shared random value between the sender and receiver $B_2$. Furthermore, given shared randomness between devices $A$ and $B_1$ the distribution $P(y_1|x, y_2)$ can be simulated exactly using Lemma 13 where $P(y_1|x, y_2) = \sum_{m=1}^d P^{B_1}(y_1|m, \lambda) P^A(m|x, y_2, \lambda) P(\lambda)$. If using global shared randomness, the value $y_2$ can be shared between all devices $A$, $B_1$, and $B_2$, the classical simulation strategy is for device $B_2$ to simply output the shared random value $y_2$, while the signaling from device $A$ to $B_1$ is simulated using an additional shared random value $\lambda$. Interestingly, no communication is needed between $A$ and $B_2$, just shared randomness. Hence $\mathcal{Q}^{\mathrm{BCN}} \subseteq \mathcal{C}_\Lambda^{\mathrm{BCN}}$ and $\mathcal{Q}^{\mathrm{BCN}}$ can be simulated using only one-bit of communication and shared randomness. Furthermore, when $|\mathcal{Y}_2| > |\mathcal{M}_2|$ there is no way for the message alphabet $\mathcal{M}_2$ to fully communicate the value $y_1 \in \mathcal{Y}_1$, therefore, shared randomness is necessary to simulate the quantum broadcast using entanglement, but only half of the communication. □

**Lemma 16.** Any bipartite pure state, $|\psi\rangle \in \mathcal{H}_d^{B_1} \otimes \mathcal{H}_d^{B_2}$, can be expressed as $|\psi\rangle = M \otimes \mathbb{I} \left|\Phi^+\right\rangle$ where

$M = UDV$ with $U$ and $V$ being unitary operators on $\mathcal{H}_d$ and $D = \sum_i \sqrt{d}\sqrt{\lambda_i}\,|i\rangle\langle i|$ is a diagonal matrix where $\sum_i \lambda_i = 1$.

*Proof.*

$$|\psi\rangle = M \otimes \mathbb{I}\left|\Phi^+\right\rangle = UDV \otimes \mathbb{I}\left|\Phi^+\right\rangle = UD \otimes V^T\left|\Phi^+\right\rangle \tag{16.30}$$

$$= (U \otimes V^T)\sum_{j=1}^{d}\frac{1}{\sqrt{d}}\left(\sum_{i=1}^{d}\sqrt{d}\sqrt{\lambda_i}\,|i\rangle\langle i|\right)|j\rangle \otimes |j\rangle \tag{16.31}$$

$$= (U \otimes V^T)\sum_{j=1}^{d}\sqrt{\lambda_j}\,|j\rangle \otimes |j\rangle \tag{16.32}$$

$$= \sum_{j=1}^{d}\sqrt{\lambda_j}\left|\phi_j^{B_1}\right\rangle \otimes \left|\phi_j^{B_2}\right\rangle \tag{16.33}$$

where in the last line $|\psi\rangle$ is expressed in its Schmidt decomposition. □

## *16.4 Strong Nonclassicality in Quantum Broadcast Networks*

Now consider the case where $|\mathcal{X}| = |\mathcal{Y}_1| = |\mathcal{Y}_2| = 4$ and $|\mathcal{M}_1| = |\mathcal{M}_2| = 2$ such that the two receivers each have 4 outputs.. In total the 4-22-44 classical broadcast polytope has 7744 vertices, which is too much to efficiently compute the polytope in full. Nevertheless, we construct 2 interesting witnesses for broadcast nonclassicality. We write these witnesses as

$$(a)\;\; 2 \geq \begin{bmatrix} 1&0&0&0\\ 0&1&0&0\\ 0&0&0&0\\ 0&0&0&0\\ 0&1&0&0\\ 1&0&0&0\\ 0&0&0&0\\ 0&0&0&0\\ 0&0&0&0\\ 0&0&0&0\\ 0&0&1&0\\ 0&0&0&1\\ 0&0&0&0\\ 0&0&0&0\\ 0&0&0&1\\ 0&0&1&0 \end{bmatrix}, \quad (b)\;\; 8 \geq \begin{bmatrix} 3&0&0&0\\ 0&3&0&0\\ 0&0&0&3\\ 1&1&0&2\\ 1&1&0&2\\ 1&1&0&2\\ 2&1&0&2\\ 1&3&0&2\\ 1&0&2&0\\ 0&1&2&0\\ 0&0&0&3\\ 1&1&0&2\\ 2&1&0&2\\ 1&2&0&2\\ 1&1&1&2\\ 2&2&1&2 \end{bmatrix} \tag{16.34}$$

We find that first inequality $(a)$ bounds all behaviors for quantum communication and entanglement assisted receivers. The second inequality (b) is a tight facet of the $\mathcal{C}_\Lambda^{\text{BCN}}$. We derive it using the linear program in Eq. (4.22) where the nonclassical reference behavior $\mathbf{P}^{NC}$ was obtained allowing the receivers to share a PR box, a static resource that is stronger than entanglement.

When using variational optimization to maximize the inequality in Eq. (16.34).b) we find that unassisted quantum signaling does not violate the inequality, but entanglement-assisted receivers with quantum or classical signaling can violate the inequality. Both quantum and classical broadcasts with entangled receivers can achieve a maximal score of 8.5, which corresponds to a violation of 0.5 beyond the classical bound.

To summarize, we provide an example of a broadcast network in which entanglement-assisted receivers produce strongly nonclassical behaviors. However, we find no strongly nonclassical example for quantum signaling. The derived example was obtained by using as reference, a nonclassical behavior produced when the receivers share a PR box, so perhaps it is not surprising that entanglement-assisted receivers can also yield strongly nonclassical behaviors. Thus, these violations could be used to verify that receivers in a broadcast network share entanglement.

# Chapter 17

# Nonclassicality in Multipoint Communication Networks

**Definition 106. Multipoint Communication Network:** (MPN), A communication network having mulitple senders and multiple receivers.

In this chapter we investigate the nonclassicality that occurs in multipoint communication networks when quantum resources are used. We restrict our focus to a handful of network topologies for which linear nonclassicality witnesses can be derived and nonclassical behaviors can be optimized (see Fig. 17.1).

## *17.1 Nonclassicality Witnesses for Multipoint Networks*

We consider multipoint networks that have two senders and two receivers with intermediate interference nodes. To derive nonclassicality witnesses, we obtain the set of vertices for the considered network topology, then we apply the linear program in Eq. (4.22) to derive a facet of the multipoint network polytope. The linear program requires a nonclassical behavior $\mathbf{P}^{NC} \notin \mathcal{C}_{\Lambda}^{\text{Net}}$, which we obtain by considering the behaviors that are impossible to obtain using the fixed set of classical communication resources. To begin, we define the sets of classical behaviors for each of the multipoint networks given in Fig. 17.1.d,e,f).

**Definition 107. The set of Classical Interference Network Behaviors** $\mathcal{C}^{\text{IN}}$, A behavior belongs to the set of classical interference network (IN) behaviors if and only if it decomposes as

$$\mathbf{P}^{\text{IN}} = (\mathbf{P}^{C_1} \otimes \mathbf{P}^{C_2})\mathbf{P}^{B}(\mathbf{P}^{A_1}\mathbf{P}^{A_2}) \tag{17.1}$$

where $\mathbf{P}^{\text{IN}} \in \mathcal{P}_{\mathcal{Y}_1\times\mathcal{Y}_2|\mathcal{X}_1\times\mathcal{X}_2}$, $\mathbf{P}^{A_i} \in \mathcal{P}_{\mathcal{A}_i|\mathcal{X}_i}$, $\mathbf{P}^{B} \in \mathcal{P}_{\mathcal{B}_1\times\mathcal{B}_2|\mathcal{A}_1\times\mathcal{A}_2}$, and $\mathbf{P}^{C_i} \in \mathcal{P}_{\mathcal{Y}_i|\mathcal{B}_i}$.

**Definition 108. The Set of Classical Compressed Interference Network Behaviors:** $\mathcal{C}^{\text{CIN}}$, A behavior belongs to the set of classical compressed interference networks (CIN) behaviors if and only if it decomposes as

$$\mathbf{P}^{\text{IN}} = (\mathbf{P}^{D_1} \otimes \mathbf{P}^{D_2})\mathbf{P}^{C}\mathbf{P}^{B}(\mathbf{P}^{A_1}\mathbf{P}^{A_2}) \tag{17.2}$$

where $\mathbf{P}^{\text{IN}} \in \mathcal{P}_{\mathcal{Y}_1\times\mathcal{Y}_2|\mathcal{X}_1\times\mathcal{X}_2}$, $\mathbf{P}^{A_i} \in \mathcal{P}_{\mathcal{A}_i|\mathcal{X}_i}$, $\mathbf{P}^{B} \in \mathcal{P}_{\mathcal{B}|\mathcal{A}_1\times\mathcal{A}_2}$, $\mathbf{P}^{C} \in \mathcal{P}_{\mathcal{C}_1\times\mathcal{C}_2|\mathcal{B}}$, and $\mathbf{P}^{D_i} \in \mathcal{P}_{\mathcal{Y}_i|\mathcal{C}_i}$.

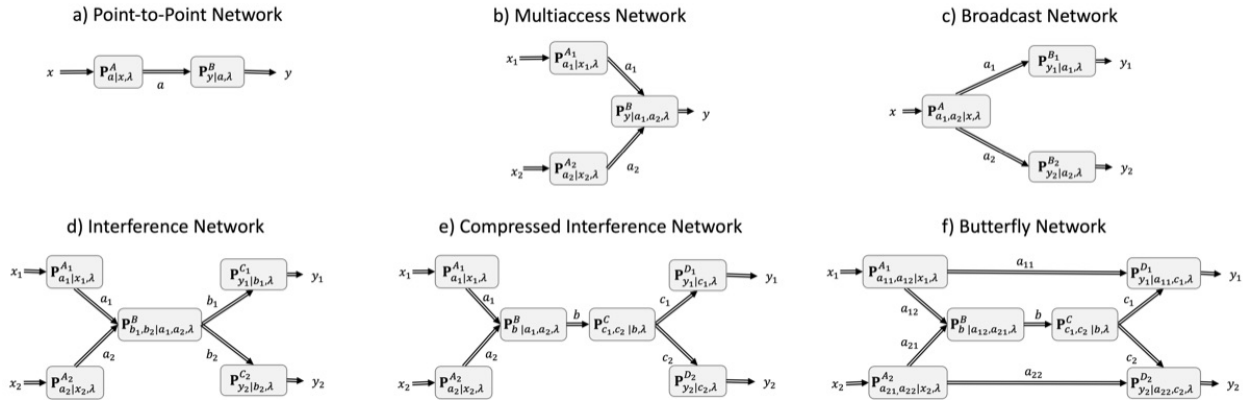



Figure 17.1: **Multipoint Communication Network DAGs:** The top row shows communication networks studied previously in this work and the bottom row shows the multipoint communication network investigated in this chapter. a) Point-to-point signaling network, b) multiaccess network, c) broadcast network, d) interference network, e) compressed interference network, and f) butterfly network.

**Definition 109. The Set of Classical Butterfly Network Behaviors:** $\mathcal{C}^{\text{BFN}}$, A behavior belongs to the set of classical butterfly network (BFN) behaviors if and only if it decomposes

$$\mathbf{P}^{\text{BFN}} = (\mathbf{P}^{D_1} \otimes \mathbf{P}^{D_2})(\mathbb{I}_{|\mathcal{A}_{11}|} \otimes (\mathbf{P}^{C}\mathbf{P}^{B}) \otimes \mathbb{I}_{|\mathcal{A}_{22}|})(\mathbf{P}^{A_1}\mathbf{P}^{A_2}) \tag{17.3}$$

where $\mathbf{P}^{\text{BFN}} \in \mathcal{P}_{\mathcal{Y}_1\times\mathcal{Y}_2|\mathcal{X}_1\times\mathcal{X}_2}$, $\mathbf{P}^{A_i} \in \mathcal{P}_{\mathcal{A}_{i1}\times\times\mathcal{A}_{i2}|\mathcal{X}_i}$, $\mathbf{P}^{B} \in \mathcal{P}_{\mathcal{B}|\mathcal{A}_1\times\mathcal{A}_2}$, $\mathbf{P}^{C} \in \mathcal{P}_{\mathcal{C}_1\times\mathcal{C}_2|\mathcal{B}}$, and $\mathbf{P}^{D_i} \in \mathcal{P}_{\mathcal{Y}_i|\mathcal{A}_{ii}\times\mathcal{C}_i}$. Note that the identity matrices are included because the butterfly network has two classical communication wires that are unused during layers $B$ and $C$.

The behaviors of each of these networks are described without the use of global shared randomness, meaning that $\mathcal{C}^{\text{Net}}$ is generally nonconvex. However, when an unlimited amount of global shared randomness is considered, the set of classical network behaviors is convex and forms the respective classical network polytope $\mathcal{C}^{\text{Net}}_{\Lambda}$. The vertices of the classical network polytope are exactly the set of deterministic behaviors $\mathcal{V}^{\text{Net}} = \{\mathbf{P} \in \mathcal{C}^{\text{Net}} \mid P(\vec{y}|\vec{x}) \in \mathbb{B}\ \forall\ \vec{x} \in \mathcal{X},\ \vec{y} \in \mathcal{Y}\}$. The facet inequalities bounding the network polytope then constitute linear nonclassicality witnesses whose violations witnesses strong nonclassicality.

In the case where $|\mathcal{X}_1| = |\mathcal{X}_2| = |\mathcal{Y}_1| = |\mathcal{Y}_2| = 3$ and all communication has a signaling dimension of $\kappa(\text{id}^{\text{Tx}\to\text{Rx}}) = 2$, the complete set of vertices can be derived for each of the multipoint communication networks shown in Fig. 17.1. The number of vertices is too large to efficiently compute the complete set of facets bounding the respective classical network polytope, however, the linear program in Eq. (4.22) can still be used to obtain example facets. The linear program requires a nonclassical behavior for which to derive a facet that witnesses its violation.

To derive examples of nonclassical behaviors, we simply consider behaviors that require additional communication to implement (see Table 17.1). Each of these behaviors corresponds to a task that is not possible given the restricted communication. Although, these nonclassical behaviors could be interpreted as linear games in which one point is awarded for the correct answer and zero points for an incorrect answer, these linear games do not represent facets of the classical network polytope. Therefore, to derive proper facets of the signaling polytope, we use these behaviors as nonclassical reference points for which the linear program in Eq. (4.22) can derive a facet that bounding the classical network polytope. Since the derived linear inequalities are tight nonclassicality witnesses, the serve as ideal operational tests for strong nonclassicality in

multipoint communication networks.

## *17.2 Strong Nonclassicality in Multipoint Communication Networks*

We now cater our analysis for each of the multipoint networks considered in Fig. 17.1. For each network we list the derived linear nonclassicality witnesses, and show when quantum communication or entanglement is added that violations can be obtained in certain cases. Thus we show by example, that strong nonclassicality can be witnessed in multipoint communication networks indicating that operational advantages can be derived in the multipoint network setting.

Overall, we find many example of strong nonclassicality when either unassisted quantum signaling is used, or quantum signaling is used with entanglement-assisted senders or receivers. However, the violations we find are only lower bounds of the possible quantum violations of these nonclassicality witnesses. The reason being that the applied variational quantum optimization methods are not guaranteed to find the global optimum. Furthermore, the free operations within the respective variational ansatzes for the network may not enable the maximal violation to be achieved. Nevertheless, a violation to any of these inequalities indicates that strong nonclassicality is present. Therefore, these violations correspond directly with an operational resource advantage.

### *17.2.1 Nonclassicality in the Interference Network*

Consider the interference network (Fig. 17.1.d) having $|\mathcal{X}_1| = |\mathcal{X}_2| = |\mathcal{Y}_1| = |\mathcal{Y}_2| = 3$ and one bit of communication used in each wire. Enumerating the vertices we find a total of 17,289 vertices. Then we derive a facet inequality of $\mathcal{C}_\Lambda^{\text{IN}}$ for each behavior in Table 17.1. The derived faced inequalities are listed in Table 17.2 where by the relabeling symmetry, any local permutation of inputs or outputs yields a similar inequality.

We then consider three distinct quantum interference network ansatzes. First, we replace all classical communication with one qubit of quantum communication. Then we consider the case where the senders are assisted with quantum entanglement (EATx). Finally, we consider the case where the receivers are assisted with quantum entanglement (EARx). In all entanglement-assisted cases, quantum communication is used. Applying our variational optimization framework yields the violations shown in Fig. 17.2. We find that each of our considered quantum signaling variational ansatzes were able to violate at least one of the nonclassicality inequalities. The largest violations were achieved for the entanglement-assisted senders. Note that the results in Fig. 17.2 show lower bounds on the possible violation, but the maximal violation for each resource configuration is not necessarily known.

### *17.2.2 Nonclassicality in the Compressed Interference Network*

Consider the compressed interference network (Fig. 17.1.e) having $|\mathcal{X}_1| = |\mathcal{X}_2| = |\mathcal{Y}_1| = |\mathcal{Y}_2| = 3$ and one bit of communication used in each wire. Enumerating the vertices we find a total of 3,681 vertices. Then we derive a facet inequality of $\mathcal{C}_\Lambda^{\text{CIN}}$ for each behavior in Table 17.1. The derived faced inequalities are listed in Table 17.3 where by the relabeling symmetry, any local permutation of inputs or outputs yields a similar inequality.

We then consider three distinct quantum compressed interference network ansatzes. First, we replace all classical communication with one qubit of quantum communication. Then we consider the case where the

$$\mathbf{P}_{\text{Mult-0}} = \begin{pmatrix} 1&1&1&1&0&0&1&0&0\\ 0&0&0&0&1&0&0&0&0\\ 0&0&0&0&0&1&0&1&0\\ 0&0&0&0&0&0&0&0&0\\ 0&0&0&0&0&0&0&0&1\\ 0&0&0&0&0&0&0&0&0\\ 0&0&0&0&0&0&0&0&0\\ 0&0&0&0&0&0&0&0&0\\ 0&0&0&0&0&0&0&0&0 \end{pmatrix} \quad \mathbf{P}_{\text{Mult-1}} = \begin{pmatrix} 1&0&0&0&0&0&0&0&0\\ 0&1&0&1&0&0&0&0&0\\ 0&0&1&0&0&0&1&0&0\\ 0&0&0&0&1&0&0&0&0\\ 0&0&0&0&0&0&0&0&0\\ 0&0&0&0&0&1&0&1&0\\ 0&0&0&0&0&0&0&0&0\\ 0&0&0&0&0&0&0&0&0\\ 0&0&0&0&0&0&0&0&1 \end{pmatrix}$$

$$\mathbf{P}_{\text{Add-0}} = \begin{pmatrix} 1&0&0&0&0&0&0&0&0\\ 0&1&0&1&0&0&0&0&0\\ 0&0&1&0&1&0&1&0&0\\ 0&0&0&0&0&1&0&1&0\\ 0&0&0&0&0&0&0&0&1\\ 0&0&0&0&0&0&0&0&0\\ 0&0&0&0&0&0&0&0&0\\ 0&0&0&0&0&0&0&0&0\\ 0&0&0&0&0&0&0&0&0 \end{pmatrix} \quad \mathbf{P}_{\text{Diff}} = \begin{pmatrix} 1&0&0&0&1&0&0&0&1\\ 0&0&0&0&0&0&0&0&0\\ 0&0&0&0&0&0&0&0&0\\ 0&0&0&0&0&0&0&0&0\\ 0&1&0&1&0&1&0&1&0\\ 0&0&0&0&0&0&0&0&0\\ 0&0&0&0&0&0&0&0&0\\ 0&0&0&0&0&0&0&0&0\\ 0&0&1&0&0&0&1&0&0 \end{pmatrix}$$

$$\mathbf{P}_{\text{Comp}} = \begin{pmatrix} 1&0&0&0&1&0&0&0&1\\ 0&0&0&0&0&0&0&0&0\\ 0&0&0&0&0&0&0&0&0\\ 0&0&0&0&0&0&0&0&0\\ 0&0&0&0&0&0&0&0&0\\ 0&1&1&0&0&1&0&0&0\\ 0&0&0&0&0&0&0&0&0\\ 0&0&0&1&0&0&1&1&0\\ 0&0&0&0&0&0&0&0&0 \end{pmatrix} \quad \mathbf{P}_{\text{Perm}} = \begin{pmatrix} 1&0&0&0&0&0&0&0&0\\ 0&1&0&0&0&0&0&0&0\\ 0&0&1&0&0&0&0&0&0\\ 0&0&0&0&1&0&0&0&0\\ 0&0&0&0&0&1&0&0&0\\ 0&0&0&1&0&0&0&0&0\\ 0&0&0&0&0&0&0&0&1\\ 0&0&0&0&0&0&1&0&0\\ 0&0&0&0&0&0&0&1&0 \end{pmatrix}$$

$$\mathbf{P}_{\text{Swap}} = \begin{pmatrix} 1&0&0&0&0&0&0&0&0\\ 0&0&0&1&0&0&0&0&0\\ 0&0&0&0&0&0&1&0&0\\ 0&1&0&0&0&0&0&0&0\\ 0&0&0&0&1&0&0&0&0\\ 0&0&0&0&0&0&0&1&0\\ 0&0&1&0&0&0&0&0&0\\ 0&0&0&0&0&1&0&0&0\\ 0&0&0&0&0&0&0&0&1 \end{pmatrix} \quad \mathbf{P}_{\text{CV}} = \mathbb{I}_9 = \begin{pmatrix} 1&0&0&0&0&0&0&0&0\\ 0&1&0&0&0&0&0&0&0\\ 0&0&1&0&0&0&0&0&0\\ 0&0&0&1&0&0&0&0&0\\ 0&0&0&0&1&0&0&0&0\\ 0&0&0&0&0&1&0&0&0\\ 0&0&0&0&0&0&1&0&0\\ 0&0&0&0&0&0&0&1&0\\ 0&0&0&0&0&0&0&0&1 \end{pmatrix}$$

Table 17.1: Nonclassical behaviors for multipoint networks having $|\mathcal{X}_1| = |\mathcal{X}_2| = |\mathcal{Y}_1| = |\mathcal{Y}_2| = 3$ and one-bit of communication for each wire. $\mathbf{P}_{\text{Mult}-0}$, multiplication starting from zero, $y_1y_2 = x_1 * x_2$ where $y_1, y_2, x_1, x_2 \in \{0, 1, 2\}$. $\mathbf{P}_{\text{Mult}-1}$, multiplication start from one, $y_1y_2 = x_1 * x_2$ where $y_1, y_2, x_1, x_2 \in \{1, 2, 3\}$. $\mathbf{P}_{\text{Add-0}}$ addition, $y_1y_2 = x_1 + x_2$ where $y_1, y_2, x_1, x_2 \in \{0, 1, 2\}$. $\mathbf{P}_{\text{Diff}}$, subtraction where $y_1 = y_2 = |x_1 - x_2|$ where $y_1, y_2, x_1, x_2 \in \{0, 1, 2\}$. $\mathbf{P}_{\text{Comp}}$, comparison where $y_1 = y_1 = 0$ if $x_1 = x_2$, $y_1 = y_2 + 1$ if $x_1 > x_2$, and $y_2 = y_1 + 1$ if $x_2 > x_1$. $\mathbf{P}_{\text{Perm}}$, $y_1 = x_1$ and $y_2 = \pi_{y_1}(x_2)$ where $\pi_{y_1}$ is a permutation depending on the output $y_1$. $\mathbf{P}_{\text{Swap}}$, the output swaps the input as $y_2 = x_1$ and $y_1 = x_2$. $\mathbf{P}_{\text{CV}}$, noiseless communication where $y_1 = x_1$ and $y_2 = x_3$

.

$$13 \geq \mathbf{G}^{\mathrm{IN}}_{\text{Mult-0}} = \begin{pmatrix} 1 & 2 & 1 & 3 & 0 & 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 & 3 & 0 & 0 & 1 & 2 \\ 1 & 0 & 0 & 1 & 1 & 3 & 0 & 2 & 0 \\ 0 & 1 & 1 & 0 & 1 & 2 & 1 & 1 & 2 \\ 0 & 1 & 1 & 0 & 1 & 2 & 1 & 1 & 2 \\ 0 & 1 & 1 & 1 & 1 & 2 & 1 & 1 & 2 \\ 1 & 0 & 0 & 1 & 3 & 1 & 0 & 0 & 2 \\ 1 & 1 & 0 & 1 & 1 & 2 & 0 & 1 & 2 \\ 1 & 1 & 1 & 2 & 2 & 2 & 1 & 1 & 1 \end{pmatrix} \qquad 13 \geq \mathbf{G}^{\mathrm{IN}}_{\text{Mult-1}} = \begin{pmatrix} 3 & 0 & 1 & 0 & 2 & 1 & 1 & 0 & 1 \\ 0 & 3 & 0 & 2 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 4 & 0 & 2 & 1 & 2 & 0 & 0 \\ 2 & 1 & 2 & 0 & 2 & 1 & 0 & 1 & 1 \\ 2 & 1 & 2 & 1 & 1 & 2 & 0 & 1 & 1 \\ 2 & 1 & 2 & 0 & 0 & 3 & 0 & 2 & 0 \\ 2 & 1 & 2 & 0 & 2 & 1 & 0 & 0 & 1 \\ 2 & 1 & 2 & 0 & 2 & 1 & 0 & 0 & 1 \\ 2 & 2 & 3 & 1 & 1 & 2 & 1 & 1 & 0 \end{pmatrix}$$

$$14 \geq \mathbf{G}^{\mathrm{IN}}_{\text{Add}} = \begin{pmatrix} 3 & 0 & 1 & 0 & 2 & 1 & 1 & 0 & 1 \\ 0 & 3 & 0 & 2 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 4 & 0 & 2 & 1 & 2 & 0 & 0 \\ 2 & 1 & 2 & 0 & 0 & 3 & 0 & 2 & 0 \\ 1 & 0 & 3 & 1 & 1 & 2 & 1 & 1 & 1 \\ 1 & 2 & 2 & 1 & 1 & 2 & 0 & 1 & 1 \\ 2 & 1 & 2 & 0 & 2 & 1 & 0 & 0 & 1 \\ 1 & 1 & 3 & 0 & 2 & 1 & 1 & 0 & 1 \\ 2 & 2 & 3 & 1 & 1 & 2 & 1 & 1 & 0 \end{pmatrix} \qquad 11 \geq \mathbf{G}^{\mathrm{IN}}_{\text{Diff}} = \begin{pmatrix} 2 & 0 & 0 & 0 & 2 & 0 & 1 & 0 & 1 \\ 0 & 0 & 2 & 1 & 1 & 2 & 1 & 0 & 1 \\ 1 & 0 & 2 & 0 & 0 & 2 & 1 & 0 & 1 \\ 0 & 0 & 2 & 1 & 1 & 2 & 1 & 0 & 1 \\ 0 & 0 & 2 & 2 & 0 & 2 & 0 & 1 & 0 \\ 0 & 0 & 2 & 2 & 1 & 2 & 0 & 0 & 1 \\ 1 & 0 & 2 & 0 & 0 & 2 & 1 & 0 & 1 \\ 0 & 0 & 2 & 2 & 1 & 2 & 0 & 0 & 1 \\ 1 & 1 & 3 & 1 & 1 & 2 & 1 & 0 & 0 \end{pmatrix}$$

$$12 \geq \mathbf{G}^{\mathrm{IN}}_{\text{Comp}} = \begin{pmatrix} 2 & 0 & 0 & 0 & 3 & 0 & 0 & 0 & 1 \\ 0 & 1 & 1 & 1 & 0 & 3 & 1 & 1 & 0 \\ 1 & 1 & 1 & 0 & 0 & 3 & 0 & 1 & 0 \\ 0 & 2 & 1 & 0 & 2 & 2 & 1 & 0 & 1 \\ 1 & 1 & 1 & 0 & 0 & 3 & 1 & 1 & 0 \\ 0 & 3 & 1 & 0 & 1 & 3 & 1 & 0 & 0 \\ 1 & 2 & 1 & 1 & 1 & 2 & 0 & 0 & 1 \\ 0 & 1 & 2 & 2 & 0 & 1 & 1 & 1 & 0 \\ 1 & 2 & 2 & 1 & 2 & 2 & 1 & 0 & 0 \end{pmatrix} \qquad 13 \geq \mathbf{G}^{\mathrm{IN}}_{\text{Perm}} = \begin{pmatrix} 3 & 0 & 0 & 0 & 1 & 1 & 0 & 1 & 1 \\ 0 & 3 & 2 & 1 & 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 4 & 1 & 1 & 1 & 0 & 1 & 0 \\ 1 & 1 & 2 & 1 & 2 & 0 & 0 & 1 & 1 \\ 1 & 2 & 2 & 1 & 0 & 3 & 0 & 1 & 0 \\ 1 & 2 & 2 & 2 & 0 & 0 & 0 & 0 & 1 \\ 1 & 2 & 2 & 1 & 1 & 1 & 0 & 1 & 1 \\ 0 & 1 & 3 & 0 & 1 & 2 & 1 & 0 & 0 \\ 2 & 2 & 3 & 1 & 1 & 2 & 0 & 1 & 0 \end{pmatrix}$$

$$13 \geq \mathbf{G}^{\mathrm{IN}}_{\text{Swap}} = \begin{pmatrix} 3 & 0 & 0 & 0 & 2 & 0 & 1 & 0 & 1 \\ 1 & 1 & 2 & 2 & 0 & 0 & 1 & 0 & 1 \\ 0 & 1 & 3 & 0 & 1 & 1 & 2 & 0 & 0 \\ 0 & 3 & 0 & 2 & 0 & 0 & 0 & 1 & 1 \\ 1 & 1 & 2 & 0 & 2 & 0 & 0 & 1 & 1 \\ 1 & 0 & 3 & 1 & 0 & 1 & 0 & 2 & 0 \\ 0 & 0 & 4 & 1 & 1 & 0 & 1 & 1 & 0 \\ 1 & 2 & 2 & 0 & 0 & 2 & 1 & 1 & 0 \\ 2 & 2 & 3 & 1 & 1 & 1 & 1 & 1 & 0 \end{pmatrix} \qquad 13 \geq \mathbf{G}^{\mathrm{IN}}_{\text{CV}} = \begin{pmatrix} 3 & 0 & 0 & 0 & 2 & 0 & 1 & 0 & 1 \\ 0 & 3 & 0 & 2 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 4 & 1 & 1 & 0 & 1 & 1 & 0 \\ 1 & 1 & 2 & 2 & 0 & 0 & 1 & 0 & 1 \\ 1 & 1 & 2 & 0 & 2 & 0 & 0 & 1 & 1 \\ 1 & 2 & 2 & 0 & 0 & 2 & 1 & 1 & 0 \\ 0 & 1 & 3 & 0 & 1 & 1 & 2 & 0 & 0 \\ 1 & 0 & 3 & 1 & 0 & 1 & 0 & 2 & 0 \\ 2 & 2 & 3 & 1 & 1 & 1 & 1 & 1 & 0 \end{pmatrix}$$

Table 17.2: Derived nonclassicality witnesses for the interference network. Each inequality $(\gamma, \mathbf{G})$ is expressed as $\gamma \geq \mathbf{G}$.

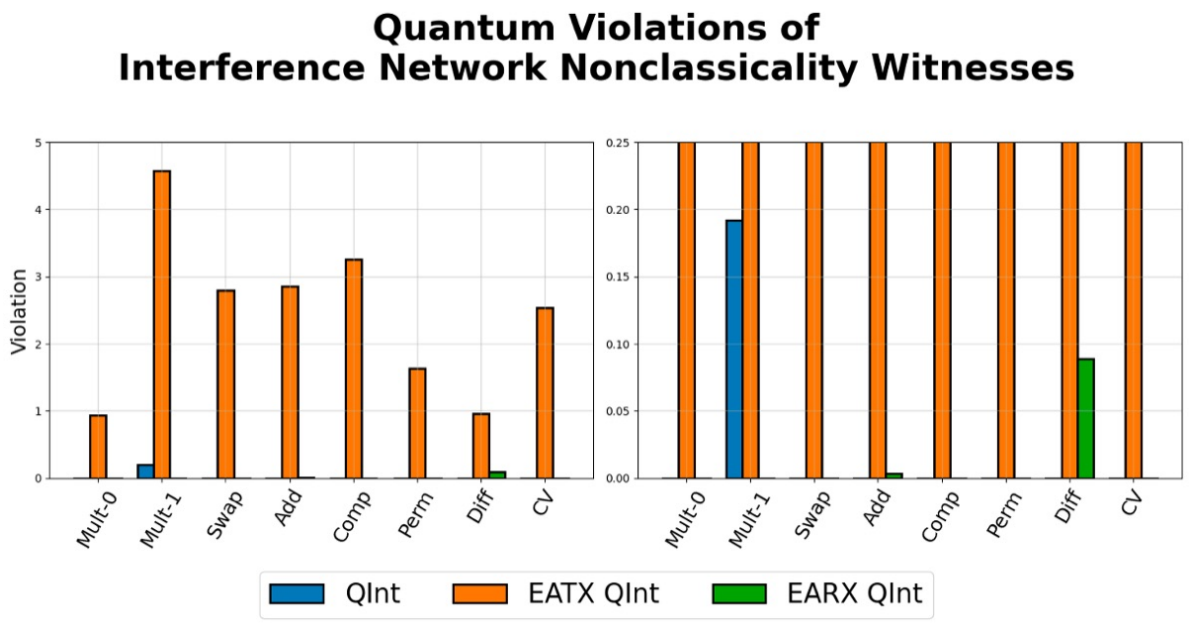



Figure 17.2: Quantum violations of classical bounds for the interference network. The $x$-axis labels the nonclassicality witness while the $y$-axis shows the violation where the classical bound is fixed to zero. The left plot shows the full violations while the right plot shows the same data but zoomed in to show detail. The blue bars show the case where unassisted quantum signaling is used. The orange bars show the case when quantum signaling is used and the senders are assisted with entanglement. The green bars show the case when quantum signaling is used and the receivers are assisted with entanglement.

senders are assisted with quantum entanglement (EATx). Finally, we consider the case where the receivers are assisted with quantum entanglement (EARx). In all entanglement-assisted cases, quantum communication is used. Applying our variational optimization framework yields the violations shown in Fig. 17.3. For each nonclassicality witness, we find a quantum violation except in the for the inequality $(12, \mathbf{G}^{\text{CIN}}_{\text{Mult-0}})$. The entanglement-assisted senders are able to violate each of the example nonclassicality witnesses. We find the entanglement-assisted receivers to only be able to violate the nonclassicality witness $(12, \mathbf{G}^{\text{CIN}}_{\text{Comp}})$. Remarkably, unassisted quantum communication is able to achieve violations in most cases, and for the $\mathbf{G}^{\text{CIN}}_{\text{Swap}}$ and $\mathbf{G}^{\text{CIN}}_{\text{Diff}}$ the unassisted quantum signaling achieves stronger violations than the entanglement-assisted senders.

### *17.2.3 Nonclassicality in the Butterfly Network*

Consider the butterfly network (Fig. 17.1.f) having $|\mathcal{X}_1| = |\mathcal{X}_2| = |\mathcal{Y}_1| = |\mathcal{Y}_2| = 3$ and one bit of communication used in each wire. Enumerating the vertices we find a total of 690,813 vertices. Then we derive a facet inequality of $\mathcal{C}^{\text{BFN}}_{\Lambda}$ for each behavior in Table 17.1. The derived faced inequalities are listed in Table 17.4 where by the relabeling symmetry, any local permutation of inputs or outputs yields a similar inequality.

We then consider three distinct quantum compressed interference network ansatzes. First, we replace all classical communication with one qubit of quantum communication. Then we consider the case where the senders are assisted with quantum entanglement (EATx). Finally, we consider the case where the receivers are assisted with quantum entanglement (EARx). In all entanglement-assisted cases, quantum communication is used. Applying our variational optimization framework yields the violations shown in Fig. 17.4. We find no examples in which unassisted quantum signaling is able to violate a nonclassicality witness. The

$$12 \geq \mathbf{G}^{\text{CIN}}_{\text{Mult-0}} = \begin{pmatrix} 1 & 2 & 2 & 1 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 3 & 0 & 1 & 2 & 1 \\ 0 & 0 & 0 & 0 & 0 & 3 & 1 & 3 & 0 \\ 0 & 1 & 2 & 0 & 1 & 2 & 1 & 2 & 1 \\ 0 & 1 & 2 & 0 & 1 & 2 & 1 & 2 & 1 \\ 0 & 1 & 2 & 0 & 1 & 2 & 1 & 2 & 1 \\ 0 & 1 & 2 & 0 & 1 & 2 & 1 & 2 & 1 \\ 0 & 1 & 2 & 0 & 1 & 2 & 1 & 2 & 1 \\ 1 & 1 & 2 & 0 & 2 & 2 & 1 & 2 & 0 \end{pmatrix} \quad 13 \geq \mathbf{G}^{\text{CIN}}_{\text{Mult-1}} = \begin{pmatrix} 4 & 0 & 1 & 1 & 0 & 2 & 0 & 0 & 1 \\ 1 & 3 & 0 & 3 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 3 & 0 & 2 & 1 & 1 & 0 & 0 \\ 2 & 1 & 1 & 2 & 2 & 1 & 0 & 0 & 1 \\ 2 & 1 & 2 & 2 & 1 & 2 & 0 & 0 & 1 \\ 2 & 1 & 1 & 2 & 0 & 3 & 0 & 1 & 0 \\ 2 & 1 & 2 & 2 & 1 & 2 & 0 & 0 & 1 \\ 2 & 1 & 2 & 2 & 1 & 2 & 0 & 0 & 1 \\ 3 & 2 & 2 & 2 & 1 & 2 & 0 & 0 & 0 \end{pmatrix}$$

$$13 \geq \mathbf{G}^{\text{CIN}}_{\text{Add}} = \begin{pmatrix} 4 & 0 & 1 & 1 & 0 & 2 & 0 & 0 & 1 \\ 1 & 3 & 0 & 3 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 3 & 0 & 2 & 1 & 1 & 0 & 0 \\ 2 & 1 & 1 & 2 & 0 & 3 & 0 & 1 & 0 \\ 2 & 1 & 2 & 2 & 1 & 2 & 0 & 0 & 1 \\ 2 & 1 & 2 & 2 & 1 & 2 & 0 & 0 & 1 \\ 2 & 1 & 2 & 2 & 1 & 2 & 0 & 0 & 1 \\ 2 & 1 & 2 & 2 & 1 & 2 & 0 & 0 & 1 \\ 3 & 2 & 2 & 2 & 1 & 2 & 0 & 0 & 0 \end{pmatrix} \quad 11 \geq \mathbf{G}^{\text{CIN}}_{\text{Diff}} = \begin{pmatrix} 2 & 0 & 0 & 0 & 2 & 0 & 0 & 0 & 1 \\ 0 & 1 & 3 & 1 & 1 & 2 & 0 & 0 & 1 \\ 0 & 1 & 3 & 1 & 1 & 2 & 0 & 0 & 1 \\ 0 & 1 & 3 & 1 & 1 & 2 & 0 & 0 & 1 \\ 0 & 1 & 3 & 2 & 0 & 2 & 0 & 1 & 0 \\ 0 & 1 & 3 & 1 & 1 & 2 & 0 & 0 & 1 \\ 0 & 1 & 3 & 1 & 1 & 2 & 0 & 0 & 1 \\ 0 & 1 & 3 & 1 & 1 & 2 & 0 & 0 & 1 \\ 1 & 2 & 3 & 1 & 1 & 2 & 0 & 0 & 0 \end{pmatrix}$$

$$12 \geq \mathbf{G}^{\text{CIN}}_{\text{Comp}} = \begin{pmatrix} 3 & 0 & 0 & 0 & 2 & 0 & 0 & 0 & 1 \\ 1 & 0 & 3 & 2 & 1 & 2 & 0 & 0 & 1 \\ 1 & 0 & 3 & 2 & 1 & 2 & 0 & 0 & 1 \\ 1 & 0 & 3 & 2 & 1 & 2 & 0 & 0 & 1 \\ 1 & 0 & 3 & 2 & 1 & 2 & 0 & 0 & 1 \\ 0 & 2 & 3 & 1 & 0 & 3 & 0 & 0 & 0 \\ 1 & 0 & 3 & 2 & 1 & 2 & 0 & 0 & 1 \\ 0 & 0 & 3 & 3 & 0 & 2 & 0 & 1 & 0 \\ 2 & 1 & 3 & 2 & 1 & 2 & 0 & 0 & 0 \end{pmatrix} \quad 9 \geq \mathbf{G}^{\text{CIN}}_{\text{Perm}} = \begin{pmatrix} 3 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 \\ 0 & 3 & 0 & 0 & 0 & 1 & 0 & 0 & 1 \\ 0 & 1 & 2 & 0 & 1 & 0 & 0 & 1 & 0 \\ 1 & 1 & 0 & 1 & 2 & 0 & 0 & 0 & 1 \\ 1 & 1 & 0 & 1 & 0 & 2 & 0 & 1 & 0 \\ 0 & 2 & 0 & 2 & 1 & 0 & 0 & 0 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 1 & 2 & 1 & 0 & 0 \\ 2 & 2 & 1 & 1 & 1 & 1 & 0 & 0 & 0 \end{pmatrix}$$

$$9 \geq \mathbf{G}^{\text{CIN}}_{\text{Swap}} = \begin{pmatrix} 3 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 \\ 0 & 2 & 0 & 2 & 1 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 1 & 2 & 1 & 0 & 0 \\ 0 & 3 & 0 & 0 & 0 & 1 & 0 & 0 & 1 \\ 1 & 1 & 0 & 1 & 2 & 0 & 0 & 0 & 1 \\ 1 & 1 & 0 & 1 & 0 & 2 & 0 & 1 & 0 \\ 0 & 1 & 2 & 0 & 1 & 0 & 0 & 1 & 0 \\ 1 & 1 & 0 & 1 & 0 & 2 & 0 & 1 & 0 \\ 2 & 2 & 1 & 1 & 1 & 1 & 0 & 0 & 0 \end{pmatrix} \quad 9 \geq \mathbf{G}^{\text{CIN}}_{\text{CV}} = \begin{pmatrix} 3 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 \\ 0 & 3 & 0 & 0 & 0 & 1 & 0 & 0 & 1 \\ 0 & 1 & 2 & 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 2 & 0 & 2 & 1 & 0 & 0 & 0 & 1 \\ 1 & 1 & 0 & 1 & 2 & 0 & 0 & 0 & 1 \\ 1 & 1 & 0 & 1 & 0 & 2 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 1 & 2 & 1 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 & 2 & 0 & 1 & 0 \\ 2 & 2 & 1 & 1 & 1 & 1 & 0 & 0 & 0 \end{pmatrix}$$

Table 17.3: Derived nonclassicality witnesses for the compressed interference network. Each inequality $(\gamma, \mathbf{G})$ is expressed as $\gamma \geq \mathbf{G}$.

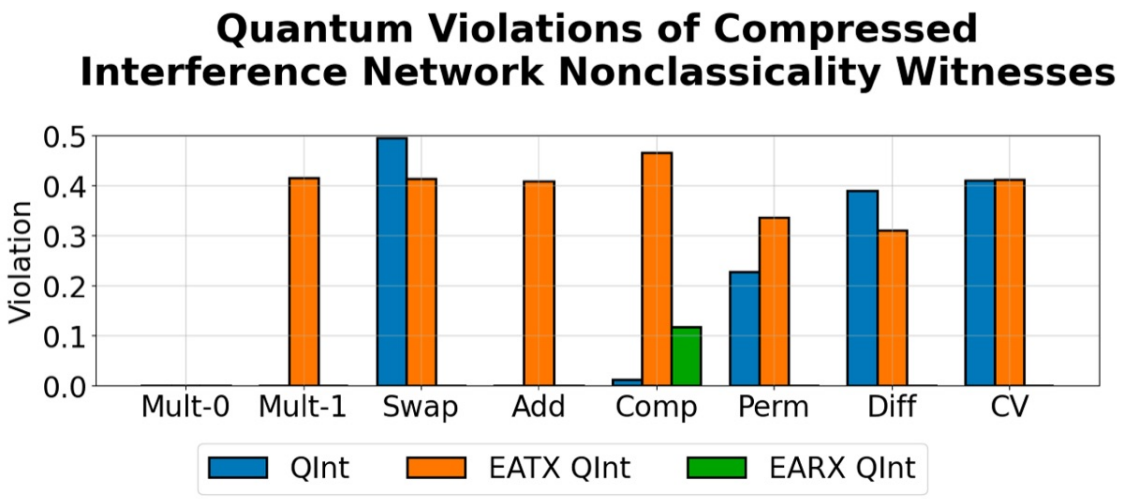



Figure 17.3: Quantum violations of classical bounds for the compressed interference network. The $x$-axis labels the nonclassicality witness while the $y$-axis shows the violation where the classical bound is fixed to zero. The blue bars show the case where unassisted quantum signaling is used. The orange bars show the case when quantum signaling is used and the senders are assisted with entanglement. The green bars show the case when quantum signaling is used and the receivers are assisted with entanglement.

entanglement-assisted senders are found to violate $\mathbf{G}^{\text{BFN}}_{\text{Mult-0}}$, $\mathbf{G}^{\text{BFN}}_{\text{Comp}}$, and $\mathbf{G}^{\text{BFN}}_{\text{Diff}}$. These variational ansatzes are able to achieve the strongest violations. The entnaglement-assisted receiver variational ansatz is found to violate the $\mathbf{G}^{\text{BFN}}_{\text{Mult-1}}$ nonclassicality witness.

$$11 \geq \mathbf{G}^{\mathrm{BFN}}_{\text{Mult-0}} = \begin{pmatrix} 1 & 3 & 1 & 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 2 & 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 & 1 & 0 & 2 & 0 \\ 0 & 2 & 1 & 0 & 1 & 1 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 & 2 & 0 & 0 & 0 & 2 \\ 0 & 2 & 1 & 1 & 1 & 1 & 0 & 1 & 1 \\ 0 & 2 & 1 & 0 & 1 & 1 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 & 1 & 0 & 1 & 0 & 1 \\ 1 & 2 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \end{pmatrix} \qquad 11 \geq \mathbf{G}^{\mathrm{BFN}}_{\text{Mult-1}} = \begin{pmatrix} 2 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 3 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 & 2 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 & 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & 1 & 0 & 2 & 0 & 1 & 0 \\ 0 & 0 & 1 & 1 & 0 & 2 & 0 & 2 & 0 \\ 1 & 1 & 1 & 1 & 1 & 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 2 & 1 & 1 & 1 & 0 & 1 \\ 1 & 1 & 1 & 2 & 1 & 1 & 1 & 1 & 1 \end{pmatrix}$$

$$10 \geq \mathbf{G}^{\mathrm{BFN}}_{\text{Add}} = \begin{pmatrix} 1 & 0 & 1 & 0 & 1 & 1 & 1 & 1 & 0 \\ 0 & 2 & 0 & 1 & 0 & 0 & 0 & 0 & 2 \\ 1 & 1 & 2 & 0 & 1 & 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 0 & 0 & 2 & 1 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 2 \\ 0 & 1 & 2 & 0 & 0 & 1 & 1 & 1 & 1 \\ 1 & 0 & 1 & 0 & 1 & 1 & 1 & 1 & 1 \\ 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 2 \\ 1 & 1 & 2 & 0 & 1 & 1 & 1 & 1 & 1 \end{pmatrix} \qquad 12 \geq \mathbf{G}^{\mathrm{BFN}}_{\text{Diff}} = \begin{pmatrix} 3 & 0 & 0 & 0 & 2 & 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 1 & 1 & 0 & 0 & 1 & 1 \\ 2 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 \\ 2 & 0 & 1 & 0 & 0 & 2 & 0 & 1 & 0 \\ 1 & 0 & 1 & 2 & 0 & 1 & 0 & 2 & 0 \\ 2 & 1 & 1 & 1 & 1 & 2 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 & 0 & 2 & 0 & 1 & 1 \\ 2 & 1 & 2 & 1 & 1 & 2 & 1 & 1 & 0 \end{pmatrix}$$

$$14 \geq \mathbf{G}^{\mathrm{BFN}}_{\text{Comp}} = \begin{pmatrix} 3 & 0 & 0 & 0 & 2 & 0 & 1 & 0 & 1 \\ 2 & 1 & 0 & 0 & 1 & 0 & 1 & 1 & 1 \\ 2 & 1 & 1 & 0 & 0 & 1 & 0 & 1 & 0 \\ 1 & 1 & 2 & 0 & 0 & 2 & 0 & 1 & 1 \\ 0 & 2 & 2 & 0 & 0 & 2 & 0 & 1 & 1 \\ 0 & 3 & 2 & 0 & 0 & 3 & 0 & 0 & 0 \\ 1 & 1 & 2 & 0 & 1 & 1 & 1 & 1 & 0 \\ 1 & 2 & 2 & 1 & 1 & 1 & 1 & 1 & 0 \\ 2 & 2 & 3 & 1 & 1 & 2 & 1 & 1 & 0 \end{pmatrix} \qquad 20 \geq \mathbf{G}^{\mathrm{BFN}}_{\text{Perm}} = \begin{pmatrix} 2 & 1 & 0 & 1 & 3 & 0 & 1 & 1 & 2 \\ 1 & 5 & 2 & 0 & 0 & 2 & 1 & 1 & 0 \\ 0 & 3 & 3 & 1 & 3 & 3 & 0 & 2 & 0 \\ 1 & 2 & 0 & 1 & 4 & 0 & 1 & 0 & 2 \\ 0 & 3 & 1 & 1 & 1 & 3 & 1 & 1 & 0 \\ 0 & 3 & 3 & 2 & 1 & 2 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 & 3 & 0 & 1 & 1 & 1 \\ 0 & 3 & 1 & 0 & 0 & 2 & 2 & 1 & 1 \\ 2 & 3 & 3 & 1 & 3 & 3 & 1 & 2 & 0 \end{pmatrix}$$

$$12 \geq \mathbf{G}^{\mathrm{BFN}}_{\text{Swap}} = \begin{pmatrix} 2 & 0 & 1 & 0 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 2 & 0 & 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 2 & 0 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 2 & 0 & 1 & 0 & 0 \\ 0 & 1 & 1 & 1 & 1 & 1 & 0 & 1 & 1 \\ 0 & 0 & 3 & 0 & 0 & 2 & 0 & 1 & 0 \\ 0 & 0 & 2 & 0 & 0 & 3 & 0 & 1 & 0 \\ 1 & 1 & 2 & 1 & 1 & 2 & 1 & 1 & 1 \end{pmatrix} \qquad 18 \geq \mathbf{G}^{\mathrm{BFN}}_{\mathrm{CV}} = \begin{pmatrix} 3 & 0 & 1 & 0 & 3 & 1 & 0 & 0 & 0 \\ 1 & 3 & 0 & 0 & 3 & 1 & 0 & 0 & 0 \\ 0 & 0 & 3 & 0 & 4 & 1 & 0 & 0 & 1 \\ 1 & 0 & 1 & 3 & 2 & 1 & 0 & 0 & 0 \\ 3 & 1 & 1 & 1 & 3 & 0 & 0 & 0 & 0 \\ 3 & 1 & 2 & 0 & 0 & 2 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 3 & 1 & 2 & 1 & 1 \\ 2 & 0 & 1 & 0 & 0 & 1 & 2 & 1 & 1 \\ 4 & 1 & 2 & 1 & 4 & 1 & 1 & 0 & 0 \end{pmatrix}$$

Table 17.4: Derived nonclassicality witnesses for the butterfly network. Each inequality $(\gamma, \mathbf{G})$ is expressed as $\gamma \geq \mathbf{G}$.

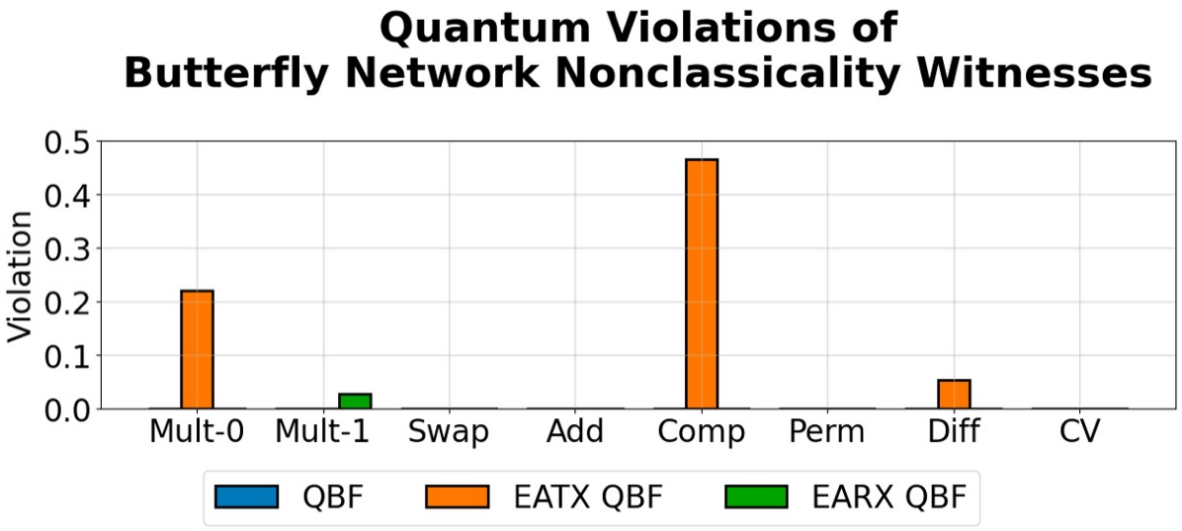



Figure 17.4: Quantum violations of classical bounds for the butterfly network. The $x$-axis labels the nonclassicality witness while the $y$-axis shows the violation where the classical bound is fixed to zero. The blue bars show the case where unassisted quantum signaling is used. The orange bars show the case when quantum signaling is used and the senders are assisted with entanglement. The green bars show the case when quantum signaling is used and the receivers are assisted with entanglement.

# Chapter 18

# Variational Quantum Networking

The world's first quantum networks are actively being developed. These rudimentary quantum networks can perform noisy quantum communication between devices, but are limited in their ability to store and process quantum information [52]. Communications protocols such as quantum key distribution [180], [185] and quantum random number generation [371] are key applications for such quantum prepare and measure networks, however, many of these protocols require the underlying device hardware to be fully characterized. In small networks characterization may not be a major problem because the hardware may be owned and operated by a single organization allowing trust to be established with each of the devices. Furthermore, the number of qubits may be small and tomography can be performed efficiently allowing the network and its noise to completely characterized.

As a network scales, however, its devices may be owned by many different parties, not all of whom can be trusted. Moreover, a user could access a network remotely without having the ability to inspect the network's underlying devices. In such cases, the unknown, untrusted, or uncharacterized devices should be treated as black-box that accepts classical data and produces classical data. Thus it becomes a challenge to certify that network devices are using quantum resources such that quantum protocols such as key distribution or randomness generation are secure. One solution is to certify the quantum resources, functionality, and security using nonclassicaliity witnesses to observe quantum behaviors using only the classical data produced by the network devices. Such *device-independent* protocols have been studied extensively within the context of quantum key distribution (DIQKD) [182], [247]–[251], [253]–[255] and quantum random number generation (DIQRNG) [186], [187], [189]–[191], [245], [246]. Since nonclassical quantum behaviors cannot be reproduced by a similar classical systems, nonclassicality can be used to assert the presence of quantum resources and therefore, the security of DIQKD and DIQRNG. Similar device-independent approaches are taken in *self-testing* of quantum devices [226], [233], [237]–[239], a certification procedure in which the quality quantum resources and functionality of quantum devices is goal is tested and verified. Such device-independent security and certification protocols have also been extended to quantum networks **di**, [164], [240], [241], [243], [244], [257], [259].

As quantum networks scale, it is expected that quantum repeaters [112], [113] will perform entanglement swapping protocols [110], [111], [198] that distribute entanglement arbitrarily amongst network devices. Hence, quantum networks will allow for arbitrary connections to be made between distant devices on the network. As a result, network devices will constantly need to be integrated and protocols such as DIQKD and DIQRNG

will need be established with remote devices. Unfortunately, when quantum devices are integrated in a noisy quantum network, challenges arise in establishing protocol's and maintaining their performance over time.

One challenge in establishing any protocol between two or more quantum devices is that the devices may not agree on the orientation of the computational basis $\{|z\rangle\}_z$. In general, the computational basis can be chosen arbitrarily and may differ between different devices. Furthermore, the quantum channel through which the quantum states are transmitted may apply an unknown unitary or other noise model that applies a transformation that reorients the basis states. For any communications protocol to implemented, the networking devices need to establish a common basis to orient their state encodings, operations and measurement. A second challenge is that the noise in quantum hardware may alter the optimal state encodings and decodings used for quantum communications. Similarly, physical measurements, unitaries, or state preparations may have biases such that certain encodings and decodings have higher fidelity than others. To develop robust protocols in the presence of such noisy conditions, it will be important to exploiting these hardware biases.

As the noise model changes over time, the networking devices will need to continually reorient their common bases and itentify the optimal state preparations and measurements for the given task. Hence it is not enough to establish a protocol, but it must be maintained over extended periods of time. Thus, to develop robust integrations between networking devices, the devices will need to be tuned with regularity to adapt to a dynamic and unknown noise model. In order to scale quantum network applications, this task must be automated.

Variational quantum optimization (VQO) is a promising approach for automating the integrations between network devices. This hybrid quantum-classical algorithm tunes the settings of a quantum devices to optimize their performance at a particular task using differential programming methods [103], [372], [373]. Recently, VQO techniques have been shown to optimize quantum networks around their inherent noise [102], [299]. In this dissertation, we assert that these VQO methods can be used to automate the integration and maintain the connection between network devices such that the devices perform the optimally given the network's noise. Since VQO methods rely only upon the classical data and a limited characterization of the quantum system, they can be applied generically throughout quantum network applications.

**Definition 110. Variational Quantum Networking:** The use of variational quantum optimization techniques to automatically establish and maintain integrations between quantum network devices in a hardware-agnostic manner.

In this chapter we introduce variational quantum networking as a novel paradigm for quantum network automation and design. The main idea is that variational quantum optimization can be used to optimize quantum networking devices against their inherent noise enabling networking tasks to be automated. Furthermore, the same variational techniques can also be applied on quantum computing hardware to simulate large-scale networks with greater efficiency than a classical computer. Thus, variational principles applied to quantum networks can serve as both a design tool and a means of automation. Thus, it is important that as quantum networks develop, compatibility with variational methods is integrated into the design of their hardware components and their communication protocols.

To support this argument, we introduce some first principles in quantum network design. We then formally introduce our variational quantum networking framework and discuss how our variational quantum optimization framework adheres to these fundamental design principles. Next, we give concrete examples on a classical simulator showing the success of our general framework. Finally, we give simple examples where variational quantum optimization is applied on both quantum computing and networking hardware. Overall,

variational quantum networking shows promise of advantage in quantum network automation and design, however, there are some challenges that must be overcome in practice.

## *18.1 First Principles in Quantum Network Design*

As quantum networks are developed it is important to bear in mind that it will be used by humans. Thus, we must design quantum networks from a human-centric perspective.

As an example, we can consider the internet as a classical network that has been successfully scaled globally to billions of users world-wide. To achieve this accomplishment, classical networks rely upon layers of abstraction that remove the end-users in the network from the physical data bits that are being transmitted between network devices. To list these layers from by increasing abstraction from the underlying physical hardware we have the physical layer, the link layer, the network layer, the transport layer, and the application layer. To enable interoperability between network devices, each layer is built upon a standardized protocol. Furthermore, when working in any given layer, the details of how the lower layers operate can largely be ignored.

Similarly, quantum networks will need to rely on layers of abstraction to create a user friendly, hardware-agnostic interface. Although, the design of the quantum networking stack and its supporting standardized protocols are far from being realized, we can still discuss the general goals that the layers of abstraction in a quantum network should achieve. Without asserting what the quantum network stack should look like, we can describe first principles of design that the abstraction of a quantum networking stack should provide. Note that in the following descriptions, the user is understood as a human instructing the network what to do.

- **Network Commands should be Declarative not Imperative:** A user should tell a network what to do, not how to do it. This is important because networks can become very complex at a low-level, even for simple tasks. Rather than forcing the user to cater their protocol to the network's low-level structure, the network should be responsible for catering its low-level resources to the task described by the user.

- **Networks should be Adaptive not Static:** In a noisy quantum network, the only constant is change. Failure to adapt implies constant failure. Rather than making the user responsible for adapting their protocol to an evolving network, the network should adapt itself such that the user's protocol is maintained.

- **Interfaces should be Standardized Not Specialized:** A specialized interface works once whereas a standardized protocol works infinitely. If networks devices are to make arbitrary connections with each other, their interfaces ought to be standard, otherwise the networking devices are not interoperable.

- **Protocols are Hardware-agnostic not Hardware-specific:** Low-level implementation should not dictate high-level implementation. There is no reason why networks cannot be hybridized to apply many different quantum hardware platforms. Furthermore, two networks implemented on different hardware platforms should not require users to design their applications around a particular hardware platform.

These design principles are not only important for making human-centric networking applications, but

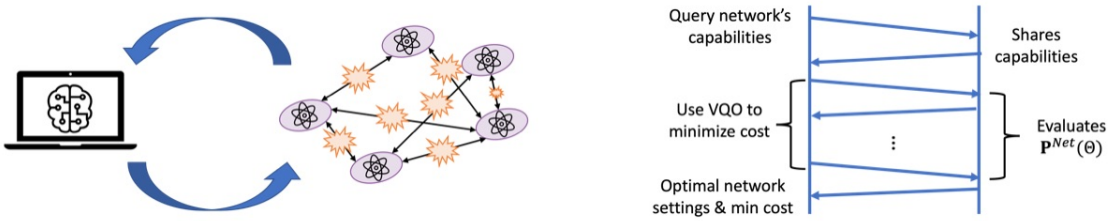



Figure 18.1: (Left) A noisy Network is optimized using variational quantum optimization. (Right) A schematic for the variational quantum networking protocol is given where left bar denotes the client who wishes to establish a protocol and the right bar denotes the noisy network on which the protocol is to be implemented.

also for network automation. Indeed, the entire operation of a quantum network should not require any manual intervention. All networking tasks should be automated, or else quantum networks cannot scale.

## *18.2 Variational Quantum Network Protocol Design and Automation*

In this section we describe a general variational quantum networking protocol. In general, variational quantum networking protocols can be used to assist component design, automate the certification and characterization of quantum resources, establish protocols, and maintain network performance. To support these claims, we give simple yet practical examples demonstrating their success. We begin by describing a generic variational quantum networking protocol, which we base on our hybrid variational quantum optimization algorithm (see Algorithm 1).

**Algorithm 5. Generic Variational Quantum Networking Protocol:** A hybrid protocol in which a classical optimizer a quantum network for a particular task. Given a network behavior $\mathbf{P}^{\text{Net}}$, the network's performance at the task is quantified using a cost function $\text{Cost}\left(\mathbf{P}^{\text{Net}}\right)$ were the network performs its task optimally when the cost function is minimized. The optimization objective is then expressed as $\min_{\Theta} \text{Cost}\left(\mathbf{P}^{\text{Net}}(\Theta)\right)$ where $\mathbf{P}^{\text{Net}}(\Theta)$ is the quantum network's behavior.

As input the algorithm requires a cost function $\text{Cost} : \mathcal{P}_{\mathcal{Y}|\mathcal{X}} \to \mathbb{R}$ and a performance goal $C \in \mathbb{R}$. As hyper parameters, he algorithm accepts a step size $\eta$ and the maximum number of iterations `max_iter`.

1. The controller queries the network for its topology $\text{Net}(\vec{N}, \vec{E})$ and device capabilities, which may include each device's communication resources, processing resources, and the parameterization for its operations.

2. Given the network's topology and resources, the optimizer constructs a suitable variational ansatz circuit $U^{\text{Net}}(\Theta_{\vec{x}})$.

3. Initialize the settings $\Theta$ at random, and set `iter` $= 0$.

4. While `iter` $<$ `max_iter` do:

   (a) Evaluate $\nabla_{\Theta}\text{Cost}\left(\mathbf{P}(\Theta)\right)$ using a hardware-compatible method such as the parameter-shift rule.

   (b) Update the settings as $\Theta = \Theta - \eta\nabla_{\Theta}\text{Cost}\left(\mathbf{P}(\Theta)\right)$.

   (c) Evaluate the cost for the updated settings $\text{Cost}\left(\mathbf{P}(\Theta)\right)$.

   (d) If $\text{Cost}\left(\mathbf{P}(\Theta)\right) \leq C$, then exit while loop, otherwise, increment `iter` $= 1+$ `iter`.

5. Return the minimal cost $\text{Cost}\left(\mathbf{P}(\Theta)\right)$ and the optimal settings $\Theta$.

Upon completion of the variational quantum optimization procedure, the quantum network performs its task for settings $\Theta$ with a performance certified as $\text{Cost}\left(\mathbf{P}(\Theta)\right)$.

Note that the generic variational quantum networking in Algorithm 5 is not guaranteed to obtain a global optimum. However, the optimized quantum network is guaranteed to perform its task as quantified by the cost function. Thus, not only does the algorithm train a quantum network for a particular task, it also returns the cost as a certificate quantifying the performance. Therefore, it is important that the user specifying the cost function also specifies a performance goal $C$. If the returned certificate is not satisfactory where $\text{Cost}\left(\mathbf{P}^{\text{Net}}(\Theta)\right) > C$ then, the optimized settings $\Theta$ must be rejected and the algorithm must be rerun. This is because the network did not perform its task within its performance goal. This may occur in a few cases including when a local optimum is found, the network is too noisy to perform its task, or the network fails to follow the optimization instructions requested by the optimizer. The latter of which could indicate that the network has malicious intent.

Variational quantum networking protocols can provide several advantages when applied on noisy quantum networks. Namely, the network's noise model does not need to be characterized prior to establishing the protocol through VQO. Rather the quantum networks free operations will be optimized to minimize the cost function in the presence of the unknown noisy process. Furthermore, if the noise changes over time, variational quantum optimization can be used to adapt the protocol to the changing noise model. Another advantage is that variational quantum networking is a hardware-agnostic protocol archetype, meaning that it could be applied within in hardware platform providing a network that has tunable parameters. Finally, variational quantum networking protocol is a declarative approach to establishing protocols on quantum networks. That is, the network can be automatically established by describing only the desired result. On the other hand, an imperative approach to networking would require a user would need to explicitly command the network how to perform the protocol, which would typically require the user to have characterized the network.

### *18.2.1 Security Concerns for Variational Quantum Networking*

Suppose now that the network is malicious. Then how would such a variational optimization protocol fare? And are the protocols it trains secure? The goal of the malicious network is for the user to accept the optimized protocol despite not having the topology or resources that it claims to have. For instance, let the malicious network be classical, its aim is then to simulate the quantum network in question to trick the user into thinking that the malicious network uses quantum resources. Such an attack amounts to a simulation based attack. The simulation attack is successful if $\text{Cost}\left(\mathbf{P}_{\text{Net}}(\Theta^\star)\right) \leq C$ at the end of the variational quantum optimization routine in Algorithm 5 because the user will accept the optimal settings $\Theta^\star$. That is, if the malicious network can simulate the network behavior such that it performs the required task with sufficient quality, the user will proceed to use the network for their desired application. If that application were to use quantum key distribution to share secret key, the malicious classical network would simply copy the secret keys without the user knowing.

For such a simulation-based attack to be successful, the network must simulate the expected behavior meticulously such that all of the evaluated gradients are correct. Otherwise, if the gradients are not evaluated properly, then the gradient will no longer point in the direction of steepest ascent causing the optimization to fail. Thus, the simulation-based attack may be difficult to carry out because it requires that the adversary be able to accurately simulate the network for the entire domain of its parameters. However, a failed simulation

attack could be thought of as a denial of service attack, in which the adversary intercepts the users instructions sent to the network and attempts to simulate the gradient. However, if the optimization fails the user will not accept establish a connection with the adversary. In this manner, the adversary can prevent a user from establishing any quantum protocols on a quantum network.

Although the variational quantum networking protocol in Algorithm 5 allows the malicious network to define its own parameterization and structure, the attack should only be successful if the the user trusts the network blindly. If the user does not trust the network, they can use the variational quantum optimization loop to implement device-independent certification and verification schemes that can expose a malicious network pretending to have resources and functionality it does not. As a result, a network user can defend against such simulation-based attack by implementing a device-independent handshake protocol for certifying and the quantum resources provided by the network.

### *18.2.2 Establishing Trust for Device-Independent Cryptography*

**Definition 111. Device-Independent Handshake Protocol:** A suite of device-independent tests that certify the quantum resources and topology of a quantum network.

The concept of a device-independent handshake protocol is very general and must be catered to specific applications and systems at hand. In general, we describe key elements that are important to consider in such protocols. Furthermore, if a subset of the quantum network devices are trusted, then the trusted devices can be used in device-independent certification protocols test the quantum properties of the untrusted devices building trust throughout the network.

Many device-independent cryptography tasks rely upon a loophole-free violation of a nonclassicality witness [247], which means that a classical simulation-based attack cannot reproduce the network's behavior. For instance, when two devices are nonsignaling or spacelike separated, a violation of the CHSH inequality or similar nonclassicalilty witness can assert that entanglement is present between the two devices. However, since one bit of classical communication between the measurement devices is sufficient to simulate CHSH violations [262], [263] it is important that the devices are actually spacelike separated. Thus, any device-independent handshake protocol must first verify the location of respective devices. Quantum position verification [374] provides a means for verifying that a device exists at a particular position. Hence, this technique can be used to verify that two measurement devices are indeed spacelike separated.

Even when network devices have been certified as using quantum resources and their positions have been verified, the user must still be cautious in the use of quantum devices. The main concern is that network devices are treated as black-boxes, hence, the user knows their classical inputs and outputs, but not underlying process. Although, a user may certify the use of quantum resources, there is nothing preventing the black-box from copying and sharing its classical input-output data with an adversary. In a device-independent key distribution protocol, this data would include the secret key. To overcome this challenge, trust needs to be placed on the devices used in the protocol.

The most practical trust models will likely resemble those used for classical networking and ecommerce [375]–[377]. Essentially, trust refers to the assurance that a network device will perform the task its asked to do in much a similar way as customers and vendors perform a transaction of resources on the world-wide web. In an ecommerce network a community of users builds trust with vendors through repeated transactions. For example, when users enter credit card information on a website, they trust that the website is not going to exploit that data. If the private information is leaked, then the service provider is held accountable in

the court of law and the users lose trust in the service provider. For a service provider to be successful in this ecommerce ecosystem, they must uphold the trust the trust of their users. In quantum networks, the service providers are likened to quantum networking devices being used in a protocol. Thus, through repeated interactions with a community of users, trust can be established in a quantum networking device. Overall, the classical data produced by remote quantum networking devces is only as secure as the trusst in the device's manufacturer and the device's owner. The device-independent protocols merely certify that the uncharacterized network devices are indeed quantum, which in turn, guarantees that random number generation is random and that key distribution is eavesdropper free.

## *18.3 Variational Quantum Networking on the Classical Simulator*

### *18.3.1 Designing State Preparations and Measurements Using Principles in State Discrimination*

In this section we demonstrate how variational optimization can be combined with quantum state discrimination to train quantum devices to implement certain pure state preparations or positive operator-valued measures (POVMs) having rank-one measurements. We begin by providing a brief overview of quantum state discrimination, and then show how POVM measurements can be trained using these principles.

Consider a point-to-point quantum communication system where the sender device $A$ is given the input $x \in \mathcal{X}$ drawn from a uniform distribution $P(x) = 1/|\mathcal{X}|$ and the receiver device $B$ outputs the classical value $y \in \mathcal{Y}$ where $|\mathcal{X}| = |\mathcal{Y}|$. The goal in state discrimination is maximize the probability that the receiver outputs the senders input where the success probability is expressed as

$$P_{\text{Succ}} = \frac{1}{|\mathcal{X}|} \sum_{x \in \mathcal{X}} P(y = x|x). \tag{18.1}$$

The challenge is that the channel $\text{id}^{A\to B}$ connecting the sender to the receiver has signaling dimension $\kappa_\Lambda(\text{id}^{A\to B}) = d$ where $d < |\mathcal{X}|$. Note that this communication scenario is discussed in detail in Chapter 14 and that the success probability is related the communication value of a classical channel $\text{cv}(\mathbf{P})$ and maximum likelihood game $\mathbf{G}^{\text{ML}} = \mathbb{I}_{|\mathcal{X}|}$ as

$$P_{\text{Succ}} = \frac{1}{|\mathcal{X}|} \left\langle \mathbf{G}^{\text{ML}}, \mathbf{P} \right\rangle = \text{cv}(\mathbf{P}). \tag{18.2}$$

In the quantum setting, the $|\mathcal{X}|$ inputs are encoded into a quantum sate ensemble $\{\rho_x \in D(\mathcal{H}^A)\}_{x=\mathcal{X}}$ and the outputs are decoded using a POVM measurement $\{\Pi_y^B\}_{y\in\mathcal{Y}} \in \text{POVM}(\mathcal{H}^B)$ where $\sum_{y\in\mathcal{Y}} \Pi_y^B = \mathbb{I}_{|\mathcal{X}|}$ and $\Pi_y^B \geq 0$. Note that in general the rank of each POVM element is bounded as $\text{Rank}\left(\Pi_{y=x}^B\right) \leq \text{Rank}\left(\rho_x^A\right)$ [91]. Thus, if an ensemble of pure states are prepared where $\rho_x = |\psi_x\rangle\langle\psi_x|$, then the optimal POVM elements take the form $\gamma_y |\phi_y\rangle\langle\phi_y|$where $|\phi_y\rangle \in \mathcal{H}_d^B$, $\gamma_y \in [0, 1]$, and $\sum_{y\in\mathcal{Y}} \gamma_y = d$. In this rank-one case, the success probability is calculated as

$$P_{\text{Succ}} = \frac{1}{|\mathcal{X}|} \sum_{x\in\mathcal{X}} P(y = x|x) = \sum_{x\in\mathcal{X}} \gamma_{y=x} |\langle\phi_{y=x}|\psi_x\rangle|^2. \tag{18.3}$$

The goal in quantum state discrimination is to find the optimal state encodings and POVM decodings that

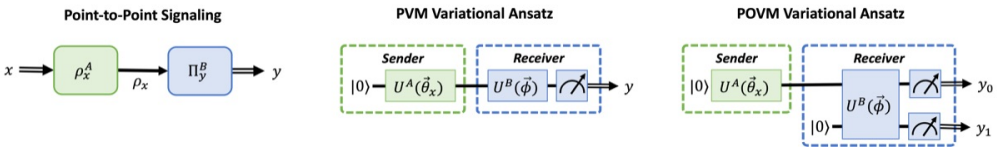



Figure 18.2: (Left) The DAG describing the considered point-to-point signaling scenario setting. (Center) The single-qubit variational ansatz circuit parameterizing all PVM measurements. (Right) The two-qubit variational ansatz circuit parameterizing all POVM measurements.

maximize the success probability. This problem can be rephrased in terms of the optimization objective,

$$P^{\star}_{\text{Succ}} = \max_{\substack{\{\Pi_y\}_{y\in\mathcal{Y}}\in\text{POVM}(\mathcal{H}^B) \\ \{|\psi_x\rangle\in\mathcal{H}^A\}_{x\in\mathcal{X}}}} \frac{1}{|\mathcal{X}|}\sum_{x\in\mathcal{X}} \gamma_{y=x}|\langle\phi_{y=x}|\psi_x\rangle|^2 \quad \text{s.t.} \quad \text{Rank}\left(\Pi_y\right) = 1. \tag{18.4}$$

Note that the maximal discrimination success probability is bounded as $P^{\star}_{\text{Succ}} \leq d$ where $d$ is the Hilbert space dimension of the encoded state ensemble $\{|\psi_x\rangle\}_{x\in\mathcal{X}}$. The optimal success probability can be achieved using either classical (orthonormal) state encodings where $\langle\psi_x|\psi_{x'}\rangle = \delta_{x,x'}$, or using non-orthogonal state encodings where $\langle\psi_x|\psi_{x'}\rangle > 0$. When non-orthogonal state encodings are used the maximal success probability can be ascertained only if a set of POVM scalars $\{\gamma_y\}_y$ exists such that $\{\gamma_x|\psi_x\rangle\langle\psi_x|\}_{x\in\mathcal{X}}$ form a valid POVM. Such a measurement is clearly optimal,

$$\frac{1}{n}\sum_{x\in\mathcal{X}} P(x|x) = \frac{1}{n}\sum_{x\in\mathcal{X}} \gamma_x|\left\langle\psi_x|\psi_x\right\rangle|^2 = d, \tag{18.5}$$

where the requirement $\sum_y \gamma_y = d$ is applied. Furthermore, it can be shown that parallel alignment between $|\psi_x\rangle$ and $|\phi_{y=x}\rangle$ is a requirement for the optimality of non-orthogonal POVM elements [91].

POVMs are valuable across a wide-range of applications in quantum information processing. Indeed, they are necessary to obtain many of the examples of nonclassicality derived for multipoint communication networks. Despite their importance, POVM measurements are not usually supported by quantum hardware, which usually implements projective measurements. In principle, POVMs can be constructed from projective measurements by adding ancillary qubits and performing a Naimark dilation on the measurement operators [149]. However, this procedure typically results in a quantum circuit that may be difficult to implement. Hence we argue that variational quantum optimization can be used to design POVM circuits on quantum computers or quantum networking hardware. More precisely, we consider a design task in which we want a quantum networking device to implement a particular rank-one POVM $\{\gamma_y|\phi\rangle\langle\phi|_y\}_{y\in\mathcal{Y}}$. Once implemented, the device implementing the POVM can be used for the desired measurement task.

To solve the state discrimination optimization problem in Eq. (18.4), we apply variational quantum optimization to train a quantum circuit to implement a POVM. The variational ansatz circuits for both qubit PVM and POVM measurements are shown in Fig. 18.2. To train the variational measurement ansatz to implement a particular measurement, the sender prepares the states $|\psi_x\rangle = |\phi_{y=x}\rangle$. This state ensemble is fixed and can be referred to as the training ensemble because it guides the measurement ansatz to the desired POVM. The cost function in the optimization is simply $\text{Cost}\left(\mathbf{P}(\Theta)\right) = -P_{\text{Succ}}(\Theta)$ where $\Theta$ parameterizes the measurement. If our optimization achieves the maximal discrimination score of $P^{\star}_{\text{Succ}} = 1$ within some small tolerable error, then the quantum circuit implements the POVM $\{\gamma_y|\phi_y\rangle\langle\phi_y|\}_{y\in\mathcal{Y}}$.

As an examples we plot train trine and symmetric informationally complete (SIC) POVM measurements.

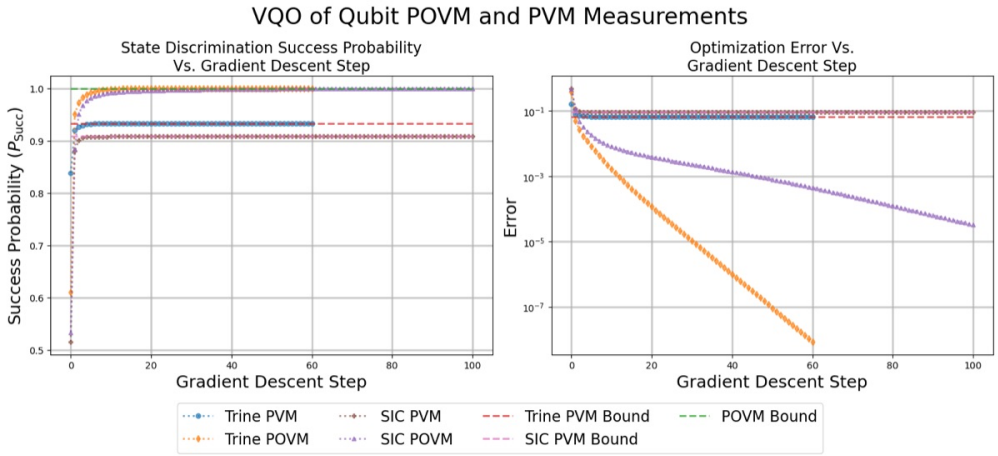



Figure 18.3: Training a qubit POVM using principles in state discrimination. PVM and POVM measurements are optimized for maximal state discrimination of both trine and SIC qubit ensembles. POVM measurements converge to the maximal success probability whereas the PVM measurements converge to their respective bounds a defined by the Helstromm measurement.

The trine state ensemble is,

$$|\psi_1\rangle = |0\rangle\,, \quad |\psi_2\rangle = \frac{1}{2}(|0\rangle + \sqrt{3}\,|1\rangle), \quad |\psi_3\rangle = \frac{1}{2}(|0\rangle - \sqrt{3}\,|1\rangle). \tag{18.6}$$

The maximal state discrimination success probability is then achieved with the POVM,

$$\Pi_1 = \frac{2}{3}\,|\psi_1\rangle\langle\psi_1|\,, \quad \Pi_2 = \frac{2}{3}\,|\psi_2\rangle\langle\psi_2|\,, \quad \Pi_3 = \frac{2}{3}\,|\psi_3\rangle\langle\psi_3|\,. \tag{18.7}$$

The SIC state ensemble consists of four states whose Bloch vectors point to the vertices of a tetrahedron inscribed on the Bloch sphere. This state ensemble is symmetric such that each $\langle\psi_x|\psi_{x'}\rangle$ is constant for all $x$ and $x'$. The SIC states are expressed as

$$|\psi_1\rangle = |0\rangle\,, \quad |\psi_2\rangle = \frac{1}{\sqrt{3}}(|0\rangle + \sqrt{2}\,|1\rangle)\,|\psi_3\rangle = \frac{1}{\sqrt{3}}(|0\rangle + \sqrt{2}e^{i2\pi/3}\,|1\rangle), \quad |\psi_2\rangle = \frac{1}{\sqrt{3}}(|0\rangle + \sqrt{2}e^{i4\pi/3}\,|1\rangle). \tag{18.8}$$

The optimal POVM is then,

$$\Pi_1 = \frac{1}{2}\,|\psi_1\rangle\langle\psi_1|\,, \quad \Pi_2 = \frac{1}{2}\,|\psi_2\rangle\langle\psi_2|\,, \quad \Pi_3 = \frac{1}{2}\,|\psi_3\rangle\langle\psi_3|\,, \quad \Pi_4 = \frac{1}{2}\,|\psi_4\rangle\langle\psi_4|\,. \tag{18.9}$$

We now train both qubit PVM and POVM measurement ansatzes to find that POVM ansatzes converge to $P^{\star}_{\text{Succ}} = 1$ while the PVM ansatzes converge to the Helstrom bound for qubit PVM state discrimination. This bound is $P^{Helstrom}_{\text{Succ}} = \frac{1}{2}(1 + \sin(\theta))$ where $\theta = \frac{\pi}{3}$ in the trine case and $\theta = \arccos(\frac{1}{\sqrt{3}})$ in the SIC ensemble case. Our results are plotted in Fig. 18.3

### *18.3.2 Variational Certification and Self-Testing*

When constructing quantum networks, it is important to verify and test the functionality of the individual devices and their connections. Such an approach is ideally device-independent such that the testing and certification tasks rely only on the observed statistics. Hence, we provide an operational framework for

self-testing that is based upon variational quantum networking.

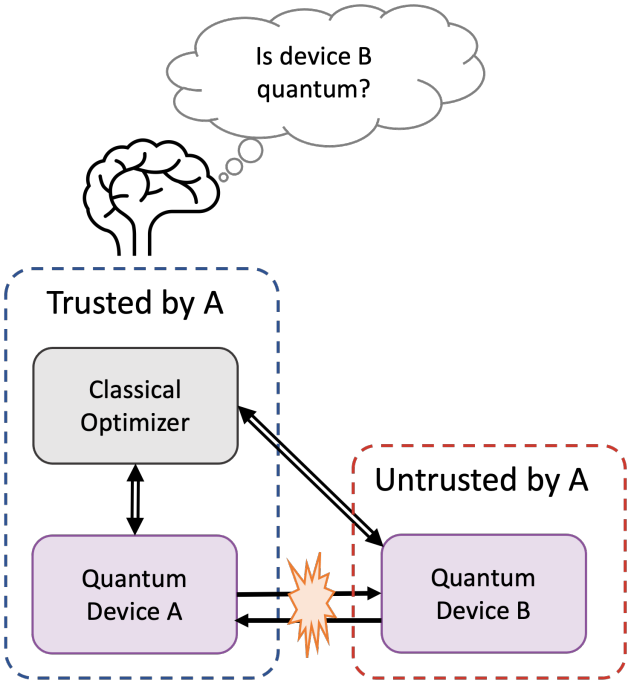



Figure 18.4: An observer holds a classical optimizer and wishes to certify the quantum functionality of an untrusted, uncharacterized device $B$, using a trusted and characterized quantum device $A$. Devices $A$ and $B$ have some noisy quantum communication between them, and both devices can be queried and controlled by the classical optimizer.

Consider the setting in which a researchers wants to test the quality of a device B's quantum functionality, by demonstrating nonclassical behaviors (see Fig. 18.4) between the untrusted device $B$ and a trusted device $A$. By demonstrating nonclassical behaviors, between the trusted and untrusted devices, the researcher can certify the quantum functionality implemented by device $B$. The variational quantum networking protocol would be implemented as

1. The optimizer queries device B for its quantum communication and measurement capabilities.
2. A relevant nonclassicality witness is selected to test device B's functionality.
3. The optimizer uses VQO to demonstrate nonclassicality.
4. Since device A's resources are known, violations occur only if device B faithfully executes its commands.
5. The protocol is successful if the violation is sufficiently large.

The protocol above is incredibly general and can be catered towards any nonclassicality witness that can be cast as a cost function and optimized by variational methods. The protocol requires trust in one or more of the involved devices to verify that the appropriate communication resources are being used. Although, we don't elaborate upon the specific certification tasks, the nonclassicality results throughout this dissertation highlight how variational quantum optimization can be successful at optimizing nonclassicality in the presence of noise.

A challenge with such methods is that noise in the quantum system may be difficult to separate from malicious behavior. Future work can investigate the possible attacks that could be applied to such variational quantum network certification frameworks.

### *18.3.3 Training Quantum Teleportation and Entanglement Swapping Protocols*

Teleportation and entanglement swapping are critical for distributing entanglement and communicating quantum information in networks. These protocols require LOCC resources. The examples shown in the section are obtained from the qNetVO Github repository [108].

The first task we examine is quantum teleportation (see Fig. 18.5). The optimization objective is to maximize the average teleportation, which we calculate for a set of six training states $\rho_{in} \in \{|0\rangle, |1\rangle, |+\rangle, |-\rangle, |\tilde{+}\rangle, |\tilde{-}\rangle\}$. We find that the average fidelity of the trained protocol approaches the maximum of 1, showing that a teleportation protocol is implemented by the optimized ansatz circuit.

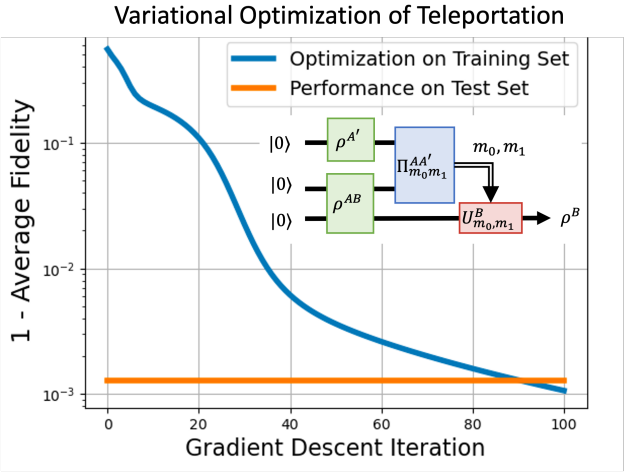



Figure 18.5: We minimize the cost $\text{Cost}(\Theta) = 1 - \frac{1}{6}\sum_{\rho_{in}} F(\rho_{in}, \rho_{out}) = 1 - \frac{1}{6}\sum_{\rho_{in}} \text{Tr}\left[\sqrt{\sqrt{\rho_{out}}\rho_{in}\sqrt{\rho_{out}}}\right]^2$, which is one minus the average teleportation fidelity over the set of input training states $\rho_{in} \in \{|0\rangle, |1\rangle, |+\rangle, |-\rangle, |\tilde{+}\rangle, |\tilde{-}\rangle\}$ . The inset shows the three-qubit variational ansatz circuit that is optimized. Since we perform the optimization on a classical simulator, the fidelity $F(\rho_{in}, \rho_{out})$ can be computed explicitly. The blue curve shows the minimization of the cost function while the horizontal orange line shows the average teleportation fidelity when the trained protocol is tested against 1000 input states randomly distributed about Bloch sphere.

One challenge faced by our teleportation training approach is that the fidelity cannot be obtained without knowing the density matrix of the quantum state. In other words, a full tomography would need to be performed on the teleportation output for each test state. We provide an alternative solution in our example of training an entanglement swapping protocol.

entanglement swapping is an application of quantum teleportation where the goal is to teleport one part of an entangled state from the central device in the bilocal network to an external device (see Fig. 18.6). As a cost function we use the violation of the CHSH inequality $\text{Cost}(\Theta) = -S_{CHSH}(\Theta)$ taken between the two edge nodes in the bilocal network. By demonsterating a CHSH violation, the entanglement between the two edge nodes can be certified. Note that we optimize the entanglement swapping protocol without having prior knowledge of the proper measurements, decoding operations, or state preparations. Furthermore, by demonstrating a successful optimization in the presence of noise, we show that we can train entanglement

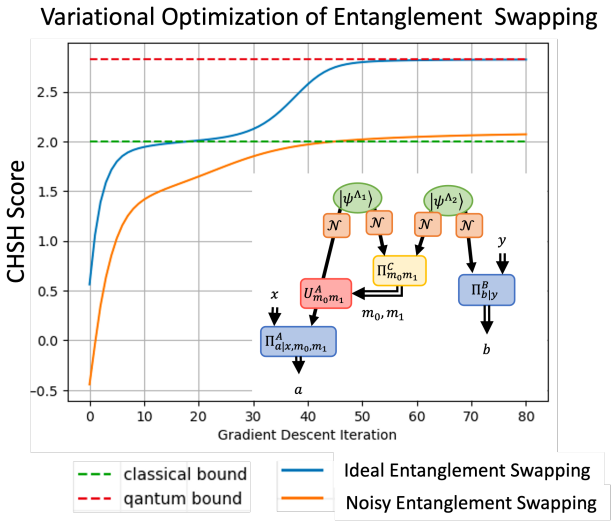



Figure 18.6: An optimization of entanglement swapping via violation of the CHSH inequality. The inset shows the network DAG that is optimized. The dashed green line shows the classical bound while the dashed red line shows the quantum bound. The blue curve shows the optimization trajectory in the noiseless case, while the the orange line shows the optimization when there is qubit noise is incorporated. In both cases, the classical bound is violated demonstrating that entanglement was successfully distributed.

swapping even in noisy setting.

### *18.3.4 Variational Characterization of the correlation structure in the n-Local Setting*

Consider the scenario in which an unknown quantum state prepared by sources in a network is measured by a collection of trusted measurement devices (see Fig. 18.7). The goal is to use the classical data obtained from the measurement devices to infer the correlation structure of the network. That is, how many independent sources are there in the network and which measurement devices are linked to each source. The content of this section is based largely upon the results of my collaboration with Argonne National Lab [378].

Quantum entanglement is a foundational resource in many quantum technologies [379]. The ability to infer the entanglement structure in a quantum network is an important characterization task because generalized entanglement protocols could be used to distribute complex entangled states to measurement devices across a network [110], [111]. While the detection and certification of multipartite entangled states has been heavily studied [380]–[386], less is known about the detection and certification when the network consists of many independent entanglement sources. Furthermore, to be practical in the near-term, such a entanglement witnessing approaches must be efficient and must not rely upon quantum memory or complex multi-qubit measurements.

A device-independent approach is therefore promising for practical characterization of quantum network topology. To this end, several methods have recently been developed that rely the classical data to infer quantum network topology. The typical approach is to collect classical data from the measurement devices and then tested for compatibility with various correlation structures. Numerous network compatibility tests have been developed including the violations of entropic bounds [387]–[389], network Bell inequalities [124], [272], [274], [278], and quantum Finner inequalities [390], [391], as well as semidefinite tests on covariance matrices [392]–[394] and inflation techniques [395], [396]. Although these tests can be implemented efficiently

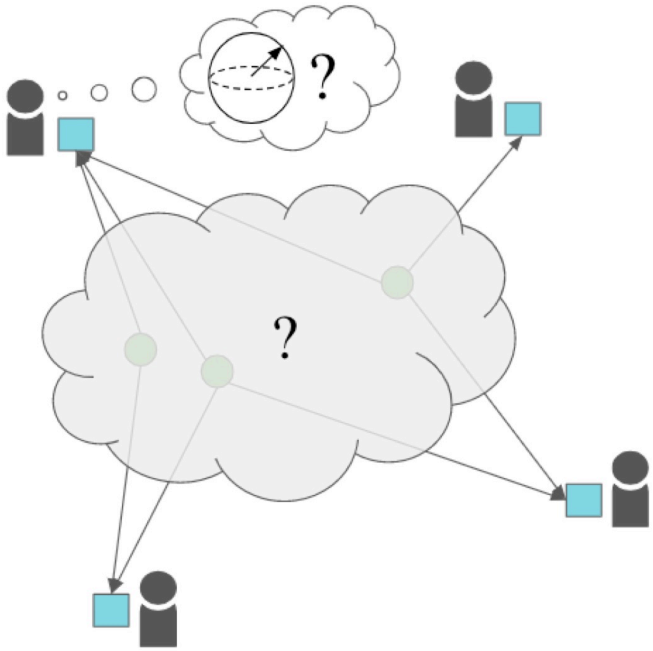



Figure 18.7: [378] In the $n$-local setting, an unknown network state prepared is distributed amongst measurement devices in a quantum network. The users operating the measurement devices would like to infer the topology of the network using only their measured data.

in practice, their main limitations are that the applied measurements are not considered to be tunable, leading to the possibility that poorly selected measurements prevent certain network correlation structures to be observed. Similarly, the tests for network compatibility assert whether one network topology can produce the observed correlations, however, there may exist multiple network topologies that are compatible with the observed network behavior creating challenges in inferring the topology with certainty.

A seperate appraoch defines the *characteristic vector* of a network, $\vec{Q} = (S(\rho^{A_j}) = -\text{Tr}\left[\rho^{A_j}\log_2(\rho^{A_j})\right])$, as the collection of von Neumann entropies evaluated at each independent measurement device. The characteristic vector is then used to infer whether two networks are equivalent under local unitary transformations [397] provided that no two devices share more than one source and that all sources emit the GHZ state

$$\left|\psi^{\text{GHZ}}\right\rangle = \frac{1}{\sqrt{2}}(|0\ldots0\rangle + |1\ldots1\rangle). \tag{18.10}$$

However, an analysis using only the von Neumann entropies at local devices will fail to distinguish between two different network topologies whose measurement nodes each receive the same number of qubits (see 18.8). As a solution, Yang et al. [397] propose using the multipartite Shannon mutual information to distinguish different network topologies. For two parties, the Shannon mutual information is defined as $I(A_i : A_j) = H(\mathbf{P}^{A_i}) + H(\mathbf{P}^{A_j}) - H(\mathbf{P}^{A_i,A_j})$ where $H(\mathbf{P}^{A_i}) = -\sum_{x\in\mathcal{X}_i} P_x \log_2(P_x)$ is the Shannon entropy and the behavior $\mathbf{P}^{A_i,A_j}$ is vector describing the joint probability distribution measured for the two devices while $\mathbf{P}^{A_i}$ and $\mathbf{P}^{A_j}$ are behaviors representing the marginal probability distribution of each measurement devices. One challenge faced by this method is that the Shannon mutual information depends upon the basis measurements performed at each independent devices. A poor choice of basis measurement may fail to witness correlations within the network.

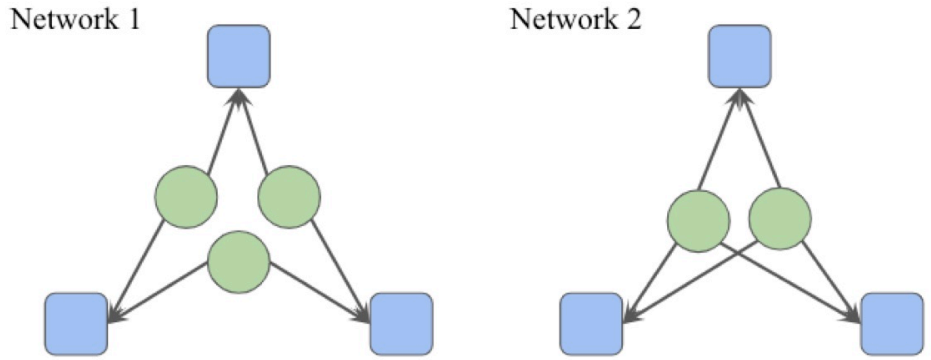



Figure 18.8: [378] Two networks having three measurement devices (blue squares), each measuring two qubits. If the sources (green circles) emit GHZ states, then the von Neumann entropy at each measurement device is identical between the two networks.

In reference [378] we extend the approach of Yang *et al.* [397] by develop a variational topology inference scheme. In our approach we apply variational quantum optimization to obtain the von Neumann entropy and maximize the mutual information in a hardware-compatible manner. Furthermore, when access to individual qubit data is known, we provide efficient an algorithm for efficient network inference. Our developed methods can be understood as a variational quantum networking protocol that can be used to certify or verify the correlation structure in the quantum network.

In our variational inference scheme we optimize the measurements of a quantum network to infer the correlation structure between devices. We demonstrate two approaches, one based on entropy and the other based on covariance. To begin, we introduce the characteristic matrix

$$Q = \begin{pmatrix} S(\rho^{A_1}) & I(A_1, A_2) & \dots & I(A_1, A_n) \\ I(A_2, A_1) & S(\rho^{A_2}) & \vdots & I(A_2, A_n) \\ \vdots & \dots & \ddots & \vdots \\ I(A_n, A_1) & I(A_n, A_2) & \ddots & S(\rho^{A_n}) \end{pmatrix}. \tag{18.11}$$

To obtain the diagonal of the characterize matrix, we express the von Neumann entropy as the minimal Shannon entropy

$$S(\rho^{A_j}) = \min_{\Pi^{A_j} \in \text{POVM}(\mathcal{H}^{A_j})} H(\mathbf{P}^{A_j}) \quad \text{where} \quad \mathbf{P}^{A_j} = \sum_{x_i} \text{Tr}\left[\Pi^{A_j} \rho^{A_j}\right] |x_i\rangle\,, \tag{18.12}$$

then the diagonal of the characteristic matrix is expressed as the minimization,

$$\vec{Q} = \min_{\{\Pi^{A_j} \in \text{POVM}(\mathcal{H}^{A_j})\}_{j=1}^{|\vec{A}|}} \sum_{j=1}^{|\vec{A}|} H(\mathbf{P}^{A_j}). \tag{18.13}$$

If the optimization in Eq. (18.13) is applied as a cost function, the von Neumann entropy of the state local to each measurement device can be obtain using variational quantum optimization.

We consider two types of optimized mutual information quantities in the off-diagonal of the characteristic

matrix. The optimized Shannon mutual information can be expressed as the collection of pairs of measurement devices

$$\max_{\{\Pi^{A_j}\in \mathrm{POVM}\mathcal{H}^{A_j}\}_{j=1}^{|\vec{A}|}} \sum_{i<j=1}^{n} I(\mathbf{P}^{A_i};\mathbf{P}^{A_j}) \tag{18.14}$$

where the behavior $\mathbf{P}^{A_i}$ are evaluated as the marginals of the joint distribution $\mathbf{P}^{\vec{A}}$ measured across all devices. We also consider the *measured mutual information*

$$I_m(A_i;A_j) = \max_{\substack{\Pi^{A_i}\in \mathrm{POVM}(\mathcal{H}^{A_i}) \\ \Pi^{A_j}\in \mathrm{POVM}(\mathcal{H}^{A_j})}} H(\mathbf{P}^{A_i}) + H(\mathbf{P}^{A_j}) - H(\mathbf{P}^{A_i,A_j}) \tag{18.15}$$

where the joint probability distribution is

$$P^{A_i,A_j}_{x_i,x_j} = \mathrm{Tr}\left[\Pi^{A_i}_{x_i}\otimes \Pi^{A_j}_{x_j}\rho^{A_iA_j}\right]. \tag{18.16}$$

The measured mutual information is distinct from the Shannon mutual information because for each bipartite pair of devices, the measurements are optimized independently. On the other hand, the Shannon mutual information optimizes one measurement $\Pi^{\mathrm{Net}} = \bigotimes_{j=1}^{n}\Pi^{A_j}$, from which the bipartite mutual informations $I(\mathbf{P}^{A_i};\mathbf{P}^{A_j})$ are derived. For this reason, the measured mutual information is quadratically slower to evaluate than the Shannon mutual information. However, there exist quantum states that the measured mutual information is generally larger than the Shannon mutual information, $I(\mathbf{P}^{A_i};\mathbf{P}^{A_j}) \leq I_m(\mathbf{P}^{A_i};\mathbf{P}^{A_j})$ because there might not be a single measurement that can obtain the maximal Shannon mutual information for all two-qubit pairings.

Covariance-based inference poses is a similar approach that can be used for inferring quantum network topology. However, instead of the characteristic matrix, we consider the covariance matrix

$$C = \begin{pmatrix} \mathrm{Var}(A_1) & \mathrm{Cov}(A_1,A_2) & \dots & \mathrm{Cov}(A_1,A_n) \\ \mathrm{Cov}(A_2,A_1) & \mathrm{Var}(A_2) & \vdots & \mathrm{Cov}(A_2,A_n) \\ \vdots & \dots & \ddots & \vdots \\ \mathrm{Cov}(A_n,A_1) & \mathrm{Cov}(A_n,A_2) & \ddots & \mathrm{Var}(A_n) \end{pmatrix} \tag{18.17}$$

where $\mathrm{Var}(A_i) = \left\langle (O^{A_i})^2\right\rangle_{\rho^{A_i}} + \left\langle O^{A_i}\right\rangle^2_{\rho^{A_i}}$ is the variance of the observable at device $A_i$ and the covariance between two observables at devices $A_i$ and $A_j$ is $\mathrm{Cov}(A_i,A_j) = \left\langle O^{A_i}\otimes O^{A_j}\right\rangle_{\rho^{A_iA_j}} - \left\langle O^{A_i}\right\rangle_{\rho^{A_i}}\left\langle O^{A_j}\right\rangle_{\rho^{A_j}}$ and $O^{A_j}$ is an Hermitian observable on the Hilbert space $\mathcal{H}^{A_i}$. Similar to the Shannon mutual information, the covariance quantifies the amount of correlation between the data of two devices, likewise the variance quantifies the amount of uncertainty in the measured data. To infer network topology with the greatest accuracy, we want to maximize each of the elements in the covariance matrix $C$. This objective can be quantified into the objective

$$\max_{\{\Pi^{A_i}\in \mathrm{POVM}(\mathcal{H}^{A_i})\}_{i=1}^{|\vec{A}|}} \mathrm{Tr}\left[C^TC\right] \tag{18.18}$$

where we consider the square of the euclidean distance to account for the fact that a negative covariance still indicates correlation whereas a zero covariance indicates no correlation.

In a noisy quantum network, the von Neumann entropy, mutual information, and covariance will decrease. Therefore, variational optimization can help obtain these values. To infer network correlations, we assert that

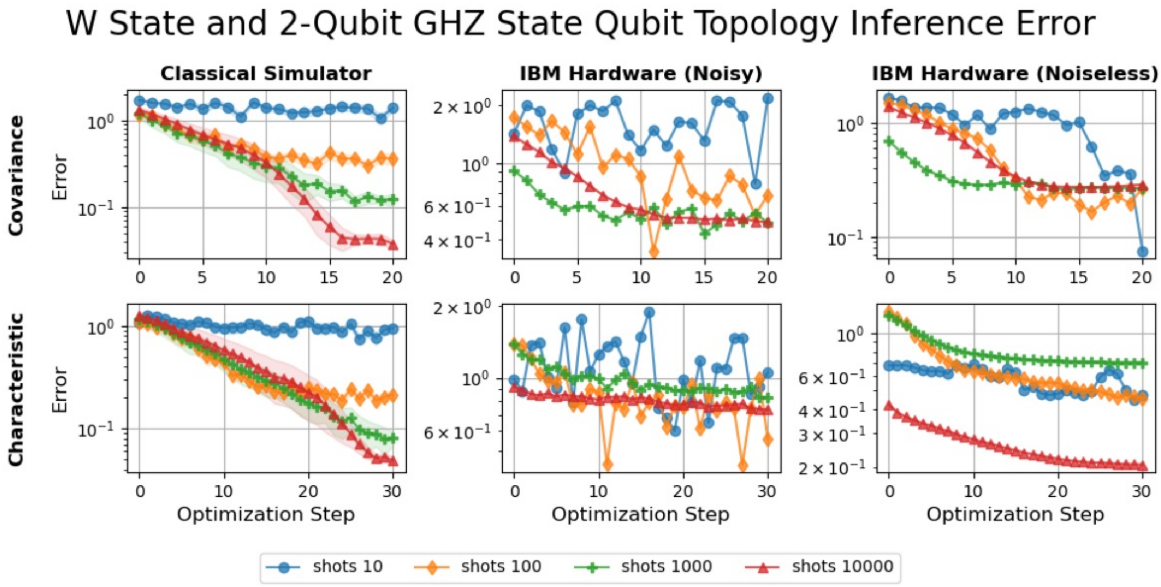



Figure 18.9: Variational quantum optimization of the covariance and characteristic matrices for a W state and two-qubit GHZ state prepared on IBM Hardware. In each plot the blue line with circle markers shows the 10-shot optimization, the orange line with diamond markers shows the 100-shot optimization, the green line with plus sign markers shows the 1,000-shot optimization, and the red line with triangle markers shows the 10,000-shot optimization. The $x$-axis of each plot shows the optimization step while the $y$-axis shows the inference error calculated as the Euclidean distance between the ideal covariance/characteristic matrix and the matrix in each optimization step. The top row plots the covariance inference matrix optimization while the bottom row plots the characteristic inference matrix optimization. The left column plots data collected from a noiseless, finite-shot classical simulator where data is averaged over 10 trial optimizations and the standard error is plotted as the respective shaded regions. The center column plots data obtained from optimization on IBM's noisy 5-qubit `ibmq_belem` device. The right column plots the settings optimized on the quantum hardware evaluated on a noiseless analytic classical simulator.

if either $\text{Cov}(A_i, A_j)$ or $I_m(A_i, A_j)$ is nonzero then a correlation must exist between the two devices. However, since there are statistical fluctuations, a minimal threshold of correlation should be establish such that the correlations must be strong enough to pass the threshold to count as correlations. Furthermore, a taking large number of shots can help the minimize the amount of error during inference. When measurements are separable across qubits, the network topology can be inferred with great accuracy because the covariance, von Neumann entropy, and mutual information are all bounded by one. This means that a precise correlation structure can be inferred.

We now apply on IBM's quantum hardware our variational scheme for inferring network topology using local qubit measurements [378]. As an example we consider the 5-qubit state preparation $|\psi\rangle = |W\rangle \otimes |\Phi\rangle$ where the $|W\rangle$ is the W state $|W\rangle = \frac{1}{\sqrt{3}}(|001\rangle + |010\rangle + |100\rangle)$ and $|\Phi^+\rangle = \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle)$ is a two qubit maximally entangled state. We also apply our variational network inference scheme to 5-qubit GHZ states and the 5-qubit zero state (see our work in Ref. [378]). In each example, a known state is prepared while we optimize a variational ansatz that parameterizes arbitrary qubit projective measurements as $\Pi_x = U^\dagger(\vec{\theta})\,|x\rangle\langle x|\,U(\vec{\theta})$ where $\vec{\theta} \in \mathbb{R}^3$. We then compare the performance of our variational inference scheme on both noisy IBM hardware and noiseless classical simulator when 10, 100, 1,000, and 10,000 shots are considered.

To quantify the performance of the optimization, we consider the inference error, which we define as the

Euclidean distance

$$d(C, C^\star) = \sqrt{\mathrm{Tr}\left[(C^\star - C)^\intercal (C^\star - C)\right]} \tag{18.19}$$

where $C$ is the optimized covariance matrix and $C^\star$ is the ideal covariance matrix for the given state. The distance in Eq. (18.19) can similarly quantify the error in the characteristic matrix $d(Q, Q^\star)$. Note that this inference error quantifier only works in our numerical experiment because we prepare a known state. In practice, the state preparation is not known and the performance of the optimization cannot be quantified by Eq. (18.19).

For the considered state $|\psi\rangle = |W\rangle \otimes |\Phi\rangle$, the ideal qubit covariance and characteristic matrices are

$$C^\star = \begin{pmatrix} 1 & \frac{2}{3} & \frac{2}{3} & 0 & 0 \\ \frac{2}{3} & 1 & \frac{2}{3} & 0 & 0 \\ \frac{2}{3} & \frac{2}{3} & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 & 1 \end{pmatrix} \quad \text{and} \quad Q^\star = \begin{pmatrix} S_W & I_W & I_W & 0 & 0 \\ I_W & S_W & I_W & 0 & 0 \\ I_W & I_W & S_W & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 & 1 \end{pmatrix} \tag{18.20}$$

where $I_W \equiv \mathcal{I}_m(q_i; q_j) \approx 0.349976$ and $S_W \equiv S(W^{q_i}) =\approx 0.918296$ for any of the qubits $q_i$ and $q_j$ of the state $|W\rangle$. To obtain the ideal covariance matrix and mutual information in Eq. (18.20), it is sufficient to measure all qubits in the $\{|+\rangle, |-\rangle\}$ basis. While this calculation is straightforward for the $|\Phi\rangle$ state, we will be more explicit with the $W^{ABC} = |W\rangle\langle W|$ state, for which the reduced density matrices are

$$W^A = \mathrm{Tr}_{BC}\left[W^{ABC}\right] = \frac{2}{3}|0\rangle\langle 0| + \frac{1}{3}|1\rangle\langle 1| \tag{18.21}$$

and

$$W^{AB} = \mathrm{Tr}_C\left[W^{ABC}\right] = \frac{1}{3}|00\rangle\, 00 + \frac{2}{3}\left|\Psi^+\right\rangle\left\langle\Psi^+\right| \tag{18.22}$$

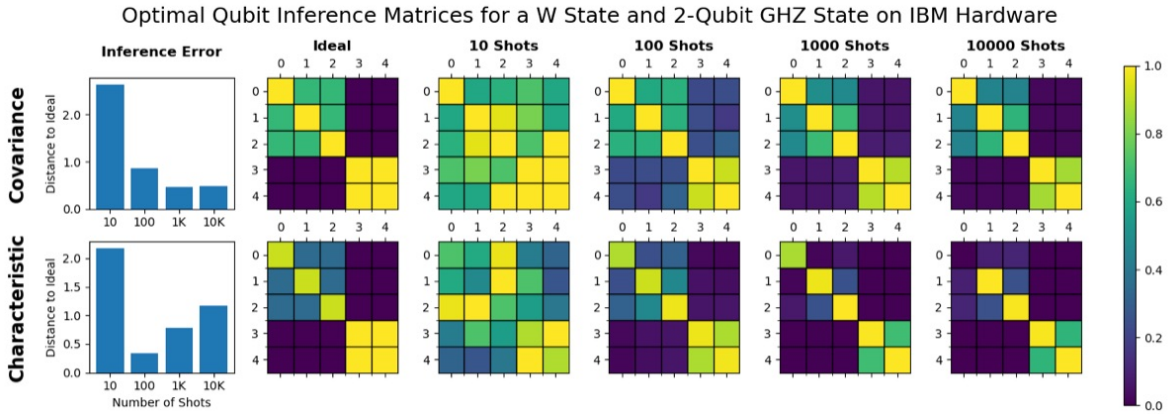



Figure 18.10: In the top row, we plot the maximal value achieved for each term of the covariance matrix across all optimization steps. In the bottom row, we plot the maximal mutual informations (off-diagonals) and minimal Shannon entropies (diagonal) across all optimization steps. The bar graph plots the Euclidean distance between the ideal and inferred matrices for each distinct number of shots.

where $|\Psi^+\rangle = (|01\rangle + |10\rangle)/\sqrt{2}$. When the observable $\sigma_x$ is measured, the covariance and variance are

$$\text{Cov}(A, B) = \text{Tr}\left[(\sigma_x \otimes \sigma_x)\, W^{AB}\right] = \frac{2}{3} \tag{18.23}$$

and $\text{Var}(A) = 1$. The qubit von Neumann entropy can then be explicitly calculated from Eq. (18.21) as

$$S_W \equiv S(W^A) = -\frac{2}{3}\log\frac{2}{3} - \frac{1}{3}\log\frac{1}{3} \approx 0.918296, \tag{18.24}$$

while the measured mutual information is found to be $I_W = \mathcal{I}_m(A, B) =\approx 0.349976$ where each party measures the observable $\sigma_x$.

Since all qubits are measured in the same basis, the measured mutual information is equivalent to the classical mutual information for all qubit pairs, that is $\mathcal{I}_m(q_i, q_j) = I(q_i, q_j)$. Thus, for this example, it is sufficient to optimize the classical bipartite mutual information rather than the measured mutual information. This simplification provides significant speed ups when running on quantum hardware because only one circuit needs to be evaluated to collect the mutual information of all qubit pairs as opposed to the 10 circuits needed to optimize the measured mutual information for each qubit pair independently. Note that queue times for the IBM hardware are the main bottleneck in our variational network inference scheme.

In Eq. (18.9), we plot the inference error as the covariance and characteristic matrices are optimized with respect to the state $|\psi\rangle = |W\rangle \otimes |\Phi\rangle$. To investigate the relation between the number of shots and the inference error, we consider optimizations where 10, 100, 1,000, and 10,000 shots are used when collecting data from the quantum computer. We expect that the inference error should decrease as the number of shots increases.

As a baseline, we first run our numerical experiment on `PennyLane`'s `default.qubit` classical simulator, which is a noiseless, finite-shot simulation of a quantum computer. The data shown in the left column of Fig. 18.9 plots the mean inference error from 10 independent optimizations for both the covariance and characteristic matrices. As expected, the amount inference error decreases as the number of shots increases. Furthermore, we find that the covariance and characteristic matrices do not always find a global optimum in which all matrix terms converge to their maximal theoretical values. As a result, the mean does not approach zero, reflecting the importance of running the optimization algorithm multiple times. Moreover, we find that the covariance matrix is optimized in fewer iterations than the characteristic matrix.

We then test our optimization on the IBM's `ibmq_belem` 5-qubit quantum computer, which exhibits a considerable amount of noise. When we run the optimization on an IBM quantum computer, the optimization steps become expensive to run due to the queue wait times. As a result, we are only able to run one optimization for each number of shots. We plot the optimization results in the middle column of Fig. 18.9. For the 10-shot case, it is not clear that the the inference error is decreasing. For larger numbers of shots, the inference error decreases throughout the optimization. To further validate the optimization results, we take the settings optimized on the noisy IBM hardware and reevaluate them on a noiseless, infinite-shot classical simulator. Overall, there is a general improvement in the inference error as shown in the rightmost column of Fig. 18.9 when compared with the data collected from the noisy IBM hardware. Although the data from the quantum hardware may seem noisy, the optimization over qubit measurements is indeed decreasing the inference error.

In Fig. 18.10 we show the optimal covariance and characteristic where the optimal matrix values are taken over all optimization steps. As the number of shots increases, so does the optimization's ability to resolve the correlation structure with greater accuracy. On the other hand, when the number of shots is small, statistical

fluctuations can lead to stronger correlations than present, leading to false-positive correlations. For example, in the 10-shot case of Fig. 18.10 the zero terms of the covariance and characteristic matrices are optimized to be quite large, which would lead a researcher to infer that the two qubits are correlated. Also note that the IBM hardware can have significant errors on a given qubit for example, in the 10,000 shot characteristic matrix the top left matrix term remains close to zero despite taking more shots than other more successful trials. From experiments on the simulator, we know that mutual information generally is more difficult to optimize than covariance. Moreover, a noisy device might not be stable for the amount of time needed to acquire 10000 shots. This instability leads to noise that inhibits the optimization's ability to make progress toward an optimum, particularly so for computing the mutual information.

As a final remark, we note that our variational quantum network topology inference scheme provides several improvements to existing techniques. First, previous methods do not typically consider the measurements to be tunable. By optimizing measurements, our variational methods could improve the effectiveness of existing topology inference approaches in which the violation of entropic bounds or Bell inequalities are used to test for network topology [124], [272], [273], [276], [387]–[391], [395], [396]. Second, quantities such as the von Neumann entropy or measured mutual information cannot be obtained for unknown quantum states. However, our variational quantum optimization methods are hardware compatible and can be used to approximate these quantities through optimization. Thus, entropy-based inference methods can be applied in practice and . Finally, our methods are hardware agnostic. The parameter-shift rule requires only a parameterized description of the unitary applied by the measurement device, how the unitary is physically implemented is not relevant.

### *18.3.5 Establishing Device-Independent Key Distribution Protocols in Quantum Networks*

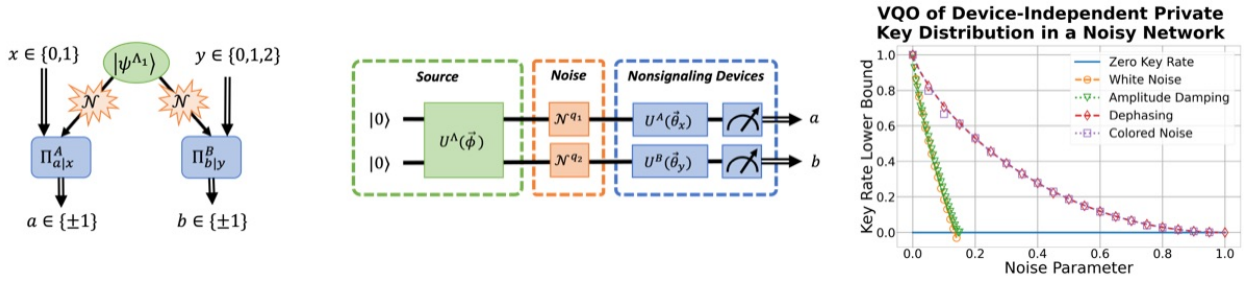



Figure 18.11: (Left) A DAG depicting the noisy DIQKD network. (Center) The variational ansatz circuit that is optimized. (Right) Maximization of the Devetak-Winter key rate (Eq. (18.26)) in the presence of noise using VQO and the cost function expressed in Eq. (18.27). The $y$-axis plots the key rate while $x$-axis plots the noise parameter for each qubit or source noise model. Each color and marker corresponds to a different noise model. The line plots the theoretcal maximal key rate derived from the maximal noisy CHSH violation in Section **??** while the markers plot the maximum obtain using VQO.

In this section, we share numerical results in which variational quantum optimization is used to establish a device-independent quantum key distribution (DIQKD) protocols in a noisy network setting. In the standard device-independent key distribution setting, device $A$ is given an input $x \in \mathcal{X} = \{0, 1\}$ and device $B$ is given an input $y \in \mathcal{Y} = \{0, 1, 2\}$ where each device outputs the binary value $a \in \mathcal{A} = \mathbb{B}$ and $b \in \mathcal{B} = \mathbb{B}$ respectively. In this way the CHSH scenario is extended by adding a third input to device $B$. The user then draws the inputs $x$ and $y$ from a uniform distribution and passes each input to its respective measurement device, then the user records the output from each device. After the data is collected, the user tests that the inputs

$x, y \in \{0, 1\}$ violate the CHSH inequality $S_{\text{CHSH}}$, while the the input $y = 2$ on device $B$ is used to distribute the key, which is shared only when $x = 0$ and $y = 2$. The probability of error is defined by the bit error rate

$$Q = P(a \neq b | x = 0, y = 2). \tag{18.25}$$

The key rate is lower bounded by the Devetak-Winter rate [250], [251],

$$r \geq r_{DW} = 1 - h(Q) - h\left(\frac{1 + \sqrt{(S_{\text{CHSH}}/2)^2 - 1}}{2}\right), \tag{18.26}$$

where $h(x) = -x \log_2(x) - (1 - x) \log_2(1 - x)$ is the binary entropy. Thus, the key rate is dependent upon only the CHSH violation $S_{\text{CHSH}}$ and the bit error rate $Q$. Then, our optimization objective is to maximize the key rate which requires $Q$ to be small while $S_{\text{CHSH}}$ is large. Therefore, we minimize the cost function

$$\text{Cost}(\Theta) = -S_{\text{CHSH}}(\Theta) + Q(\Theta), \tag{18.27}$$

where $\Theta$ is the network settings. This cost function is used rather than Eq. (18.26) because if $S_{\text{CHSH}} < 2$, then the square root in Eq. (18.26) is taken of a negative number, which breaks the automatic differentiation in our VQO framework.

In Fig. 18.11, we show that our VQO methods can optimize a device on a classical simulator with a static noise model. The theoretical key rate on the source depolarizing noise model is taken from previous works [250], [251] where $S_{CHSH} = 2\sqrt{2}(1 - 2Q)$ where $Q = (\frac{8\gamma}{15})$. For the remaining noise models, we apply directly the result of Section **??** to derive the theoretical scores. For the source colored noise, we find $S_{CHSH} = 2\sqrt{1 + (1 - \gamma)^2}$ and $Q = 0$ because colored noise preserves a classical correlation between the devices. Likewise for the qubit dephasing channel we find $Q = 0$ and the max CHSH violation to be $S_{CHSH} = 2\sqrt{1 + (1 - \gamma)^2}$. For the uniform qubit amplitude damping channel we find that $S_{CHSH} = 2\sqrt{2(1 - \gamma)^2}$ and $Q = (1 - \gamma + \gamma^2 - \sqrt{1 - 2\gamma + 2\gamma^2})/2 + (1 - \gamma)\gamma/2$ which is the sum of the two smallest eigenvalues of $\mathcal{A}_\gamma \otimes \mathcal{A}_\gamma(|\Phi^+\rangle\langle\Phi^+|)$.

## *18.4 Variational Quantum Networking on Quantum Hardware*

Up to this point, all variational quantum optimization results have been obtained using a classical simulator. As a proof of principle we now demonstrate that variational quantum optimization can be implemented on quantum computing and networking hardware. In doing so, we show explicitly how variational quantum networking can be applied in practice. As supporting evidence, we provide three example including violations of $n$-locality in star and chain networks, inference of network topology in the $n$-local setting, and inversion of a noisy fiber transformation. The first two examples are obtained from the open-access IBM quantum computers while the third example is obtained from an experimental quantum optics apparatus. Finally, we discuss the practical challenges and advantages of applying variational quantum optimization on quantum hardware.

In all cases, the quantum hardware is trusted and its resources are characterized. This means that we can skip the step in variational quantum networking where the network hardware is queried for its topology and resource capabilities. Furthermore, in all cases, the classical optimizer serves as a controller that tunes the network to perform the desired task. In this way, we show how the variational quantum optimization portion

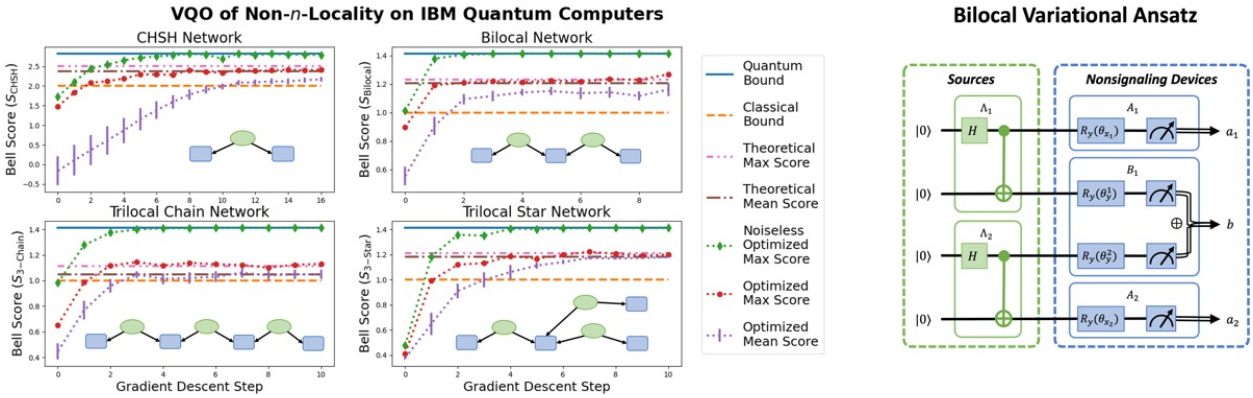



Figure 18.12: [102] (Left) **VQO of Non-$n$-Locality on IBM Quantum Computers.** Violations of the star and chain $n$-locality inequalities are maximized in quantum star and chain networks in the non-signaling $n$-local setting. The $x$-axis shows the gradient descent step in the optimization while the $y$-axis shows the Bell score. The noiseless quantum bound is shown by the solid blue line and the classical bound is shown by the dashed orange line. For each Bell inequality, we aggregate data across several different optimization runs. At the end of each optimization, the theoretically optimal score is evaluated on noisy hardware to serve as a baseline. The mean theoretical score is shown by the dash-dotted brown line and the max theoretical score is shown by the dash-dot-dotted pink line. In each step, the max score across all optimization is shown by the dotted red line with circle markers while the mean score across all optimizations is shown by the dotted purple line with error bars showing the standard error. The dotted green line with diamond markers shows the noiseless optimal score obtained by running the settings for the maximal score in each step on a noiseless classical simulator. (Right) **Bilocal Variational Ansatz:** A simple hardware-efficient variational ansatz circuit. When applying VQO on hardware, we consider a simple ansatz. Each source (green) prepares the Bell state $|\Phi^+\rangle$ using a Hadamard and CNOT gate. Each nonsignaling device (blue) applies a local rotation about the $y$-axis to each qubit before measurement in the computational basis. When a measurement device contains more than one local qubit, the XOR is taken to convert the bit string into a binary output.

of the variational quantum networking scheme plays out when implemented on hardware.

### *18.4.1 Maximizing Violations of Chain and Star n-Locality Inequalities on an IBM Quantum Computer*

In this section, we demonstrate that our VQO techniques can maximize violations of $n$-locality in star an chain quantum networks. As discussed in Chapter 13, these $n$-locality violations are important for self-testing quantum network resources and implementing device-independent protocols in quantum star and chain networks. Hence maximizing this form of nonclassicality on quantum network hardware can be important at a low-level for testing measurement and state preparation devices, as well as at a high-level where device-independent protocols can be established between devices in a quantum network. Note that the results in this section are adapted from our work in Reference [102].

Using IBM quantum computers, we implement the variational quantum optimization algorithm in Algorithm 1. The cost functions for $n$-local star and chain networks are then expressed as

$$\min_{\Pi^{\text{Net}} \in \text{POVM}(\mathcal{H}^{\text{Net}})} -S_{n\text{-Star}\{\mathbf{G}^{\text{CHSH}_i}\}_i}(|\psi\rangle\langle\psi|^{\text{Net}}) \quad \text{and} \quad \min_{\Pi^{\text{Net}} \in \text{POVM}(\mathcal{H}^{\text{Net}})} -S_{n\text{-Chain}\{\mathbf{G}^{\text{CHSH}}\}}(|\psi\rangle\langle\psi|^{\text{Net}}) \tag{18.28}$$

where $|\psi\rangle^{\text{Net}} = \bigotimes_{i=1}^{n} |\Phi^+\rangle\langle\Phi^+|^{\Lambda_i}$ and $\Pi^{\text{Net}} = \bigotimes_{j=1}^{n} \Pi_{x_j}^{A_j} \otimes \Pi_y^{B_j}$ where local measurements are assumed to be separable across qubits. For the cases of $n \in \{1, 2, 3\}$, we optimize the measurements on IBM hardware

such that the cost functions in Eq. (18.28) are minimized. In each optimization we use the variational ansatz depicted in Fig. 18.12 in which a Bell state $|\Phi^+\rangle = \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle)$ is prepared at each source and a local qubit measurements are free to rotate in the $xz$-plane of the Bloch sphere. Although this ansatz does not express the complete set of measurements or state preparations, it corresponds to physically relevant experimental setups in which entangled photon polarization encoded qubits are used [193]–[197].

We plot our VQO results in Fig. 18.12. In our optimizations, each circuit is evaluated using 6000 shot and initialized with settings drawn randomly from a uniform distribution. The CHSH case was optimized using the 5-qubit `ibmq_belem` device while the bilocal, trilocal chain, and trilocal star networks were optimized using the 7-qubit `ibmq_casablanca` and/or `ibmq_jakarta` devices. The CHSH plot shown in the upper-left of Fig. 18.12 aggregates data from 11 separate optimizations using a step size of $\eta = 0.12$. The bilocal plot shown in the upper-right of Fig. 18.12 aggregates data from 5 optimizations and using a step size ranging from $\eta = 1.4$ to $\eta = 1.5$. The trilocal chain plot shown in the bottom-left of Fig. 18.12 aggregates data from 6 optimization using a step size ranging from $\eta = 1.6$ to $\eta = 2$. The trilocal star plot shown in the bottom-right of Fig. 18.12 aggregates data from 5 optimizations using a step sizes ranging from $\eta = 1.6$ to $\eta = 1.8$.

As a baseline measure of noise on the IBM quantum computers, we calculate the theoretical score using the optimal settings in the noiseless case. That is, we use the settings that would naively be selected by a researcher for an ideal quantum network. Since the quantum computing hardware is noisy, the non-$n$-local correlations deteriorate causing a separation between the noiseless quantum bound (Blue) and the noisy mean theoretical bound (Brown). In particular, the bottom-left plot shows the theoretical quantum violation of the trilocal chain inequality to be close to the classical bound, especially when compared with the trilocal star network (bottom-right). This difference is due to the fact that the $S_{n\text{-Star}\{\mathbf{G}^{\text{CHSH}_i}\}_i}$ is more robust to noise than $S_{n\text{-Chain}\{\mathbf{G}^{\text{CHSH}}\}}$ [270], [271]. Finally, the noise on the quantum computers is not constant and fluctuates throughout the day [398], hence, the optimal settings in the noiseless case do not always produce the same expected value.

All plots in Fig. 18.12 show the mean optimized score exceeding the classical bound. Thus, VQO finds settings non-$n$-local correlations in all studied cases. In most cases, the error bars shrink as the optimization step increases indicating a consistent convergence to an optimum. In the bilocal network optimization, the significant error bars on the final step are likely a result of the step size being too large caused by the cost to overshoot the minimum. In the trilocal chain and star optimizations, the mean optimization score converges to the mean theoretical score showing that the optimization consistently finds the theoretical maximum. In the bilocal and CHSH optimizations, the mean optimization score does not reach the mean theoretical score. This is the result of some optimizations finding the local optimum of the classical bound.

In all plots, the max optimized score converges to a value consistent with the max theoretical score. In all cases except for the CHSH optimization, an optimized score is found that exceed the max theoretical score. While there is a statistical chance that that the optimized score may be larger than the max theoretical score, it is also possible that our VQO framework may be finding optimal settings tailored for the quantum hardware. That is, there may be biases in the gate operations and measurement bases that VQO may be optimizing against. On the other hand, the theoretical settings are naïve to any such hardware biases.

Finally, we find that the settings optimized on the IBM hardware correspond to the optimal settings in the noiseless case. In reference to when we take the settings for the maximal score in each optimization step and run them on a noiseless classical simulator, they maximally violate the Bell inequality in question. While this feature is remarkable, its value is questionable. We may be able to obtain the optimal settings on a large NISQ device, but the quantum computer will still output a nonoptimal answer that cannot be checked on a

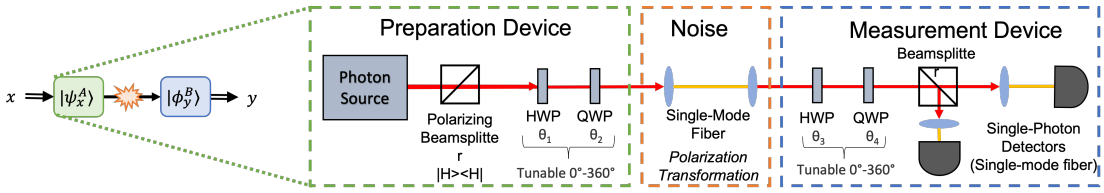



Figure 18.13: A simple prepare-and-measure network is implemented in a polarization encoded photonic qubit system. The photon source output horizontally polarized photons whose polarization can be tuned via half-waveplate (HWP) and quarter-waverplates (QWP) on rotational stages. An unknown unitary transformation is applied by a fiber optic cable. The measurement device uses a HWP and QWP to decode the photon polarization encoding into the computational basis implemented as single photon detectors observering horizontally and vertically polarized light.

classical computer due to computing constraints. Further study of this optimization feature is important as there may exist an application where the optimal settings are valuable to know on their own.

The two main challenges encountered in this experiment include inefficiencies of queue wait times, the network latency of remote execution, and the noise on IBM's quantum hardware. In total, we find that our experiments take dramatically longer to perform on IBM hardware than on a laptop computer, even when IBM's hardware is reserved. This poor performance of VQO reflects the fact that existing quantum computers are far from providing any advantage, especially on small system sizes that can be efficiently computed. Furthermore, we note that the star $n$-locality inequality in Eq. (13.14) requires $|\mathcal{Y}|\prod_{i=1}^{n}|\mathcal{X}_i|$ correlator terms to be evaluated, each having the form $\langle O_{x_1}^{A_1}\dots O_{x_n}^{A_n}O_y^B\rangle_{\rho^{\tilde{\Lambda}}}$. Thus, we observe that the star $n$-locality requires a number of circuits to be evaluated that scales as $O(|\mathcal{X}|^{n+1})$ where we take $|\mathcal{X}| = |\mathcal{X}_i| = |\mathcal{Y}|$. Hence these methods cannot be scaled to large networks without significant parallelization.

### *18.4.2 Variational Quantum Optimization of Experimental Quantum Optics*

As an important demonstration, we show how our variational quantum optimization methods can be deployed on quantum networking hardware. Namely, we implement our methods on a single qubit using polarization encoded qubits. The simple task that we demonstrate is inverting the noise model on a fiber optics cable (see Fig. 18.13).

As a proof of concept, we demonstrate fiber inversion by inputting horizontally polarized light into the system and extinguishing the photon count on the vertically polarized single photon detector. To simplify the experiment, we use only the HWP and QWP rotational stages in the preparation device. These components allow for any qubit unitary to applied to the input photons. The goal is that $QWP(\theta_2)HWP(\theta_1) = U_{Fiber}^{\dagger}$ such that we minimize

$$P(V|H) = \left|\left\langle V|U_{Fiber}^{\dagger}QWP(\theta_2)HWP(\theta_1)|H\right\rangle\right|^2. \tag{18.29}$$

In Fig. 18.14, we show that the optimization successfully tunes the waveplates such that the fiber noise is inverted. This is seen by the convergence to a minimum in the cost landscape. We find that the optimization does not converge to the minimum in practice, falling short of the expected minimum of detecting only background photon counts. We suspect that this deviation is due to imperfect optical components within the experimental apparatus.

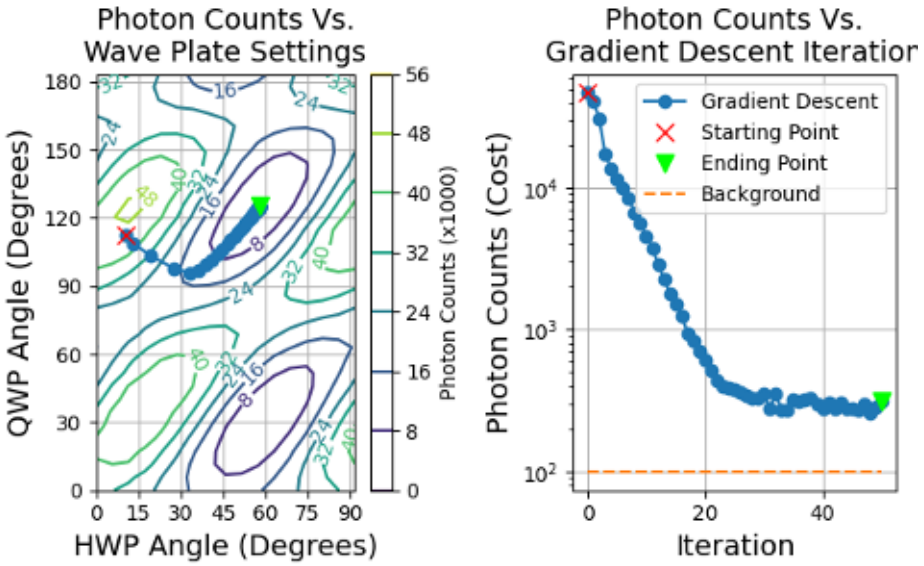



Figure 18.14: (Left) A contour plot of the cost landscape is constructed by measuring $P(V|H)$ on the experimental apparatus for fixed $\theta_1$ and $\theta_2$ as in Eq. (18.29). The blue line plots the trajectory of the optimization as it travels from a peak to the minimum of the cost landscape. (Right) The performance of the optimization is plotted on a log scale with respect to the expected background counts.

## *18.5 Scaling Variational Quantum Networking Applications*

Variational quantum networking may promise interesting technological advantages, however, can these techniques be scaled to large quantum networks? In this section, we investigate the potential bottle-necks limiting variational quantum optimization in quantum computing and networking systems. Since variational quantum networking is an extension of variational quantum optimization, it inherits the scaling challenges and bottlenecks discussed in Section 11.3.2 and Section 11.3.3. The main challenges being large numbers of circuits to evaluate on the network, barren plateaus [309]–[312], and the general NP-hardness of the classical optimizer [313]. The latter two challenges regard the difficulty of solving the optimization problem. As new techniques are developed that address these problems in the field of variational quantum algorithms, these techniques can be applied directly to the setting of variational quantum networking. Thus, we focus primarily on the challenge of scaling the number of circuit evaluations.

In a quantum network, the number of circuits needed to evaluate the cost function scales as $O(|\mathcal{X}|)$ while the number of circuits needed to evaluate a gradient using the parameter-shift rule scales as $O(|\mathcal{X}| \times |\Theta|)$ where $|\mathcal{X}|$ is the number of classical inputs to the network, and $|\Theta|$ is the number of network parameters. Furthermore, for $n$ devices having classical inputs $|\mathcal{X}| = \prod_{i=1}^{n} |\mathcal{X}_i|$ where $|\mathcal{X}_i|$ is the number of inputs for the $i^{th}$ device, and in the worst case, $|\Theta|$ $O(\text{Dim}\left(\mathcal{H}^{\text{Dev}_i}\right))$ where $\text{Dim}\left(\mathcal{H}^{\text{Dev}_i}\right)$ scales exponentially with the number of qubits in the $i^{th}$ device. Thus, the number of circuits to evaluate in a generic variational quantum networking protocol can grow exponentially in both the number of devices and the number of qubits used by each device.

While in Section 11.3.2 we argue that broad parallelization across quantum computers can be used to offset the number of circuit evaluations, the same cannot be said for quantum networks because there typically won't be many copies of a network across which circuit evaluations can be parallelized, that is, we must

collect all of the network data in serial. Therefore, there are two immediate strategies that can help scale variational quantum optimization techniques for quantum networks: decreasing the number of circuits to run and decreasing the cost of running each circuit.

The number of circuits can be decreased by decreasing the number of relevant network inputs $|\mathcal{X}|$, and decreasing the number of parameters $|\Theta|$ in the variational ansatz. Since the number of network inputs grows exponentially with the number of devices, decreasing the number of inputs of a single device by one can subtract a significant number of circuit evaluations from the total cost. Moreover, if fewer inputs are considered for each device not only does $|\mathcal{X}|$ decrease, but so does $|\Theta|$. Recall that the unitary applied by the $i^{th}$ network device in the $j^{th}$ network layer is $U^{\mathrm{Dev}_{i,j}}(\Theta_{x_{i,j}})$, where in total the device has $|\Theta_{i,j}| = |\mathcal{X}_{i,j}| \times |\Theta_{x_{i,j}}|$ parameters associated with it. Thus by removing one input, from one device, the number of settings decreases by $|\Theta_{x_{i,j}}|$. Likewise, by considering a more restrictive ansatz, the number of settings considered can be diminished significantly. For example, rather than implementing arbitrary $N$ qubit unitary operators that take $4^N - 1$ parameters, an ansatz that scales polynomially with $N$ could instead be considered.

A complementary approach that can also improve the performance of variational quantum networking is to decrease the cost of evaluating a quantum circuit on quantum networking hardware. To this end, there are only a few important parameters to consider, the number of shots needed for the desired performance, the rate at which shots can be sampled, and the time it takes to update the parameters across the network such that a new circuit can be run. It naturally follows that the time it takes to evaluate a circuit is exactly $t = N_{\mathrm{Shot}}/R_{\mathrm{Shot}}$ where $N_{\mathrm{Shot}}$ is the number of shots and $R_{\mathrm{Shot}}$ is the shot rate or number of shots per second. Our results show by example, that as few as 10 shots per circuit evaluation can be used to evaluate gradients and traverse towards the optimum, however, the large stochastic fluctuations about the mean causes the precision of the optimization to decrease. Increasing the number of shots then improves the numerical precision of the optimization. However, the number of shots that can reasonably collected is limited by the shot rate $R_{\mathrm{Shot}}$, which is a detail specific to a particular quantum network implementation.

Furthermore, running a quantum circuit has an associated overhead of adjusting the network parameters such that the appropriate operations are applied in the network for a given classical input. As an example, in our variational quantum optics experiment, the half-wave plates and quarter-wave plates used to encode quantum state preparations and measurements are placed on rotational stages that rotate at a limited rate. Since the parameter-shift rule requires a macroscopic shift, the time it takes to rotate each wave plate from one orientation to the next is considerable and serves as the bottleneck in our experimental apparatus. In this example, there is a high cost to running circuits with different settings. Hence it is extremely important to decrease the time it takes to reorient the network for a given set of settings $\Theta$.

Overall, there are significant challenges that may limit the scale at which variational quantum networking techniques can be applied. However, it is important to note that there is an incredible amount of utility in using these methods to solve smaller problems. The main advantage being that the network can establish and maintain protocols on quantum networks that have uncharacterized noise that evolves over time.

# Part IV

# Discussion and Conclusion

# Chapter 19

# VQO: a Tool for Discovery, Design, and Automation

Variational quantum optimization (VQO) is a key method supporting this work. We develop the qNetVO Python package [108] to simulate and optimize noisy quantum communication networks in flexible, hardware-agnostic manner. Throughout this work, we demonstrate our VQO framework's ability to to find optimal solutions in the presence of noise, operate on either quantum hardware or classical simulator, and even find optima at low shot limits. We therefore provide a comprehensive optimization framework that allows for meaningful scientific investigated on the classical simulator, while also being extensible to quantum computing and networking hardware as these quantum technologies improve. On quantum computers our methods show promise of having design advantages in simulation and optimization while on quantum networks, our methods show promise of advantages beyond classical approaches in characterization and automation.

## *19.1 The Flexibility and Hardware-Agnosticism of VQO*

One advantage of our VQO methods are that they are generic and hardware-agnostic. Although we mainly optimize nonclassicality, VQO techniques can be applied to a broad domain of optimization problems. Furthermore, these hybrid computing methods are compatible with any quantum system that is tunable over a set of optimization parameters.It is important that hybrid quantum-classical algorithms be hardware-agnostic because a wide range of quantum computing platforms are already available and will continue to be developed.

In Chapter 11 we provide a generic description of VQO, in which the goal is to minimize a cost function $\text{Cost}\left(\mathbf{P}(\Theta)\right)$. The cost function relies only on the behavior of the network $\mathbf{P}(\Theta)$ obtained for the settings $\Theta$ and is therefore agnostic to the physical hardware. This means that the underlying system producing the behavior can be treated as a black-box and any black-box optimization procedure technique could be applied. In this work, we consider a standard gradient descent algorithm. On classical hardware we apply backpropagation to evaluate the gradient while on quantum hardware, we apply the parameter-shift rule.

## 19.2 Variational Quantum Optimization as a Tool for Discovery and Design

Simulation and optimization are important tools in science and engineering. As noisy quantum networks are developed, our VQO methods show promise of being able to scale beyond classical approaches for simulating and optimizing quantum systems.

Currently, our VQO methods are best run on classical hardware due to the inherent noise and cost of using quantum hardware. The majority of our results were obtained using PennyLane's `default.qubit` classical simulator [210] to simulate the quantum network system. These simulated results were obtained on a laptop computer, and in most cases, the entire optimization can run within a few minutes. Even when applied on a classical simulator running on a laptop computer, VQO has demonstrated its value as a convenient tool for gaining theoretical insights and designing protocols in quantum networks.

Indeed we apply our methods in the qNetVO software and demonstrate its application in optimizing noisy quantum networks. In Chapter 12 and Chapter 13 we demonstrate that our VQO framework can reproduce known maximal violations in both noiseless and noisy cases. In Chapter 14 we show that VQO can be used to certify the signaling dimension of a quantum channel, moreover, our VQO methods find novel entanglement-assisted classical communication strategies that demonstrate an increase to the signaling dimension of a signaling system. In Chapter 15, Chapter 16, and Chapter 17 we demonstrate our software's ability to optimize nonclassicality in a wide range of quantum resource configurations. Although the upper bound on the quantum violations may not be known, our software is able to find examples of nonclassicality demonstrating operational advantage.

Although our results were obtained on a laptop computer, scaling to larger systems will require improvements. One way to scale is to use a more efficient classical simulator. For example, making use of parallel processing or using more powerful computational resources could help scale our methods to larger networks with more qubits and parameters. However, the fundamental scaling challenges of quantum systems will likely to persist for large quantum networks with complex entanglement structures. In principle, such systems can be simulated more efficiently on a quantum computer. Therefore, as larger and less noisy quantum computers become available, it might be possible for our hybrid simulation and optimization techniques to be scaled beyond classical computing approaches. By being hardware-agnostic, our VQO methods can make use of the classical simulators now, and when quantum computers become available, our approaches can be extended seamlessly to any quantum computing platform.

One drawback of using quantum computers to simulate or optimize a quantum network is that there might not be a one-to-one correspondence between the noise and operations in the quantum network versus the quantum computer. Thus, a protocol designed on a quantum computer may not perform as well on a quantum network because the noise model might be different. To mitigate this issue, it would be important to characterize the network's noise so that it can be accurately simulated.

## 19.3 A Path Towards Quantum Network Automation

VQO can also serve as a means to automate quantum network protocols without needing to characterize a network's underlying noise. In Chapter 13, we find for $n$-local nonsignaling networks that our VQO methods are able to find the maximally nonclassical behaviors even when noise is present. That is, if a networking application relies upon a quantum violation of $n$-locality, our hybrid VQO algorithm can automate the network devices to establish those maximally nonclassical behaviors without needing to characterize the noise.

We develop this approach to network automation in Chapter 18 where we introduce variational quantum networking as a design archetype that provides a declarative, hardware-agnostic protocol for quantum network automation. We argue that automation will be necessary to establish and maintain protocols between devices in a noisy network environment. To support our argument we show examples of both establishing quantum network protocols and applying our methods on quantum hardware.

The main advantage of this approach is that the underlying noise in the quantum network does not need to be characterized, which can be a difficult task. Furthermore, variational quantum networking is a hardware-agnostic protocol that relies only upon the classical data produced by the devices, and the causal structure encoded by the network's DAG. In principle, variational quantum networking could be applied to quantum networks on any hardware platform, and even hybrid networks integrating multiple quantum hardware architectures.

## *19.4 The Scaling Challenges of VQO*

Although VQO show promise of advantage on both quantum computing and networking hardware, practical challenges will emerge when scaling this hybrid computing approach. As discussed in Chapter 11 and Chapter 18, increasing the number of parameters or the size of the domain leads to a polynomial increase in the number of circuits that must be run to evaluate the gradient. When run on a quantum computer, gradient evaluation can be parallelized broadly. However, if VQO is running on a quantum network to automate its protocols, then parallelization won't likely be available. Separately, variational quantum optimization applied to quantum networks will likely suffer the same issues with barren plateaus as observed generally for variational quantum algorithms.

## *19.5 The Nonoptimality of VQO*

It is important to acknowledge that our VQO framework does not obtain the global optimum with certainty. The gradient descent algorithm is neither guaranteed to converge within a given number of iterations, or find a global optimum. Despite the possibility that VQO does not find a global optimum, we argue from a practical standpoint that the minimized cost quantifies the performance of the optimized system. Hence it serves a certificate of performance that a user can decide to either accept or reject.

Consider the VQO algorithm as a black-box. The input to the black-box encodes a cost function and a variational ansatz. As output the black-box provides the optimal settings $\Theta^\star$ and the minimized cost $\text{Cost}\,(\Theta^\star)$. From a practical perspective, the minimized cost $\text{Cost}\,(\Theta^\star)$ serves as a certificate quantifying the performance of the example solution. The quality of the optimization result is then explicit, if $\text{Cost}\,(\Theta^\star)$ can be tolerated then the black-box can be applied.

Unfortunately, it cannot be determined from the optimization results whether or not a better solution exists. If the cost is not sufficiently minimized, then the algorithm can be repeated many times with randomized initial conditions in hopes of stumbling upon a better solution. In theoretical investigations, it would be useful to know whether the algorithm has obtain the global minimum of the cost. Therefore it is important to derive a bound on the minimal cost. In doing so, the distance to the optimal solution then encodes the quality of the optimization.

Although the results of variational quantum optimization may not be optimal, the minimized cost provides a certificate of its quality characterizing the black-boxes performance. It therefore falls upon the researcher

applying the algorithm to assert whether or not the demonstrated performance is sufficient for the application.

## *19.6 The Outlook of VQO*

Future work can build off our variational quantum optimization framework for quantum networks in a number of ways. Our simulation and optimization software can be extended to new applications beyond nonclassicality. Furthermore, our framework can also be tested and developed on to quantum networking software. It would be interesting to see how generic variational quantum networking protocols compare to hardware-specific approaches for establishing and maintaining protocols in quantum systems.

Given the current momentum behind the study of hybrid quantum-classical computing and variational quantum algorithms, their applications and implementation will continue to be iterated upon and developed. Furthermore, the quantum computing experience will likely offer more streamlined support to these hybrid algorithms. As these advancements are made in the field of variational quantum algorithms, these advancements can be applied to our VQO framework for quantum networks.

# Chapter 20

# Nonclassicality in Noisy Quantum Networks

We investigate nonclassicality over a range noisy quantum network scenarios. We explicitly model noisy quantum operations, as well as the resource constraints of noisy quantum networks. The networks we consider have few layers of processing, few devices, few qubits, and operate in the low shot regime.

Overall, we find examples of nonclassicality in all studied communication networks, using various entanglement and quantum communication resources. These examples demonstrate that quantum resources are able to provide broad operational advantages in communication limited scenarios. Furthermore, we find that these nonclassical behaviors are robust to some amount of noise, and can be demonstrated using minimal quantum resources. Hence we assert that nonclassicality can serve as a merit of resource quality and operational advantage in real-world quantum networks.

From a physical perspective, these nonclassical network behaviors, especially in multipoint networks, show novel ways that quantum physics defies the classical assumptions of locality, causality, and realism. Although, we expect that nonclassicality experiments in multipoint communications networks will reveal that the predictions of quantum theory are correct, these scenarios reveal new ways in which quantum mechanics deviates from classical. In the near-term quantum networks, nonclassicality can be used for testing and certifying quantum and LOCC resources. Furthermore, Nonclassicality will also be important for applications in information security and distributed computing.

## *20.1 Weak Nonclassicality in Networks*

Throughout this investigation, we define weak nonclassicality as being quantum behaviors that cannot be reproduced using a classical network having a similar DAG. However, when global shared randomness is held by all network devices, then a weakly nonclassical behavior can be simulated by the classical network. Any classical network behavior $\mathcal{C}^{\text{Net}}$ can be simulated using equivalent quantum communication resources $\mathcal{Q}^{\text{Net}}$, however, weak nonclassicality pertains specifically to the case $\mathcal{C}^{\text{Net}} \subseteq \mathcal{Q}^{\text{Net}} \subseteq \mathcal{C}_{\Lambda}^{\text{Net}}$.

Throughout this investigation, we study three key examples of weak nonclassicality. In Chapter 13, investigate quantum violations of $n$-locality in nonsignaling networks, in which the quantum violations of $n$-locality are weakly nonclassical. Then in Chapter 14 we find that weak nonclassicality exists in simple qubit

point-to-point signaling scenario. We prove in Theorem 13 that the separation between quantum signaling behaviors can require an unbounded amount of classical communication so simulate. Similarly, we show in Chapter 16 that weakly nonclassicality can be observed in the simplest broadcast setting in which an entangled state is measured by two separate receivers. We show in Theorem 15 that this example can be likened to the weak nonclassicality in point-to-point signaling scenarios. Thus, we find that an unbounded amount of classical communication can be required to simulate the behavior of measurements on bipartite entanglement.

In general, weak nonclassicality manifests in systems where local independence is assumed between devices, but not necessarily on the device's inputs. In essence, weak nonclassicality corresponds to the randomness intrinsic to quantum mechanics. Thus it is not surprising that quantum systems behave as if there is some amount of shared randomness present. Indeed, if an unlimited amount of global shared randomness is present, then the resource advantages granted by weak nonclassicality vanish. Therefore, the extra classical communication needed to simulate weakly nonclassical behaviors can be precisely understood as the randomness, that must be shared amongst devices.

Weak nonclassicality is interesting because it demonstrates a *static* resource advantage, meaning that additional shared randomness is needed to simulate weakly nonclassical behaviors. Unlimited shared randomness may not be a free resources in network systems, it could be difficult to coordinate or synchronize devices so that it could be used effectively, or perhaps, devices may not have had the chance to distribute or store randomness. In such cases, weak nonclassicality can demonstrate a significant resource advantage because an *unlimited* amount of classical communication may be needed to reproduce quantum behaviors.

The resource advantage of weak nonclassicality can be applied in self-testing schemes where devices can be trusted to not share randomness. In point-to-point signaling or broadcast scenarios, weak nonclassicality can be important for certifying the signaling dimension of a channel, or testing the functionality of state preparation and measurement devices. Furthermore, In the $n$-local nonsignaling setting, weak nonclassicality can be used to test the measurements and sources in the network. Furthermore, weakly nonclassical behaviors can produce particular random distributions with greater efficiency than classical systems. Beyond self-testing and generating random distributions, the applications of weak nonclassicality are limited because it does not allow information to be transmitted more efficiently, it allows only correlation to be transmitted more efficiently.

## *20.2 Strong Nonclassicality in Networks*

Throughout this investigation we introduce strong nonclassicality as being quantum network behaviors that cannot be simulated by a similar classical network having the same DAG, even when an unlimited amount of global shared randomness is held by the devices. More concisely, strong nonclassicality describes the case where $\mathcal{C}_{\Lambda}^{\text{Net}} \subseteq \text{Conv}\left(\mathcal{Q}^{\text{Net}}\right)$ where the convex hull is taken to ensure that the classical set is contained by the quantum. Strong nonclassicality corresponds to an interesting resource advantage because additional classical communication resources are required to simulate strongly nonclassical quantum network behaviors.

In this investigation, we find strong nonclassicality in all considered networks, but not all quantum resource configurations in those netoworks can lead to strong nonclassicality. In Chapter 12 we study the premier example of strong nonclassicality in which a collection of nonsignaling devices share entanglement. In Chapter 14 we find in point-to-point signaling systems, that strong nonclassicality can be observed in both entanglement-assisted classical communication and entanglement-assisted quantum communication.

The former is interesting because, although entanglement-assisted classical communication cannot boost the capacity of a noiseless quantum channel, it can produce nonclassical behaviors. The latter example is unsurprising, however, because entanglement-assisted quantum communication allows dense coding protocols to be implemented allowing the communication capacity to double. On the contrary, in Chapter 15 we investigate multiple access networks in which strong nonclassicality can be observed when quantum communication is unassisted by entanglement. This resource advantage is demonstrated explicitly by the random access coding task, however, our observed nonclassicality is far more general. Furthermore, entanglement-assisted senders in the multi-access networks can also lead to strongly nonclassical behaviors. In Chapter 16, we find no examples in broadcast networks where unassisted quantum communication can lead to strong nonclassicality. However, we do find cases in which entanglement-assisted receivers are able to generate strongly nonclassical behaviors. Finally, in Chapter 17, we study multipoint communication networks in which we find several examples of strong nonclassicality for quantum communication, as well as for entanglement-assisted senders, receivers, and communication.

It is interesting that in multi-access networks and interference networks that entanglement is not necessary for generating strong nonclassicality. As demonstrated in the random access coding task, even a single qubit can generate strong nonclassicality provided that incompatible measurements can be applied. Thus, what appears to be an indicator of when strong nonclassicality can manifest without entanglement is the presence of locally independent inputs. Note that all considered networks except broadcast and point-to-point networks have locally independent inputs. However, these examples indicate that there might be more to strong nonclassicality than local independence of inputs. That is, entanglement-assisted resource can be applied in these scenarios lacking local independence of inputs to produce strongly nonclassical behaviors.

When entanglement is used to assist classical communication resources, interesting examples of nonclassicality are found. For example, dense coding can boost the communication capacity of a network's channels increasing the amount of classical signaling leading to strong nonclassicality and operational advantage. Furthermore, in the random access code or the point-to-point signaling examples, if entanglement is shared between the sender and receiver, classical communication can be used to achieve strong nonclassicality. Similarly, when two classical senders share entanglement in the multiple access channel, the resulting quantum violations of the classical bounds are equal to or greater than the violations achieved when the senders instead use quantum communication. Moreover, in the broadcast network, entangled receivers are able to achieve strong nonclassicality, although, these broadcast nonclassicality may simply be a special case of the nonsignaling network scenario. Nevertheless, entanglement-assisted networks are able to exhibit strongly nonclassical behaviors beyond what quantum communication can achieve.

At this point, the foremost application of strong nonclassicality is device-independent certification and security protocols. Here quantum networking devices and resources are tested in their ability to violate the classical bounds for their given communication resources. In this work, we derive novel nonclassicality witnesses that serve as operational tests for quantum resource configurations. Notably, in Chapter 14 we apply the complete set of qubit signaling dimension bounds to test the nonclassicality of a fundamental LOCC resource, entanglement-assisted classical communication. Similarly, in multiple access networks we derive the finger printing inequality as a test for entanglement-assisted classical senders, and in the broadcast network we derive a classicality inequality that can test entanglement assisted receivers. Thus, we derive operational tests for certifying LOCC resources in networks and quantifying their operational advantage. In the future, strongly nonclassical correlations, may also lead to distributed information processing advantages, or device-independent networking protocols.

## *20.3 The Noise Robustness of Nonclassicality*

A main focus of this work is to investigate the noise robustness of nonclassicality across different network topologies. In doing so, the operational advantage of noisy quantum networks can be quantified. In general, we find that noise robustness depends upon the network's causal structure and considered resources. Furthermore, not all operational tests of nonclassicality demonstrate the same noise robustness.

As discussed in Chapter 10, it is difficult to characterize the nonclassicality noise robustness because the classical and quantum sets of behaviors need to be characterized. Otherwise, it cannot be assert whether or not $\mathcal{Q}^{\text{Net}} \subseteq \mathcal{C}^{\text{Net}}$. A more practical approach is to investigate the noise robustness of nonclassicality with respect to particular nonclassicality witness. Then the noise robustness of nonclassicality with respect to the particular witness is characterized by the critical noise parameter $\gamma_0$ at which the network can no longer exhibit nonclassicality. Furthermore, we find it convenient to rescale the nonclassicality witnesses such that their maximal possible score is one and their classical bound is zero.

For different inequalities bounding the same network, we find that their each individual inequality has a distinct noise robustness. Indeed this can be seen explicitly in point-to-point signaling systems (Chapter 14) and multi-access networks (Chapter 15), in which we consider a white noise model on the single receiver. Since there is just one receiver, mixing white noise into the quantum network behavior is equivalent to a depolarizing channel operating upon the measured state. We find that as the amount of white noise increases, the maximal violation decreases linearly with the noise parameter, however, the rate at which the maximal violation deteriorates is unique to each given inequality. As a result, not all nonclassicality witnesses have the same noise robustness.

The network's causal structure can also have a significant impact on the noise robustness of nonclassicality. As an example, in Chapter 13 we compare the star and chain network in the $n$-local setting. Although equivalent noise models can be applied to the sources of each network, the star network shows a greater noise robustness in the violation of the compared $n$-locality inequalities. In general, if one network has a greater robustness to noise than another, then property should be considered when designing quantum network systems to maximize the operational advantage of noisy networks.

In Chapter 13 we also consider how different types of noise affect nonclassicality, and in many cases, we derive the critical noise parameter at which the network's nonclassicality is broken with respect to the considered $n$-locality inequality. Overall, we find that some noise models are more destructive than others. For instance, depolarizing or white noise are the most destructive noise models. Furthermore, depolarizing deteriorates quantum states and measurements without bias meaning that no preferred encodings or decodings can be chosen to mitigate the effects of noise. In general however, noise models will typically have a bias that can be exploited. For this reason, channels like dephasing or bit flip channels are less destructive than depolarizing channels. Furthermore, the type of noise can even dictate whether maximally entangled state preparations are optimal for nonclassicality or not. For example, in Chapter 12, we show that the maximal violation of the CHSH scenario is a nonmaximally entangled state when amplitude damping noise is present on both sides of the entangled state. Additionally, we show that this property pertains to nonunital qubit channels, but not unital qubit channels, for which CHSH nonclassicality is always optimized by a maximally entangled state preparation.

Intuitively, as noise decreases the amount of nonclassicality, the operational value should also decrease. This can be quantified directly as noisy networks being easier to simulate than noiseless networks. As a trivial example, consider any network having a white noise behavior $\mathbf{P} = \sum_{x,y} \frac{1}{|\mathcal{Y}|} |y\rangle\langle x|$. Such a system requires

no classical communication or shared randomness to simulate because each network device can just output white noise.

In the case of weak nonclassicality, we find that the presence of small amounts of white noise can significantly decrease the classical simulation cost. In Chapter 14, we show by example that when white noise is incorporated into the weakly nonclassical planar symmetric behavior, the advantage where an unbounded amount of classical communication is needed vanishes. In these noisy cases, we find only a small advantage where a classical unassisted signaling dimension of 3 is needed to simulate a qubit with a considerable amount of depolarizing noise. Nevertheless, there is still a measurable quantum resource advantage even in the presence of white noise.

In general, when strong nonclassicality is considered, a resource advantage persists until the classical bound can no longer be violated. Thus, any statistically relevant violation of a classical bound demonstrates that additional quantum communication is needed. However, as the distance between the nonclassical behavior and the classical set decreases, so does the simulation cost. To see this, consider the case where a behavior can be written as a convex combination of behaviors $\mathbf{P}^Q = \mu\mathbf{P}^C + (1-\mu)\mathbf{V}^{NC}$ where $\mathbf{P}^Q \in \mathcal{Q}^{\text{Net}}$ is a nonclassical quantum behavior, $\mathbf{P}^C$ is a classical behavior on the boundary of $\mathcal{C}_\Lambda^{\text{Net}}$ and $\mathbf{V}^{NC}$ is a deterministic behavior not contained by $\mathcal{C}_\Lambda^{\text{Net}}$ that requires the signaling dimension $d$ to increase by exactly one to be simulated. Then to simulate the network, $d$ classical messages are needed with probability $\mu$ where $d+1$ are needed with probability $1-\mu$. As result, the simulation advantage can be considered to be $\mu d + (1-\mu)(d+1) = d + (1-\mu)$.

## *20.4 How might network design be informed by our results?*

As noisy quantum networks are developed, it will be important to maximize their noise robustness. As discussed in the previous section, it the noise robustness is guided by the network's topology, applied resources, physical noise model, and operational tasks.

At the highest-level of design are network topology and their expected operational tasks. Indeed, network topologies should be chosen to facilitate nonclassicality that is robust to noise. Similarly, to demonstrate operational advantage, quantum network tasks should be carefully selected to operate within the expected noise robustness.

At a lower level, the applied communication resources and physical implementation should be considered. In principle, the noise of one quantum hardware may be preferred over another because it is less destructive to nonclassicality. Likewise, there may be distinct noise associated with entanglement distribution as opposed to quantum communication. These noisy resources will need to be compared to determine which is more suitable for achieving the network's task.

Finally, a main point of this work is that variational optimization methods can be applied on quantum network hardware to train nonclassical behaviors in quantum networks. Since the optimal state encodings and measurements depend on the particular noise model, it is important that networks allow for tunable state preparations and measurements such that they can be recalibrated for optimal performance over a range of noise models and tasks.

## *20.5 Quantifying Operational Advantage by Nonclassicality*

Our motivating question is to better understand how the operational advantage of quantum systems deteriorates in the presence of noise. To investigate this problem, we argue that quantum nonclassicality

serves as a quantifier of operational advantage. It is no doubt that nonclassical behaviors require additional classical communication resources to reproduce. Thus, there is a direct connection between a behavior being nonclassical and a resource advantage. However, it is important to ask: what advantage does a nonclassical behavior provide?

Precisely, the resource advantage pertains to a specific task, which is simulating the nonclassical behavior. That is, our merit for operational advantage requires a system to produce a behavior that could not have been produced classically. However, there is no particular reason why the nonclassical behaviors of quantum mechanics ought to align with tasks that humans believe are useful and valuable.

Indeed, many quantum applications do not rely upon nonclassical behaviors at all. Nevertheless, these applications rely upon the existence of quantum properties such as superposition, entanglement, and measurement contextuality. Thus, the value of nonclassicality is not necessarily derived from the nonclassical behaviors themselves, but from the fact that those nonclassical behaviors can only exist from the use of quantum resources, assuming that the causal structure and communication limitations hold. Therefore, the practical application of nonclassicality is to judge the quality of quantum resources before they are applied.

It is also important to acknowledge the respective costs of quantum and classical communication resources. Classical signaling resources are cheap and robust as demonstrated by our existing information technology infrastructure. On the contrary, quantum communications technologies are currently experimental and require high-cost systems to implement. Now if we compare nonclassical behaviors to classical, we find a communication resource advantage. However, it doesn't usually require many bits to simulate a nonclassical quantum system, at least not in the studied networks. As a result, a little bit extra classical communication, which is cheap, can simulate quantum nonclassicality, which is expensive. Thus, the operational advantage of nonclassicality becomes a bit meaningless if the resource limitations are not respected. Thus, when nonclassicality is applied in multipoint communication networks care needs to be taken to ensure that no classical communication side-channels exist. However, this is not very realistic because classical communication is necessary for coordinating network devices.

## *20.6 The Advantages and Drawbacks of our Nonclassicality Framework*

The main advantages for our nonclassicality framework are that it is generic and hardware-agnostic. The procedure of deriving nonclassicality witnesses can be applied to any network topology. Throughgh optimization we are also able to find examples of nonclassicality in most quantum resource configurations applied in quantum networks. Therefore, nonclassicality can quantify the operational advantage in nearly any network configuration. Furthermore, nonclassicality is a device-independent quantifier because it does not require any knowledge of the physical hardware. Thus, the quality of communication resources can be evaluated independent of the physical implementation. This feature is important because it allows the nonclassicality to be compared between two networks having different physical implementations.

Another key advantage is that nonclassicality is practical for real-world applications because it has minimal resource requirements. Unlike resource complexity advantages that require many resources to demonstrate, nonclassicality can demonstrate advantage using simple systems. Therefore, when the number of resources is small, nonclassicality serves as a better quantifier of operational advantage than complexity. Furthermore, many nonclassicality witnesses can tolerate some amount of noise, allowing them to demonstrate operational advantage in realistic networks. However, it is less clear whether the complexity advantages known for quantum networks in the noiseless setting remain in the presence of noise, and what their noise robustness is.

One of the main challenges to overcome when working with nonclassicality is the exponential growth in the network polytope as the network scales. Both the dimension and the number of vertices scale exponentially with the number independent device inputs. This presents a major algorithm challenge to enumerating the vertices and deriving nonclassicality witnesses. Similarly, the large number of facet inequalities bounding any given network polytope make it difficult to determine whether a behavior is contained by the network polytope. However, given the incredible amount of relabeling symmetry in network polytopes, one workaround may be exploit this symmetry in theoretical arguments or computation to derive general nonclassicality witnesses for arbitrary network scenarios.

Another drawback is that our framework of nonclassicality requires the causal structure of the network to be known along with the amount of communication between each device. In practice, this requires a lot of trust to be placed on the network. While we argue in Chapter 18 that one trusted device can certify the quantum properties of the other, a device-independent characterization of an untrusted device is limited in what it can learn with certainty. However, if the network can be trusted, then nonclassicality can indeed quantify the operational advantage.

Furthermore, a cornerstone of our theoretical framework for nonclassicality is that we operate in the discrete memoryless setting, which might not be true for quantum networks. Note that quantum devices could store their classical data in memory where this data might correspond to an input or output of the network. The device could then update its operation based upon these results. If the network is trusted, then the devices can be trusted to operate without memory, however, trust is difficult to establish. More investigation would be needed to understand how the stored classical data could be used to imitate nonclassicality.

## *20.7 Future Nonclassicality Research*

Our framework of nonclassicality can generally be applied to any quantum network. Although nonsignaling networks have been heavily studied, there is still much to learn about nonclassicality in multipoint communication networks. In this direction, our work can be extended in a few ways.

First, we consider mainly qubit systems, leaving open questions about how our results generalize beyond qubit systems. We expect that nonclassicality will be present at any dimension, but it will be interesting to investigate the interplay between the signaling dimension and the nonclassicality noise robustness in a network.

Second, future theoretical works can further characterize the nonclassicality of multipoint communication networks. We show that nonclassicality can be found throughout many different resource configurations, however, we do not prove or derive the optimal encodings or decodings for nonclassicality. Furthermore, through studying these nonclassical behaviors, new self-testing and device-independent protocols may be derived, leading to new applications in quantum information science. Finally, deriving general nonclassicality witnesses for multipoint communication networks with many senders and receivers will be important for scaling our approaches.

Third, future works can demonstrate experimentally the nonclassical behaviors of multipoint communication networks. It remains to be investigated whether loop-hole free demonstrations of multipoint network nonclassicality can be achieved. Such tests could further support the predictions quantum theory, and demonstrate the quantum advantage of multipoint communication networks.

# Chapter 21

# Conclusion

Future quantum networks have the potential to revolutionize science, technology, and society, however, noise in quantum systems limits their capabilities. Thus, we seek to better understand how the operational advantage of quantum network technologies deteriorates in the presence of noise. In this dissertation we develop a comprehensive approach for quantifying operational advantage in terms of nonclassicality and investigate the noise robustness of nonclassicality in noisy quantum networks.

To quantify the operational advantage of quantum networks we develop a framework of operational nonclassicality, in which bounds are derived that contain the set of classical network behaviors. These bounds serve as operational tests, or games, that can witness nonclassicality using only the measurement statistics of the network. Nonclassical behaviors demonstrate operational advantage because additional classical resources are required to simulate the nonclassical behavior, hence, quantum communication resources have greater information processing power. We argue that quantum nonclassicality serves as a practical quantifier of quantum advantage because it is device-independent, depending only upon the network's causal structure and communication resources. In practice however, establishing nonclassical correlations in noisy quantum networks can be difficult due to challenges in characterization, simulation, and optimization of quantum networks.

To address these challenges, we develop a hybrid quantum-classical computing framework in which a quantum network, its causal structure, and its free operations are encoded into a variational quantum circuit ansatz. To simulate the network, the ansatz circuit can be evaluated either on classical or quantum hardware. The network can then be optimized using variational quantum optimization techniques. On a classical simulator and quantum hardware, we that these variational optimization techniques can find the optimal state preparations, unitary operations, and measurements for maximizing nonclassicality in the presence of noise. Furthermore, these methods are extensible to quantum network hardware directly enabling networks with uncharacterized noise models to be optimized for general networking tasks. Applying this concept further, we introduce variational quantum networking as a design archetype by which adaptive protocols can be patterned. In application, variational quantum networking enables automation in establishing and maintaining network protocols in noisy quantum systems.

Overall, our VQO approaches demonstrate the ability to reproduce known nonclassicality results and obtain new maximal violations in settings where nonclassicality has not previously been studied. Namely, we derive novel examples of nonclassicality in multipoint communication scenarios that apply quantum resources.

Our results shed light on the interplay of classical and quantum communication resources and how they can lead to quantum advantage. Furthermore through comparing the noise robustness of nonclassicality in different networks, we can identify certain network topologies that may be more advantageous in the presence of noise.

From a basic science perspective, our result confirm much of what is already known–there exists intrinsic randomness within quantum systems. This randomness can be viewed in our examples of weak nonclassicality, in which sharing an unlimited amount of randomness between devices allows them to reproduce the nonclassical behavior. More compelling, however, is the existence of strong nonclassicality in nonsignaling and signaling scenarios alike. The existence of nonclassicality in quantum systems implies not only that there is a quantum advantage, but also that the classical assumptions of locality, causality, or realism break down. Perhaps further study of the nonclassicality in multipoint communication scenarios will help elucidate the underlying mechanisms of nonclassicality, however, it may be more fruitful to apply the operational advantage of nonclassicality.

From a practical perspective, the first and most practical application of nonclassicality is to use it for self-testing and device-independent certification of quantum resources and functionality. The reason being that such certification tasks form the basis of integrating quantum network devices and quantifying the quality and operational advantage of the consumed quantum resources. Furthermore, nonclassicality may later more advanced applications in device-independent processing and distributed computation.

In conclusion, the operational advantage of quantum networks is quantified by their nonclassicality. Even in the presence of noise, a network exhibiting a nonclassical behavior demonstrates an advantage. Furthermore, we identify variational quantum optimization as a promising technology for automating noisy quantum protocols, gaining an improved robustness to uncharacterized noise that evolves over time.

# Part V

# Appendix

# Appendix A

# Proofs for Complete Sets of Signaling Polytope Facets

The proofs in this section are adapted from our work in reference [114].

## A.1 Proof of Theorem 10

*Proof.* Our proofs to parts (i) and (ii) follow the same approach. In both cases we want to show that the signaling polytope $\mathcal{C}_{\Lambda}^{\mathcal{X}\xrightarrow{d}\mathcal{Y}}$ is equivalent to some convex polytope that is bound by certain signaling dimension witnesses. We prove this fact by showing that the extreme points of the latter are also extreme points of the former; the converse has already been shown in Eq. (14.29). Recall that the extreme points of $\mathcal{C}_{\Lambda}^{\mathcal{X}\xrightarrow{d}\mathcal{Y}}$ consist of all extreme points of $\mathcal{P}_{\mathcal{Y}|\mathcal{X}}$ having rank no greater than $d$. In other words, $\mathbf{P}$ is extremal in $\mathcal{C}_{\Lambda}^{\mathcal{X}\xrightarrow{d}\mathcal{Y}}$ if and only if it is column stochastic with 0/1 elements and at most $d$ nonzero rows.

We rely heavily on the following general characterization of extreme points.

**Proposition 18.** Let $\mathcal{S} \subset \mathbb{R}^{n' \times n}$ be some convex polytope in $\mathbb{R}|\mathcal{Y}| \times |\mathcal{X}|$. Then $\mathbf{P}$ is an extreme point of $\mathcal{S}$ if and only if there does not exist some $\mathbf{D} \in \mathbb{R}^{|\mathcal{Y}|\times|\mathcal{X}|}$ such that $\mathbf{P} \pm \mathbf{D} \in \mathcal{S}$.

In our application of Proposition 18, we will refer to $\mathbf{D} \in \mathbb{R}^{|\mathcal{Y}|\times|\mathcal{X}|}$ as a "valid" perturbation of $\mathbf{P}$ if $\mathbf{P} \pm \mathbf{D} \in \mathcal{S}$; hence if $\mathbf{D}$ is a valid perturbation then $\mathbf{P}$ cannot be extremal.

Some other terminology used in our proofs is the following. For a channel $\mathbf{P} \in \mathcal{P}_{\mathcal{Y}|\mathcal{X}}$, an element $P(y|x)$ is called *non-extremal* if it lies in the open interval $(0, 1)$. We say that $P(y|x)$ is a *row maximizer* if it attains the largest value in row $y$ of $\mathbf{P}$. It is further called a *unique row maximizer* if there are no other elements in row $y$ having this value. Finally, we define the maximum likelihood estimation (ML) sum

$$\phi(\mathbf{P}) := \sum_{y=1}^{|\mathcal{Y}|} \|\mathbf{r}_y\|_\infty, \tag{A.1}$$

where $\mathbf{r}_y$ denotes row $y$ of $\mathbf{P}$ and $\|\mathbf{r}_y\|_\infty$ is its row maximizer. Then the maximum likelihood estimation (ML)

polytope can be expressed as

$$\mathcal{M}^{\mathcal{X}\overset{d}{\to}\mathcal{Y}} = \{\mathbf{P} \in \mathcal{P}_{\mathcal{Y}|\mathcal{X}} \mid \phi(\mathbf{P}) \leq d\}. \tag{A.2}$$

Note that $\phi$ is a convex function so that if $\phi(\mathbf{P}) = d$ with $\mathbf{P} = \sum_\lambda p_\lambda \mathbf{V}_\lambda$ for extreme points $\mathbf{V}_\lambda \in \mathcal{M}^{\mathcal{X}\overset{d}{\to}\mathcal{Y}}$ and non-negative numbers $p_\lambda$, then necessarily $\phi(\mathbf{V}_\lambda) = d$ for every $\lambda$.

Now we prove Theorem 10(i) in Section A.2.1, and we prove Theorem 10(ii) Section A.2.2. □

## *A.2 Proof of Theorem 10*

### *A.2.1 Proof of Theorem 10(i)*

*Proof.* The proof of Theorem 10(i) follows immediately from the following lemma due to the convexity of the ML and signaling polytopes.

**Lemma 17.** For arbitrary $\mathcal{X}$ and $\mathcal{Y}$, the extreme points of $\mathcal{M}^{\mathcal{X}\overset{d}{\to}\mathcal{Y}}$ where $d = |\mathcal{Y}| - 1$ are also extreme points of $\mathcal{C}_\Lambda^{\mathcal{X}\overset{d}{\to}\mathcal{Y}}$.

*Proof.* We first show the conclusion of Lemma 17 is true for any extreme point $\mathbf{V}$ of $\mathcal{M}^{\mathcal{X}\overset{d}{\to}\mathcal{Y}}$ having $d = |\mathcal{Y}| - 1$ and ML sum $\phi(\mathbf{V}) < |\mathcal{Y}| - 1$. If $\mathbf{V}$ is not extremal in $\mathcal{P}_{\mathcal{Y}|\mathcal{X}}$, then $\mathbf{V}$ must have at least one column $x$ with two non-extremal elements $V(y_1|x)$ and $V(y_2|x)$. However, we could then take two perturbations $V(y_1|x) \to V(y_1|x) \pm \epsilon$ and $V(y_2|x) \to V(y_2|x) \mp \epsilon$ with $\epsilon$ chosen sufficiently small so that the ML sum remains $\phi(\mathbf{V}) < |\mathcal{Y}| - 1$ and the numbers remain non-negative. Hence by contradiction, $\mathbf{V}$ must be extremal in $\mathcal{P}_{\mathcal{Y}|\mathcal{X}}$ and $\text{Rank}\,(\mathbf{V}) < |\mathcal{Y}| - 1$.

When $d = |\mathcal{Y}| - 1$, let us then consider an extremal point $\mathbf{V}$ of $\mathcal{M}^{\mathcal{X}\overset{d}{\to}\mathcal{Y}}$ for which $\phi(\mathbf{V}) = |\mathcal{Y}| - 1$. Since $\phi(\mathbf{V}) = |\mathcal{Y}| - 1$ is an integer and $\mathbf{V}$ has $|\mathcal{Y}|$ rows, then $\mathbf{V}$ must have at least two non-extremal row maximizers, which are possibly in different columns. We will again introduce perturbations, but care is needed to ensure that the perturbations are valid; *i.e.* the perturbed behaviors must remain in $\mathcal{M}^{\mathcal{X}\overset{d}{\to}\mathcal{Y}}$. There are two cases to consider:

**Case (a)**: Suppose that two non-extremal row maximizers occur in the same column: say $V(y_1|x)$ and $V(y_2|x)$ are both row maximizers in column $x$. Since these values will account for the contributions of rows $y_1$ and $y_2$ in the ML sum, and since there are only $|\mathcal{Y}|$ total rows in this sum, we must have that all other row maximizers are $+1$. Hence we introduce perturbations $V(y_1|x) \to V(y_1|x) \pm \epsilon$ and $V(y_2|x) \to V(y_2|x) \mp \epsilon$. If $V(y_1|x)$ and $V(y_2|x)$ are unique row maximizers, then this perturbation is valid. On the other hand, if there are columns $x', x''$ such that $V(y_1|x) = V(y_1|x')$ and/or $V(y_2|x) = V(y_2|x'')$ (with possibly $x' = x''$), then we must also introduce a corresponding perturbation $V(y_1|x') \to V(y_1|x') \pm \epsilon$ and/or $V(y_2|x'') \to V(y_2|x'') \mp \epsilon$. To preserve normalization in columns $x'$ and/or $x''$, we will have to introduce an off-setting perturbation to some other row in $x'$ and/or $x''$. This can always be done since either $x' = x''$, or $x'$ and/or $x''$ have a non-extremal element in some other row which is not a row maximizer (since all other row maximizers are $+1$).

**Case (b)**: No column has two non-extremal row maximizers, and $\mathbf{V}$ has at least two non-extremal row maximizers that belong to different columns. For each row $y$ with a non-extremal row maximizer, add perturbations $\pm\epsilon_y$ to all the row maximizers in that row. Since each column has at most one row maximizer,

a normalization-preserving perturbation $\mp\epsilon_y$ can be added to another non-extremal element in any column having a row maximizer in row $y$. Finally, choose the $\epsilon_y$ so that $\sum_{y\in\mathcal{Y}}\epsilon_y = 0$. □

□

### *A.2.2 Proof of Theorem 10(ii)*

*Proof.* We now turn to the ambiguous polytopes $\mathcal{A}_{\cap}^{\mathcal{X}\overset{d}{\to}\mathcal{Y}} := \cap_{k=|\mathcal{X}|}^{|\mathcal{Y}|}\mathcal{A}_k^{\mathcal{X}\overset{d}{\to}\mathcal{Y}}$ where $d = |\mathcal{X}| - 1$. Recall that $\mathcal{A}_k^{\mathcal{X}\overset{d}{\to}\mathcal{Y}}$ is the polytope of behaviors $\mathbf{P} \in \mathcal{P}_{\mathcal{Y}|\mathcal{X}}$ satisfying all signaling dimension witnesses of the form

$$\langle \mathbf{G}_k^{\mathcal{X}\overset{d}{\to}\mathcal{Y}}, \mathbf{P}\rangle \leq d = |\mathcal{X}| - 1, \tag{A.3}$$

where $\mathbf{G}_k^{\mathcal{X}\overset{d}{\to}\mathcal{Y}}$ has $k$ guessing rows and $(|\mathcal{Y}|-k)$ ambiguous rows. In this case, all the elements in an ambiguous row are equal to $\frac{1}{|\mathcal{X}|-d+1} = \frac{1}{2}$.

To prove Theorem 10(ii) we apply the following lemma to show that the extreme points of $\mathcal{A}_{\cap}^{\mathcal{X}\overset{d}{\to}\mathcal{Y}}$ are the same as those of $\mathcal{C}_{\Lambda}^{\mathcal{X}\overset{d}{\to}\mathcal{Y}}$ when $d = |\mathcal{X}| - 1$. Then by convexity of $\mathcal{A}_{\cap}^{\mathcal{X}\overset{d}{\to}\mathcal{Y}}$ and $\mathcal{C}_{\Lambda}^{\mathcal{X}\overset{d}{\to}\mathcal{Y}}$, we must have $\mathcal{A}_{\cap}^{\mathcal{X}\overset{d}{\to}\mathcal{Y}} = \mathcal{C}_{\Lambda}^{\mathcal{X}\overset{d}{\to}\mathcal{Y}}$.

**Lemma 18.** For arbitrary $|\mathcal{Y}| \geq |\mathcal{X}|$ and $d = |\mathcal{X}| - 1$, the extreme points of $\mathcal{A}_{\cap}^{\mathcal{X}\overset{d}{\to}\mathcal{Y}}$ are the extreme points of $\mathcal{C}_{\Lambda}^{\mathcal{X}\overset{d}{\to}\mathcal{Y}}$.

*Proof.* We first argue that the conclusion of Lemma 18 is true for any extreme point $\mathbf{V}$ of $\mathcal{A}_{\cap}^{\mathcal{X}\overset{d}{\to}\mathcal{Y}}$ such that $\langle \mathbf{G}_k^{\mathcal{X}\overset{d}{\to}\mathcal{Y}}, \mathbf{V}\rangle < |\mathcal{X}| - 1$ for all $\mathbf{G}_k^{\mathcal{X}\overset{d}{\to}\mathcal{Y}}$ and all integers $|\mathcal{X}| \leq k \leq |\mathcal{Y}|$. Analogously to Lemma 17, if $\mathbf{V}$ has at least one column $x$ with two non-extremal elements $V(y_1|x)$ and $V(y_2|x)$, we can take two sufficiently small perturbations $V(y_1|x) \to V(y_1|x) \pm \epsilon$ and $V(y_2|x) \to V(y_2|x) \mp \epsilon$ and still satisfy all the constraints of Eq. (A.3). Hence, $\mathbf{V}$ must be an extreme element of $\mathcal{P}_{\mathcal{Y}|\mathcal{X}}$. In this case, $\text{Rank}\,(\mathbf{V}) < |\mathcal{X}| - 1$ since $\phi(\mathbf{P}) < |\mathcal{X}| - 1$, and so $\mathbf{V} \in \mathcal{C}_{\Lambda}^{\mathcal{X}\overset{d}{\to}\mathcal{Y}}$.

It remains to prove the conclusion of Lemma 18 whenever Eq. (A.3) is tight for some $\mathcal{A}_k^{\mathcal{X}\overset{d}{\to}\mathcal{Y}}$. The lengthiest part of this argument is when $k = |\mathcal{Y}|$ and tightness in Eq. (A.3) corresponds to the ML sum equaling $|\mathcal{X}| - 1$. In this case, Proposition 20 below shows that $\mathbf{V}$ must be an extreme point of $\mathcal{C}_{\Lambda}^{\mathcal{X}\overset{d}{\to}\mathcal{Y}}$. However, before proving this result, we apply it to show that Lemma 18 holds whenever Eq. (A.3) is tight for some other $\mathbf{G}_k^{\mathcal{X}\overset{d}{\to}\mathcal{Y}}$ with $k < |\mathcal{Y}|$. Specifically, we will perform a lifting technique on any vertex $\mathbf{V}$ satisfying $\langle \mathbf{G}_k^{\mathcal{X}\overset{d}{\to}\mathcal{Y}}, \mathbf{V}\rangle = |\mathcal{X}| - 1$ and reduce it to the case of the ML sum equaling $(|\mathcal{X}| - 1)$.

Suppose that $\phi(\mathbf{V}) < |\mathcal{X}| - 1$ yet there exists some $\mathbf{G}_k^{\mathcal{X}\overset{d}{\to}\mathcal{Y}}$ such that $\langle \mathbf{G}_k^{\mathcal{X}\overset{d}{\to}\mathcal{Y}}, \mathbf{V}\rangle = |\mathcal{X}| - 1$. The matrix $\mathbf{G}_k^{\mathcal{X}\overset{d}{\to}\mathcal{Y}}$ identifies $(|\mathcal{Y}|-k)$ ambiguous rows, and suppose that $y$ is an ambiguous row such that $\frac{1}{2}\|\mathbf{r}_y\|_1 > \|\mathbf{r}_y\|_\infty$, with $\mathbf{r}_y$ being the $y^{th}$ row of $\mathbf{V}$. To be concrete, let us suppose without loss of generality that the components of row $\mathbf{r}_y$ are arranged in non-increasing order (*i.e.* $V(y|x_i) \geq V(y|x_{i+1})$), and let $k$ be the smallest index such that

$$\frac{1}{2}\left(-\sum_{i=1}^{k-1} V(y|x_i) + \sum_{i=k}^{|\mathcal{X}|} V(y|x_i)\right) \leq V(y|x_k). \tag{A.4}$$

By the assumption $\frac{1}{2}\|\mathbf{r}_y\|_1 > \|\mathbf{r}_y\|_\infty$, we have $k > 1$. Also, since $k$ is the smallest integer satisfying Eq. (A.4), we have

$$\frac{1}{2}\left(-\sum_{i=1}^{k-2} V(y|x_i) + \sum_{i=k-1}^{|\mathcal{X}|} V(y|x_i)\right) > V(y|x_{k-1}). \tag{A.5}$$

Subtracting $V(y|x_{k-1})$ from both sides of this equation implies that the LHS of Eq. (A.4) is strictly positive. Hence, there exists some $\lambda \in (0,1]$ such that

$$\lambda V(y|x_k) = \frac{1}{2}\left(-\sum_{i=1}^{k-1} V(y|x_i) + \sum_{i=k}^{|\mathcal{X}|} V(y|x_i)\right). \tag{A.6}$$

Consider then the new matrix $\widetilde{\mathbf{V}}$ formed from $\mathbf{V}$ by splitting row $y$ into $k$ rows as follows:

$$\mathbf{r}_y \to \begin{bmatrix} V(y|x_1) & 0 & \cdots & 0 & (1-\lambda)V(y|x_k) & V(y|x_{k+1}) & \cdots & V(y|x_n) \\ 0 & V(y|x_2) & \cdots & 0 & 0 & 0 & \cdots & 0 \\ \vdots & & \vdots & & \vdots & & \vdots & \\ 0 & 0 & \cdots & V(y|x_{k-1}) & 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 & \lambda V(y|x_k) & 0 & \cdots & 0 \end{bmatrix}. \tag{A.7}$$

Notice that we can obtain $\mathbf{V}$ from $\widetilde{\mathbf{V}}$ by coarse-graining over these rows. Moreover, this decomposition was constructed so that

$$\sum_{i=1}^{k} \|\widetilde{\mathbf{r}}_{y_i}\|_\infty = \sum_{i=1}^{k-1} V(y|x_i) + \lambda V(y|x_k) = \|\mathbf{r}_y\|_1, \tag{A.8}$$

where the $\widetilde{\mathbf{r}}_{y_i}$ are the rows in Eq. (A.7). Essentially this transformation allows us to replace an ambiguous row with a collection of guessing rows so that the overall guessing score does not change.

We perform this row splitting process on all ambiguous rows of $\mathbf{V}$ thereby obtaining a new matrix $\widetilde{\mathbf{V}}$ such that $\phi(\widetilde{\mathbf{V}}) = |\mathcal{X}| - 1$. If $|\mathcal{Y}|$ is the total number of rows in $\widetilde{\mathbf{V}}$, then $\widetilde{\mathbf{V}}$ will be an element of $\mathcal{A}_\cap^{\mathcal{X}\xrightarrow{d}\mathcal{Y}}$. We decompose $\widetilde{\mathbf{V}}$ into a convex combination of extremal points of $\mathcal{A}_\cap^{\mathcal{X}\xrightarrow{d}\mathcal{Y}}$ as $\widetilde{\mathbf{V}} = \sum_\lambda p_\lambda \widetilde{\mathbf{V}}_\lambda$. By the convexity of $\phi$, it follows that $\phi(\widetilde{\mathbf{V}}_\lambda) = |\mathcal{X}| - 1$, and we can therefore apply Proposition 20 below on the channels $\widetilde{\mathbf{V}}_\lambda$ to conclude that they are extreme points of $\mathcal{C}_\Lambda^{\mathcal{X}\xrightarrow{d}\mathcal{Y}}$. Consequently, each $\widetilde{\mathbf{V}}_\lambda$ has only one nonzero element per row. Let $\mathbf{R}$ denote the coarse-graining map such that $\mathbf{V} = \mathbf{R}\widetilde{\mathbf{V}}$, and apply

$$\mathbf{V} = \mathbf{R}\widetilde{\mathbf{V}} = \sum_\lambda p_\lambda \mathbf{R}\widetilde{\mathbf{V}}_\lambda. \tag{A.9}$$

However, by the assumption that $\mathbf{V}$ is extremal, this is only possible if $\mathbf{R}\widetilde{\mathbf{V}}_\lambda$ is the same for every $\lambda$. As a result, any two $\widetilde{\mathbf{V}}_\lambda$ and $\widetilde{\mathbf{V}}_{\lambda'}$ can differ only in rows that coarse-grain into the same rows by $\mathbf{R}$. From this it follows that $\mathbf{V}$ can have no more than one nonzero element per column and $\text{Rank}\,(\mathbf{V}) \leq |\mathcal{X}| - 1$. Hence we've shown that the extreme points of $\mathcal{A}_\cap^{\mathcal{X}\xrightarrow{d}\mathcal{Y}}$ are indeed extreme points of the signaling polytope $\mathcal{C}_\Lambda^{\mathcal{X}\xrightarrow{d}\mathcal{Y}}$.

To complete the proof of Lemma 18, we establish the case when $\phi(\mathbf{V}) = |\mathcal{X}| - 1$, as referenced above. We begin by proving the partial result provided by Proposition 19 and then, use this result to prove Proposition 20.

**Proposition 19.** If $\mathbf{V}$ is an extreme point of $\mathcal{A}_\cap^{\mathcal{X}\xrightarrow{d}\mathcal{Y}}$ satisfying $\phi(\mathbf{V}) = |\mathcal{X}| - 1$, then each column of $\mathbf{V}$ must have at least one unique row maximizer or it has only one nonzero element.

*Proof.* Suppose on the contrary that some column $x$ has more than one nonzero element yet no unique row maximizer. Let $\mathcal{S}_x \subset \mathcal{Y}$ be the set of rows for which column $x$ contains a row maximizer. Since only one row maximizer per row contributes to the ML sum, and the elements of column $x$ sum to one, we can satisfy $\phi(\mathbf{V}) = |\mathcal{X}| - 1$ if and only if both conditions hold:

(i) each row $y$ in $\mathcal{S}_x$ has only two nonzero elements $V(y|x)$ and $V(y|x_y)$ for some column $x_y \neq x$;

(ii) every other nonzero element in $\mathbf{V}$ outside of column $x$ and the rows in $\mathcal{S}_x$ are unique row maximizers.

With this structure, we introduce three cases of valid perturbations.

**Case (a)**: $V(y_1|x)$ and $V(y_2|x)$ are non-extremal elements in column $x$ with $y_1, y_2 \notin \mathcal{S}_x$. Then $V(y_1|x) \to V(y_1|x) \pm \epsilon$ and $V(y_2|x) \to V(y_2|x) \mp \epsilon$ is a valid perturbation. Indeed, even if we consider $y_1$ or $y_2$ as ambiguous rows, there is at most one other element in each of these rows (property (i) above), and so this perturbation would not violate any of the inequalities in (A.3).

**Case (b)**: $V(y_1|x)$ and $V(y_2|x)$ are non-extremal elements in column $x$ with $y_1 \in \mathcal{S}_x$ and $y_2 \notin \mathcal{S}_x$. Then $V(y_1|x) = V(y_1|x_{y_1})$ for some other column $x_{y_1} \neq x$. By normalization, there will be another element $V(y_3|x_{y_1})$ in column $x_{y_1}$ that by property (ii) is a unique row maximizer. Hence, we introduce perturbations

$$\begin{aligned} &V(y_1|x) \to V(y_1|x) \pm \epsilon && V(y_1|x_{y_1}) \to V(y_1|x_{y_1}) \pm \epsilon \\ &V(y_2|x) \to V(y_2|x) \mp \epsilon && \\ & && V(y_3|x_{y_1}) \to V(y_3|x_{y_1}) \mp \epsilon. \end{aligned} \tag{A.10}$$

For clarity, the line spacing is chosen here so that elements on the same vertical line correspond to elements in the same row of $\mathbf{V}$. By properties (i) and (ii), these perturbations do not increase the ML sum, nor are they able to violate any of the other inequalities in (A.3).

**Case (c)**: $V(y_1|x)$ and $V(y_2|x)$ are non-extremal elements in column $x$ with $y_1, y_2 \in \mathcal{S}_x$. Then $V(y_1|x) = V(y_1|x_{y_1})$ and $V(y_2|x) = V(y_2|x_{y_2})$ for some other columns $x_{y_1}, x_{y_2} \neq x$ (with possibly $x_{y_1} = x_{y_2}$). By normalization, there will be elements $V(y_3|x_{y_1})$ and $V(y_4|x_{y_2})$ in columns $x_{y_1}$ and $x_{y_2}$ respectively that are unique row maximizers (again by property (ii)). Note this requires that $y_1, y_2, y_3, y_4$ are all distinct rows. Hence, we introduce perturbations

$$\begin{aligned} &V(y_1|x) \to V(y_1|x) \pm \epsilon && V(y_1|x_{y_1}) \to V(y_1|x_{y_1}) \pm \epsilon && \\ &V(y_2|x) \to V(y_2|x) \mp \epsilon && && V(y_2|x_{y_2}) \to V(y_2|x_{y_2}) \mp \epsilon \\ & && V(y_3|x_{y_1}) \to V(y_3|x_{y_1}) \mp \epsilon && \\ & && && V(y_4|x_{y_2}) \to V(y_4|x_{y_2}) \pm \epsilon, \end{aligned} \tag{A.11}$$

Normalization is preserved under these perturbations and all the inequalities in (A.3) are satisfied.

As we have shown valid perturbations in all three cases under the assumption that some column has non-extremal elements with no unique row maximizer, the proposition follows.

□

**Proposition 20.** If $\mathbf{V}$ is an extreme point of $\mathcal{A}_{\cap}^{\mathcal{X} \overset{d}{\to} \mathcal{Y}}$ satisfying $\phi(\mathbf{V}) = |\mathcal{X}| - 1$, then $\mathbf{V}$ is an extreme point of $\mathcal{C}_{\Lambda}^{\mathcal{X} \overset{d}{\to} \mathcal{Y}}$.

*Proof.* Suppose that $\mathbf{V}$ has some column $x_1$ containing more than one nonzero element (if no such column can be found, then the proposition is proven). Let $V(y_1|x_1) \in (0, 1)$ denote a unique row maximizer, which is assured to exist by Proposition 19. We again proceed by considering two cases.

**Case (a)**: Column $x_1$ contains only one row maximizer $V(y_1|x_1)$ and all other elements in the column are not row maximizers. Then there must exist another column $x_1'$ that also contains at least two nonzero elements.

Indeed, if on the contrary all other columns only had one nonzero element each, then it would be impossible for $\phi(\mathbf{V}) = |\mathcal{X}| - 1$. If $x_1'$ only contains row maximizers, then proceed to case (b) and replace $x_1$ with $x_1'$. Otherwise, $x_1'$ does not only contain row maximizers; rather it has a unique row maximizer $V(y_3|x_1')$ in row $y_3$ and a nonzero element $V(y_4|x_1')$ in row $y_4$ that is not a row maximizer. Thus, we can introduce the valid perturbations

$$\begin{aligned} V(y_1|x_1) &\to V(y_1|x_1) \pm \epsilon \\ V(y_2|x_1) &\to V(y_2|x_1) \mp \epsilon \end{aligned} \qquad \begin{aligned} V(y_3|x') &\to V(y_3|x_1') \mp \epsilon \\ V(y_4|x') &\to V(y_4|x_1') \mp \epsilon \end{aligned} \tag{A.12}$$

where $V(y_2|x_1)$ denotes another nonzero element in $x_1$ (with possibly $y_2 = y_3, y_4$ and/or $y_3 = y_1$). It can be verified that all inequalities in (A.3) are preserved under these perturbations.

**Case (b)**: Column $x_1$ only contains row maximizers, with $V(y_2|x_1)$ being another one in addition to $V(y_1|x_1)$. If $V(y_2|x_1)$ is a unique row maximizer, then valid perturbations can be made to both $V(y_1|x_1)$ and $V(y_2|x_1)$. On the other hand, suppose that $V(y_2|x_1)$ is a non-unique row maximizer, and let $V(y_2|x_2) = V(y_2|x_1)$ be another row maximizer in column $x_2$. There can be no other nonzero elements in row $y_2$. Indeed, if there were another column, say $x_3$, such that $V(y_2|x_3) > 0$, then we would have

$$\frac{1}{2}\|\mathbf{r}_{y_2}\|_1 \geq \frac{1}{2}\Big(V(y_2|x_1) + V(y_2|x_2) + V(y_2|x_3)\Big) > V(y_2|x_2) = \|\mathbf{r}_{y_2}\|_\infty, \tag{A.13}$$

and so

$$\langle \mathbf{G}_{|\mathcal{Y}|-1}^{\mathcal{X} \overset{d}{\to} \mathcal{Y}}, \mathbf{V} \rangle > \phi(\mathbf{V}) = |\mathcal{X}| - 1 = d, \tag{A.14}$$

where the one ambiguous row in $\mathbf{G}_{|\mathcal{Y}|-1}^{\mathcal{X} \overset{d}{\to} \mathcal{Y}}$ is $y_2$. Hence, the only nonzero elements in row $y_2$ are $V(y_2|x_1)$ and $V(y_2|x_2)$. Let $V(y_3|x_2)$ be a unique row maximizer in column $x_2$.

We must be able to find another column $x_3$ with more than one nonzero element, one of which is a unique row maximizer and the other which is a non-unique row maximizer. For if this were not the case, then any other column in $\mathbf{V}$ would either have a unique row maximizer equaling one, or it would have at least two elements, one being a unique row maximizer and the others not being row maximizers. However, the latter possibility was covered in case (a) and was shown to be impossible for an extremal $\mathbf{V}$. For the former, if all then other $|\mathcal{X}|-2$ columns outside of $x_1$ and $x_2$ contain unique row maximizers equaling one, then they would collectively contribute an amount of $|\mathcal{X}| - 2$ to the ML sum. Since every element in column $x_1$ is a row maximizer, and $V(y_3|x_2)$ is a row maximizer in column $x_2$, we would have $\phi(\mathbf{V}) > (|\mathcal{X}| - 2) + 1 + V(y_3|x_2) > |\mathcal{X}| - 1$. Hence, there must exist another column $x_3$ with a non-unique row maximizer $V(y_5|x_3)$ that is shared with column $x_4$ (which may be equivalent to either $x_1$ or $x_2$). Letting $V(y_4|x_3)$ and $V(y_6|x_4)$ denote unique row maximizers

in columns $x_3$ and $x_4$, respectively, we can perform the valid perturbations

$$\begin{aligned}
&V(y_1|x_1) \to V(y_1|x_1) \pm \epsilon \\
&V(y_2|x_1) \to V(y_2|x_1) \mp \epsilon \quad V(y_2|x_2) \to V(y_2|x_2) \mp \epsilon \\
&\qquad\qquad\qquad V(y_3|x_2) \to V(y_3|x_2) \pm \epsilon \\
&\qquad\qquad\qquad\qquad\qquad\qquad V(y_4|x_3) \to V(y_4|x_3) \pm \epsilon \\
&\qquad\qquad\qquad\qquad\qquad\qquad V(y_5|x_3) \to V(y_5|x_3) \mp \epsilon \quad V(y_5|x_4) \to V(y_2|x_2) \mp \epsilon \\
&\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\quad V(y_6|x_4) \to V(y_6|x_4) \pm \epsilon.
\end{aligned} \tag{A.15}$$

Note that $y_1, y_2, y_3, y_4, y_5, y_6$ are all distinct rows since each row in $\mathbf{V}$ can have at most one pair of non-unique row maximizers while rows $y_1, y_3, y_4, y_6$ contain unique row maximizers. This assures that the perturbations do not violate the inequalities in (A.3).

As cases (a) and (b) exhaust all possibilities, we see that $\mathbf{V}$ can only have one nonzero element per column. From this the conclusion of Proposition 20 follows. □

This completes the proof of Lemma 18. □

□

### *A.2.3 Proof of Theorem 11*

*Proof.* In this section we prove the Theorem 11 to analyze the signaling polytope $\mathcal{C}_\Lambda^{\mathcal{X}\xrightarrow{d}\mathcal{Y}}$ when $d = 2$, $|\mathcal{Y}| = 4$, and $|\mathcal{X}|$ is arbitrary. To begin we define the polyhedron of behaviors

$$\mathcal{C}(\gamma, \mathbf{G}) := \{\mathbf{P} \in \mathcal{P}_{\mathcal{Y}|\mathcal{X}} \mid \langle \mathbf{G}, \mathbf{P} \rangle = \sum_{x=\mathcal{X}} \sum_{y=\mathcal{Y}} G_{y,x} P(y|x) \leq \gamma\} \tag{A.16}$$

for any signaling dimension witness $(\gamma, \mathbf{G})$ with $\mathbf{G} \in \mathbb{R}^{|\mathcal{Y}|\times|\mathcal{X}|}$ and $\gamma \in \mathbb{R}$. Since $\mathcal{C}_\Lambda^{\mathcal{X}\xrightarrow{d}\mathcal{Y}}$ is a convex polytope, there exists a finite number of polyhedra $\{\mathcal{C}(\gamma_m, \mathbf{G}_m)\}_{m=1}^r$ such that

$$\mathcal{C}_\Lambda^{\mathcal{X}\xrightarrow{d}\mathcal{Y}} = \bigcap_{m=1}^{r} \mathcal{C}(\gamma_m, \mathbf{G}_m). \tag{A.17}$$

**Remark.** Without loss of generality, we can assume that the matrices $\mathbf{G}_m$ contain non-negative elements. Indeed, if $G_{y,x} < 0$ is the smallest element in column $x$ of $\mathbf{G}_m$, then we replace each element in column $x$ as $G_{y',x} \to G_{y',x} + G_{y,x}$ and shift $\gamma \to \gamma + G_{y,x}$. Hence the smallest element in column $x$ of $\mathbf{G}_m$ becomes $G_{y,x} = 0$.

The proof of Theorem 11 is a consequence of Lemma 19 and Lemma 20 described below, as well as our numerical results for the $d = 2$, $|\mathcal{Y}| = 4$, and $|\mathcal{X}| = 6$ signaling polytope facets [349] (see Fig. 14.1). First, by Lemma 19, any signaling dimension witness $(\gamma, \mathbf{G})$ bounding $\mathcal{C}_\Lambda^{\mathcal{X}\xrightarrow{d}\mathcal{Y}}$ can be reduced to a new signaling dimension witness $(\hat{\gamma}, \hat{\mathbf{G}})$ having at most 2 nonzero elements in each column. The reduced inequality $(\hat{\gamma}, \hat{\mathbf{G}})$ satisfies $\mathcal{C}_\Lambda^{\mathcal{X}\xrightarrow{d}\mathcal{Y}} \subset \mathcal{C}(\hat{\gamma}, \hat{\mathbf{G}}) \subset \mathcal{C}(\gamma, \mathbf{G})$, and thus bounds $\mathcal{C}_\Lambda^{\mathcal{X}\xrightarrow{d}\mathcal{Y}}$ more tightly than the inequality $(\gamma, \mathbf{G})$. Next, we use Lemma 20 to show that for any integer $|\mathcal{X}|$, $d = 2$, and $|\mathcal{Y}| = 4$, a tight signaling dimension witness of $\mathcal{C}_\Lambda^{\mathcal{X}\xrightarrow{d}\mathcal{Y}}$ has at most six nonzero columns. The presence of all-zero columns implies that this inequality is an

input lifting of a tight signaling dimension witness for the signaling polytope having $|\mathcal{X}| = 6$. Therefore, the complete set of tight signaling dimension witnesses bounding any signaling polytope with $|\mathcal{Y}| = 4$ outputs and $d = 2$ is the set of all input liftings and permutations of the generator facets of shown in Table 14.1.

**Lemma 19.** If $\mathcal{C}_\Lambda^{\mathcal{X}\overset{d}{\to}\mathcal{Y}} \subset \mathcal{C}(\gamma, \mathbf{G})$, then there exists a polyhedron $\mathcal{C}(\hat{\gamma}, \hat{\mathbf{G}})$ with $\hat{\mathbf{G}}$ having at most $(|\mathcal{Y}| - d)$ nonzero elements in each column and satisfying

$$\mathcal{C}_\Lambda^{\mathcal{X}\overset{d}{\to}\mathcal{Y}} \subset \mathcal{C}(\hat{\gamma}, \hat{\mathbf{G}}) \subset \mathcal{C}(\gamma, \mathbf{G}). \tag{A.18}$$

*Proof.* Suppose $\mathcal{C}_\Lambda^{\mathcal{X}\overset{d}{\to}\mathcal{Y}} \subset \mathcal{C}(\gamma, \mathbf{G})$ and consider an arbitrary $x \in \mathcal{X}$. For convenience, let us relabel the elements of the $x^{th}$ column of $\mathbf{G}$ in non-increasing order; *i.e* $G_{y,x} \geq G_{y+1,x}$. Every vertex $\mathbf{V}$ of $\mathcal{C}_\Lambda^{\mathcal{X}\overset{d}{\to}\mathcal{Y}}$ will satisfy

$$\gamma \geq \sum_{x',y} G_{y,x'} V(y|x') = \sum_y G_{y,x} V(y|x) + \sum_{x' \neq x, y} G_{y,x'} V(y|x') = \sum_y G_{y,x} V(y|x) + f(\mathbf{G}, \mathbf{V}, x), \tag{A.19}$$

where $f(\mathbf{G}, \mathbf{V}, x) := \sum_{x' \neq x, y} G_{y,x'} V(y|x')$. A key observation is that, for every vertex $\mathbf{V}$ of $\mathcal{C}_\Lambda^{\mathcal{X}\overset{d}{\to}\mathcal{Y}}$,

$$\gamma \geq G_{d,x} + f(\mathbf{G}, \mathbf{V}, x). \tag{A.20}$$

We prove this observation using Eq. (A.19). First consider any vertex $\mathbf{V}$ such that $V(y|x) = \delta_{d'y}$ with $d' \geq d$. Then Eq. (A.19) shows that $\gamma \geq G_{d',x} + f(\mathbf{G}, \mathbf{V}, x) \geq G_{d,x} + f(\mathbf{G}, \mathbf{V}, x)$, since we have labeled the elements in non-increasing order. On the other hand, consider a vertex $\mathbf{V}$ for which $V(y|x) = \delta_{d'y}$ with $d' < d$. Since vertices can be formed with $d$ nonzero rows, we can choose another vertex $\mathbf{V}'$ that is identical to $\mathbf{V}$ in all columns $x' \neq x$, and yet for column $x$ it satisfies $V'(y|x) = \delta_{d''y}$ with $d \leq d''$. Hence applying Eq. (A.19) to vertex $\mathbf{V}'$ yields

$$\gamma \geq G_{d'',x} + f(\mathbf{V}', x) \geq G_{d,x} + f(\mathbf{G}, \mathbf{V}', x) = G_{d,x} + f(\mathbf{G}, \mathbf{V}, x), \tag{A.21}$$

where the last line follows from the fact that $\mathbf{V}$ and $\mathbf{V}'$ only differ in column $x$.

Having established Eq. (A.20), we next form a new matrix $\hat{\mathbf{G}}$ which is obtained from $\mathbf{G}$ by replacing its $x^{th}$ column with

$$(\hat{G}_{y,x})_y^T := (\overbrace{0, 0 \cdots, 0}^{d}, G_{d+1,x} - G_{d,x}, \cdots, G_{n',x} - G_{d,x})^T. \tag{A.22}$$

Letting $\hat{\gamma} = \gamma - G_{d,x}$, for any vertex $\mathbf{V}$ we have

$$\begin{aligned}
\sum_{x',y} \hat{G}_{y,x'} V(y|x') &= \sum_y \hat{G}_{y,x} V(y|x) + f(\mathbf{G}, \mathbf{V}, x) \\
&= \begin{cases} f(\mathbf{G}, \mathbf{V}, x) & \text{if } V(y|x) = \delta_{d'y} \text{ with } d' \leq d \\ G_{d',x} - G_{d,x} + f(\mathbf{G}, \mathbf{V}, x) & \text{if } V(y|x) = \delta_{d'y} \text{ with } d' > d \end{cases} \\
&\leq \hat{\gamma},
\end{aligned} \tag{A.23}$$

where the last inequality follows from Eq. (A.20) (in the first case) and Eq. (A.19) (in the second case).

Hence, we have that $\mathcal{C}_\Lambda^{\mathcal{X}\xrightarrow{d}\mathcal{Y}} \subset \mathcal{C}(\hat{\gamma}, \hat{\mathbf{G}})$. Conversely, if $\mathbf{P} \in \mathcal{C}(\hat{\gamma}, \hat{\mathbf{G}})$, then

$$\begin{aligned}
\gamma - G_{d,x} \geq \sum_{x',y} \hat{G}_{y,x'} P(y|x') &= \sum_y \hat{G}_{y,x} P(y|x) + \sum_{x' \neq x, y} \hat{G}_{y,x'} P(y|x') \\
&= \sum_{y=d+1}^{n'} (G_{y,x} - G_{d,x}) P(y|x) + \sum_{x' \neq x, y} \hat{G}_{y,x'} P(y|x') \\
&= -G_{d,x}(1 - \sum_{y=1}^{d} P(y|x)) + \sum_{y=d+1}^{n'} G_{y,x} P(y|x) + \sum_{x' \neq x, y} \hat{G}_{y,x'} P(y|x') \\
&\geq -G_{d,x} + \sum_{y=1}^{n'} G_{y,x} P(y|x) + \sum_{x' \neq x, y} \hat{G}_{y,x'} P(y|x') \\
&= -G_{d,x} + \sum_{x',y} G_{y,x'} P(y|x'). \qquad \text{(A.24)}
\end{aligned}$$

Therefore, $\mathbf{P} \in \mathcal{C}(\gamma, \mathbf{G})$ and so $\mathcal{C}_\Lambda^{\mathcal{X}\xrightarrow{d}\mathcal{Y}} \subset \mathcal{C}(\hat{\gamma}, \hat{\mathbf{G}}) \subset \mathcal{C}(\gamma, \mathbf{G})$. Note that if $\mathbf{G}$ has only non-negative elements then so will $\hat{\mathbf{G}}$. □

**Lemma 20.** For any finite number of inputs $|\mathcal{X}|$ were $d = 2$ and $|\mathcal{Y}| = 4$

$$\mathcal{C}_\Lambda^{\mathcal{X}\xrightarrow{d}\mathcal{Y}} = \bigcap_{m=1}^{s} \mathcal{C}(\gamma_m, \mathbf{G}_m) \qquad \text{(A.25)}$$

with each $\mathbf{G}_m$ having at most six nonzero columns.

*Proof.* As a consequence of Lemma 19, we can always find a complete set of polyhedra $\{\mathcal{C}(\hat{\gamma}_m, \hat{\mathbf{G}}_m)\}_{m=1}^s$ such that

$$\mathcal{C}_\Lambda^{\mathcal{X}\xrightarrow{d}\mathcal{Y}} = \bigcap_{m=1}^{s} \mathcal{C}(\hat{\gamma}_m, \hat{\mathbf{G}}_m)$$

where $|\mathcal{Y}| = 4$ and $d = 2$ such that each $\hat{\mathbf{G}}_m$ has no more than positive elements in each column and the rest being zero. Our goal is to show that the number of such columns can be reduced to six. The key steps in our reduction are given by the following two propositions.

**Proposition 21.** Consider the matrices

$$\hat{\mathbf{G}} = \begin{bmatrix} a & b & \cdot & \cdots \\ c & d & \cdot & \cdots \\ 0 & 0 & \cdot & \cdots \\ 0 & 0 & \cdot & \cdots \end{bmatrix}, \qquad \hat{\mathbf{G}}' = \begin{bmatrix} a-c & b+c & \cdot & \cdots \\ 0 & d+c & \cdot & \cdots \\ 0 & 0 & \cdot & \cdots \\ 0 & 0 & \cdot & \cdots \end{bmatrix}, \qquad a \geq c \geq 0, \qquad \text{(A.26)}$$

which differ only in the first two columns. Then $\mathcal{C}_\Lambda^{\mathcal{X}\xrightarrow{d}\mathcal{Y}} \subset \mathcal{C}(\hat{\gamma}, \hat{\mathbf{G}})$ if and only if $\mathcal{C}_\Lambda^{\mathcal{X}\xrightarrow{d}\mathcal{Y}} \subset \mathcal{C}(\hat{\gamma}, \hat{\mathbf{G}}')$.

*Proof.* When $d = 2$ and $|\mathcal{Y}| = 4$, every vertex $\mathbf{V}$ of $\mathcal{C}_\Lambda^{\mathcal{X}\xrightarrow{d}\mathcal{Y}}$ will have support in only two rows. If $\mathbf{V}$ has support in the first two rows, then its upper left corner will have one of the forms $\left(\begin{smallmatrix}1&1\\0&0\end{smallmatrix}\right)$, $\left(\begin{smallmatrix}1&0\\0&1\end{smallmatrix}\right)$, $\left(\begin{smallmatrix}0&1\\1&0\end{smallmatrix}\right)$, $\left(\begin{smallmatrix}0&0\\1&1\end{smallmatrix}\right)$. In each of these cases, $\langle \hat{\mathbf{G}}, \mathbf{V} \rangle \leq \hat{\gamma} \Leftrightarrow \langle \hat{\mathbf{G}}', \mathbf{V} \rangle \leq \hat{\gamma}$.

The other possibility is that $\mathbf{V}$ has support in only one of the first two rows. This leads to upper left corners of the form $\left(\begin{smallmatrix}1&1\\0&0\end{smallmatrix}\right)$, $\left(\begin{smallmatrix}0&0\\1&1\end{smallmatrix}\right)$, $\left(\begin{smallmatrix}1&0\\0&0\end{smallmatrix}\right)$, $\left(\begin{smallmatrix}0&1\\0&0\end{smallmatrix}\right)$, $\left(\begin{smallmatrix}0&0\\1&0\end{smallmatrix}\right)$, $\left(\begin{smallmatrix}0&0\\0&1\end{smallmatrix}\right)$. Suppose now that $\mathcal{C}_\Lambda^{\mathcal{X}\xrightarrow{d}\mathcal{Y}} \subset \mathcal{C}(\hat{\gamma}, \hat{\mathbf{G}})$. If a vertex $\mathbf{V}$

of $\mathcal{C}_\Lambda^{\mathcal{X}\xrightarrow{d}\mathcal{Y}}$ has form $\left(\begin{smallmatrix}1&0\\0&0\end{smallmatrix}\right)$ in the upper left corner, non-negativity of $c$ implies that $\langle\hat{\mathbf{G}}',\mathbf{V}\rangle \le \hat{\gamma}$. A somewhat less trivial case is any vertex $\mathbf{V}_1$ having form $\left(\begin{smallmatrix}0&1\\0&0\end{smallmatrix}\right)$ in the upper left corner. Here we need to use the fact that there exists a vertex $\mathbf{V}_2$ with $\left(\begin{smallmatrix}1&1\\0&0\end{smallmatrix}\right)$ in the upper left corner but is identical to $\mathbf{V}_1$ in all other columns. Hence we have

$$\hat{\gamma} \ge \langle\hat{\mathbf{G}},\mathbf{V}_2\rangle = a+b+\kappa \quad \Rightarrow \quad \langle\hat{\mathbf{G}}',\mathbf{V}_1\rangle = b+c+\kappa \le a+b+\kappa \le \hat{\gamma}, \tag{A.27}$$

where $\kappa$ is the contribution of the other columns to the inner product, and we have used the assumption that $a \ge c$. Similar reasoning shows that $\langle\hat{\mathbf{G}}',\mathbf{V}\rangle \le \hat{\gamma}$ for all other vertices $\mathbf{V}$. Conversely, by an analogous case-by-case consideration, we can establish that $\mathcal{C}_\Lambda^{\mathcal{X}\xrightarrow{d}\mathcal{Y}} \in \mathcal{C}(\hat{\gamma},\hat{\mathbf{G}}')$ implies $\langle\hat{\mathbf{G}},\mathbf{V}\rangle \le \hat{\gamma}$ for all vertices $\mathbf{V}$ of $\mathcal{C}_\Lambda^{\mathcal{X}\xrightarrow{d}\mathcal{Y}}$. □

**Proposition 22.** Consider the matrices

$$\hat{\mathbf{G}} = \begin{bmatrix} a & b & \cdot & \cdots \\ 0 & 0 & \cdot & \cdots \\ 0 & 0 & \cdot & \cdots \\ 0 & 0 & \cdot & \cdots \end{bmatrix}, \qquad \hat{\mathbf{G}}' = \begin{bmatrix} a+b & 0 & \cdot & \cdots \\ 0 & 0 & \cdot & \cdots \\ 0 & 0 & \cdot & \cdots \\ 0 & 0 & \cdot & \cdots \end{bmatrix}, \qquad \hat{\mathbf{G}}'' = \begin{bmatrix} 0 & a+b & \cdot & \cdots \\ 0 & 0 & \cdot & \cdots \\ 0 & 0 & \cdot & \cdots \\ 0 & 0 & \cdot & \cdots \end{bmatrix}. \tag{A.28}$$

which differ only in the first two columns. Then for $d=2$, $|\mathcal{Y}|=4$, and arbitrary $|\mathcal{X}|$, $\mathcal{C}_\Lambda^{\mathcal{X}\xrightarrow{d}\mathcal{Y}} \subset \mathcal{C}(\hat{\gamma},\hat{\mathbf{G}})$ if and only if $\mathcal{C}_\Lambda^{\mathcal{X}\xrightarrow{d}\mathcal{Y}} \in \mathcal{C}(\hat{\gamma},\hat{\mathbf{G}}') \cap \mathcal{C}(\hat{\gamma},\hat{\mathbf{G}}'')$.

*Proof.* This proof considers the vertices of $\mathcal{C}_\Lambda^{\mathcal{X}\xrightarrow{d}\mathcal{Y}}$ and applies the same reasoning as the proof of Proposition 21. □

Continuing with the proof of Lemma 20, suppose that $\mathcal{C}_\Lambda^{\mathcal{X}\xrightarrow{d}\mathcal{Y}} \in \mathcal{C}(\hat{\gamma}_m,\hat{\mathbf{G}}_m)$ with each column of $\hat{\mathbf{G}}_m$ having no more than two nonzero rows. We can group the columns into six groups according to which two rows have zero (it may be that a column has more than two zeros, in which case we just select one group to place it in). By repeatedly applying Proposition 21, we can replace $\hat{\mathbf{G}}_m$ with a matrix $\hat{\mathbf{G}}'_m$ such that each group has at most one column with two nonzero elements; the rest of the columns in that group have at most just one nonzero element. We then repeatedly apply Proposition 22 to remove multiple columns with the same single nonzero row. In the end, we arrive at the following:

$$\mathcal{C}_\Lambda^{\mathcal{X}\xrightarrow{d}\mathcal{Y}} \in \mathcal{C}(\hat{\gamma}_m,\hat{\mathbf{G}}_m) \qquad \Leftrightarrow \qquad \mathcal{C}_\Lambda^{\mathcal{X}\xrightarrow{d}\mathcal{Y}} \in \bigcap_j \mathcal{C}(\hat{\gamma}_m,\hat{\mathbf{G}}_{m,j}), \tag{A.29}$$

where each $\hat{\mathbf{G}}_{m,j}$ has at most ten nonzero columns corresponding to the different ways that no more than two nonzero elements can occupy a column. That is, up to a permutation of columns, each $\hat{\mathbf{G}}_{m,j}$ will have the form

$$\hat{\mathbf{G}}_{m,j} = \begin{bmatrix} a_1 & b_1 & c_1 & 0 & 0 & 0 & g & 0 & 0 & 0 & 0 & \cdots \\ a_2 & 0 & 0 & d_1 & e_1 & 0 & 0 & h & 0 & 0 & 0 & \cdots \\ 0 & b_2 & 0 & d_2 & 0 & f_1 & 0 & 0 & i & 0 & 0 & \cdots \\ 0 & 0 & c_2 & 0 & e_2 & f_2 & 0 & 0 & 0 & j & 0 & \cdots \end{bmatrix}. \tag{A.30}$$

The final step is to remove the block of diagonal elements $[g,h,i,j]$. To do this, observe that we absorb any of these diagonal elements into an earlier column, provided that the row contains the largest element in that

column. For example, if $f_2 > f_1$, then we can replace $\hat{\mathbf{G}}_{m,j}$ with

$$\hat{\mathbf{G}}'_{m,j} = \begin{bmatrix} a_1 & b_1 & c_1 & 0 & 0 & 0 & g & 0 & 0 & 0 & 0 & \cdots \\ a_2 & 0 & 0 & d_1 & e_1 & 0 & 0 & h & 0 & 0 & 0 & \cdots \\ 0 & b_2 & 0 & d_2 & 0 & f_1 & 0 & 0 & i & 0 & 0 & \cdots \\ 0 & 0 & c_2 & 0 & e_2 & f_2 + j & 0 & 0 & 0 & 0 & 0 & \cdots \end{bmatrix}, \tag{A.31}$$

and we can easily see that $\mathcal{C}_\Lambda^{\mathcal{X}\xrightarrow{d}\mathcal{Y}} \in \mathcal{C}(\hat{\gamma}_m, \hat{\mathbf{G}}_{m,j})$ if and only if $\mathcal{C}_\Lambda^{\mathcal{X}\xrightarrow{d}\mathcal{Y}} \in \mathcal{C}(\hat{\gamma}_m, \hat{\mathbf{G}}'_{m,j},)$. By considering the maximum element in each of the first six columns, we can perform this replacement for at least three of the four elements $[g, h, i, j]$. If we can do this for all four elements, then the proof is complete. On the other hand, if we can only remove three of these elements, then we will obtain a matrix $\hat{\mathbf{G}}''_{m,j}$ of the form (up to row/column permutations)

$$\hat{\mathbf{G}}''_{m,j} = \begin{bmatrix} a_1 & b_1 & c_1 & 0 & 0 & 0 & g & 0 & 0 & 0 & 0 & \cdots \\ a_2 & 0 & 0 & d_1 & e_1 & 0 & 0 & 0 & 0 & 0 & 0 & \cdots \\ 0 & b_2 & 0 & d_2 & 0 & f_1 & 0 & 0 & 0 & 0 & 0 & \cdots \\ 0 & 0 & c_2 & 0 & e_2 & f_2 & 0 & 0 & 0 & 0 & 0 & \cdots \end{bmatrix} \tag{A.32}$$

with $a_1, b_1, c_2$ not having the largest values in their respective columns. In this case, we construct the matrix

$$\hat{\mathbf{G}}'''_{m,j} = \begin{bmatrix} a_1 + g & b_1 + g & c_1 + g & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \cdots \\ a_2 + g & 0 & 0 & d_1 & e_1 & 0 & 0 & 0 & 0 & 0 & 0 & \cdots \\ 0 & b_2 + g & 0 & d_2 & 0 & f_1 & 0 & 0 & 0 & 0 & 0 & \cdots \\ 0 & 0 & c_2 + g & 0 & e_2 & f_2 & 0 & 0 & 0 & 0 & 0 & \cdots \end{bmatrix}, \tag{A.33}$$

from which it can be verified that $\mathcal{C}_\Lambda^{\mathcal{X}\xrightarrow{d}\mathcal{Y}} \subset \mathcal{C}(\hat{\gamma}_m, \hat{\mathbf{G}}''_{m,j})$ if and only if $\mathcal{C}_\Lambda^{\mathcal{X}\xrightarrow{d}\mathcal{Y}} \subset \mathcal{C}(\hat{\gamma}_m + 2g, \hat{\mathbf{G}}'''_{m,j})$. □

Thus we conclude our proof of Lemma 20. □

# Appendix B

# Proofs for Tight Signaling Dimension Witnesses

The derivations of tight signaling dimension witnesses contained by this section verify Lemma 6 by showing that all behaviors $\mathbf{P} \in \mathcal{C}_{\Lambda}^{\mathcal{X} \xrightarrow{d} \mathcal{Y}}$ satisfy $\langle \mathbf{G}, \mathbf{P} \rangle \leq \gamma$ and then constructing set of $\text{Dim}\left(\mathcal{C}_{\Lambda}^{\mathcal{X} \xrightarrow{d} \mathcal{Y}}\right) = |\mathcal{X}|(|\mathcal{Y}| - 1)$ affinely independent vertices $\{\mathbf{V} \in \mathcal{V}^{\mathcal{X} \xrightarrow{d} \mathcal{Y}} | \ \langle \mathbf{G}, \mathbf{V} \rangle = \gamma\}$. To assist with the enumeration of affinely independent vertices, we introduce a simple construction for affinely independent vectors with 0/1 elements.

**Lemma 21.** [114] Consider an $n$-element binary vector $\vec{b}_k \in \mathbb{B}^n$ with $n_0$ null elements and $n_1$ unit elements where $n_0 + n_1 = n$. A set of $n$ affinely independent vectors $\{\vec{b}_k\}_{k=1}^n$ is constructed as follows:

- Let $\vec{b}_1$ be the binary vector where the first $n_0$ elements are null and the next $n_1$ elements are unit values.
- For $k \in [2, n_0 + 1]$, $\vec{b}_k$ is derived from $\vec{b}_1$ by swapping the unit element at index $(n_0 + 1)$ with the null element at index $(k - 1)$.
- For $k \in [n_0 + 2, n]$, $\vec{b}_k$ is derived from $\vec{b}_1$ by swapping the null element at index $n_0$ with the unit element at index $k$.

For example, when $n = 5$, $n_0 = 2$, and $n_1 = 3$ the enumeration yields

$$\left\{\vec{b}_1 = [0, 0, 1, 1, 1], \quad \vec{b}_2 = [1, 0, 0, 1, 1], \quad \vec{b}_3 = [0, 1, 0, 1, 1], \quad \vec{b}_4 = [0, 1, 1, 0, 1], \quad \vec{b}_5 = [0, 1, 1, 1, 0]\right\}. \tag{B.1}$$

*Proof.* To verify the affine independence of $\{\vec{b}\}_{k=1}^n$ it is sufficient to show the linear independence of $\{\vec{b}_1 - \vec{b_k}\}_{k=2}^n$. Note that each $(\vec{b}_1 - \vec{b}_k)$ has two nonzero elements, one of which occurs at an index that is zero for the remaining vectors $(\vec{b}_1 - \vec{b}_{k'})$ where $k \neq k'$. Therefore, the vectors in $\{\vec{b}_1 - \vec{b_k}\}_{k=2}^n$ are linearly independent and $\{\vec{b}_k\}_{k=1}^n$ is affinely independent. □

## *B.1 Maximum Likelihood Facets of Signaling Polytopes*

**Proposition 23.** [114] The *maximum likelihood* (ML) signaling dimension witness $(d, \mathbf{G}_{\text{ML}}^{|\mathcal{Y}|})$ is a facet of all signaling polytopes $\mathcal{C}_{\Lambda}^{\mathcal{X} \xrightarrow{d} \mathcal{Y}}$ with $|\mathcal{Y}| = |\mathcal{X}|$ and $|\mathcal{Y}| > d > 1$.

*Proof.* To prove that $(d, \mathbf{G}_{\text{ML}}^{|\mathcal{Y}|})$ is a tight bound of $\mathcal{C}_\Lambda^{\mathcal{X}\xrightarrow{d}\mathcal{Y}}$ we construct a set of $\text{Dim}\left(\mathcal{C}_\Lambda^{\mathcal{X}\xrightarrow{d}\mathcal{Y}}\right) = |\mathcal{Y}|(|\mathcal{Y}|-1)$ affinely independent vertices $\{\mathbf{V} \in \mathcal{V}^{\mathcal{X}\xrightarrow{d}\mathcal{Y}} | \ \langle \mathbf{G}_{\text{ML}}^{|\mathcal{Y}|}, \mathbf{P}\rangle = d\}$. Taking $\mathbf{G}_{\text{ML}}^{|\mathcal{Y}|}$ to be the $|\mathcal{Y}| \times |\mathcal{Y}|$ identity matrix, a vertex $\mathbf{V}$ satisfies $d = \langle \mathbf{G}_{\text{ML}}^{|\mathcal{Y}|}, \mathbf{V}\rangle$ when $d$ unit elements of $\mathbf{V}$ lie along the diagonal. In this case, $(|\mathcal{Y}|-d)$ unit elements of $\mathbf{V}$ can be freely distributed in the remaining columns of the $d$ selected rows. For simplicity, we place all free elements in a single row with index $y \in \mathcal{Y}$ which we refer to as the target row. In the target row, we set $V(y|y) = 1$ while the off-diagonals, $V(y|x \neq y)$ with $x \in \mathcal{X} = \mathcal{Y}$ contain $(|\mathcal{Y}|-d)$ unit elements and $(d-1)$ null elements. Lemma 21 describes a construction of $(|\mathcal{Y}|-1)$ affinely independent vectors $\{\vec{b}_k\}_{k\in[|\mathcal{Y}|-1]}$ to set as the off-diagonals in the target row. Then, for each $x \in \mathcal{X} = \mathcal{Y}$ where $V(y|x \neq y) = 0$, we set $V(x|x) = 1$. This procedure obtains the upper bound in Eq. (14.34) and constructs an affinely independent vertex for each of the $(|\mathcal{Y}|-1)$ binary vectors in $\{\vec{b}_k\}_{k\in[|\mathcal{Y}|-1]}$. For example, targeting row $y = 3$ of $\mathbf{G}_{\text{ML}}^{5}$ when $d = 3$ yields four vertices,

$$\mathbf{V} \in \left\{ \begin{bmatrix} 1&0&0&0&0\\ 0&1&0&0&0\\ 0&0&1&1&1\\ 0&0&0&0&0\\ 0&0&0&0&0 \end{bmatrix}, \begin{bmatrix} 1&0&0&0&0\\ 0&0&0&0&0\\ 0&1&1&0&1\\ 0&0&0&1&0\\ 0&0&0&0&0 \end{bmatrix}, \begin{bmatrix} 1&0&0&0&0\\ 0&0&0&0&0\\ 0&1&1&1&0\\ 0&0&0&0&0\\ 0&0&0&0&1 \end{bmatrix}, \begin{bmatrix} 0&0&0&0&0\\ 0&1&0&0&0\\ 1&0&1&0&1\\ 0&0&0&1&0\\ 0&0&0&0&0 \end{bmatrix} \right\}. \tag{B.2}$$

Repeating the procedure for each $y \in \mathcal{Y}$ results in $|\mathcal{Y}|(|\mathcal{Y}|-1)$ affinely independent vertices. The vertices enumerated for each target row $y$ are affinely independent from all other target rows because the free unit elements are only allowed in the target row. As a final note, this procedure does not work in the case where $d = 1$ because there are only $|\mathcal{Y}|$ vertices in $\mathcal{V}^{\mathcal{X}\xrightarrow{d}\mathcal{Y}}$ or the case where $d = |\mathcal{Y}|$ because only one vertex $\mathbf{V} = \mathbf{G}_{\text{ML}}^{|\mathcal{Y}|}$ maximizes Eq. (14.34). Since $|\mathcal{Y}|(|\mathcal{Y}|-1) = \text{Dim}\left(\mathcal{C}^{\mathcal{X}\xrightarrow{d}\mathcal{Y}}\right)$ affinely independent vertices are constructed, $(d, \mathbf{G}_{\text{ML}}^{|\mathcal{Y}|})$ is proven to be a tight bound of all signaling polytopes with $|\mathcal{Y}| > d > 1$. □

## *B.2 Ambiguous Guessing Facets of Signaling Polytopes*

**Proposition 24.** [114] The inequality $(\gamma_?^{|\mathcal{Y}|,d}, \mathbf{G}_?^{|\mathcal{Y}|,d})$ is a facet of $\mathcal{C}_\Lambda^{\mathcal{X}\xrightarrow{d}\mathcal{Y}}$ when $|\mathcal{X}| = |\mathcal{Y}| - 1$ and $|\mathcal{Y}| - 2 \geq d \geq 2$.

*Proof.* To prove that $(\gamma_?^{|\mathcal{Y}|,d}, \mathbf{G}_?^{|\mathcal{Y}|,d})$ is a facet of $\mathcal{C}_\Lambda^{\mathcal{X}\xrightarrow{d}\mathcal{Y}}$ when $|\mathcal{X}| = |\mathcal{Y}| - 1$, we construct a set of $\text{Dim}\left(\mathcal{C}_\Lambda^{\mathcal{X}\xrightarrow{d}\mathcal{Y}}\right) = (|\mathcal{Y}|-1)^2$ affinely independent vertices $\{\mathbf{V} \in \mathcal{V}^{\mathcal{X}\xrightarrow{d}\mathcal{Y}} | \ d(|\mathcal{Y}|-d) = \langle \mathbf{G}_?^{|\mathcal{Y}|,d}, \mathbf{V}\rangle\}$. Using Proposition **??** we enumerate $(|\mathcal{Y}|-1)(|\mathcal{Y}|-2)$ affinely independent vertices that optimize the $\mathbf{G}_{\text{ML}}^{(|\mathcal{Y}|-1)}$ block. The remaining vertices use the ambiguous row and $(d-1)$ guessing rows. For these vertices, the ambiguous row has $(d-1)$ null elements and $(|\mathcal{Y}|-d)$ unit elements, hence, Lemma 21 can be used to enumerate $(|\mathcal{Y}|-1)$ affinely independent arrangements of the ambiguous row. For each of the $(|\mathcal{Y}|-1)$ arrangements, a vertex $\mathbf{V}_?$ is constructed by setting each $V_?(x|x) = 1$ where $x < |\mathcal{Y}|$ and $V_?(|\mathcal{Y}||x) = 0$. Combining the $(|\mathcal{Y}|-1)(|\mathcal{Y}|-2)$ vertices from the $\mathbf{G}_{\text{ML}}^{(|\mathcal{Y}|-1)}$ block and the $(|\mathcal{Y}|-1)$ vertices from the ambiguous row, a total of $(|\mathcal{Y}|-1)^2$ affinely independent vertices are found. Therefore, $(\gamma_?^{|\mathcal{Y}|,d}, \mathbf{G}_?^{|\mathcal{Y}|,d})$ is a tight signaling dimension witness of $\mathcal{C}_\Lambda^{\mathcal{X}\xrightarrow{d}\mathcal{Y}}$. The constraint that $(|\mathcal{Y}|-2) \geq d$ follows from the fact that, if $d \geq (|\mathcal{Y}|-1)$, then no optimal vertices use the ambiguous row resulting in an insufficient number of vertices to justify the facet. □

## *B.3 k-Guessing Facets of Signaling Polytopes*

**Proposition 25.** [114] The inequality $(\gamma_{\mathrm{K}}^{|\mathcal{Y}|,k,d}, \mathbf{G}_{\mathrm{K}}^{|\mathcal{Y}|,k})$ is a facet of $\mathcal{C}_{\Lambda}^{\mathcal{X}\xrightarrow{d}\mathcal{Y}}$ when $|\mathcal{X}| = \binom{|\mathcal{Y}|}{k}$, $|\mathcal{Y}| - 2 \geq k \geq 1$, and $d = |\mathcal{Y}| - k$.

*Proof.* To prove that $(\gamma_{\mathrm{K}}^{|\mathcal{Y}|,k,d}, \mathbf{G}_{\mathrm{K}}^{|\mathcal{Y}|,k})$ is a facet of $\mathcal{C}_{\Lambda}^{\mathcal{X}\xrightarrow{d}\mathcal{Y}}$ we construct a set of $\mathrm{Dim}\left(\mathcal{C}_{\Lambda}^{\mathcal{X}\xrightarrow{d}\mathcal{Y}}\right) = (|\mathcal{Y}|-1)\binom{|\mathcal{Y}|}{k}$ affinely independent vertices $\{\mathbf{V} \in \mathcal{V}^{\mathcal{X}\xrightarrow{d}\mathcal{Y}} |\ \gamma_{\mathrm{K}}^{|\mathcal{Y}|,k,d} = \langle \mathbf{G}_{\mathrm{K}}^{|\mathcal{Y}|,k}, \mathbf{V}\rangle\}$. Observe that separating the first row from the rest of $\mathbf{G}_{\mathrm{K}}^{|\mathcal{Y}|,k}$ results in a block matrix having the form,

$$\mathbf{G}_{\mathrm{K}}^{|\mathcal{Y}|,k} = \left[\begin{array}{c|c} \vec{1} & \vec{0} \\ \hline \mathbf{G}_{\mathrm{K}}^{(|\mathcal{Y}|-1),(k-1)} & \mathbf{G}_{\mathrm{K}}^{(|\mathcal{Y}|-1),k} \end{array}\right] \tag{B.3}$$

$$e.g.\quad \mathbf{G}_{\mathrm{K}}^{5,2} = \left[\begin{array}{cccc|cccccc} 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ \hline 1 & 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 1 & 1 \end{array}\right] = \left[\begin{array}{c|c} \vec{1} & \vec{0} \\ \hline \mathbf{G}_{\mathrm{K}}^{4,1} & \mathbf{G}_{\mathrm{K}}^{4,2} \end{array}\right], \tag{B.4}$$

where $\vec{0}$ and $\vec{1}$ are row vectors containing zeros and ones, and we refer to $\mathbf{G}_{\mathrm{K}}^{(|\mathcal{Y}|-1),(k-1)}$ and $\mathbf{G}_{\mathrm{K}}^{(|\mathcal{Y}|-1),k}$ as left and right $k$-guessing blocks respectively. The left and right $k$-guessing blocks suggest a recursive approach to our construction of affinely independent vertices. Namely, we construct $\binom{|\mathcal{Y}|}{k}$ vertices by targeting the first row of $\mathbf{G}_{\mathrm{K}}^{|\mathcal{Y}|,k}$ while Proposition 25 is recursively applied to enumerate the remaining vertices using the left and right $k$-guessing blocks. The recursion requires two base cases to be addressed:

1. For $d = 2$ and $|\mathcal{Y}| = k + d$, the construction of affinely independent vertices is given by Proposition 26.
2. For $k = 1$, the construction of affinely independent vertices is given by Proposition 23.

An iteration of this recursive construction proceeds as follows:

1. We construct an affinely independent vertex for each of the $\binom{|\mathcal{Y}|}{k}$ elements in the first row of $\mathbf{G}_{\mathrm{K}}^{|\mathcal{Y}|,k}$. For each index $x_1'$ in the $\vec{1}$ block, a vertex $\mathbf{V}_1$ is constructed by setting all $V_1(1|x) = 1$ where $x \neq x_1'$ and $V_1(y|x_1') = 1$ where $y > 1$ is the smallest row index such that $G_{y,x_1'} = 1$. The remaining rows of $\mathbf{V}_1$ are filled to maximize the right $k$-guessing block. Then, for each index $x_0'$ in the $\vec{0}$ block, a vertex $\mathbf{V}_0$ is constructed by setting $V_0(1|x_0') = 1$ and all $V_0(1|x) = 1$ where $G_{1,x} = 1$. The remaining $(d-1)$ rows of $\mathbf{V}_0$ are filled to maximize the right $k$-guessing block. This procedure enumerates $\binom{|\mathcal{Y}|}{k}$ affinely independent vertices.
2. The remaining $(|\mathcal{Y}|-2)\binom{|\mathcal{Y}|}{k}$ vertices are found by individually targeting the left and right $k$-guessing blocks. To construct a vertex $\mathbf{V}_L$ using the left block $\mathbf{G}_{\mathrm{K}}^{(|\mathcal{Y}|-1)),(k-1)}$, the first row of $\mathbf{V}_L$ is not used. The left block is then a $(k-1)$-guessing game with $(|\mathcal{Y}|-1)$ outputs where $d = (|\mathcal{Y}|-1) - (k-1) = |\mathcal{Y}| - k$, hence, Proposition 25 holds and $(|\mathcal{Y}|-2)\binom{|\mathcal{Y}|-1}{k-1}$ affinely independent vertices are enumerated using the described recursive process. Note that for each vertex of form $\mathbf{V}_L$, the remaining elements are filled to maximize the right $k$-guessing block $\mathbf{G}_{\mathrm{K}}^{(|\mathcal{Y}|-1),k}$. Similarly, to construct a vertex $\mathbf{V}_R$ using the right block $\mathbf{G}_{\mathrm{K}}^{(|\mathcal{Y}|-1),k}$, we set all elements $V_R(1|x) = 1$ where $G_{1,x}^{|\mathcal{Y}|,k} = 1$. The remaining $(d-1)$ rows of $\mathbf{V}_R$ are filled by optimizing the $\mathbf{G}_{\mathrm{K}}^{(|\mathcal{Y}|-1),k}$ block. Since $d = |\mathcal{Y}| - k$ and $(d-1) = (|\mathcal{Y}|-1) - k$, Proposition

25 holds, and recursively applying this procedure constructs $(|\mathcal{Y}|-2)\binom{|\mathcal{Y}|-1}{k}$ vertices of form $\mathbf{V}_R$ using the right $k$-guessing block.

3. Vertices of forms $\mathbf{V}_0$, $\mathbf{V}_1$, $\mathbf{V}_L$ and $\mathbf{V}_R$ are easily verified to be affinely independent. Summing these vertices yields $(|\mathcal{Y}|-2)\binom{|\mathcal{Y}|-1}{k-1}+(|\mathcal{Y}|-2)\binom{|\mathcal{Y}|-1}{k}+\binom{|\mathcal{Y}|}{k}=(|\mathcal{Y}|-1)\binom{|\mathcal{Y}|}{k}$ affinely independent vertices, therefore, the $k$-guessing signaling dimension witness is tight when $|\mathcal{Y}|=k+d$.

□

**Proposition 26.** [114] The *k-guessing game* signaling dimension witness $(\gamma_{\mathrm{K}}^{|\mathcal{Y}|,k,d},\mathbf{G}_{\mathrm{K}}^{|\mathcal{Y}|,k})$ is a tight bound for all signaling polytopes $\mathcal{C}_{\Lambda}^{\mathcal{X}\overset{d}{\to}\mathcal{Y}}$ with $|\mathcal{X}|=\binom{|\mathcal{Y}|}{k}$, $d=2$, and $k=|\mathcal{Y}|-2$.

*Proof.* To prove that the inequality is tight, we construct a set containing $(|\mathcal{Y}|-1)\binom{|\mathcal{Y}|}{k}$ affinely independent vertices $\{\mathbf{V}\in\mathcal{V}^{\mathcal{X}\overset{d=2}{\to}\mathcal{Y}}\mid\langle\mathbf{G}_{\mathrm{K}}^{|\mathcal{Y}|,(|\mathcal{Y}|-2)},\mathbf{V}\rangle=\binom{|\mathcal{Y}|}{|\mathcal{Y}|-2}-1\}$. To help illustrate this proof, we use the example of $(\gamma_{\mathrm{K}}^{5,3,2},\mathbf{G}_{\mathrm{K}}^{5,3})$ where

$$\mathbf{G}_{\mathrm{K}}^{5,3}=\begin{bmatrix}1&1&1&1&1&1&0&0&0&0\\1&1&1&0&0&0&1&1&1&0\\1&0&0&1&1&0&1&1&0&1\\0&1&0&1&0&1&1&0&1&1\\0&0&1&0&1&1&0&1&1&1\end{bmatrix}. \tag{B.5}$$

and $\gamma_{\mathrm{K}}^{5,3,2}=\binom{5}{3}-1$. Since $d=2$, we consider vertices $\mathbf{V}\in\mathcal{V}^{\mathcal{X}\overset{d=2}{\to}\mathcal{Y}}$ having $\mathrm{Rank}\,(\mathbf{V})=2$ where each vertex uses two rows $y$ and $y'$ where $y<y'$. In general, each of the $\binom{|\mathcal{Y}|}{2}$ two-row selections from $\mathbf{G}_{\mathrm{K}}^{|\mathcal{Y}|,(|\mathcal{Y}|-2)}$ have a unique column $x_0$ containing null elements both rows $y$ and $y'$. Therefore, for each unique pair $y$ and $y'$, two affinely independent vertices $\mathbf{V}_1$ and $\mathbf{V}_2$ are constructed by setting $V_1(y|x_0)=1$ and $V_2(y'|x_0)=1$ while the remaining terms are arranged such that all unit elements in row $y$ and the remaining elements in row $y'$ are selected to achieve the optimal score. Performing this procedure for the first two rows of $\mathbf{G}_{\mathrm{K}}^{5,3}$ ($y=1$ and $y'=2$) constructs the vertices

$$\mathbf{V}_1=\begin{bmatrix}1&1&1&1&1&1&0&0&0&1\\0&0&0&0&0&0&1&1&1&0\\0&0&0&0&0&0&0&0&0&0\\0&0&0&0&0&0&0&0&0&0\\0&0&0&0&0&0&0&0&0&0\end{bmatrix},\quad \mathbf{V}_2=\begin{bmatrix}1&1&1&1&1&1&0&0&0&0\\0&0&0&0&0&0&1&1&1&1\\0&0&0&0&0&0&0&0&0&0\\0&0&0&0&0&0&0&0&0&0\\0&0&0&0&0&0&0&0&0&0\end{bmatrix} \tag{B.6}$$

where $x_0=10$ in this example. Repeating this procedure for each of the $\binom{|\mathcal{Y}|}{2}$ row selections produces, $2\binom{|\mathcal{Y}|}{2}=2\binom{|\mathcal{Y}|}{k}$ affinely independent vertices, one for each null element in $\mathbf{G}_{\mathrm{K}}^{|\mathcal{Y}|,(|\mathcal{Y}|-2)}$.

The remaining vertices are constructed by selecting a target row $y\in[|\mathcal{Y}|-1]$. In the target row, for each $x'$ where $G_{y,x'}=1$ a vertex $\mathbf{V}_3$ is constructed by setting $V_3(y|x)=1$ for all $x\neq x'$ that satisfy $G_{y,x}=1$. A secondary row $y'>y$ of $\mathbf{V}_3$ is chosen where $y'$ is the smallest index satisfying $G_{y',x'}=1$. We then set $V(y'|x')=1$ while the remaining elements of $\mathbf{V}_3$ are set to achieve the optimal score. For selected rows $y$ and $y'$, the null column at index $x_0$ is set in the target row as $V_3(y|x_0)=1$. For example, consider $\mathbf{G}_{\mathrm{K}}^{5,3}$ with

the target row as $y = 1$ and $x' = 4$ we construct the vertex,

$$\mathbf{V}_3 = \begin{bmatrix} 1 & 1 & 1 & 0 & 1 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 1 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \tag{B.7}$$

Note that all secondary row indices $y' \leq y + 3$ are required to construct a vertex $\mathbf{V}_3$ for each unit element in the target row $y$. Let $\Delta y = y' - y$, then $\sum_{\Delta y=1}^{3} \binom{|\mathcal{Y}|-1-\Delta y}{d+1-\Delta y}$ vertices are constructed for target row $y$. For $y = |\mathcal{Y}| - 2$ and $y = |\mathcal{Y}| - 1$, the sum terminates at $\Delta y = 2$ and $\Delta y = 1$ respectively because the vertices are only affinely independent if the secondary row has index $y' > y$. Thus, this process produces

$$\sum_{\Delta y=1}^{3} (|\mathcal{Y}| - \Delta y) \binom{|\mathcal{Y}| - 1 - \Delta y}{d + 1 - \Delta y} = (|\mathcal{Y}| - 3) \binom{|\mathcal{Y}|}{d} \tag{B.8}$$

affinely independent vertices where the identities $\frac{l}{m}\binom{l}{m} = \binom{l-1}{m-1}$ and $\frac{l+1-m}{m}\binom{l}{m} = \binom{l}{m-1}$ are used to convert the binomial coefficients to the form $\binom{|\mathcal{Y}|}{d} = \binom{|\mathcal{Y}|}{k}$. Combining the vertices of form $\mathbf{V}_1$, $\mathbf{V}_2$, and $\mathbf{V}_3$ yields a set of $2\binom{|\mathcal{Y}|}{k} + (|\mathcal{Y}| - 3)\binom{|\mathcal{Y}|}{k} = (|\mathcal{Y}| - 1)\binom{|\mathcal{Y}|}{k}$ affinely independent vertices. Therefore, when $d = 2$ and $k = |\mathcal{Y}| - 2$, $(\binom{|\mathcal{Y}|}{|\mathcal{Y}|-2} - 1, \mathbf{G}_{\mathrm{K}}^{|\mathcal{Y}|,(|\mathcal{Y}|-2)})$ is a tight signaling dimension witness of the signaling polytope $\mathcal{C}_{\Lambda}^{\mathcal{X} \xrightarrow{d} \mathcal{Y}}$ where $|\mathcal{X}| = \binom{|\mathcal{Y}|}{k}$. □

## *B.4 Anti-Guessing Facets of Signaling Polytopes*

**Proposition 27.** [114] The inequality $(\gamma_{\mathrm{A}}^{\varepsilon,d}, \mathbf{G}_{\mathrm{A}}^{\varepsilon,m'})$ is a facet of $\mathcal{C}_{\Lambda}^{\mathcal{X} \xrightarrow{d} \mathcal{Y}}$ where $|\mathcal{X}| = |\mathcal{Y}|$, $|\mathcal{Y}| - 2 \geq d \geq 2$, and $|\mathcal{Y}| - d + 1 \geq \varepsilon \geq 3$.

*Proof.* To prove the tightness of the anti-guessing facet inequality we derive a row-by-row construction showing that $\mathrm{Dim}\left(\mathcal{C}_{\Lambda}^{\mathcal{X} \xrightarrow{d} \mathcal{Y}}\right) = |\mathcal{Y}|(|\mathcal{Y}| - 1)$ affinely independent vertices $\{\mathbf{V} \in \mathcal{V}^{\mathcal{X} \xrightarrow{d} \mathcal{Y}} | \langle \mathbf{G}_{\mathrm{A}}^{\varepsilon,m'}, \mathbf{V} \rangle = \gamma_{\mathrm{A}}^{\varepsilon,d}\}$. For convenience, we refer to the first $\varepsilon$ rows of $\mathbf{G}_{\mathrm{A}}^{\varepsilon,m'}$ as *anti-guessing* rows and the remaining $m'$ rows as *guessing* rows. We treat anti-guessing and guessing rows individually because each admits its own vertex construction whose vertices are affinely independent each other. To help illustrate this proof, we draw upon the example where $\varepsilon = m' = d = 3$,

$$\mathbf{G}_{\mathrm{A}}^{3,3} = \left[\begin{array}{ccc|ccc} 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{array}\right] \quad \text{and} \quad \gamma^{3,3} = 4, \tag{B.9}$$

however, the construction generalizes to anti-guessing facets of arbitrary size.

For a target anti-guessing row $y \in [1, \varepsilon]$ we construct $(|\mathcal{Y}| - 1)$ vertices where $(\varepsilon - 1)$ vertices are constructed using the $\mathbf{G}_A^{\varepsilon}$ block and $m'$ vertices are constructed using the $\hat{0}$ block in the top right. Note that a vertex can only achieve the upper bound $\gamma_{\mathrm{A}}^{\varepsilon,d}$ if two or less anti-guessing rows are used. A vertex $\mathbf{V}_{\mathrm{A}}$ is constructed using the $\mathbf{G}_{\mathrm{A}}^{\varepsilon,m'}$ block by setting $V_{\mathrm{A}}(y|x) = 1$ for all $x$ that satisfy $G_{y,x}^{\varepsilon,m'} = 1$ and selecting a secondary row

$y' \neq y$ with $y' \in [1, \varepsilon]$ and setting $V_{\mathrm{A}}(y'|x') = 1$ where $x'$ is the index of the null element in the target row $G^{\varepsilon,m'}_{y,x'} = 0$. All remaining elements of $\mathbf{V}_{\mathrm{A}}$ are set so that the first $(d-2)$ diagonal elements of the $\mathbf{G}^{m'}_{\mathrm{ML}}$ block are selected and any remaining terms are set as unit elements in the target row. An affinely independent vertex is constructed for each of the $(\varepsilon - 1)$ choices of secondary row $y'$. For example, when targeting row $y = 1$ we enumerate two vertices

$$\mathbf{V}_{\mathrm{A}} \in \left\{ \left[\begin{array}{ccc|ccc} 1 & 1 & 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{array}\right], \left[\begin{array}{ccc|ccc} 1 & 1 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{array}\right] \right\}. \tag{B.10}$$

For a target anti-guessing row $y$, an additional $m'$ vertices with form $\mathbf{V}_{\mathrm{A},0}$ are constructed using the $\hat{0}$ block in the top right. If $m' > (d-1)$, we set the target row as $V_{\mathrm{A},0}(y|x) = 1$ where $x \in [1, \varepsilon]$. The remaining $(d-1)$ rows are then used to maximize the $\mathbf{G}^{m'}_{\mathrm{ML}}$ block. Using Lemma 21 a set of $m'$ affinely independent vectors $\{\vec{b}_k\}^{m'}_{k=1}\}$ with $(d-1)$ null elements and $(m'-d+1)$ unit elements can be constructed and used in the $\hat{0}$ block of $\mathbf{V}_{\mathrm{A},0}$ by setting $V_{\mathrm{A},0}(y|[\varepsilon+1, |\mathcal{Y}|]) = \vec{b}_k$. All remaining null elements in the target row of $\mathbf{V}_{\mathrm{A},0}$ are then set along the diagonal of the $\mathbf{G}^{m'}_{\mathrm{ML}}$ block. Since there are $m'$ choices of $\vec{b}_k$, that many affinely independent vertices can be constructed. For example, when targeting row $y = 1$ we enumerate 3 vertices,

$$\mathbf{V}_{\mathrm{A},0} \in \left\{ \left[\begin{array}{ccc|ccc} 1 & 1 & 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{array}\right], \left[\begin{array}{ccc|ccc} 1 & 1 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{array}\right], \left[\begin{array}{ccc|ccc} 1 & 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{array}\right] \right\}. \tag{B.11}$$

If $m' = (d-1)$, a secondary anti-guessing row $y'$ is selected where the anti-guessing rows are set as $V_{\mathrm{A},0}(y|x) = 1$ and $V_{\mathrm{A},0}(y'|x')$ where $x, x' \in [1, \varepsilon]$ and $G^{\varepsilon,m'}_{y,x} = 1$ and $G^{\varepsilon,m'}_{y,x'} = 0$. The remainder of the procedure is the same as the $m' > (d-1)$ case. Note that in the $m' = (d-1)$ case one of the $\mathbf{V}_{\mathrm{A},0}$ vertices is redundant of a $\mathbf{V}_{\mathrm{A}}$ vertex. To reconcile this conflict another vertex must be added which maximizes $\mathbf{G}^{m'}_{\mathrm{ML}}$ with $V(y|x) = 1$ for all $x \in [1, \varepsilon]$ and $V(x'|x') = 1$ for all $x' \in [\varepsilon+1, |\mathcal{Y}|]$. By this procedure $(\varepsilon - 1) + m' = (|\mathcal{Y}| - 1)$ affinely independent vertices are constructed for each target row $y \in [1, \varepsilon]$. Thus, $\varepsilon(|\mathcal{Y}| - 1)$ affinely independent vertices are constructed for the anti-guessing rows of $\mathbf{G}^{\varepsilon,m'}_{\mathrm{A}}$.

For a target guessing row $y \in [\varepsilon+1, |\mathcal{Y}|]$ we construct $(|\mathcal{Y}| - 1)$ vertices where $\varepsilon$ are constructed using the $\hat{0}$ block in the lower left and $(m'-1)$ vertices using the $\mathbf{G}^{m'}_{\mathrm{ML}}$ block. Starting with the lower left $\hat{0}$ block we construct a vertex $\mathbf{V}_{\mathrm{ML},0}$ for each $x \in [1, \varepsilon]$ by setting $V_{\mathrm{ML},0}(y|x) = 1$ and $V_{\mathrm{ML},0}(y|y) = 1$. Of the remaining $(d-1)$ rows one is used to maximize the $\mathbf{G}^{\varepsilon}_{\mathrm{A}}$ block and $(d-2)$ rows maximize the $\mathbf{G}^{m'}_{\mathrm{ML}}$ block. Any unspecified unit terms of $\mathbf{V}_{\mathrm{ML},0}$ are set in the target row $y$. Since there are $\varepsilon$ values of $x$ to consider, this procedure produces $\varepsilon$ affinely independent vertices. For example, when targeting row $y = 4$ we enumerate 3 vertices,

$$\mathbf{V}_{\mathrm{ML},0} \in \left\{ \left[\begin{array}{ccc|ccc} 0&0&0&0&0&0\\ 0&0&0&0&0&0\\ 0&1&1&0&0&0\\ \hline 1&0&0&1&0&1\\ 0&0&0&0&1&0\\ 0&0&0&0&0&0 \end{array}\right], \left[\begin{array}{ccc|ccc} 0&0&0&0&0&0\\ 1&0&1&0&0&0\\ 0&0&0&0&0&0\\ \hline 0&1&0&1&0&1\\ 0&0&0&0&1&0\\ 0&0&0&0&0&0 \end{array}\right], \left[\begin{array}{ccc|ccc} 1&1&0&0&0&0\\ 0&0&0&0&0&0\\ 0&0&0&0&0&0\\ \hline 0&0&1&1&0&1\\ 0&0&0&0&1&0\\ 0&0&0&0&0&0 \end{array}\right] \right\}. \tag{B.12}$$

Next, we use the $\mathbf{G}_{\mathrm{ML}}^{m'}$ block to a vertex $\mathbf{V}_{\mathrm{ML}}$. If $m' > (d-1)$, then we set $V_{\mathrm{ML}}(1|x) = 1$ for all $x \in [1, \varepsilon]$ and use the procedure in Proposition 23 to enumerate $(m'-1)$ affinely independent vertices that optimize the $\mathbf{G}_{\mathrm{ML}}^{m'}$ block in the target row. If $m' = (d-1)$, then two anti-guessing rows are selected to maximize the $\mathbf{G}_{\mathrm{A}}^{\varepsilon}$ block while the procedure in Proposition 23 is used for the remaining $(d-2)$ rows are used to construct $(m'-1)$ affinely independent vertices that optimize the $\mathbf{G}_{\mathrm{ML}}^{m'}$ block in the target row. For example, when targeting row $y = 4$ we enumerate 2 vertices,

$$\mathbf{V}_{\mathrm{ML}} \in \left\{ \left[\begin{array}{ccc|ccc} 1&1&1&0&0&0\\ 0&0&0&0&0&0\\ 0&0&0&0&0&0\\ \hline 0&0&0&1&0&1\\ 0&0&0&0&1&0\\ 0&0&0&0&0&0 \end{array}\right], \left[\begin{array}{ccc|ccc} 1&1&1&0&0&0\\ 0&0&0&0&0&0\\ 0&0&0&0&0&0\\ \hline 0&0&0&1&1&0\\ 0&0&0&0&0&0\\ 0&0&0&0&0&1 \end{array}\right] \right\}. \tag{B.13}$$

Each guessing row produces $\varepsilon + (m'-1) = (|\mathcal{Y}|-1)$ affinely independent vertices, thus in total, we have $m'(|\mathcal{Y}|-1)$ vertices enumerated for the guessing rows.

Combining the procedures for the guessing and anti-guessing rows, we construct a total of $\varepsilon(|\mathcal{Y}|-1) + m'(|\mathcal{Y}|-1) = |\mathcal{Y}|(|\mathcal{Y}|-1)$ affinely independent vertices. Therefore, we prove that $(\mathbf{G}_{\mathrm{A}}^{\varepsilon,m'}, \gamma_{\mathrm{A}}^{\varepsilon,d})$ is a tight bound. We now address the bounds on $d$ and $\varepsilon$. The lower bound $\varepsilon \geq 3$ follows from the fact that $\mathbf{G}_{\mathrm{A}}^{2} = \mathbf{G}_{\mathrm{ML}}^{2}$ meaning the anti-guessing game is indistinguishable from the maximum likelihood game. The upper bound $|\mathcal{Y}| - d + 1 \geq \varepsilon$ follows from the fact that $m' \geq (d-1)$ must be satisfied or $|\mathcal{Y}|(|\mathcal{Y}|-1)$ affinely independent vertices cannot be found because the entire diagonal of the $\mathbf{G}_{\mathrm{ML}}^{m'}$ block must be used by every vertex to satisfy $\langle \mathbf{G}_{\mathrm{A}}^{\varepsilon,m'}, \mathbf{V} \rangle = \varepsilon + d - 2$. The upper bound $|\mathcal{Y}| - 2 \geq d$ results from the lower bound on $\varepsilon$ and the fact that $d$ cannot be so large the $|\mathcal{Y}| - d + 1 < 3$. □

## *B.5 Bounds on the Signaling Dimension of a Quantum Replacer Channel, Proof of Theorem 12*

In this section we provide two propositions that support the proof of Theorem 12. Recall that a $d$-dimensional partial replacer channel is a quantum channel having the form

$$\mathcal{R}_{\mu}(X) = \mu X + (1-\mu)\mathrm{Tr}\left[X\right]\sigma, \tag{B.14}$$

where $1 \geq \mu \geq 0$, $\sigma$ is some fixed density matrix, and $X$ is a quantum state on a $d$-dimensional Hilbert space. Note that the partial erasure channel $\mathcal{E}_{\mu}$ corresponds to $\sigma$ being an erasure flag $|E\rangle\langle E|$, where $|E\rangle$

is orthogonal to $\{|1\rangle, \cdots, |d\rangle\}$. We first show that the lower bound of $\kappa_\Lambda(\mathcal{R}_\mu) \geq \lceil \mu d + (1-\mu) \rceil$ (see Eq. (14.52)) is not improved by any choice of states $\{\rho_x\}_{x\in\mathcal{X}}$, POVM $\{\Pi_y\}_{y\in\mathcal{Y}}$, or ambiguous guessing game $\mathbf{G}_{k=|\mathcal{Y}|}^{\mathcal{X}\overset{d}{\to}\mathcal{Y}}$ as defined in Eq. (14.35).

**Proposition 28.** [114] The maximum likelihood score for any classical channel $\mathbf{P}_{\mathcal{R}_\mu} \in \mathcal{Q}^{\mathcal{X}\to\mathcal{Y}}(\mathcal{R}_\mu)$ generated using a partial replacer channel $\mathcal{R}_\mu$ is bounded as

$$\langle \mathbf{G}_{\text{ML}}, \mathbf{P}_{\mathcal{R}_\mu} \rangle \leq \mu d + (1-\mu) \tag{B.15}$$

where $\mathbf{G}_{\text{ML}}$ is any maximum likelihood facet satisfying Proposition **??**(i).

*Proof.* In this proof, we first consider the unlifted maximum likelihood $\mathbf{G}_{\text{ML}}^{|\mathcal{Y}|} = \mathbb{I}_{|\mathcal{Y}|}$ where $|\mathcal{X}| = |\mathcal{Y}|$ (see Section 14.3.4), and then generalize across all input/output liftings taking $\mathbf{G}_{\text{ML}}^{|\mathcal{Y}|} \to \mathbf{G}_{\text{ML}} \in \mathbb{R}^{m'\times m}$ where $m', m \geq |\mathcal{Y}|$. To begin we maximize $\langle \mathbf{G}_{\text{ML}}^{|\mathcal{Y}|}, \mathbf{P}_{\mathcal{R}_\mu} \rangle$ over the quantum states $\{\rho_x\}_{x\in\mathcal{X}}$ and POVM $\{\Pi_y\}_{y\in\mathcal{Y}}$,

$$\begin{aligned}
\max \langle \mathbf{G}_{\text{ML}}^{|\mathcal{Y}|}, \mathbf{P}_{\mathcal{R}_\mu} \rangle &= \max_{\{\rho_x\}_x, \{\Pi_y\}_y} \sum_{x=y} \text{Tr}\left[\Pi_y \mathcal{R}_\mu(\rho_x)\right] && \text{(B.16)}\\
&= \max_{\{\rho_x\}_x, \{\Pi_y\}_y} \sum_{x=y} \mu \text{Tr}\left[\Pi_y \rho_x\right] + (1-\mu)\text{Tr}\left[\Pi_y \sigma\right] && \text{(B.17)}\\
&\leq \max_{\{\Pi_y\}_y} \sum_{y} \mu \text{Tr}\left[\Pi_y\right] + (1-\mu)\text{Tr}\left[\Pi_y \sigma\right] && \text{(B.18)}\\
&= \mu d + (1-\mu), && \text{(B.19)}
\end{aligned}$$

where line (B.18) uses the fact that $\text{Tr}\left[\Pi_y \rho_x\right] \leq \text{Tr}\left[\Pi_y\right]$ for any choice of $\Pi_y$ and $\rho_x$ while line (B.19) results from $\sum_y \text{Tr}\left[\Pi_y\right] = d$ and $\sum_y \text{Tr}\left[\Pi_y \sigma\right] = \text{Tr}\left[\sigma\right] = 1$. A simple example that achieves this bound is the scenario where Alice sends orthogonal states $\{|x\rangle\langle x|\}_{x=1}^d$ and Bob measures with a similar POVM $\{|y\rangle\langle y|\}_{y=1}^d$, then

$$\begin{aligned}
\langle \mathbf{G}_{\text{ML}}^{d}, \mathbf{P}_{\mathcal{R}_\mu} \rangle &= \sum_{y=x=1}^{d} \text{Tr}\left[|y\rangle\langle y| \, \mathcal{R}_\mu\big(\, |x\rangle\langle x| \,\big)\right] && \text{(B.20)}\\
&= \mu \sum_{y=1}^{d} \text{Tr}\left[|y\rangle \langle y\rangle \, y \, \langle y|\right] + (1-\mu) \sum_{y=1}^{d} \text{Tr}\left[|y\rangle\langle y| \, \sigma\right] && \text{(B.21)}\\
&= \mu d + (1-\mu). && \text{(B.22)}
\end{aligned}$$

In general, the upper bound is achieved whenever $\Pi_y \rho_x = \Pi_y$ for all $x \in \mathcal{X}$ and $y \in \mathcal{Y}$. Note that this requires $\text{Rank}\left(\Pi_y\right) = \text{Rank}\left(\rho_x\right) = 1$ and $\Pi_y || \rho_x$.

To extend the bound $\langle \mathbf{G}_{\text{ML}}^{|\mathcal{Y}|}, \mathbf{P}_{\mathcal{R}_\mu} \rangle \leq \mu d + (1-\mu)$ to all liftings of $\mathbf{G}_{\text{ML}}^{|\mathcal{Y}|}$, we make two observations. First, note that the input lifting taking $\mathbf{G}_{\text{ML}}^{|\mathcal{Y}|} \to \mathbf{G}''_{\text{ML}} \in \mathbb{R}^{|\mathcal{Y}|\times m}$ contains $(m - |\mathcal{X}|)$ all-zero columns. These all-zero columns of $\mathbf{G}''_{\text{ML}}$ do not contribute to the inner product $\langle \mathbf{G}''_{\text{ML}}, \mathbf{P}_{\mathcal{R}_\mu} \rangle$, and therefore, cannot increase the inner product beyond $\mu d + (1-\mu)$. Second, observe that the output lifting taking $\mathbf{G}_{\text{ML}}^{|\mathcal{Y}|} \to \mathbf{G}'_{\text{ML}} \in \mathbb{R}^{(|\mathcal{Y}|+1)\times|\mathcal{X}|}$ requires a new POVM $\{\Pi'_y\}_{y=1}^{|\mathcal{Y}|+1}$ which must satisfy $\sum_{y=1}^{|\mathcal{Y}|+1} \Pi'_y = \mathbb{I}_d$. Furthermore, one column $x$ of $\mathbf{G}'_{\text{ML}}$ has two nonzero elements in rows $y$ and $y'$ where $G'_{y,x} = G'_{y',x} = 1$. In this case, two POVM elements $\Pi'_y$ and $\Pi'_{y'}$ are both optimized against the state $\rho_x$. However, the constraint $\text{Tr}\left[\Pi'_{y'}\rho_x\right] + \text{Tr}\left[\Pi'_y \rho_x\right] \leq 1$ holds for any choice of $\rho_x$ and POVM. Therefore, the inner product $\langle \mathbf{G}'_{\text{ML}}, \mathbf{P}_{\mathcal{R}_\mu} \rangle \leq \mu d + (1-\mu)$. The argument applied

for the output lifting holds in general where one or more columns $x$ contain at least two non-zero elements. Thus, the upper bound in Eq. (B.19) holds for any input/output lifting taking $\mathbf{G}_{\text{ML}}^{|\mathcal{Y}|} \to \mathbf{G}_{\text{ML}} \in \mathbb{R}^{m' \times m}$ where $\min\{m, m'\} \geq |\mathcal{Y}|$. This concludes the proof. □

The upper bound on the maximum likelihood score from Proposition 28 serves as a lower bound on the signaling dimension of the partial replacer channel $\kappa_\Lambda(\mathcal{R}_\mu)$. This follows from the fact that if $\mathbf{P}_{\mathcal{R}_\mu} \notin \mathcal{M}_r^{\mathcal{X}\to\mathcal{Y}}$, then $\kappa_\Lambda(\mathcal{R}_\mu) > r$ Furthermore, the integer nature of the signaling dimension implies that $\kappa_\Lambda(\mathcal{R}_\mu) \geq \lceil \mu d + (1-\mu) \rceil$. We now turn to certify the signaling dimension of the partial erasure channel.

**Proposition 29.** [114] The signaling dimension of of a $d$-dimensional partial erasure channel is,

$$\kappa_\Lambda(\mathcal{E}_\mu) = \min\{d, \lceil \mu d + 1 \rceil\}. \tag{B.23}$$

*Proof.* Let the classical channel $\mathbf{P}_{\mathcal{E}_\mu}$ be induced by the partial erasure channel $\mathcal{E}_\mu$ via Eq. (14.10) for any collection of quantum states $\{\rho_x\}_x$ and POVM $\{\Pi_y\}_y$. The transition probabilities are then expressed

$$P_{\mathcal{E}_\mu}(y|x) = \mu P_{\text{id}_d}(y|x) + (1-\mu) P_{|E\rangle}(y), \tag{B.24}$$

where $P_{\text{id}_d}(y|x) = \text{Tr}\left[\Pi_y \rho_x\right]$ and $P_{|E\rangle}(y) = \text{Tr}\left[\Pi_y |E\rangle\langle E|\right]$. Since the simulation protocol for partial replacer channels can faithfully simulate $\mathbf{P}_{\mathcal{E}_\mu}$, the upper bound $\kappa_\Lambda(\mathcal{R}_\mu) \leq \lceil \mu d + 1 \rceil$ holds (see the proof of Theorem 12 in the main text). Therefore, $\min\{d, \lceil \mu d + 1 \rceil\} \geq \kappa_\Lambda(\mathcal{E}_\mu)$. To establish a lower bound on $\kappa_\Lambda(\mathcal{E}_\mu)$ we consider the channel $\mathbf{P}_{\mathcal{E}_\mu} \in \mathcal{P}^{d\to(d+1)}$ generated by the scenario where Alice sends the computational basis states $\{|x\rangle\langle x|\}_{i=1}^d$ and Bob measures with the POVM $\{|y\rangle\langle y|\}_{y=1}^{d+1}$ where $|d+1\rangle = |E\rangle$,

$$\begin{aligned}
\mathbf{P}_{\mathcal{E}_\mu} &= \sum_{x=1}^{d} \sum_{y=1}^{d+1} \text{Tr}\left[|y\rangle\langle y| \, \mathcal{E}_\mu(|x\rangle\langle x|)\right] |y\rangle\langle x| && \text{(B.25)} \\
&= \sum_{x=1}^{d} \sum_{y=1}^{d+1} \left(\mu \text{Tr}\left[|y\rangle \langle y\rangle \, x \, \langle x|\right] + (1-\mu) \text{Tr}\left[|y\rangle \langle y\rangle \, E \, \langle E|\right]\right) |y\rangle\langle x| && \text{(B.26)} \\
&= \mu \sum_{x=1}^{d} |x\rangle\langle x| + (1-\mu) \sum_{x=1}^{d} |E\rangle\langle x| \, . && \text{(B.27)}
\end{aligned}$$

As demonstrated in Proposition 28, $\mathbf{P}_{\mathcal{E}_\mu}$ achieves the maximum likelihood upper bound for partial replacer channels, $\langle \mathbf{G}_{\text{ML}}, \mathbf{P}_{\mathcal{E}_\mu} \rangle = \mu d + (1-\mu)$. In fact, this bound also holds for non-orthogonal quantum states $\{\rho_x\}_{x\in\mathcal{X}}$ where $|\mathcal{X}| > d$.

To improve the lower bound on $\kappa_\Lambda(\mathcal{E}_\mu)$ beyond Proposition 28, we consider the ambiguous polytope $\mathcal{A}_{(|\mathcal{Y}|-1),r}^{(|\mathcal{Y}|-1)\to|\mathcal{Y}|}$ with ambiguous guessing facets $\mathbf{G}_?^{|\mathcal{Y}|,r}$ that are tight signaling dimension witnesses of $\mathcal{C}_r^{\mathcal{X}\to\mathcal{Y}}$ (see Appendix **??**). Our goal is to find the smallest integer $r$ such that $\mathbf{P}_{\mathcal{E}_\mu} \in \mathcal{A}_{(|\mathcal{Y}|-1),r}^{(|\mathcal{Y}|-1)\to|\mathcal{Y}|}$, that is, $\langle \mathbf{G}_?^{|\mathcal{Y}|,r}, \mathbf{P}_{\mathcal{E}_\mu} \rangle \leq r(|\mathcal{Y}| - r)$ is satisfied. Consider the erasure channel $\mathbf{P}_{\mathcal{E}_\mu} \in \mathcal{P}^{d\to(d+1)}$ described by Eq. (B.27). We find that the inequality

$$r(|\mathcal{Y}| - r) \geq \langle \mathbf{G}_?^{|\mathcal{Y}|,r}, \mathbf{P}_{\mathcal{E}_\mu} \rangle = (|\mathcal{Y}| - r)\mu d + (1-\mu)(|\mathcal{Y}| - 1), \tag{B.28}$$

is violated if $\mathbf{P}_{\mathcal{E}_\mu} \notin \mathcal{A}^{(|\mathcal{Y}|-1)\to|\mathcal{Y}|}_{(|\mathcal{Y}|-1),r}$ for $|\mathcal{Y}|-2 \geq r \geq 2$. Note that in our example $|\mathcal{Y}| = d+1$, however, this procedure holds for any $|\mathcal{Y}| > d$. Rearranging inequality (B.28) into the form,

$$0 \geq r^2 - r(\mu d + |\mathcal{Y}|) + \mu d|\mathcal{Y}| + (1-\mu)(|\mathcal{Y}|-1), \tag{B.29}$$

allows us to find the values of $r$ for which inequality (B.28) is satisfied by solving for the zeros $r_\pm$ of the quadratic on the RHS of Eq. (B.29),

$$r_\pm = \frac{1}{2}(\mu d + |\mathcal{Y}|) \pm \frac{1}{2}\sqrt{(|\mathcal{Y}| - \mu d)^2 - 4(1-\mu)(|\mathcal{Y}|-1)}. \tag{B.30}$$

Since the parabola of Eq. (B.29) is concave up, all integer values of $r \in [r_-, r_+]$ satisfy inequality (B.28). Furthermore, the smallest integer for which the inequality is satisfied is $r = \lceil r_- \rceil$. Therefore, the signaling dimension is bounded as

$$\kappa_\Lambda(\mathcal{E}_\mu) \geq r = \left\lceil \frac{1}{2}(\mu d + |\mathcal{Y}|) - \frac{1}{2}\sqrt{|\mathcal{Y}| - \mu d)^2 - 4(1-\mu)|\mathcal{Y}| - 1)} \right\rceil. \tag{B.31}$$

The value of $r$ in Eq. (B.31) satisfies the facet inequality (B.28) for all allowed values of $|\mathcal{Y}|$, $\mu$, and $d$. Note that $|\mathcal{Y}|$ is a free parameter which we can choose as any integer $|\mathcal{Y}| \geq d+1$. In our example, $\mathbf{P}_{\mathcal{E}_\mu}$ has $|\mathcal{Y}| = (d+1)$ which obtains the lower bound $\kappa_\Lambda(\mathcal{E}_\mu) \geq \lceil \mu d + 1 \rceil$. To see this, we substitute $|\mathcal{Y}| = (d+1)$ into Eq. (B.31) and perform some algebra,

$$\begin{aligned}
r &= \left\lceil \frac{1}{2}\Big(\mu d + d + 1\Big) \quad - \frac{1}{2}\sqrt{\Big(d(1-\mu)+1\Big)^2 - 4d(1-\mu)} \right\rceil \\
&= \left\lceil \frac{1}{2}\Big(\mu d + d + 1\Big) - \frac{1}{2}\sqrt{\Big(d(1-\mu)-1\Big)^2} \right\rceil && \text{(B.32)} \\
&= \left\lceil \frac{1}{2}\Big(\mu d + d + 1\Big) - \frac{1}{2}\Big(d(1-\mu)-1\Big) \right\rceil && \text{(B.33)} \\
&= \left\lceil \frac{1}{2}(\mu d + 1) + \frac{1}{2}(\mu d + 1) \right\rceil && \text{(B.34)} \\
&= \lceil \mu d + 1 \rceil. && \text{(B.35)}
\end{aligned}$$

Hence $\kappa_\Lambda(\mathcal{E}_\mu) \geq r = \lceil \mu d + 1 \rceil$. Additionally, substituting $n' > d+1$ into Eq. (B.31) results in a necessarily smaller value or $r$ therefore $|\mathcal{Y}| = d+1$ is a maximum.

It is important to note that the lower bound $\kappa_\Lambda(\mathcal{E}_\mu) \geq \lceil \mu d + 1 \rceil$ only holds for $r \leq |\mathcal{Y}| - 2$ because $\mathbf{G}_?^{|\mathcal{Y}|,r}$ is not a facet for signaling polytopes $\mathcal{C}_\Lambda^{\mathcal{X}\overset{d}{\to}\mathcal{Y}}$ with $d = r > |\mathcal{Y}| - 2$ and $|\mathcal{X}| = |\mathcal{Y}| - 1$. Therefore, we must consider the edge case where $r = |\mathcal{Y}| - 1 = |\mathcal{X}|$, that is, the case where the trivial upper bound of Eq. (14.20) is obtained. From Theorem 10 Condition (ii) we know that $\mathcal{C}_\Lambda^{\mathcal{X}\overset{|\mathcal{X}|-1}{\to}\mathcal{Y}} = \cap k = |\mathcal{X}|^{|\mathcal{Y}|}\mathcal{A}^{\mathcal{X}\to\mathcal{Y}}_{k,|\mathcal{X}|-1}$. It follows for the edge case $r = |\mathcal{Y}| - 1 = |\mathcal{X}|$ that if a channel $\mathbf{P}_{\mathcal{E}_\mu} \notin \mathcal{A}^{(|\mathcal{Y}|-1)\to|\mathcal{Y}|}_{(|\mathcal{Y}|-1),r}$, then $\kappa_\Lambda^{\mathcal{X}\to\mathcal{Y}}(\mathcal{E}_\mu) = \min\{|\mathcal{X}|, |\mathcal{Y}|\} = (|\mathcal{Y}| - 1)$. Hence $\kappa_\Lambda(\mathcal{E}_\mu)$ is proven to be tight with the upper bound. To illustrate this case we consider inequality (B.29) and substitute $r = |\mathcal{Y}| - 2$,

$$0 \geq (|\mathcal{Y}|^2 - 4|\mathcal{Y}| + 4) - (|\mathcal{Y}| - 2)(\mu d + |\mathcal{Y}|) + \mu d|\mathcal{Y}| + (1-\mu)(|\mathcal{Y}| - 1) \tag{B.36}$$

$$\geq -4|\mathcal{Y}| + 4 + 2\mu d + 2|\mathcal{Y}| + (1-\mu)|\mathcal{Y}| - (1-\mu). \tag{B.37}$$

Next, we substitute $d = |\mathcal{Y}| - 1$ into Eq. (B.37) as this is the edge case we wish to consider,

$$0 \geq -4|\mathcal{Y}| + 4 + 2\mu(|\mathcal{Y}| - 1) + 2|\mathcal{Y}| + (1-\mu)|\mathcal{Y}| - (1-\mu) \tag{B.38}$$

$$\geq |\mathcal{Y}|(\mu - 1) + 3 - \mu \tag{B.39}$$

$$\geq 3 - |\mathcal{Y}| + \mu(|\mathcal{Y}| - 1). \tag{B.40}$$

Rearranging inequality (B.40), we find that it is satisfied *iff*, $\frac{|\mathcal{Y}|-3}{|\mathcal{Y}|-1} \geq \mu$. Therefore, when $\mu > \frac{|\mathcal{Y}|-3}{|\mathcal{Y}|-1}$, inequality (B.28) is violated and, by Theorem 10(ii) we certify that $\kappa_\Lambda(\mathcal{E}_\mu) = d$. Considering this edge case, we arrive at the conclusion that $\kappa_\Lambda(\mathcal{E}_\mu) \geq \min\{d, \lceil \mu d + 1 \rceil\}$ which is exactly the upper bound $\min\{d, \lceil \mu d + 1 \rceil\} \geq \kappa_\Lambda(\mathcal{E}_\mu)$. That is, the signaling dimension of the erasure channel is bounded tightly from above and below from which it follows, $\kappa_\Lambda(\mathcal{E}_\mu) = \min\{d, \lceil \mu d + 1 \rceil\}$. □

# References


[1] U. S. Code, *National quantum initiative act*, https://www.congress.gov/bill/115th-congress/house-bill/6227, 2018.

[2] Q. Zhang, F. Xu, L. Li, N.-L. Liu, and J.-W. Pan, "Quantum information research in china," *Quantum Science and Technology*, vol. 4, no. 4, p. 040 503, Nov. 2019. DOI: 10.1088/2058-9565/ab4bea. [Online]. Available: https://doi.org/10.1088/2058-9565/ab4bea.

[3] B. Sussman, P. Corkum, A. Blais, D. Cory, and A. Damascelli, "Quantum canada," *Quantum Science and Technology*, vol. 4, no. 2, p. 020 503, Feb. 2019. DOI: 10.1088/2058-9565/ab029d. [Online]. Available: https://doi.org/10.1088/2058-9565/ab029d.

[4] Y. Yamamoto, M. Sasaki, and H. Takesue, "Quantum information science and technology in japan," *Quantum Science and Technology*, vol. 4, no. 2, p. 020 502, Feb. 2019. DOI: 10.1088/2058-9565/ab0077. [Online]. Available: https://doi.org/10.1088/2058-9565/ab0077.

[5] T. M. Roberson and A. G. White, "Charting the australian quantum landscape," *Quantum Science and Technology*, vol. 4, no. 2, p. 020 505, Feb. 2019. DOI: 10.1088/2058-9565/ab02b4. [Online]. Available: https://doi.org/10.1088/2058-9565/ab02b4.

[6] P. Knight and I. Walmsley, "UK national quantum technology programme," *Quantum Science and Technology*, vol. 4, no. 4, p. 040 502, Oct. 2019. DOI: 10.1088/2058-9565/ab4346. [Online]. Available: https://doi.org/10.1088/2058-9565/ab4346.

[7] M. Riedel, M. Kovacs, P. Zoller, J. Mlynek, and T. Calarco, "Europe's quantum flagship initiative," *Quantum Science and Technology*, vol. 4, no. 2, p. 020 501, Feb. 2019. DOI: 10.1088/2058-9565/ab042d. [Online]. Available: https://doi.org/10.1088/2058-9565/ab042d.

[8] E. Gibney, "Quantum gold rush: The private funding pouring into quantum start-ups," *Nature*, vol. 574, no. 7776, pp. 22–24, Oct. 2019. DOI: 10.1038/d41586-019-02935-4. [Online]. Available: https://doi.org/10.1038/d41586-019-02935-4.

[9] R. de Wolf, "The potential impact of quantum computers on society," *Ethics and Information Technology*, vol. 19, no. 4, pp. 271–276, Sep. 2017. DOI: 10.1007/s10676-017-9439-z. [Online]. Available: https://doi.org/10.1007/s10676-017-9439-z.

[10] NobelPrize.org, *The nobel prize in physics 2022*, N. P. O. A. 2023, Ed. [Online]. Available: https://www.nobelprize.org/prizes/physics/2022/summary/.

[11] A. Einstein, B. Podolsky, and N. Rosen, "Can quantum-mechanical description of physical reality be considered complete?" *Physical review*, vol. 47, no. 10, p. 777, 1935.

[12] J. S. Bell, "On the einstein podolsky rosen paradox," *Physics Physique Fizika*, vol. 1, no. 3, p. 195, 1964.

[13] J. F. Clauser, M. A. Horne, A. Shimony, and R. A. Holt, "Proposed experiment to test local hidden-variable theories," *Phys. Rev. Lett.*, vol. 23, pp. 880–884, 15 Oct. 1969. DOI: 10.1103/PhysRevLett.23.880. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevLett.23.880.

[14] A. Aspect, P. Grangier, and G. Roger, "Experimental tests of realistic local theories via bell's theorem," *Phys. Rev. Lett.*, vol. 47, pp. 460–463, 7 Aug. 1981. DOI: 10.1103/PhysRevLett.47.460. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevLett.47.460.

[15] B. Hensen, H. Bernien, A. E. Dréau, A. Reiserer, N. Kalb, M. S. Blok, J. Ruitenberg, R. F. L. Vermeulen, R. N. Schouten, C. Abellán, W. Amaya, V. Pruneri, M. W. Mitchell, M. Markham, D. J. Twitchen, D. Elkouss, S. Wehner, T. H. Taminiau, and R. Hanson, "Loophole-free bell inequality violation using electron spins separated by 1.3 kilometres," *Nature*, vol. 526, no. 7575, pp. 682–686, Oct. 2015. DOI: 10.1038/nature15759. [Online]. Available: https://doi.org/10.1038/nature15759.

[16] L. K. Shalm, E. Meyer-Scott, B. G. Christensen, P. Bierhorst, M. A. Wayne, M. J. Stevens, T. Gerrits, S. Glancy, D. R. Hamel, M. S. Allman, K. J. Coakley, S. D. Dyer, C. Hodge, A. E. Lita, V. B. Verma, C. Lambrocco, E. Tortorici, A. L. Migdall, Y. Zhang, D. R. Kumor, W. H. Farr, F. Marsili, M. D. Shaw, J. A. Stern, C. Abellán, W. Amaya, V. Pruneri, T. Jennewein, M. W. Mitchell, P. G. Kwiat, J. C. Bienfang, R. P. Mirin, E. Knill, and S. W. Nam, "Strong loophole-free test of local realism," *Phys. Rev. Lett.*, vol. 115, p. 250 402, 25 Dec. 2015. DOI: 10.1103/PhysRevLett.115.250402. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevLett.115.250402.

[17] M. Giustina, M. A. M. Versteegh, S. Wengerowsky, J. Handsteiner, A. Hochrainer, K. Phelan, F. Steinlechner, J. Kofler, J.-Å. Larsson, C. Abellán, W. Amaya, V. Pruneri, M. W. Mitchell, J. Beyer, T. Gerrits, A. E. Lita, L. K. Shalm, S. W. Nam, T. Scheidl, R. Ursin, B. Wittmann, and A. Zeilinger, "Significant-loophole-free test of bell's theorem with entangled photons," *Phys. Rev. Lett.*, vol. 115, p. 250 401, 25 Dec. 2015. DOI: 10.1103/PhysRevLett.115.250401. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevLett.115.250401.

[18] R. P. Feynman *et al.*, "Simulating physics with computers," *Int. j. Theor. phys*, vol. 21, no. 6/7, 1982.

[19] S. Lloyd, "Universal quantum simulators," *Science*, vol. 273, no. 5278, pp. 1073–1078, Aug. 1996. DOI: 10.1126/science.273.5278.1073. [Online]. Available: https://doi.org/10.1126/science.273.5278.1073.

[20] I. Buluta and F. Nori, "Quantum simulators," *Science*, vol. 326, no. 5949, pp. 108–111, Oct. 2009. DOI: 10.1126/science.1177838. [Online]. Available: https://doi.org/10.1126/science.1177838.

[21] K. L. Brown, W. J. Munro, and V. M. Kendon, "Using quantum computers for quantum simulation," *Entropy*, vol. 12, no. 11, pp. 2268–2307, Nov. 2010. DOI: 10.3390/e12112268. [Online]. Available: https://doi.org/10.3390/e12112268.

[22] I. M. Georgescu, S. Ashhab, and F. Nori, "Quantum simulation," *Rev. Mod. Phys.*, vol. 86, pp. 153–185, 1 Mar. 2014. DOI: 10.1103/RevModPhys.86.153. [Online]. Available: https://link.aps.org/doi/10.1103/RevModPhys.86.153.

[23] Y. Cao, J. Romero, and A. Aspuru-Guzik, "Potential of quantum computing for drug discovery," *IBM Journal of Research and Development*, vol. 62, no. 6, 6:1–6:20, 2018. DOI: 10.1147/JRD.2018.2888987.

[24] B. Bauer, S. Bravyi, M. Motta, and G. K.-L. Chan, "Quantum algorithms for quantum chemistry and quantum materials science," *Chemical Reviews*, vol. 120, no. 22, pp. 12 685–12 717, Oct. 2020. DOI: 10.1021/acs.chemrev.9b00829. [Online]. Available: https://doi.org/10.1021/acs.chemrev.9b00829.

[25] A. Montanaro, "Quantum algorithms: An overview," *npj Quantum Information*, vol. 2, no. 1, Jan. 2016. DOI: 10.1038/npjqi.2015.23. [Online]. Available: https://doi.org/10.1038/npjqi.2015.23.

[26] A. Ambainis, "Quantum search algorithms," *ACM SIGACT News*, vol. 35, no. 2, pp. 22–35, 2004.

[27] P. W. Shor, "Polynomial-time algorithms for prime factorization and discrete logarithms on a quantum computer," *SIAM review*, vol. 41, no. 2, pp. 303–332, 1999.

[28] V. Giovannetti, S. Lloyd, and L. Maccone, "Advances in quantum metrology," *Nature Photonics*, vol. 5, no. 4, pp. 222–229, Mar. 2011. DOI: 10.1038/nphoton.2011.35. [Online]. Available: https://doi.org/10.1038/nphoton.2011.35.

[29] J. A. et al., "Enhanced sensitivity of the LIGO gravitational wave detector by using squeezed states of light," *Nature Photonics*, vol. 7, no. 8, pp. 613–619, Jul. 2013. DOI: 10.1038/nphoton.2013.177. [Online]. Available: https://doi.org/10.1038/nphoton.2013.177.

[30] A. Kellerer, "Quantum telescopes," *Astronomy and Geophysics*, vol. 55, no. 3, pp. 3.28–3.32, Jun. 2014. DOI: 10.1093/astrogeo/atu126. [Online]. Available: https://doi.org/10.1093/astrogeo/atu126.

[31] A. R. Kurek, T. Pieta, T. Stebel, A. Pollo, and A. Popowicz, "Quantum telescope: Feasibility and constraints," *Opt. Lett.*, vol. 41, no. 6, pp. 1094–1097, Mar. 2016. DOI: 10.1364/OL.41.001094. [Online]. Available: https://opg.optica.org/ol/abstract.cfm?URI=ol-41-6-1094.

[32] P. Kómár, E. M. Kessler, M. Bishof, L. Jiang, A. S. Sørensen, J. Ye, and M. D. Lukin, "A quantum network of clocks," *Nature Physics*, vol. 10, no. 8, pp. 582–587, Jun. 2014. DOI: 10.1038/nphys3000. [Online]. Available: https://doi.org/10.1038/nphys3000.

[33] M. A. Taylor and W. P. Bowen, "Quantum metrology and its application in biology," *Physics Reports*, vol. 615, pp. 1–59, Feb. 2016. DOI: 10.1016/j.physrep.2015.12.002. [Online]. Available: https://doi.org/10.1016/j.physrep.2015.12.002.

[34] N. Aslam, H. Zhou, E. K. Urbach, M. J. Turner, R. L. Walsworth, M. D. Lukin, and H. Park, "Quantum sensors for biomedical applications," *Nature Reviews Physics*, vol. 5, no. 3, pp. 157–169, Feb. 2023. DOI: 10.1038/s42254-023-00558-3. [Online]. Available: https://doi.org/10.1038/s42254-023-00558-3.

[35] Q.-C. Sun, Y.-L. Mao, S.-J. Chen, W. Zhang, Y.-F. Jiang, Y.-B. Zhang, W.-J. Zhang, S. Miki, T. Yamashita, H. Terai, X. Jiang, T.-Y. Chen, L.-X. You, X.-F. Chen, Z. Wang, J.-Y. Fan, Q. Zhang, and J.-W. Pan, "Quantum teleportation with independent sources and prior entanglement distribution over a network," *Nature Photonics*, vol. 10, no. 10, pp. 671–675, Sep. 2016. DOI: 10.1038/nphoton.2016.179. [Online]. Available: https://doi.org/10.1038/nphoton.2016.179.

[36] D. Cozzolino, B. D. Lio, D. Bacco, and L. K. Oxenløwe, "High-dimensional quantum communication: Benefits, progress, and future challenges," *Advanced Quantum Technologies*, vol. 2, no. 12, p. 1 900 038, Oct. 2019. DOI: 10.1002/qute.201900038. [Online]. Available: https://doi.org/10.1002/qute.201900038.

[37] S. Herbert, "Increasing the classical data throughput in quantum networks by combining quantum linear network coding with superdense coding," *Phys. Rev. A*, vol. 101, p. 062 332, 6 Jun. 2020. DOI: 10.1103/PhysRevA.101.062332. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevA.101.062332.

[38] J. Preskill, "Quantum Computing in the NISQ era and beyond," *Quantum*, vol. 2, p. 79, Aug. 2018, ISSN: 2521-327X. DOI: 10.22331/q-2018-08-06-79. [Online]. Available: https://doi.org/10.22331/q-2018-08-06-79.

[39] R. Van Meter and S. J. Devitt, "The path to scalable distributed quantum computing," *Computer*, vol. 49, no. 9, pp. 31–42, 2016. DOI: 10.1109/MC.2016.291.

[40] F. A. et al., "Quantum supremacy using a programmable superconducting processor," *Nature*, vol. 574, no. 7779, pp. 505–510, Oct. 2019. DOI: 10.1038/s41586-019-1666-5. [Online]. Available: https://doi.org/10.1038/s41586-019-1666-5.

[41] F. Pan, K. Chen, and P. Zhang, "Solving the sampling problem of the sycamore quantum circuits," *Phys. Rev. Lett.*, vol. 129, p. 090 502, 9 Aug. 2022. DOI: 10.1103/PhysRevLett.129.090502. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevLett.129.090502.

[42] C. Elliott, A. Colvin, D. Pearson, O. Pikalo, J. Schlafer, and H. Yeh, "Current status of the DARPA quantum network (invited paper)," in *SPIE Proceedings*, E. J. Donkor, A. R. Pirich, and H. E. Brandt, Eds., SPIE, May 2005. DOI: 10.1117/12.606489. [Online]. Available: https://doi.org/10.1117/12.606489.

[43] M. Peev, C. Pacher, R. Alléaume, C. Barreiro, J. Bouda, W. Boxleitner, T. Debuisschert, E. Diamanti, M. Dianati, J. Dynes, *et al.*, "The secoqc quantum key distribution network in vienna," *New Journal of Physics*, vol. 11, no. 7, p. 075 001, 2009.

[44] M. Sasaki, M. Fujiwara, H. Ishizuka, W. Klaus, K. Wakui, M. Takeoka, S. Miki, T. Yamashita, Z. Wang, A. Tanaka, *et al.*, "Field test of quantum key distribution in the tokyo qkd network," *Optics express*, vol. 19, no. 11, pp. 10 387–10 409, 2011.

[45] S.-K. Liao, W.-Q. Cai, J. Handsteiner, B. Liu, J. Yin, L. Zhang, D. Rauch, M. Fink, J.-G. Ren, W.-Y. Liu, *et al.*, "Satellite-relayed intercontinental quantum network," *Physical review letters*, vol. 120, no. 3, p. 030 501, 2018.

[46] S. Wengerowsky, S. K. Joshi, F. Steinlechner, H. Hübel, and R. Ursin, "An entanglement-based wavelength-multiplexed quantum communication network," *Nature*, vol. 564, no. 7735, pp. 225–228, Dec. 2018. DOI: 10.1038/s41586-018-0766-y. [Online]. Available: https://doi.org/10.1038/s41586-018-0766-y.

[47] J. F. Dynes, A. Wonfor, W. W. .-S. Tam, A. W. Sharpe, R. Takahashi, M. Lucamarini, A. Plews, Z. L. Yuan, A. R. Dixon, J. Cho, Y. Tanizawa, J. .-P. Elbers, H. Greißer, I. H. White, R. V. Penty, and A. J. Shields, "Cambridge quantum network," *npj Quantum Information*, vol. 5, no. 1, Nov. 2019. DOI: 10.1038/s41534-019-0221-4. [Online]. Available: https://doi.org/10.1038/s41534-019-0221-4.

[48] Y.-A. Chen, Q. Zhang, T.-Y. Chen, W.-Q. Cai, S.-K. Liao, J. Zhang, K. Chen, J. Yin, J.-G. Ren, Z. Chen, S.-L. Han, Q. Yu, K. Liang, F. Zhou, X. Yuan, M.-S. Zhao, T.-Y. Wang, X. Jiang, L. Zhang, W.-Y. Liu, Y. Li, Q. Shen, Y. Cao, C.-Y. Lu, R. Shu, J.-Y. Wang, L. Li, N.-L. Liu, F. Xu, X.-B. Wang, C.-Z. Peng, and J.-W. Pan, "An integrated space-to-ground quantum communication network over 4, 600 kilometres," *Nature*, vol. 589, no. 7841, pp. 214–219, Jan. 2021. DOI: 10.1038/s41586-020-03093-8. [Online]. Available: https://doi.org/10.1038/s41586-020-03093-8.

[49] J.-H. Kim, J.-W. Chae, Y.-C. Jeong, and Y.-H. Kim, "Quantum communication with time-bin entanglement over a wavelength-multiplexed fiber network," *APL Photonics*, vol. 7, no. 1, p. 016 106, 2022. DOI: 10.1063/5.0073040. [Online]. Available: https://doi.org/10.1063/5.0073040.

[50] V. Scarani and C. Kurtsiefer, "The black paper of quantum cryptography: Real implementation problems," *Theoretical Computer Science*, vol. 560, pp. 27–32, Dec. 2014. DOI: 10.1016/j.tcs.2014.09.015. [Online]. Available: https://doi.org/10.1016/j.tcs.2014.09.015.

[51] N. S. Agency, *Quantum key distribution (qkd) and quantum cryptography qc*. [Online]. Available: https://www.nsa.gov/Cybersecurity/Quantum-Key-Distribution-QKD-and-Quantum-Cryptography-QC/.

[52] S. Wehner, D. Elkouss, and R. Hanson, "Quantum internet: A vision for the road ahead," *Science*, vol. 362, no. 6412, 2018. DOI: 10.1126/science.aam9288.

[53] H. J. Kimble, "The quantum internet," *Nature*, vol. 453, no. 7198, pp. 1023–1030, Jun. 2008. DOI: 10.1038/nature07127.

[54] W. Kozlowski and S. Wehner, "Towards large-scale quantum networks," in *Proceedings of the Sixth Annual ACM International Conference on Nanoscale Computing and Communication*, 2019, pp. 1–7.

[55] A. S. Cacciapuoti, M. Caleffi, F. Tafuri, F. S. Cataliotti, S. Gherardini, and G. Bianchi, "Quantum internet: Networking challenges in distributed quantum computing," *IEEE Network*, vol. 34, no. 1, pp. 137–143, 2020. DOI: 10.1109/MNET.001.1900092.

[56] D. Cuomo, M. Caleffi, and A. S. Cacciapuoti, "Towards a distributed quantum computing ecosystem," *IET Quantum Communication*, vol. 1, no. 1, pp. 3–8, Jul. 2020. DOI: 10.1049/iet-qtc.2020.0002. [Online]. Available: https://doi.org/10.1049/iet-qtc.2020.0002.

[57] A. Ekert and C. Macchiavello, "Quantum error correction for communication," *Phys. Rev. Lett.*, vol. 77, pp. 2585–2588, 12 Sep. 1996. DOI: 10.1103/PhysRevLett.77.2585. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevLett.77.2585.

[58] C. H. Bennett, D. P. DiVincenzo, J. A. Smolin, and W. K. Wootters, "Mixed-state entanglement and quantum error correction," *Phys. Rev. A*, vol. 54, pp. 3824–3851, 5 Nov. 1996. DOI: 10.1103/PhysRevA.54.3824. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevA.54.3824.

[59] C. H. Bennett, H. J. Bernstein, S. Popescu, and B. Schumacher, "Concentrating partial entanglement by local operations," *Phys. Rev. A*, vol. 53, pp. 2046–2052, 4 Apr. 1996. DOI: 10.1103/PhysRevA.53.2046. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevA.53.2046.

[60] C. H. Bennett, G. Brassard, S. Popescu, B. Schumacher, J. A. Smolin, and W. K. Wootters, "Purification of noisy entanglement and faithful teleportation via noisy channels," *Phys. Rev. Lett.*, vol. 76, pp. 722–725, 5 Jan. 1996. DOI: 10.1103/PhysRevLett.76.722. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevLett.76.722.

[61] M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information*. Cambridge University Press, 2009. DOI: 10.1017/cbo9780511976667. [Online]. Available: https://doi.org/10.1017/cbo9780511976667.

[62] F. Rozp ędek, T. Schiet, L. P. Thinh, D. Elkouss, A. C. Doherty, and S. Wehner, "Optimizing practical entanglement distillation," *Phys. Rev. A*, vol. 97, p. 062 333, 6 Jun. 2018. DOI: 10.1103/PhysRevA.97.062333. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevA.97.062333.

[63] C. Č. Brukner, M. Żukowski, J.-W. Pan, and A. Zeilinger, "Bell's inequalities and quantum communication complexity," *Phys. Rev. Lett.*, vol. 92, p. 127 901, 12 Mar. 2004. DOI: 10.1103/PhysRevLett.92.127901. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevLett.92.127901.

[64] H. Buhrman, R. Cleve, S. Massar, and R. de Wolf, "Nonlocality and communication complexity," *Rev. Mod. Phys.*, vol. 82, pp. 665–698, 1 Mar. 2010. DOI: 10.1103/RevModPhys.82.665. [Online]. Available: https://link.aps.org/doi/10.1103/RevModPhys.82.665.

[65] Z.-A. Jia, L. Wei, Y.-C. Wu, and G.-C. Guo, "Quantum advantages of communication complexity from bell nonlocality," *Entropy*, vol. 23, no. 6, 2021, ISSN: 1099-4300. [Online]. Available: https://www.mdpi.com/1099-4300/23/6/744.

[66] R. Raussendorf, "Contextuality in measurement-based quantum computation," *Phys. Rev. A*, vol. 88, p. 022 322, 2 Aug. 2013. DOI: 10.1103/PhysRevA.88.022322. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevA.88.022322.

[67] M. Howard, J. Wallman, V. Veitch, and J. Emerson, "Contextuality supplies the "magic" for quantum computation," *Nature*, vol. 510, no. 7505, pp. 351–355, Jun. 2014. DOI: 10.1038/nature13460. [Online]. Available: https://doi.org/10.1038/nature13460.

[68] J. Bermejo-Vega, N. Delfosse, D. E. Browne, C. Okay, and R. Raussendorf, "Contextuality as a resource for models of quantum computation with qubits," *Phys. Rev. Lett.*, vol. 119, p. 120 505, 12 Sep. 2017. DOI: 10.1103/PhysRevLett.119.120505. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevLett.119.120505.

[69] I. L. Chuang and M. A. Nielsen, "Prescription for experimental determination of the dynamics of a quantum black box," *Journal of Modern Optics*, vol. 44, no. 11-12, pp. 2455–2467, 1997. DOI: 10.1080/09500349708231894. eprint: https://www.tandfonline.com/doi/pdf/10.1080/09500349708231894. [Online]. Available: https://www.tandfonline.com/doi/abs/10.1080/09500349708231894.

[70] D. W. Leung, "Choi's proof as a recipe for quantum process tomography," *Journal of Mathematical Physics*, vol. 44, no. 2, p. 528, 2003. DOI: 10.1063/1.1518554. [Online]. Available: https://doi.org/10.1063/1.1518554.

[71] R. Harper, S. T. Flammia, and J. J. Wallman, "Efficient learning of quantum noise," *Nature Physics*, vol. 16, no. 12, pp. 1184–1188, Aug. 2020. DOI: 10.1038/s41567-020-0992-8. [Online]. Available: https://doi.org/10.1038/s41567-020-0992-8.

[72] E. Onorati, T. Kohler, and T. Cubitt, "Fitting quantum noise models to tomography data," *arXiv preprint arXiv:2103.17243*, 2021.

[73] R. Levy, D. Luo, and B. K. Clark, "Classical shadows for quantum process tomography on near-term quantum computers," *arXiv preprint arXiv:2110.02965*, 2021.

[74] J. Kunjummen, M. C. Tran, D. Carney, and J. M. Taylor, "Shadow process tomography of quantum channels," *Phys. Rev. A*, vol. 107, p. 042 403, 4 Apr. 2023. DOI: 10.1103/PhysRevA.107.042403. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevA.107.042403.

[75] X. Xu, S. Benjamin, J. Sun, X. Yuan, and P. Zhang, "A herculean task: Classical simulation of quantum computers," *arXiv preprint arXiv:2302.08880*, 2023.

[76] R. Orús, "Tensor networks for complex quantum systems," *Nature Reviews Physics*, vol. 1, no. 9, pp. 538–550, Aug. 2019. DOI: 10.1038/s42254-019-0086-7. [Online]. Available: https://doi.org/10.1038/s42254-019-0086-7.

[77] G. Vidal, "Efficient classical simulation of slightly entangled quantum computations," *Phys. Rev. Lett.*, vol. 91, p. 147 902, 14 Oct. 2003. DOI: 10.1103/PhysRevLett.91.147902. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevLett.91.147902.

[78] F. Verstraete, J. J. García-Ripoll, and J. I. Cirac, "Matrix product density operators: Simulation of finite-temperature and dissipative systems," *Phys. Rev. Lett.*, vol. 93, p. 207 204, 20 Nov. 2004. DOI: 10.1103/PhysRevLett.93.207204. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevLett.93.207204.

[79] Y.-Y. Shi, L.-M. Duan, and G. Vidal, "Classical simulation of quantum many-body systems with a tree tensor network," *Phys. Rev. A*, vol. 74, p. 022 320, 2 Aug. 2006. DOI: 10.1103/PhysRevA.74.022320. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevA.74.022320.

[80] J. Jordan, R. Orús, G. Vidal, F. Verstraete, and J. I. Cirac, "Classical simulation of infinite-size quantum lattice systems in two spatial dimensions," *Phys. Rev. Lett.*, vol. 101, p. 250 602, 25 Dec. 2008. DOI: 10.1103/PhysRevLett.101.250602. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevLett.101.250602.

[81] T. Barthel, C. Pineda, and J. Eisert, "Contraction of fermionic operator circuits and the simulation of strongly correlated fermions," *Phys. Rev. A*, vol. 80, p. 042 333, 4 Oct. 2009. DOI: 10.1103/PhysRevA.80.042333. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevA.80.042333.

[82] P. Corboz, G. Evenbly, F. Verstraete, and G. Vidal, "Simulation of interacting fermions with entanglement renormalization," *Phys. Rev. A*, vol. 81, p. 010 303, 1 Jan. 2010. DOI: 10.1103/PhysRevA.81.010303. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevA.81.010303.

[83] R. Orús, "Exploring corner transfer matrices and corner tensors for the classical simulation of quantum lattice systems," *Phys. Rev. B*, vol. 85, p. 205 117, 20 May 2012. DOI: 10.1103/PhysRevB.85.205117. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevB.85.205117.

[84] C. Huang, F. Zhang, M. Newman, X. Ni, D. Ding, J. Cai, X. Gao, T. Wang, F. Wu, G. Zhang, H.-S. Ku, Z. Tian, J. Wu, H. Xu, H. Yu, B. Yuan, M. Szegedy, Y. Shi, H.-H. Zhao, C. Deng, and J. Chen, "Efficient parallelization of tensor network contraction for simulating quantum computation," *Nature Computational Science*, vol. 1, no. 9, pp. 578–587, Sep. 2021. DOI: 10.1038/s43588-021-00119-7. [Online]. Available: https://doi.org/10.1038/s43588-021-00119-7.

[85] F. Pan and P. Zhang, "Simulation of quantum circuits using the big-batch tensor network method," *Phys. Rev. Lett.*, vol. 128, p. 030 501, 3 Jan. 2022. DOI: 10.1103/PhysRevLett.128.030501. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevLett.128.030501.

[86] A. W. Sandvik and G. Vidal, "Variational quantum monte carlo simulations with tensor-network states," *Phys. Rev. Lett.*, vol. 99, p. 220 602, 22 Nov. 2007. DOI: 10.1103/PhysRevLett.99.220602. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevLett.99.220602.

[87] M. Espig, W. Hackbusch, S. Handschuh, and R. Schneider, "Optimization problems in contracted tensor networks," 2011.

[88] H.-J. Liao, J.-G. Liu, L. Wang, and T. Xiang, "Differentiable programming tensor networks," *Phys. Rev. X*, vol. 9, p. 031 041, 3 Sep. 2019. DOI: 10.1103/PhysRevX.9.031041. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevX.9.031041.

[89] D. Gottesman, "The heisenberg representation of quantum computers," *arXiv preprint quant-ph/9807006*, 1998.

[90] S. Aaronson and D. Gottesman, "Improved simulation of stabilizer circuits," *Phys. Rev. A*, vol. 70, p. 052 328, 5 Nov. 2004. DOI: 10.1103/PhysRevA.70.052328. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevA.70.052328.

[91] Y. Eldar, A. Megretski, and G. Verghese, "Designing optimal quantum detectors via semidefinite programming," *IEEE Transactions on Information Theory*, vol. 49, no. 4, pp. 1007–1012, 2003. DOI: 10.1109/TIT.2003.809510.

[92] S. Boyd, S. Boyd, L. Vandenberghe, and C. U. Press, *Convex Optimization*, ser. Berichte über verteilte messysteme pt. 1. Cambridge University Press, 2004, ISBN: 9780521833783. [Online]. Available: https://books.google.com/books?id=mYm0bLd3fcoC.

[93] M. Navascués, S. Pironio, and A. Acín, "Bounding the set of quantum correlations," *Phys. Rev. Lett.*, vol. 98, p. 010 401, 1 Jan. 2007. DOI: 10.1103/PhysRevLett.98.010401. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevLett.98.010401.

[94] P. Skrzypczyk and D. Cavalcanti, *Semidefinite Programming in Quantum Information Science*, ser. 2053-2563. IOP Publishing, 2023, ISBN: 978-0-7503-3343-6. DOI: 10.1088/978-0-7503-3343-6. [Online]. Available: https://dx.doi.org/10.1088/978-0-7503-3343-6.

[95] E. Chitambar and G. Gour, "Quantum resource theories," *Rev. Mod. Phys.*, vol. 91, p. 025 001, 2 Apr. 2019. DOI: 10.1103/RevModPhys.91.025001. [Online]. Available: https://link.aps.org/doi/10.1103/RevModPhys.91.025001.

[96] S. Abramsky, R. S. Barbosa, and S. Mansfield, "Contextual fraction as a measure of contextuality," *Phys. Rev. Lett.*, vol. 119, p. 050 504, 5 Aug. 2017. DOI: 10.1103/PhysRevLett.119.050504. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevLett.119.050504.

[97] K. Mordasewicz and J. Kaniewski, "Quantifying incompatibility of quantum measurements through non-commutativity," *Journal of Physics A: Mathematical and Theoretical*, vol. 55, no. 26, p. 265 302, Jun. 2022. DOI: 10.1088/1751-8121/ac71eb. [Online]. Available: https://doi.org/10.1088/1751-8121/ac71eb.

[98] S. G. A. Brito, B. Amaral, and R. Chaves, "Quantifying bell nonlocality with the trace distance," *Phys. Rev. A*, vol. 97, p. 022 111, 2 Feb. 2018. DOI: 10.1103/PhysRevA.97.022111. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevA.97.022111.

[99] D.-L. Deng, "Machine learning detection of bell nonlocality in quantum many-body systems," *Phys. Rev. Lett.*, vol. 120, p. 240 402, 24 Jun. 2018. DOI: 10.1103/PhysRevLett.120.240402. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevLett.120.240402.

[100] A. Canabarro, S. Brito, and R. Chaves, "Machine learning nonlocal correlations," *Phys. Rev. Lett.*, vol. 122, p. 200 401, 20 May 2019. DOI: 10.1103/PhysRevLett.122.200401. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevLett.122.200401.

[101] K. Bharti, T. Haug, V. Vedral, and L.-C. Kwek, "How to teach ai to play bell non-local games: Reinforcement learning," *arXiv preprint arXiv:1912.10783*, 2019.

[102] B. Doolittle, T. R. Bromley, N. Killoran, and E. Chitambar, "Variational quantum optimization of nonlocality in noisy quantum networks," *IEEE Transactions on Quantum Engineering*, pp. 1–28, 2023. DOI: 10.1109/TQE.2023.3243849.

[103] M. Cerezo, A. Arrasmith, R. Babbush, S. C. Benjamin, S. Endo, K. Fujii, J. R. McClean, K. Mitarai, X. Yuan, L. Cincio, and P. J. Coles, "Variational quantum algorithms," *Nature Reviews Physics*, vol. 3, no. 9, pp. 625–644, Aug. 2021. DOI: 10.1038/s42254-021-00348-9. [Online]. Available: https://doi.org/10.1038/s42254-021-00348-9.

[104] M. Schuld, V. Bergholm, C. Gogolin, J. Izaac, and N. Killoran, "Evaluating analytic gradients on quantum hardware," *Phys. Rev. A*, vol. 99, p. 032 331, 3 Mar. 2019. DOI: 10.1103/PhysRevA.99.032331. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevA.99.032331.

[105] A. Mari, T. R. Bromley, and N. Killoran, "Estimating the gradient and higher-order derivatives on quantum hardware," *Phys. Rev. A*, vol. 103, p. 012 405, 1 Jan. 2021. DOI: 10.1103/PhysRevA.103.012405. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevA.103.012405.

[106] O. Kyriienko and V. E. Elfving, "Generalized quantum circuit differentiation rules," *Phys. Rev. A*, vol. 104, p. 052 417, 5 Nov. 2021. DOI: 10.1103/PhysRevA.104.052417. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevA.104.052417.

[107] D. Wierichs, J. Izaac, C. Wang, and C. Y.-Y. Lin, "General parameter-shift rules for quantum gradients," *Quantum*, vol. 6, p. 677, Mar. 2022. DOI: 10.22331/q-2022-03-30-677. [Online]. Available: https://doi.org/10.22331/q-2022-03-30-677.

[108] B. Doolittle and T. Bromley, *Qnetvo: The quantum network variational optimizer*, https://github.com/ChitambarLab/qNetVO, version v0.4.1, Mar. 2022. DOI: 10.5281/zenodo.6345834. [Online]. Available: https://github.com/ChitambarLab/qNetVO.

[109] N. Brunner, D. Cavalcanti, S. Pironio, V. Scarani, and S. Wehner, "Bell nonlocality," *Rev. Mod. Phys.*, vol. 86, pp. 419–478, 2 Apr. 2014. DOI: 10.1103/RevModPhys.86.419. [Online]. Available: https://link.aps.org/doi/10.1103/RevModPhys.86.419.

[110] M. Żukowski, A. Zeilinger, M. A. Horne, and A. K. Ekert, ""event-ready-detectors" bell experiment via entanglement swapping," *Phys. Rev. Lett.*, vol. 71, pp. 4287–4290, 26 Dec. 1993. DOI: 10.1103/PhysRevLett.71.4287. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevLett.71.4287.

[111] S. Bose, V. Vedral, and P. L. Knight, "Multiparticle generalization of entanglement swapping," *Phys. Rev. A*, vol. 57, pp. 822–829, 2 Feb. 1998. DOI: 10.1103/PhysRevA.57.822. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevA.57.822.

[112] H.-J. Briegel, W. Dür, J. I. Cirac, and P. Zoller, "Quantum repeaters: The role of imperfect local operations in quantum communication," *Phys. Rev. Lett.*, vol. 81, pp. 5932–5935, 26 Dec. 1998. DOI: 10.1103/PhysRevLett.81.5932. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevLett.81.5932.

[113] N. Sangouard, C. Simon, H. de Riedmatten, and N. Gisin, "Quantum repeaters based on atomic ensembles and linear optics," *Rev. Mod. Phys.*, vol. 83, pp. 33–80, 1 Mar. 2011. DOI: 10.1103/RevModPhys.83.33. [Online]. Available: https://link.aps.org/doi/10.1103/RevModPhys.83.33.

[114] B. Doolittle and E. Chitambar, "Certifying the classical simulation cost of a quantum channel," *Phys. Rev. Res.*, vol. 3, p. 043 073, 4 Oct. 2021. DOI: 10.1103/PhysRevResearch.3.043073. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevResearch.3.043073.

[115] P. E. Frenkel and M. Weiner, "Classical information storage in an n-level quantum system," *Communications in Mathematical Physics*, vol. 340, no. 2, pp. 563–574, Sep. 2015. DOI: 10.1007/s00220-015-2463-0. [Online]. Available: https://doi.org/10.1007/s00220-015-2463-0.

[116] ——, "On entanglement assistance to a noiseless classical channel," *Quantum*, vol. 6, p. 662, Mar. 2022, ISSN: 2521-327X. DOI: 10.22331/q-2022-03-01-662. [Online]. Available: https://doi.org/10.22331/q-2022-03-01-662.

[117] R. A. Costa, M. Langberg, and J. Barros, "One-shot capacity of discrete channels," in *2010 IEEE International Symposium on Information Theory*, 2010, pp. 211–215. DOI: 10.1109/ISIT.2010.5513244.

[118] G. Ziegler, *Lectures on Polytopes*, ser. Graduate Texts in Mathematics. Springer New York, 2012.

[119] A. El Gamal and Y.-H. Kim, *Network information theory*. Cambridge university press, 2011.

[120] J. Pearl, "Causal inference in statistics: An overview," 2009.

[121] F. V. Jensen and T. D. Nielsen, *Bayesian networks and decision graphs*. Springer, 2007, vol. 2.

[122] M. Dall'Arno, S. Brandsen, A. Tosini, F. Buscemi, and V. Vedral, "No-hypersignaling principle," *Phys. Rev. Lett.*, vol. 119, p. 020 401, 2 Jul. 2017. DOI: 10.1103/PhysRevLett.119.020401. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevLett.119.020401.

[123] T. Heinosaari, O. Kerppo, and L. Leppäjärvi, "Communication tasks in operational theories," *arXiv preprint arXiv:2003.05264*, 2020.

[124] A. Tavakoli, A. Pozas-Kerstjens, M.-X. Luo, and M.-O. Renou, "Bell nonlocality in networks," *Reports on Progress in Physics*, vol. 85, no. 5, p. 056 001, Mar. 2022. DOI: 10.1088/1361-6633/ac41bb. [Online]. Available: https://dx.doi.org/10.1088/1361-6633/ac41bb.

[125] G. M. Ziegler, "Lectures on 0/1-polytopes," in *Polytopes-combinatorics and computation*, Springer, 2000, pp. 1–41.

[126] I. Pitowsky, "Correlation polytopes: Their geometry and complexity," *Mathematical Programming*, vol. 50, pp. 395–414, 1991. DOI: 10.1007/BF01594946. [Online]. Available: https://doi.org/10.1007/BF01594946.

[127] S. Pironio, "Lifting bell inequalities," *Journal of Mathematical Physics*, vol. 46, no. 6, p. 062 112, Jun. 2005. DOI: 10.1063/1.1928727. [Online]. Available: https://doi.org/10.1063/1.1928727.

[128] T. Christof and G. Reinelt, "Decomposition and parallelization techniques for enumerating the facets of combinatorial polytopes," *International Journal of Computational Geometry & Applications*, vol. 11, no. 04, pp. 423–437, 2001.

[129] D. Rosset, J.-D. Bancal, and N. Gisin, "Classifying 50 years of bell inequalities," *Journal of Physics A: Mathematical and Theoretical*, vol. 47, no. 42, p. 424 022, Oct. 2014. DOI: 10.1088/1751-8113/47/42/424022. [Online]. Available: https://doi.org/10.1088/1751-8113/47/42/424022.

[130] Q. Huangfu and J. J. Hall, "Parallelizing the dual revised simplex method," *Mathematical Programming Computation*, vol. 10, no. 1, pp. 119–142, 2018.

[131] I. Dunning, J. Huchette, and M. Lubin, "Jump: A modeling language for mathematical optimization," *SIAM Review*, vol. 59, no. 2, pp. 295–320, 2017.

[132] J. Bezanson, A. Edelman, S. Karpinski, and V. B. Shah, "Julia: A fresh approach to numerical computing," *SIAM Review*, vol. 59, no. 1, pp. 65–98, 2017. DOI: 10.1137/141000671.

[133] M. Zukowski, D. Kaszlikowski, A. Baturo, and J.-A. Larsson, "Strengthening the bell theorem: Conditions to falsify local realism in an experiment," *arXiv preprint quant-ph/9910058*, 1999.

[134] D. Kaszlikowski, P. Gnaci ński, M. Żukowski, W. Miklaszewski, and A. Zeilinger, "Violations of local realism by two entangled n-dimensional systems are stronger than for two qubits," *Phys. Rev. Lett.*, vol. 85, pp. 4418–4421, 21 Nov. 2000. DOI: 10.1103/PhysRevLett.85.4418. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevLett.85.4418.

[135] T. Christof and A. Löbel, *Porta*, http://porta.zib.de/, version v1.4.1, 1997. [Online]. Available: http://porta.zib.de/.

[136] B. Doolittle and B. Legat, *Xporta.jl*, https://github.com/JuliaPolyhedra/XPORTA.jl, version v0.1.3, Apr. 2020. [Online]. Available: https://github.com/JuliaPolyhedra/XPORTA.jl.

[137] B. Legat, R. Deits, G. Goretkin, T. Koolen, J. Huchette, D. Oyama, and M. Forets, *Juliapolyhedra/polyhedra.jl: V0.6.16*, version v0.6.16, Jun. 2021. DOI: 10.5281/zenodo.4993670. [Online]. Available: https://doi.org/10.5281/zenodo.4993670.

[138] B. Doolittle, *Bellscenario.jl*, https://github.com/ChitambarLab/BellScenario.jl, version v0.1.2, Dec. 2020. DOI: 10.5281/zenodo.5061693. [Online]. Available: https://github.com/ChitambarLab/BellScenario.jl.

[139] M. Planck, "On the law of distribution of energy in the normal spectrum," *Annalen der physik*, vol. 4, no. 553, p. 1, 1901.

[140] A. Einstein, *Über einen die erzeugung und verwandlung des lichtes betreffenden heuristischen gesichtspunkt*, 1905.

[141] N. Bohr, "I. on the constitution of atoms and molecules," *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, vol. 26, no. 151, pp. 1–25, 1913.

[142] L. de Broglie, "Waves and quanta," *Nature*, vol. 112, no. 2815, pp. 540–540, Oct. 1923. DOI: 10.1038/112540a0. [Online]. Available: https://doi.org/10.1038/112540a0.

[143] P. A. M. Dirac, *The principles of quantum mechanics*, 27. Oxford university press, 1981.

[144] J. Von Neumann, *Mathematical foundations of quantum mechanics: New edition*. Princeton university press, 2018, vol. 53.

[145] J. J. Sakurai and E. D. Commins, *Modern quantum mechanics, revised edition*. American Association of Physics Teachers, 1995.

[146] B. Doolittle and Contributors, *QBase.jl: A base library for quantum information*. https://github.com/ChitambarLab/QBase.jl, version v0.2.2, Aug. 2020. DOI: 10.5281/zenodo.4717994. [Online]. Available: https://github.com/chitambarlab/QBase.jl.

[147] W. F. Stinespring, "Positive functions on c*-algebras," *Proceedings of the American Mathematical Society*, vol. 6, no. 2, pp. 211–216, 1955. DOI: https://doi.org/10.2307/2032342.

[148] K. Kraus, *States, Effects, and Operations: Fundamental Notions of Quantum Theory (Lecture Notes in Physics, 190)*, B. A., J. D. Dollard, and W. H. Wootters, Eds. Springer, Oct. 1, 1983, p. 151, ISBN: 978-3-5401-2732-1.

[149] M. Neumark, "Self-adjoint extensions of the second kind of a symmetric operator," *Bull. Acad. Sci. USSR. Ser. Math.*, vol. 4, pp. 90–104, 1940.

[150] A. Einstein *et al.*, "On the electrodynamics of moving bodies," *Annalen der physik*, vol. 17, no. 10, pp. 891–921, 1905.

[151] A. Aspect, P. Grangier, and G. Roger, "Experimental tests of realistic local theories via bell's theorem," *Phys. Rev. Lett.*, vol. 47, pp. 460–463, 7 Aug. 1981. DOI: 10.1103/PhysRevLett.47.460. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevLett.47.460.

[152] J. A. Wheeler, "The "past" and the "delayed-choice" double-slit experiment," in *Mathematical Foundations of Quantum Theory*, Elsevier, 1978, pp. 9–48. DOI: 10.1016/b978-0-12-473250-6.50006-6. [Online]. Available: https://doi.org/10.1016/b978-0-12-473250-6.50006-6.

[153] R. P. Feynman, R. B. Leighton, and M. Sands, *The Feynman lectures on physics; New millennium ed.* New York, NY: Basic Books, 2010, Originally published 1963-1965. [Online]. Available: https://cds.cern.ch/record/1494701.

[154] V. Jacques, E. Wu, F. Grosshans, F. Treussart, P. Grangier, A. Aspect, and J.-F. Roch, "Experimental realization of wheeler's delayed-choice gedanken experiment," *Science*, vol. 315, no. 5814, pp. 966–968, Feb. 2007. DOI: 10.1126/science.1136303. [Online]. Available: https://doi.org/10.1126/science.1136303.

[155] A. Aspect, J. Dalibard, and G. Roger, "Experimental test of bell's inequalities using time-varying analyzers," *Phys. Rev. Lett.*, vol. 49, pp. 1804–1807, 25 Dec. 1982. DOI: 10.1103/PhysRevLett.49.1804. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevLett.49.1804.

[156] Q.-C. Sun, Y.-F. Jiang, B. Bai, W. Zhang, H. Li, X. Jiang, J. Zhang, L. You, X. Chen, Z. Wang, Q. Zhang, J. Fan, and J.-W. Pan, "Experimental demonstration of non-bilocality with truly independent sources and strict locality constraints," *Nature Photonics*, vol. 13, no. 10, pp. 687–691, Aug. 2019. DOI: 10.1038/s41566-019-0502-7. [Online]. Available: https://doi.org/10.1038/s41566-019-0502-7.

[157] D. Wu, Q. Zhao, C. Wang, L. Huang, Y.-F. Jiang, B. Bai, Y. Zhou, X.-M. Gu, F.-M. Liu, Y.-Q. Mao, Q.-C. Sun, M.-C. Chen, J. Zhang, C.-Z. Peng, X.-B. Zhu, Q. Zhang, C.-Y. Lu, and J.-W. Pan, "Closing the locality and detection loopholes in multiparticle entanglement self-testing," *Phys. Rev. Lett.*, vol. 128, p. 250 401, 25 Jun. 2022. DOI: 10.1103/PhysRevLett.128.250401. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevLett.128.250401.

[158] R. Chaves, G. B. Lemos, and J. Pienaar, "Causal modeling the delayed-choice experiment," *Phys. Rev. Lett.*, vol. 120, p. 190 401, 19 May 2018. DOI: 10.1103/PhysRevLett.120.190401. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevLett.120.190401.

[159] R. Ionicioiu, T. Jennewein, R. B. Mann, and D. R. Terno, "Is wave–particle objectivity compatible with determinism and locality?" *Nature Communications*, vol. 5, no. 1, Sep. 2014. DOI: 10.1038/ncomms5997. [Online]. Available: https://doi.org/10.1038/ncomms5997.

[160] R. Rossi, "Restrictions for the causal inferences in an interferometric system," *Phys. Rev. A*, vol. 96, p. 012 106, 1 Jul. 2017. DOI: 10.1103/PhysRevA.96.012106. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevA.96.012106.

[161] E. Polino, I. Agresti, D. Poderini, G. Carvacho, G. Milani, G. B. Lemos, R. Chaves, and F. Sciarrino, "Device-independent test of a delayed choice experiment," *Phys. Rev. A*, vol. 100, p. 022 111, 2 Aug. 2019. DOI: 10.1103/PhysRevA.100.022111. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevA.100.022111.

[162] H.-L. Huang, Y.-H. Luo, B. Bai, Y.-H. Deng, H. Wang, Q. Zhao, H.-S. Zhong, Y.-Q. Nie, W.-H. Jiang, X.-L. Wang, J. Zhang, L. Li, N.-L. Liu, T. Byrnes, J. P. Dowling, C.-Y. Lu, and J.-W. Pan, "Compatibility of causal hidden-variable theories with a delayed-choice experiment," *Phys. Rev. A*, vol. 100, p. 012 114, 1 Jul. 2019. DOI: 10.1103/PhysRevA.100.012114. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevA.100.012114.

[163] C. Budroni, A. Cabello, O. Gühne, M. Kleinmann, and J.-Å. Larsson, "Kochen-specker contextuality," *Rev. Mod. Phys.*, vol. 94, p. 045 007, 4 Dec. 2022. DOI: 10.1103/RevModPhys.94.045007. [Online]. Available: https://link.aps.org/doi/10.1103/RevModPhys.94.045007.

[164] I. Šupi ć, J.-D. Bancal, Y. Cai, and N. Brunner, "Genuine network quantum nonlocality and self-testing," *Phys. Rev. A*, vol. 105, p. 022 206, 2 Feb. 2022. DOI: 10.1103/PhysRevA.105.022206. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevA.105.022206.

[165] D. Cavalcanti, M. L. Almeida, V. Scarani, and A. Acin, "Quantum networks reveal quantum nonlocality," *Nature communications*, vol. 2, no. 1, pp. 1–6, 2011. DOI: 10.1038/ncomms1193. [Online]. Available: https://doi.org/10.1038/ncomms1193.

[166] M.-X. Luo, "Nonlocality of all quantum networks," *Phys. Rev. A*, vol. 98, p. 042 317, 4 Oct. 2018. DOI: 10.1103/PhysRevA.98.042317. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevA.98.042317.

[167] M.-O. Renou, E. Bäumer, S. Boreiri, N. Brunner, N. Gisin, and S. Beigi, "Genuine quantum nonlocality in the triangle network," *Phys. Rev. Lett.*, vol. 123, p. 140 401, 14 Sep. 2019. DOI: 10.1103/PhysRevLett.123.140401. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevLett.123.140401.

[168] P. Contreras-Tejada, C. Palazuelos, and J. I. de Vicente, "Genuine multipartite nonlocality is intrinsic to quantum networks," *Phys. Rev. Lett.*, vol. 126, p. 040 501, 4 Jan. 2021. DOI: 10.1103/PhysRevLett.126.040501. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevLett.126.040501.

[169] X. Coiteux-Roy, E. Wolfe, and M.-O. Renou, "No bipartite-nonlocal causal theory can explain nature's correlations," *Phys. Rev. Lett.*, vol. 127, p. 200 401, 20 Nov. 2021. DOI: 10.1103/PhysRevLett.127.200401. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevLett.127.200401.

[170] A. Pozas-Kerstjens, N. Gisin, and A. Tavakoli, "Full network nonlocality," *Phys. Rev. Lett.*, vol. 128, p. 010 403, 1 Jan. 2022. DOI: 10.1103/PhysRevLett.128.010403. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevLett.128.010403.

[171] A. G. Lamas and E. Chitambar, "Multipartite nonlocality in clifford networks," *arXiv preprint arXiv:2208.04731*, 2022. [Online]. Available: https://arxiv.org/abs/2208.04731.

[172] W. Heisenberg, "Uber den anschaulichen inhalt der quantentheoretischen kinematik und mechanik," *Zeitschrift fur Physik*, vol. 43, no. 3-4, pp. 172–198, Mar. 1927. DOI: 10.1007/bf01397280. [Online]. Available: https://doi.org/10.1007/bf01397280.

[173] A. S. Holevo, "Bounds for the quantity of information transmitted by a quantum communication channel," *Problemy Peredachi Informatsii*, vol. 9, no. 3, pp. 3–11, 1973.

[174] R. V. Meter, "Quantum networking and internetworking," *IEEE Network*, vol. 26, no. 4, pp. 59–64, 2012. DOI: 10.1109/MNET.2012.6246754.

[175] C. Simon, "Towards a global quantum network," *Nature Photonics*, vol. 11, no. 11, pp. 678–680, Oct. 2017. DOI: 10.1038/s41566-017-0032-0. [Online]. Available: https://doi.org/10.1038/s41566-017-0032-0.

[176] T. Proctor, P. Knott, and J. Dunningham, "Networked quantum sensing," *arXiv preprint arXiv:1702.04271*, 2017.

[177] E. T. Khabiboulline, J. Borregaard, K. De Greve, and M. D. Lukin, "Optical interferometry with quantum networks," *Phys. Rev. Lett.*, vol. 123, p. 070 504, 7 Aug. 2019. DOI: 10.1103/PhysRevLett.123.070504. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevLett.123.070504.

[178] D. Gottesman, T. Jennewein, and S. Croke, "Longer-baseline telescopes using quantum repeaters," *Phys. Rev. Lett.*, vol. 109, p. 070 503, 7 Aug. 2012. DOI: 10.1103/PhysRevLett.109.070503. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevLett.109.070503.

[179] E. T. Khabiboulline, J. Borregaard, K. De Greve, and M. D. Lukin, "Quantum-assisted telescope arrays," *Phys. Rev. A*, vol. 100, p. 022 316, 2 Aug. 2019. DOI: 10.1103/PhysRevA.100.022316. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevA.100.022316.

[180] C. H. Bennett and G. Brassard, "Quantum cryptography: Public key distribution and coin tossing," *arXiv preprint arXiv:2003.06557*, 1984.

[181] C. Elliott, "Building the quantum network," *New Journal of Physics*, vol. 4, pp. 46–46, Jul. 2002. DOI: 10.1088/1367-2630/4/1/346. [Online]. Available: https://doi.org/10.1088/1367-2630/4/1/346.

[182] U. Vazirani and T. Vidick, "Fully device independent quantum key distribution," *Communications of the ACM*, vol. 62, no. 4, pp. 133–133, 2019.

[183] A. Broadbent, J. Fitzsimons, and E. Kashefi, "Universal blind quantum computation," in *2009 50th Annual IEEE Symposium on Foundations of Computer Science*, IEEE, Oct. 2009. DOI: 10.1109/focs.2009.36. [Online]. Available: https://doi.org/10.1109/focs.2009.36.

[184] S. Pirandola, U. L. Andersen, L. Banchi, M. Berta, D. Bunandar, R. Colbeck, D. Englund, T. Gehring, C. Lupo, C. Ottaviani, J. L. Pereira, M. Razavi, J. S. Shaari, M. Tomamichel, V. C. Usenko, G. Vallone, P. Villoresi, and P. Wallden, "Advances in quantum cryptography," *Adv. Opt. Photon.*, vol. 12, no. 4, pp. 1012–1236, Dec. 2020. DOI: 10.1364/AOP.361502. [Online]. Available: http://opg.optica.org/aop/abstract.cfm?URI=aop-12-4-1012.

[185] V. Scarani, H. Bechmann-Pasquinucci, N. J. Cerf, M. Du šek, N. Lütkenhaus, and M. Peev, "The security of practical quantum key distribution," *Rev. Mod. Phys.*, vol. 81, pp. 1301–1350, 3 Sep. 2009. DOI: 10.1103/RevModPhys.81.1301. [Online]. Available: https://link.aps.org/doi/10.1103/RevModPhys.81.1301.

[186] R. Colbeck and A. Kent, "Private randomness expansion with untrusted devices," *Journal of Physics A: Mathematical and Theoretical*, vol. 44, no. 9, p. 095 305, Feb. 2011. DOI: 10.1088/1751-8113/44/9/095305. [Online]. Available: https://doi.org/10.1088/1751-8113/44/9/095305.

[187] H.-W. Li, M. Pawlowski, Z.-Q. Yin, G.-C. Guo, and Z.-F. Han, "Semi-device-independent randomness certification using $n$-1 quantum random access codes," *Phys. Rev. A*, vol. 85, p. 052 308, 5 May 2012. DOI: 10.1103/PhysRevA.85.052308. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevA.85.052308.

[188] G. Cañas, J. Cariñe, E. S. Gómez, J. F. Barra, A. Cabello, G. B. Xavier, G. Lima, and M. Pawłowski, "Experimental quantum randomness generation invulnerable to the detection loophole," *arXiv preprint arXiv:1410.3443*, 2014.

[189] W.-Z. Liu, M.-H. Li, S. Ragy, S.-R. Zhao, B. Bai, Y. Liu, P. J. Brown, J. Zhang, R. Colbeck, J. Fan, *et al.*, "Device-independent randomness expansion against quantum side information," *Nature Physics*, vol. 17, no. 4, pp. 448–451, 2021.

[190] Y. Liu, Q. Zhao, M.-H. Li, J.-Y. Guan, Y. Zhang, B. Bai, W. Zhang, W.-Z. Liu, C. Wu, X. Yuan, *et al.*, "Device-independent quantum random-number generation," *Nature*, vol. 562, no. 7728, pp. 548–551, 2018. DOI: 10.1038/s41586-018-0559-3. [Online]. Available: https://doi.org/10.1038/s41586-018-0559-3.

[191] Y. Liu, X. Yuan, M.-H. Li, W. Zhang, Q. Zhao, J. Zhong, Y. Cao, Y.-H. Li, L.-K. Chen, H. Li, T. Peng, Y.-A. Chen, C.-Z. Peng, S.-C. Shi, Z. Wang, L. You, X. Ma, J. Fan, Q. Zhang, and J.-W. Pan, "High-speed device-independent quantum random number generation without a detection loophole," *Phys. Rev. Lett.*, vol. 120, p. 010 503, 1 Jan. 2018. DOI: 10.1103/PhysRevLett.120.010503. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevLett.120.010503.

[192] M. Avesani, H. Tebyanian, P. Villoresi, and G. Vallone, "Semi-device-independent heterodyne-based quantum random-number generator," *Phys. Rev. Appl.*, vol. 15, p. 034 034, 3 Mar. 2021. DOI: 10.1103/PhysRevApplied.15.034034. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevApplied.15.034034.

[193] G. Carvacho, F. Andreoli, L. Santodonato, M. Bentivegna, R. Chaves, and F. Sciarrino, "Experimental violation of local causality in a quantum network," *Nature communications*, vol. 8, no. 1, pp. 1–6, 2017. DOI: https://doi.org/10.1038/ncomms14775.

[194] D. J. Saunders, A. J. Bennet, C. Branciard, and G. J. Pryde, "Experimental demonstration of nonbilocal quantum correlations," *Science Advances*, vol. 3, no. 4, e1602743, 2017. DOI: 10.1126/sciadv.1602743.

[195] F. Andreoli, G. Carvacho, L. Santodonato, M. Bentivegna, R. Chaves, and F. Sciarrino, "Experimental bilocality violation without shared reference frames," *Phys. Rev. A*, vol. 95, p. 062 315, 6 Jun. 2017. DOI: 10.1103/PhysRevA.95.062315. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevA.95.062315.

[196] Q.-C. Sun, Y.-F. Jiang, B. Bai, W. Zhang, H. Li, X. Jiang, J. Zhang, L. You, X. Chen, Z. Wang, *et al.*, "Experimental demonstration of non-bilocality with truly independent sources and strict locality constraints," *Nature Photonics*, vol. 13, no. 10, pp. 687–691, 2019. DOI: https://doi.org/10.1038/s41566-019-0502-7.

[197] D. Poderini, I. Agresti, G. Marchese, E. Polino, T. Giordani, A. Suprano, M. Valeri, G. Milani, N. Spagnolo, G. Carvacho, *et al.*, "Experimental violation of n-locality in a star quantum network," *Nature communications*, vol. 11, no. 1, pp. 1–8, 2020. DOI: https://doi.org/10.1038/s41467-020-16189-6.

[198] C. H. Bennett, G. Brassard, C. Crépeau, R. Jozsa, A. Peres, and W. K. Wootters, "Teleporting an unknown quantum state via dual classical and einstein-podolsky-rosen channels," *Phys. Rev. Lett.*, vol. 70, pp. 1895–1899, 13 Mar. 1993. DOI: 10.1103/PhysRevLett.70.1895. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevLett.70.1895.

[199] M. Epping, H. Kampermann, and D. Bruß, "Robust entanglement distribution via quantum network coding," *New Journal of Physics*, vol. 18, no. 10, p. 103 052, Oct. 2016. DOI: 10.1088/1367-2630/18/10/103052. [Online]. Available: https://doi.org/10.1088/1367-2630/18/10/103052.

[200] M. Pompili, S. L. N. Hermans, S. Baier, H. K. C. Beukers, P. C. Humphreys, R. N. Schouten, R. F. L. Vermeulen, M. J. Tiggelman, L. dos Santos Martins, B. Dirkse, S. Wehner, and R. Hanson, "Realization of a multinode quantum network of remote solid-state qubits," *Science*, vol. 372, no. 6539, pp. 259–264, Apr. 2021. DOI: 10.1126/science.abg1919. [Online]. Available: https://doi.org/10.1126/science.abg1919.

[201] S. L. N. Hermans, M. Pompili, H. K. C. Beukers, S. Baier, J. Borregaard, and R. Hanson, "Qubit teleportation between non-neighbouring nodes in a quantum network," *Nature*, vol. 605, no. 7911, pp. 663–668, May 2022. DOI: 10.1038/s41586-022-04697-y. [Online]. Available: https://doi.org/10.1038/s41586-022-04697-y.

[202] M. Pompili, C. D. Donne, I. te Raa, B. van der Vecht, M. Skrzypczyk, G. Ferreira, L. de Kluijver, A. J. Stolk, S. L. N. Hermans, P. Pawełczak, W. Kozlowski, R. Hanson, and S. Wehner, "Experimental demonstration of entanglement delivery using a quantum network stack," *npj Quantum Information*, vol. 8, no. 1, Oct. 2022. DOI: 10.1038/s41534-022-00631-2. [Online]. Available: https://doi.org/10.1038/s41534-022-00631-2.

[203] R. Raussendorf, D. E. Browne, and H. J. Briegel, "Measurement-based quantum computation on cluster states," *Phys. Rev. A*, vol. 68, p. 022 312, 2 Aug. 2003. DOI: 10.1103/PhysRevA.68.022312. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevA.68.022312.

[204] V. Danos, E. D'Hondt, E. Kashefi, and P. Panangaden, "Distributed measurement-based quantum computation," *Electronic Notes in Theoretical Computer Science*, vol. 170, pp. 73–94, Mar. 2007. DOI: 10.1016/j.entcs.2006.12.012. [Online]. Available: https://doi.org/10.1016/j.entcs.2006.12.012.

[205] H. J. Briegel, D. E. Browne, W. Dür, R. Raussendorf, and M. V. den Nest, "Measurement-based quantum computation," *Nature Physics*, vol. 5, no. 1, pp. 19–26, Jan. 2009. DOI: 10.1038/nphys1157. [Online]. Available: https://doi.org/10.1038/nphys1157.

[206] P. Pathumsoot, T. Matsuo, T. Satoh, M. Hajdu šek, S. Suwanna, and R. Van Meter, "Modeling of measurement-based quantum network coding on a superconducting quantum processor," *Phys. Rev. A*, vol. 101, p. 052 301, 5 May 2020. DOI: 10.1103/PhysRevA.101.052301. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevA.101.052301.

[207] T.-C. Wei, "Measurement-based quantum computation," *arXiv preprint arXiv:2109.10111*, 2021.

[208] D. Llewellyn, Y. Ding, I. I. Faruque, S. Paesani, D. Bacco, R. Santagati, Y.-J. Qian, Y. Li, Y.-F. Xiao, M. Huber, M. Malik, G. F. Sinclair, X. Zhou, K. Rottwitt, J. L. O'Brien, J. G. Rarity, Q. Gong, L. K. Oxenlowe, J. Wang, and M. G. Thompson, "Chip-to-chip quantum teleportation and multi-photon entanglement in silicon," *Nature Physics*, vol. 16, no. 2, pp. 148–153, Dec. 2019. DOI: 10.1038/s41567-019-0727-x. [Online]. Available: https://doi.org/10.1038/s41567-019-0727-x.

[209] L. S. Madsen, F. Laudenbach, M. F. Askarani, F. Rortais, T. Vincent, J. F. F. Bulmer, F. M. Miatto, L. Neuhaus, L. G. Helt, M. J. Collins, A. E. Lita, T. Gerrits, S. W. Nam, V. D. Vaidya, M. Menotti, I. Dhand, Z. Vernon, N. Quesada, and J. Lavoie, "Quantum computational advantage with a programmable photonic processor," *Nature*, vol. 606, no. 7912, pp. 75–81, Jun. 2022. DOI: 10.1038/s41586-022-04725-x. [Online]. Available: https://doi.org/10.1038/s41586-022-04725-x.

[210] V. Bergholm, J. Izaac, M. Schuld, C. Gogolin, S. Ahmed, V. Ajith, M. S. Alam, G. Alonso-Linaje, B. AkashNarayanan, A. Asadi, *et al.*, "Pennylane: Automatic differentiation of hybrid quantum-classical computations," *arXiv preprint arXiv:1811.04968*, 2018.

[211] M. Beth Ruskai, S. Szarek, and E. Werner, "An analysis of completely-positive trace-preserving maps on m2," *Linear Algebra and its Applications*, vol. 347, no. 1, pp. 159–187, 2002, ISSN: 0024-3795. DOI: https://doi.org/10.1016/S0024-3795(01)00547-X. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S002437950100547X.

[212] J. Cornwall, *Group theory in physics*, 1984.

[213] A. Cabello, A. Feito, and A. Lamas-Linares, "Bell's inequalities with realistic noise for polarization-entangled photons," *Phys. Rev. A*, vol. 72, p. 052 112, 5 Nov. 2005. DOI: 10.1103/PhysRevA.72.052112. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevA.72.052112.

[214] V. Scarani, "The device-independent outlook on quantum physics," *Acta Physica Slovaca*, vol. 62, no. 4, pp. 347–409, 2012.

[215] F. Costa and S. Shrapnel, "Quantum causal modelling," *New Journal of Physics*, vol. 18, no. 6, p. 063 032, Jun. 2016. DOI: 10.1088/1367-2630/18/6/063032. [Online]. Available: https://doi.org/10.1088/1367-2630/18/6/063032.

[216] J.-M. A. Allen, J. Barrett, D. C. Horsman, C. M. Lee, and R. W. Spekkens, "Quantum common causes and quantum causal models," *Phys. Rev. X*, vol. 7, p. 031 021, 3 Jul. 2017. DOI: 10.1103/PhysRevX.7.031021. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevX.7.031021.

[217] A. K. Suresh, M. Frembs, and E. G. Cavalcanti, "A semantics for counterfactuals in quantum causal models," *arXiv preprint arXiv:2302.11783*, 2023.

[218] R. Chaves and T. Fritz, "Entropic approach to local realism and noncontextuality," *Phys. Rev. A*, vol. 85, p. 032 113, 3 Mar. 2012. DOI: 10.1103/PhysRevA.85.032113. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevA.85.032113.

[219] T. Fritz and R. Chaves, "Entropic inequalities and marginal problems," *IEEE Transactions on Information Theory*, vol. 59, no. 2, pp. 803–817, 2013. DOI: 10.1109/TIT.2012.2222863.

[220] M. Tomamichel and E. Hänggi, "The link between entropic uncertainty and nonlocality," *Journal of Physics A: Mathematical and Theoretical*, vol. 46, no. 5, p. 055 301, Jan. 2013. DOI: 10.1088/1751-8113/46/5/055301. [Online]. Available: https://doi.org/10.1088/1751-8113/46/5/055301.

[221] S. Raeisi, P. Kurzyński, and D. Kaszlikowski, "Entropic tests of multipartite nonlocality and state-independent contextuality," *Phys. Rev. Lett.*, vol. 114, p. 200 401, 20 May 2015. DOI: 10.1103/PhysRevLett.114.200401. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevLett.114.200401.

[222] N. Miklin, A. A. Abbott, C. Branciard, R. Chaves, and C. Budroni, "The entropic approach to causal correlations," *New Journal of Physics*, vol. 19, no. 11, p. 113 041, Nov. 2017. DOI: 10.1088/1367-2630/aa8f9f. [Online]. Available: https://doi.org/10.1088/1367-2630/aa8f9f.

[223] G. Brassard, A. Broadbent, and A. Tapp, "Quantum pseudo-telepathy," *Foundations of Physics*, vol. 35, no. 11, pp. 1877–1907, 2005. DOI: 10.1007/s10701-005-7353-4. [Online]. Available: https://doi.org/10.1007/s10701-005-7353-4.

[224] J. Silman, S. Machnes, and N. Aharon, "On the relation between bell's inequalities and nonlocal games," *Physics Letters A*, vol. 372, no. 21, pp. 3796–3800, 2008, ISSN: 0375-9601. DOI: 10.1016/j.physleta.2008.03.001. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0375960108003836.

[225] C. Palazuelos and T. Vidick, "Survey on nonlocal games and operator space theory," *Journal of Mathematical Physics*, vol. 57, no. 1, p. 015 220, 2016. DOI: 10.1063/1.4938052. [Online]. Available: https://doi.org/10.1063/1.4938052.

[226] I. Šupić and J. Bowles, "Self-testing of quantum systems: A review," *Quantum*, vol. 4, p. 337, Sep. 2020. DOI: 10.22331/q-2020-09-30-337. [Online]. Available: https://doi.org/10.22331/q-2020-09-30-337.

[227] R. Gallego, N. Brunner, C. Hadley, and A. Acín, "Device-independent tests of classical and quantum dimensions," *Phys. Rev. Lett.*, vol. 105, p. 230 501, 23 Nov. 2010. DOI: 10.1103/PhysRevLett.105.230501. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevLett.105.230501.

[228] N. Brunner, M. Navascués, and T. Vértesi, "Dimension witnesses and quantum state discrimination," *Phys. Rev. Lett.*, vol. 110, p. 150 501, 15 Apr. 2013. DOI: 10.1103/PhysRevLett.110.150501. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevLett.110.150501.

[229] J. Ahrens, M. Badziąg Piotr anlid Pawłowski, M. Żukowski, and M. Bourennane, "Experimental tests of classical and quantum dimensionality," *Phys. Rev. Lett.*, vol. 112, p. 140 401, 14 Apr. 2014. DOI: 10.1103/PhysRevLett.112.140401. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevLett.112.140401.

[230] V. D'Ambrosio, F. Bisesto, F. Sciarrino, J. F. Barra, G. Lima, and A. Cabello, "Device-independent certification of high-dimensional quantum systems," *Phys. Rev. Lett.*, vol. 112, p. 140 503, 14 Apr. 2014. DOI: 10.1103/PhysRevLett.112.140503. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevLett.112.140503.

[231] J. I. de Vicente, "Shared randomness and device-independent dimension witnessing," *Phys. Rev. A*, vol. 95, p. 012 340, 1 Jan. 2017. DOI: 10.1103/PhysRevA.95.012340. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevA.95.012340.

[232] C.-E. Bardyn, T. C. H. Liew, S. Massar, M. McKague, and V. Scarani, "Device-independent state estimation based on bell's inequalities," *Phys. Rev. A*, vol. 80, p. 062 327, 6 Dec. 2009. DOI: 10.1103/PhysRevA.80.062327. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevA.80.062327.

[233] A. Tavakoli, J. ę. Kaniewski, T. Vértesi, D. Rosset, and N. Brunner, "Self-testing quantum states and measurements in the prepare-and-measure scenario," *Phys. Rev. A*, vol. 98, p. 062 307, 6 Dec. 2018. DOI: 10.1103/PhysRevA.98.062307. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevA.98.062307.

[234] G. Moreno, R. Nery, C. de Gois, R. Rabelo, and R. Chaves, "Semi-device-independent certification of entanglement in superdense coding," *Phys. Rev. A*, vol. 103, p. 022 426, 2 Feb. 2021. DOI: 10.1103/PhysRevA.103.022426. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevA.103.022426.

[235] A. Go čanin, I. Šupi ć, and B. Daki ć, "Sample-efficient device-independent quantum state verification and certification," *PRX Quantum*, vol. 3, p. 010 317, 1 Feb. 2022. DOI: 10.1103/PRXQuantum.3.010317. [Online]. Available: https://link.aps.org/doi/10.1103/PRXQuantum.3.010317.

[236] R. Rabelo, M. Ho, D. Cavalcanti, N. Brunner, and V. Scarani, "Device-independent certification of entangled measurements," *Phys. Rev. Lett.*, vol. 107, p. 050 502, 5 Jul. 2011. DOI: 10.1103/PhysRevLett.107.050502. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevLett.107.050502.

[237] E. S. Gómez, S. Gómez, P. González, G. Cañas, J. F. Barra, A. Delgado, G. B. Xavier, A. Cabello, M. Kleinmann, T. Vértesi, and G. Lima, "Device-independent certification of a nonprojective qubit measurement," *Phys. Rev. Lett.*, vol. 117, p. 260 401, 26 Dec. 2016. DOI: 10.1103/PhysRevLett.117.260401. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevLett.117.260401.

[238] J.-D. Bancal, N. Sangouard, and P. Sekatski, "Noise-resistant device-independent certification of bell state measurements," *Phys. Rev. Lett.*, vol. 121, p. 250 506, 25 Dec. 2018. DOI: 10.1103/PhysRevLett.121.250506. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevLett.121.250506.

[239] A. Tavakoli, M. Smania, T. Vértesi, N. Brunner, and M. Bourennane, "Self-testing nonprojective quantum measurements in prepare-and-measure experiments," *Science Advances*, vol. 6, no. 16, Apr. 2020. DOI: 10.1126/sciadv.aaw6664. [Online]. Available: https://doi.org/10.1126/sciadv.aaw6664.

[240] J.-D. Bancal, N. Sangouard, and P. Sekatski, "Noise-resistant device-independent certification of bell state measurements," *Phys. Rev. Lett.*, vol. 121, p. 250 506, 25 Dec. 2018. DOI: 10.1103/PhysRevLett.121.250506. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevLett.121.250506.

[241] M. O. Renou, J. ę. Kaniewski, and N. Brunner, "Self-testing entangled measurements in quantum networks," *Phys. Rev. Lett.*, vol. 121, p. 250 507, 25 Dec. 2018. DOI: 10.1103/PhysRevLett.121.250507. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevLett.121.250507.

[242] C. M. Lee, "Device-independent certification of non-classical measurements via causal models," *arXiv preprint arXiv:1806.10895*, 2018.

[243] M.-X. Luo, "A nonlocal game for witnessing quantum networks," *npj Quantum Information*, vol. 5, no. 1, pp. 1–6, 2019. DOI: 10.1038/s41534-019-0203-6. [Online]. Available: https://doi.org/10.1038/s41534-019-0203-6.

[244] I. Šupić, J. Bowles, M.-O. Renou, A. Acín, and M. J. Hoban, "Quantum networks self-test all entangled states," *Nature Physics*, Feb. 2023. DOI: 10.1038/s41567-023-01945-4. [Online]. Available: https://doi.org/10.1038/s41567-023-01945-4.

[245] S. Pironio, A. Acín, S. Massar, A. B. de La Giroday, D. N. Matsukevich, P. Maunz, S. Olmschenk, D. Hayes, L. Luo, T. A. Manning, *et al.*, "Random numbers certified by bell's theorem," *Nature*, vol. 464, no. 7291, pp. 1021–1024, 2010.

[246] H.-W. Li, Z.-Q. Yin, Y.-C. Wu, X.-B. Zou, S. Wang, W. Chen, G.-C. Guo, and Z.-F. Han, "Semi-device-independent random-number expansion without entanglement," *Phys. Rev. A*, vol. 84, p. 034 301, 3 Sep. 2011. DOI: 10.1103/PhysRevA.84.034301. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevA.84.034301.

[247] A. K. Ekert, "Quantum cryptography based on bell's theorem," *Phys. Rev. Lett.*, vol. 67, pp. 661–663, 6 Aug. 1991. DOI: 10.1103/PhysRevLett.67.661. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevLett.67.661.

[248] D. Mayers and A. Yao, "Quantum cryptography with imperfect apparatus," in *Proceedings 39th Annual Symposium on Foundations of Computer Science (Cat. No. 98CB36280)*, IEEE, 1998, pp. 503–509.

[249] J. Barrett, L. Hardy, and A. Kent, "No signaling and quantum key distribution," *Phys. Rev. Lett.*, vol. 95, p. 010 503, 1 Jun. 2005. DOI: 10.1103/PhysRevLett.95.010503. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevLett.95.010503.

[250] A. Acín, N. Brunner, N. Gisin, S. Massar, S. Pironio, and V. Scarani, "Device-independent security of quantum cryptography against collective attacks," *Phys. Rev. Lett.*, vol. 98, p. 230 501, 23 Jun. 2007. DOI: 10.1103/PhysRevLett.98.230501. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevLett.98.230501.

[251] S. Pironio, A. Acín, N. Brunner, N. Gisin, S. Massar, and V. Scarani, "Device-independent quantum key distribution secure against collective attacks," *New Journal of Physics*, vol. 11, no. 4, p. 045 021, Apr. 2009. DOI: 10.1088/1367-2630/11/4/045021. [Online]. Available: https://doi.org/10.1088/1367-2630/11/4/045021.

[252] M. Pawłowski and N. Brunner, "Semi-device-independent security of one-way quantum key distribution," *Phys. Rev. A*, vol. 84, p. 010 302, 1 Jul. 2011. DOI: 10.1103/PhysRevA.84.010302. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevA.84.010302.

[253] B. W. Reichardt, F. Unger, and U. Vazirani, "Classical command of quantum systems," *Nature*, vol. 496, no. 7446, pp. 456–460, 2013.

[254] R. Arnon-Friedman, F. Dupuis, O. Fawzi, R. Renner, and T. Vidick, "Practical device-independent quantum cryptography via entropy accumulation," *Nature communications*, vol. 9, no. 1, pp. 1–11, 2018.

[255] R. Schwonnek, K. T. Goh, I. W. Primaatmaja, E. Y.-Z. Tan, R. Wolf, V. Scarani, and C. C.-W. Lim, "Device-independent quantum key distribution with random key basis," *Nature Communications*, vol. 12, no. 1, May 2021. DOI: 10.1038/s41467-021-23147-3. [Online]. Available: https://doi.org/10.1038/s41467-021-23147-3.

[256] Y.-L. Tang, H.-L. Yin, Q. Zhao, H. Liu, X.-X. Sun, M.-Q. Huang, W.-J. Zhang, S.-J. Chen, L. Zhang, L.-X. You, Z. Wang, Y. Liu, C.-Y. Lu, X. Jiang, X. Ma, Q. Zhang, T.-Y. Chen, and J.-W. Pan, "Measurement-device-independent quantum key distribution over untrustful metropolitan network," *Phys. Rev. X*, vol. 6, p. 011 024, 1 Mar. 2016. DOI: 10.1103/PhysRevX.6.011024. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevX.6.011024.

[257] C. M. Lee and M. J. Hoban, "Towards device-independent information processing on general quantum networks," *Phys. Rev. Lett.*, vol. 120, p. 020 504, 2 Jan. 2018. DOI: 10.1103/PhysRevLett.120.020504. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevLett.120.020504.

[258] Y.-L. Yang, Y.-G. Yang, Y.-H. Zhou, W.-M. Shi, and D. Li, "Measurement-device-independent quantum wireless network communication," *Quantum Information Processing*, vol. 21, no. 4, Apr. 2022. DOI: 10.1007/s11128-022-03469-x. [Online]. Available: https://doi.org/10.1007/s11128-022-03469-x.

[259] M.-X. Luo, "Fully device-independent model on quantum networks," *Phys. Rev. Research*, vol. 4, p. 013 203, 1 Mar. 2022. DOI: 10.1103/PhysRevResearch.4.013203. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevResearch.4.013203.

[260] Y. Xiang, "Multipartite quantum cryptography based on the violation of svetlichn's inequality," *The European Physical Journal D*, vol. 77, no. 2, p. 31, 2023. DOI: 10.1140/epjd/s10053-023-00613-9. [Online]. Available: https://doi.org/10.1140/epjd/s10053-023-00613-9.

[261] I. Newman, "Private vs. common random bits in communication complexity," *Information processing letters*, vol. 39, no. 2, pp. 67–71, 1991.

[262] D. Bacon and B. F. Toner, "Bell inequalities with auxiliary communication," *Phys. Rev. Lett.*, vol. 90, p. 157 904, 15 Apr. 2003. DOI: 10.1103/PhysRevLett.90.157904. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevLett.90.157904.

[263] B. F. Toner and D. Bacon, "Communication cost of simulating bell correlations," *Phys. Rev. Lett.*, vol. 91, p. 187 904, 18 Oct. 2003. DOI: 10.1103/PhysRevLett.91.187904. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevLett.91.187904.

[264] M. Pawłowski, T. Paterek, D. Kaszlikowski, V. Scarani, A. Winter, and M. Żukowski, "Information causality as a physical principle," *Nature*, vol. 461, no. 7267, pp. 1101–1104, Oct. 2009. DOI: 10.1038/nature08400. [Online]. Available: https://doi.org/10.1038/nature08400.

[265] C. H. Bennett and S. J. Wiesner, "Communication via one- and two-particle operators on einstein-podolsky-rosen states," *Phys. Rev. Lett.*, vol. 69, pp. 2881–2884, 20 Nov. 1992. DOI: 10.1103/PhysRevLett.69.2881. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevLett.69.2881.

[266] E. Chitambar and F. Leditzky, "On the duality of teleportation and dense coding," *arXiv preprint arXiv:2302.14798*, 2023.

[267] T. Fritz, "Beyond bell's theorem: Correlation scenarios," *New Journal of Physics*, vol. 14, no. 10, p. 103 001, Oct. 2012. DOI: 10.1088/1367-2630/14/10/103001. [Online]. Available: https://doi.org/10.1088/1367-2630/14/10/103001.

[268] C. Branciard, N. Gisin, and S. Pironio, "Characterizing the nonlocal correlations created via entanglement swapping," *Phys. Rev. Lett.*, vol. 104, p. 170 401, 17 Apr. 2010. DOI: 10.1103/PhysRevLett.104.170401. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevLett.104.170401.

[269] C. Branciard, D. Rosset, N. Gisin, and S. Pironio, "Bilocal versus nonbilocal correlations in entanglement-swapping experiments," *Phys. Rev. A*, vol. 85, p. 032 119, 3 Mar. 2012. DOI: 10.1103/PhysRevA.85.032119. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevA.85.032119.

[270] A. Tavakoli, P. Skrzypczyk, D. Cavalcanti, and A. Acín, "Nonlocal correlations in the star-network configuration," *Phys. Rev. A*, vol. 90, p. 062 109, 6 Dec. 2014. DOI: 10.1103/PhysRevA.90.062109. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevA.90.062109.

[271] K. Mukherjee, B. Paul, and D. Sarkar, "Correlations in $n$-local scenario," *Quantum Information Processing*, vol. 14, no. 6, pp. 2025–2042, Apr. 2015. DOI: 10.1007/s11128-015-0971-7. [Online]. Available: https://doi.org/10.1007/s11128-015-0971-7.

[272] R. Chaves, "Polynomial bell inequalities," *Phys. Rev. Lett.*, vol. 116, p. 010 402, 1 Jan. 2016. DOI: 10.1103/PhysRevLett.116.010402. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevLett.116.010402.

[273] A. Tavakoli, "Bell-type inequalities for arbitrary noncyclic networks," *Phys. Rev. A*, vol. 93, p. 030 101, 3 Mar. 2016. DOI: 10.1103/PhysRevA.93.030101. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevA.93.030101.

[274] D. Rosset, C. Branciard, T. J. Barnea, G. Pütz, N. Brunner, and N. Gisin, "Nonlinear bell inequalities tailored for quantum networks," *Phys. Rev. Lett.*, vol. 116, p. 010 403, 1 Jan. 2016. DOI: 10.1103/PhysRevLett.116.010403. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevLett.116.010403.

[275] L. Yang, X. Qi, and J. Hou, "Nonlocal correlations in the tree-tensor-network configuration," *Phys. Rev. A*, vol. 104, p. 042 405, 4 Oct. 2021. DOI: 10.1103/PhysRevA.104.042405. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevA.104.042405.

[276] D. Rosset, C. Branciard, T. J. Barnea, G. Pütz, N. Brunner, and N. Gisin, "Nonlinear bell inequalities tailored for quantum networks," *Phys. Rev. Lett.*, vol. 116, p. 010 403, 1 Jan. 2016. DOI: 10.1103/PhysRevLett.116.010403. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevLett.116.010403.

[277] R. Chaves, "Polynomial bell inequalities," *Phys. Rev. Lett.*, vol. 116, p. 010 402, 1 Jan. 2016. DOI: 10.1103/PhysRevLett.116.010402. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevLett.116.010402.

[278] A. Tavakoli, "Bell-type inequalities for arbitrary noncyclic networks," *Phys. Rev. A*, vol. 93, p. 030 101, 3 Mar. 2016. DOI: 10.1103/PhysRevA.93.030101. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevA.93.030101.

[279] M.-X. Luo, "Computationally efficient nonlinear bell inequalities for quantum networks," *Phys. Rev. Lett.*, vol. 120, p. 140 402, 14 Apr. 2018. DOI: 10.1103/PhysRevLett.120.140402. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevLett.120.140402.

[280] A. Ambainis, A. Nayak, A. Ta-Shma, and U. Vazirani, "Dense quantum coding and a lower bound for 1-way quantum automata," in *Proceedings of the thirty-first annual ACM symposium on Theory of Computing*, 1999, pp. 376–383.

[281] A. Ambainis, D. Leung, L. Mancinska, and M. Ozols, "Quantum random access codes with shared randomness," *arXiv preprint arXiv:0810.2937*, 2008.

[282] A. Tavakoli, A. Hameedi, B. Marques, and M. Bourennane, "Quantum random access codes using single d-level systems," *Phys. Rev. Lett.*, vol. 114, p. 170 502, 17 Apr. 2015. DOI: 10.1103/PhysRevLett.114.170502. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevLett.114.170502.

[283] M. Quadeer, M. Banik, A. Ambainis, and A. Rai, "Strong supremacy of quantum systems as communication resource," *arXiv preprint arXiv:1703.01996*, 2017.

[284] C. Carmeli, T. Heinosaari, and A. Toigo, "Quantum random access codes and incompatibility of measurements," *EPL (Europhysics Letters)*, vol. 130, no. 5, p. 50 001, Jun. 2020. DOI: 10.1209/0295-5075/130/50001. [Online]. Available: https://doi.org/10.1209/0295-5075/130/50001.

[285] B. S. Cirel'son, "Quantum generalizations of bell's inequality," *Letters in Mathematical Physics*, vol. 4, no. 2, pp. 93–100, Mar. 1980. DOI: 10.1007/bf00417500. [Online]. Available: https://doi.org/10.1007/bf00417500.

[286] R. Pal and S. Ghosh, "Non-locality breaking qubit channels: The case for CHSH inequality," *Journal of Physics A: Mathematical and Theoretical*, vol. 48, no. 15, p. 155 302, Mar. 2015. DOI: 10.1088/1751-8113/48/15/155302. [Online]. Available: https://doi.org/10.1088/1751-8113/48/15/155302.

[287] Y. Zhang, R. A. Bravo, V. O. Lorenz, and E. Chitambar, "Channel activation of CHSH nonlocality," *New Journal of Physics*, vol. 22, no. 4, p. 043 003, Apr. 2020. DOI: 10.1088/1367-2630/ab7bef. [Online]. Available: https://doi.org/10.1088/1367-2630/ab7bef.

[288] Martín Abadi, Ashish Agarwal, Paul Barham, Eugene Brevdo, Zhifeng Chen, Craig Citro, Greg S. Corrado, Andy Davis, Jeffrey Dean, Matthieu Devin, Sanjay Ghemawat, Ian Goodfellow, Andrew Harp, Geoffrey Irving, Michael Isard, Y. Jia, Rafal Jozefowicz, Lukasz Kaiser, Manjunath Kudlur, Josh Levenberg, Dandelion Mané, Rajat Monga, Sherry Moore, Derek Murray, Chris Olah, Mike Schuster, Jonathon Shlens, Benoit Steiner, Ilya Sutskever, Kunal Talwar, Paul Tucker, Vincent Vanhoucke, Vijay Vasudevan, Fernanda Viégas, Oriol Vinyals, Pete Warden, Martin Wattenberg, Martin Wicke, Yuan Yu, and Xiaoqiang Zheng, *TensorFlow: Large-scale machine learning on heterogeneous systems*, Software available from tensorflow.org, 2015. [Online]. Available: https://www.tensorflow.org/.

[289] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, A. Desmaison, A. Kopf, E. Yang, Z. DeVito, M. Raison, A. Tejani, S. Chilamkurthy, B. Steiner, L. Fang, J. Bai, and S. Chintala, "Pytorch: An imperative style, high-performance deep learning library," in *Advances in Neural Information Processing Systems 32*, Curran Associates, Inc., 2019, pp. 8024–8035. [Online]. Available: http://papers.neurips.cc/paper/9015-pytorch-an-imperative-style-high-performance-deep-learning-library.pdf.

[290] S. Ruder, "An overview of gradient descent optimization algorithms," *arXiv preprint arXiv:1609.04747*, 2016.

[291] L. Bottou, F. E. Curtis, and J. Nocedal, "Optimization methods for large-scale machine learning," *SIAM Review*, vol. 60, no. 2, pp. 223–311, 2018. DOI: 10.1137/16M1080173.

[292] A. Griewank *et al.*, "On automatic differentiation," *Mathematical Programming: recent developments and applications*, vol. 6, no. 6, pp. 83–107, 1989.

[293] A. G. Baydin, B. A. Pearlmutter, A. A. Radul, and J. M. Siskind, "Automatic differentiation in machine learning: A survey," *Journal of Machine Learning Research*, vol. 18, no. 153, pp. 1–43, 2018. [Online]. Available: http://jmlr.org/papers/v18/17-468.html.

[294] J. Grabmeier, E. Kaltofen, and V. Weispfenning, *Computer algebra handbook: foundations, applications, systems*. Springer, 2003.

[295] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning representations by back-propagating errors," *nature*, vol. 323, no. 6088, pp. 533–536, 1986. DOI: https://doi.org/10.1038/323533a0.

[296] S. Suresh, S. Omkar, and V. Mani, "Parallel implementation of back-propagation algorithm in networks of workstations," *IEEE Transactions on Parallel and Distributed Systems*, vol. 16, no. 1, pp. 24–34, 2005. DOI: 10.1109/TPDS.2005.11.

[297] X. Sierra-Canto, F. Madera-Ramirez, and V. Uc-Cetina, "Parallel training of a back-propagation neural network using cuda," in *2010 Ninth International Conference on Machine Learning and Applications*, 2010, pp. 307–312. DOI: 10.1109/ICMLA.2010.52.

[298] J. Stokes, J. Izaac, N. Killoran, and G. Carleo, "Quantum Natural Gradient," *Quantum*, vol. 4, p. 269, May 2020, ISSN: 2521-327X. DOI: 10.22331/q-2020-05-25-269. [Online]. Available: https://doi.org/10.22331/q-2020-05-25-269.

[299] Z. Ma, P. Gokhale, T.-X. Zheng, S. Zhou, X. Yu, L. Jiang, P. Maurer, and F. T. Chong, "Adaptive circuit learning for quantum metrology," in *2021 IEEE International Conference on Quantum Computing and Engineering (QCE)*, 2021, pp. 419–430. DOI: 10.1109/QCE52317.2021.00063.

[300] T. Peng, A. W. Harrow, M. Ozols, and X. Wu, "Simulating large quantum circuits on a small quantum computer," *Phys. Rev. Lett.*, vol. 125, p. 150 504, 15 Oct. 2020. DOI: 10.1103/PhysRevLett.125.150504. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevLett.125.150504.

[301] F. Barratt, J. Dborin, M. Bal, V. Stojevic, F. Pollmann, and A. G. Green, "Parallel quantum simulation of large systems on small NISQ computers," *npj Quantum Information*, vol. 7, no. 1, May 2021. DOI: 10.1038/s41534-021-00420-3. [Online]. Available: https://doi.org/10.1038/s41534-021-00420-3.

[302] R. Gerritsma, G. Kirchmair, F. Zähringer, E. Solano, R. Blatt, and C. F. Roos, "Quantum simulation of the dirac equation," *Nature*, vol. 463, no. 7277, pp. 68–71, Jan. 2010. DOI: 10.1038/nature08688. [Online]. Available: https://doi.org/10.1038/nature08688.

[303] B. P. Lanyon, C. Hempel, D. Nigg, M. Müller, R. Gerritsma, F. Zähringer, P. Schindler, J. T. Barreiro, M. Rambach, G. Kirchmair, M. Hennrich, P. Zoller, R. Blatt, and C. F. Roos, "Universal digital quantum simulation with trapped ions," *Science*, vol. 334, no. 6052, pp. 57–61, Oct. 2011. DOI: 10.1126/science.1208001. [Online]. Available: https://doi.org/10.1126/science.1208001.

[304] R. Barends, L. Lamata, J. Kelly, L. García-Álvarez, A. G. Fowler, A. Megrant, E. Jeffrey, T. C. White, D. Sank, J. Y. Mutus, B. Campbell, Y. Chen, Z. Chen, B. Chiaro, A. Dunsworth, I.-C. Hoi, C. Neill, P. J. J. O'Malley, C. Quintana, P. Roushan, A. Vainsencher, J. Wenner, E. Solano, and J. M. Martinis, "Digital quantum simulation of fermionic models with a superconducting circuit," *Nature Communications*, vol. 6, no. 1, Jul. 2015. DOI: 10.1038/ncomms8654.

[305] E. A. Martinez, C. A. Muschik, P. Schindler, D. Nigg, A. Erhard, M. Heyl, P. Hauke, M. Dalmonte, T. Monz, P. Zoller, and R. Blatt, "Real-time dynamics of lattice gauge theories with a few-qubit quantum computer," *Nature*, vol. 534, no. 7608, pp. 516–519, Jun. 2016. DOI: 10.1038/nature18318. [Online]. Available: https://doi.org/10.1038/nature18318.

[306] B. Rost, B. Jones, M. Vyushkova, A. Ali, C. Cullip, A. Vyushkov, and J. Nabrzyski, "Simulation of thermal relaxation in spin chemistry systems on a quantum computer using inherent qubit decoherence," *arXiv preprint arXiv:2001.00794*, 2020.

[307] B. Rost, L. Del Re, N. Earnest, A. F. Kemper, B. Jones, and J. K. Freericks, "Demonstrating robust simulation of driven-dissipative problems on near-term quantum computers," *arXiv preprint arXiv:2108.01183*, 2021.

[308] A. M. Childs, D. Maslov, Y. Nam, N. J. Ross, and Y. Su, "Toward the first quantum simulation with quantum speedup," *Proceedings of the National Academy of Sciences*, vol. 115, no. 38, pp. 9456–9461, Sep. 2018. DOI: 10.1073/pnas.1801723115. [Online]. Available: https://doi.org/10.1073/pnas.1801723115.

[309] J. R. McClean, S. Boixo, V. N. Smelyanskiy, R. Babbush, and H. Neven, "Barren plateaus in quantum neural network training landscapes," *Nature Communications*, vol. 9, no. 1, Nov. 2018. DOI: 10.1038/s41467-018-07090-4. [Online]. Available: https://doi.org/10.1038/s41467-018-07090-4.

[310] M. Cerezo, A. Sone, T. Volkoff, L. Cincio, and P. J. Coles, "Cost function dependent barren plateaus in shallow parametrized quantum circuits," *Nature Communications*, vol. 12, no. 1, Mar. 2021. DOI: 10.1038/s41467-021-21728-w. [Online]. Available: https://doi.org/10.1038/s41467-021-21728-w.

[311] S. Wang, E. Fontana, M. Cerezo, K. Sharma, A. Sone, L. Cincio, and P. J. Coles, "Noise-induced barren plateaus in variational quantum algorithms," *Nature Communications*, vol. 12, no. 1, Nov. 2021. DOI: 10.1038/s41467-021-27045-6. [Online]. Available: https://doi.org/10.1038/s41467-021-27045-6.

[312] M. Cerezo and P. J. Coles, "Higher order derivatives of quantum neural networks with barren plateaus," *Quantum Science and Technology*, vol. 6, no. 3, p. 035 006, Jun. 2021. DOI: 10.1088/2058-9565/abf51a. [Online]. Available: https://doi.org/10.1088/2058-9565/abf51a.

[313] L. Bittel and M. Kliesch, "Training variational quantum algorithms is np-hard," *Phys. Rev. Lett.*, vol. 127, p. 120 502, 12 Sep. 2021. DOI: 10.1103/PhysRevLett.127.120502. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevLett.127.120502.

[314] J. F. Gonthier, M. D. Radin, C. Buda, E. J. Doskocil, C. M. Abuan, and J. Romero, "Identifying challenges towards practical quantum advantage through resource estimation: The measurement roadblock in the variational quantum eigensolver," *arXiv preprint arXiv:2012.04001*, p. 8, 2020.

[315] A. Suprano, D. Zia, E. Polino, T. Giordani, L. Innocenti, A. Ferraro, M. Paternostro, N. Spagnolo, and F. Sciarrino, "Dynamical learning of a photonics quantum-state engineering process," *Advanced Photonics*, vol. 3, no. 06, Dec. 2021. DOI: 10.1117/1.ap.3.6.066002. [Online]. Available: https://doi.org/10.1117/1.ap.3.6.066002.

[316] D. Poderini, E. Polino, G. Rodari, A. Suprano, R. Chaves, and F. Sciarrino, "Ab initio experimental violation of bell inequalities," *Phys. Rev. Research*, vol. 4, p. 013 159, 1 Feb. 2022. DOI: 10.1103/PhysRevResearch.4.013159. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevResearch.4.013159.

[317] W. C. Myrvold, J. Christian, and N. Gisin, "Bell inequalities: Many questions, a few answers," *Quantum Reality, Relativistic Causality, and Closing the Epistemic Circle: Essays in Honour of Abner Shimony*, pp. 125–138, 2009.

[318] A. Tavakoli, M. O. Renou, N. Gisin, and N. Brunner, "Correlations in star networks: From bell inequalities to network inequalities," *New Journal of Physics*, vol. 19, no. 7, p. 073 003, Jul. 2017. DOI: 10.1088/1367-2630/aa7673. [Online]. Available: https://dx.doi.org/10.1088/1367-2630/aa7673.

[319] R. Horodecki, P. Horodecki, and M. Horodecki, "Violating bell inequality by mixed spin-12 states: Necessary and sufficient condition," *Physics Letters A*, vol. 200, no. 5, pp. 340–344, 1995, ISSN: 0375-9601. DOI: https://doi.org/10.1016/0375-9601(95)00214-N. [Online]. Available: https://www.sciencedirect.com/science/article/pii/037596019500214N.

[320] O. Gamel, "Entangled bloch spheres: Bloch matrix and two-qubit state space," *Phys. Rev. A*, vol. 93, p. 062 320, 6 Jun. 2016. DOI: 10.1103/PhysRevA.93.062320. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevA.93.062320.

[321] A. Tavakoli, N. Gisin, and C. Branciard, "Bilocal bell inequalities violated by the quantum elegant joint measurement," *Phys. Rev. Lett.*, vol. 126, p. 220 401, 22 Jun. 2021. DOI: 10.1103/PhysRevLett.126.220401. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevLett.126.220401.

[322] L. Yang, X. Qi, and J. Hou, "Quantum nonlocality in any forked tree-shaped network," *Entropy*, vol. 24, no. 5, 2022, ISSN: 1099-4300. [Online]. Available: https://www.mdpi.com/1099-4300/24/5/691.

[323] Y. Yang, S. Xiao, and H. Cao, "Nonlocality of a type of multi-star-shaped quantum networks," *Journal of Physics A: Mathematical and Theoretical*, vol. 55, no. 2, p. 025 303, 2022. DOI: 10.1088/1751-8121/ac3f89.

[324] K. Mukherjee, "Detecting nontrilocal correlations in a triangle network," *Phys. Rev. A*, vol. 106, p. 042 206, 4 Oct. 2022. DOI: 10.1103/PhysRevA.106.042206. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevA.106.042206.

[325] P. Bej, A. Ghosal, A. Roy, S. Mal, and D. Das, "Creating quantum correlations in generalized entanglement swapping," *Phys. Rev. A*, vol. 106, p. 022 428, 2 Aug. 2022. DOI: 10.1103/PhysRevA.106.022428. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevA.106.022428.

[326] C.-X. Huang, X.-M. Hu, Y. Guo, C. Zhang, B.-H. Liu, Y.-F. Huang, C.-F. Li, G.-C. Guo, N. Gisin, C. Branciard, and A. Tavakoli, "Entanglement swapping and quantum correlations via symmetric joint measurements," *Phys. Rev. Lett.*, vol. 129, p. 030 502, 3 Jul. 2022. DOI: 10.1103/PhysRevLett.129.030502. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevLett.129.030502.

[327] G. Carvacho, E. Roccia, M. Valeri, F. B. Basset, D. Poderini, C. Pardo, E. Polino, L. Carosini, M. B. Rota, J. Neuwirth, S. F. C. da Silva, A. Rastelli, N. Spagnolo, R. Chaves, R. Trotta, and F. Sciarrino, "Quantum violation of local causality in an urban network using hybrid photonic technologies," *Optica*, vol. 9, no. 5, pp. 572–578, May 2022. DOI: 10.1364/OPTICA.451523. [Online]. Available: https://opg.optica.org/optica/abstract.cfm?URI=optica-9-5-572.

[328] S. Gupta, S. Datta, and A. S. Majumdar, "Preservation of quantum nonbilocal correlations in noisy entanglement-swapping experiments using weak measurements," *Phys. Rev. A*, vol. 98, p. 042 322, 4 Oct. 2018. DOI: 10.1103/PhysRevA.98.042322. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevA.98.042322.

[329] K. Mukherjee, I. Chakrabarty, and G. Mylavarapu, "Persistency of non n-local correlations in linear networks," *arXiv preprint arXiv:2208.06861*, 2022.

[330] N. Gisin, Q. Mei, A. Tavakoli, M. O. Renou, and N. Brunner, "All entangled pure quantum states violate the bilocality inequality," *Phys. Rev. A*, vol. 96, p. 020 304, 2 Aug. 2017. DOI: 10.1103/PhysRevA.96.020304. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevA.96.020304.

[331] F. Andreoli, G. Carvacho, L. Santodonato, R. Chaves, and F. Sciarrino, "Maximal qubit violation of n-locality inequalities in a star-shaped quantum network," *New Journal of Physics*, vol. 19, no. 11, p. 113 020, 2017. DOI: https://doi.org/10.1088/1367-2630/aa8b9b.

[332] A. Kundu, M. K. Molla, I. Chattopadhyay, and D. Sarkar, "Maximal qubit violation of $n$-local inequalities in a quantum network," *Phys. Rev. A*, vol. 102, p. 052 222, 5 Nov. 2020. DOI: 10.1103/PhysRevA.102.052222. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevA.102.052222.

[333] B. Doolittle and E. Chitambar, "Regarding the maximal qubit violations of n-locality in star and chain networks," *arXiv preprint arXiv:2212.06915*, 2022.

[334] R. Kumar and A. K. Pan, "Generalized n-locality inequalities in linear-chain network for arbitrary inputs scenario and their quantum violations," *Annalen der Physik*, p. 2 200 182, 2022. DOI: 10.1002/andp.202200182. [Online]. Available: https://doi.org/10.1002/andp.202200182.

[335] J. Körner and A. Orlitsky, "Zero-error information theory," *IEEE Transactions on Information Theory*, vol. 44, no. 6, pp. 2207–2229, 1998. DOI: 10.1109/18.720537.

[336] R. Duan, *Super-activation of zero-error capacity of noisy quantum channels*, 2009. eprint: arXiv:0906.2527.

[337] T. S. Cubitt, J. Chen, and A. W. Harrow, "Superactivation of the asymptotic zero-error classical capacity of a quantum channel," *IEEE Transactions on Information Theory*, vol. 57, no. 12, pp. 8114–8126, 2011. DOI: 10.1109/TIT.2011.2169109.

[338] T. S. Cubitt, D. Leung, W. Matthews, and A. Winter, "Zero-error channel capacity and simulation assisted by non-local correlations," *IEEE Transactions on Information Theory*, vol. 57, no. 8, pp. 5509–5523, 2011. DOI: 10.1109/TIT.2011.2159047.

[339] R. Duan and A. Winter, "No-signalling-assisted zero-error capacity of quantum channels and an information theoretic interpretation of the lovász number," *IEEE Transactions on Information Theory*, vol. 62, no. 2, pp. 891–914, 2016. DOI: 10.1109/TIT.2015.2507979.

[340] I. Devetak and A. Winter, "Distilling common randomness from bipartite quantum states," *IEEE Transactions on Information Theory*, vol. 50, no. 12, pp. 3183–3196, 2004. DOI: 10.1109/TIT.2004.838115.

[341] I. Devetak, A. W. Harrow, and A. J. Winter, "A resource framework for quantum shannon theory," *IEEE Transactions on Information Theory*, vol. 54, no. 10, pp. 4587–4618, 2008. DOI: 10.1109/TIT.2008.928980.

[342] C. Bennett, P. Shor, J. Smolin, and A. Thapliyal, "Entanglement-assisted capacity of a quantum channel and the reverse shannon theorem," *IEEE Transactions on Information Theory*, vol. 48, no. 10, pp. 2637–2655, Oct. 2002. DOI: 10.1109/tit.2002.802612. [Online]. Available: https://doi.org/10.1109/tit.2002.802612.

[343] X. Wang and M. M. Wilde, "Cost of quantum entanglement simplified," *Phys. Rev. Lett.*, vol. 125, p. 040 502, 4 Jul. 2020. DOI: 10.1103/PhysRevLett.125.040502.

[344] K. Fang, X. Wang, M. Tomamichel, and M. Berta, "Quantum channel simulation and the channel's smooth max-information," *IEEE Transactions on Information Theory*, vol. 66, no. 4, pp. 2129–2140, 2020. DOI: 10.1109/TIT.2019.2943858.

[345] M. Dall'Arno, E. Passaro, R. Gallego, and A. Acín, "Robustness of device-independent dimension witnesses," *Phys. Rev. A*, vol. 86, p. 042 312, 4 Oct. 2012. DOI: 10.1103/PhysRevA.86.042312. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevA.86.042312.

[346] J. Ahrens, P. Badziąg, A. Cabello, and M. Bourennane, "Experimental device-independent tests of classical and quantum dimensions," *Nature Physics*, vol. 8, no. 8, pp. 592–595, Jun. 2012. DOI: 10.1038/nphys2333.

[347] M. Dall'Arno, S. Brandsen, F. Buscemi, and V. Vedral, "Device-independent tests of quantum measurements," *Phys. Rev. Lett.*, vol. 118, p. 250 501, 25 Jun. 2017. DOI: 10.1103/PhysRevLett.118.250501.

[348] T. Heinosaari and O. Kerppo, "Communication of partial ignorance with qubits," *Journal of Physics A: Mathematical and Theoretical*, vol. 52, no. 39, p. 395 301, Sep. 2019. DOI: 10.1088/1751-8121/ab3ae4. [Online]. Available: https://doi.org/10.1088/1751-8121/ab3ae4.

[349] B. Doolittle, *SignalingDimension.jl*, https://github.com/ChitambarLab/SignalingDimension.jl, version v0.1.2, Feb. 2021. DOI: 10.5281/zenodo.5224630. [Online]. Available: https://github.com/ChitambarLab/SignalingDimension.jl.

[350] T. S. Cubitt, D. Leung, W. Matthews, and A. Winter, "Zero-error channel capacity and simulation assisted by non-local correlations," *IEEE Transactions on Information Theory*, vol. 57, no. 8, pp. 5509–5523, 2011. DOI: 10.1109/TIT.2011.2159047.

[351] J. E. Cohen and U. G. Rothblum, "Nonnegative ranks, decompositions, and factorizations of nonnegative matrices," *Linear Algebra and its Applications*, vol. 190, pp. 149–168, 1993, ISSN: 0024-3795. DOI: https://doi.org/10.1016/0024-3795(93)90224-C. [Online]. Available: https://www.sciencedirect.com/science/article/pii/002437959390224C.

[352] T. Lee, Z. Wei, and R. de Wolf, "Some upper and lower bounds on PSD-rank," *Mathematical Programming*, vol. 162, no. 1-2, pp. 495–521, Jul. 2016. DOI: 10.1007/s10107-016-1052-0. [Online]. Available: https://doi.org/10.1007/s10107-016-1052-0.

[353] H. Fawzi, J. Gouveia, P. A. Parrilo, R. Z. Robinson, and R. R. Thomas, "Positive semidefinite rank," *Mathematical Programming*, vol. 153, no. 1, pp. 133–177, Jul. 2015. DOI: 10.1007/s10107-015-0922-1. [Online]. Available: https://doi.org/10.1007/s10107-015-0922-1.

[354] A. Barvinok, *A Course in Convexity*. American Mathematical Society, 2002.

[355] E. Davies, "Information and quantum measurement," *Information Theory, IEEE Transactions on*, vol. 24, no. 5, pp. 596–599, Sep. 1978, ISSN: 0018-9448. DOI: 10.1109/TIT.1978.1055941.

[356] S. A. Vavasis, "On the complexity of nonnegative matrix factorization," *SIAM Journal on Optimization*, vol. 20, no. 3, pp. 1364–1377, Jan. 2010. DOI: 10.1137/070709967. [Online]. Available: https://doi.org/10.1137/070709967.

[357] A. CICHOCKI and A.-H. PHAN, "Fast local algorithms for large scale nonnegative matrix and tensor factorizations," *IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences*, vol. E92-A, no. 3, pp. 708–721, 2009. DOI: 10.1587/transfun.e92.a.708. [Online]. Available: https://doi.org/10.1587/transfun.e92.a.708.

[358] C. Févotte and J. Idier, "Algorithms for nonnegative matrix factorization with the $\beta$-divergence," *Neural Computation*, vol. 23, no. 9, pp. 2421–2456, 2011. DOI: 10.1162/NECO_a_00168.

[359] O. Grisel, A. Mueller, Lars, A. Gramfort, G. Louppe, P. Prettenhofer, T. J. Fan, M. Blondel, V. Niculae, J. Nothman, A. Joly, G. Lemaitre, J. Vanderplas, L. Estève, manoj kumar, J. du Boisberranger, H. Qin, N. Hug, N. Varoquaux, R. Layton, J. H. Metzen, A. Jalali, R. (Venkat) Raghav, J. Schönberger, R. Yurchak, J. Jerphanion, T. D. la Tour, W. Li, C. Marmo, and C. Woolam, *Scikit-learn/scikit-learn: Scikit-learn 1.2.2*, version 1.2.2, Mar. 2023. DOI: 10.5281/zenodo.7711792. [Online]. Available: https://doi.org/10.5281/zenodo.7711792.

[360] E. Chitambar, I. George, B. Doolittle, and M. Junge, "The communication value of a quantum channel," *IEEE Transactions on Information Theory*, vol. 69, no. 3, pp. 1660–1679, 2023. DOI: 10.1109/TIT.2022.3218540.

[361] B. Doolittle and I. George, *Cvchannel.jl: Convex optimization methods for evaluating the communication value of a quantum channel*. Version v0.1, Sep. 2021. DOI: 10.5281/zenodo.5517863. [Online]. Available: https://github.com/ChitambarLab/CVChannel.jl.

[362] L. B. Beasley and T. J. Laffey, "Real rank versus nonnegative rank," *Linear Algebra and its Applications*, vol. 431, no. 12, pp. 2330–2335, 2009, Special Issue in honor of Shmuel Friedland, ISSN: 0024-3795. DOI: https://doi.org/10.1016/j.laa.2009.02.034. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0024379509001256.

[363] E. Van der Meulen, "A survey of multi-way channels in information theory: 1961-1976," *IEEE Transactions on Information Theory*, vol. 23, no. 1, pp. 1–37, 1977.

[364] M. Huang, Y. Zhang, and G. Hou, "Classical capacity of a quantum multiple-access channel," *Phys. Rev. A*, vol. 62, p. 052 106, 5 Oct. 2000. DOI: 10.1103/PhysRevA.62.052106. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevA.62.052106.

[365] A. Winter, "The capacity of the quantum multiple-access channel," *IEEE Transactions on Information Theory*, vol. 47, no. 7, pp. 3059–3065, 2001. DOI: 10.1109/18.959287.

[366] M.-H. Hsieh, I. Devetak, and A. Winter, "Entanglement-assisted capacity of quantum multiple-access channels," *IEEE Transactions on Information Theory*, vol. 54, no. 7, pp. 3078–3090, 2008. DOI: 10.1109/TIT.2008.924726.

[367] J. Yard, P. Hayden, and I. Devetak, "Capacity theorems for quantum multiple-access channels: Classical-quantum and quantum-quantum capacity regions," *IEEE Transactions on Information Theory*, vol. 54, no. 7, pp. 3091–3113, 2008. DOI: 10.1109/TIT.2008.924665.

[368] J.-Q. Qin, J.-T. Wang, Y.-L. Yu, and X.-B. Wang, "General theory of quantum fingerprinting network," *Phys. Rev. Research*, vol. 3, p. 033 039, 3 Jul. 2021. DOI: 10.1103/PhysRevResearch.3.033039. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevResearch.3.033039.

[369] Y. Zhang, X. Chen, and E. Chitambar, "Building multiple access channels with a single particle," *Quantum*, vol. 6, p. 653, Feb. 2022. DOI: 10.22331/q-2022-02-16-653. [Online]. Available: https://doi.org/10.22331/q-2022-02-16-653.

[370] F. Leditzky, M. A. Alhejji, J. Levin, and G. Smith, "Playing games with multiple access channels," *Nature Communications*, vol. 11, no. 1, Mar. 2020. DOI: 10.1038/s41467-020-15240-w. [Online]. Available: https://doi.org/10.1038/s41467-020-15240-w.

[371] M. Herrero-Collantes and J. C. Garcia-Escartin, "Quantum random number generators," *Reviews of Modern Physics*, vol. 89, no. 1, p. 015 004, 2017.

[372] N. Moll, P. Barkoutsos, L. S. Bishop, J. M. Chow, A. Cross, D. J. Egger, S. Filipp, A. Fuhrer, J. M. Gambetta, M. Ganzhorn, A. Kandala, A. Mezzacapo, P. Müller, W. Riess, G. Salis, J. Smolin, I. Tavernelli, and K. Temme, "Quantum optimization using variational algorithms on near-term quantum devices," *Quantum Science and Technology*, vol. 3, no. 3, p. 030 503, Jun. 2018. DOI: 10.1088/2058-9565/aab822. [Online]. Available: https://doi.org/10.1088/2058-9565/aab822.

[373] K. Mitarai, M. Negoro, M. Kitagawa, and K. Fujii, "Quantum circuit learning," *Phys. Rev. A*, vol. 98, p. 032 309, 3 Sep. 2018. DOI: 10.1103/PhysRevA.98.032309. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevA.98.032309.

[374] H. Buhrman, N. Chandran, S. Fehr, R. Gelles, V. Goyal, R. Ostrovsky, and C. Schaffner, "Position-based quantum cryptography: Impossibility and constructions," *SIAM Journal on Computing*, vol. 43, no. 1, pp. 150–178, 2014.

[375] D. Manchala, "E-commerce trust metrics and models," *IEEE Internet Computing*, vol. 4, no. 2, pp. 36–44, 2000. DOI: 10.1109/4236.832944.

[376] B. J. Corbitt, T. Thanasankit, and H. Yi, "Trust and e-commerce: A study of consumer perceptions," *Electronic Commerce Research and Applications*, vol. 2, no. 3, pp. 203–215, 2003, Selected Papers from the Pacific Asia Conference on Information Systems, ISSN: 1567-4223. DOI: https://doi.org/10.1016/S1567-4223(03)00024-3. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1567422303000243.

[377] A. Salam, L. Iyer, P. Palvia, and R. Singh, "Trust in e-commerce," *Communications of the ACM*, vol. 48, no. 2, pp. 72–77, Feb. 2005. DOI: 10.1145/1042091.1042093. [Online]. Available: https://doi.org/10.1145/1042091.1042093.

[378] D. T. Chen, B. Doolittle, J. M. Larson, Z. H. Saleem, and E. Chitambar, "Inferring quantum network topology using local measurements," *arXiv preprint arXiv:2212.07987*, 2022.

[379] J. P. Dowling and G. J. Milburn, "Quantum technology: The second quantum revolution," *Philosophical Transactions of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences*, vol. 361, no. 1809, A. G. J. MacFarlane, Ed., pp. 1655–1674, Jun. 2003. DOI: 10.1098/rsta.2003.1227. [Online]. Available: https://doi.org/10.1098/rsta.2003.1227.

[380] B. M. Terhal, "Detecting quantum entanglement," *Theoretical Computer Science*, vol. 287, no. 1, pp. 313–335, 2002. DOI: 10.1016/S0304-3975(02)00139-1. [Online]. Available: https://doi.org/10.1016/S0304-3975(02)00139-1.

[381] O. Gühne and G. Tóth, "Entanglement detection," *Physics Reports*, vol. 474, no. 1-6, pp. 1–75, 2009. DOI: 10.1016/j.physrep.2009.02.004. [Online]. Available: https://doi.org/10.1016/j.physrep.2009.02.004.

[382] M. Huber, F. Mintert, A. Gabriel, and B. C. Hiesmayr, "Detection of high-dimensional genuine multipartite entanglement of mixed states," *Physical Review Letters*, vol. 104, p. 210 501, 21 May 2010. DOI: 10.1103/PhysRevLett.104.210501. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevLett.104.210501.

[383] C. Harney, S. Pirandola, A. Ferraro, and M. Paternostro, "Entanglement classification via neural network quantum states," *New Journal of Physics*, vol. 22, no. 4, p. 045 001, Apr. 2020. DOI: 10.1088/1367-2630/ab783d. [Online]. Available: https://dx.doi.org/10.1088/1367-2630/ab783d.

[384] M.-X. Luo and S.-M. Fei, "Robust multipartite entanglement without entanglement breaking," *Physical Review Research*, vol. 3, p. 043 120, 4 Nov. 2021. DOI: 10.1103/PhysRevResearch.3.043120. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevResearch.3.043120.

[385] K. Hansenne, Z.-P. Xu, T. Kraft, and O. Gühne, "Symmetries in quantum networks lead to no-go theorems for entanglement distribution and to verification techniques," *Nature Communications*, vol. 13, no. 1, Jan. 2022. DOI: 10.1038/s41467-022-28006-3. [Online]. Available: https://doi.org/10.1038/s41467-022-28006-3.

[386] F. Scala, S. Mangini, C. Macchiavello, D. Bajoni, and D. Gerace, "Quantum variational learning for entanglement witnessing," in *International Joint Conference on Neural Networks*, IEEE, Jul. 2022. DOI: 10.1109/ijcnn55064.2022.9892080.

[387] J. Henson, R. Lal, and M. F. Pusey, "Theory-independent limits on correlations from generalized bayesian networks," *New Journal of Physics*, vol. 16, no. 11, p. 113 043, Nov. 2014. DOI: 10.1088/1367-2630/16/11/113043. [Online]. Available: https://dx.doi.org/10.1088/1367-2630/16/11/113043.

[388] R. Chaves, C. Majenz, and D. Gross, "Information–theoretic implications of quantum causal structures," *Nature Communications*, vol. 6, no. 1, pp. 1–8, 2015. DOI: 10.1038/ncomms6766. [Online]. Available: https://doi.org/10.1038/ncomms6766.

[389] M. Weilenmann and R. Colbeck, "Analysing causal structures with entropy," *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 473, no. 2207, p. 20 170 483, 2017. DOI: 10.1098/rspa.2017.0483. [Online]. Available: https://doi.org/10.1098/rspa.2017.0483.

[390] M.-O. Renou, Y. Wang, S. Boreiri, S. Beigi, N. Gisin, and N. Brunner, "Limits on correlations in networks for quantum and no-signaling resources," *Physical Review Letters*, vol. 123, p. 070 403, 7 Aug. 2019. DOI: 10.1103/PhysRevLett.123.070403. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevLett.123.070403.

[391] M.-X. Luo, "Network configuration theory for all networks," *arXiv:2107.05846*, 2021. DOI: 10.48550/arXiv.2107.05846.

[392] A. Kela, K. V. Prillwitz, J. Aberg, R. Chaves, and D. Gross, "Semidefinite tests for latent causal structures," *IEEE Transactions on Information Theory*, vol. 66, no. 1, pp. 339–349, Jan. 2020. DOI: 10.1109/tit.2019.2935755. [Online]. Available: https://doi.org/10.1109/tit.2019.2935755.

[393] J. Åberg, R. Nery, C. Duarte, and R. Chaves, "Semidefinite tests for quantum network topologies," *Physical Review Letters*, vol. 125, p. 110 505, 11 Sep. 2020. DOI: 10.1103/PhysRevLett.125.110505. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevLett.125.110505.

[394] T. Kraft, C. Spee, X.-D. Yu, and O. Gühne, "Characterizing quantum networks: Insights from coherence theory," *Physical Review A*, vol. 103, p. 052 405, 5 May 2021. DOI: 10.1103/PhysRevA.103.052405. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevA.103.052405.

[395] E. Wolfe, R. W. Spekkens, and T. Fritz, "The inflation technique for causal inference with latent variables," *Journal of Causal Inference*, vol. 7, no. 2, p. 20 170 020, 2019. DOI: doi:10.1515/jci-2017-0020. [Online]. Available: https://doi.org/10.1515/jci-2017-0020.

[396] E. Wolfe, A. Pozas-Kerstjens, M. Grinberg, D. Rosset, A. Acín, and M. Navascués, "Quantum inflation: A general approach to quantum causal compatibility," *Physical Review X*, vol. 11, p. 021 043, 2 May 2021. DOI: 10.1103/PhysRevX.11.021043. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevX.11.021043.

[397] X. Yang, Y.-H. Yang, and M.-X. Luo, "Strong entanglement distribution of quantum networks," *Physical Review Research*, vol. 4, no. 1, p. 013 153, 2022. DOI: 10.1103/PhysRevResearch.4.013153.

[398] S. Dasgupta and T. S. Humble, "Stability of noisy quantum computing devices," *arXiv preprint arXiv:2105.09472*, 2021.