Device-independent Quantum Key Distribution and the Selection of Bell Inequalities

Inaugural Dissertation

for the attainment of the title of

Doctor of Natural Sciences (Dr. rer. nat.)

in the Faculty of Mathematics and Natural Sciences
at the Heinrich-Heine-Universität Düsseldorf

presented by

Sarnava Datta

from Balurghat, India

Düsseldorf, December 2022

from the Institute for Theoretical Physics III
at the Heinrich-Heine-Universität Düsseldorf

Published by permission of the
Faculty of Mathematics and Natural Sciences at
Heinrich-Heine-Universität Düsseldorf

Supervisor: Prof. Dr. Dagmar Bruß
Co-Supervisor: PD. Dr. Hermann Kampermann

Date of the oral examination:

Declaration of Authorship

I declare under oath that I have compiled my dissertation independently and without any undue assistance by third parties under consideration of the "Principles for the Safeguarding of Good Scientific Practice at Heinrich-Heine-Universität Düsseldorf".

Place, Date	Sarnava Datta

To my parents,
Sanjay and Susmita

Abstract

In recent years, we have seen significant developments in quantum cryptography and notably in quantum key distribution (QKD). A QKD protocol enables two parties to generate a shared secret key via an insecure quantum channel and an authenticated public classical channel. Though QKD provides information-theoretic security, it requires complete characterization of the sources and devices. Such settings are vulnerable to side-channel attacks. Device-independent quantum key distribution (DIQKD) is introduced to avoid such problems offering the strictest form of security. A DIQKD protocol does not depend on the intrinsic properties of the devices and channels involved in the protocol and can be performed with untrusted or uncharacterized devices. The central tenet behind a DIQKD protocol is the observation of nonlocal correlations, certified by a Bell inequality violation. The length of the final secure key depends on the Bell inequality and the amount of its violation. Thus, the selection of a suitable Bell inequality in a DIQKD protocol is essential. In most DIQKD protocols, the Bell inequality is chosen beforehand. We introduce a DIQKD protocol where a Bell inequality is designed from the complete measurement statistics which is one central topic of this thesis. From the input-output probability distribution, we construct an optimized Bell inequality that leads to the maximum Bell violation for the particular measurement settings. We use this optimized Bell inequality and the corresponding violation to bound the secret key rate by upper bounding the guessing probability (or lower bounding the min-entropy) using the NPA hierarchy [NPA07, NPA08]. We study our protocol for a general Bell scenario, i.e. for any number of measurement inputs and measurement outcomes and random measurement settings. We also perform finite-size secret key analysis under the assumption of collective attacks.

In the second part, we introduce a novel method to estimate the guessing probability, which is a crucial parameter in many device-independent cryptographic processes and can also serve as a witness for nonlocal correlations. We utilize neural network architectures and supervised machine learning

to estimate the guessing probability from the measurement statistics. Additionally, we use deep learning models to select suitable Bell inequalities from the input-output probability distribution that can then be used in a DIQKD protocol. Our results offer a novel method and a proof of principle for the relevance of deep learning for estimating the guessing probability, selecting a suitable Bell inequality and broadening the understanding of nonlocality.

Zusammenfassung

In den letzten Jahren gab es bedeutende Entwicklungen in der Quantenkryptografie, insbesondere in der Quantenschlüsselverteilung (QKD). Ein QKD-Protokoll ermöglicht es zwei Parteien, einen gemeinsamen geheimen Schlüssel über einen unsicheren Quantenkanal und einen authentifizierten öffentlichen klassischen Kanal zu erzeugen. Obwohl QKD informationstheoretische Sicherheit bietet, gilt dies nur, wenn eine vollständige Beschreibung über alle im Protokoll verwendeten Geräte vorliegt. Da dies häufig nicht gegeben ist, sind QKD Protokolle anfällig für so genannte Seitenkanalangriffe. Die geräteunabhängige Quantenschlüsselverteilung (DIQKD) wurde eingeführt um solche Probleme zu vermeiden und bietet die strengste Form der Sicherheit. Ein DIQKD-Protokoll hängt nicht von den intrinsischen Eigenschaften der am Protokoll beteiligten Geräte und Kanäle ab und kann mit nicht vertrauenswürdigen oder nicht charakterisierten Geräten durchgeführt werden. Der zentrale Grundsatz eines DIQKD-Protokolls ist die Beobachtung von nichtlokalen Korrelationen, die durch eine Verletzung der Bellschen Ungleichung bestätigt werden. Die Länge des endgültigen sicheren Schlüssels hängt von der Bell-Ungleichung und dem Ausmaß ihrer Verletzung ab. Daher ist die Auswahl einer geeigneten Bell-Ungleichung in einem DIQKD-Protokoll von entscheidender Bedeutung. In den meisten DIQKD-Protokollen wird die Bell-Ungleichung im Voraus gewählt. In dieser Arbeit stellen wir ein DIQKD-Protokoll vor, bei dem eine Bell-Ungleichung aus der vollständigen Messstatistik entwickelt wird. Aus der Eingabe-Ausgabe-Wahrscheinlichkeitsverteilung konstruieren wir eine optimierte Bell-Ungleichung, die zur maximalen Bell-Verletzung für die jeweiligen Messeinstellungen führt. Wir verwenden diese optimierte Bell-Ungleichung und die entsprechende Verletzung, um eine untere Schanke an die Rate des geheimen Schlüssels zu finden, indem wir die Erkennungswahrscheinlichkeit nach oben begrenzen (oder die Min-Entropie nach unten begrenzen), indem wir die NPA-Hierarchie verwenden. Wir untersuchen unser Protokoll für ein allgemeines Bell-Szenario, d. h. für eine beliebige Anzahl von Messeingängen

und Messergebnissen und zufällige Messeinstellungen. Wir führen auch eine Analyse des geheimen Schlüssels endlicher Länge unter der Annahme von kollektiven Angriffen durch.

Im zweiten Teil stellen wir eine neuartige Methode zur Schätzung der Erkennungswahrscheinlichkeit vor, die ein entscheidender Parameter in vielen geräteunabhängigen kryptographischen Verfahren ist und auch als Zeuge für nichtlokale Korrelationen dienen kann. Wir setzen neuronale Netzwerkarchitekturen und überwachtes maschinelles Lernen ein, um die Erkennungswahrscheinlichkeit aus den Messstatistiken zu schätzen. Darüber hinaus verwenden wir Deep-Learning-Modelle, um geeignete Bell-Ungleichungen aus der Eingabe-Ausgabe-Wahrscheinlichkeitsverteilung auszuwählen, die dann in einem DIQKD-Protokoll verwendet werden können. Unsere Ergebnisse bieten eine neuartige Methode und einen Grundsatzbeweis für die Relevanz von Deep Learning für die Schätzung der Erkennungswahrscheinlichkeit, die Auswahl einer geeigneten Bell-Ungleichung und die Erweiterung des Verständnisses von Nichtlokalität.

Acknowledgment

I want to express my gratitude to a lot of people who supported me along the way, either directly or indirectly, before we go into the intricacies of this thesis. Therefore, allow me to give credit where credit is due.

First of all, I want to express my gratitude to Dagmar Bruß, my supervisor, and Hermann Kampermann, my co-supervisor, for giving me a chance to develop into the scientist I am today. They gave me the privilege of working alongside them in a kind, friendly and productive environment. Their expertise and scientific thinking have been the most valuable resources for me at every level of my research. They continuously encourage me to strive for excellence in scientific research and writing, and I sincerely appreciate that. They assisted me in developing skills that I will undoubtedly carry with me for the rest of my life. For the rest of my life, they will undoubtedly be an inspiration to me.

I would also like to thank my past and present colleagues in Düsseldorf: Giulio Gianfelici, Julia Kunzelmann, Federico Grasselli, Carlo Liorni, Giacomo Carrara, Gláucia Murta, Lucas Tendick, Nikolai Wyderka, Thomas Wagner, Justus Neumann, Felix Bischof, and Timo Holz together with Lennart Bittel, Raphael Brieger, and our IT administrator Jens Bremer. I had enlightening and beneficial conversations with them about scientific and non-scientific subjects. Furthermore, I thank Lucas, Thomas, Gláucia, Giacomo, Julia, and Federico for their assistance in editing and proofreading this thesis and Lennart for translating its abstract. I also extend my gratitude to all other members of our research group.

In addition to the academic community, my parents, Sanjay, Susmita, and my sister Sarnalina deserve special recognition because none of this would have been possible without them. I shall always be grateful to them for their assistance. I want to emphasize how proud I am of my parents for starting off so modestly and achieving such wonderful accomplishments.

Contents

List of Figures

1 Introduction

Applications in the subject of cryptography enhance consumer access to privacy, authentication, and confidentiality. Secure communication is an important sub-field that aims to provide private communication between parties while ensuring that no unauthorized party can access the message. Numerous methods to encrypt messages arose over the years but were always broken afterwards. The field of cryptography has a long history of successes and failures.

In 1917, Vernam introduced the so-called One-Time Pad encryption, which protects against an eavesdropper with unlimited computational capacity [Ver26]. If the parties do not use the same key twice, this scheme cannot theoretically be thwarted. Three decades later, Shannon demonstrated the Vernam scheme's efficacy, showing that no other encryption technique utilizes a smaller key [Sha49]. A safe key distribution method is needed to enable secure communication using encryption schemes. Due to this limitation, most of today's cryptographic applications rely on systems whose security cannot be proven in theory and instead depends on our understanding of the complexity of particular tasks [RSA78]. Strictly speaking, it is possible to beat these strategies, but it will take substantial computing power. Quantum cryptography provides a solution to the issue of using the principles of quantum mechanics for secret communication. Unquestionably, quantum key distribution (QKD), whose goal is to provide a safe encryption key to the honest parties that desire to communicate, is the most significant aspect of quantum cryptography. Bennet and Brassard proposed the first QKD protocol in 1984, the famous BB84 protocol [BB84]. It uses the no-cloning theorem [WZ82] and properties of quantum physics to establish

secure communication.

A decade after the discovery of QKD, Peter Shor made an important discovery: large numbers can theoretically be factorized effectively [Sho99] as long as several quantum systems can be manipulated coherently. The current generation of conventional cryptographic methods is based on large integer factorization. It is computationally challenging and time-consuming. Shor's algorithm makes it possible to solve a large number in polynomial time. Even while quantum computers are not yet a reality, the mere thought that they might be developed has sparked worries about how the security of various cryptographic techniques may be in jeopardy. The cautious approach accelerates QKD research. Numerous QKD techniques [Eke91, Ben92, Bru98, Ren08, LMC05, GLLP04, SP00, SBPC+09, MQZL05, LCT14, TLGR12, SARG04, IWY02, IWY03, GRZ+04, SBG+05, WBC+14, LGPRC13, Lev15, TKI03, Koa04, KP03] have been proposed since the BB84 protocol. Despite groundbreaking work and offering information-theoretic security on paper, it is necessary to thoroughly characterize the apparatus, sources, and/or the channel between the parties. Since the theoretical description is needed to establish a secure key, these protocols are called device-dependent (DD) QKD protocols. Any experimental deviation the theoretical description does not account for could allow a malicious eavesdropper to undermine the protocol's security.

Realistically, it is often not plausible to completely characterize a device. The devices could even be prepared by a malicious eavesdropper (Eve). Furthermore, QKD is also vulnerable to hacking attempts that we will call "quantum hacking". Quantum hacking refers to attacks that force devices to behave differently from the model used in the security proof. For example, the security of many prepare-and-measure protocols (e.g. BB84) is required to presume that the source's emitting quantum states are well-characterized. Ref.[VMH01, GFK+06] proposed the so-called Trojan horse attack where Eve injects a bright light into the source utilized in the protocol by using reflectivity. Thus it modifies the emitted signals and gathers additional information about the modulation of the reflected light. Blinding attack [Mak09, LWW+10] is another form of quantum hacking. In this attack, Eve manipulates the detector by shining intense light into it, causing the detector only to activate if the receiver selects the same basis as Eve. Furthermore, some side-channels are opened without any active attacks by Eve. Interested readers can look into [XMZ+20] to see a detailed description of the side-channels and known attacks.

This calls for a new standard of security that is agnostic to the inter-

nal working of the devices. In [MY98], the authors introduce a device-independent (DI) way to certify the security of a cryptographic protocol where the protocol is independent of the exact internal workings of the quantum devices required for QKD. Security is established utilizing the nonlocal nature of input-output correlations. Bell inequalities [BCP+14] are indispensable for DI security as their violation verifies the quantum nature of input-output correlations. The secret key length will also depend on the estimated violation of the Bell inequality.

Therefore, in DIQKD, the selection of Bell inequalities has paramount importance. The parties typically choose the Bell inequality before the protocol's commencement [PAB+09, MPA11, AFDF+18, AFRV19]. However, [DKB22a, BSS14, NSPS14] proposes a DIQKD scenario in which the Bell inequality is created using the complete measurement statistics. In [NSPS14, BSS14], the Bell inequality designed is designed from the measurement statistics in such a way that it leads to the maximal device-independent secret key rate (DISKR) for that specific setup in the asymptotic scenario. The authors of [DKB22a] construct a Bell inequality that leads to the maximum Bell violation for that particular measurement setting of Alice and Bob. The optimized Bell inequality, tailored to the measurement statistics, and the corresponding violation, can be used to bound the achievable DISKR by solving semidefinite optimization problems. In [DKB22b], it is shown that one can also utilize trained deep learning models to obtain an optimal Bell inequality, which can then be employed in a DIQKD protocol. Deep learning networks can also be exploited to directly estimate the guessing probability. The introduction of deep learning makes the estimation of the optimal Bell inequality and guessing probability faster while achieving a very high degree of accuracy and low statistical error.

Structure of the thesis

This thesis aims to describe our research on Device-independent Quantum Key Distribution and the selection of Bell inequalities clearly and comprehensively. In order to achieve this, our manuscript displays the logical structure below.

In Chap. 2, we introduce basic notions of many concepts in quantum information theory and the required mathematical tools. The concepts of quantum information theory are essential to comprehend the rest of this thesis. We provide definitions for various entropies that encapsulate various

measures of information. These entropies and their operational meaning will undoubtedly play a significant role in the latter part of this thesis.

Chap. 3 is devoted to Bell inequalities and nonlocality. First, we thoroughly discuss different types of correlation. Then, we discuss the Bell inequalities and study the most famous Bell inequality, the CHSH inequality. After that, we present a numerical tool that derives a Bell inequality from the measurement outcomes of an experiment. Finally, we finish the chapter with a brief discussion of semi-definite programming and NPA hierarchy, a numerical tool required to characterize the quantum correlations.

We introduce quantum key distribution (QKD) in Chap. 4. We define the notions of security in the quantum cryptographic scenario and explain what attacks an adversary can make. We then describe the steps of a generic QKD protocol and create a platform for finite key security analysis. Additionally, we describe the paradigmatic BB84 and Entanglement-based BB84 protocol and prove its security under the most general circumstances.

The topic of Chap. 5 is Device-independent (DI) QKD. In recent years, much research has been done into the DI security of quantum cryptography systems. We discuss the assumptions made in a DIQKD protocol. We clarify how a DIQKD protocol's security relates to Bell inequality violations. From there, we present the Clauser-Horn-Shimony-Holt (CHSH) inequality-based DIQKD protocol and provide an analytical expression for the asymptotic secret key rate. We also present the analytical expression of the asymptotic secret key rate in a more general Bell scenario pertinent to this thesis. Additionally, we give a quick overview of the most current, cutting-edge experimental realizations of DIQKD.

The choice of Bell inequality certainly plays a central role in DIQKD. In Chap. 6, we introduce a DIQKD scenario in which an optimal Bell inequality is constructed from the complete measurement statistics rather than fixing a specific Bell inequality beforehand.

Chap. 7 introduces the machine learning approach to deal with the guessing probability estimation problem. First, we provide a brief overview of machine learning and a feed-forward neural network. We then use trained deep learning models to estimate the guessing probability and the associated optimal Bell inequality from the random bipartite quantum probability distribution.

We conclude with Chap. 8 and give an outlook for future research based on our work.

The original publications of our research manuscripts are provided in the appendices.

2

Preliminaries

Quantum mechanics is the physical model we use to characterize the quantum-cryptography protocols in this thesis. We presume the reader is familiar with the fundamental ideas of quantum mechanics because this thesis does not provide the opportunity to cover quantum mechanics and other related topics in detail. We further assume that the reader is conversant with the fundamentals of statistics, including random variables, expectations, probability distributions, and linear algebra. However, in this chapter, we will quickly go through a few of those ideas using the Dirac notation, operators, quantum measurements, and the postulates of quantum mechanics. We also briefly explain the density operator formalism, which is helpful when approaching quantum mechanics from the viewpoint of information theory. Finally, we also discuss the entropies that characterize information-processing tasks prevalent in quantum cryptography. This chapter is mostly inspired from [NC10, Ren08, Tom15, Wil13, Gra21].

2.1 Dirac Notation & Hilbert Spaces

In this thesis, we particularly consider discrete quantum systems with finite $d \in \mathbb{N}$ inherent degrees of freedom.

Definition 2.1 (Hilbert Spaces)**.** *Quantum systems are associated with a Hilbert space. A Hilbert space $\mathcal{H}$ is a vector space equipped with a scalar product, denoted by $\langle \cdot | \cdot \rangle$, over the field $\mathbb{C}$ of complex numbers that is complete with respect to the norm induced by the scalar product.*

The simplest conceivable Hilbert space is that of a single spin. It is spanned by the two vectors $|0\rangle$ and $|1\rangle$ or $|\uparrow\rangle$ and $|\downarrow\rangle$; i.e. the spin points up or down. The state of a quantum mechanical system with d degrees of freedom is represented by a d-dimensional normalized vector $|\psi\rangle$ of state space $\mathcal{H}$. The notation $|\psi\rangle$ was introduced by Dirac in 1939 [Dir39]. This vector symbol $|\psi\rangle$ is called a *ket*. To rigorously define the inner product of the Hilbert space $\mathcal{H}$, we introduce the dual version of $\mathcal{H}$.

Definition 2.2 (Dual vectors)**.** *For a Hilbert space $\mathcal{H}$ over the field $\mathbb{C}$, the dual space $\mathcal{H}^*$ is the vector space of all linear maps $\mathcal{H} \to \mathbb{C}$. Elements of the dual space are denoted by $\langle\phi|$ and are called the dual (or bra) vector.*

The action of $\langle\phi| \in \mathcal{H}^*$ on a vector $|\psi\rangle \in \mathcal{H}$ can be written as:

$$\langle\phi| : |\psi\rangle \to \langle\phi|\psi\rangle \in \mathbb{C}\,, \tag{2.1}$$

i.e. they map the state to a scalar entity. This operation also defines the inner product $\langle\phi|\psi\rangle$ of the states $|\psi\rangle, |\phi\rangle \in \mathcal{H}$. Two vectors $|\psi\rangle, |\phi\rangle \in \mathcal{H}$ is said to be *orthogonal* if their inner product $\langle\phi|\psi\rangle = 0$. The inner product induces the norm $\|.\|$ on $\mathcal{H}$, via $\| |\psi\rangle \| := \sqrt{\langle\psi|\psi\rangle}$. We will call a state $|\psi\rangle$ normalized if $\| |\psi\rangle \| = 1$. This orthogonality and normality lead to orthonormal states.

Definition 2.3 (Orthonormal States)**.** *A set of vectors $\{|\psi_i\rangle\}$ is called orthonormal and is exclusively composed of normalized and mutually orthogonal vectors: if $\langle\psi_i|\psi_j\rangle = \delta_{ij}$, where δ_{ij} is the Kronecker delta.*

2.2 Operators

Definition 2.4 (Linear Operator)**.** *A linear operator L is a linear map from Hilbert space $\mathcal{H}_A$ to $\mathcal{H}_B$ that takes elements of $\mathcal{H}_A$, $|\psi\rangle \in \mathcal{H}_A$ to $\mathcal{H}_B$: $L|\psi\rangle \in \mathcal{H}_B$. A linear operator can be represented as a matrix in a pair of orthonormal bases for $\mathcal{H}_A$ and $\mathcal{H}_B$, $\{|e_i\rangle\}_{i=1}^{d_A}$ and $\{|f_j\rangle\}_{j=1}^{d_B}$, respectively, where d_A and d_B are the dimensions of $\mathcal{H}_A$ and $\mathcal{H}_B$. The matrix representation for L is then given by*

$$L = \sum_{i,j} L_{i,j}\, |f_j\rangle \langle e_i|\ , \tag{2.2}$$

where $L_{i,j} = \langle f_j|L|e_i\rangle$.

We define $\mathcal{L}(\mathcal{H}_A, \mathcal{H}_B)$ as the set of linear operators from $\mathcal{H}_A$ to $\mathcal{H}_B$ and $\mathcal{L}(\mathcal{H})$ as the set of linear operators that map from $\mathcal{H}$ to $\mathcal{H}$. The adjoint or

Hermitian conjugate of an operator that maps from $\mathcal{H}_A$ to $\mathcal{H}_B$ is denoted by $L^\dagger$ and is defined by:

$$\langle\psi|L|\phi\rangle = \left(\langle\phi|L^\dagger|\psi\rangle\right)^* \quad \text{for} \quad |\phi\rangle \in \mathcal{H}_A, |\psi\rangle \in \mathcal{H}_B\,, \tag{2.3}$$

where * is the complex conjugate.

An operator $L \in \mathcal{L}(\mathcal{H})$ is hermitian if $L^\dagger = L$. A positive semidefinite operator M is a linear operator, that is Hermitian and satisfies

$$\langle\psi|M|\psi\rangle \geq 0\,, \quad \forall\psi \in \mathcal{H}\,. \tag{2.4}$$

A unitary operator U is a linear operator, that satisfies

$$UU^\dagger = U^\dagger U = \mathbb{1}\,, \tag{2.5}$$

where $\mathbb{1}$ is the identity operator. For an orthonormal basis $\{|e_i\rangle\}$, the identity operator can be written as $\mathbb{1} = \sum_i |e_i\rangle\langle e_i|$. A unitary transformation also links any two orthonormal bases $\{e_i\}_{i=1}^d$ and $\{f_i\}_{i=1}^d$ in the Hilbert space $\mathcal{H}$, i.e.

$$|f_i\rangle = U\,|e_i\rangle\ .$$

Two bases are called mutually unbiased if $|\langle e_i|f_j\rangle|^2 = \frac{1}{d}$, for every i and j.

Projectors are one specific type of linear operator. The projectors are the operators $\Pi \in \mathcal{L}(\mathcal{H})$ that satisfy $\Pi^2 = \Pi$. For a set of orthonormal states $\{|\phi_i\rangle\}$, projector can be expressed as $\sum_i \left|\phi_i\right\rangle\!\left\langle\phi_i\right|$ that is not necessarily complete, i.e. $\sum_i \left|\phi_i\right\rangle\!\left\langle\phi_i\right| \leq 1$.

Definition 2.5 (Eigenvalues & Eigenstates)**.** *Let $A \in \mathcal{L}(\mathcal{H})$ is a linear operator on a Hilbert space $\mathcal{H}$. If a scalar $a \in \mathbb{C}$ and $|a\rangle \in \mathcal{H}$ with $|a\rangle \neq 0$ satisfy the following equation*

$$A\,|a\rangle = a\,|a\rangle\ ,$$

then we say that a is the eigenvalue of A and $|a\rangle$ is eigenstate of A belonging to the eigenvalue a.

The eigenvalues $\{a_i\}$ of an operator $A \in \mathcal{L}(\mathcal{H})$ are the solution of the following characteristic equation:

$$\det(A - a\mathbb{I}) = 0\,, \tag{2.6}$$

where det is the determinant function for matrices. For a Hermitian operator $A \in \mathcal{L}(\mathcal{H})$ on a finite d dimensional Hilbert space $\mathcal{H}$, the eigenvalues $\{a_i\}$ are

real and the eigenstates form an orthonormal basis $\{|a_i\rangle_{i=1}^d\}$. The operator A can also be represented in spectral decomposition form which reads:

$$A = \sum_{i=1}^{d} a_i \, |a_i\rangle\langle a_i| \tag{2.7}$$

We will now discuss the trace function.

Definition 2.6 (Trace)**.** *Given an orthonormal basis $\{|e_i\rangle\}$ for a Hilbert space $\mathcal{H}$, the trace of a Hermitian operator, A, is defined as*

$$\mathrm{Tr}(A) = \sum_i \langle e_i|A|e_i\rangle \ . \tag{2.8}$$

The trace is independent of the choice of orthonormal basis, since if the basis is chosen to be the eigenvectors of A then $\mathrm{Tr}(A)$ is the sum of the eigenvalues of A. We can write A in the form of Eq. (2.7). Thus, for any unitary U, it holds that

$$U^\dagger A U = \sum_i a_i \, |f_i\rangle\langle f_i| \ , \tag{2.9}$$

where $|f_i\rangle = U\,|a_i\rangle$. Note that the set of states $\{|a_i\rangle\}$ are orthonormal as:

$$\begin{aligned}\langle f_i|f_j\rangle &= \langle a_i|\, U^\dagger U\, |a_j\rangle \\ &= \langle a_i|a_j\rangle \\ &= \delta_{ij} \ .\end{aligned}$$

Thus, a_i are the eigenvalues for $U^\dagger A U$ as well as A. This means that for any orthonormal basis $\{|e_i\rangle\}$, there exists a unitary U such that $|e_i\rangle = U\,|a_i\rangle$ so that

$$\begin{aligned}\mathrm{Tr}(A) &= \sum_i \langle e_i|A|e_i\rangle \\ &= \sum_i \langle a_i|U^\dagger A U|a_i\rangle \\ &= \sum a_i \ .\end{aligned}$$

Thus the trace does not depend on the basis $\{|e_i\rangle\}$ used to calculate the trace. The trace is linear such that

$$\mathrm{Tr}(\alpha A + \beta B) = \alpha \,\mathrm{Tr}\, A + \beta \,\mathrm{Tr}\, B \ .$$

Moreover, the trace of a product of operators is invariant under cyclic permutation of the operators, i.e.

$$\mathrm{Tr}(ABC) = \mathrm{Tr}(BCA) = \mathrm{Tr}(CAB)$$

for any $A, B, C \in \mathcal{L}(\mathcal{H})$. However, the trace is not, in general, invariant under non-cyclic permutations.

2.3 Density Operator Formalism

A convenient and practical description of quantum states is provided by the density operator formalism. We call a *density operator* w.r.t. a fixed basis a *density matrix*.

Definition 2.7 (Density operator)**.** *Let $\{|\psi_i\rangle\}$ be a set of quantum states and $\{p_i\}$ a probability distribution for $i = 1, \cdots, n$, i.e. $\sum_{i=1}^{n} p_i = 1$ and $p_i \in [0, 1]$ for all i. The operator*

$$\rho = \sum_{i=1}^{n} p_i \, |\psi_i\rangle \, \langle\psi_i| \tag{2.10}$$

is called a mixed state if $p_i < 1$ for all i. We call ρ a pure state if there exists one index i' for which $p_{i'} = 1$, i.e. $\rho = |\psi\rangle\langle\psi|$.

The matrix representation of ρ is called the density matrix. The set of all density operators on a Hilbert space $\mathcal{H}$ is denoted by $\mathcal{S}(\mathcal{H})$. From Def. 2.7, it follows that $\mathcal{S}(\mathcal{H})$ is generated by a convex combination of all pure density operators $|\psi\rangle\langle\psi|$. Moreover, a convex combination of any density operators, $\rho = \sum_i p_i \rho_i$, is also a density operator and represents a statistical ensemble $\{p_i, \rho_i\}$, where ρ_i may be pure or mixed. Thus a density operator can be characterized by the following theorem.

Theorem 2.1 (Characterization of density operators)**.** *A Hermitian operator ρ is a density operator for some ensemble $\{p_i, \rho_i\}$ (i.e. $\rho = \sum_i p_i \rho_i$) if and only if*

$$\rho \geq 0, \quad \textit{and} \quad \mathrm{Tr}(\rho) = 1.$$

The criterion which determines whether a state ρ is pure or mixed is given by its purity, i.e. $\mathrm{Tr}(\rho^2)$. A state is pure if $\mathrm{Tr}(\rho^2) = 1$, and mixed if $\mathrm{Tr}(\rho^2) < 1$. Since density operators provide an alternate formulation for quantum mechanics, we will state the first postulate of quantum mechanics in terms of density operators.

Postulate 2.1. *Quantum systems associated with a Hilbert space $\mathcal{H}$ are known as the state space. The state of a quantum system can be completely described by the density operator $\rho \in \mathcal{S}(\mathcal{H})$ (i.e. positive operators with trace one on the Hilbert space H). The density operator ρ of a quantum system in a statistical ensemble $\{p_i, \rho_i\}$ reads: $\sum_i p_i \rho_i$.*

2.4 The Qubit, Pauli Operators and Bloch Sphere

In many quantum information applications, the fundamental quantum system is a two-level system called quantum bit or qubit. Physical realizations of qubits are, for example, two orthogonal polarizations of a photon, a photon that can be found in one of two distinct paths or the spin state of spin-$\frac{1}{2}$ particles.

The state space of a qubit is a two-dimensional Hilbert space, $\mathcal{H}_2$. The commonly used basis for $\mathcal{H}_2$ is the computational basis $\{|0\rangle, |1\rangle\}$. Thus, any pure qubit state is represented by a superposition of the form:

$$|\psi\rangle = \alpha\,|0\rangle + \beta\,|1\rangle \quad \text{with} \quad |\alpha|^2 + |\beta|^2 = 1, \quad \alpha, \beta \in C\,. \tag{2.11}$$

Pure quantum states that differ by a global phase factor are physically equivalent, so we can consider the parameterization into the spherical coordinates:

$$|\psi\rangle = \cos\frac{\theta}{2}|0\rangle + e^{i\phi}\sin\frac{\theta}{2}|1\rangle\,, \tag{2.12}$$

with $0 \leqslant \theta \leqslant \pi$ and $0 \leqslant \phi \leqslant 2\pi$. The angles (θ, ϕ) describe a unique mapping of the pure quantum state $|\psi\rangle$ with a point on the surface of a three dimensional sphere, also called *Bloch sphere*. See Fig. 2.1 for visualization.

Any mixed qubit state is represented by a density operator ρ acting on $\mathcal{H}_2$ and can be expressed as a combination of the identity operator $\mathbb{1}$ and the Pauli operators σ_x, σ_y and σ_z :

$$\rho = \frac{\mathbb{1} + \vec{r}\cdot\vec{\sigma}}{2}, \quad \vec{r} \in \mathbb{R}^3 : \left\|\vec{r}\right\| \leq 1 \tag{2.13}$$

where $\vec{\sigma} = \left(\sigma_x, \sigma_y, \sigma_z\right)^T$. The matrix representation of the Pauli operators with respect to the computational basis reads:

$$\sigma_x = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \quad \sigma_y = \begin{bmatrix} 0 & -\mathrm{i} \\ \mathrm{i} & 0 \end{bmatrix}, \quad \sigma_z = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \tag{2.14}$$

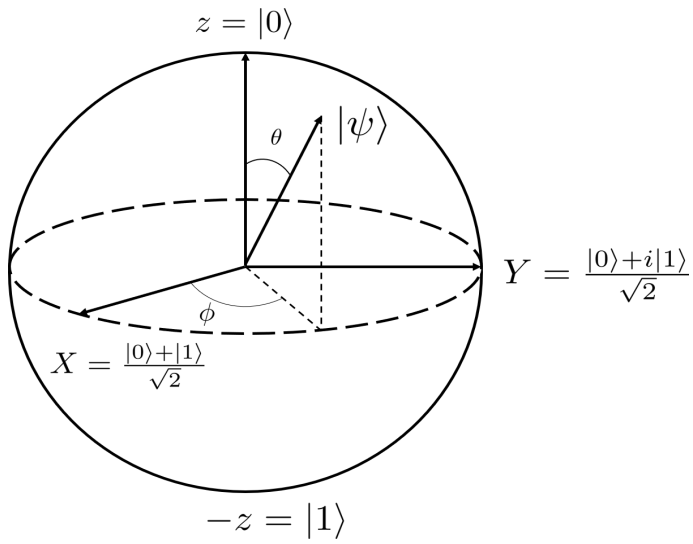

Figure 2.1: Representation of a state $|\psi\rangle$ in a Bloch sphere.

Note that, depending on the context, we also indicate the Pauli operators as $\sigma_x = \sigma_1 = X, \sigma_y = \sigma_2 = Y$ and $\sigma_z = \sigma_3 = Z$. The Pauli operators are hermitian, traceless operators with eigenvalues ±1, and satisfy the following relation:

$$\sigma_i \sigma_j = \delta_{i,j} \mathbb{1} + \sum_{k=1}^{3} \varepsilon_{ijk} \sigma_k \tag{2.15}$$

where ε_{ijk} is the Levi-Civita symbol, which is equal to +1 (-1) if the triple (i, j, k) is a cyclic (anti-cyclic) permutation of $(1, 2, 3)$, and zero otherwise.

The representation of Eq. (2.13) has great utility in many computations. The vector $\left\|\vec{r}\right\|$ individuates a point inside the Bloch sphere. The purity of a qubit state can be readily computed as: $\mathrm{Tr}\left[\rho^2\right] = \left(1 + \left\|\vec{r}\right\|^2\right)/2$. Thus, the norm of the vector r indicates whether the state is pure or mixed. For pure state, $\left\|\vec{r}\right\| = 1$ and it lies on the surface of the Bloch sphere. For a mixed state, $\left\|\vec{r}\right\| < 1$, and it resides in the interior of the Bloch sphere. When $\left\|\vec{r}\right\| = 0$, the state is said to be maximally mixed and located at the centre of the Bloch sphere.

2.5 Quantum Measurement

Some sort of interaction is required to extract information from a physical system. A measurement describes the interaction of an apparatus with the quantum system under study. It builds a bridge between quantum states on one side and classical outcomes on the other. Here in this section, we

describe a quantum measurement [NC10].

Postulate 2.2 (Quantum Measurement)**.** *A quantum measurement of a physical quantity of a system in quantum state ρ is described by a set of measurement operators $\{M_m\}$, satisfying the completeness relation $\sum_m M_m^\dagger M_m = \mathbb{1}$. The index m refers to the measurement outcome that may occur in the experiment. The probability of observing the measurement outcome m is given by*

$$\Pr(m) = \mathrm{Tr}\left(M_m^\dagger M_m \rho\right) , \tag{2.16}$$

and the state of the system after the measurement is

$$\rho_m = \frac{M_m \rho M_m^\dagger}{\mathrm{Tr}\left(M_m^\dagger M_m \rho\right)} . \tag{2.17}$$

Postulate 2.2 provides the most general description of a quantum measurement. There are two special cases of quantum measurements which are of particular interest in quantum information.

2.5.1 Projective Measurement

An observable is a Hermitian operator on the state space of the observed system. The projective measurement can be specified by an observable M with a spectral decomposition

$$M = \sum_m m P_m , \tag{2.18}$$

where P_m is the projector onto the eigenspace of the observable M with eigenvalue m. The possible outcomes m of the measurement correspond to the eigenvalues of the observable. Projective measurements can be understood as a special case of Postulate 2.2. Suppose the measurement operators in Postulate 2.2, in addition to satisfying the completeness relation $\sum_m P_m^\dagger P_m = I$, also satisfy the conditions that P_m are orthogonal projectors, i.e. the P_m are Hermitian: $P_m^\dagger = P_m$, and $P_m P_n = \delta_{m,n} P_n$. If the projectors are all rank-one $P_m = |m\rangle\langle m|$, the measurement is called a *von Neumann measurement*.
The average values for projective measurement outcomes can be directly written as:

$$\langle M \rangle := \sum_m m \Pr(m) = \sum_m m \mathrm{Tr}\left[P_m \rho\right] = \mathrm{Tr}[M\rho] . \tag{2.19}$$

2.5.2 Positive Operator-valued Measurement (POVM)

A positive operator-valued measurement (POVM) is defined by a set of positive operators $\{E_m\}$, $E_m \geq 0$ acting on the state space following the completeness relation: $\sum_m E_m = \mathbb{1}$. Then the probability of obtaining outcome m when measuring the system in state ρ is given by:

$$\Pr(m) = \mathrm{Tr}\left[E_m \rho\right] . \tag{2.20}$$

POVMs are a special case of Postulate 2.2. When the measurement operators are given by $M_m = \sqrt{E_m}$, which implies $M_m^\dagger M_m = E_m$, the Postulate 2.2 and the definition of POVM coincides. Note that POVM measurements also include projective measurements, i.e. for $M_m = P_m$ with $P_m P_n = \delta_{m,n} P_m$ we find $E_m = P_m$.
Finally, we remark that, although projective measurements are particular cases of POVMs, any POVM measurement on a d-dimensional Hilbert space $\mathcal{H}$ can be expressed as a projective measurement on a Hilbert space $\mathcal{H}'$ of higher dimension d', i.e. $d < d'$. This result is known as the *Naimark extension* [DJR05, Per06, Par12].

2.6 Composition of Quantum System, Entanglement and Separability

So far, we only consider a single quantum system. Now, we will discuss the composite system achieved via the *tensor* or *Kronecker* product.

Postulate 2.3 (Composite system)**.** *Let $\mathcal{H}_i$ be the state space of the i^{th} subsystem for $i = 1, \cdots, n$. Then the state space of the composite system $\mathcal{H}$ is given by*

$$\mathcal{H} = \mathcal{H}_1 \otimes \mathcal{H}_2 \otimes \cdots \otimes \mathcal{H}_n$$

The Postulate 2.3 allows us to introduce the concept of separability and entanglement, which plays a vital role in many quantum information protocols. Consider a bipartite scenario where Alice (Bob) prepares the pure state $\rho_A \in \mathcal{S}(\mathcal{H}_A)$ ($\rho_B \in \mathcal{S}(\mathcal{H}_B)$). The composite system $\rho_{AB} \in \mathcal{S}(\mathcal{H}_A \otimes \mathcal{H}_B)$ will be called a product state if the global state can be written as

$$\rho_{AB} = \rho_A \otimes \rho_B . \tag{2.21}$$

In terms of quantum state $|\psi_A\rangle \in \mathcal{H}_A$ belongs to Alice and $|\psi_B\rangle \in \mathcal{H}_B$ belongs

to Bob, the composite state $|\psi_{AB}\rangle \in \mathcal{H}_{AB}$ is a product state if the global state can be written as

$$|\psi_{AB}\rangle = |\psi_A\rangle \otimes |\psi_B\rangle \ . \tag{2.22}$$

For brevity, $|\psi_A\rangle \otimes |\psi_B\rangle$ will be denoted as $|\psi_A\psi_B\rangle$. Such product state can be prepared independently by Alice and Bob by means of *Local Operation and Classical Communication*. This statement does not hold for the pure non-product state, i.e. they cannot be prepared locally and will show some degree of correlation when measured. In the more complex case of mixed states, suppose Alice and Bob locally prepare the state $\rho_A^i \in \mathcal{S}(\mathcal{H}_A)$ and $\rho_B^i \in \mathcal{S}(\mathcal{H}_B)$ with some probability p_i. The state of the bipartite system will be called a product state if it is a convex combination of the product states, i.e. if it can be written as:

$$\rho = \sum_i p_i \rho_A^i \otimes \rho_B^i \, . \tag{2.23}$$

Definition 2.8 (Separability and Entanglement)**.** *A quantum state $\rho_{AB} \in \mathcal{S}(\mathcal{H}_A \otimes \mathcal{H}_B)$ is called separable if there exists a convex combination of pure product states $\left|\psi_A^i\right\rangle \otimes \left|\psi_B^i\right\rangle$, with $\left|\psi_A^i\right\rangle \in \mathcal{H}_A$ and $\left|\psi_B^i\right\rangle \in \mathcal{H}_B$, such that:*

$$\rho_{AB} = \sum_i p_i \left|\psi_A^i, \psi_B^i\right\rangle \left\langle \psi_A^i, \psi_B^i\right| \, . \tag{2.24}$$

Otherwise, ρ_{AB} is called entangled.

Examples of bipartite qubit entangled states are the Bell states

$$\begin{aligned} \left|\phi^{\pm}\right\rangle &= \frac{1}{\sqrt{2}} \left(|00\rangle \pm |11\rangle\right) \, , \\ \left|\psi^{\pm}\right\rangle &= \frac{1}{\sqrt{2}} \left(|01\rangle \pm |10\rangle\right) \, . \end{aligned} \tag{2.25}$$

They form a maximally entangled basis, known as the Bell basis, of four dimensional Hilbert space of two qubits. A generalization of these Bell states in the multipartite scenario is represented by *Greenberger-Horne-Zeilinger* (GHZ) states [HHHH09].

Theorem 2.2 (Schimdt decomposition)**.** *Let $|\psi_{AB}\rangle \in \mathcal{H}_A \otimes \mathcal{H}_B$ be the pure state of a bipartite system. Then there exists an orthonormal basis $\{|\alpha_i\rangle\}$, $i = \{1, \cdots, d_A\}$,*

of $\mathcal{H}_A$ and an orthonormal basis $\{|\beta_j\rangle\}$, $j = \{1, \cdots, d_B\}$, of $\mathcal{H}_B$ such that:

$$|\psi_{AB}\rangle = \sum_{k=1}^{d} \lambda_k |\alpha_k, \beta_k\rangle \tag{2.26}$$

where λ_k are positive real coefficients called Schmidt coefficients and $d \leq \min(d_A, d_B)$ is the Schmidt rank.

Proof. See [NC10] for detailed proof. □

The number of non-zero Schmidt coefficients is called the *Schmidt rank*. If the Schmidt rank is one, the state $|\psi_{AB}\rangle$ is separable. A state is entangled if and only if its Schmidt rank is strictly bigger than 1. Given a state ρ_{AB} of a composite system, a natural question that arises is how the state of the respective subsystems ρ_A and ρ_B can be accessed. To answer this question, we introduce the reduced density operator.

Definition 2.9 (Reduced density operator)**.** *Let ρ_{AB} be the state of a bipartite quantum system composed of two Hilbert spaces $\mathcal{H}_A$ and $\mathcal{H}_B$. Then the reduced density operator representing the state on subsystem A is given by:*

$$\rho_A = \mathrm{tr}_B[\rho_{AB}]\,, \tag{2.27}$$

where tr_B denotes the partial trace on subsystem B.

For a pure two-qubit state $\rho_{\mathrm{AB}} = \left|\phi^+\right\rangle \left\langle\phi^+\right|$, with $\left|\phi^+\right\rangle := \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle)$, the reduced density matrix of system A is given by

$$\begin{aligned} \rho_{\mathrm{A}} &= \mathrm{tr}_B[\rho_{AB}] \\ &= \sum_{i=0}^{1} (\mathbb{1} \otimes \langle i|_B) \rho_{\mathrm{AB}} (\mathbb{1} \otimes |i\rangle_B) \\ &= \frac{1}{2}(|0\rangle\langle 0| + |1\rangle\langle 1|) \\ &= \frac{1}{2}\mathbb{1}\,. \end{aligned} \tag{2.28}$$

Hence, the partial trace of a pure state can result in a completely mixed state. This example leads us to a general statement regarding the inverse transformation. Remarkably, every mixed state can be considered as the reduced state of a pure entangled state in a larger system. This process is called purification and is widely used in quantum cryptography.

Theorem 2.3 (Purification). *Let ρ_A on $\mathcal{H}_A$ be the state of a quantum system A. Then there exists an auxiliary system E with state space $\mathcal{H}_E$ and a pure state $|\psi_{AE}\rangle \in \mathcal{H}_A \otimes \mathcal{H}_E$, called a purification of ρ_A, such that:*

$$\mathrm{Tr}_E\left[\, |\psi_{AE}\rangle \langle\psi_{AE}| \,\right] = \rho_A \tag{2.29}$$

Proof. See [NC10] for detailed proof. □

Note that all purifications $|\psi_{AE}\rangle$ of ρ_A are related by unitaries on E.

2.7 Quantum Channels

Quantum channels (also called quantum operations) describe the general evolution of a quantum state. The following provides a prescription of the characterization of such evolution of a quantum state.

Postulate 2.4. *The evolution of a closed quantum system is determined by a unitary transformation U. If $\rho(t_1)$ is the state of a system at time t_1, and $\rho(t_2)$ is the state of a system at time t_2, then $\rho(t_1)$ and $\rho(t_2)$ follows:*

$$\rho(t_2) = U\rho(t_1)U^\dagger\,, \tag{2.30}$$

where U is the unitary operator on $\mathcal{H}$ following $UU^\dagger = U^\dagger U = \mathbb{1}$, with $\mathbb{1}$ being the identity operator on $\mathcal{H}$.

The final state of a quantum system ρ after a quantum operation $\mathcal{E}$ is given by $\mathcal{E}(\rho)$. A quantum channel $\mathcal{E}$ needs to obey the following axiomatic properties [NC10]:

- The probability with which the process described by ε occurs is specified by $\mathrm{Tr}(\mathcal{E}(\rho))$. Thus, for any state ρ, $0 \leqslant \mathrm{Tr}(\mathcal{E}(\rho)) \leqslant 1$.
- $\mathcal{E}$ is a convex linear map over the set of density operators. That translates to, for a set of probabilities $\{p_i\}$,

$$\mathcal{E}\left(\sum_i p_i\rho_i\right) = \sum_i p_i\mathcal{E}(\rho_i)\,.$$

- $\mathcal{E}$ is a completely positive (CP) map. This means, for every input state ρ, $\mathcal{E}(\rho)$ must be positive. Furthermore, for a composite state $\rho_{AB} \in \mathcal{S}(\mathcal{H}_A \otimes \mathcal{H}_B)$, the operator $\left(\mathbb{1}_A \otimes \mathcal{E}\right)(\rho_{AB}) \in \mathcal{S}(\mathcal{H}_A \otimes \mathcal{H}_B)$ is also positive.

The first property represents the quantum measurement as a quantum operation. The normalized state after the quantum operation is defined as $\mathcal{E}(\rho)/\mathrm{Tr}\big(\mathcal{E}(\rho)\big)$. A quantum channel is called trace-preserving if $\mathrm{Tr}\big(\mathcal{E}(\rho)\big) = \mathrm{Tr}(\rho) = 1$, and non-trace preserving if $\mathrm{Tr}\big(\mathcal{E}(\rho)\big) < 1$. The second property stems from the Postulate 2.1. If a system is in one of state from the ensemble $\{p_i, \rho_i\}$, after applying the quantum information $\mathcal{E}$, it will be in the ensemble $\{p_i, \mathcal{E}(\rho_i)\}$. It specifies that the evolution of ρ is consistent with each subsystem ρ_i. The third property ensures that a quantum channel produces a valid density operator even when acting on a subsystem of a larger, possibly entangled, system.
Quantum channels can be presented in an elegant way using Kraus operators.

Theorem 2.4 (Kraus Theorem)**.** *A quantum operation $\mathcal{E}(\rho)$, $\rho \in \mathcal{S}(\mathcal{H})$, following all three properties stated above can be represented by linear operators K_i satisfying*

$$\sum_i K_i^\dagger K_i \leq \mathbb{1}\,, \tag{2.31}$$

such that

$$\mathcal{E}(\rho) = \sum_i K_i \rho K_i^\dagger\,. \tag{2.32}$$

$\{K_i\}$'s are called the Kraus operators. The number of Kraus operators cannot be larger than d^2, where d is the dimension of $\mathcal{H}$.

The proof of the theorem can be found in [NC10]. Kraus decomposition of a quantum channel is not unique. However, it provides us with an analytical expression for a generic quantum channel.

2.7.1 Depolarizing Noise

An important example of a quantum channel is the depolarizing noise channel. Consider the depolarizing noise channel as $\mathcal{E}_{depol}$ acting on $\rho \in \mathcal{S}(\mathcal{H})$. Then the state after the operation $\mathcal{E}_{depol}(\rho)$ reads:

$$\mathcal{E}_{depol}(\rho) = (1-p)\rho + \frac{p}{3}\sum_{i=1}^{3} \sigma_i \rho \sigma_i^\dagger\,, \tag{2.33}$$

with Kraus operators

$$\begin{aligned} k_0 &= \sqrt{1-p}\mathbb{1}\,, \\ k_i &= \sqrt{p/3}\sigma_i \quad \text{for} \quad i \in \{1,2,3\}\,, \end{aligned} \tag{2.34}$$

and $p \in [0,1]$ denotes the noise parameter. The resulting state of Eq. (2.33) can be interpreted as follows: the state ρ is unchanged with probability $1-p$ or is affected by one of the qubit errors with probability $\frac{p}{3}$ each. In many quantum information protocols, applications of $\mathcal{E}_{depol}$ on the qubits are used to test the robustness of the protocol against noise.

2.8 Entropy

Entropy measures are fundamental tools in classical and quantum information. It quantifies the randomness, i.e. it measures how much uncertainty there is in the state of a physical system from the observer's perspective. Here we review some entropy measures used in the remainder of this thesis.

Definition 2.10 (Shannon Entropy)**.** *Let $\{p_x\}$ be a probability distribution of a random variable X with possible outcomes $\{x\}$. Then the Shannon entropy of X (or of the distribution $\{p_x\}$) is given by:*

$$H(X) = H\left(\{p_x\}\right) = -\sum_x p_x \log p_x \tag{2.35}$$

In the definition (and throughout this thesis), logarithms indicated by 'log' are taken to base two, while 'ln' indicates a natural logarithm. $H(X)$ quantifies how much information we gain, on average, after measuring/learning the value of X. An alternative view is that $H(X)$ measures the uncertainty about X before we learn its value.
For a binary-valued random variable X, the Shannon entropy is called binary entropy and reads:

$$h(p) := -p\log p - (1-p)\log(1-p)\,. \tag{2.36}$$

Given two random variables X (with possible outcomes $\{x\}$) and Y (with possible outcomes $\{y\}$) jointly distributed according to $\{p(x,y)\}$, the joint

Shannon entropy of X and Y defined as:

$$H(XY) = \sum_{x,y} p(x,y) \log p(x,y)\,, \tag{2.37}$$

and the conditional entropy reads:

$$H(X|Y) = H(XY) - H(Y)\,. \tag{2.38}$$

The conditional entropy of X given Y quantifies how uncertain we are, on average, is about the value of X, given that we learned the value of Y. Since the uncertainty on both random variables X and Y is greater than the uncertainty on Y, the conditional entropy follows $H(X|Y) \geq 0$. Finally, the mutual information $H(X:Y)$ measures the amount of information we gain on X by observing the value of Y.

$$H(X:Y) = H(X) - H(X|Y) = H(X) + H(Y) - H(XY)\,. \tag{2.39}$$

Shannon's noiseless coding theorem provides an operational interpretation of Shannon's entropy. If a source is emitting a sequence of independent and identically distributed random variables $X_1, X_2, \cdots, X_n$ according to a probability distribution P_X, asymptotically the amount of bits needed per source to store the data by encoding it in a bitstring without losing information is given by $H(X)$. If $\ell^{\varepsilon}_{\text{compr}}(X)$ is the minimum length (measured in terms of bits) needed to compress X such that X can be recovered without losing information, except for an error probability ε, then the compression rate

$$r_{\text{compr}}(X) := \lim_{\varepsilon \to 0} \lim_{n \to \infty} \frac{\ell^{\varepsilon}_{\text{compr}}\ (X_1 X_2 \cdots X_n)}{n} \tag{2.40}$$

of the example is equal to $H(X)$ [KRS09]. While the Shannon entropy measures the uncertainty associated with a classical probability distribution, the von Neumann entropy deals with the quantum states described by the density operators.

Definition 2.11 (von Neumann Entropy)**.** *The von Neumann entropy of a quantum state ρ is defined as:*

$$H(\rho) = -\text{Tr}[\rho \log \rho]\,. \tag{2.41}$$

If $\{\lambda_i\}$ are the eigenvalues of ρ, then von Neumann entropy can be written as

$$H(\rho) = -\sum_i \lambda_i \log \lambda_i \,. \tag{2.42}$$

Often, the von Neumann entropy of a system A of state ρ is denoted as $H(A)_\rho$. One can interpret the von Neumann entropy of ρ as the Shannon entropy of the probability distribution defined by its eigenvalues.
Additionally, von Neumann entropy satisfies the following properties.

- For every state ρ in d dimensional Hilbert space, $0 \leq S(\rho) \leq \log_2 d$. $H(\rho) = 0$ for a pure state and $H(\rho) = \log_2 d$ if the state is maximally mixed.

- *Additivity*: Let $\mathcal{H}_A$ and $\mathcal{H}_B$ be two Hilbert spaces. Suppose there are density operators $\rho \in \mathcal{H}_A$ and $\sigma \in \mathcal{H}_B$. Then additivity of the entropy implies
$$H(\rho \otimes \sigma) = H(\rho) + H(\sigma).$$

- *Subadditivity*: Let ρ_{AB} be a bipartite state on tensor product Hilbert space $\mathcal{H}_A \otimes \mathcal{H}_B$. Then subadditivity is described as
$$H\left(\rho_{AB}\right) \leq H\left(\rho_A\right) + H\left(\rho_B\right).$$

- *Concavity*: Let $\{p_i\}$ be a probability distribution, i.e. $0 \leq p_i \leq 1$ and $\sum_i p_i = 1$. Then the concavity is described as
$$H\left(\rho\right) \geq \sum_i p_i H\left(\rho_i\right),$$
where $\rho = \sum_i p_i \rho_i$.

- *Strong subadditivity*: Let $\rho_{ABC} \in \mathcal{H}_A \otimes \mathcal{H}_B \otimes \mathcal{H}_C$, the strong subadditivity is described as
$$H(\rho_{ABC}) + H(\rho_B) \leq H(\rho_{AB}) + H(\rho_{BC})$$

2.8.1 Holevo Bound

Now consider two parties, Alice and Eve. Suppose Alice has a random variable X with values $\{x_i\}$ according to a probability distribution $\{p_i\}$. Alice then prepares a quantum state, chosen from a set $\{\rho_i\}$ according to

$\{p_i\}$, and gives the state to Eve. Eve's task is to access the value x_i of X. Eve performs a measurement described by POVM elements $\{E_y\}$ on the states and obtains Y. The amount of accessible information, i.e. the amount of information that Eve can get about the variable X, is the maximum value of the mutual information $H(X:Y)$ over all the possible measurements that Eve can do. For any measurement, her accessible information is upper bounded by the Holevo quantity $\chi(X:Y)$ [Hol73], i.e.

$$H(X:Y) \leqslant \chi(X:Y) := H(\rho) - \sum_i p_i H(\rho_i) \ , \tag{2.43}$$

where $\rho = \sum_i p_i \rho_i$.

2.9 Min- and Max-Entropy

Min- and max- entropy was introduced in [Ren08]. It plays a crucial role in the security of quantum cryptographic protocols.

Definition 2.12 (Min-entropy [Tom15, KRS09]). *Let ρ_{AB} be a bipartite density operator. The min-entropy of A conditioned on B is defined by*

$$H_{\min}(A|B)_\rho := -\log \min \left[\mathrm{Tr}(\sigma_B) : \sigma_B \geq 0, (id_A \otimes \sigma_B) - \rho_{AB} \geq 0 \right] . \tag{2.44}$$

An operational interpretation of the min-entropy, proposed in [KRS09], suggests the importance of this entropy measure for quantum cryptography. Consider the classical quantum state

$$\rho_{AE} = \sum_a p(a) |a\rangle\langle a| \otimes \rho_E^a \ , \tag{2.45}$$

Consider the classical-quantum state where $\{|a\rangle\}$ forms an orthonormal basis for system A and can represent a classical random variable A that assumes value a with probability $p(a)$, and ρ_E^a is a general quantum state on system E that may depend on the specific value of a. The guessing probability, $p_{\text{guess}}(A|E)$, is the optimal probability with which someone that has access to system E can correctly guess the value of the variable A :

$$p_{\text{guess}}(A|E)_\rho = \sup_{\{M_E^a\}} \sum_a p(a) \mathrm{Tr}\left(M_E^a \rho_E^a\right) \ , \tag{2.46}$$

where the supremum is over all possible measurements, described by the set of POVMs $\{M_E^a\}$ on the system E. It was shown in [KRS09] that, similarly

to the classical case, the conditional min-entropy $H_{\min}(A|E)$ of a classical variable A is directly related to the guessing probability $p_{\text{guess}}(A|E)_\rho$:

$$H_{\min}(A|E)_\rho = -\log p_{\text{guess}}(A|E)_\rho \,. \tag{2.47}$$

Definition 2.13 (Max-entropy [Tom15, KRS09]). *Let ρ_{AB} be a bipartite density operator and let ρ_{ABC} be a purification of ρ_{AB}. Then max-entropy of A conditioned on B of the state ρ_{AB} is also defined as:*

$$H_{\max}(A|B)_\rho = -H_{\min}(A|C)_\rho \tag{2.48}$$

In general, the relation between min- and max-entropy and von Neumann entropy of a bipartite density operator ρ_{AB} reads:

$$H_{\min}(A|B) \leq H(A|B) \leq H_{\max}(A|B) \,. \tag{2.49}$$

2.10 Smooth Min- and Max-Entropy

In quantum information processing and cryptographic tasks, one must deal with ϵ-error probability. Thus, ϵ-smooth versions of min- and max- entropies are introduced. The smooth entropies of a state ρ are defined as optimizations over the min- and max-entropies of states $\tilde{\rho}$ that is close to ρ in the purified distance. Thus, we define the trace distance, and the purified distance [Tom15, NC10] between two quantum states.

Definition 2.14 (Trace distance [Tom15]). *The trace distance between two density operators $\rho, \sigma \in \mathcal{S}(\mathcal{H})$ is defined as:*

$$\delta(\rho, \sigma) = \|\rho - \sigma\|_{\text{Tr}} = \frac{1}{2} \|\rho - \sigma\|_1 \,, \tag{2.50}$$

where $\|X\|_1 = \text{Tr}\left(|X|\right) = \text{Tr}\left(\sqrt{X^\dagger X}\right)$.

The definition of trace distance can be extended to sub-normalized states (positive operators with trace smaller or equal to 1). For two sub-normalized density operators $\hat{\rho}$ and $\hat{\sigma}$, the generalized trace distance reads:

$$\delta\left(\hat{\rho}, \hat{\sigma}\right) = \|\hat{\rho} - \hat{\sigma}\|_{\text{Tr}} = \frac{1}{2} \|\hat{\rho} - \hat{\sigma}\|_1 + \frac{1}{2} \left|\text{Tr}\left(\hat{\rho} - \hat{\sigma}\right)\right| \tag{2.51}$$

The trace distance is related to the distinguishability between two states. The probability of distinguishing between two states ρ and σ with a single

measurement is bounded by $\frac{1}{2}(1+\delta(\rho,\sigma))$.

Definition 2.15 (Purified distance [Tom15]). *For two sub-normalized density operators $\hat{\rho}$ and $\hat{\sigma}$, the purified distance reads:*

$$\mathcal{D}(\hat{\rho},\hat{\sigma}) := \sqrt{1-\overline{F}(\hat{\rho},\hat{\sigma})}\,, \tag{2.52}$$

where $\overline{F}$ is the generalized fidelity

$$\overline{F}(\rho,\sigma) := \left(\mathrm{Tr}(\sqrt{\sqrt{\hat{\rho}}\hat{\sigma}\sqrt{\hat{\rho}}}) + \sqrt{(1-\mathrm{Tr}\hat{\rho})(1-\mathrm{Tr}\hat{\sigma})}\right)^2 . \tag{2.53}$$

Now, We can define the smooth entropies.

Definition 2.16 (Smoothed min and max-entropy [Tom15, VDTR13]). *For a quantum state ρ_{AB} and $\epsilon \geq 0$, the smooth min-entropy of system A conditioned on B is defined as*

$$H_{\min}^{\epsilon}(A|B) := \max_{\tilde{\rho}_{AB}\in\mathcal{B}^{\epsilon}(\rho_{AB})} H_{\min}(A|B)_{\tilde{\rho}_{AB}}\,, \tag{2.54}$$

and, the smooth max-entropy of system A conditioned on B is defined as

$$H_{\max}^{\epsilon}(A|B) := \min_{\tilde{\rho}_{AB}\in\mathcal{B}^{\epsilon}(\rho_{AB})} H_{\max}(A|B)_{\tilde{\rho}_{AB}}\,. \tag{2.55}$$

Sub-normalized operators are also taken into account in the optimization. $\mathcal{B}^{\epsilon}$ is an ϵ-ball of sub-normalized operators around the state ρ_{AB} defined in terms of the purified distance, i.e.

$$\mathcal{B}^{\epsilon}(\rho) = \{\tilde{\rho}\geq 0 : \mathrm{Tr}\,(\tilde{\rho})\leq 1 \text{ and } \mathcal{D}\,(\rho,\tilde{\rho})\leq\epsilon\} \tag{2.56}$$

Similar to Eq. (2.48), the duality relation also exists for smooth min- and max- entropies [KRS09, Tom15]. Let ρ_{ABC} be a pure quantum state, then

$$H_{\max}^{\epsilon}(A|B)_{\rho} = -H_{\min}^{\epsilon}(A|C)_{\rho}\,. \tag{2.57}$$

The smooth min- and max- entropies satisfy some important properties. One important property is the data-processing inequality [Tom15], which reads:

$$H_{\min}^{\epsilon}(A|B)_{\rho} \leq H_{\min}^{\epsilon}\,(A|B')_{\tau}\ , \tag{2.58}$$

$$H_{\max}^{\epsilon}(A|B)_{\rho} \leq H_{\max}^{\epsilon}\,(A|B')_{\tau}\ , \tag{2.59}$$

where $\tau_{AB'} = I_A \otimes \mathcal{E}(\rho_{AB})$ and $\mathcal{E}$ is a CPTP (B, B') channel. The data-processing inequality states that if we process the quantum side information B through a CP trace-preserving map $\mathcal{E}$, we always increase our uncertainty on A. Another important property is the asymptotic equipartition property (AEP) which links the smooth entropies to the Shannon/von Neumann entropy:

$$\begin{aligned} \lim_{\epsilon\to 0}\lim_{n\to\infty}\frac{1}{n}H^{\epsilon}_{\min}(A^n|B^n)_{\rho^{\otimes n}} &= H(A|B)_\rho\,, \\ \lim_{\epsilon\to 0}\lim_{n\to\infty}\frac{1}{n}H^{\epsilon}_{\max}(A^n|B^n)_{\rho^{\otimes n}} &= H(A|B)_\rho\,. \end{aligned} \tag{2.60}$$

Here we see that the smooth min- and max-entropies converge to the von Neumann entropy in the limit of several copies of a quantum state.

Now, we will discuss the operational interpretation of smooth entropies and their implication in quantum cryptographic schemes. Recall that, for a random variable X and $\varepsilon \geq 0$, $\ell^{\varepsilon}_{\text{compr}}(X)$ (see Eq. (2.40)) is the minimum length of encoding of the random variable X, from which the value of X can be recovered with probability at least $1-\varepsilon$ without losing any information. This quantity is actually equal to the ε-smooth max-entropy of X

$$\ell^{\varepsilon}_{\text{compr}}(X) = H^{\varepsilon'}_{\max}(X) + O(\log 1/\varepsilon) \tag{2.61}$$

for some $\varepsilon' \in \left[\frac{\varepsilon}{2}, 2\varepsilon\right]$. This result generalizes Shannon's noiseless coding theorem Eq. (2.40) to a scenario where the number of realizations of X is finite. Shannon's theorem can be recovered as an asymptotic limit of the Eq. (2.61) for X consisting of many independent and identically distributed pieces $X_1, \ldots, X_n$, i.e.

$$r_{\text{compr}}(X) = \lim_{\varepsilon\to 0}\lim_{n\to\infty}\frac{\ell^{\varepsilon}_{\text{compr}}\,(X_1\cdots X_n)}{n} \tag{2.62}$$

$$= \lim_{\varepsilon\to 0}\lim_{n\to\infty}\frac{1}{n}H^{\varepsilon}_{\max}(X_1\cdots X_n) \tag{2.63}$$

$$= H(X)\,. \tag{2.64}$$

Here we employ Eq. (2.40) for the first, Eq. (2.61) for the second and asymptotic equipartition property of Eq. (2.60) for the last equality.
Consider a cryptographic scenario where two parties, namely Alice and Bob, want to establish a shared secret key over a noisy channel. In the presence of noise, Bob has access to the probability distribution $P(X|Y)$ of the possible keys X held by Alice conditioned on his noisy side information Y. In a

quantum key distribution (QKD) protocol, Alice and Bob perform an error correction (EC) protocol. After the EC step, Alice and Bob share the same secret key with high probability. The result of Eq. (2.61) can be used in this scenario. During EC, Alice sends $\ell^{\varepsilon}_{\text{compr}}(X|Y)$ amount of information to Bob that allows him to correctly guess her key, with an error probability ε. This information is equal to $\ell^{\varepsilon}_{\text{compr}}(X|Y) \approx H^{\varepsilon'}_{\max}(X|Y)$ [Gra21].

Consider the same cryptographic scenario as above. Suppose that an eavesdropper, Eve, has access to quantum side information E correlated with Alice's key X. The goal is to extract a secure key $f(X)$. In a cryptographic context, $f(X)$ is distributed uniformly relative to the side information E held by Eve. In QKD protocol, this process is known as privacy amplification. The maximum number of uniform and independent bits that can be extracted from X is directly given by the smooth min-entropy of X. Precisely, if $\ell^{\varepsilon}_{\text{extr}}(X|E)$ is the maximum length of $f(X)$, computed from X and which is ε-close to a bitstring Z uniform and independent of E [KRS09], then it holds:

$$\ell^{\varepsilon}_{\text{extr}}(X|E) = H^{\varepsilon'}_{\min}(X|E) + O(\log 1/\varepsilon)\,. \tag{2.65}$$

Both error correction and privacy amplification are fundamental tasks in any QKD protocol. We will discuss these processes in later chapters and express the final secret key length in terms of smooth min- and max- entropy.

3

Bell Inequalities and Nonlocality

Bell's theorem [Bel64, BB04] states that quantum physics is incompatible with local hidden-variable (LHV) theories [EPR35]. Via the LHV assumptions, Bell derived inequalities consisting of correlator functions bounded in any LHV theory [Bel64]. A violation of such bounds by any correlations unequivocally proves their non-classical nature of them. Quantum theory allows for such correlations and therefore contradicts at least one of the assumptions of an LHV theory. The merit of Bell inequalities lies in their ability to identify nonlocal correlations.

This chapter is structured as follows. We start with discussing different types of correlation in Sec. 3.1. In Sec. 3.2, we discuss Bell inequalities and review the most famous Bell inequality, the CHSH inequality in Sec. 3.3. Then we introduce a numerical tool that derives a Bell inequality from the measurement outcomes of an experiment in Sec. 3.4. Finally, we finish the chapter by introducing a numerical tool called NPA hierarchy to characterize the quantum set in Sec. 3.5. We also provide a brief introduction to semi-definite programming in that context.

3.1 Classical, No-signalling, and Quantum Correlations

In this section, we present the mathematical characterization of different kinds of correlations. Following [BCP+14], we consider a bipartite setting of two distant observers, i.e. Alice and Bob. Alice and Bob perform measurements on a shared physical system (e.g. an entangled particle) using the

measurement device in their possession. Each party selects locally an input (a measurement setting) that produces an output. Abstractly describing the circumstance, we can say that Alice and Bob have access to a *black box*. Each side chooses the inputs locally, and the box generates an output. This scenario is referred to as a Bell scenario. Suppose, Alice performs measurement

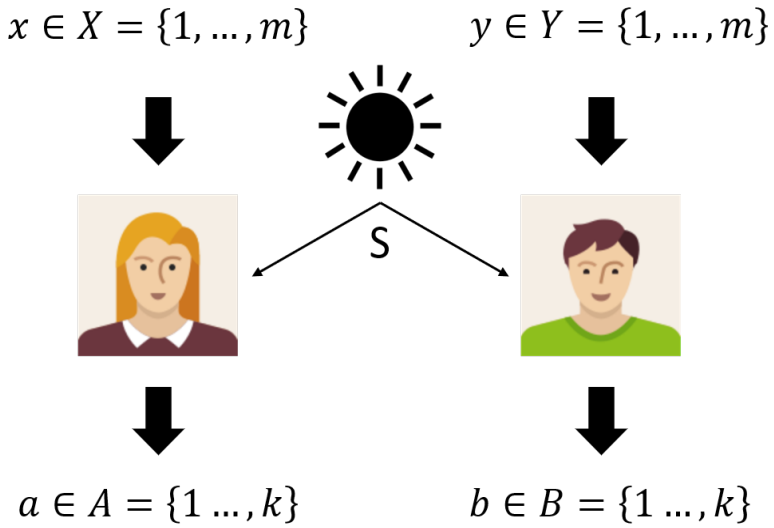

Figure 3.1: Schematic description of a Bell scenario consisting of two parties, Alice and Bob. A source S repeatedly distributes a state to both parties, which perform measurements on their share of the global state specified by an input $x \in X = \{1, 2, \cdots, m\}$ (for Alice) and $y \in Y \in \{1, 2, \cdots, m\}$ (for Bob). Each measurement yields one of k different outcomes $a \in A = \{1, 2, \cdots, k\}$ (for Alice) and $b \in B = \{1, 2, \cdots, k\}$ (for Bob). This scenario is denoted as $[m, k]$ Bell scenario.

specified by inputs $x \in X = \{1, \cdots, m\}$ and Bob performs the measurement denoted by $y \in Y = \{1, \cdots, m\}$. The measurements produce the outputs $a \in A = \{1, \cdots, k\}$ (for Alice) and $b \in B = \{1, \cdots, k\}$ (for Bob), see Fig. 3.1 for visualization. We denote this scenario as $[m, k]$ Bell scenario. The Bell setting is completely characterized by the set $\mathbf{P} := \{P(ab|xy)\} \subset \mathbb{R}^{m^2k^2}$ of all joint conditional probabilities, which we refer to as a behavior. Thus, the constraints are imposed by the conditions of positivity $P(ab|xy) \geq 0 \; \forall a, b, x, y$; and the normalization $\sum_{a,b=1}^{k} P(ab|xy) = 1$ for all x and y. The existence of a given physical model behind the correlations obtained in a Bell scenario translates into additional constraints on the behaviors $P(ab|xy)$. There are three main types of correlation that can be distinguished.

3.1.1 No-signalling Correlations

A first natural limitation on behaviors **P** are the no-signalling constraints [Cir80, PR94]. They are expressed as

$$\sum_{b=1}^{k} P(ab|xy) = P(a|x) \qquad \forall a, x, y \quad \text{and}$$
$$\sum_{a=1}^{k} P(ab|xy) = P(b|y) \qquad \forall b, x, y\,. \tag{3.1}$$

These relations state that the marginal probability distribution of Alice is independent of Bob's input y and vice versa. In particular, if Alice and Bob are space-like separated, the no-signalling constraints in Eq. (3.1) guarantee that Alice and Bob cannot use their devices for instantaneous signalling, preventing a direct conflict with relativity. The set of all correlations satisfying the no-signalling constraints forms a convex polytope $\mathcal{NS}$.

3.1.2 Classical Correlation

We say that a behavior **P** (=$\{P(ab|xy)\}$) is local if one can write it in the following form:

$$P(a, b|x, y) = \int_{\Lambda} q(\lambda) P(a|x, \lambda) P(b|y, \lambda) d\lambda\,. \tag{3.2}$$

where λ are the hidden variables that completely describe the system under consideration. It takes value in a space Λ and is distributed according to the probability density $q(\lambda)$. $P(a|x, \lambda)$ ($P(b|y, \lambda)$) are local probability response functions for Alice (Bob). Operationally λ can be conceived as the shared randomness where Alice (Bob) will select an outcome a (b) based on her (his) measurement setting x (y) and λ.

The set of all probabilities with local/classical origin, i.e. the probabilities that can be reproduced within a classical or locally real theory, forms a convex polytope [Pit89, Fin82, Pit91]. We denote this polytope as $\mathcal{P}$. The polytope $\mathcal{P}$ can be characterized by its extremal points $\mathbf{v}_p \in \mathbb{R}^{m^2k^2}$, where $p = \{1, 2, \cdots, k^{2m}\}$, so-called vertices. The vertices $\mathbf{v}_p$ have entries from

the set {0,1} and correspond to the deterministic strategies[1]. Every classical correlation $\mathbf{P}_{cl} \in \mathcal{P}$ can be written as a convex combination of all the vertices of the extremal points, i.e.

$$\mathbf{P}_{cl} = \sum_{p=1}^{k^{2m}} \lambda_p \mathbf{v}_p \tag{3.3}$$

where $\lambda_x \geq 0$ and $\sum_{p=1}^{k^{2m}} \lambda_p = 1$. This subsequently implies that any correlation that cannot be written in the form of Eq. (3.3) is nonlocal.

3.1.3 Quantum Correlation

Quantum correlations are the set of correlations that is achievable in quantum mechanics. A behavior is quantum if there exist a quantum state $|\Psi_{AB}\rangle \in \mathcal{H}_A \otimes \mathcal{H}_B$ of arbitrary dimension and measurement operators (POVM elements) $\{M_{a|x}\}$ and $\{M_{b|y}\}$ which describes the performed measurement such that

$$P(ab|xy) = \langle \Psi_{AB}| M_{a|x} \otimes M_{b|y} |\Psi_{AB}\rangle \ . \tag{3.4}$$

The set of all quantum behaviors forms a convex set $\mathcal{Q}$, but it is not a polytope. The boundaries of $\mathcal{Q}$ are still unknown in spite of analytical efforts to describe it [PPK+09, FSA+13].

The sets $\mathcal{P}$, $\mathcal{Q}$ and $\mathcal{NS}$ are closed, bounded and convex, and obey the following relation:

$$\mathcal{P} \subsetneq \mathcal{Q} \subsetneq \mathcal{NS} \, . \tag{3.5}$$

The local and no-signalling sets are polytopes. Thus, these can be character-

[1] In a deterministic strategy, the local response functions $p(a|x,\lambda)$ and $p(b|y,\lambda)$ only take values from the set $\{0,1\}$. In a deterministic model, the hidden variable λ specifies the assignment of one of the potential outputs to each input. Let $\lambda = (a_1, \ldots, a_m; b_1, \ldots, b_m)$ indicates an assignment of outputs correspond to their input; a_x and b_y denote the output for each of the inputs $x = 1, \ldots, m$ and $y = 1, \ldots, m$. Then the corresponding deterministic behavior/strategy $\mathbf{P}^{\lambda}_{\text{det}} \in \mathcal{P}$ is denoted as:

$$P^{\lambda}_{\text{det}}(ab|xy) = \begin{cases} 1 & \text{if } a = a_x \text{ and } b = b_y \\ 0 & \text{otherwise.} \end{cases}$$

For the $[m, k]$ Bell scenario, there are k^{2m} possible output assignments. Therefore, k^{2m} such local deterministic strategies can exist; thus $\lambda = \{1, \cdots, k^{2m}\}$.

ized by the convex hull of a finite number of extremal points, the vertices. However, the characterization of the quantum set is not so straightforward. To characterize the quantum set, one needs to use the NPA hierarchy, which we will discuss later in this chapter.

3.2 Bell Inequalities

In 1964, John Stewart Bell published an article on the EPR paradox [Bel64] where he gave a precise mathematical characterization of *local realism*. It enabled him to obtain the following result:

Theorem 3.1. *No physical theory of local hidden variables can ever reproduce all the predictions of quantum mechanics.*

To demonstrate his conclusion, Bell derived an inequality that must be fulfilled by all local correlations (as defined in Eq. (3.2)); but some quantum correlations do not comply with this inequality. Namely, he presented a hyperplane that separates a quantum behavior $\mathbf{P}_Q \in \mathcal{Q}$ from the entire set of local behaviors $\mathcal{P}$. Since then, many such inequalities were derived and termed 'Bell inequalities'. The boundaries (or facets) of the classical polytope $\mathcal{P}$ sharply divide the classical and the quantum set and therefore represent a Bell inequality. Bell inequalities that correspond to these facets of $\mathcal{P}$ are called the facet Bell inequalities [BCP+14]. There are also other types of Bell inequalities that only act as a hyperplane, but not a facet of $\mathcal{P}$. See Fig. 3.2 for visualization. The generic form of a Bell inequality is an inequality that is linear in $\mathbf{P}$ [BCP+14]:

$$\sum_{a,b,x,y} C(ab|xy)P(ab|xy) \leq \mathcal{I}_L \,. \tag{3.6}$$

We denote a Bell inequality as B in this thesis. A Bell inequality B is specified by the coefficients $C(ab|xy) \in \mathbb{R}$ (see Eq. (3.6)). Here, $\mathcal{I}_L$ is the classical bound which is the maximal value over all local correlations. We denote $\sum_{a,b,x,y} C(ab|xy)P(ab|xy)$ as the Bell value $B[\mathbf{P}]$, i.e. $B[\mathbf{P}] = \sum_{a,b,x,y} C(ab|xy)P(ab|xy)$. A behavior with classical origin (i.e. $\{P(ab|xy)\} \in \mathcal{P}$ or cannot be decomposed as Eq. (3.3)) cannot violate any of these Bell inequalities by design. Thus the violation of a Bell inequality certifies the presence of nonlocality.

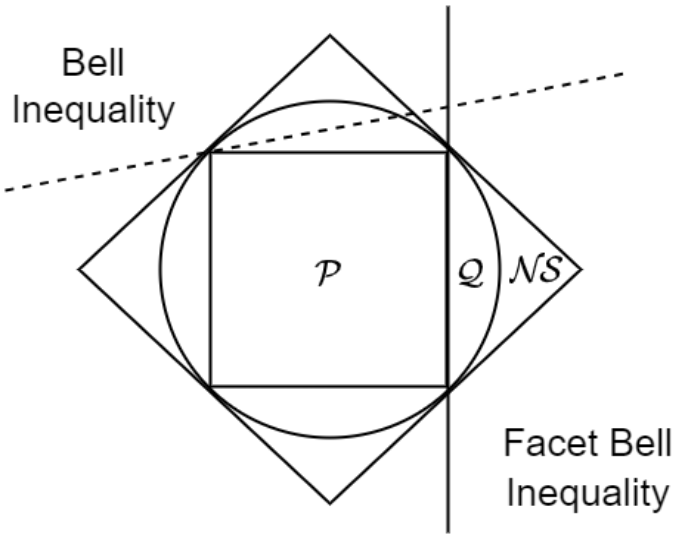

Figure 3.2: Illustration of the sets of boxes. $\mathcal{P}$, $\mathcal{Q}$, and $\mathcal{NS}$ denote the sets of classical, quantum, and no-signalling correlations, respectively. All sets are convex, $\mathcal{P}$ and $\mathcal{NS}$ are polytopes, and the relation $\mathcal{P} \subsetneq \mathcal{Q} \subsetneq \mathcal{NS}$ holds. Bell inequalities can be used to separate classical correlations from quantum ones. The solid line represents a facet Bell inequality. In contrast, the dotted line represents a Bell inequality which is only a hyperplane, not a facet of $\mathcal{P}$.

3.3 CHSH Inequality

Here, we discuss the most simple yet nontrivial Bell inequality, CHSH inequality [CHSH69]. It was introduced by *John Clauser*, *Michael Horne*, *Abner Shimony* and *Richard Holt* in 1969 and is abbreviated as *CHSH inequality*. It has significant importance in the cryptographic scenario. Most device-independent quantum cryptographic protocols depend on a Bell inequality violation. Therefore, these protocols are tailored to or directly depend on the CHSH inequality since it is the most simple bipartite Bell scenario. For this reason, we are paying close attention to this Bell setup.

In this case, the Bell scenario is defined as follows: Alice and Bob have two measurement settings, and each measurement setting has two measurement outcomes. Suppose Alice has the input settings $x \in X = \{0, 1\}$. Similarly, Bob's input settings are denoted as $y \in Y = \{0, 1\}$. The output is denoted as $a \in A = \{0, 1\}$ for Alice, and $b \in B = \{0, 1\}$ Bob. Then, the CHSH inequality reads:

$$\mathcal{S}_{\text{CHSH}} := \langle A_0 B_0 \rangle + \langle A_0 B_1 \rangle + \langle A_1 B_0 \rangle - \langle A_1 B_1 \rangle \leqslant 2\,, \tag{3.7}$$

where

$$\langle A_x B_y \rangle = p(a = b|xy) - p(a \neq b|xy) \tag{3.8}$$

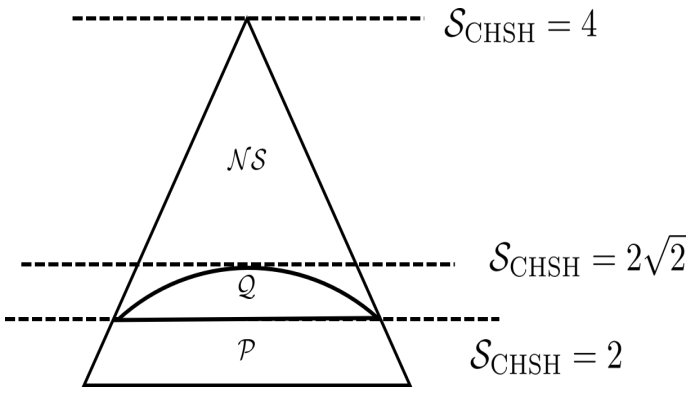

Figure 3.3: Illustration of the CHSH inequality as a hyperplane and its maximum attainable values for classical, quantum and no-signalling set.

represents the correlation of the outputs a of Alice and b of Bob when they perform the measurement labeled by x and y, respectively. In any LHV model, the CHSH value $\mathcal{S}_{\text{CHSH}}$ cannot exceed the value of 2. However, it is not the case for correlations with a quantum origin, as we show in the following. If Alice and Bob share the Bell state $|\phi^+\rangle$ (see Eq. (2.25)) on which they perform the following measurement settings:

$$\begin{aligned} A_0 &= \sigma_z\,, & A_1 &= \sigma_x\,, \\ B_0 &= \frac{\sigma_z + \sigma_x}{\sqrt{2}}\,, & B_1 &= \frac{\sigma_z - \sigma_x}{\sqrt{2}}\,, \end{aligned} \tag{3.9}$$

they obtain $\mathcal{S}_{\text{CHSH}} = 2\sqrt{2}$. It violates the classical bound and thus demonstrates the nonlocality of quantum correlations. This value is the highest achievable by a quantum behavior proven by Tsirelson [Cir80]. Here we present the following theorem:

Theorem 3.2 (Tsirelson bound [Cir80])**.** *Let $\{A_x\}$ and $\{B_y\}$ with $x \in \{0, 1, \cdots, m_a - 1\}$ and $y \in \{0, 1, \cdots, m_b - 1\}$ be two sets of observables whose eigenvalues lie in $[-1, 1]$. Then for any state $|\psi\rangle \in \mathcal{H}_A \otimes \mathcal{H}_B$, there exist real normalized vectors $\boldsymbol{v}_0, \cdots, \boldsymbol{v}_{m_a-1}, \boldsymbol{w}_0, \cdots, \boldsymbol{w}_{m_b-1} \in \mathbb{R}^{m_a+m_b}$, such that for all $x \in \{0, \ldots, m_a - 1\}$ and $y \in \{0, \ldots, m_b - 1\}$[2] the expectation value can be written as:*

$$\left\langle A_x B_y \right\rangle_{|\psi\rangle} = \langle\psi| A_x \otimes B_y |\psi\rangle = \boldsymbol{v}_x^T \boldsymbol{w}_y\,. \tag{3.10}$$

Tsirelson proved the maximum quantum value for the CHSH inequality using this theorem. We validate this using the method described in [EKB13].

[2]Note that, in our explanation of Bell scenarios in Sec. 3.1, we have used $m_a = m_b = m$.

For CHSH scenario, $m_a = m_b = 2$. For this scenario, there exist unit vectors $\boldsymbol{v}_0, \boldsymbol{v}_1, \boldsymbol{w}_0, \boldsymbol{w}_1 \in \mathbb{R}^4$, such that

$$\begin{aligned} \mathcal{S}_{\text{CHSH}} &= \sum_{x,y=0}^{1} (-1)^{x\cdot y} \langle A_x B_y \rangle_{|\psi\rangle} \\ &= \sum_{x,y=0}^{1} (-1)^{x\cdot y} \boldsymbol{v}_x^T \boldsymbol{w}_y \\ &= \boldsymbol{V}^T G \boldsymbol{W} \,, \end{aligned} \tag{3.11}$$

where $\boldsymbol{V} = (\boldsymbol{v}_0, \boldsymbol{v}_1)^T$, $\boldsymbol{W} = (\boldsymbol{w}_0, \boldsymbol{w}_1)^T$, and $G = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \otimes \mathbb{1}_4$. $\mathbb{1}_4$ represents an identity matrix of dimension 4. An upper bound on Eq. (3.11) is established by

$$\begin{aligned} \boldsymbol{V}^T G \boldsymbol{W} &\leq |\boldsymbol{V}^T G \boldsymbol{W}| \\ &\leq |\boldsymbol{V}| \|G\|_2 |\boldsymbol{W}| \\ &= \sqrt{2}\sqrt{2}\sqrt{2} = 2\sqrt{2} \,. \end{aligned} \tag{3.12}$$

In Eq. (3.12), we introduce the Spectral norm $\|\cdot\|_2$ which is a matrix norm induced by the Euclidean vector norm. Since G is symmetric, $\|G\|_2$ is given by its largest absolute eigenvalue, which is $\sqrt{2}$. We have used the following relations: $\boldsymbol{V} \cdot (G\boldsymbol{W}) \leq |\boldsymbol{V}||G\boldsymbol{W}|$ and $G\boldsymbol{W} \leq \|G\|_2 \boldsymbol{W}$ to achieve the second inequality of Eq. (3.12). The Euclidean norm of $\boldsymbol{V}$, $|\boldsymbol{V}| = \sqrt{|\boldsymbol{v}_1|^2 + |\boldsymbol{v}_2|^2} = \sqrt{2}$, and $\boldsymbol{W}$, $|\boldsymbol{W}| = \sqrt{|\boldsymbol{w}_1|^2 + |\boldsymbol{w}_2|^2} = \sqrt{2}$ (because $\boldsymbol{v}_1$, $\boldsymbol{v}_2$, $\boldsymbol{w}_1$, $\boldsymbol{w}_2$ are unit vectors) are used to achieve the equality of the second last step. The bound in Eq. (3.12) is tight, as we already discussed the measurement settings and the quantum state (see Eq. (3.9)) that can achieve the Bell value $2\sqrt{2}$. For no-signalling behavior, the Popescu-Rohrlich (PR) box [PR94] is given by:

$$P^{\text{PR}}(ab|xy) = \frac{1}{2} \delta_{a\oplus b, x\cdot y} \,. \tag{3.13}$$

$P^{\text{PR}}(ab|xy)$ achieves the maximal value $\mathcal{S}_{\text{CHSH}} = 4$ for no-signalling behavior. See Fig. 3.3 for a pictorial representation.

3.4 Designing Bell Inequalities from Probability Distribution

Bell inequalities are a versatile tool used in many quantum information-theoretic tasks, such as nonlocality and entanglement detection, quantum cryptography, etc. However, one cannot use one Bell inequality for every task. Moreover, Bell inequality suited for one may not be optimal for another. Thus, designing Bell inequalities appropriate for that specific task is crucial. In this section, we discuss how we construct a Bell inequality from a given probability distribution **P** that leads to the maximal violation for that particular **P**.

Consider the $[m, k]$ Bell scenario (see Sec. 3.1 for details) where the parties receive the measurement data that obeys the probability distribution **P** which reads:

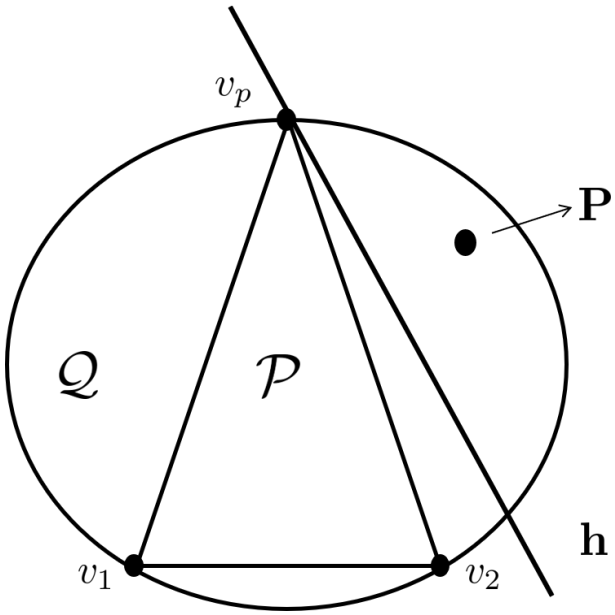

Figure 3.4: Schematic representation of a Bell inequality, specified by the vector **h** defining a hyperplane, separates all vertices v_p from the observed probability distribution **P** (showed by the black point situated outside the classical polytope $\mathcal{P}$).

$$\mathbf{P} := \{P(ab|xy)\}\,, \tag{3.14}$$

where $x \in X$, $y \in Y$, $a \in A$ and $b \in B$. Recall that the Bell inequalities correspond to hyperplanes in the probability space that separate the classical correlation polytope $\mathcal{P}$ from the set of all genuine quantum correlations $\mathcal{Q}\backslash\mathcal{P}$.

Such hyperplanes are specified by a normal vector $\mathbf{h} \in \mathbb{R}^{m^2k^2}$. If $\mathbf{P} \in \mathcal{Q} \setminus \mathcal{P}$, there exists at least one hyperplane $\mathbf{h}$ that separates all the vertices $\mathbf{v}_p$ of $\mathcal{P}$ from the probability distribution $\mathbf{P}$. If there exists no such hyperplane, then the observed measurement statistics $\mathbf{P} \notin \mathcal{P}$.

To obtain the hyperplane vector $\mathbf{h}$ (corresponding to the Bell inequality) that leads to maximal Bell violation for the measurement data $\mathbf{P}$, we need to formulate the following optimization problem [DKB22a]:

$$\begin{aligned} \max_{\mathbf{h},c} \quad & \mathbf{h}^T\mathbf{P} - c \\ \text{subject to} \quad & \mathbf{h}^T\mathbf{v}_p \leq c \quad \forall \quad p \in \{1, \cdots, k^{2m}\} \\ & -1 \leq h_i \leq 1 \quad \forall \quad i \in \{1, \cdots, m^2k^2\} \end{aligned} \tag{3.15}$$

with the classical bound c. The additional constraint imposed on the elements of h_i of the hyperplane vector keeps the maximization bounded. The hyperplane found in this manner has the form

$$\mathbf{h} = \{h(ab|xy)\}\,, \tag{3.16}$$

where $x \in X$, $y \in Y$, $a \in A$ and $b \in B$. Thus, the Bell inequality found by the optimization and specified by the hyperplane vector $\mathbf{h}$ is given as

$$\sum_{a,b,x,y} h(ab|xy)P(ab|xy) \leq c\,. \tag{3.17}$$

The Bell inequality, represented by Eq. (3.17), leads to the maximal Bell violation for $\mathbf{P}$; see Fig. 3.4 for visualization. Note that, if $\mathbf{P} \in \mathcal{P}$, the optimization problem of Eq. (3.15) is infeasible and no Bell inequality can be found.

This kind of construction of Bell inequality from the measurement data, designed to attain the maximal Bell violation for that specific measurement statistics, has significant importance. The Bell inequality derived in this manner can be used to bound the device-independent detection efficiency [SKB17]. This method can also be used to obtain the bounds on secret key rates in device-independent QKD scenarios [DKB22a], which we will discuss in detail in the latter part of this thesis.

3.5 NPA Hierarchy

Recall that we indicated in Sec. 3.1.3 that it is challenging to fully describe the set of quantum correlations. Here, we illustrate a powerful and incredibly

flexible versatile technique named the Navascués-Pironio-Acín (NPA) hierarchy [NPA07, NPA08] that can be used to characterize the set of quantum correlations Q.

3.5.1 Introduction to Semidefinite Programming

First, we will quickly recap the basics of semidefinite programming (SDP) [VB96]. We will follow the formulation of [NPA08, BBV04]. SDP is a subset of convex optimization [BBV04], i.e. the task of optimizing convex functions over convex sets. Many of the problems arising in quantum information theory are of this kind: entanglement distillation [Rai01], distinguishing separable and entangled states [DPS02], and the unambiguous discrimination of non-orthogonal quantum states [Eld03], to name just a few examples, can be aided by SDP.
In SDPs, a linear objective function is optimized over convex constraint functions. It can be formulated as

$$\begin{aligned} &\text{maximize:} \quad \mathrm{Tr}(F_0 Z)\,, \\ &\text{subject to:} \quad \mathrm{Tr}(F_i Z) = c_i\,, \quad \forall i \in \{1, \cdots, s\} \\ &\qquad\qquad\quad Z \geq 0\,. \end{aligned} \tag{3.18}$$

This is known as *primal problem*. The problem variable is the Hermitian matrix $Z \in C^{r\times r}$ and the problem parameters or problem data are the Hermitian matrices $F_0, F_i \in C^{r\times r}$ and scalars c_i. The variable Z is primal feasible if $\mathrm{Tr}(F_i Z) = c_i \; \forall \; i \in \{1, \cdots, s\}$ and $Z \geq 0$. It is strictly primal feasible if $Z > 0$ instead of $Z \geq 0$.
Every primal problem has it's *dual*. The *dual* program is a minimization of a linear function of $\mathbf{x} = (x_1, \cdots, x_s)^T$ subject to the restrictions imposed by an affine combination of F_i,

$$\begin{aligned} &\text{minimize:} \quad \mathbf{c}^T \mathbf{x}\,, \\ &\text{subject to:} \quad F(\mathbf{x}) = \sum_{i=1}^{s} x_i F_i - F_0 \geq 0\,. \end{aligned} \tag{3.19}$$

The variable $\mathbf{x}$ is *dual feasible* if $F(\mathbf{x}) \geq 0$ and *strictly dual feasible* if $F(\mathbf{x}) > 0$. The key property of the dual program is that it yields useful bounds on the optimal value for the primal solution and vice versa. To see this, define the

optimal primal solution

$$p^* := \sup\{\mathrm{Tr}(F_0 Z) \,|Z \geq 0,\ \mathrm{Tr}(F_i Z) = c_i \ \forall\ i \in \{1, \cdots, s\}\}\,, \tag{3.20}$$

and the optimal dual solution

$$d^* := \inf\{\mathbf{c}^T \mathbf{x} \mid F(\mathbf{x}) \geq 0\}\,. \tag{3.21}$$

Let us take a dual feasible point x and a primal feasible point Z. Then

$$\begin{aligned} \mathbf{c}^T \mathbf{x} - \mathrm{Tr}(F_0 Z) &= \sum_{i=1}^{s} \mathrm{Tr}(F_i Z) x_i - \mathrm{Tr}(F_0 Z) \\ &= \mathrm{Tr}(F(x) Z) \geq 0\,. \end{aligned} \tag{3.22}$$

This proves that the optimal primal value p^* and the optimal dual value d^* satisfy $p^* \leq d^*$. This is called weak duality. For strong duality, it holds $p^* = d^*$. The existence of a strictly feasible primal point Z or dual point $\mathbf{x}$ is a sufficient condition for this strong duality [VB96], which is also known as Slater's theorem.
SDPs can be efficiently solved using with standard ready-to-use tools like CVX [GB14], Mosek [ApS19], Sedumi [Stu99], SDPT3 [TTT99], Yalmip [Löf04] etc. Now we will discuss the Navascués-Pironio-Acín Hierarchy, also known as NPA hierarchy.

3.5.2 Navascués-Pironio-Acín Hierarchy

Ref. [NPA07, NPA08] address the following problem: Is a behavior $\mathbf{P}$ of quantum origin? In other words, do there exist local measurement operators and a quantum state that can reproduce the behavior $\mathbf{P}$? Since the dimensions of the quantum systems are unbounded, full characterization is difficult. Therefore, a series of weaker conditions are considered rather than just searching for a generic quantum realization for a given behavior. A behavior is considered to be of quantum origin if it meets the requirements at all levels. On the other hand, if a prerequisite is not met, we can discount a quantum origin for the behavior.

Let $|\psi\rangle \in \mathcal{H}$ be a pure state and $\{E_{a|x}\}, \{E_{b|y}\}$ be sets of projective measurement operators belongs to Alice and Bob, respectively. Let us recall that a quantum behavior is a set of conditional probabilities $\mathbf{P} := \{P(ab|xy)\}$

such that

$$P(ab|xy) = \left\langle \psi \left| E_{a|x} \otimes E_{b|y} \right| \psi \right\rangle ,$$

for all a, b, x, y.

Let's define a set of observables plus the identity

$$\mathcal{E} = \mathbb{1} \cup \{E_{a|x}\} \cup \{E_{b|y}\} . \tag{3.23}$$

Let $\mathcal{O} := \{O_1, \cdots, O_n\}$ be a set of n operators O_i. Each O_i is a linear combination of products of the projectors in the set $\mathcal{E}$. Consider linear equations of the form

$$\sum_{ij} (F_k)_{ij} \langle \psi | O_i^\dagger O_j | \psi \rangle = g_k(\mathbf{P}) , \tag{3.24}$$

where $k = \{1, \cdots, s\}$, and $g_k(\mathbf{P})$ are linear functions of the probabilities,

$$g_k(\mathbf{P}) = (g_k)_0 + \sum_{a,b,x,y} (g_k)_{abxy} P(ab|xy) . \tag{3.25}$$

Let $\mathcal{S}_q$ be the set that contains operators which are nontrivial products of the projectors $E_{a|x}$ and $E_{b|y}$. The sets of $\mathcal{S}_q$ are defined in the following way:

$$\begin{aligned} \mathcal{S}_0 &= \{\mathbb{1}\} \\ \mathcal{S}_1 &= \mathcal{S}_0 \cup \{E_{a|x}\} \cup \{E_{b|y}\} \\ \mathcal{S}_2 &= \mathcal{S}_1 \cup \{E_{a|x} E_{a'|x}\} \cup \{E_{a|x} E_{b|y}\} \cup \{E_{b|y} E_{b'|y}\} \\ \mathcal{S}_3 &= \mathcal{S}_2 \cup \cdots \end{aligned} \tag{3.26}$$

Thus, the set $\mathcal{S}_q$ is the set of all products of observables up to length q. From the construction itself, it is evident that

$$\mathcal{S}_0 \subseteq \mathcal{S}_1 \subseteq \mathcal{S}_2 \subseteq \mathcal{S}_3 \subseteq \cdots \tag{3.27}$$

and every operator $O_i \in \mathcal{O}$ can be written as a linear combination of operators from $\mathcal{S}_q$ for sufficiently large q.

Navascués, Pironio and Acín [NPA07, NPA08] proved that it is a necessary and sufficient condition for an unspecified behavior $\mathbf{P}$ to have a quantum realization if there exists a certificate Γ such that

$$\sum_{ij} (F_k)_{ij} \Gamma_{ij} = g_k(\mathbf{P}) \tag{3.28}$$

where $k = \{1, \cdots, s\}$ and Γ is a complex, hermitian positive semidefinite matrix. The coefficients Γ_{ij} is defined as:

$$\Gamma_{ij} := \left\langle \psi \left| O_i^\dagger O_j \right| \psi \right\rangle , \tag{3.29}$$

which form the moment matrix $\Gamma \in \mathbb{C}^{n \times n}$ associated to the set $\boldsymbol{O}$.

The existence of such a certificate can be verified by the solution of the following SDP [NPA08]:

$$\begin{aligned} &\text{maximize} \quad \lambda \\ &\text{subject to} \quad \mathrm{Tr}\,(F_k \Gamma) = g_k(\mathbf{P}), \quad \forall k \in \{1, \ldots, s\}, \\ &\qquad\qquad \Gamma - \lambda 1 \geqslant 0\,. \end{aligned} \tag{3.30}$$

A positive solution $\lambda \geqslant 0$ of this SDP implies the existence of positive semidefinite matrix $\Gamma \geq \lambda 1$. On the other hand, a negative solution $\lambda < 0$ implies that the given behavior $\mathbf{P}$ has a non-quantum origin.

The Eq. (3.30) represents the *primal* problem of the SDP. As mentioned in Sec. 3.5.1, one can solve Eq. (3.30) both in their *primal* and *dual* forms. The *dual* of Eq. (3.30) is:

$$\begin{aligned} &\text{minimize} \quad \sum_k y_k g_k(P) \\ &\text{subject to} \quad F(y) = \sum_k y_k F_k \geq 0, \\ &\qquad\qquad \sum_k y_k \,\mathrm{tr}(F_k) = 1\,. \end{aligned} \tag{3.31}$$

A certificate of order q, denoted as Γ^q, is associated to the set $\mathcal{S}_q$. Recall the set of operators $\mathcal{S}_q$ follow Eq. (3.27). Thus, the family of certificate $\Gamma^1, \Gamma^2, \cdots, \Gamma^q, \cdots$ represents a hierarchy of conditions a quantum probability must follow. Each condition in the hierarchy is stricter than the previous one. Let $\{\mathbf{P}_q\}$ define a set of behaviors for which a certificate of order q exists. $\{\mathbf{P}_q\}$ defines an associated subspace $\boldsymbol{Q}_q$ of the probability space that contains $\boldsymbol{Q}$. The existence of family of certificates $\Gamma^1, \Gamma^2, \cdots, \Gamma^q, \cdots$ will give rise to the sequence

$$\boldsymbol{Q}_1 \supseteq \boldsymbol{Q}_2 \supseteq \cdots \supseteq \boldsymbol{Q}_q \supseteq \cdots \tag{3.32}$$

of outer approximations of the quantum set $\boldsymbol{Q}$. Certificates of higher order provide a more accurate approximation of the quantum set $\boldsymbol{Q}$ and $\lim_{q \to \infty} \boldsymbol{Q}_q = \boldsymbol{Q}$. Thus any non-quantum behavior $\mathbf{P}$ will fail at some level

$q < \infty$.
Now, we can state the central result of [NPA08].

Theorem 3.3. *Let* $\boldsymbol{P}$ *be a behavior such that there exists a certificate* Γ^q *of order* q *for all* $q \geqslant 1$. *Then* $\boldsymbol{P}$ *belongs to* $\boldsymbol{Q}$.

Proof. See Ref. [NPA08] for the proof. Also, see Fig. 3.5 for visualization. □

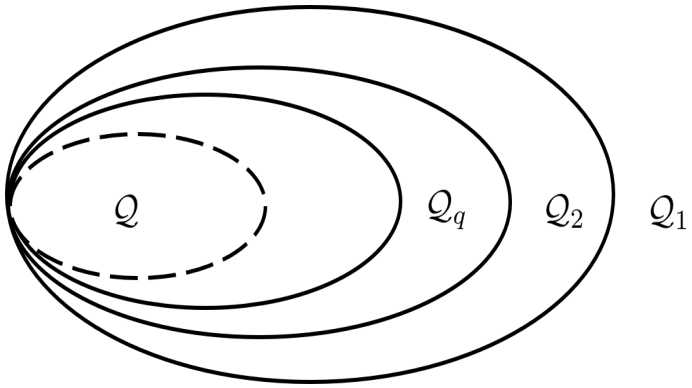

Figure 3.5: Geometrical interpretation of a quantum set. Q is the set of all quantum behaviors. Q_n denotes the set of all behaviors for which a certificate of order n exists. Certificates of higher order provide a more accurate approximation of the quantum set.

Note that the numerical resources to check whether the probability distribution $\mathbf{P}$ satisfies the criteria increase with the hierarchy because the size of Γ^q increases with the hierarchy (due to the addition of constraints). Fortunately, the objective function of Eq. (3.30) (or Eq. (3.31)) often converges (attains a high level of numerical precision) at a low hierarchy level (typically 2). Usually, the computation can be terminated once the desired level of numerical precision is reached. The quantumness of a certain behavior is then validated within this accuracy. As a constraint for an optimization task, one can use the publicly downloadable MATLAB-toolbox QETLAB [Joh16] and Python module ncpol2sdpa [Wit15] to test the fulfillment of the hierarchy to some given level.

The NPA hierarchy has a wide range of applications. For example, it enables us to establish an upper bound on the Tsirelon bound [Cir80] of any given Bell inequality. It also helps us to bound the guessing probability of a device-independent quantum key distribution protocol [MPA11], which we will discuss in Chap. 5.

4

General Concepts in Quantum Key Distribution

The security of classical cryptographic protocols relies on the assumption that the adversary has limited computational resources. Thus the security is based on certain mathematical problems which are 'hard' to solve. One such example is the RSA encryption [RSA78], which is based on the fact that prime factorization of large numbers cannot be done in polynomial time. In contrast, the security of quantum cryptography relies on intrinsic principles of nature, as described by quantum mechanics. Therefore, assuming that quantum mechanics is correct, the security offered by quantum cryptography is everlasting and secure against retroactive attacks.

Quantum key distribution (QKD) is a cryptographic task in which two (or more) honest parties, Alice and Bob, wish to establish a shared secret string of bits unknown to any third party, including a potential eavesdropper, Eve. The security of a QKD protocol uses several ideas from quantum physics, information theory, and computer science. Here we deconstruct the notion of security for QKD into its parts and elaborate on the steps of the protocol. We adopted the concepts of this chapter from [SBPC+09, Ren08, SR08a, SR08b, Gra21]. First, we will discuss the types of attack an eavesdropper can perform on QKD protocols in Sec. 4.1. Then, we precisely define security in Sec. 4.2. In Sec. 4.3, we discuss the quantum and classical post-processing steps used in QKD protocols. We will show how these results can be used to help reduce the security definition to a different kind of problem. Afterwards, we discuss the BB84 protocol (in Sec. 4.4) and the entanglement-based BB84 protocol (in Sec. 4.5) in detail, as it represents the origin of (bipartite) QKD.

Then we conclude the chapter by calculating the asymptotic secret key rate for the BB84 protocol in Sec. 4.6.

4.1 Eve's Attack

In a bipartite cryptographic scenario, the involved parties, Alice and Bob, aim to establish a secret key for secure communication. An adversary, Eve, is a third party who can eavesdrop on the public communication between the parties. Additionally, Eve may interfere with the quantum channels and explore correlations with the established key between Alice and Bob. Here we discuss what Eve can or cannot do since the security of a QKD protocol can be proven under these circumstances. Three different classes of attacks for Eve are considered in the literature [SBPC+09]: individual attacks, collective attacks, and coherent attacks. The first two attack strategies pose some restrictions to Eve, while the third one is the most general attack allowed by quantum mechanics. When facing the task of proving security for a QKD protocol, a first attempt may be made to prove security against individual and collective attacks before proving the security under coherent attacks.

- **Individual attacks**: These attacks are the least powerful attacks for Eve. The eavesdropper can only attack each round of the protocol individually. In this case, she is assumed to have no quantum memory, so her best strategy is to measure her quantum side information at each round.

- **Collective attacks**: It assumes that Eve performs the same attack in each round of the protocol; i.e. for different rounds, quantum side information of Eve is identically and independently distributed (IID) for different rounds. Different from individual attacks, Eve has a quantum memory. Therefore, she can store her quantum side information at each round and use it to carry out a global operation at the end of the protocol.

- **Coherent attacks**: It is the most general type of attack. There are no assumptions on the capabilities of the eavesdropper, except that the laws of quantum mechanics bound her. Eve is allowed to perform global operations on her quantum side information. In this case, the states distributed to the parties in every round may have arbitrary correlations with previous and future rounds.

4.2 Security Definitions

Before we describe how to prove security, it is essential to define what we mean by security. QKD's goal is to ensure that Alice and Bob share a key that no eavesdropper has any information about. This definition is too strong as we can only achieve approximate security, which is adequate for practical purposes. For this purpose, we need to define some security parameters [Ren08, AFRV19, MvDR+19]. The security of quantum key distribution can be split into two conditions: correctness and secrecy. The definition of correctness is straightforwardly motivated since we want to ensure that Alice's and Bob's keys are almost always the same. We just require that the probability of their keys being different is low. To construct this definition, we assume Alice's and Bob's keys at the end of the protocol as K_A and K_B, respectively. Then we define the correctness of the QKD protocol as the following:

Definition 4.1 (Correctness)**.** *A QKD protocol is ϵ_{corr}-correct if the final key K_A of Alice differs from the final key K_B of Bob with probability at most ϵ_{corr}, i.e.*

$$\Pr(K_A \neq K_B) \leq \epsilon_{corr} \,. \tag{4.1}$$

The other aspect of the security of a QKD protocol is secrecy. Secrecy for QKD is the notion that the eavesdropper, Eve, does not have any information about Alice's key. Secrecy is defined as the distance between the shared state of Alice and Eve in the real protocol and the ideal protocol.

Definition 4.2 (Secrecy)**.** *For any $\epsilon_{sec} \geq 0$, a QKD protocol is ϵ_{sec}-secret w.r.t the adversary E if the joint state of Alice's key and Eve's side information satisfies:*

$$\Pr(\Omega) \cdot \frac{1}{2} \left\| \rho_{K_A E_T|\Omega} - \tau_{K_A} \otimes \rho_{E_T|\Omega} \right\|_1 \leq \epsilon_{sec} \,, \tag{4.2}$$

where $\rho_{K_A E_T|\Omega}$ is the state that describes the correlation between Alice's final secret key K_A and the total information available to Eve E_T given that the protocol did not abort, while τ_{K_A} is the maximally mixed state on K_A. Here $\Pr(\Omega)$ is the probability of not aborting the protocol.

The definition of secrecy can be operationally interpreted as follows. Let $\rho_{K_A E_T|\Omega}$ be the final state of the real QKD protocol and $\tau_{K_A} \otimes \rho_{E_T|\Omega}$ be the final state of the ideal protocol, which acts exactly like the real protocol except that it outputs a perfectly secret key for Alice, i.e. uniformly distributed and independent of Eve's system. Then, the real protocol is ϵ_{sec}-secure if

a distinguisher, who has access to all the inputs and outputs of the real and ideal protocols (including Eve's system), cannot distinguish the two protocols except for a probability at most ϵ_{sec} [PR22].

The secrecy of Alice's key K_A alone does not guarantee that even Bob's key K_B is secret unless we combine it with a statement on the correctness of the protocol. Therefore we define the security of a QKD protocol as follows.

Definition 4.3 (Security). *If a protocol is ϵ_{corr}-correct and ϵ_{sec}-secret, then it is ϵ^s_{QKD}-correct and secret for any $\epsilon^s_{QKD} \geq \epsilon_{corr} + \epsilon_{sec}$.*

Lastly, another necessary condition of QKD is *completeness* or *robustness*. Completeness states that there should exist an honest implementation for which the probability of aborting the protocol is minuscule (ϵ^c_{QKD}). Mathematically, it can be expressed as $\Pr(\Omega) \geq 1 - \epsilon^c_{QKD}$, where $\Pr(\Omega)$ is the probability of not aborting the protocol.

Now that we state all the security definitions, we will define the secret key rate of the QKD protocol.

Definition 4.4 (Secret key rate). *Secret key rate r is the ratio*

$$r = \frac{l}{n}, \tag{4.3}$$

where l is the length of the secret key produced by a ϵ-secure QKD protocol and n is the total number of uses of the quantum channel.

The above rate is evaluated in *bits/round*. Additionally, the time required to perform one round (or use of the quantum channel) can also be taken into account to present the secret key rate in *bits/s*.
The goal of the security analysis of a QKD protocol is to derive the secret key rate as a function of these security parameters (Def. 4.1, Def. 4.2, Def. 4.3, Def. 4.4) that Alice and Bob can estimate during the execution of the protocol. Now we will discuss the structure of a general QKD protocol.

4.3 General QKD Protocol

Almost all QKD protocols follow the same general structure. We will focus on bipartite QKD, where two parties, Alice and Bob, aim to share a secret random string.
First, the protocol has a quantum stage where quantum states are prepared and distributed. Then the parties perform local measurements. After that,

the classical stages follow. The classical stage, usually called classical post-processing, is generally divided into three parts, i.e. parameter estimation, information reconciliation (also called error correction), and privacy amplification. These steps are performed on measurement outcomes of the parties. The parameter estimation step is used to estimate some global properties of the shared string, which is crucial for calculating the secret key rate. Alice and Bob performed the information reconciliation step to correct the errors between their strings, which may have been caused by an eavesdropper or noise in the channel and devices they used. In privacy amplification, Alice and Bob ensure that any residual knowledge an eavesdropper has gained is removed. Alice and Bob need to communicate classically for the classical post-processing, and they need to know that an eavesdropper does not interfere with this communication. Hence, they use an authenticated classical channel.

The communication between Alice and Bob can happen in two ways. The first way, called direct reconciliation, is if Alice only sends classical information about her string to Bob, and Bob does not tell Alice anything about his string. If the roles of Alice and Bob are reversed so that Bob only sends classical information about his string to Alice, then this is called reverse reconciliation. Direct and reverse reconciliation are one-way classical post-processing. Even though the communication is one-way, the other party, such as Bob, in direct reconciliation, may need to communicate some auxiliary information to Alice, such as whether they should abort or continue the protocol (see below for more information on aborting). They can also implement post-processing by using two-way communication, where Alice and Bob send information to each other about their strings. Typically one-way communication is considered since it is usually easier to analyze and sufficient to perform the post-processing. Throughout this thesis, we will assume that direct reconciliation is being performed. Now we will discuss all the general steps in detail.

4.3.1 Preparation, Distribution and Measurement

It is the quantum stage of the QKD protocol. In this step, a source distributes the quantum state to Alice and Bob. Alternatively, Alice could prepare the quantum state and send it to Bob. This step is repeated n times.

Alice and Bob perform local measurements once they receive the quantum state in every round of the protocol. Alice and Bob generally use one of the two different sets of measurement settings. Alice and Bob choose one

set of measurement settings that are more frequently employed. These measurement rounds are called the *key generation (KG)* rounds. The outcomes of KG rounds will constitute the final secret key between the parties. The other set of measurements are used for *test* rounds, also referred to as *parameter estimation (PE)* rounds. The outcomes of PE rounds will be used to estimate some relevant parameters of the protocol. The parties use a pre-shared key to determine whether a round is a KG round or a PE round. As an alternative, the parties can employ a sifting phase to select rounds that include the same measurement setting [TL17].

The local measurements are specified by the inputs x_i and y_i, respectively, for the i^{th} round of the protocol once they receive the quantum state. The local measurements produce classical outcomes. Alice and Bob record the outcomes a_i and b_i, respectively. After n rounds of measurement, Alice and Bob hold the input bit strings X^n and Y^n and the output bit strings A^n and B^n, respectively, which are the collections of the input and output of their measurement.

Now we will discuss the classical steps of a QKD protocol in reverse chronological order starting with privacy amplification and then information reconciliation and parameter estimation.

4.3.2 Privacy Amplification

The last step of classical post-processing is privacy amplification. In this step, Alice and Bob want to turn their equal string of bits, which may be partially known to an eavesdropper, into a shorter, completely secure string of bits. In order to do that, they are going to make use of a 2-universal family of hash functions.

A hash function $f : \{0,1\}^n \to \{0,1\}^\ell$ is a function that maps a longer string of bits into a shorter string, i.e. $\ell \leq n$. We will be interested in particular families of hash functions that satisfy a property called 2-universality.

Definition 4.5 (2-universal hash function)**.** *Consider a family hash function* $\mathcal{F} = \{f : \{0,1\}^n \to \{0,1\}^\ell\}$. *$\mathcal{F}$ will be called 2-universal if*

$$\Pr_{f\in\mathcal{F}}\left[f(x) = f\left(x'\right)\right] = \frac{1}{2^\ell}, \tag{4.4}$$

for every two strings $x, x' \in \{0,1\}^n$ with $x \neq x'$ and f is randomly chosen from $\mathcal{F}$. There always exist a 2-universal family of hash functions [CW79] for $\ell \leq n$.

Now, we will state the *Leftover Hashing Lemma* [RW05, TSSR11, TLGR12,

MvDR+19]. The Leftover Hashing Lemma guarantees that the resulting key of Alice after the application of the hash function, i.e. after privacy amplification, is almost uncorrelated from Eve's side information. In particular, the Leftover Hashing Lemma provides an upper bound on the distance between the real state $\rho_{K_A FE}$ and the ideal state $\tau_{K_A} \otimes \rho_{FE}$, which depends on the length of the final key and Eve's uncertainty about Alice's bit-string before privacy amplification.

Theorem 4.1 (Leftover Hashing Lemma)**.** *Let $\rho_{A^n E}$ be a cq-state, where the classical register A^n stores an n-bit string, and let $\mathcal{F}$ be a 2-universal family of hash functions, from $\{0,1\}^n$ to $\{0,1\}^\ell$, that maps A^n into K_A, then*

$$\left\| \rho_{K_A FE - \tau_{K_A} \otimes \rho_{FE}} \right\|_{\text{tr}} \leq \frac{1}{2} 2^{-\frac{1}{2}\left(H_{\min}(A^n|E)_\rho - \ell\right)}, \tag{4.5}$$

where F is a classical register that stores the hash function f (see Def. 4.5) .

Proof. See [Ren08, TSSR11, Tom15] for proof and more details. □

In general, the smooth min-entropy can be much higher than the min-entropy. Thus, instead of min-entropy, we reframe the leftover hashing lemma in terms of smooth min-entropy. However, we have to pay a linear term in the security parameter for this relaxation.

Theorem 4.2 (Leftover Hashing Lemma with smooth min-entropy)**.** *Let $\rho_{A^n E}$ be a classical quantum state. Let $\mathcal{H}$ be a 2-universal family of hash functions, from $\{0,1\}^n$ to $\{0,1\}^l$, that maps the classical n-bit string A^n into K_A. Then*

$$\|\rho_{K_A FE} - \tau_{K_A} \otimes \rho_{FE}\|_{\text{tr}} \leq 2^{-\frac{1}{2}\left(H^\epsilon_{min}(A^n|E)_\rho - l\right)} + 2\epsilon\,. \tag{4.6}$$

Proof. See [TSSR11, Tom15, TL17] for the proof and more detail. □

The Leftover hashing lemma provides a tool to bound the distance of the state of the protocol after privacy amplification to the ideal state. Thus we can write

$$\begin{aligned} &\Pr(\Omega) \cdot \left\| \rho_{K_A E_T|\Omega} - \tau_{K_A} \otimes \rho_{E_T|\Omega} \right\|_{\text{tr}} \\ &= \left\| \rho_{K_A E_T \wedge \Omega} - \tau_{K_A} \otimes \rho_{F E_T \wedge \Omega} \right\|_{\text{tr}} \\ &\leq \frac{1}{2} 2^{-\frac{1}{2}\left(H^\epsilon_{\min}(A^n|E_T)_{\rho\wedge\Omega} - \ell\right)} + 2\epsilon \end{aligned} \tag{4.7}$$

where $\rho_{K_A E_T \wedge \Omega} = \Pr(\Omega) \cdot \rho_{K_A E_T|\Omega}$ is a subnormalized state. Note that, by choosing

$$\ell \leq H^\epsilon_{\min}\left(A^n \mid E_T\right)_\rho - 2\log\left(\frac{1}{2\epsilon_{PA}}\right) \tag{4.8}$$

we achieve a ϵ_{sec}-secret key with $\epsilon_{sec} = \epsilon_{PA} + 2\epsilon$ (see Def. 4.2). Here we have used, $H^{\epsilon}_{\min}(A^n \mid E_T)_{\rho_{\wedge\Omega}} \geq H^{\epsilon}_{\min}(A^n \mid E_T)_{\rho}$ [TL17]. This is to deal with the fact that usually, one can estimate the smooth min-entropy of the normalized state rather than that of the subnormalized state conditioned on not aborting.

4.3.3 Information Reconciliation

In the previous section, we have seen that the key length is determined by the smooth entropy of Alice's string of raw bits conditioned on the information available to the eavesdropper using the leftover hashing Lemma (see Theorem 4.2). However, we have to make sure that the protocol is correct. Alice and Bob try to correct the errors between their strings, which may have been caused by an eavesdropper or noise in the channel and devices they used. They want to communicate a minimal amount of relevant information to each other over the classically authenticated channel so that they can correct any errors. This step is often called the Error Correction (EC) step. We will use the names Information Reconciliation and Error Correction interchangeably.

Consider the following scenario at this point in the protocol. Alice has a bitstring A^n and Bob has a bitstring B^n that may be different from A^n, while Eve has a quantum state ρ_E that may have correlations with Alice and Bob's information. Alice wants to send some function of her key to Bob so that Bob can use this information and B^n to reconstruct A^n. To do so, the parties perform an EC procedure so that Bob can compute a guess $\hat{A}^n$ of Alice's bitstring A^n. This error correction process reveals some information over the public channel.

Note that Bob does not have access to Alice's system, so neither Alice nor Bob know if the error correction succeeded or not. We can use two-universal hash functions for this checking procedure. The checking procedure will be done in several steps such as

- Alice (uniformly at random) chooses a two-universal hash function from a family of such functions and computes a hash $f_{\text{EC}}(A^n)$ of length $\lceil \log\left(\frac{1}{\epsilon_{EC}}\right) \rceil$. Alice then sends the hash function f_{EC} and the hash values (evaluation of the function) $f_{\text{EC}}(A^n)$ to Bob over the public channel.
- Bob uses the function f_{EC} and apply in his key to compute $f_{\text{EC}}(\hat{A}^n)$.
- If the hash values are equal, i.e. $f_{\text{EC}}(A^n) = f_{\text{EC}}(\hat{A}^n)$, then with high probability, Alice's and Bob's keys are the same. Otherwise, they will

abort the protocol.

Due to the definition of the families of hash functions (Def. 4.5), it is clear that the QKD protocol is ϵ_{EC}-correct. The defining feature of a two-universal hash function is the probability that two outputs of length $\lceil \log\left(\frac{1}{\epsilon_{EC}}\right) \rceil$ coincide, given that the inputs are different, is small, namely: $2^{-\lceil \log\left(\frac{1}{\epsilon_{EC}}\right) \rceil}$. In formulas, we have that:

$$\begin{aligned}
&\Pr\left[f_{\text{EC}}\left(A^n\right) = f_{\text{EC}}\left(\hat{A}^n\right), A^n \neq \hat{A}^n\right] \\
&\leq \Pr\left[f_{\text{EC}}\left(A^n\right) = f_{\text{EC}}\left(\hat{A}^n\right) \mid A^n \neq \hat{A}^n\right] \\
&\leq 2^{-\lceil \log(1/\epsilon_{\text{EC}}) \rceil} \\
&\leq \epsilon_{\text{EC}} \,.
\end{aligned} \tag{4.9}$$

For security, we need that the keys that are put through the hash function in privacy amplification are correct. If the error-corrected keys after the error correction step are the same (which happens with at least $1 - \epsilon_{\text{EC}}$ probability), then their hashes are guaranteed to be the same. This implies that the protocol is ϵ_{EC} even after privacy amplification:

$$\begin{aligned}
&\Pr\left[K_A \neq K_B\right] \\
&= \Pr\left[K_A \neq K_B, f_{\text{EC}}(A^n) = f_{\text{EC}}(\hat{A}^n)\right] \\
&\leq \Pr\left[A^n \neq \hat{A}^n, f_{\text{EC}}(A^n) = f_{\text{EC}}(\hat{A}^n)\right] \\
&\leq \epsilon_{\text{EC}} \,.
\end{aligned} \tag{4.10}$$

This is how the protocol guarantees ϵ_{EC} correctness.

It is also important to know how much information has been leaked to Eve during the error correcting code. Typically, all of the bits of communication sent from Alice to Bob in the error correction protocol are considered to be leaked bits of information to Eve. The amount of communication will depend on the particular error correcting code used. Let O_{EC} denotes all the classical communication in the error correction step. It includes the communication for the error correction, the hash function f_{EC}, and the hashed values of Alice's key, which are communicated via a public channel. Note that the total information of Eve E_T in Eq. (4.8) includes the classical communication O_{EC} and the side information of Eve E. To remove the dependence on the

information exchanged by Alice and Bob during error correction [TLGR12]

$$H_{\min}^{\epsilon}(A^n \mid EO_{EC})_{\rho} \geq H_{\min}^{\epsilon}(A^n \mid E)_{\rho} - \text{leak}_{\text{EC}}\,, \tag{4.11}$$

where leak_{EC} is the amount of bits communicated by Alice and Bob during information reconciliation. One can also looks into [RW05, Ren08, SR08a, SR08b, RR12, TMMPE14] for more detail.

The minimum leakage of a one-way information reconciliation protocol can be bounded using max-entropy following [RW05, MvDR+19]

$$\text{leak}_{\text{EC}} \leq H_{\max}^{\frac{\epsilon'_{\text{EC}}}{2}}(A^n \mid B^n) + \log\left(\frac{8}{\epsilon'_{\text{EC}}} + \frac{2}{2-\epsilon'_{\text{EC}}}\right) + \log\left(\frac{1}{\epsilon_{\text{EC}}}\right). \tag{4.12}$$

Now, the Asymptotic equipartition property [Tom15] reads:

Theorem 4.3 (Asymptotic equipartition property [TCR09])**.** *Let $\rho = \rho_{AE}^{\otimes n}$ be an IID state. Then for $n \geq \frac{8}{5}\log\frac{2}{\epsilon^2}$*

$$H_{min}^{\epsilon}(A^n|E^n)_{\rho_{AE}^{\otimes n}} > nH(A|E)_{\rho_{AE}} - \sqrt{n}\delta(\epsilon,\chi)\,,$$

and similarly

$$H_{max}^{\epsilon}(A^n|E^n)_{\rho_{AE}^{\otimes n}} < nH(A|E)_{\rho_{AE}} + \sqrt{n}\delta(\epsilon,\chi)\,,$$

where $\delta(\epsilon,\chi) = 4\log(\chi)\sqrt{\log\frac{2}{\epsilon^2}}$ and $\chi = \sqrt{2^{-H_{min}(A|E)_{\rho_{AE}}}} + \sqrt{2^{H_{max}(A|E)_{\rho_{AE}}}} + 1$.

Using Theorem 4.3, the min-entropy of Eq. (4.11) and the max entropy of Eq. (4.12) can be converted to n single round von Neumann entropy under the assumption of collective attacks. For coherent attacks, one can use the Post-selection technique [CKR09], uncertainty principle [TR11] and entropy accumulation theorem [DFR20].

Thus by combining Eq. (4.8), Eq. (4.11), Eq. (4.12) and by using Theorem 4.3, one obtains the following asymptotic secret key rate for a generic QKD protocol:

$$r_{\infty} = \lim_{n\to\infty}\frac{l}{n} = H(A|E)_{\rho} - H(A|B)_{\rho}\,.$$

4.3.4 Parameter Estimation

The first step of classical post-processing is parameter estimation. Parameter estimation can depend on the model of the protocol. Depending on the

protocol, the parties need to estimate different parameters to gain statistical knowledge about the output strings of their respective measurements.

In this step, the parties will use a small sample of the strings to estimate a global property of those strings so that they can bound the $H(A|E)_\rho$ and $H(A|B)_\rho$, discussed in the previous section. These bounds are generally statistical inequalities that state that the statistical property of the sample must be close to the statistical property of the entire set if a random subset of data is known. In terms of sample size, the proximity is exponential.

Example 1: One example of parameter estimation is the Bell violation estimation which is an important step in device-independent QKD protocols (for detailed discussion, see Chap. 5 and Chap. 6) which is then used to bound the von Neumann entropy. Consider B (specified by the coefficients $\{h(ab|xy)\}$) is the Bell inequality the parties decided to use in the DIQKD protocol, and $B[\mathbf{P}]$ is the Bell value corresponding to the probability distribution $\mathbf{P}$. In this scenario, the Bell inequality reads:

$$\sum_{a,b,x,y} h(ab|xy)P(ab|xy) \leq c\,, \tag{4.13}$$

with the Bell value $B[\mathbf{P}] = \sum_{a,b,x,y} h(ab|xy)P(ab|xy)$. However, in an experiment, the parties do not have access to probabilities but frequencies. To carry out the Bell value estimation from the frequencies, Alice and Bob publicly announce their measurement settings and outcomes of the parameter estimation round's data (see Sec. 4.3.1). They will calculate the Bell value $B[\hat{\mathbf{P}}]$ using the frequencies $\hat{\mathbf{P}}$, where $\hat{\mathbf{P}} = \{\hat{P}(ab|xy)\}$. $\hat{P}(ab|xy)$ is defined as:

$$\hat{P}(ab|xy) = \frac{N(a,b,x,y)}{N(x,y)}\,,$$

where $N(a,b,x,y)$ is the number of occurrences of the input-output pair. $N(x,y)$ is the number of instances when Alice chooses measurement $x \in X$ and Bob chooses measurement $y \in Y$. The Bell value $B[\hat{\mathbf{P}}]$ is a function of the joint frequencies as

$$B[\hat{\mathbf{P}}] = \sum_{a\in A, b\in B, x\in X, y\in Y} h(ab|xy)\hat{P}(ab|xy)\,. \tag{4.14}$$

To estimate the deviation of $B[\hat{\mathbf{P}}]$ (the Bell value obtained by the frequencies) from the real Bell value $B[\mathbf{P}]$, Alice and Bob use Hoeffding's inequality

[Hoe63, Hoe94].

Theorem 4.4. *Let $X_1, X_2, \cdots, X_n$ be independent random variables strictly bounded by the intervals $[a_i, b_i]$, i.e. $a_i \leq X_i \leq b_i$. We define*

$$\bar{X} = \frac{1}{n}(X_1 + X_2 + \cdots + X_n)\,.$$

Then, Hoeffding's inequality reads

$$\Pr\left(\bar{X} - E[\bar{X}] \geq t\right) \leq \exp\left\{-\frac{2n^2t^2}{\sum_{i=1}^{n}(b_i - a_i)^2}\right\}.$$

Let $c_i := b_i - a_i$ and $c_i \leq C \;\forall\; i$. Then, Hoeffding's inequality reads

$$\Pr\left(\bar{X} - E[\bar{X}] \geq t\right) \leq \exp\left\{-\frac{2n^2t^2}{nC^2}\right\} = \exp\left\{-\frac{2nt^2}{C^2}\right\}.$$

Let $\chi(e)$ be an indicator function for a particular event e, i.e. $\chi(e) = 1$ if the event e is observed, $\chi(e) = 0$ otherwise. We consider a random variable

$$\hat{B}_i = \sum_{a\in A, b\in B, x\in X, y\in Y} h(ab|xy)\frac{\chi(a_i = a, b_i = b, x_i = x, y_i = y)}{\hat{p}(x_i = x, y_i = y)}\,,$$

where, $\hat{p}(x_i = x, y_i = y) = \frac{N_{x,y}}{N}$ is the input joint frequency distribution. We get $\frac{1}{N}\sum_{i=1}^{N}\hat{B}_i = B[\hat{\mathbf{P}}]$ and $E[\frac{1}{N}\sum_{i=1}^{N}\hat{B}_i] = B[\mathbf{P}]$, where $E[\cdot]$ denotes the expectation value. Defining

$$q_{\min} = \min_{a,b,x,y}\frac{h(ab|xy)}{\hat{p}(x_i = x, y_i = y)}\,,$$
$$q_{\max} = \max_{a,b,x,y}\frac{h(ab|xy)}{\hat{p}(x_i = x, y_i = y)}\,,$$

we have $q_{\min} \leq \hat{B}_i \leq q_{\max}$, where $a \in A$, $b \in B$, $x \in X$, $y \in Y$. We define, $\gamma := q_{max} - q_{min}$. By using Hoeffding's inequality (see Theorem 4.4), the parties can bound the deviation δ of the Bell value obtained by the frequencies from the asymptotic value by a probability:

$$\Pr\left(B[\mathbf{P}] \geq B[\hat{\mathbf{P}}] - \delta\right) \geq 1 - \epsilon, \tag{4.15}$$

with

$$\epsilon = \exp\left\{\left(-\frac{2N\delta^2}{\gamma^2}\right)\right\}. \tag{4.16}$$

For a given ϵ of a QKD protocol, one can calculate the confidence interval δ using Theorem 4.4. The parties will use $B[\hat{\mathbf{P}}] - \delta$ as the Bell value in the QKD protocol since the asymptotic Bell value $B[\mathbf{P}]$ is larger than $B[\hat{\mathbf{P}}] - \delta$ with ϵ error probability. For asymptotic scenario, $B[\mathbf{P}]$ and $B[\hat{\mathbf{P}}]$ coincides as $N \to \infty, \epsilon, \delta \to 0$ (see Eq. (4.16)).

Example 2: Another example of parameter estimation is to estimate the QBER (Quantum bit error rate) Q. To accomplish this step, Alice sends Bob a small sample of her key generation measurement rounds' outcomes through the authenticated classical channel. She can also publicly announce her string via a public channel. From the sample, the parties will calculate the QBER $\hat{Q}$ of the key generation round and bound it by using the following theorem:

Theorem 4.5. *[GKB19, YC19] Let $\mathcal{X}_{n+k}$ be a random binary string of $n+k$ bits, $\mathcal{X}_k$ be a random sample (without replacement) of m entries from the string $\mathcal{X}_{n+k}$ and $\mathcal{X}_n$ be the remaining bit string. Λ_k and Λ_n are the frequencies of bit value 1 in string $\mathcal{X}_k$ and $\mathcal{X}_n$, respectively. For any $\epsilon_1 > 0$, it holds the upper tail inequality:*

$$\Pr[\Lambda_n \geq \Lambda_k + \gamma_1(n, k, \Lambda_k, \epsilon_1)] > \epsilon_1\,, \tag{4.17}$$

where $\gamma_1(a, b, c, d)$ is the positive root of

$$\ln\binom{bc}{b} + \ln\binom{ac + a\gamma_1(a, b, c, d)}{a} = \ln\binom{(a+b)c + a\gamma_1(a, b, c, d)}{a+b} + \ln d\,.$$

For $\epsilon_2 > 0$, we have the lower tail inequality:

$$\Pr[\Lambda_n \leq \Lambda_k - \gamma_2(n, k, \Lambda_k, \epsilon_2)] > \epsilon_2\,, \tag{4.18}$$

where $\gamma_2(a, b, c, d)$ is the positive root of

$$\ln\binom{bc}{b} + \ln\binom{ac - a\gamma_2(a, b, c, d)}{a} = \ln\binom{(a+b)c - a\gamma_2(a, b, c, d)}{a+b} + \ln d\,.$$

Using Theorem 4.5, the parties can deduce that the QBER characterizing the key generation round is not larger than $\hat{Q}+\gamma$ (estimated QBER + statistical correction) with very high probability and use $\hat{Q} + \gamma$ as the QBER of the key

generation rounds for the QKD protocol. In the asymptotic limit where both n and k diverge, we have that Q and $\hat{Q}$ coincide. If Alice and Bob see their Bell violation or error rate exceeding the allowed threshold, they abort the protocol. Otherwise, they continue.

Now, we will discuss the BB84 QKD protocol. It is probably the most analyzed protocol, not only due to it being the first, but also due to its simplicity and symmetry. The BB84 protocol has several security proofs that apply under various assumptions [Ren08, KGR05, RGK05, TLGR12, GLLP04]. Here, we mostly follow the protocol stated in [SBPC+09, Gra21].

4.4 BB84 Protocol

The quantum key distribution protocol that goes under the name BB84 [BB84] was proposed by Bennet and Brassard (hence the name). The security principle lies in the fact that Alice encodes the key bits in non-orthogonal states and sends them to Bob. Therefore if Eve attempts to learn the key bits by intercepting the states in the quantum channel, she can not avoid introducing noise as she cannot perfectly distinguish the states [NC10].

Let's say Alice has a single photon source with well-defined spectral properties in her possession. Alice and Bob align their polarizers in two distinct polarization bases. One polarization base is specified by the horizontal/vertical directions $(0°/90°)$ which is associated with the eigenbasis $\{|0\rangle, |1\rangle\}$ of Pauli operator Z. The other polarization base is specified by the diagonal/anti-diagonal directions $(+45°/-45°)$, which is associated with the eigenbasis $\{|+\rangle, |-\rangle\}$ of Pauli operator X, where $|\pm\rangle = (|0\rangle \pm |1\rangle)/\sqrt{2}$. Alice and Bob equate the non-orthogonal states $|0\rangle$ and $|+\rangle$ ($|1\rangle$ and $|-\rangle$) with the bit value 0(1). Any interference with the quantum channel by Eve to gain information causes disturbance in the transmitted signal because of the no-cloning theorem. Alice and Bob can identify the signal's disturbance, ensuring the protocol's security. The BB84 protocol consists of the following steps:

1. In every round of the protocol, the parties do the following:
    - Alice chooses random bits x and a.
    - If $x = 0$, Alice uses the Z -basis to encode a (she prepares the state $|0\rangle$ if $a = 0$, and prepares $|1\rangle$ if $a = 1$). Similarly, if $x = 1$, Alice uses the X -basis to encode a (she prepares the state $|+\rangle$ if $a = 0$, and prepares $|-\rangle$ if $a = 1$).

- Alice sends the prepared state to Bob through the insecure quantum channel.
- Bob announces whether he received the state.
- Bob randomly chooses a bit y and records the outcome b. If Bob measures in the same basis Alice used to prepare the photon, he learns the bit she encoded on that photon, provided that the signal has not been altered. If Bob measures in the complementary basis, he obtains a random bit since the two bases are mutually unbiased.

2. **Sifting:** Alice and Bob publicly announce their choices of basis, x and y, and compare them. They discard the rounds in which Bob measured in a different basis than the one prepared by Alice, i.e. when $x \neq y$. In absence of errors due to noise or eavesdropping, the strings of Alice and Bob would coincide.

3. **Parameter estimation:** Alice and Bob use a fraction of the rounds (in which both measured in the same basis) in order to estimate the quantum bit error rates (QBERs) Q_X and Q_Z.

$$\begin{aligned} Q_z &= P(a \neq b \mid x = y = 0)\,, \\ Q_x &= P(a \neq b \mid x = y = 1)\,. \end{aligned} \tag{4.19}$$

4. **Information reconciliation:** Alice and Bob perform one-way error correction and communicate over the authenticated public channel in order to correct their string of bits. At the end of this step, Alice and Bob should hold the same bit-string.

5. **Privacy amplification:** Alice and Bob use an extractor on the error corrected strings to generate shorter, but completely secret strings of ℓ bits, which are their final keys.

Given that one party prepares and communicates quantum states while the other measures them, the protocol above is described in a prepare-and-measure form. The last three steps of the BB84 protocol are needed since the channel between Alice and Bob might be noisy, and all the observed noise is attributed to the actions of Eve to ensure security.

One example of an eavesdropping technique used by Eve is the intercept-resend attack. Eve randomly selects the X or Z basis for each round. Eve cannot completely clone non-orthogonal states; thus, she must make a measurement to gather information. By doing this, she unintentionally introduces errors in rounds where Alice's choice and her choice conflict. Eve's

interference appears in measurement results for Bob (who is awaiting a signal) in the absence of noise, which can only be explained by a third party's interaction.

An analogous entanglement-based explanation is far more practical when demonstrating the security of a QKD system or calculating its key rate.

4.5 Entanglement Based Version

In the entanglement-based version of the BB84 protocol [Ben92], Alice is in control of a quantum source and both parties hold measurement devices with two inputs, $x, y \in \{0, 1\}$. This protocol consists of the following steps:

1. In every round of the protocol, the parties do the following:

   - A source prepares the Bell state $\left|\phi^+\right\rangle$ and distributes to Alice and Bob.
   - Alice chooses an input $x \in \{0, 1\}$ uniformly at random. If $x = 0$, Alice measures her part of the system in the Z-basis. She measures in the X-basis if $x = 1$. Alice records her outcome a.
   - Similarly, Bob chooses a random bit y. If $y = 0$, Bob measures her part of the system in the Z-basis. If $y = 1$ he measures in the X-basis. He also records his outcome b.

2. **Classical post-processing:** Classical post-processing step consists of sifting, parameter estimation, error correction and privacy amplification which are the same as the BB84 protocol (see Sec. 4.4) discussed previously.

4.6 Asymptotic Secret-Key Rate of BB84 Protocol

Now we will calculate the secret key rate for the BB84 protocol, which represents the value achievable by the secret key rate in the asymptotic limit of infinitely many uses of the quantum channel. Here we only consider the asymptotic secret key rate. The finite size effects are discussed in [Ren08, SR08a, Gra21].

Assuming one-way classical post-processing, the asymptotic secret key

rate of a QKD protocol is lower bounded by Devatek-Winter rate [DW05]

$$\begin{aligned} r_{DW} &\geqslant H(A:B) - \chi(A:E) \\ &= H(A:B) - \left(H\left(\rho_{\mathrm{E}}\right) - \sum_{a=\pm 1} P(a) H\left(\rho_{\mathrm{E}|a}\right) \right) \end{aligned} \tag{4.20}$$

which is the difference between the mutual information $H(A:B)$, Eq. (2.39), (between Alice and Bob) and the Holevo quantity $\chi(A:E)$, Eq. (2.43), (between Alice and Eve). The reduced state $\rho_{\mathrm{E}} = \sum_a P(a) H\left(\rho_{\mathrm{E}|a}\right)$ of Eve is a mixture of states $\rho_{\mathrm{E}|a}$, conditioned on the value of Alice's signal. The Devetak-Winter rate has the following interpretation: The mutual information describes the amount of information shared by Alice and Bob. Due to the action of Eve, this information is only partially secure, which is why the Holevo quantity, an upper bound on Eve's accessible information, is subtracted. In this worst-case scenario, the security of the remaining information is ensured. However, the Eq. (4.20) can be recast as [SR08a, SR08b, Ren08]

$$r = H(A|E) - H(A|B)\,. \tag{4.21}$$

We have already seen this expression in Sec. 4.3.3. Now we will calculate the secret key rate in terms of the estimated QBER Q_z and Q_x (see Sec. 4.4). We use an asymmetric variant of the BB84 protocol in the computation, where the raw key is only derived from Z basis measurements, and PE is applied to the X outcomes (together with a fraction of Z outcomes) for the purpose of simplicity.

We assume that the marginal distributions of Alice's and Bob's outcomes are symmetrized. This assumption does not alter the correlation of the raw keys of Alice and Bob, as well as the observed QBERs. Indeed, if this is not the case, Alice and Bob can symmetrize them by deciding to flip their outcomes with a probability of $1/2$ for each protocol round. They can agree on which outcomes to flip over the classical public channel. Note that this does not change Eve's information since she listens to the public channel. Hence security proof can be restricted to the symmetrized scenario. Due to the symmetrization of the marginals, the conditional Shannon entropy $H(A|B)$ can be expressed exclusively in terms of QBER Q_z as follows:

$$H(A|B) = h(Q_z)\,, \tag{4.22}$$

where h is the binary entropy, Eq. (2.36).

For the calculation of the $H(A|E)$, we follow [RGK05, KGR05]. Due to the symmetry of the BB84 protocol, it is not restrictive to assume that the final state Alice and Bob share is Bell diagonal. That is,

$$\tilde{\rho}_{\mathrm{AB}} = \lambda_{00}\Phi_{00} + \lambda_{01}\Phi_{01} + \lambda_{10}\Phi_{10} + \lambda_{11}\Phi_{11}\,, \tag{4.23}$$

where $\Phi_{ij} = |\phi_{ij}\rangle\langle\phi_{ij}|$, and $0 \leq \lambda_{ij} \leq 1$, $\sum_{i,j}\lambda_{ij} = 1$ for $i, j \in \{0,1\}$. Here, we relabelled the Bell states according to

$$|\phi^+\rangle \rightarrow |\phi_{00}\rangle\,, \tag{4.24}$$
$$|\phi^-\rangle \rightarrow |\phi_{10}\rangle\,, \tag{4.25}$$
$$|\psi^+\rangle \rightarrow |\phi_{01}\rangle\,, \tag{4.26}$$
$$|\psi^-\rangle \rightarrow |\phi_{11}\rangle\,. \tag{4.27}$$

The state $\tilde{\rho}_{\mathrm{AB}}$ can be obtained by the following operation:

$$\tilde{\rho}_{\mathrm{AB}} = \frac{1}{4}\Big(\rho_{AB} + X\otimes X\rho_{AB}X\otimes X + Y\otimes Y\rho_{AB}Y\otimes Y + Z\otimes Z\rho_{AB}Z\otimes Z\Big)\,. \tag{4.28}$$

The State $\tilde{\rho}_{AB}$ preserves the Bell diagonal elements, i.e.

$$\big\langle\Phi_{ij}\big|\rho_{AB}\big|\Phi_{ij}\big\rangle = \big\langle\Phi_{ij}\big|\tilde{\rho}_{AB}\big|\Phi_{ij}\big\rangle = \lambda_{ij} \tag{4.29}$$

for $i, j \in \{0,1\}$. The assumption about the Bell diagonal state is not restrictive because $H(A|E)_{\rho_{AB}} \geq H(A|E)_{\tilde{\rho}_{AB}}$ and the observed QBERs are also unaffected; see [Gra21] for detailed explanation. Now given the parties share $\tilde{\rho}_{AB}$, Q_z and Q_x relate to the Bell coefficients by

$$\begin{aligned} Q_z &= \lambda_{01} + \lambda_{11}\,, \\ Q_x &= \lambda_{10} + \lambda_{11}\,. \end{aligned} \tag{4.30}$$

The global pure state $|\phi_{ABE}\rangle$ of Alice, Bob and Eve given that Eve holds the purifying system of $\tilde{\rho}_{AB}$ reads:

$$|\phi_{ABE}\rangle = \sum_{i,j=0}^{1}\sqrt{\lambda_{ij}}\,|\phi_{ij}\rangle_{AB}\otimes|e_{ij}\rangle_{E}\,, \tag{4.31}$$

where $|e_{ij}\rangle_E$ is an orthonormal basis in $\mathcal{H}_E$. From Eq. (4.31), $H(A|E)_{\tilde{\rho}_{AB}}$ can

be expressed as:

$$H(A|E)_{\tilde{\rho}_{AB}} = 1 + h(Q_z) - H(\{\lambda_{ij}\})\,, \tag{4.32}$$

where $H(\{\lambda_{ij}\}) = \sum_{ij} = -\lambda_{ij} \log_2 \lambda_{ij}$. Since we have to consider worst case scenario, we have to minimize Eq. (4.32) with the constraints of QBER in Eq. (4.30). The result of the minimization is as follows [SBPC+09]:

$$\begin{aligned} H(A|E) &= 1 + h(Q_z) - \big(h(Q_x) + h(Q_z)\big) \\ &= 1 - h(Q_x)\,. \end{aligned} \tag{4.33}$$

Putting Eq. (4.33) and Eq. (4.22) together, we obtain the asymptotic key rate for the BB84 protocol:

$$r_{\mathrm{BB84}} = 1 - h(Q_x) - h(Q_z)\,. \tag{4.34}$$

5

Device-independent Approach to Quantum Key Distribution

There have been many different QKD protocols [Eke91, Ben92, Bru98, Ren08, LMC05, GLLP04, SP00, SBPC+09, MQZL05, LCT14, TLGR12] introduced since the advent of the QKD in 1984 with the famous BB84 protocol [BB84]. A complete description of the devices, sources, and/or communication channels between the parties is required for the security of these device-dependent protocols. For instance, if the devices are making measurements in four dimensional systems instead of qubits, the protocol can be broken easily [PAB+09]. It is challenging to fully describe a device utilized in the experiment in a real-world context. Even a malicious spy may have prepared the device (Eve). Additionally, it was shown how to hack into existing implementations to take advantage of experimental imperfections [LWW+10, GLLL+11, ZFQ+08]. Device-independent quantum key distribution (DIQKD) was introduced to overcome these drawbacks. In DIQKD, the security does not require any assumptions about the inherent properties of the devices or the dimension of the Hilbert space of the quantum signals and thus claims the highest level of security in quantum cryptography. This method of relaxing the assumptions about the underlying equipment prevents the eavesdropper from being able to exploit the inadequacies of the involved devices, thus effectively eliminating the threat of side-channel attacks.

In DIQKD, the quantum apparatuses are treated as black boxes that produce classical outputs given an input. The only relevant information is the statistics of inputs and outputs. The security of DI protocols requires a

DI witness to certify nonlocal correlations. Thus, it is crucial to incorporate a Bell test (Bell inequality violation) in a DIQKD protocol. The observed Bell violation quantifies the degree of nonlocality. Higher Bell violation equates to a lower degree of correlation with any other system due to the monogamy of entanglement [CKW00]. Monogamy was already explicitly established in the Ekert protocol [Eke91]. This is the underlying physical basis that provides DI security.

The concept of certifying a certain device behavior through input-output correlations was initially proposed in [MY98]. The first quantitative step towards formalizing the security of DIQKD was achieved by bounding the information of a no-signalling eavesdropper about a single signal between Alice and Bob [BHK05]. The subsequent publications assume that the devices behave identically and independently (IID) throughout the measurement rounds [ABG+07, AGM06, PAB+09, HR10, HRW10] to prove the security of the DIQKD protocol. However, the IID assumption (collective attacks) is generally not justified in the DI scenario. The first security proof in the fully DI scenario, i.e. without the IID assumption, is presented in [VV14]. Later DI protocols without the IID assumptions (coherent attack) are proved in [AFDF+18, AFRV19] using the entropy accumulation [DFR20]. Recently, DIQKD protocols are also developed in the directions of multiparty setting (also called conference key agreement) in [RMW18, RMW19, HKB20, GMKB21].

We open this chapter with Sec. 5.1 discussing the assumptions of the DIQKD scenario. Afterwards, we review the general structure of a DIQKD protocol that uses CHSH violation as a Bell test in Sec. 5.2. In Sec. 5.3, an analytical bound on the asymptotic secret key rate is specified based on [PAB+09]. Sec. 5.4 outlines how numerical bounds on the DI secret key rate can be achieved in a general Bell setting [MPA11] using NPA Hierarchy. We end the chapter with a discussion about the current status of experimental realizations of DIQKD in Sec. 5.5.

5.1 Assumptions of DIQKD

The assumptions made in a BB84 implementation are significantly loosened in the device-independent situation. However, it is important to note which assumptions remain in implementing a DI protocol. We assume [PAB+09, AFDF+18]:

1. Alice's and Bob's physical locations are isolated. No unwanted infor-

mation can leak out to the outside.

2. Alice and Bob each have access to independent and trusted random number generators.

3. Trustworthy computers carry out the local classical computations, and all classical public communication takes place over an authenticated channel.

4. They have trusted classical devices (e.g. memories and computing devices) to store and process the classical data generated by their quantum apparatuses.

5. State preparation is independent of the measurements performed on them.

6. Quantum Mechanics is correct.

Another presumption frequently employed in security proofs is that the experiment rounds are independent and identically distributed (IID). The IID implementation assumes that the devices behave independently and in the same way in every round of the protocol. It also assumes that the states are distributed in the same for every round of the protocol. In summary, the state of the n rounds can be written as $\rho_{A^nB^nE} = \rho_{ABE}^{\otimes n}$. Note that collective attacks from the eavesdropper assume the IID assumption, whereas coherent attacks do not satisfy the same.

5.2 Device-independent Quantum Key Distribution Protocol

The First ideas of device-independent QKD arouse from the E91 protocol [Eke91], which uses the CHSH inequality [CHSH69] violation to check the presence of an eavesdropper. Recall, the CHSH inequality reads:

$$\mathcal{S}_{\text{CHSH}} := \langle A_0B_0\rangle + \langle A_0B_1\rangle + \langle A_1B_0\rangle - \langle A_1B_1\rangle\,, \tag{5.1}$$

where $\langle A_xB_y\rangle = p(a = b|xy) - p(a \neq b|xy)$ represents the correlation of the outputs a of Alice and b of Bob when they perform the measurements labeled by x and y, respectively.

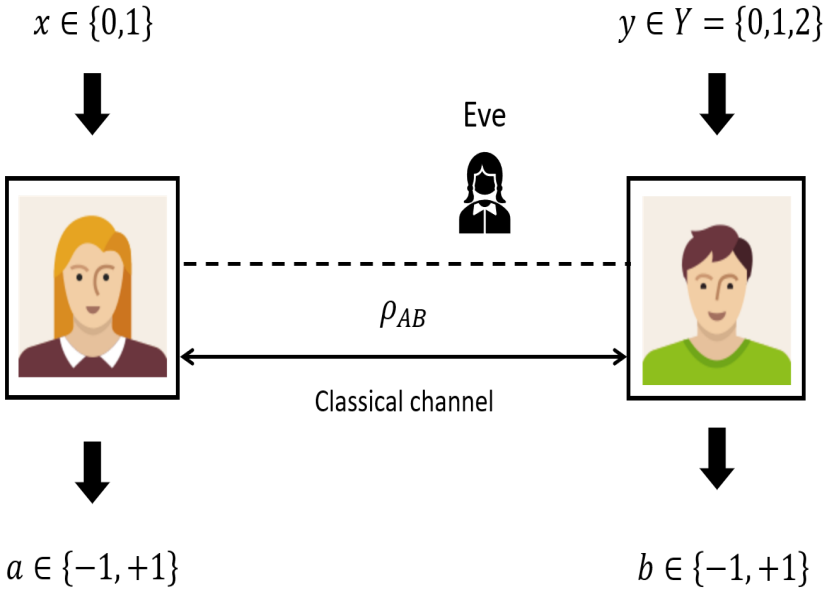

Figure 5.1: Schematic representation of a DIQKD scenario. A source (usually assumed to be controlled by Eve) distributes states to Alice and Bob. Alice can choose between two different measurements with $x \in \{0,1\}$ with outcomes $a \in \{-1,+1\}$, while Bob can choose between three different measurements $y \in \{0,1,2\}$, also with outcomes $b \in \{-1,+1\}$. A classical, authenticated communication channel is also available to Alice and Bob. The measurement devices in the DIQKD scenario are untrusted; therefore, they are under Eve's control in the worst case.

Now assume, Alice has an uncharacterized measurement device with two inputs $x \in X = \{0,1\}$, which outputs $a \in A = \{-1,+1\}$ upon measurement. Likewise, Bob has another uncharacterized measurement device with three inputs $y \in Y = \{0,1,2\}$ with dichotomic outcomes $b \in B = \{-1,+1\}$. See Fig. 5.1 for a pictorial representation. We presume that the device's behavior conforms to the IID assumption, and Eve performs collective attacks. Specifically, we assume that the total state shared by the three parties has the product form $|\psi_{\mathbf{ABE}}\rangle = |\psi_{ABE}\rangle^{\otimes n}$ and Eve holds the purification.

The protocol involves the following steps:

1. In every round of protocol the parties do the following:

   - An unknown state ρ_{AB} is distributed between Alice and Bob.
   - In each round, the parties measure their share of quantum state ρ_{AB}. There are two types of measurement rounds, the key generation rounds and the parameter estimation round. The parameter estimation round's measurement is much less frequent. A pre-

shared random key determines the type of each round. Alice and Bob choose the input $(x, y) = (0, 2)$ for key generation measurement rounds. In parameter estimation measurement rounds, the parties randomly choose their inputs $x, y \in \{0, 1\}$.

- The parties record their inputs and outputs as (x_i, y_i) and (a_i, b_i), respectively. After N rounds of measurement, Alice and Bob hold the input bit strings X^N and Y^N and the output bit strings A^N and B^N, respectively.

2. **Parameter estimation**: Using the measurement outcomes of parameter estimation measurement rounds, Alice and Bob estimate the CHSH value $\mathcal{S}_{\text{CHSH}}$. They also use a portion of key generation rounds to estimate the QBER, which reads:

$$Q = p(a \neq b | x = 0, y = 2) .$$

The rest of the rounds are used for the raw keys.

3. **Information reconciliation**: Alice and Bob implemented an one way error correction protocol to correct their bit-strings. At the end of this step, both Alice and Bob will have the same bit-string.

4. **Privacy amplification**: Alice and Bob utilize an extractor to convert the error-corrected, partially secret bit-strings into shorter, fully secret (unknown to the eavesdropper) strings of ℓ bits, which are their final keys.

5.3 Asymptotic Device-independent Secret Key Rate

Consider the DIQKD protocol described above in Sec. 5.2. The asymptotic secret-key rate generated by the protocol stated above is lower bounded by the Devetak-Winter [DW05] rate:

$$r_\infty \geqslant r_{DW} = H(A_0 : B_2) - \chi(A_0 : E) , \tag{5.2}$$

which is the difference between the mutual information between Alice and Bob and the Holevo quantity between Alice and Eve. We can assume w.l.o.g. that the marginal probabilities are uniform, i.e. $\langle A_x \rangle = \langle B_y \rangle = 0$. It can be achieved by classical post-processing and it has no impact on the values of QBER Q and the CHSH violation $\mathcal{S}_{\text{CHSH}}$. Therefore, the mutual information

is given by $H(A_0 : B_2) = 1 - h(Q)$, where h is the binary entropy. The Holevo quantity reads:

$$\chi(A_0 : E) = H(\rho_E) - \frac{1}{2} \sum_{a_0=\pm 1} H\left(\rho_{E|a_0}\right) , \tag{5.3}$$

where $\rho_E = \mathrm{Tr}_{AB}\left(|\psi_{ABE}\rangle \langle \psi_{ABE}| \right)$ denotes Eve's quantum state after tracing out Alice and Bob's share of the state, and $\rho_{E|a_0}$ is Eve's quantum state when Alice has obtained the result a_0 for the measurement A_0.

The goal is now to upper bound the Holevo quantity in terms of CHSH inequality violation $\mathcal{S}_{\mathrm{CHSH}}$.

Theorem 5.1 (Upper bound of Holevo quantity)**.** *Let a quantum state $|\psi\rangle_{ABE}$ and the set of measurement operators $\{A_0, A_1, B_0, B_1\}$ yields the CHSH violation of $\mathcal{S}_{\mathrm{CHSH}}$. Then after Alice and Bob have symmetrized their marginals,*

$$\chi(A_0 : E) \leqslant h\left(\frac{1 + \sqrt{\mathcal{S}^2_{\mathrm{CHSH}}/4 - 1}}{2} \right) \tag{5.4}$$

Proof. We refer the reader to Ref. [PAB+09] for the proof. □

From the upper bound of Holevo quantity, it follows:

$$r_\infty \geqslant 1 - h(Q) - h\left(\frac{1 + \sqrt{\mathcal{S}^2_{\mathrm{CHSH}}/4 - 1}}{2} \right) . \tag{5.5}$$

5.4 Quantifying Device-independent Secret Key Rate via Min-entropy

In this section, we describe how the min-entropy and NPA hierarchy (see Sec. 3.5.2 for details) can be used to lower bound the DI secret key rate (DISKR) in terms of an observed Bell inequality violation. We follow the methods described in [MPA11]. The bipartite asymptotic DISKR can also be expressed as [MvDR+19, AFDF+18]

$$r_\infty \geqslant H(A|X, E) - H(A|B) . \tag{5.6}$$

In a general Bell scenario, the conditional von Neumann entropy $H(A|X,E)$ is hard to calculate analytically. As already mentioned (see Eq. (2.49)), conditional von Neumann entropy can be lower bounded by the conditional min-entropy as: $H(A|X,E) \geq H_{\min}(A|X,E)$. In Eq. (2.47), we mentioned that min-entropy can be expressed in terms of guessing probability, i.e. $H_{\min}(A|X,E) = -\log_2 p_{\text{guess}}(A|X,E)$. Our objective is to find an upper bound on $p_{\text{guess}}(A|X,E)$ to cover the worst-case scenario. Because $-\log_2(x)$ is a monotonically decreasing function of x, it provides a lower bound on $H_{\min}(A|X,E)$.

In [MPA11], it is shown that $p_{\text{guess}}(A|X,E)$ can be upper bounded as a function of the Bell value. Let $\mathcal{G}$ denote the Bell operator associated with the DIQKD setting and $B[\mathbf{P}]$ be the observed Bell value (here B is the Bell inequality, and $\mathbf{P}$ is the probability distribution). We can write $p_{\text{guess}}(A|X,E) \leqslant G_x\left(B[\mathbf{P}]\right)$, where G_x is a concave and monotonically decreasing function of the Bell value $B[\mathbf{P}]$. The bound $G_x\left(B[\mathbf{P}]\right)$ can always be established with the NPA hierarchy [NPA07, NPA08] by solving the following semidefinite programme:

$$\begin{aligned} &\max_{\rho_{\text{AB}},\{A(a|x)\},\{B(b|y)\}} p_{\text{guess}}(A|X,E) \\ &\text{subject to: } \operatorname{Tr}\left(\mathcal{G}\rho_{\text{AB}}\right) = B[\mathbf{P}]\,. \end{aligned} \tag{5.7}$$

The solution of the SDP in Eq. (5.7) provides the maximum possible value for $p_{\text{guess}}(A|X,E)$ for a fixed parameter $B[\mathbf{P}]$. The maximization is performed over all states ρ_{AB} and observables $A(a|x)$, $B(b|y)$. The constraint corresponds to the checking of a Bell inequality violation. This establishes the existence of some randomness which is the fundamental tenet of the device-independent paradigm. $\mathcal{G}$ is the Bell operator, associated with the DIQKD setting, defined as:

$$\mathcal{G} = \sum_{a,b,x,y} C(ab|xy)A(a|x)B(b|y)\,, \tag{5.8}$$

where $C(ab|xy)$ are the coefficients defining the Bell inequality B (see Eq. (3.6)). This implies, a lower bound on the DI secret-key rate in the asymptotic limit is given by

$$r_\infty \geqslant -\log_2\left(G_x\left(B[\mathbf{P}]\right)\right) - H(A|B)\,, \tag{5.9}$$

where $G_x\left(B[\mathbf{P}]\right)$ is obtained from the solution of the SDP in Eq. (5.7).

We did not specify anything about the observables and states; merely an

observed Bell inequality violation is required. Bounds by SDPs of the form in Eq. (5.7) are thus device-independent and valid against the most general adversary. They are, however, often overly pessimistic since we bound min-entropy instead of von Neumann entropy. It is an open problem to obtain tighter bounds in a general setting, numerically as well as analytically. Recent developments are made where sophisticated methods of bounding the conditional von Neumann entropy [TSG+21, SGP+21, BFF21a, BFF21b] are introduced instead of bounds based on min-entropy. Another way to proceed is to use a tailored Bell inequality (that depends on the measurement statistics of Alice and Bob) for DIQKD [DKB22a, NSPS14, BSS14], which we will discuss in the next chapter.

5.5 Experimental Realization of DIQKD

Contrary to the device-dependent QKD, implementing a DIQKD in an experimental setup is quite challenging. There are several issues with this subject. One main bottleneck is to perform a detection loophole-free Bell test at a significant spatial distance while achieving adequate Bell inequality violation[1] and a low QBER. Note that the devices must operate at a decent clock-rate to suppress the finite-size effect to an acceptable tolerance level.

Loophole-free Bell test is the main ingredient of a DIQKD protocol. All the Bell experiments performed so far can be divided into two broad categories, i.e. Photonic experiments [LYL+18, LLR+21, LZZ+21, ZSB+20, CMA+13, GVW+15, SZB+21, SMSC+15, BKG+18] and the Heralded entanglement system [NDN+22, ZvLR+21, RBG+17]. In the photonic experiments, photon pair with entangled degrees of freedom is prepared, incorporating polarisation via spontaneous parametric down-conversion. Then Alice and Bob each measure their part of entangled photons using single-photon detectors. Fully photonic systems realize a high clock-rate and low QBER, but they experience low CHSH value. That's why one could not generate a non-zero device-independent secret key using a photonic set even though some achieve high enough CHSH violation for device-independent randomness generation [BKG+18, ZSB+20, LLR+21]. However, recently in [LZZ+21, LZZ+22], the authors performed a photonic DIQKD experiment asymptotically secure when Eve is restricted to collective attacks. In this ex-

[1] All existing loophole-free Bell experiments, including all the DIQKD proof-of-principle demonstrations, are based on the CHSH inequality to date.

periment, the CHSH value $\mathcal{S}_{\text{CHSH}} \approx 2.0472$ is achieved using a post-selection technique across a transmission distance of 220 meters while using single-photon detectors with efficiency $\geq 87\%$.

On the other hand, loophole-free Bell experiments using heralded entanglement involve preparing an entangled state between a long-lived quantum system and a photon. Examples of long-lived systems include trapped ions [MMO+07], atoms [HKO+12], NV-centre [BHP+13] and quantum dots [DSG+16]. Alice and Bob store the long-lived quantum system in their respective isolated laboratories while sending the photonic system for a Bell state measurement in a heralding station. A successful Bell state measurement ensures Entanglement swapping. The long-lived systems can then be measured. Detection efficiencies are high in this experimental setup compared to the photonic systems. Therefore, heralded entanglement systems provide high CHSH value and low QBER compared to the photonic systems. However, the clock-rate is lacking in this scenario due to the slow heralding rate. Recently using a trapped-ions-based heralded entanglement setup [NDN+21, NDN+22], a non-zero device-independent secret key is generated while achieving $\mathcal{S}_{\text{CHSH}} \approx 2.64$ and QBER $Q \approx 1.8\%$ over 2 meters even when the heralding rate is low. In [ZvLR+21, ZvLR+22], the authors demonstrate another DIQKD setup using heralded entangled atoms. They were able to generate an asymptotic DISKR of about 0.07 per entanglement generation event when the parties were 400 m apart. However, when finite-size effects are taken into account, the block size is too small to produce a DISKR.

6

Device-independent Quantum Key Distribution and Post-selection of Bell Inequality

Device-independent quantum key distribution (DIQKD) protocols are based on the violation of a loophole-free Bell inequality. Usually, a pre-specified Bell inequality B is chosen before performing the DIQKD protocol. For a given input-output probability distribution $\mathbf{P}$, the Bell value of a Bell inequality B is denoted as $B[\mathbf{P}]$. It is then possible, in principle, to compute an upper bound on the length of the device-independent secret key rate (DISKR) of the input-output behavior through the quantity $B[\mathbf{P}]$. Thus, one can notice that the choice of Bell inequality deeply influences the length of the generated key. Therefore, different DIQKD protocols have been proposed, where the Bell inequality is not agreed upon beforehand but constructed from the observed probability distribution of the measurement outcomes [DKB22a, NSPS14, BSS14]. In [NSPS14, BSS14], the Bell inequality, constructed from the measurement statistics, is designed in such a way that it leads to the maximal DISKR for that precise setup in the asymptotic scenario. The authors of [DKB22a] follow a two-step process. First, they construct a Bell inequality from the input-output probability distribution that leads to the maximum Bell violation for that particular measurement setting of Alice and Bob. Then they use the optimized Bell inequality and the corresponding violation to bound the DISKR.

The content of this chapter is based on the work in [DKB22a], and it

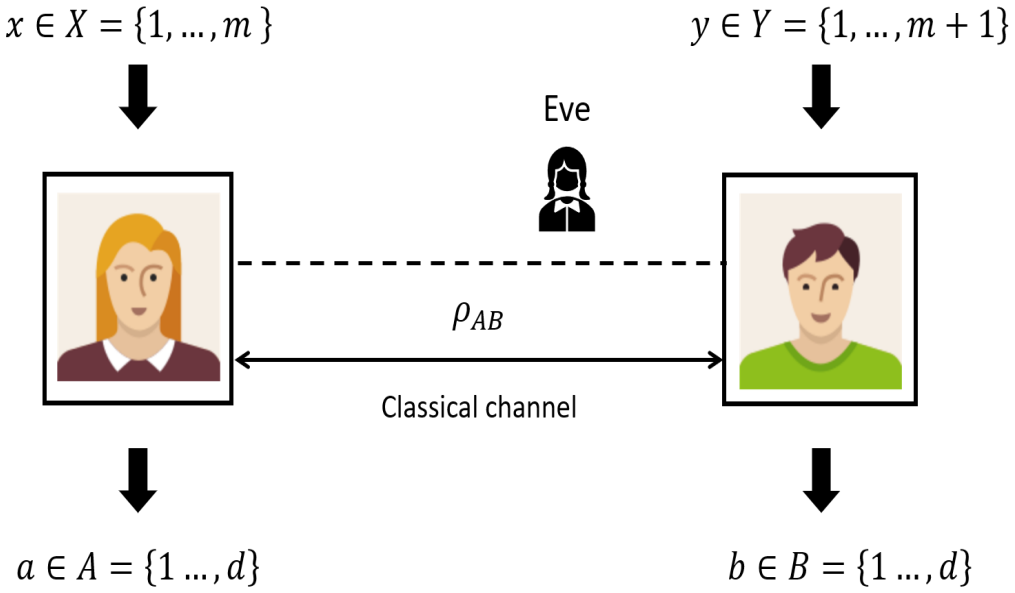

Figure 6.1: Schematic representation of a DIQKD scenario with m measurement settings per party, where each measurement setting has d outcomes. A source (usually assumed to be controlled by Eve) distributes states to Alice and Bob. Alice can choose between m different measurements with $x \in \{1, \cdots, m\}$ with outcomes $a \in \{1, \cdots, d\}$, while Bob can choose between $m+1$ different measurements $y \in \{1, \cdots, m+1\}$, also with outcomes $b \in \{1, \cdots, d\}$. An authenticated classical communication channel is also available to Alice and Bob.

is a central result of this thesis. We start by describing a DIQKD protocol where the Bell inequality is constructed from the measurement statistics, c.f. Sec. 6.1. In Sec. 6.2, we provide the mathematical expression of the secret key rate, including finite size effects. We end this chapter by explaining the utility of this approach in Sec. 6.3.

6.1 DIQKD Protocol from a Post-selected Bell Inequality

Here we follow the protocol of Ref. [DKB22a]. Consider the IID (identically and independently distributed) scenario, where the devices will behave independently and identically in each round. The state distributed between the parties is also the same for each round of the protocol. Alice has m measurement inputs $x \in \{1, \cdots, m\}$ and each of the inputs has d corresponding outputs $a \in \{1, \cdots, d\}$. Bob instead has $m+1$ measurement

inputs $y \in \{1, \cdots, m+1\}$. Each measurement input of Bob also has d outputs $b \in \{1, \cdots, d\}$. See Fig. 6.1 for a pictorial representation. The protocol consists of the following steps:

1. In every round of the protocol, the parties do the following:

   - A state ρ_{AB} is distributed between Alice and Bob.
   - In each round of the protocol, the parties, according to a preshared key T, choose a random $T_i = \{0, 1\}$ such that $\Pr(T_i = 1) = \xi$. Depending on T_i, they then perform one of two different types of measurement:
     (a) If $T_i = 0$, Alice and Bob choose the measurement input ($x = 1, y = m + 1$) and use the outcomes to generate the raw key. We call these *key generation (KG) rounds*.
     (b) If $T_i = 1$, Alice and Bob choose the measurement inputs $x \in \{1, \cdots, m\}$ and $y \in \{1, \cdots, m\}$, respectively, uniformly at random. We use the outcome of these measurements for parameter estimation. These cases will be denoted as *parameter estimation (PE) rounds*.
   - The parties record their inputs and outputs as (x_i, y_i) and (a_i, b_i), respectively. After N rounds of measurement, Alice and Bob hold the input bit strings X^N and Y^N and the output bit strings A^N and B^N, respectively.

2. Alice and Bob publicly announce the outcomes of the PE rounds. They randomly divide[1] these outcomes into three sets. From the first set, Alice and Bob estimate the frequencies $\hat{\mathbf{P}}_1 = \{\hat{P}(ab|xy)\}$ (see Eq. (3.14)). The protocol aborts if $\hat{\mathbf{P}}_1 \in \mathcal{P}$. Otherwise, they construct an optimal Bell inequality B by solving the optimization problem of Eq. (3.15). The Bell inequality B is specified by the coefficients $\{h(ab|xy)\}_{x,y=1,\cdots,m}^{a,b=1,\cdots,d}$.

   From the measurement outcomes of the second set, the parties measure the Bell value $B[\hat{\mathbf{P}}_2]$. The deviation of the estimated Bell value $B[\hat{\mathbf{P}}_2]$ from the real Bell value $B[\mathbf{P}]$ can be bounded using Hoeffding's

[1] Alice is assumed to hold a random number generator. According to this random number generator, she specifies to which set each PE round's measurement outcomes belong.

Inequality (see Eq. (4.15)):

$$\Pr\left(B[\mathbf{P}] \geq B[\hat{\mathbf{P}}_2] - \delta_{est}\right) \geq 1 - \epsilon_{est}\,, \tag{6.1}$$

where $\epsilon_{est} = \exp\left\{\left(-\frac{2N\xi\delta_{est}^2}{3\gamma^2}\right)\right\}$. $\frac{N\xi}{3}$ is the number of measurement rounds used to estimate the Bell value $B[\hat{\mathbf{P}}_2]$.

Alice and Bob calculate the Bell value $B[\hat{\mathbf{P}}_3]$ from the outcomes of the third set. They use the Bell inequality B and corresponding violation $B[\hat{\mathbf{P}}_2] - \delta_{est}$ as a hypothesis in the experiment. The protocol aborts if the Bell value $B[\hat{\mathbf{P}}_3]$ is smaller than $B[\hat{\mathbf{P}}_2] - \delta_{est}$ for an honest implementation.

3. To estimate the QBER Q of the key generation rounds, Alice and Bob publicly announce the measurement outcomes of $N\eta$ randomly sampled key generation rounds. Using the tail inequality of the Theorem 4.5, the QBER of the raw key can be bounded as:

$$\Pr\left[Q \geq \hat{Q} + \gamma_{est}\left(N(1-\xi-\eta), N\eta, \hat{Q}, \epsilon_{est}^{\gamma}\right)\right] > \epsilon_{est}^{\gamma}\,, \tag{6.2}$$

where $\hat{Q}$ is the estimated QBER calculated from the sample Alice sends to Bob and $\gamma_{est}\left(N(1-\xi-\eta), N\eta, \hat{Q}, \epsilon_{est}^{\gamma}\right)$ is the positive root of the following equation:

$$\begin{aligned} &\ln\binom{N(1-\xi-\eta)\hat{Q} + N(1-\xi-\eta)\gamma_{est}}{N(1-\xi-\eta)} + \ln\binom{N\eta\hat{Q}}{N\eta} \\ &= \ln\binom{(N(1-\xi)\hat{Q} + N(1-\xi-\eta)\gamma_{est}}{N(1-\xi)} + \ln\epsilon_{est}^{\gamma}\,. \end{aligned} \tag{6.3}$$

From the Eq. (6.2), one can conclude that QBER Q of the key generation rounds is smaller than estimated QBER $\hat{Q}$ plus a statistical correction γ_{est} with very high probability of $1 - \epsilon_{est}^{\gamma}$.

4. Alice and Bob employ a one-way error correction (EC) protocol to obtain identical raw keys K_A and K_B from their bit strings A^N and B^N. During the whole EC process, Alice communicates O_{EC}[2] to Bob

[2] O_{EC} denotes all the classical communication. It consists of the information leaked during EC, the hash function, and the hash values that Alice sent to Bob to verify if EC is successful.

such that he can guess the outcomes A^N of Alice. The probability with which the EC protocol aborts for an honest implementation is at most ϵ_{EC}^c. If the EC does not abort, the parties obtain the error-corrected raw keys K_A (for Alice) and K_B (for Bob). Given that the protocol does not abort, the probability of Alice and Bob holding different raw keys is at most ϵ_{EC}, i.e. $\Pr(K_A \neq K_B) \leq \epsilon_{EC}$. See also Sec. 4.3.3 and Appendix of [DKB22a] for details.

The hash values of keys belonging to Alice and Bob are different with a high probability if the QBER Q of the key generation round is greater than the estimated QBER $\hat{Q} + \gamma_{est}$ [Gra21]. Since this event, which happens with probability at most ϵ_{est}^{γ}, leads to the abortion of the implement EC protocol, we can conclude that $\epsilon_{EC}^{C} \leq \epsilon_{est}^{\gamma}$, with ϵ_{EC}^{C} being the abortion probability of the EC protocol.

5. Alice and Bob apply a privacy amplification protocol and obtain a secure final key $\tilde{K}_A = \tilde{K}_B$ of length ℓ.

6.2 Device-independent Secret Key Rate

As already discussed, to provide a lower bound on the DISKR, one has to estimate two terms: the conditional von Neumann entropy $H(A|X,E)$ and the error correction information $H(A|B)$ of the raw key. As already mentioned in the previous chapter, $H(A|X,E)$ can be bounded using the guessing probability $P_{\text{guess}}(A|X,E)$ utilizing the min-entropy. The guessing probability $P_{\text{guess}}(A|X,E)$ can be upper bounded by a function G_x of the estimated Bell violation $B[\mathbf{P}]$ [MPA11]

$$P_{\text{guess}}(A|X,E) \leq G_x\left(B\left[\mathbf{P}\right]\right) . \tag{6.4}$$

We recall that the protocol aborts if $B[\hat{\mathbf{P}}_3] < B[\hat{\mathbf{P}}_2] - \delta_{est}$. Since $B[\hat{\mathbf{P}}_3]$ is calculated from a finite number of rounds, we need to use Hoeffding's inequality to infer the real Bell violation. We define a confidence interval δ_{con}, and the associated error probability ϵ_{con}. We bound the probability of wrongly accepting the hypothesis with the error probability ϵ_{con} by:

$$\begin{aligned} &\Pr\left(B\left[\hat{\mathbf{P}}_2\right] - \delta_{est} \geq B\left[\hat{\mathbf{P}}_3\right] + \delta_{con}\right) < \epsilon_{con} \\ \Rightarrow &\Pr\left(B\left[\hat{\mathbf{P}}_2\right] - \delta_{est} - \delta_{con} \geq B\left[\hat{\mathbf{P}}_3\right]\right) < \epsilon_{con} . \end{aligned} \tag{6.5}$$

Therefore, given that Alice and Bob do not abort the protocol, we infer that the Bell violation of the system under consideration is higher than $B[\hat{\mathbf{P}}_2] - \delta_{est} - \delta_{con}$ (with maximum ϵ_{con} error probability). Considering the worst possible scenario, we use the Bell violation $B[\hat{\mathbf{P}}_2] - \delta_{est} - \delta_{con}$ to upper bound the guessing probability $P_{\text{guess}}(A|X,E)$ by solving the semi-definite programme:

$$\begin{aligned} &\max_{\rho_{AB},\{A(a|x)\},\{B(b|y)\}} \quad P_{\text{guess}}(A|X,E) \\ &\text{subject to:} \quad \text{Tr}(\rho\mathcal{G}) = B\left[\hat{\mathbf{P}}_2\right] - \delta_{est} - \delta_{con}\,, \end{aligned} \tag{6.6}$$

using NPA hierarchy [NPA07, NPA08]. The Bell operator $\mathcal{G}$ is given by

$$\mathcal{G} = \sum_{a,b,x,y} h(ab|xy)A(a|x)B(b|y)\,,$$

where $\{h(ab|xy)\}$ are the coefficients of the Bell inequality B. In Eq. (6.6), $A(a|x)$ and $B(b|y)$ are the measurement operators of Alice and Bob, respectively, and ρ_{AB} is the state shared between them. Hence the conditional min-entropy $H_{\min}(A|X,E)$ can be bounded by [KRS09]

$$\begin{aligned} H_{\min}(A|X,E)_\rho &= -\log_2 P_{\text{guess}}\left(A|X,E\right) \\ &\geq -\log_2 G_x\left(B\left[\hat{\mathbf{P}}_2\right] - \delta_{est} - \delta_{con}\right)\,, \end{aligned} \tag{6.7}$$

where the function G is defined in Eq. (6.4).

To bound the error correction information, we need to estimate the key generation round's QBER Q. In Sec. 6.1, we show that we can upper bound key generation round's QBER with at least $1 - \epsilon_{est}^{\gamma}$ probability by $\hat{Q} + \gamma_{est}$. Considering the worst possible scenario, we can upper bound the von Neumann entropy $H(A|B)$ as:

$$H(A|B) \leq f\left(\hat{Q} + \gamma_{est}\right)\,, \tag{6.8}$$

where $f(x) = h(x) + x\log_2(d-1)$. Here, d is the number of outcomes per measurement in the Bell scenario [BMF+16] and h is the binary entropy function.

Using the bound on the min-entropy (see Eq. (6.7)) and the QBER (see Eq. (6.8)), we can derive the finite-size secret key rate of a ϵ_{DIQKD}^{s}-sound, ϵ_{DIQKD}^{c}-complete DIQKD protocol for collective attacks. In fact either the protocol in Sec. 6.1 aborts with probability higher than $1 - (\epsilon_{con} + \epsilon_{EC}^{c})$ or a

$(2\epsilon_{EC} + \epsilon_s + \epsilon_{PA})$)-correct-and-secret key of length

$$\begin{aligned} l \leqslant N \Big[& -\log_2 G_x \left(B\left[\hat{\mathbf{P}}_2\right] - \delta_{\text{est}} - \delta_{\text{con}} \right) \\ & -(1-\xi-\eta) f\left(\hat{Q} + \gamma_{\text{est}}\right) - (\xi+\eta)\log_2 d \Big] \\ & - \sqrt{N} \left(4 \log\left(2\sqrt{2^{\log_2 d}} + 1\right) \left(\sqrt{\log \frac{8}{\epsilon'^2_{\text{EC}}}} + \sqrt{\log \frac{2}{\epsilon_s^2}} \right) \right) \\ & - \log\left(\frac{8}{\epsilon'^2_{\text{EC}}} + \frac{2}{2-\epsilon'_{\text{EC}}} \right) - \log \frac{1}{\epsilon_{\text{EC}}} - 2\log \frac{1}{2\epsilon_{\text{PA}}}. \end{aligned} \tag{6.9}$$

can be generated. Here $\epsilon^c_{DIQKD} \leq \epsilon_{est} + \epsilon^{\gamma}_{est}$ (for an honest implementation) and $\epsilon^s_{DIQKD} \leq 2\epsilon_{EC} + \epsilon_s + \epsilon_{PA}$. The expression in Eq. (6.9) is derived in [DKB22a] where we explain the complete secret key analysis in detail; also see Appendix A. Table 6.1 lists all parameters of the DIQKD protocol.

6.3 Applications

The potential and versatility of this approach have been illustrated with several examples for different numbers of measurement settings and different numbers of outcomes in [DKB22a], also see Sec. A. In the case of the $[2, 2]$ Bell scenario, this method recovers the standard CHSH inequality when using a maximally entangled Bell state mixed with white noise and the CHSH measurement settings; see Eq. (3.9) for details. Thus, the secret key rate generated coincides with the one of [MPA11] in which a predetermined standard CHSH inequality is used. Though this method finds a hyperplane equivalent to the CHSH inequality, it identifies the particular facet Bell inequality with the maximal violation, which is then used in the DIQKD protocol. Other facets (also equivalent to CHSH inequality) may admit local hidden variable models leading to zero DISKR. For the $[2, d]$ Bell scenario (2 measurement settings per party, d outcomes each), this method recovers *CGLMP inequality* [CGL+02] when using the maximally entangled state of two qudits affected by white noise with probability p, i.e. the state

$$\rho = (1-p)|\psi\rangle\langle\psi| + p\frac{\mathbb{1}}{d^2}, \tag{6.10}$$

where $|\psi\rangle = \sum_{i=0}^{d-1} \frac{1}{\sqrt{d}}|ii\rangle$, and the measurement settings in Eq. (24) of [DKB22a] (also see Appendix A). Similar to the previous case, this method

Table 6.1: Parameters of the DIQKD protocol

N	Number of measurement rounds in the protocol
ξ	Fraction of parameter estimation rounds performd out of the total number of rounds
η	Fraction of measurement rounds used to estimate the QBER
ϵ_s	Smoothing parameter
$\epsilon_{EC}, \epsilon'_{EC}$	Error probabilities of the error correction protocol
ϵ^c_{EC}	Probability of abortion of error correction protocol
δ_{est}	Width of the statistical interval for the Bell violation hypothesis test
ϵ_{est}	Error probability of the Bell violation hypothesis test
δ_{con}	Confidence interval for the Bell test
ϵ_{con}	Error probability of the Bell violation estimation
γ_{est}	Width of the statistical interval for the QBER estimation
ϵ^{γ}_{est}	Error probability of the QBER estimation
ϵ_{PA}	Error probability of the privacy amplification protocol
ϵ^c_{DIQKD}	Completeness parameter of the DIQKD protocol
ϵ^s_{DIQKD}	Soundness parameter of the DIQKD protocol

finds the facet with the maximal violation; again, other equivalent facets may be subjected to local hidden variable models leading to a zero DISKR.

This procedure also has advantages w.r.t. the CHSH scenario (corresponds to the $[2, 2]$ Bell scenario) when the parties use non-optimal measurement settings. In the case of $[2, 2]$ Bell scenario, DISKR is calculated via the analytical expression of Eq. (5.5) that uses the lower bound on the von Neumann entropy and thus generates higher DISKR than bounds based on min-entropy. However, suppose non-optimal measurement settings are used. In that case, this method can generate higher DISKR by employing additional measurement settings per party (making it a $[3, 2]$ Bell scenario) than using any subset of two measurement settings per party (and using the analytical expression of Eq. (5.5)). Another edge of this method is even if the measurement statistics obtained by two non-optimal measurement settings per party lead to a zero DISKR, adding another measurement setting per party and adopting this strategy can lead to a non-zero DISKR. Examples of these cases are shown in [DKB22a].

Moreover, when applying this approach to random measurements scenario where Alice's and Bob's device performs random measurements, the probability of generating non-zero DISKR increases with the number of measurement settings per party and decreases with the introduction of noise in the shared state. This phenomenon is in coherence with nonlocal volume [LCMA18, DRGP+17, dRGP+20, FP15, FDRV+18, BN18], which increases when more measurement settings for each party are used for the pure bipartite entangled state and shrinks with the introduction of noise in the shared state.

As mentioned before, in [NSPS14, BSS14], the authors introduced another approach to bound the DISKR by directly using the complete measurement data. In the asymptotic regime, this procedure corresponds to constructing a Bell inequality that leads to the maximal DISKR for the precise setup. However, small changes in the parameters (e.g. imperfections in the measurement directions) or the measured probability distribution lead to different Bell inequalities corresponding to the optimal secret key rate. On the other hand, the Bell inequality derived from the approach of [DKB22a] is robust and stable against small fluctuations of the measurement directions or in the shared state. It can also generate a non-zero secret key by performing fewer measurement rounds because of the smaller effect of statistical corrections in the Bell inequality violation. The statistical corrections become insignificant for a high number of measurement rounds, such that the method of [NSPS14, BSS14] yields a higher secret key.

The typical number of measurement rounds to generate a non-zero key varies between 10^6 to 10^8 for the $[m, 2]$ Bell scenario and is of the order 10^6 for the $[2, d]$ Bell scenario. Note that the protocol in Sec. 6.1 uses min-entropy to bound the von Neumann entropy. However, sophisticated methods of bounding von Neumann entropy [BFF21a, BFF21b] could increase the secret key rate. Other methods like advantage distillation [TLR20], noisy pre-processing [HST+20], and random post-selection [XZZP21] can also be applied in this framework to improve the DISKR generation.

7

Upper bound on the Guessing Probability using Machine Learning

Recent advances in machine learning (ML) have dramatically changed science and society. Making computers act without being explicitly instructed is the aim of machine learning. The self-driving car, fraud detection, speech recognition, drug discovery, and predicting the 3-D structure of proteins based on their genetic sequence are some areas where ML is being used [Lav15, AlQ19]. Recently, ML has been heavily utilized in different domains of physics, including quantum physics, such as phase transition [YAK18, BCMT17, Wan16, CCMK17, Wet17, HSS17, VNLH17, CM17, DLS17], black hole detection [AAA+16, Pas16], topological codes [TM17], glassy dynamics [SCS+16], gravitational lenses [HLM17], anti-de Sitter/conformal field theory (AdS/CFT) [HSTT18], string theory [CHKN17], Monte Carlo simulation [LQMF17, HW17], tensor network [CCX+18, HM+21], many-body physics [SRN17, CT17] and wave analysis [BBC+13]. Inspired by this progress, techniques from machine learning have been used to solve analytically or numerically complex problems in quantum information. In [CBC19], an ensemble of multilayer perceptrons and genetic algorithms are combined to detect and characterize nonlocality. In [KCC+20], the authors use neural networks to solve the causal inference (whether an observed probability distribution can be reproduced using only classical resources) problem. They encode the causal structure into the architecture of a neural network to see whether the target distribution is 'learned'. The neural network serves as an oracle, showing that if a behavior can be 'learned', it is classical. In [Den18], the authors use the restricted Boltzmann machine (RBM) architecture to

find quantum nonlocality in many-body systems. In [BHVK19], the authors used reinforcement learning to train AI to play Bell nonlocal games and obtain maximal quantum Bell violation for various Bell inequalities. Since Bell inequalities cannot be a reliable tool for entanglement detection, Bell inequalities and feed-forward neural networks are blended to use them as a dependable state classifier that can segregate an entangled state from a separable one in [MY18]. Reinforcement learning with RBM is incorporated to detect the entangled states [HPFP20]. Given the full tomographic data, random states of two qutrits are classified as separable, entangled with positive partial transpose, or entangled with negative partial transpose using an automated machine learning model [GCDM21]. In [LHL+18], the authors use various classical machine learning techniques and ensemble training via bootstrap aggregating (bagging) to detect entanglement in quantum states. Inspired by these successes in implementing machine learning in quantum information, we introduce deep learning methods to estimate the upper bound of the guessing probability.

Measurement statistics of a Bell experiment are called 'nonlocal' whenever the obtained correlations defy the underlying premises of local realism [Bel64]. Device-independent quantum key distribution (DIQKD) [ABG+07, PAB+09, AFRV19, AFDF+18, BHK05, MPA11, VV14, MRC+14, AMP06, MvDR+19, HKB20, HR10] and device-independent randomness generation (DIRNG) [PAM+10, NSBSP18, PM13, BSS14, NSPS14, BKB17, AMP12, AM16, SC18] leverage these nonlocal correlations to certify the private randomness generated from the quantum states. A key component of DIQKD and DIRNG is the estimation of global and local randomness. The assessment of the guessing probability is often a crucial problem in quantifying randomness. Additionally, it can serve as an indicator of nonlocality. The probability that an adversary will correctly predict the outcome of a party's measurement is known as the guessing probability. The adversary cannot predict the outcome with certainty if the guessing probability is less than 1. It suggests that the system contains intrinsic randomness. Bounding the guessing probability is a difficult task. The guessing probability cannot be calculated explicitly. One can upper bound the guessing probability by solving a semidefinite optimization problem. Typically, the Bell inequality and associated quantum Bell violation is used as a constraint in the optimization of guessing probability [MPA11, PAM+10]. One needs to make use of the hierarchical structure of the quantum correlation set (NPA hierarchy) [NPA07, NPA08] to resolve the semidefinite optimization problem. The complexity of this optimization problem is increasing as the Bell scenario

becomes more complex (more measurements and outcomes), making it exceedingly difficult computationally. To circumvent this problem, inspired by the recent progress in utilizing ML in quantum information, we develop deep learning models that can predict the guessing probability and the optimal Bell inequality (used to calculate the guessing probability) from the quantum probability distribution. We use the supervised machine learning method to develop deep learning models. First, we randomly sample probability distributions from the quantum correlation set and use it as the input of the training data for supervised learning. With this data, we calculate the upper bound of the guessing probability using the two-step method of [DKB22a]; see also Sec. 3.4 and Sec. 5.4 for details. We use the optimal Bell inequality and the guessing probability as the output of the training data. After adequate training, the model can accurately identify the pattern and estimate the guessing probability and the corresponding optimal Bell inequality with high accuracy and low average statistical error.

This chapter's content, an important result of this thesis, is based on [DKB22b]. This chapter is organized as follows. We briefly introduce machine learning and feed-forward neural network in Sec. 7.1. We explain the process of sampling quantum probability distribution from the quantum correlation space in Sec. 7.2. This sampled quantum probability distribution is then used as input data in the deep learning models. We illustrate the deep learning models we used to assess the data in Sec. 7.3 and discuss the utility and performance in Sec. 7.4.

7.1 Introduction to Machine Learning

Machine learning (ML) is a sub-field of artificial intelligence (AI) [SSBD14, GBC16]. The goal of machine learning is to enable a computer to complete a certain task without direct guidance from an outside source. In the words of Mitchell: *"A computer program is said to learn from experience E for some class of tasks T and performance measure P, if the performance at tasks in T, as measured by P, improves with E"* [M+97]. Thus learning of the 'machine' happens whenever $P(T) \propto E$. In a nutshell, Machine learning is a technique used to build intricate models to generate predictions for issues that are challenging for fixed algorithms to handle. The machine learning model is frequently viewed as a "black box", difficult or even impossible to comprehend, and humans are content to accept the trained machine's response as correct.

Here, we provide a brief but insightful description of the relevant steps

of machine learning. These include the Tasks, the Experiences, the machine learning algorithms, and also the performance measures for validation. The structure of this discussion is inspired by [GBC16].

7.1.1 The Task, T

Machine learning enables us to tackle tasks that are too difficult to solve with fixed programs written and designed by human beings. Machine learning tasks are usually formulated regarding the machine's ability to learn and process an example. An example is a collection of features that have been quantitatively measured from some object or event that we want the machine learning system to process. Let's say each instance of data consists of a set of features. We represent the example as a vector $\mathbf{x} \in \mathbb{R}^n$, and each entry of the vector x_i is one feature. For example, the features of an image are usually the pixel values in that particular image.

Some examples of machine learning tasks are classification, regression, denoising, anomaly detection, density estimation, imputation of missing values, machine translation, etc. Here, we only focus on the two most common tasks: classification and regression. In a classification task, the program trains itself to learn a function $f : \mathbb{R}^n \to \{1, \cdots, k\}$, where k is a finite and (typically) pre-established integer number. The learned program determines which of the K categories the given input belongs to. In general, the model returns a normalized probability distribution over the k classes, and the suggested class is the one with the highest probability. In a regression task, the computer program is trained to predict a numerical value for a given output. Therefore the algorithm aims to model an appropriate function $f : \mathbb{R}^n \to \mathbb{R}$.

7.1.2 The Experience, E

Learning algorithms are broadly divided into three classes, supervised, unsupervised learning, and reinforcement learning.

Algorithms for supervised machine learning are designed to learn from the labelled examples. The training data for supervised learning algorithms consists of inputs (X) and the desired results (y). The algorithm will look for patterns in the data during training to predict the expected outputs, such as $y = f(X)$. Because an algorithm learning from the training dataset can be compared to a teacher supervising the learning process, it is known as supervised learning. The algorithm iteratively produces predictions on the training data and is corrected by the teacher because we know the right

answers. When the algorithm performs to an acceptable standard, learning ceases. The program's primary objective is to estimate the $p(y|X)$ input-output conditional probability distribution. Some examples of supervised learning algorithms are Regression, Logistic Regression, Naive Bayes Classifiers, KNN (K nearest neighbors) [Alt92, Fix51, Dud78], Decision Trees [Qui86, BFOS84], Random forest [Bre01, Ho95], XGBoost [CG16] etc. Since supervised learning helps to optimize performance with the help of experience, it can be applied to solve various types of real-world computation problems. However, one disadvantage of this learning method is the collection of labelled examples since collecting big data is quite challenging.

Algorithms for unsupervised learning deal with unlabelled data. The model works independently to discover undetected patterns and information. Unsupervised machine learning techniques aim to learn the probability distribution $P(x)$ of the examples x. K-Means Clustering, Principal Component Analysis, and Hierarchical Clustering are some examples of unsupervised learning algorithms.

Reinforcement learning is about taking a sequence of apt actions to achieve the maximal reward in a particular situation. Neither data nor label is available in this scenario. The artificial intelligence agent faces a game-like circumstance; employs trial and error to come up with the final solution. During the task, the AI agent gets either rewards or penalties for every action it performs. The aim is to maximize the rewards throughout the entire course. Note that only the reward policy is set; no hints are given to the AI agent to solve the task. The model must determine how to complete the objective, starting with entirely arbitrary trials and ending with sophisticated strategies. Some autonomous driving functions that include reinforcement learning are trajectory optimization, motion planning, dynamic pathing, controller optimization, and scenario-based learning policies for highways. Another example of reinforcement learning was demonstrated by Deepmind's AlphaZero, which has defeated the best human player in the board game Go [SSS+17].

7.1.3 The Performance, P

The learners' ability to perform well for new inputs is a crucial component of machine learning. We split the dataset into two sets to accomplish this. To train the machine, we utilize samples from one set called the training set. We evaluate the performance using the data from the second batch, which typically accounts for 20%–25% of the total data. This set is called the test

set.

One frequently assesses a learner's performance in terms of a loss function. An event or the values of one or more variables are mapped onto a real number that intuitively represents some "cost" related to the event via a loss function. Because of this, we also refer to these functions as cost functions and use both terms interchangeably. A loss function is what a machine learning model aims to reduce during the training process.

For classification tasks, we measure *accuracy* as our performance measure. The percentage of cases for which the model generates the correct output is defined as accuracy. For regression tasks, one often uses Mean Squared Error (MSE) or Mean Absolute Error (MAE), which corresponds to the mean L1 and L2 norms, respectively, as loss functions. MSE is defined as:

$$MSE(y, \hat{y}) = \frac{1}{N} \sum_{i=0}^{N} \left(y_i - \hat{y}_i\right)^2 , \tag{7.1}$$

and MAE reads:

$$MAE(y, \hat{y}) = \frac{1}{N} \sum_{i=0}^{N} \left|y_i - \hat{y}_i\right| , \tag{7.2}$$

where $\hat{y}$ is the estimated output of the model.

Both MSE and MAE have benefits and drawbacks. The MSE is excellent for guaranteeing that our trained model does not contain any outlier predictions with significant mistakes because the squaring component of the MSE gives these errors more weight. However, the squaring portion of the function amplifies the mistake if our model makes a single abysmal forecast. As a result, the MSE cost function is less resistant to outliers due to this characteristic. Thus, by lowering the MSE loss function, we can typically accept a number of minor errors throughout the learning process but no huge errors. Since we are considering the absolute value in the case of MAE, all errors will have the exact linear weighting. As a result, unlike the MSE, we will not give our outliers much consideration, and our cost function offers a broad understanding and assessment of how well our model is functioning. However, occasionally, the outliers' enormous errors have the same weight as smaller errors. It could lead to our model performing most of the time admirably while sometimes producing a few highly inaccurate forecasts.

Thus, ideal loss functions are varied from project to project. Since we will perform a regression task, we train the model with the MSE loss function, which is implemented via the Python Scikit-learn package [PVG+11].

However, we use both MSE and MAE as performance measures over the test set.

7.1.4 Overfitting and Underfitting

An ML model generally has access to the training data. While training, we calculate some error measures. It is called the *training error*. If a model generalizes any new input data from the problem domain appropriately, it is said to be a good machine learning model. The ability to perform well on previously unseen data of an ML model is called *generalization*. During the training process, our goal is to reduce the training error corresponding to an optimization problem. However, machine learning is different from an optimization process. It has to forecast future data that the data model has never encountered. In an ML model, we want a lower *generalization error*, also called *test error*, simultaneously. Before diving further, let us understand two important terms: bias and variance. Building accurate models and avoiding overfitting and underfitting errors would be easier with a thorough grasp of these flaws.

- *Bias*: Bias is the difference between the model's prediction and the actual value we are attempting to predict. High bias models oversimplify the model and pay very little attention to the training data.
- *Variance*: Variance specifies the variation in the prediction if different training data is used. Simply said, variance indicates how much a random variable deviates from its predicted value. A model with a significant variance pays close attention to the training data and does not generalize to new data. As a result, these models have significant error rates on test data while performing exceptionally well on training data.

Underfitting: A machine learning model is underfitting when it cannot recognize the underlying pattern in the data. It refers to a model that neither models the training data nor generalizes to new data. Its occurrence indicates that our model or method does not adequately suit the data. It typically occurs when we try to develop a linear model with fewer non-linear data or when we have insufficient non-linear data to build an accurate model. The machine learning model will likely produce a lot of incorrect predictions in these circumstances since the rules are too simple and flexible such that one can apply them to sparse data.

Some reasons for underfitting are high bias and low variance, the small size of the training dataset, the simple model, and uncleaned and noisy training data. To reduce underfitting, one should take the following steps, e.g. increase model complexity, increase the number of features, perform feature engineering, remove noise from the data, and increase training duration.
Overfitting: When a statistical model fails to produce reliable predictions on test data but fits the training data well, it is said to be overfitted. When a model overfits, it learns the information and noise in the training data to the point where it adversely affects the model's performance on new data. It indicates that the machine learns concepts from the noise or random fluctuations in the training data, which don't apply to new data. Thus, it poses a difficulty for the models' capacity to generalize.
The non-parametric and non-linear approaches are the root causes of overfitting since these types of machine learning algorithms have more latitude in how they develop the model based on the dataset, making it possible for them to produce highly irrational models. If we have linear data, employing a linear algorithm is one way to prevent overfitting; if we use decision trees, utilizing parameters like the maximal depth is another. Overfitting is a problem when the evaluation of machine learning algorithms on training data differs from the evaluation of unknown data. Some common reasons for overfitting are building complex models and huge size differences in training and test data. One can employ numerous techniques to reduce overfitting, such as increasing training data, reducing model complexity, early stopping, and using dropouts in neural networks.
A model is said to have a good fit for the data when it provides predictions with no error, which is the ideal situation. A sweet spot between overfitting and underfitting allows for this condition. We must examine our model's performance over time as it gains knowledge from the training dataset to comprehend it. Our model will continue to learn as time goes on, so the training error and testing error data will continue to drop. The presence of noise and less valuable features will make the model more prone to overfitting if it is allowed to learn for an excessively long time. As a result, the test error will start to increase, and our model's performance will decline. We will halt just before the errors increase to get a good match. The model is proficient at this point in both our unseen testing dataset and training datasets.

Now we will discuss feed forward neural networks since we are going to use them as our models later.

7.1.5 Feed Forward Neural Network

Feed forward neural networks (FFNN) are the backbone of deep learning which is a subfield of machine learning; also see Fig. 7.1. Deep learning deals with algorithms inspired by the structure and function of the brain called artificial neural networks. The basic unit of an artificial neural network is a

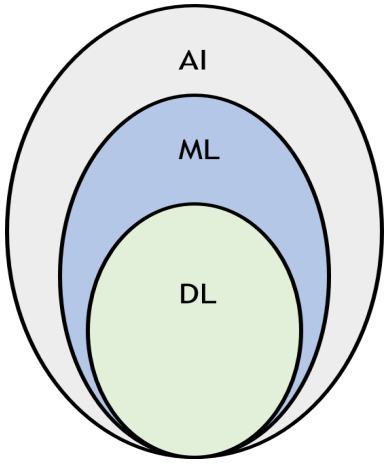

Figure 7.1: Deep learning (DL) is a sub-discipline of machine learning (ML) which is a sub-field of artificial intelligence (AI).

single artificial neuron. It is a real-valued function of the form

$$AN(\mathbf{x}) = \phi\left(\sum_i w_i x_i + b\right), \tag{7.3}$$

where w_i and b is the weights and bias of corresponding input x_i, and ϕ is a real-valued function $\phi : \mathbb{R} \to \mathbb{R}$. For visualization, see Fig. 7.2. The function ϕ is usually known as the activation function. There are many possible activation functions that exist in the literature. Here we list some important ones which are often used in practice.

- Linear: $\phi(x) = x$.
- Sigmoid: $\phi(x) = \frac{1}{1+e^{-x}}$.
- Hyperbolic tangent: $\phi(x) = \tanh(x)$.
- Rectified Linear Unit (ReLu): $\phi(x) = \max(0, x)$.

The activation function decides whether a neuron should be activated by calculating the weighted sum and adding bias to it. The non-linear trans-

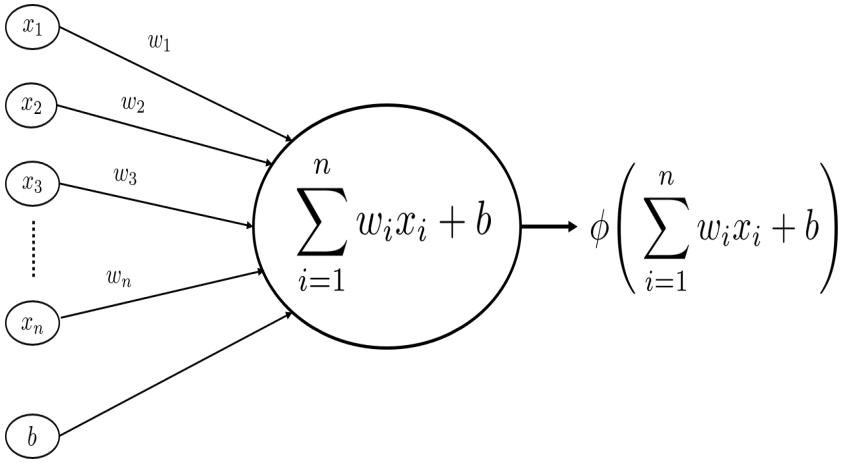

Figure 7.2: Basic unit of a feed forward neural network. Here, x_i denotes the inputs, w_i is the corresponding weight, b is the bias. ϕ is the activation function that introduces the nonlinearity in the output. The output of the node reads: $\phi\big(\sum_i w_i x_i + b\big)$.

formation of the input introduced by the activation function enables neural networks to learn and carry out complex tasks [Cyb89, HSW89].

Artificial neurons can be combined to form a complex structure that can perform complicated tasks. It can be achieved by connecting the output of a neuron as an input to another. One can think of this as a graph $G = (V, E)$ where the nodes V correspond to the artificial neurons and edges E correspond to the connections between two neurons. This whole graph G structure is called an artificial network. A typical artificial network consists of an input, hidden, and output layer. The input layer is the first layer where the data is fed into the network. The middle layers are called the hidden layers, which is the defining feature of deep learning. The last layer is called the output layer, which produces the result. Neural networks with more than one hidden layer are called deep neural networks, and machine learning of deep neural networks is called deep learning. Neural networks with no loops are called feed forward neural networks (FFNN); see Fig. 7.3 for visualization. The output information of the neurons is always fed in the forward direction; never fed back. It corresponds to a directed acyclic graph.

An FFNN's aim is approximating a function $f(\mathbf{X})$ by $f^*(\mathbf{X}, \vec{\theta})$ which maps an input $\mathbf{X}$ to an output y returning the best values of the parameters $\vec{\theta}$ (weights and biases) after the learning process. We have to fix a loss function for the training process of the network. We use MSE as our loss function to train the model. Evaluated on our whole training set, our MSE loss function

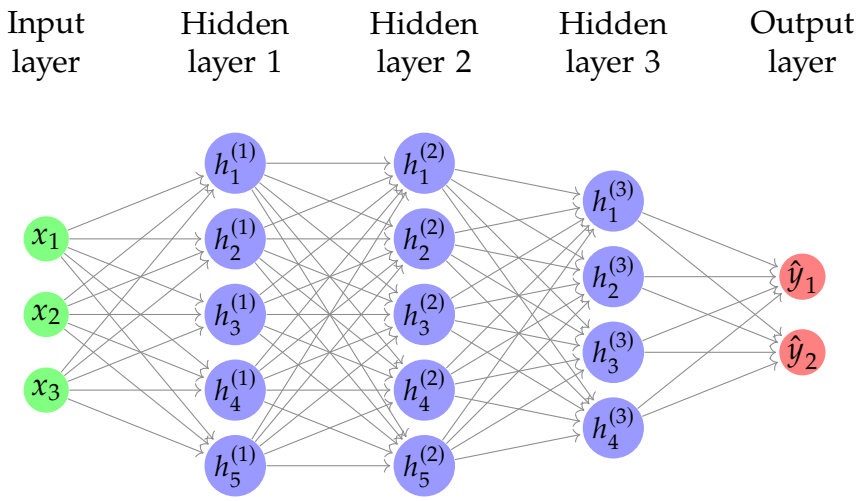

Figure 7.3: Schematic representation of a Feed Forward Neural Network (FFNN).

for training is defined as

$$\mathcal{L}^{MSE}(\vec{\theta}) = \frac{1}{m} \sum_{x \in X_{train}} (\mathbf{y}_{train} - f^*(\mathbf{X}, \vec{\theta}))^2 \,, \tag{7.4}$$

where m is the size of the training set. In principle, one has to minimize the loss, which is done using gradient descent. This is, by definition, comprised of two steps: calculating gradients of the loss function and then updating existing parameters in response to the gradients. This cycle is repeated until reaching the minima of the loss function. Calculating the gradient $\nabla_{\vec{\theta}} \mathcal{L}^{MSE}(\vec{\theta})$ is a tedious task with respect to the weight and bias parameters. It can be achieved by back-propagation [GBC16, BW91, VMR+88, HN92, LTHS88, CR95, F+88]. Following recent deep learning approaches, one can use state of the art optimizers that minimize the loss function. For our purpose, we use the ADAM optimizer [KB14], which uses the back-propagation method to calculate the gradient, already contained in several deep learning packages [Cho15, AAB+15].

7.2 Data Generation

A supervised machine learning technique is incorporated to obtain an upper bound of the guessing probability. Any supervised machine learning approach's first step is generating the training points. For this purpose, we generate random input-output measurement statistics of a $[m, k]$ (i.e. m

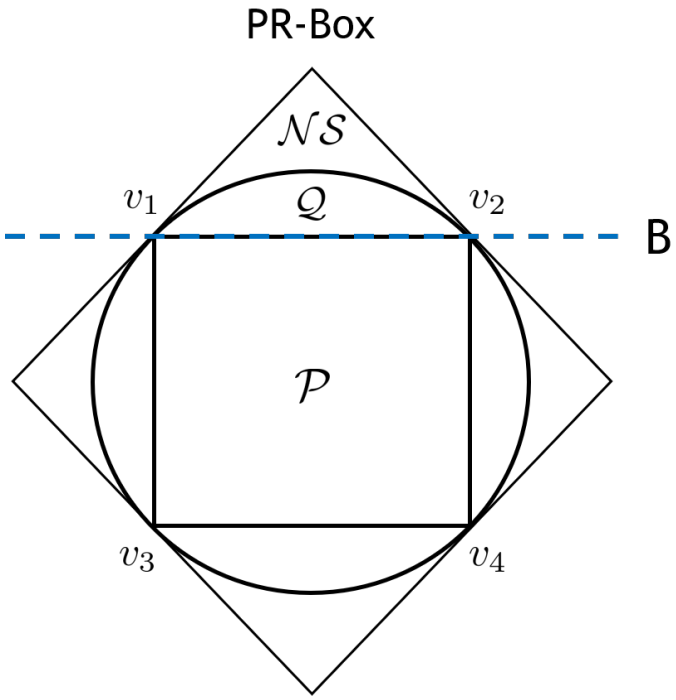

Figure 7.4: A sketch for the set of correlations. All classical probabilities form a convex polytope $\mathcal{P}$, which is embedded in the set $\mathcal{Q}$ of quantum correlations, which in turn is a subset of the no-signalling polytope $\mathcal{NS}$. v_1, v_2, v_3 and v_4 are the vertices of the local polytope. B (blue dashed line) represents the facet Bell inequality which separates the classical polytope from the quantum and no-signalling set.

measurement settings, k outcomes each) Bell scenario. The input-output measurement statistics or probability distribution is also called a 'behavior' [BCP+14], and we use these terms interchangeably. Since the guessing probability for the local behaviors is always 1 (i.e. Eve can guess the right outcome with probability 1), there is no need to train the machine to perform well on local behaviors. Thus, we only generate random quantum bipartite probability distribution and use it as the input (features) of the supervised machine learning model. *Weighted vertex sampling* method [KCB+21] can be employed to generate the quantum probability distribution.

To generate samples from the quantum set $\mathcal{Q}$, recall the set of probabilities, satisfying classical and no-signalling conditions, form the convex polytopes $\mathcal{P}$ and $\mathcal{NS}$, respectively (see Fig. 7.4 for illustration). For the $[m, k]$ Bell scenario, the classical polytope $\mathcal{P}$ is specified by k^{2m} local vertices. However, the classical polytope can also be described by its facets. These facets

represent the hyperplane (or Bell inequality) that separates any non-classical (quantum and no-signalling) behavior from the classical ones. These facets are called facet Bell inequalities or tight Bell inequalities [BCP+14]. Eight facet Bell inequalities exist for the [2, 2] scenario, all equivalent to the CHSH inequality [CHSH69]. For the [3, 2] Bell scenario, 648 facet Bell inequalities are identified. It is done using the formulation of [Fuk03]. Using [Fuk03], one can calculate all the facets of a convex polytope given its vertices. The transformation from vertex representation to the facet representation of a polytope is known as facet enumeration or convex hull problem, which uses Gaussian and Fourier-Motzkin elimination [1]. Note that all the 648 facet Bell inequalities correspond to two classes of independent facet Bell inequalities, i.e. the CHSH inequality and the $I3322$ inequality [CG04, PV10]. For the [4, 2] Bell scenario, more than 10000 facet Bell inequalities [CC19] exist. However, all facets correspond to 174 independent facet Bell inequalities [CG19].

These facet Bell inequalities are spanned by some of the local vertices of the classical polytope [2]. These vertices provide the maximum classical bound of the corresponding facet Bell inequality. Consider n local vertices span a facet Bell inequality. We denote these set of n vertices as $\left\{P_i^{\mathcal{L}}(ab|xy)\right\}_{i=1}^{n}$ and the PR-box of the corresponding facet Bell inequality as $P^{\text{PR}}(ab|xy)$, see Fig. 7.4 for better visualization. The PR-Box $P^{\text{PR}}(ab|xy)$ can be defined as the probability distribution that provides the maximal no-signalling bound of the corresponding facet Bell inequality [BLM+05, PR94]. To generate a behavior $\mathbf{P}$ from the set $\mathcal{NS}\backslash\mathcal{P}$, uniform random weighted mixtures of the $n+1$ vertices (n vertices that span the facet Bell inequality and the corresponding PR-box) with n-fold weight on the PR-box are taken. Formally, the sample behavior $\mathbf{P} \in \mathcal{NS} \setminus \mathcal{P}$ can be generated as:

$$P(ab|xy) := \frac{nw_0 P^{\text{PR}}(ab|xy) + \sum_{i=1}^{n} w_i P_i^{\mathcal{L}}(ab|xy)}{nw_0 + \sum_{i=1}^{n} w_i} \tag{7.5}$$

where the $w_i \in [0, 1]$ are uniformly drawn random numbers. While generating probability distributions, we consider all facet Bell inequalities for

[1] The list of facets consists of positivity constraints and the facet Bell inequalities. Since physical theory never violates this condition, there is no particular interest in this constraint. Only the facet Bell inequalities are chosen, discarding the positivity constraints.

[2] Eight local vertices span all the facet inequalities for the [2, 2] Bell scenario. For the [3, 2] Bell scenario, facet Bell inequalities equivalent to the CHSH inequality is spanned by thirty-two vertices. Twenty vertices span inequalities equivalent to the $I3322$ inequality.

the [2, 2] and [3, 2] Bell scenario. However, since there are more than 10000 facet Bell inequalities for the [4, 2] Bell scenario, we only restrict ourselves to generating probability distributions using the independent facet Bell inequalities.

From this set of samples, we only select the behaviors with Q_2 realization (second level of NPA hierarchy), i.e. $P(ab|xy) \in Q_2$ [NPA07, NPA08], to generate random quantum probability distribution. Here, we presume that Q_2 provides a good approximation for the original quantum set Q.

Following that, the guessing probability of the sampled correlation $\mathbf{P}$ is estimated using the two-step method of [DKB22a] solving the optimization problem of Eq. (5.7); see Sec. 3.4 and Sec. 5.4 for a detailed description. Without loss of generality, one can calculate the guessing probability of Alice's first measurement setting. This two-step process will provide us with two essential parameters: the upper bound of the guessing probability $P_g^*(a|x, E)$ and the associated Bell inequality B (specified by the hyperplane vector $\{h(ab|xy)\}$) that is used to solve the optimization process.

The goal of the deep learning models is to predict the guessing probability $P_g^*(a|x, E)$ and the optimal Bell inequality B from the probability distribution $\mathbf{P}$ using supervised machine learning. For the $[m, k]$ Bell scenario, the dataset $\{\mathbf{X}, \mathbf{y}\}$ is prepared with the input (feature)

$$\mathbf{X} := \{P(ab|xy)\}_{x,y=1,\cdots,m}^{a,b=1,\cdots,k} . \tag{7.6}$$

and the output (target)

$$\mathbf{y} := \left[\{h(ab|xy)\}_{x,y=1,\cdots,m}^{a,b=1,\cdots,k}, P_g^*(a|x, E)\right] . \tag{7.7}$$

7.3 Deep Learning Models

In this section, we discuss the deep learning models that are used to assess the dataset $\{\mathbf{X}, \mathbf{y}\}$. Two types of neural network architectures are utilized here: 'linear' FFNN and 'nonlinear' FFNN. The 'linear' FFNN consists of several layers (without any branching); see Fig. 7.5 for better visualization. In this thesis, we refer to this neural network construction as NN_1.

The input layer has m^2k^2 neurons corresponding to the elements in $\{P(ab|xy)\}$. The output layer has $m^2k^2 + 1$ neurons (m^2k^2 neurons correspond to the coefficients of the optimal Bell inequality $\{h(ab|xy)\}$, and one

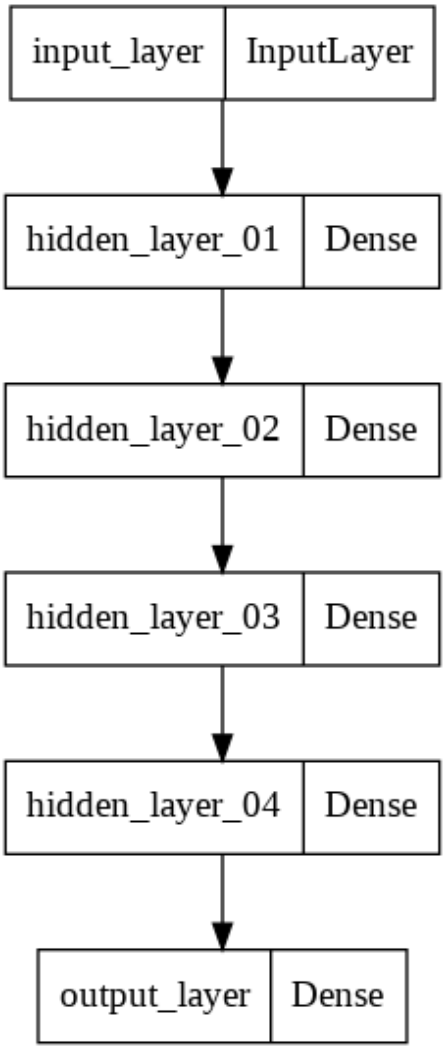

Figure 7.5: Schematic description of a linear neural network. It consists of an input layer, several hidden layers and an output layer without branching. Hidden layers and the output layer are dense layers, meaning the neurons of the layer are connected to every neuron of its preceding layer. In this construction, input layer: $\{P(ab|xy)\}$ (m^2k^2 neurons) and output layer: $\left[\{h(ab|xy)\}, P_g^*(a|x,E)\right]$ (m^2k^2+1 neurons).

corresponds to the guessing probability $P_g^*(a|x,E)$) [3]. While training the FFNN, the dataset $\{\mathbf{X},\mathbf{y}\}$ is divided into two parts following the standard approach. The first part of the dataset is for training and cross-validation (80%) [4], and the second part is for testing (20%). The authors perform 100 rounds of training using the optimizer ADAM [KB14], of which the first 50 rounds have a fixed learning rate of 0.001. For the rest of the rounds, the learning rate is reduced by 90% in every tenth round. The activation function ReLu (Rectified linear unit) is utilized in the input and the hidden layers. In the output layer, the linear activation function is used for m^2k^2 neurons that correspond to the optimal Bell inequality, and the sigmoid activation function is incorporated for the neuron that corresponds to the guessing probability. Mean Squared Error (MSE) (also see Eq. (7.1)) is used as the cost function, which is minimized during the training process.

The second type of neural network architecture that is used is the non-linear FFNN. In this network, two parallel submodels are incorporated to interpret parts of the output that share the same input. Here, the input layer has m^2k^2 neurons corresponding to the elements of the input-output probability distribution $\{P(ab|xy)\}$. Some hidden layers follow the input layer. After that, one hidden layer is bifurcated into two, creating two branches. The first branch of the network is for predicting the coefficients of the optimal Bell inequality $\{h(ab|xy)\}$, and thus has m^2k^2 neurons. The second branch of the neural network is for predicting the guessing probability. Thus, the output layer will have only one neuron corresponding to $P_g^*(a|x,E)$. In this thesis, we refer to this neural network construction as NN_2; see Fig. 7.6 for visualization. NN_2 can be built using Keras functional API [Cho15]. The activation function ReLu (Rectified linear unit) is used in the input and the hidden layers for both branches of the network. The linear activation function is used in the output layer of the first branch, while the sigmoid activation function is used in the second branch. The other details about the training of the network are the same as NN_1 neural network stated previously. Both NN_1 and NN_2 predicted an output of the form:

$$\mathbf{y}_{\text{pred}} := \left[\{h(ab|xy)^{\text{pred}}\}_{x,y=1,\cdots,m}^{a,b=1,\cdots,k}, P_g^{\text{pred}}(a|x,E)\right] ,$$

[3] If we only want to predict the upper bound of the guessing probability $P_g^*(a|x,E)$), there is just one neuron in the output layer.

[4] 80% of this set is used for training while the rest 20% is utilized for the validation of the model.

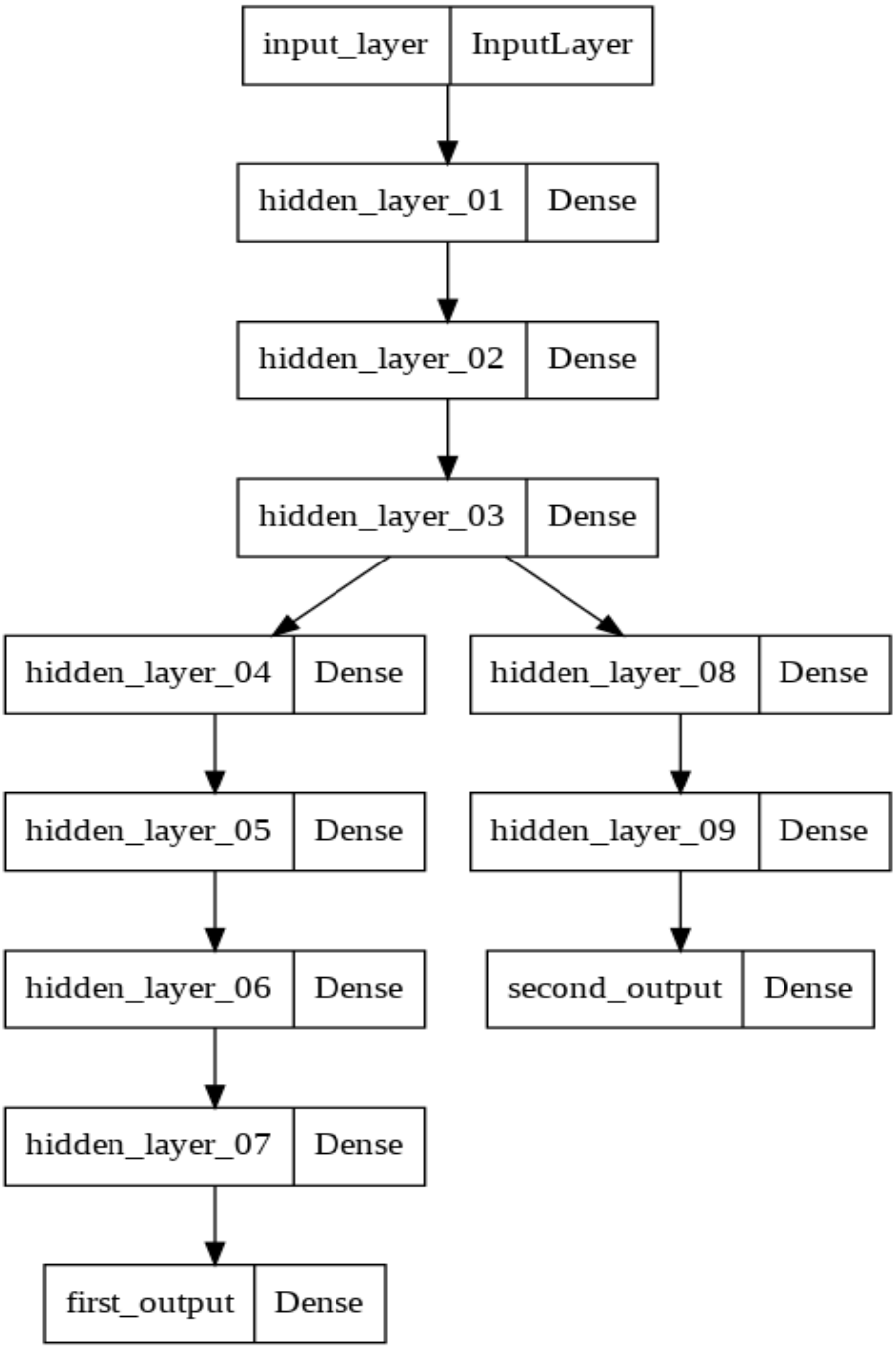

Figure 7.6: Schematic diagram of a neural network where a hidden layer is bifurcated into two different arms, which goes on predicting different parts of the output separately. In this construction, input layer: $\{P(ab|xy)\}$ (m^2k^2 neurons), first output: $\{h(ab|xy)\}$ (m^2k^2 neurons) and second output: $P_g^*(a|x,E)$ (1 neuron).

where $\{h(ab|xy)^{\text{pred}}\}_{x,y=1,\cdots,m}^{a,b=1,\cdots,k}$ are the coefficients of the predicted Bell inequality B^{pred} and $P_g^{\text{pred}}(a|x,E)$ is the predicted guessing probability.

7.4 Performance

Since the models predict two separate entities (the Bell inequality and guessing probability), the performance of the neural networks is evaluated separately for each entity. As a performance measure of predicting the guessing probability, we use mean absolute error (MAE)

$$\begin{aligned} &\text{MAE}\left(P_g^*(a|x,E), P_g^{\text{pred}}(a|x,E)\right) \\ &= \frac{1}{N_{\text{test}}} \sum_{i=1}^{N_{\text{test}}} \left| P_g^*(a|x,E)_i - P_g^{\text{pred}}(a|x,E)_i \right| , \end{aligned} \tag{7.8}$$

and mean squared error (MSE)

$$\begin{aligned} &\text{MSE}\left(P_g^*(a|x,E), P_g^{\text{pred}}(a|x,E)\right) \\ &= \frac{1}{N_{\text{test}}} \sum_{i=1}^{N_{\text{test}}} \left(P_g^*(a|x,E)_i - P_g^{\text{pred}}(a|x,E)_i \right)^2 . \end{aligned} \tag{7.9}$$

N_{test} is the number of data points in the test set and $P_g^{\text{pred}}(a|x,E)$ is the predicted guessing probability. In case of predicting the Bell inequality B (characterized by its coefficients $h(ab|xy)$), we use MAE, which reads:

$$\begin{aligned} &\text{MAE}\left(B, B^{\text{pred}}\right) \\ &= \frac{1}{m^2k^2} \frac{1}{N_{\text{test}}} \sum_{i=1}^{N_{\text{test}}} \sum_{a,b=1}^{k} \sum_{x,y=1}^{m} \left| h(ab|xy)_i - h(ab|xy)_i^{\text{pred}} \right| , \end{aligned} \tag{7.10}$$

and MSE, which reads:

$$\begin{aligned} &\text{MSE}\left(B, B^{\text{pred}}\right) \\ &= \frac{1}{m^2k^2} \frac{1}{N_{\text{test}}} \sum_{i=1}^{N_{\text{test}}} \sum_{a,b=1}^{k} \sum_{x,y=1}^{m} \left(h(ab|xy)_i - h(ab|xy)_i^{\text{pred}} \right)^2 , \end{aligned} \tag{7.11}$$

as the performance measures, where B^{pred} is the predicted Bell inequality.

We test the performance of the deep learning models in two different scenarios. The first case is when the deep learning models only predict the upper bound of the guessing probability. The guessing probability calculated from a trained DL model only estimates the upper bound with no certification. Therefore, we cannot use this estimation to bound the secret key rate. Nevertheless, the device-independent secret key rate can be bounded using a Bell inequality that produces a non-zero Bell violation for specific measurement data. Therefore, obtaining a Bell inequality with a non-zero Bell violation from the measurement statistics verifies the nonlocality of the input-output correlation and ensures that the guessing probability will be less than one. Thus, in the second scenario, we predict the guessing probability and the optimal Bell inequality (employed to bound the guessing probability).

While forecasting the guessing probability $P_g^*(a|x,E)$ only, the trained deep learning model (NN_1 neural network architecture with one output neuron in the output layer, see Fig. 7.5 for visualization) predicts the guessing probability with mean absolute error in the range of 10^{-3} to 10^{-2} (mean squared error is in the range of 10^{-5} to 10^{-3}). Such small errors without the knowledge of Bell inequality violation or the structure of the quantum correlations set are truly remarkable. The errors in the prediction of the guessing probability increased with the complexity of the Bell scenario (i.e. with the increased number of inputs and outputs) since it involves a more complex quantum set, and the number of input neurons increased with m and k in the Bell scenario; see Appendix. B for a detailed description.

For the second scenario, deep learning models are trained to predict the guessing probability $P_g^*(a|x,E)$ as well as the optimal Bell inequality B that can be used to bound the guessing probability. Both NN_1 (see Fig. 7.5) and NN_2 (see Fig. 7.6) neural network architectures are employed to perform this task. Both neural network architectures predict the guessing probability and the optimal Bell inequality with high accuracy and low error. For the task of predicting the guessing probability, MAE ranges between 10^{-4} to 10^{-2} (MSE ranges between 10^{-6} to 10^{-3}). While predicting the optimal Bell inequality B, MAE ranges between 10^{-3} to 10^{-2} (MSE ranges between 10^{-4} to 10^{-2}); see Appendix. B for a detailed description. Note that the errors in the prediction of the optimal Bell inequality increased with the complexity of the Bell scenario since the neural network has to predict more coefficients of the Bell inequality.

Another way to evaluate the quality of the predicted optimal Bell inequality B^{pred} is to use them in upper bounding the guessing probability problem,

also see Eq. (5.7). While looking into the probability of $P_g^*(a|x,E) < 1$ when $P_g^*(a|x,E)$ is calculated from the predicted Bell inequality B^{pred}, it is observed that the predicted Bell inequality can generate $P_g^*(a|x,E) < 1$ with very high probability ($\approx$ 99.5% for the $[2,2]$ Bell scenario, $\approx$ 99% for the $[3,2]$ Bell scenario, $\approx$ 94% for the $[4,2]$ Bell scenario). To further assess the predicted Bell inequality from the trained networks, one can also look into the statistical errors between the true guessing probability $P_g^*(a|x,E)$ and the guessing probability obtained using the predicted Bell inequality. The MAE ranges from 10^{-5} to 10^{-2}, while the MSE ranges from 10^{-8} to 10^{-3}; see Appendix. B for a detailed description. Similar to the previous cases, the statistical errors rise with the Bell scenario's complexity. Our deep learning method is orders of magnitude quicker compared to the usual classical solver once the deep learning model is trained. When comparing the computational runtime for the guessing probability estimation, we observe a speedup of $10^3 - 10^5$ for obtaining a prediction about a new instance compared to the runtime of the usual method for solving the semi-definite optimization problem (using the classical solver like Mosek); see Eq. (5.7) for details. When comparing the runtime comparison of predicting the optimal Bell inequality, we see a speedup of $10^3 - 10^4$ compared to the runtime of the usual method for solving the linear optimization problem (the classical solver like Mosek using PICOS [SS22] python interface); see Eq. (3.15) for details. For runtime comparison between the classical solver and the neural network approach, see Appendix B. The difference in runtime is also increasing between our approach and the classical solver method with the number of measurement settings m (or outcomes per measurement k) in the Bell scenario. This follows from the fact that the number of variables in the optimization process of Eq. (5.7) increases exponentially with the number of measurement settings (or outcomes per measurement) in the Bell scenario. Thus, it takes more computational time to perform a semi-definite program (or a linear program) using a classical solver. Only the functional output is computed by a trained neural network using its optimal weights and biases. The only factor influencing how long it takes to compute the prediction task is the neural network size.

In a nutshell, using ML and DL methods provides an alternative route to obtain the guessing probability. The fast processing of the guessing probability estimation is particularly relevant as it can be applied to randomness certification. With current technology, the Bell test event rates are around 100 kHz, which results in new data every 10μs [BKG+18]. As a result, on a single CPU, it is already too fast for traditional SDP solvers. Implementing

our deep learning strategy in those situations may be beneficial. The size of a deep neural network that can process every event during an experiment can be optimized. We also develop deep learning models and apply supervised machine learning to estimate the optimal Bell inequality, which can be utilized to bound the secret key rate via upper bounding the guessing probability for a DIQKD protocol.

While estimating the guessing probability and the optimal Bell inequality, the statistical errors increase with the Bell scenario's complexity (i.e. an increase in the number of measurements per party). Since there are more inputs and outputs in the Bell scenario, the neural network architectures may not be able to generalize the complex system with a small number of hidden layers and nodes in each layer. One can take two approaches to reduce errors. Either one creates a larger dataset to train the model or creates a sizeable neural network architecture with more hidden layers or nodes in every layer. Nevertheless, employing a larger dataset for training or training a more extensive neural network will take more time. In large networks, overfitting is another concern. Additionally, the neural network's size affects how long a new instance takes to compute. Therefore, one has to optimize the neural network architectures while balancing speed and accuracy.

Applying neural network architectures to the Bell scenarios with more measurement settings and outcomes is the logical next step. One can also expand this approach to multipartite scenarios. Investigating several different neural network architectures is another area that merits investigation for future effort. Finally, beyond the benefit of speed, one may look for new Bell inequalities using neural network architectures.

8

Conclusion & Outlook

Two main novel topics constitute the core of this dissertation. On the one hand, we discuss the device-independent (DI) quantum key distribution (QKD) using post-selected Bell inequalities. On the other hand, we illustrate the utility of neural network architectures in predicting the guessing probability and the optimal Bell inequality.

Due to data security concerns, quantum cryptography has recently attracted much attention. This is because it successfully infuses the concepts of quantum physics and computer science, which facilitates information-theoretic secure communication between two parties. However, due to the deviation between theoretical models and practical devices, the security of such systems cannot be ensured. Therefore, we consider the device-independent (DI) approach, where the security is independent of any presumptions regarding the fundamental characteristics of the devices and quantum signals. Instead, it relies on a loophole-free Bell inequality violation. Thus the selection of a suitable Bell inequality has paramount importance in DIQKD. In this context, we introduce a novel DIQKD protocol in our article [DKB22a] (which is the content of Chap. 6), where the Bell inequality is constructed from the measurement statistics of the experiment instead of using a pre-specified one. Furthermore, the Bell inequality is designed to lead to the maximum achievable Bell violation for that specific measurement statistics. We employ a semidefinite programming technique to upper bound the guessing probability while using the constructed Bell inequality and corresponding Bell violation as a constraint. The upper bound of the guessing probability is then used to bound the device-independent secret key rate (DISKR) by lower bounding the min-entropy. The hierarchical

structure of the quantum correlation set, e.g. NPA hierarchy, is employed to solve this optimization problem. Besides conceptualizing a DIQKD model using post-selected Bell inequalities, we provide a detailed finite-size secret key analysis of our proposed procedure. Our approach is flexible and can be adapted for generalized Bell scenarios, i.e., n parties, m measurement settings, and k different outcomes for each measurement setting. Implementing our method for non-optimal settings, one can generate higher DISKR than the standard CHSH inequality. Moreover, compared to related approaches [NSPS14, BSS14], our method yields a robust Bell inequality that is stable against small fluctuations in the measurement settings or the shared state and needs fewer measurement rounds to generate a non-zero DISKR. Our systematic analysis is not exhaustive. Future research must use more sophisticated methods of bounding the conditional von Neumann entropy [SGP+21, BFF21a, BFF21b], which could increase the secret key rate compared to the bounds based on the min-entropy.

In [DKB22b] (which is the content of Chap. 7), we introduce a novel method to estimate the guessing probability that uses trained deep learning models to bypass the computationally complex and cumbersome semidefinite optimization process. We have also built deep learning models to select a suitable Bell inequality that can then be employed for DISKR calculation. To generate the input data for the supervised machine learning process, we have demonstrated a method for sampling random quantum probability distribution using facet Bell inequalities for different Bell scenarios. Following our approach, one can estimate the guessing probability and the optimal Bell inequality (that can be used to get the optimal guessing probability) with extremely high precision and low average error. Moreover, our approach requires minimal computational resources, and we get a speed-up of 10^3 to 10^5 for obtaining a new prediction (about guessing probability or optimal Bell inequality) compared to the runtime of using a classical SDP solver. This approach can also be expanded to multipartite scenarios. Beyond the advantage of speed, one may use neural network architectures to look for new Bell inequalities beneficial to the guessing probability estimation problem. However, our methodology does not account for uncertainty or offer certifications of the output. It remains for future work to use techniques like probabilistic modeling [Gha15] that can certify the correctness of the model's output.

Overall, our study fits into the global movement toward the quantum revolution, which promises practical uses for quantum technology. We hope that our doctoral work has aided in the ongoing effort to turn DIQKD

protocols into practical cryptographic solutions and sparked an additional fundamental investigation into the burgeoning subject of quantum cryptography.

Bibliography

[AAA+16] Benjamin P Abbott, Richard Abbott, TD Abbott, MR Abernathy, Fausto Acernese, Kendall Ackley, Carl Adams, Thomas Adams, Paolo Addesso, RX Adhikari, et al. Observation of gravitational waves from a binary black hole merger. *Physical Review Letters*, 116(6):061102, 2016.

[AAB+15] 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, Yangqing 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, 2015. Software available from tensorflow.org.

[ABG+07] Antonio Acín, Nicolas Brunner, Nicolas Gisin, Serge Massar, Stefano Pironio, and Valerio Scarani. Device-independent security of quantum cryptography against collective attacks. *Physical Review Letters*, 98(23):230501, 2007.

[AFDF+18] Rotem Arnon-Friedman, Frédéric Dupuis, Omar Fawzi, Renato Renner, and Thomas Vidick. Practical device-independent quantum cryptography via entropy accumulation. *Nature Communications*, 9(1):1–11, 2018.

[AFRV19] Rotem Arnon-Friedman, Renato Renner, and Thomas Vidick. Simple and tight device-independent security proofs. *SIAM Journal on Computing*, 48(1):181–225, 2019.

[AGM06] Antonio Acin, Nicolas Gisin, and Lluis Masanes. From bell's theorem to secure quantum key distribution. *Physical Review Letters*, 97(12):120405, 2006.

[AlQ19] Mohammed AlQuraishi. Alphafold at casp13. *Bioinformatics*, 35(22):4862–4865, 2019.

[Alt92] Naomi S Altman. An introduction to kernel and nearest-neighbor nonparametric regression. *The American Statistician*, 46(3):175–185, 1992.

[AM16] Antonio Acín and Lluis Masanes. Certified randomness in quantum physics. *Nature*, 540(7632):213–219, 2016.

[AMP06] Antonio Acin, Serge Massar, and Stefano Pironio. Efficient quantum key distribution secure against no-signalling eavesdroppers. *New Journal of Physics*, 8(8):126, 2006.

[AMP12] Antonio Acín, Serge Massar, and Stefano Pironio. Randomness versus nonlocality and entanglement. *Physical Review Letters*, 108(10):100402, 2012.

[ApS19] Mosek ApS. Mosek optimization toolbox for matlab. *User's Guide and Reference Manual, Version*, 4, 2019.

[BB84] Charles H Bennett and Gilles Brassard. Proceedings of the IEEE International Conference on Computers, Systems and Signal Processing, 1984.

[BB04] John S Bell and John Stewart Bell. *Speakable and unspeakable in quantum mechanics: Collected papers on quantum philosophy*. Cambridge University Press, 2004.

[BBC+13] Rahul Biswas, Lindy Blackburn, Junwei Cao, Reed Essick, Kari Alison Hodge, Erotokritos Katsavounidis, Kyungmin Kim, Young-Min Kim, Eric-Olivier Le Bigot, Chang-Hwan Lee, et al. Application of machine learning algorithms to the study of noise artifacts in gravitational-wave data. *Physical Review D*, 88(6):062003, 2013.

[BBV04] Stephen Boyd, Stephen P Boyd, and Lieven Vandenberghe. *Convex optimization*. Cambridge University Press, 2004.

[BCMT17] Peter Broecker, Juan Carrasquilla, Roger G Melko, and Simon Trebst. Machine learning quantum phases of matter beyond the fermion sign problem. *Scientific Reports*, 7(1):1–10, 2017.

[BCP+14] Nicolas Brunner, Daniel Cavalcanti, Stefano Pironio, Valerio Scarani, and Stephanie Wehner. Bell nonlocality. *Reviews of Modern Physics*, 86(2):419, 2014.

[Bel64] John S Bell. On the einstein podolsky rosen paradox. *Physics Physique Fizika*, 1(3):195, 1964.

[Ben92] Charles H Bennett. Quantum cryptography using any two nonorthogonal states. *Physical Review Letters*, 68(21):3121, 1992.

[BFF21a] Peter Brown, Hamza Fawzi, and Omar Fawzi. Computing conditional entropies for quantum correlations. *Nature Communications*, 12(1):1–12, 2021.

[BFF21b] Peter Brown, Hamza Fawzi, and Omar Fawzi. Device-independent lower bounds on the conditional von neumann entropy. *arXiv preprint arXiv:2106.13692*, 2021.

[BFOS84] Leo Breiman, Jerome H Friedman, Richard A Olshen, and Charles J Stone. Classification and regression trees. belmont, ca: Wadsworth. *International Group*, 432:151–166, 1984.

[BHK05] Jonathan Barrett, Lucien Hardy, and Adrian Kent. No signaling and quantum key distribution. *Physical Review Letters*, 95(1):010503, 2005.

[BHP+13] Hannes Bernien, Bas Hensen, Wolfgang Pfaff, Gerwin Koolstra, Machiel S Blok, Lucio Robledo, Tim H Taminiau, Matthew Markham, Daniel J Twitchen, Lilian Childress, et al. Heralded entanglement between solid-state qubits separated by three metres. *Nature*, 497(7447):86–90, 2013.

[BHVK19] Kishor Bharti, Tobias Haug, Vlatko Vedral, and Leong-Chuan Kwek. How to teach ai to play bell non-local games: Reinforcement learning. *arXiv preprint arXiv:1912.10783*, 2019.

[BKB17] Felix Bischof, Hermann Kampermann, and Dagmar Bruß. Measurement-device-independent randomness generation with arbitrary quantum states. *Physical Review A*, 95(6):062305, 2017.

[BKG+18] Peter Bierhorst, Emanuel Knill, Scott Glancy, Yanbao Zhang, Alan Mink, Stephen Jordan, Andrea Rommal, Yi-Kai Liu, Bradley Christensen, Sae Woo Nam, et al. Experimentally generated randomness certified by the impossibility of superluminal signals. *Nature*, 556(7700):223–226, 2018.

[BLM+05] Jonathan Barrett, Noah Linden, Serge Massar, Stefano Pironio, Sandu Popescu, and David Roberts. Nonlocal correlations as an information-theoretic resource. *Physical Review A*, 71(2):022101, 2005.

[BMF+16] Kamil Brádler, Mohammad Mirhosseini, Robert Fickler, Anne Broadbent, and Robert Boyd. Finite-key security analysis for multilevel quantum key distribution. *New Journal of Physics*, 18(7):073030, 2016.

[BN18] Artur Barasiński and Mateusz Nowotarski. Volume of violation of Bell-type inequalities as a measure of nonlocality. *Physical Review A*, 98(2):022132, 2018.

[Bre01] Leo Breiman. Random forests. *Machine learning*, 45(1):5–32, 2001.

[Bru98] Dagmar Bruß. Optimal eavesdropping in quantum cryptography with six states. *Physical Review Letters*, 81(14):3018, 1998.

[BSS14] Jean-Daniel Bancal, Lana Sheridan, and Valerio Scarani. More randomness from the same data. *New Journal of Physics*, 16(3):033011, 2014.

[BW91] Wray L Buntine and Andreas S Weigend. Bayesian backpropagation. *Complex Systems*, 5(6):603–643, 1991.

[CBC19] Askery Canabarro, Samuraí Brito, and Rafael Chaves. Machine learning nonlocal correlations. *Physical Review Letters*, 122(20):200401, 2019.

[CC19] Thomas Cope and Roger Colbeck. Bell inequalities from no-signaling distributions. *Phys. Rev. A*, 100:022114, Aug 2019.

[CCMK17] Kelvin Ch'Ng, Juan Carrasquilla, Roger G Melko, and Ehsan Khatami. Machine learning phases of strongly correlated fermions. *Physical Review X*, 7(3):031038, 2017.

[CCX+18] Jing Chen, Song Cheng, Haidong Xie, Lei Wang, and Tao Xiang. Equivalence of restricted boltzmann machines and tensor network states. *Physical Review B*, 97(8):085104, 2018.

[CG04] Daniel Collins and Nicolas Gisin. A relevant two qubit bell inequality inequivalent to the chsh inequality. *Journal of Physics A: Mathematical and General*, 37(5):1775, 2004.

[CG16] Tianqi Chen and Carlos Guestrin. Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, pages 785–794, 2016.

[CG19] E Zambrini Cruzeiro and Nicolas Gisin. Complete list of tight bell inequalities for two parties with four binary settings. *Physical Review A*, 99(2):022104, 2019.

[CGL+02] Daniel Collins, Nicolas Gisin, Noah Linden, Serge Massar, and Sandu Popescu. Bell inequalities for arbitrarily high-dimensional systems. *Physical Review Letters*, 88(4):040404, 2002.

[CHKN17] Jonathan Carifio, James Halverson, Dmitri Krioukov, and Brent D Nelson. Machine learning in the string landscape. *Journal of High Energy Physics*, 2017(9):1–36, 2017.

[Cho15] François Chollet. keras. `https://github.com/fchollet/keras`, 2015.

[CHSH69] John F Clauser, Michael A Horne, Abner Shimony, and Richard A Holt. Proposed experiment to test local hidden-variable theories. *Physical Review Letters*, 23(15):880, 1969.

[Cir80] Boris S Cirel'son. Quantum generalizations of bell's inequality. *Letters in Mathematical Physics*, 4(2):93–100, 1980.

[CKR09] Matthias Christandl, Robert König, and Renato Renner. Postselection technique for quantum channels with applications to quantum cryptography. *Physical Review Letters*, 102(2):020504, 2009.

[CKW00] Valerie Coffman, Joydip Kundu, and William K Wootters. Distributed entanglement. *Physical Review A*, 61(5):052306, 2000.

[CM17] Juan Carrasquilla and Roger G Melko. Machine learning phases of matter. *Nature Physics*, 13(5):431–434, 2017.

[CMA+13] Brad G Christensen, Kevin T McCusker, Joseph B Altepeter, Brice Calkins, Thomas Gerrits, Adriana E Lita, Aaron Miller, Lynden K Shalm, Yanbao Zhang, Sae Woo Nam, et al. Detection-loophole-free test of quantum nonlocality, and applications. *Physical Review Letters*, 111(13):130406, 2013.

[CR95] Yves Chauvin and David E Rumelhart. *Backpropagation: theory, architectures, and applications*. Psychology Press, 1995.

[CT17] Giuseppe Carleo and Matthias Troyer. Solving the quantum many-body problem with artificial neural networks. *Science*, 355(6325):602–606, 2017.

[CW79] J Lawrence Carter and Mark N Wegman. Universal classes of hash functions. *Journal of Computer and System Sciences*, 18(2):143–154, 1979.

[Cyb89] George Cybenko. Approximation by superpositions of a sigmoidal function. *Mathematics of control, signals and systems*, 2(4):303–314, 1989.

[Den18] Dong-Ling Deng. Machine learning detection of bell nonlocality in quantum many-body systems. *Physical Review Letters*, 120(24):240402, 2018.

[DFR20] Frederic Dupuis, Omar Fawzi, and Renato Renner. Entropy accumulation. *Communications in Mathematical Physics*, 379(3):867–913, 2020.

[Dir39] P. A. M. Dirac. A new notation for quantum mechanics. *Mathematical Proceedings of the Cambridge Philosophical Society*, 35(3):416–418, 1939.

[DJR05] Thomas Decker, Dominik Janzing, and Martin Rötteler. Implementation of group-covariant positive operator valued measures by orthogonal measurements. *Journal of Mathematical Physics*, 46(1):012104, 2005.

[DKB22a] Sarnava Datta, Hermann Kampermann, and Dagmar Bruß. Device-independent secret key rates via a postselected bell inequality. *Phys. Rev. A*, 105:032451, Mar 2022.

[DKB22b] Sarnava Datta, Hermann Kampermann, and Dagmar Bruß. Upper bound on the guessing probability using machine learning. *arXiv preprint arXiv:2212.08500*, 2022.

[DLS17] Dong-Ling Deng, Xiaopeng Li, and S Das Sarma. Machine learning topological states. *Physical Review B*, 96(19):195145, 2017.

[DPS02] Andrew C Doherty, Pablo A Parrilo, and Federico M Spedalieri. Distinguishing separable and entangled states. *Physical Review Letters*, 88(18):187904, 2002.

[DRGP+17] Anna De Rosier, Jacek Gruca, Fernando Parisio, Tamás Vértesi, and Wiesław Laskowski. Multipartite nonlocality and random measurements. *Physical Review A*, 96(1):012101, 2017.

[dRGP+20] Anna de Rosier, Jacek Gruca, Fernando Parisio, Tamás Vértesi, and Wiesław Laskowski. Strength and typicality of nonlocality in multisetting and multipartite bell scenarios. *Physical Review A*, 101(1):012116, 2020.

[DSG+16] Aymeric Delteil, Zhe Sun, Wei-bo Gao, Emre Togan, Stefan Faelt, and Ataç Imamoğlu. Generation of heralded entanglement between distant hole spins. *Nature Physics*, 12(3):218–223, 2016.

[Dud78] SA Dudani. The distance-weighted k-nearest neighbor rule. *IEEE trans. on systems, man and cybernetics*, 8(4):311–313, 1978.

[DW05] Igor Devetak and Andreas Winter. Distillation of secret key and entanglement from quantum states. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 461(2053):207–235, 2005.

[EKB13] Michael Epping, Hermann Kampermann, and Dagmar Bruß. Designing bell inequalities from a tsirelson bound. *Physical review letters*, 111(24):240404, 2013.

[Eke91] Artur K Ekert. Quantum cryptography based on bell's theorem. *Physical Review Letters*, 67(6):661, 1991.

[Eld03] Yonina C Eldar. A semidefinite programming approach to optimal unambiguous discrimination of quantum states. *IEEE Transactions on Information Theory*, 49(2):446–456, 2003.

[EPR35] Albert Einstein, Boris Podolsky, and Nathan Rosen. Can quantum-mechanical description of physical reality be considered complete? *Physical Review*, 47(10):777, 1935.

[F+88] Scott E Fahlman et al. *An empirical study of learning speed in back-propagation networks*. Carnegie Mellon University, Computer Science Department, 1988.

[FDRV+18] Alejandro Fonseca, Anna De Rosier, Tamás Vértesi, Wiesław Laskowski, and Fernando Parisio. Survey on the Bell nonlocality of a pair of entangled qudits. *Physical Review A*, 98(4):042105, 2018.

[Fin82] Arthur Fine. Hidden variables, joint probability, and the Bell inequalities. *Physical Review Letters*, 48(5):291, 1982.

[Fix51] Evelyn Fix. *Discriminatory analysis: nonparametric discrimination, consistency properties*. USAF School of Aviation Medicine, 1951.

[FP15] EA Fonseca and Fernando Parisio. Measure of nonlocality which is maximal for maximally entangled qutrits. *Physical Review A*, 92(3):030101, 2015.

[FSA+13] Tobias Fritz, Ana Belén Sainz, Remigiusz Augusiak, J Bohr Brask, Rafael Chaves, Anthony Leverrier, and Antonio Acín. Local orthogonality as a multipartite principle for quantum correlations. *Nature communications*, 4(1):1–7, 2013.

[Fuk03] Komei Fukuda. Cddlib reference manual. *Report version 093a, McGill University, Montréal, Quebec, Canada*, 2003.

[GB14] Michael Grant and Stephen Boyd. CVX: Matlab software for disciplined convex programming, version 2.1. `http://cvxr.com/cvx`, March 2014.

[GBC16] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep Learning*. MIT Press, 2016. `http://www.deeplearningbook.org`.

[GCDM21] Caio BD Goes, Askery Canabarro, Eduardo I Duzzioni, and Thiago O Maciel. Automated machine learning can classify bound entangled states with tomograms. *Quantum Information Processing*, 20(3):1–18, 2021.

[GFK+06] Nicolas Gisin, Sylvain Fasel, Barbara Kraus, Hugo Zbinden, and Grégoire Ribordy. Trojan-horse attacks on quantum-key-distribution systems. *Physical Review A*, 73(2):022320, 2006.

[Gha15] Zoubin Ghahramani. Probabilistic machine learning and artificial intelligence. *Nature*, 521(7553):452–459, 2015.

[GKB19] Federico Grasselli, Hermann Kampermann, and Dagmar Bruß. Conference key agreement with single-photon interference. *New Journal of Physics*, 21(12):123002, 2019.

[GLLL+11] Ilja Gerhardt, Qin Liu, Antía Lamas-Linares, Johannes Skaar, Christian Kurtsiefer, and Vadim Makarov. Full-field implementation of a perfect eavesdropper on a quantum cryptography system. *Nature communications*, 2(1):1–6, 2011.

[GLLP04] Daniel Gottesman, H-K Lo, Norbert Lutkenhaus, and John Preskill. Security of quantum key distribution with imperfect devices. In *International Symposium on Information Theory, 2004. ISIT 2004. Proceedings.*, page 136. IEEE, 2004.

[GMKB21] Federico Grasselli, Gláucia Murta, Hermann Kampermann, and Dagmar Bruß. Entropy bounds for multiparty device-independent cryptography. *PRX Quantum*, 2(1):010308, 2021.

[Gra21] Federico Grasselli. Quantum cryptography. *Quantum science and technology. Cham: Springer*, 2021.

[GRZ+04] Nicolas Gisin, Grégoire Ribordy, Hugo Zbinden, Damien Stucki, Nicolas Brunner, and Valerio Scarani. Towards practical and fast quantum cryptography. *arXiv preprint quant-ph/0411022*, 2004.

[GVW+15] Marissa Giustina, Marijn AM Versteegh, Sören Wengerowsky, Johannes Handsteiner, Armin Hochrainer, Kevin Phelan, Fabian Steinlechner, Johannes Kofler, Jan-Åke Larsson, Carlos Abellán, et al. Significant-loophole-free test of bell's theorem with entangled photons. *Physical Review Letters*, 115(25):250401, 2015.

[HHHH09] Ryszard Horodecki, Paweł Horodecki, Michał Horodecki, and Karol Horodecki. Quantum entanglement. *Reviews of Modern Physics*, 81(2):865, 2009.

[HKB20] Timo Holz, Hermann Kampermann, and Dagmar Bruß. Genuine multipartite bell inequality for device-independent conference key agreement. *Physical Review Research*, 2(2):023251, 2020.

[HKO+12] Julian Hofmann, Michael Krug, Norbert Ortegel, Lea Gérard, Markus Weber, Wenjamin Rosenfeld, and Harald Weinfurter. Heralded entanglement between widely separated atoms. *Science*, 337(6090):72–75, 2012.

[HLM17] Yashar D Hezaveh, Laurence Perreault Levasseur, and Philip J Marshall. Fast automated analysis of strong gravitational lenses with convolutional neural networks. *Nature*, 548(7669):555–557, 2017.

[HM+21] Yichen Huang, Joel E Moore, et al. Neural network representation of tensor network and chiral states. *Physical Review Letters*, 127(17):170601, 2021.

[HN92] Robert Hecht-Nielsen. Theory of the backpropagation neural network. In *Neural Networks for Perception*, pages 65–93. Elsevier, 1992.

[Ho95] Tin Kam Ho. Random decision forests. In *Proceedings of 3rd international conference on document analysis and recognition*, volume 1, pages 278–282. IEEE, 1995.

[Hoe63] Wassily Hoeffding. Large deviations in multinomial distributions. *Math. Statist*, 34:1620, 1963.

[Hoe94] Wassily Hoeffding. Probability inequalities for sums of bounded random variables. In *The Collected Works of Wassily Hoeffding*, pages 409–426. Springer, 1994.

[Hol73] Alexander Semenovich Holevo. Bounds for the quantity of information transmitted by a quantum communication channel. *Problemy Peredachi Informatsii*, 9(3):3–11, 1973.

[HPFP20] Cillian Harney, Stefano Pirandola, Alessandro Ferraro, and Mauro Paternostro. Entanglement classification via neural network quantum states. *New Journal of Physics*, 22(4):045001, 2020.

[HR10] Esther Hänggi and Renato Renner. Device-independent quantum key distribution with commuting measurements. *arXiv preprint arXiv:1009.1833*, 2010.

[HRW10] Esther Hänggi, Renato Renner, and Stefan Wolf. Efficient device-independent quantum key distribution. In *Annual International Conference on the Theory and Applications of Cryptographic Techniques*, pages 216–234. Springer, 2010.

[HSS17] Wenjian Hu, Rajiv RP Singh, and Richard T Scalettar. Discovering phases, phase transitions, and crossovers through unsupervised machine learning: A critical examination. *Physical Review E*, 95(6):062122, 2017.

[HST+20] Melvyn Ho, Pavel Sekatski, EY-Z Tan, Renato Renner, J-D Bancal, and Nicolas Sangouard. Noisy preprocessing facilitates a photonic realization of device-independent quantum key distribution. *Physical Review Letters*, 124(23):230502, 2020.

[HSTT18] Koji Hashimoto, Sotaro Sugishita, Akinori Tanaka, and Akio Tomiya. Deep learning and the ads/cft correspondence. *Physical Review D*, 98(4):046019, 2018.

[HSW89] Kurt Hornik, Maxwell Stinchcombe, and Halbert White. Multilayer feedforward networks are universal approximators. *Neural networks*, 2(5):359–366, 1989.

[HW17] Li Huang and Lei Wang. Accelerated monte carlo simulations with restricted boltzmann machines. *Physical Review B*, 95(3):035105, 2017.

[IWY02] Kyo Inoue, Edo Waks, and Yoshihisa Yamamoto. Differential phase shift quantum key distribution. *Physical Review Letters*, 89(3):037902, 2002.

[IWY03] K Inoue, E Waks, and Y Yamamoto. Differential-phase-shift quantum key distribution using coherent light. *Physical Review A*, 68(2):022317, 2003.

[Joh16] Nathaniel Johnston. Qetlab: A matlab toolbox for quantum entanglement, version 0.9. *qetlab. com*, 2016.

[KB14] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

[KCB+21] Tamás Kriváchy, Yu Cai, Joseph Bowles, Daniel Cavalcanti, and Nicolas Brunner. High-speed batch processing of semidefinite programs with feedforward neural networks. *New Journal of Physics*, 23(10):103034, 2021.

[KCC+20] Tamás Kriváchy, Yu Cai, Daniel Cavalcanti, Arash Tavakoli, Nicolas Gisin, and Nicolas Brunner. A neural network oracle for quantum nonlocality problems in networks. *npj Quantum Information*, 6(1):1–7, 2020.

[KGR05] Barbara Kraus, Nicolas Gisin, and Renato Renner. Lower and upper bounds on the secret-key rate for quantum key distribution protocols using one-way classical communication. *Physical Review Letters*, 95(8):080501, 2005.

[Koa04] Masato Koashi. Unconditional security of coherent-state quantum key distribution with a strong phase-reference pulse. *Physical Review Letters*, 93(12):120501, 2004.

[KP03] Masato Koashi and John Preskill. Secure quantum key distribution with an uncharacterized source. *Physical Review Letters*, 90(5):057902, 2003.

[KRS09] Robert Konig, Renato Renner, and Christian Schaffner. The operational meaning of min-and max-entropy. *IEEE Transactions on Information Theory*, 55(9):4337–4347, 2009.

[Lav15] Antonio Lavecchia. Machine-learning approaches in drug discovery: methods and applications. *Drug discovery today*, 20(3):318–331, 2015.

[LCMA18] Victoria Lipinska, Florian J Curchod, Alejandro Máttar, and Antonio Acín. Towards an equivalence between maximal entanglement and maximal quantum nonlocality. *New Journal of Physics*, 20(6):063043, 2018.

[LCT14] Hoi-Kwong Lo, Marcos Curty, and Kiyoshi Tamaki. Secure quantum key distribution. *Nature Photonics*, 8(8):595, 2014.

[Lev15] Anthony Leverrier. Composable security proof for continuous-variable quantum key distribution with coherent states. *Physical Review Letters*, 114(7):070501, 2015.

[LGPRC13] Anthony Leverrier, Raúl García-Patrón, Renato Renner, and Nicolas J Cerf. Security of continuous-variable quantum key distribution against general attacks. *Physical Review Letters*, 110(3):030502, 2013.

[LHL+18] Sirui Lu, Shilin Huang, Keren Li, Jun Li, Jianxin Chen, Dawei Lu, Zhengfeng Ji, Yi Shen, Duanlu Zhou, and Bei Zeng. Separability-entanglement classifier via machine learning. *Physical Review A*, 98(1):012315, 2018.

[LLR+21] Wen-Zhao Liu, Ming-Han Li, Sammy Ragy, Si-Ran Zhao, Bing Bai, Yang Liu, Peter J Brown, Jun Zhang, Roger Colbeck, Jingyun Fan, et al. Device-independent randomness expansion against quantum side information. *Nature Physics*, 17(4):448–451, 2021.

[LMC05] Hoi-Kwong Lo, Xiongfeng Ma, and Kai Chen. Decoy state quantum key distribution. *Physical Review Letters*, 94(23):230504, 2005.

[Löf04] J. Löfberg. Yalmip: A toolbox for modeling and optimization in matlab. In *In Proceedings of the CACSD Conference*, Taipei, Taiwan, 2004.

[LQMF17] Junwei Liu, Yang Qi, Zi Yang Meng, and Liang Fu. Self-learning monte carlo method. *Physical Review B*, 95(4):041101, 2017.

[LTHS88] Yann LeCun, D Touresky, G Hinton, and T Sejnowski. A theoretical framework for back-propagation. In *Proceedings of the 1988 connectionist models summer school*, volume 1, pages 21–28. CMU, Pittsburgh, Pa: Morgan Kaufmann, 1988.

[LWW+10] Lars Lydersen, Carlos Wiechers, Christoffer Wittmann, Dominique Elser, Johannes Skaar, and Vadim Makarov. Hacking commercial quantum cryptography systems by tailored bright illumination. *Nature photonics*, 4(10):686, 2010.

[LYL+18] Yang Liu, Xiao Yuan, Ming-Han Li, Weijun Zhang, Qi Zhao, Jiaqiang Zhong, Yuan Cao, Yu-Huai Li, Luo-Kan Chen, Hao Li, et al. High-speed device-independent quantum random number generation without a detection loophole. *Physical Review Letters*, 120(1):010503, 2018.

[LZZ+21] Wen-Zhao Liu, Yu-Zhe Zhang, Yi-Zheng Zhen, Ming-Han Li, Yang Liu, Jingyun Fan, Feihu Xu, Qiang Zhang, and Jian-Wei Pan. High-speed device-independent quantum key distribution against collective attacks. *arXiv preprint arXiv:2110.01480*, 2021.

[LZZ+22] Wen-Zhao Liu, Yu-Zhe Zhang, Yi-Zheng Zhen, Ming-Han Li, Yang Liu, Jingyun Fan, Feihu Xu, Qiang Zhang, and Jian-Wei Pan. Toward a photonic demonstration of device-independent quantum key distribution. *Phys. Rev. Lett.*, 129:050502, Jul 2022.

[M+97] Tom M Mitchell et al. Machine learning. 1997. *Burr Ridge, IL: McGraw Hill*, 45(37):870–877, 1997.

[Mak09] Vadim Makarov. Controlling passively quenched single photon detectors by bright light. *New Journal of Physics*, 11(6):065003, 2009.

[MMO+07] David L Moehring, Peter Maunz, Steve Olmschenk, Kelly C Younge, Dzmitry N Matsukevich, L-M Duan, and Christopher Monroe. Entanglement of single-atom quantum bits at a distance. *Nature*, 449(7158):68–71, 2007.

[MPA11] Lluís Masanes, Stefano Pironio, and Antonio Acín. Secure device-independent quantum key distribution with causally independent measurement devices. *Nature communications*, 2:238, 2011.

[MQZL05] Xiongfeng Ma, Bing Qi, Yi Zhao, and Hoi-Kwong Lo. Practical decoy state for quantum key distribution. *Physical Review A*, 72(1):012326, 2005.

[MRC+14] Lluís Masanes, Renato Renner, Matthias Christandl, Andreas Winter, and Jonathan Barrett. Full security of quantum key distribution from no-signaling constraints. *IEEE Transactions on Information Theory*, 60(8):4973–4986, 2014.

[MvDR+19] Gláucia Murta, Suzanne B van Dam, Jérémy Ribeiro, Ronald Hanson, and Stephanie Wehner. Towards a realization of device-independent quantum key distribution. *Quantum Science and Technology*, 4(3):035011, 2019.

[MY98] Dominic Mayers and Andrew Yao. Quantum cryptography with imperfect apparatus. In *Proceedings 39th Annual Symposium on Foundations of Computer Science (Cat. No. 98CB36280)*, pages 503–509. IEEE, 1998.

[MY18] Yue-Chi Ma and Man-Hong Yung. Transforming bell's inequalities into state classifiers with machine learning. *NPJ Quantum Information*, 4(1):1–10, 2018.

[NC10] M.A. Nielsen and I.L. Chuang. *Quantum Computation and Quantum Information: 10th Anniversary Edition*. Cambridge University Press, 2010.

[NDN+21] David P Nadlinger, Peter Drmota, Bethan C Nichol, Gabriel Araneda, Dougal Main, Raghavendra Srinivas, David M Lucas, Christopher J Ballance, Kirill Ivanov, E Tan, et al. Device-independent quantum key distribution. *arXiv preprint arXiv:2109.14600*, 2021.

[NDN+22] DP Nadlinger, P Drmota, BC Nichol, G Araneda, D Main, R Srinivas, DM Lucas, CJ Ballance, K Ivanov, EY-Z Tan, et al. Experimental quantum key distribution certified by bell's theorem. *Nature*, 607(7920):682–686, 2022.

[NPA07] Miguel Navascués, Stefano Pironio, and Antonio Acín. Bounding the set of quantum correlations. *Physical Review Letters*, 98(1):010401, 2007.

[NPA08] Miguel Navascués, Stefano Pironio, and Antonio Acín. A convergent hierarchy of semidefinite programs characterizing the set of quantum correlations. *New Journal of Physics*, 10(7):073013, 2008.

[NSBSP18] Olmo Nieto-Silleras, Cédric Bamps, Jonathan Silman, and Stefano Pironio. Device-independent randomness generation from several bell estimators. *New Journal of Physics*, 20(2):023049, 2018.

[NSPS14] Olmo Nieto-Silleras, Stefano Pironio, and Jonathan Silman. Using complete measurement statistics for optimal device-independent randomness evaluation. *New Journal of Physics*, 16(1):013035, 2014.

[PAB+09] Stefano Pironio, Antonio Acin, Nicolas Brunner, Nicolas Gisin, Serge Massar, and Valerio Scarani. Device-independent quantum key distribution secure against collective attacks. *New Journal of Physics*, 11(4):045021, 2009.

[PAM+10] Stefano Pironio, Antonio Acín, Serge Massar, A Boyer de La Giroday, Dzmitry N Matsukevich, Peter Maunz, Steven Olmschenk, David Hayes, Le Luo, T Andrew Manning, et al. Random numbers certified by bell's theorem. *Nature*, 464(7291):1021–1024, 2010.

[Par12] Matteo GA Paris. The modern tools of quantum mechanics. *The European Physical Journal Special Topics*, 203(1):61–86, 2012.

[Pas16] Mario Pasquato. Detecting intermediate mass black holes in globular clusters with machine learning. *arXiv preprint arXiv:1606.08548*, 2016.

[Per06] Asher Peres. Quantum theory: concepts and methods, vol. 57. *Fundamental Theories of Physics*, 2006.

[Pit89] I Pitowski. Quantum probability. *Quantum Logic*, 1989.

[Pit91] Itamar Pitowsky. Correlation polytopes: their geometry and complexity. *Mathematical Programming*, 50(1-3):395–414, 1991.

[PM13] Stefano Pironio and Serge Massar. Security of practical private randomness generation. *Physical Review A*, 87(1):012336, 2013.

[PPK+09] Marcin Pawłowski, Tomasz Paterek, Dagomir Kaszlikowski, Valerio Scarani, Andreas Winter, and Marek Żukowski. Information causality as a physical principle. *Nature*, 461(7267):1101–1104, 2009.

[PR94] Sandu Popescu and Daniel Rohrlich. Quantum nonlocality as an axiom. *Foundations of Physics*, 24(3):379–385, 1994.

[PR22] Christopher Portmann and Renato Renner. Security in quantum cryptography. *Reviews of Modern Physics*, 94(2):025008, 2022.

[PV10] Károly F Pál and Tamás Vértesi. Maximal violation of a bipartite three-setting, two-outcome bell inequality using infinite-dimensional quantum systems. *Physical Review A*, 82(2):022116, 2010.

[PVG+11] Fabian Pedregosa, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, et al. Scikit-learn: Machine learning in python. *The Journal of Machine Learning Research*, 12:2825–2830, 2011.

[Qui86] J. Ross Quinlan. Induction of decision trees. *Machine learning*, 1(1):81–106, 1986.

[Rai01] Eric M Rains. A semidefinite program for distillable entanglement. *IEEE Transactions on Information Theory*, 47(7):2921–2933, 2001.

[RBG+17] Wenjamin Rosenfeld, Daniel Burchardt, Robert Garthoff, Kai Redeker, Norbert Ortegel, Markus Rau, and Harald Weinfurter. Event-ready bell test using entangled atoms simultaneously closing detection and locality loopholes. *Physical Review Letters*, 119(1):010402, 2017.

[Ren08] Renato Renner. Security of quantum key distribution. *International Journal of Quantum Information*, 6(01):1–127, 2008.

[RGK05] Renato Renner, Nicolas Gisin, and Barbara Kraus. Information-theoretic security proof for quantum-key-distribution protocols. *Physical Review A*, 72(1):012332, 2005.

[RMW18] Jérémy Ribeiro, Gláucia Murta, and Stephanie Wehner. Fully device-independent conference key agreement. *Physical Review A*, 97(2):022307, 2018.

[RMW19] Jérémy Ribeiro, Gláucia Murta, and Stephanie Wehner. Reply to "comment on 'fully device-independent conference key agreement'". *Physical Review A*, 100(2):026302, 2019.

[RR12] Joseph M Renes and Renato Renner. One-shot classical data compression with quantum side information and the distillation of common randomness or secret keys. *IEEE Transactions on Information Theory*, 58(3):1985–1991, 2012.

[RSA78] R. L. Rivest, A. Shamir, and L. Adleman. A method for obtaining digital signatures and public-key cryptosystems. *Commun. ACM*, 21(2):120–126, feb 1978.

[RW05] Renato Renner and Stefan Wolf. Simple and tight bounds for information reconciliation and privacy amplification. In *International conference on the Theory and Application of Cryptology and Information security*, pages 199–216. Springer, 2005.

[SARG04] Valerio Scarani, Antonio Acin, Grégoire Ribordy, and Nicolas Gisin. Quantum cryptography protocols robust against photon number splitting attacks for weak laser pulse implementations. *Physical Review Letters*, 92(5):057901, 2004.

[SBG+05] Damien Stucki, Nicolas Brunner, Nicolas Gisin, Valerio Scarani, and Hugo Zbinden. Fast and simple one-way quantum key distribution. *Applied Physics Letters*, 87(19):194108, 2005.

[SBPC+09] Valerio Scarani, Helle Bechmann-Pasquinucci, Nicolas J Cerf, Miloslav Dušek, Norbert Lütkenhaus, and Momtchil Peev. The security of practical quantum key distribution. *Reviews of Modern Physics*, 81(3):1301, 2009.

[SC18] Paul Skrzypczyk and Daniel Cavalcanti. Maximal randomness generation from steering inequality violations using qudits. *Physical Review Letters*, 120(26):260401, 2018.

[SCS+16] Samuel S Schoenholz, Ekin D Cubuk, Daniel M Sussman, Efthimios Kaxiras, and Andrea J Liu. A structural approach

to relaxation in glassy liquids. *Nature Physics*, 12(5):469–471, 2016.

[SGP+21] René Schwonnek, Koon Tong Goh, Ignatius W Primaatmaja, Ernest Y-Z Tan, Ramona Wolf, Valerio Scarani, and Charles C-W Lim. Device-independent quantum key distribution with random key basis. *Nature Communications*, 12(1):1–8, 2021.

[Sha49] Claude E Shannon. Communication theory of secrecy systems. *The Bell System Technical Journal*, 28(4):656–715, 1949.

[Sho99] Peter W Shor. Polynomial-time algorithms for prime factorization and discrete logarithms on a quantum computer. *SIAM review*, 41(2):303–332, 1999.

[SKB17] Jochen Szangolies, Hermann Kampermann, and Dagmar Bruß. Device-independent bounds on detection efficiency. *Physical Review Letters*, 118(26):260401, 2017.

[SMSC+15] Lynden K Shalm, Evan Meyer-Scott, Bradley G Christensen, Peter Bierhorst, Michael A Wayne, Martin J Stevens, Thomas Gerrits, Scott Glancy, Deny R Hamel, Michael S Allman, et al. Strong loophole-free test of local realism. *Physical Review Letters*, 115(25):250402, 2015.

[SP00] Peter W Shor and John Preskill. Simple proof of security of the bb84 quantum key distribution protocol. *Physical Review Letters*, 85(2):441, 2000.

[SR08a] Valerio Scarani and Renato Renner. Quantum cryptography with finite resources: Unconditional security bound for discrete-variable protocols with one-way postprocessing. *Physical review letters*, 100(20):200501, 2008.

[SR08b] Valerio Scarani and Renato Renner. Security bounds for quantum cryptography with finite resources. In *Workshop on Quantum Computation, Communication, and Cryptography*, pages 83–95. Springer, 2008.

[SRN17] Frank Schindler, Nicolas Regnault, and Titus Neupert. Probing many-body localization with neural networks. *Physical Review B*, 95(24):245134, 2017.

[SS22] Guillaume Sagnol and Maximilian Stahlberg. Picos: A python interface to conic optimization solvers. *Journal of Open Source Software*, 7(70):3915, 2022.

[SSBD14] Shai Shalev-Shwartz and Shai Ben-David. *Understanding machine learning: From theory to algorithms*. Cambridge university press, 2014.

[SSS+17] David Silver, Julian Schrittwieser, Karen Simonyan, Ioannis Antonoglou, Aja Huang, Arthur Guez, Thomas Hubert, Lucas Baker, Matthew Lai, Adrian Bolton, et al. Mastering the game of go without human knowledge. *Nature*, 550(7676):354–359, 2017.

[Stu99] Jos F Sturm. Using sedumi 1.02, a matlab toolbox for optimization over symmetric cones. *Optimization methods and software*, 11(1-4):625–653, 1999.

[SZB+21] Lynden K Shalm, Yanbao Zhang, Joshua C Bienfang, Collin Schlager, Martin J Stevens, Michael D Mazurek, Carlos Abellán, Waldimar Amaya, Morgan W Mitchell, Mohammad A Alhejji, et al. Device-independent randomness expansion with entangled photons. *Nature Physics*, 17(4):452–456, 2021.

[TCR09] Marco Tomamichel, Roger Colbeck, and Renato Renner. A fully quantum asymptotic equipartition property. *IEEE Transactions on Information Theory*, 55(12):5840–5847, 2009.

[TKI03] Kiyoshi Tamaki, Masato Koashi, and Nobuyuki Imoto. Unconditionally secure key distribution based on two nonorthogonal states. *Physical Review Letters*, 90(16):167904, 2003.

[TL17] Marco Tomamichel and Anthony Leverrier. A largely self-contained and complete security proof for quantum key distribution. *Quantum*, 1:14, 2017.

[TLGR12] Marco Tomamichel, Charles Ci Wen Lim, Nicolas Gisin, and Renato Renner. Tight finite-key analysis for quantum cryptography. *Nature communications*, 3(1):1–6, 2012.

[TLR20] Ernest Y-Z Tan, Charles C-W Lim, and Renato Renner. Advantage distillation for device-independent quantum key distribution. *Physical Review Letters*, 124(2):020502, 2020.

[TM17] Giacomo Torlai and Roger G Melko. Neural decoder for topological codes. *Physical review letters*, 119(3):030501, 2017.

[TMMPE14] Marco Tomamichel, Jesus Martinez-Mateo, Christoph Pacher, and David Elkouss. Fundamental finite key limits for information reconciliation in quantum key distribution. In *2014 IEEE International Symposium on Information Theory*, pages 1469–1473. IEEE, 2014.

[Tom15] Marco Tomamichel. *Quantum Information Processing with Finite Resources: Mathematical Foundations*, volume 5. Springer, 2015.

[TR11] Marco Tomamichel and Renato Renner. Uncertainty relation for smooth entropies. *Physical Review Letters*, 106(11):110506, 2011.

[TSG+21] Ernest Y-Z Tan, René Schwonnek, Koon Tong Goh, Ignatius William Primaatmaja, and Charles C-W Lim. Computing secure key rates for quantum cryptography with untrusted devices. *npj Quantum Information*, 7(1):1–6, 2021.

[TSSR11] Marco Tomamichel, Christian Schaffner, Adam Smith, and Renato Renner. Leftover hashing against quantum side information. *IEEE Transactions on Information Theory*, 57(8):5524–5535, 2011.

[TTT99] Kim-Chuan Toh, Michael J Todd, and Reha H Tütüncü. Sdpt3—a matlab software package for semidefinite programming, version 1.3. *Optimization methods and software*, 11(1-4):545–581, 1999.

[VB96] Lieven Vandenberghe and Stephen Boyd. Semidefinite programming. *SIAM review*, 38(1):49–95, 1996.

[VDTR13] Alexander Vitanov, Frederic Dupuis, Marco Tomamichel, and Renato Renner. Chain rules for smooth min-and max-entropies. *IEEE Transactions on Information Theory*, 59(5):2603–2612, 2013.

[Ver26] Gilbert S Vernam. Cipher printing telegraph systems: For secret wire and radio telegraphic communications. *Journal of the AIEE*, 45(2):109–115, 1926.

[VMH01] Artem Vakhitov, Vadim Makarov, and Dag R Hjelme. Large pulse attack as a method of conventional optical eavesdropping in quantum cryptography. *Journal of Modern Optics*, 48(13):2023–2038, 2001.

[VMR+88] Thomas P Vogl, JK Mangis, AK Rigler, WT Zink, and DL Alkon. Accelerating the convergence of the back-propagation method. *Biological Cybernetics*, 59(4-5):257–263, 1988.

[VNLH17] Evert PL Van Nieuwenburg, Ye-Hua Liu, and Sebastian D Huber. Learning phase transitions by confusion. *Nature Physics*, 13(5):435–439, 2017.

[VV14] Umesh Vazirani and Thomas Vidick. Fully device-independent quantum key distribution. *Physical Review Letters*, 113(14):140501, 2014.

[Wan16] Lei Wang. Discovering phase transitions with unsupervised learning. *Physical Review B*, 94(19):195105, 2016.

[WBC+14] Nino Walenta, Andreas Burg, Dario Caselunghe, Jeremy Constantin, Nicolas Gisin, Olivier Guinnard, Raphaël Houlmann, Pascal Junod, Boris Korzh, Natalia Kulesza, et al. A fast and versatile quantum key distribution system with hardware key distillation and wavelength multiplexing. *New Journal of Physics*, 16(1):013047, 2014.

[Wet17] Sebastian J Wetzel. Unsupervised learning of phase transitions: From principal component analysis to variational autoencoders. *Physical Review E*, 96(2):022140, 2017.

[Wil13] Mark M. Wilde. *Quantum Information Theory*. Cambridge University Press, 2013.

[Wit15] Peter Wittek. Algorithm 950: Ncpol2sdpa—sparse semidefinite programming relaxations for polynomial optimization problems of noncommuting variables. *ACM Transactions on Mathematical Software (TOMS)*, 41(3):21, 2015.

[WZ82] William K Wootters and Wojciech H Zurek. A single quantum cannot be cloned. *Nature*, 299(5886):802–803, 1982.

[XMZ+20] Feihu Xu, Xiongfeng Ma, Qiang Zhang, Hoi-Kwong Lo, and Jian-Wei Pan. Secure quantum key distribution with realistic devices. *Reviews of Modern Physics*, 92(2):025002, 2020.

[XZZP21] Feihu Xu, Yu-Zhe Zhang, Qiang Zhang, and Jian-Wei Pan. Device-independent quantum key distribution with random post selection. *arXiv preprint arXiv:2110.02701*, 2021.

[YAK18] Nobuyuki Yoshioka, Yutaka Akagi, and Hosho Katsura. Learning disordered topological phases by statistical recovery of symmetry. *Physical Review B*, 97(20):205110, 2018.

[YC19] Hua-Lei Yin and Zeng-Bing Chen. Finite-key analysis for twin-field quantum key distribution with composable security. *Scientific Reports*, 9(1):1–9, 2019.

[ZFQ+08] Yi Zhao, Chi-Hang Fred Fung, Bing Qi, Christine Chen, and Hoi-Kwong Lo. Quantum hacking: Experimental demonstration of time-shift attack against practical quantum-key-distribution systems. *Physical Review A*, 78(4):042333, 2008.

[ZSB+20] Yanbao Zhang, Lynden K Shalm, Joshua C Bienfang, Martin J Stevens, Michael D Mazurek, Sae Woo Nam, Carlos Abellán, Waldimar Amaya, Morgan W Mitchell, Honghao Fu, et al. Experimental low-latency device-independent quantum randomness. *Physical Review Letters*, 124(1):010505, 2020.

[ZvLR+21] Wei Zhang, Tim van Leent, Kai Redeker, Robert Garthoff, Rene Schwonnek, Florian Fertig, Sebastian Eppelt, Valerio Scarani, Charles C-W Lim, and Harald Weinfurter. Experimental device-independent quantum key distribution between distant users. *arXiv preprint arXiv:2110.00575*, 2021.

[ZvLR+22] Wei Zhang, Tim van Leent, Kai Redeker, Robert Garthoff, René Schwonnek, Florian Fertig, Sebastian Eppelt, Wenjamin Rosenfeld, Valerio Scarani, Charles C-W Lim, et al. A device-independent quantum key distribution system for distant users. *Nature*, 607(7920):687–691, 2022.

Device-independent secret key rates via a post-selected Bell inequality

Title:	Device-independent secret key rates via a posts-elected Bell inequality
Authors:	Sarnava Datta, Hermann Kampermann and Dagmar Bruß
Journal:	Physical Review A
Impact factor:	2.971 (2021)
Date of submission:	11 November 2021
Publication status:	Published
Contribution by S.D:	First author (Approx 85 %)

This publication corresponds to the reference [DKB22a]. A summary of the results is presented in Chap. 6. The general framework and the research objective were worked out in collaboration with my co-authors. I regularly discussed the project with my co-authors, and they repeatedly gave me valuable input. In particular, HK guided me in the introduction of semidefinite programming and DIQKD. Together with HK, I designed the DIQKD protocol. I performed all analytical calculations and carried out the finite key analysis. I created all figures in the article and wrote the MATLAB code to solve the semidefinite optimization problem of upper bounding the guessing probability. I followed DB's suggestions on what simulations to perform and how to present them in the plots. I wrote the entire manuscript, which was proofread and improved by my co-authors.

Device-independent secret key rates via a postselected Bell inequality

Sarnava Datta,* Hermann Kampermann, and Dagmar Bruß
Institut für Theoretische Physik III, Heinrich-Heine-Universität Düsseldorf, D-40225 Düsseldorf, Germany

(Received 11 November 2021; accepted 22 February 2022; published 29 March 2022)

In device-independent quantum key distribution (DIQKD) the security is not based on any assumptions about the intrinsic properties of the devices and the quantum signals but on the violation of a Bell inequality. We introduce a DIQKD scenario in which an optimal Bell inequality is constructed from the performed measurement data rather than fixing beforehand a specific Bell inequality. Our method can be employed in a general way, for any number of measurement settings and any number of outcomes. We provide an implementable DIQKD protocol and perform finite-size security key analysis for collective attacks. We compare our approach with related procedures in the literature and analyze the robustness of our protocol. We also study the performance of our method in several Bell scenarios as well as for random measurement settings.

DOI: 10.1103/PhysRevA.105.032451

I. INTRODUCTION

Data security concerns are prevalent in the modern world. One of the most prominent domains of quantum communication is quantum key distribution (QKD), which allows to distribute a secure key between two (or more) parties, namely, Alice and Bob, where the security is only based on the laws of quantum mechanics. Since the inception of QKD [1], a variety of QKD protocols [2–12] have been introduced. However, the security of these device-dependent protocols needs complete characterization of the devices, sources, and/or the channel between the parties. In a realistic scenario, the device can be not completely characterized or could even be prepared by a malicious eavesdropper (Eve). Furthermore, hacking of existing implementations that exploit experimental imperfections was demonstrated [13–15]. To overcome these drawbacks, device-independent (DI) QKD was introduced [16], where the security does not require any assumptions about the inherent properties of the devices or the dimension of the Hilbert space of the quantum signals. The security of DI protocols is based on the observation of a loophole-free Bell inequality violation [17–29] which guarantees the quantum nature of the observed data. The length of the secret key will depend on the estimated violation of the Bell inequality.

In this article we introduce a DIQKD scenario in which the Bell inequality is not agreed upon beforehand but will be constructed from the observed probability distribution of the measurement outcomes. We follow a two-step process: From the input-output probability distribution, we construct a Bell inequality that leads to the maximum Bell violation for that particular measurement setting of Alice and Bob. Then we use this optimized Bell inequality and the corresponding violation to bound the secret key rate. Note that in [30,31], the authors introduced an alternative approach to bound the device-independent secret key rate via a Bell inequality and the corresponding violation, which is also constructed from the full measurement statistics. We will relate and compare our method with theirs in the Results section (Sec. VI). In particular, we show that our procedure is advantageous in the nonasymptotic regime.

This paper is organized as follows. We start in Sec. II by briefly reviewing classical and quantum correlations. Then we explain how to obtain the optimal Bell inequality from the observed probability distribution. We lay the framework to provide a confidence interval for the Bell expectation value in Sec. III. We provide an implementable DIQKD protocol in Sec. IV and calculate the finite-size secret key rate in Sec. V. In Sec. VI we illustrate our method with several examples.

II. GENERAL FRAMEWORK

In this section we review the concept of the classical correlation polytope in Sec. II A, and, based on this, we explain in Sec. II B how to construct Bell inequalities that are maximally violated by the measurement data.

A. Set of correlations

Consider a setup for two parties[1] (namely, Alice and Bob) connected by a quantum channel. The parties perform local measurements on a joint quantum state. Let us assume that Alice and Bob have m_a and m_b measurement settings, respectively. Alice's set of measurement settings is denoted as $X = \{1, \ldots, m_a\}$, and Bob's set of measurement settings as $Y = \{1, \ldots, m_b\}$. To estimate the probability distribution from the experimental data, we have to use the measurement device N times in succession. We assume that the devices behave independently and identically (i.i.d.) in each round, i.e., the results of the ith round are independent of the past $i-1$

*Sarnava.Datta@hhu.de

[1]Note that our method can be extended in a straightforward way to n parties.

rounds. The setting of the ith round is denoted as $x_i \in X$ for Alice and $y_i \in Y$ for Bob. Each of these measurement settings has d outcomes, which are denoted as $a_i \in A = \{1, \ldots, d\}$ for Alice and $b_i \in B = \{1, \ldots, d\}$ for Bob. We call this the $[(m_a, m_b), d]$ scenario, i.e., two parties with (m_a, m_b) measurement settings and d outcomes each. When both parties have an equal number of measurement settings, i.e., $m_a = m_b = m$, we will denote this as the $[m, d]$ scenario. The joint probability of getting outcome a when Alice is using the measurement setting $x \in X$ and b when Bob uses the measurement setting $y \in Y$ is denoted as $P(A_x^a B_y^b)$. All these joint probabilities will be collected in a probability vector

$$\mathbf{P} := \left[P\left(A_x^a B_y^b\right)\right], \tag{1}$$

where $x \in X$, $y \in Y$, $a \in A$, and $b \in B$. The associated probability space is of dimension

$$D^d_{m_a,m_b} := m_a m_b d^2. \tag{2}$$

The set of all probabilities that represent a classical or locally real theory forms a convex polytope [32–34]. We denote this polytope as $\mathcal{P}$. Any probability distribution which is not contained in $\mathcal{P}$ shows nonclassical or quantum behavior and can be witnessed by the violation of a Bell inequality [35]. As illustrated in [36], the polytope of classical correlations can be characterized by its extremal points $\mathbf{v}_p$, where $p = \{1, 2, \ldots, d^{m_a+m_b}\}$, and $\mathbf{v}_p$ has entries from the set $\{0,1\}$. The extremal points of the polytope correspond to deterministic strategies. Every classical correlation $\mathbf{P}_{cl} \in \mathcal{P}$ can be written as a convex combination of all the deterministic strategies as

$$\mathbf{P}_{cl} = \sum_{p=1}^{d^{m_a+m_b}} \lambda_p \mathbf{v}_p, \tag{3}$$

where $\lambda_p \geqslant 0$ and $\sum_{p=1}^{d^{m_a+m_b}} \lambda_p = 1$. This subsequently implies that every observed probability distribution which cannot be decomposed as shown in Eq. (3) violates at least one Bell inequality.

B. Designing Bell inequalities

Consider the $[(m_a, m_b), d]$ scenario where the parties receive the measurement data $\mathbf{P}$. In order to extract a secret key from these classical measurement data, they need to violate a Bell inequality. As shown in [36], this scenario can be translated to a linear separation problem. For illustration, see Fig. 1. Bell inequalities correspond to hyperplanes in the probability space that separate the classical correlation polytope $\mathcal{P}$ from the set of all genuine quantum correlations $\mathcal{Q} \setminus \mathcal{P}$. Such hyperplanes are specified by a normal vector $\mathbf{h} \in \mathbb{R}^{D^d_{m_a,m_b}}$ with the dimension given in Eq. (2). If $\mathbf{P} \in \mathcal{Q} \setminus \mathcal{P}$, there exists at least one hyperplane $\mathbf{h}$ that separates all the vertices $\mathbf{v}_p$ of $\mathcal{P}$ from the observed probability distribution $\mathbf{P}$. We set the objective of the linear program to find the hyperplane vector $\mathbf{h}$ corresponding to the Bell inequality which is maximally violated by the measurement data $\mathbf{P}$. This optimization problem can be formulated as

$$\begin{aligned} \max_{\mathbf{h},c} \quad & \mathbf{h}^T\mathbf{P} - c \\ \text{subject to} \quad & \mathbf{h}^T\mathbf{v}_p \leqslant c \quad \forall \quad p \in \{1, \ldots, d^{m_a+m_b}\} \end{aligned}$$

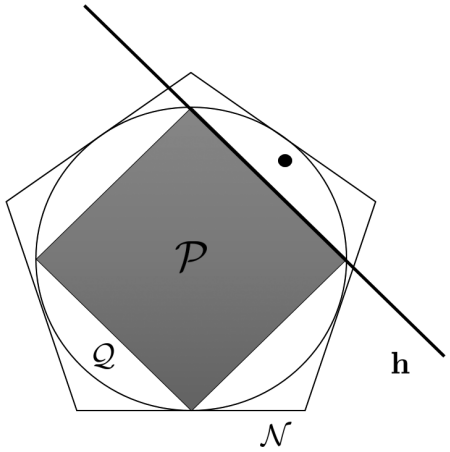

FIG. 1. A sketch for the set of correlations. All classical probabilities form a convex polytope $\mathcal{P}$, which is embedded in the set $\mathcal{Q}$ of quantum correlations, which in turn is a subset of the nonsignaling polytope $\mathcal{N}$. The Bell inequality is specified by the vector $\mathbf{h}$ defining a hyperplane which separates all vertices v_p from the observed probability distribution $\mathbf{P}$ (the black point situated outside the classical polytope $\mathcal{P}$).

$$\begin{aligned} & \mathbf{h}^T\mathbf{P} > c \\ & -1 \leqslant h_i \leqslant 1 \quad \forall \quad i \in \left\{1, \ldots, D^d_{m_a,m_b}\right\}, \end{aligned} \tag{4}$$

with the classical bound c. The additional constraint imposed on the elements of h_i of the hyperplane vector $\mathbf{h}$ keeps the maximization bounded. The chosen boundaries of h_i do not influence the result of the optimization problem besides being a global scaling factor. The hyperplane found in this manner has the form

$$\mathbf{h} = \left[h^{ab}_{A_x B_y}\right], \tag{5}$$

where $x \in X$, $y \in Y$, $a \in A$, and $b \in B$. Thus the Bell inequality found by the optimization and specified by the hyperplane vector $\mathbf{h}$ is given as

$$B[\mathbf{P}] = \sum_{a,b,x,y} h^{ab}_{A_x B_y} P\left(A_x^a B_y^b\right) \leqslant c. \tag{6}$$

Equation (6) represents the Bell inequality that is maximally violated by the observed probability distribution $\mathbf{P}$. Note that if $\mathbf{P} \in \mathcal{P}$, the optimization problem Eq. (4) is infeasible and no Bell inequality can be found.

III. STATISTICAL FLUCTUATIONS AND THEIR ESTIMATION

So far, we have concentrated on the ideal asymptotic case, that is, using the exact probabilities as entries of the observed probability distribution $\mathbf{P}$. However, in a real experiment one does not have access to probabilities but only to frequencies that are subject to statistical uncertainties and systematic errors. Since systematic errors mostly arise from specific experimental settings, we solely focus on the theoretical framework and concentrate on statistical fluctuations, as they lead to uncertainties in the observed Bell violation.

Let Alice and Bob perform N rounds of measurements. The number of instances when Alice chooses measurement $x \in X$

and Bob chooses measurement $y \in Y$ is denoted by $N_{x,y}$. In a real experiment, instead of having access to joint probabilities, we estimate them by the joint frequencies $\hat{P}(A_x^a B_y^b) = \frac{N(a,b,x,y)}{N_{x,y}}$. Here $N(a,b,x,y)$ is the number of occurrences of the corresponding input-output pair.

The Bell value $B[\hat{\mathbf{P}}]$ is a function of the joint frequencies,

$$B[\hat{\mathbf{P}}] = h^{ab}_{A_x B_y} \hat{P}(A_x^a B_y^b), \tag{7}$$

see also Eq. (6). Let $\chi(e)$ be an indicator function for a particular event e, i.e., $\chi(e) = 1$ if the event e is observed, $\chi(e) = 0$ otherwise. We introduce a random variable

$$\hat{B}_i = \sum_{a,b,x,y} h^{ab}_{A_x B_y} \frac{\chi(a_i = a, b_i = b, x_i = x, y_i = y)}{\hat{p}(x_i = x, y_i = y)},$$

where $\hat{p}(x_i = x, y_i = y) = \frac{N_{x,y}}{N}$ is the input joint frequency distribution. We get $\frac{1}{N}\sum_{i=1}^{N} \hat{B}_i = B[\hat{\mathbf{P}}]$. Defining

$$q_{\min} = \min_{a,b,x,y} \frac{h^{ab}_{A_x B_y}}{\hat{p}(x_i = x, y_i = y)},$$

$$q_{\max} = \max_{a,b,x,y} \frac{h^{ab}_{A_x B_y}}{\hat{p}(x_i = x, y_i = y)},$$

we have $q_{\min} \leqslant \hat{B}_i \leqslant q_{\max}$. We define $\gamma := q_{\max} - q_{\min}$. By using Hoeffding's inequality [37] (see Lemma 2 in Appendix A), we can bound the deviation δ of the Bell value obtained by the frequencies from the asymptotic value by a probability:

$$\Pr(B[\mathbf{P}] \geqslant B[\hat{\mathbf{P}}] - \delta) \geqslant 1 - \epsilon, \tag{8}$$

with

$$\epsilon = \exp\left(-\frac{2N\delta^2}{\gamma^2}\right). \tag{9}$$

For a given ϵ of a DIQKD protocol, one can calculate the confidence interval δ for the Bell value using Eq. (9).

IV. DIQKD MODEL AND PROTOCOL

Let us state the DIQKD protocol. We consider the i.i.d. scenario where the devices will behave independently and identically in each round. The state distributed between the parties is also the same for each round of the protocol. Alice has m measurement inputs $x \in \{1, \ldots, m\}$. Each of the inputs has d corresponding outputs $a \in \{1, \ldots, d\}$. Bob instead has $m+1$ measurement inputs $y \in \{1, \ldots, m+1\}$. Each measurement input of Bob also has d outputs $b \in \{1, \ldots, d\}$.

(1) In every round of the protocol, the parties do the following:

(a) A state ρ_{AB} is distributed between Alice and Bob.

(b) There are two types of measurement rounds, namely, raw key generation rounds and parameter estimation rounds. According to a preshared random key T, Alice and Bob choose a random $T_i = \{0, 1\}$ such that $\Pr(T_i = 1) = \xi$. If $T_i = 0$, Alice and Bob choose the measurement input $(x = 1, y = m+1)$ to generate the raw key. Otherwise, Alice and Bob choose the measurement inputs $x \in \{1, \ldots, m\}$ and $y \in \{1, \ldots, m\}$, respectively, uniformly at random. These cases will be denoted as parameter estimation rounds.

(c) The parties record their inputs and outputs as (x_i, y_i) and (a_i, b_i). After N rounds of measurement, we denote the input bit strings as X^N and Y^N and output bit strings as A^N and B^N for Alice and Bob, respectively.

(2) Alice and Bob publicly reveal their measurement outcomes of the parameter estimation rounds. They divide the parameter estimation rounds' data randomly into three sets (Alice specifies to which set each parameter estimation round's data belongs, according to a random number generator in her possession). From the first set, Alice and Bob estimate the frequencies $\hat{\mathbf{P}}_1 = [\hat{P}(A_x^a B_y^b)]$ [see Eq. (1)]. If $\hat{\mathbf{P}}_1$ is inside the classical correlation polytope $\mathcal{P}$, the protocol aborts. Otherwise, they construct an optimal Bell inequality by solving the linear optimization in Eq. (4). Then Alice and Bob use the data from the second set to calculate the Bell value $B[\hat{\mathbf{P}}_2]$. They then bound the deviation of this estimated Bell value $B[\hat{\mathbf{P}}_2]$ from the real Bell value $B[\mathbf{P}]$ by [see Eq. (8)]

$$\Pr(B[\mathbf{P}] \geqslant B[\hat{\mathbf{P}}_2] - \delta_{\text{est}}) \geqslant 1 - \epsilon_{\text{est}}, \tag{10}$$

where $\epsilon_{\text{est}} = \exp\left(-\frac{2N\xi\delta_{\text{est}}^2}{3\gamma^2}\right)$ and $\frac{N\xi}{3}$ are the number of measurement rounds used to estimate the Bell value $B[\hat{\mathbf{P}}_2]$.

The parties will use the Bell inequality and corresponding violation $B[\hat{\mathbf{P}}_2] - \delta_{\text{est}}$ as a hypothesis in the experiment. From the data of the third set, the parties calculate the Bell value $B[\hat{\mathbf{P}}_3]$. For an honest implementation, the protocol aborts if the Bell value $B[\hat{\mathbf{P}}_3]$ is smaller than $B[\hat{\mathbf{P}}_2] - \delta_{\text{est}}$.

(3) Furthermore, the parties need to estimate the Quantum bit error rate (QBER) Q to bound the error correction information. Alice and Bob publicly reveal the measurement outcomes from $N\eta$ randomly sampled key generation rounds to estimate the QBER. The QBER of the raw key can be upper bounded with high probability using the tail inequality (see Lemma 1 in Appendix A):

$$\Pr\left[Q \geqslant \hat{Q} + \gamma_{\text{est}}\left(N(1-\xi-\eta), N\eta, \hat{Q}, \epsilon_{\text{est}}^{\gamma}\right)\right] > \epsilon_{\text{est}}^{\gamma}, \tag{11}$$

where $\gamma_{\text{est}}(N(1-\xi-\eta), N\eta, \hat{Q}, \epsilon_{\text{est}}^{\gamma})$ is the positive root of the following equation:

$$\ln \binom{N(1-\xi-\eta)\hat{Q} + N(1-\xi-\eta)\gamma_{\text{est}}}{N(1-\xi-\eta)} + \ln \binom{N\eta\hat{Q}}{N\eta} = \ln \binom{(N(1-\xi)\hat{Q} + N(1-\xi-\eta)\gamma_{\text{est}}}{N(1-\xi)} + \ln \epsilon_{\text{est}}^{\gamma}. \tag{12}$$

Thus we can deduce that the QBER Q is not larger than $\hat{Q} + \gamma_{\text{est}}$ (estimated QBER + statistical correction) with very high probability of $1 - \epsilon_{\text{est}}^{\gamma}$.

(4) Alice and Bob use an one-way error correction (EC) protocol to obtain identical raw keys K_A and K_B from their bit strings A^N and B^N. During the process of error correction, Alice communicates O_{EC} (O_{EC} denotes all the classical communication in the error correction step) to Bob such that he can guess the outcomes A^N of Alice. If EC aborts, they abort the protocol. In an honest implementation, this happens with probability at most ϵ_{EC}^{c}. Otherwise, they obtain error-corrected raw keys K_A and K_B [12,38–40]. The probability that Alice and Bob do not abort but hold different raw keys $K_A \neq K_B$ is at most ϵ_{EC}. For details, see Appendix B 1.

When the real QBER Q is greater than $\hat{Q} + \gamma_{\text{est}}$ (which happens with probability $\epsilon^{\gamma}_{\text{est}}$), the hashed values of keys belonging to Alice and Bob (which is sent from Alice to Bob to check if the error correction is successful, see Appendix B 1 for details) are different with high probability [38]. This results in the abortion of the implemented error correction protocol. Thus we can upper bound the error correction abortion probability ϵ^{c}_{EC} by $\epsilon^{\gamma}_{\text{est}}$.

(5) Alice and Bob apply a privacy amplification protocol to obtain a secure final key $\tilde{K}_A = \tilde{K}_B$ of length l that is close to be uniformly random and independent of the adversary's knowledge.

V. SECRET KEY RATE

To provide a lower bound on the device-independent secret key rate, one has to estimate two terms. One is the conditional von Neumann entropy $H(A|X,E)$ and the other one is the error correction information $H(A|B)$ of the raw key [41]. To estimate the latter, one can follow the footsteps of [25,42]; the detailed derivation is shown in Appendix B. For the estimation of the conditional von Neumann entropy $H(A|X,E)$, we lower bound it by the conditional min-entropy $H_{\min}(A|X,E) = -\log_2 P_g(A|X,E)$ [see Eq. (B18)] [43], where $P_g(A|X,E)$ is Eve's guessing probability about Alice's X-measurement results conditioned on her side information E. $P_g(A|X,E)$ can be upper bounded by a function G_x of the estimated Bell violation $B[\mathbf{P}]$ [26] by solving a semidefinite program [44], i.e.,

$$P_g(A|X,E) \leqslant G_x(B[\mathbf{P}]). \tag{13}$$

In real-life experiments, one does not have access to the probabilities. Instead, one has to deal with the frequencies. In Sec. IV we discussed that the protocol will abort if the observed Bell violation $B[\hat{\mathbf{P}}_3]$ in the hypothesis testing is smaller than $B[\hat{\mathbf{P}}_2] - \delta_{\text{est}}$. We need to take into account that the observed Bell violation $B[\hat{\mathbf{P}}_3]$ is calculated from a finite number of rounds. To infer the real Bell violation of the i.i.d. implementation, we make use of Hoeffding's inequality to define a confidence interval δ_{con} and the associated error probability ϵ_{con}. We bound the probability of wrongly accepting the hypothesis with the error probability ϵ_{con} by

$$\begin{aligned}&\Pr(B[\hat{\mathbf{P}}_2] - \delta_{\text{est}} \geqslant B[\hat{\mathbf{P}}_3] + \delta_{\text{con}}) < \epsilon_{\text{con}}\\ &\Rightarrow \Pr(B[\hat{\mathbf{P}}_2] - \delta_{\text{est}} - \delta_{\text{con}} \geqslant B[\hat{\mathbf{P}}_3]) < \epsilon_{\text{con}}.\end{aligned} \tag{14}$$

Therefore given that Alice and Bob do not abort the protocol, we infer that the Bell violation of the system under consideration is higher than $B[\hat{\mathbf{P}}_2] - \delta_{\text{est}} - \delta_{\text{con}}$ (with maximum ϵ_{con} probability of error). We consider the worst possible scenario and use the Bell violation $B[\hat{\mathbf{P}}_2] - \delta_{\text{est}} - \delta_{\text{con}}$ to upper bound the guessing probability $P_g(A|X,E)$ via a semidefinite program

$$\begin{aligned}&\max_{\rho,\{A(a|x)\},\{B(b|y)\}} \quad P_g(A|X,E)\\ &\text{subject to:} \quad \text{Tr}(\rho\mathcal{G}) = B[\hat{\mathbf{P}}_2] - \delta_{\text{est}} - \delta_{\text{con}}.\end{aligned} \tag{15}$$

The guessing probability $P_g(A|X,E)$ is bounded by using the NPA hierarchy [45,46] up to level 2 in the optimization problem of Eq. (15). The optimization is performed using standard tools YALMIP [47], CVX [48–50], NCPOL2SDPA [51], and QETLAB [52]. Here we have used the SDPT3 [53] solver for solving the optimization problem of Eq. (15). One can use SEDUMI [54] or MOSEK [55] as possible alternatives. $\mathcal{G}$ is the Bell operator, defined as

$$\mathcal{G} = \sum_{a,b,x,y} h^{ab}_{A_xB_y} A(a|x)B(b|y).$$

$A(a|x)$ and $B(b|y)$ are measurement operators for Alice and Bob, respectively, and ρ is the state shared between Alice and Bob. Hence the conditional von Neumann entropy $H_{\min}(A|XYE, T=1)$ can be bounded by

$$\begin{aligned}H_{\min}(A|XYE,T=1)_{\rho} &= -\log_2 P(A|X,E)\\ &\geqslant -\log_2 G_x(B[\hat{\mathbf{P}}_2] - \delta_{\text{est}} - \delta_{\text{con}}).\end{aligned} \tag{16}$$

The function G is defined in Eq. (13). $T=1$ specifies the outcomes of the parameter estimation rounds which are used for the estimation of the min-entropy.

To bound the error correction information, we need to estimate the QBER Q, i.e., the probability that Alice's and Bob's measurement outcomes in the key generation rounds differ. In Sec. IV we have discussed that we can upper bound the QBER Q of the raw key with at least $1-\epsilon^{\gamma}_{\text{est}}$ probability by $\hat{Q} + \gamma_{\text{est}}$. In Appendix B we show that we can upper bound the von Neumann entropy $H(A|B)$ [20,38]:

$$H(A|B) \leqslant f(\hat{Q} + \gamma_{\text{est}}), \tag{17}$$

where $f(x) = h(x) + x\log_2(d-1)$. Here d is the number of outcomes per measurement in the Bell scenario [56] and h is the binary entropy function.

Using the bound on the min-entropy [see Eq. (16)] and the QBER [see Eq. (17)], we derive the finite-size secret key rate of a $\epsilon^{s}_{\text{DIQKD}}$-sound, $\epsilon^{c}_{\text{DIQKD}}$-complete (see Definition 6 and Appendix B for details) DIQKD protocol for collective attacks. The statement is as follows [42]: Either the protocol in Sec. IV aborts with probability higher than $1-(\epsilon_{\text{con}}+\epsilon^{c}_{\text{EC}})$ or an $(2\epsilon_{\text{EC}}+\epsilon_s+\epsilon_{\text{PA}})$-correct-and-secret key of length

$$\begin{aligned}l \leqslant N&(-\log_2 G_x(B[\hat{\mathbf{P}}_2] - \delta_{\text{est}} - \delta_{\text{con}})\\ &-(1-\xi-\eta)f(\hat{Q}+\gamma_{\text{est}}) + (\xi+\eta)\log_2 d)\\ &-\sqrt{N}\left(4\log_2(2\sqrt{2^{\log_2 d}+1})\left(\sqrt{\log_2\frac{8}{\epsilon'^2_{\text{EC}}}}+\sqrt{\log_2\frac{2}{\epsilon_s^2}}\right)\right)\\ &-\log_2\left(\frac{8}{\epsilon'^2_{\text{EC}}}+\frac{2}{2-\epsilon'_{\text{EC}}}\right)-\log_2\frac{1}{\epsilon_{\text{EC}}}-2\log_2\frac{1}{2\epsilon_{\text{PA}}},\end{aligned} \tag{18}$$

can be generated where $\epsilon^{c}_{\text{DIQKD}} \leqslant \epsilon_{\text{est}} + \epsilon^{\gamma}_{\text{est}}$ (for an honest implementation) and $\epsilon^{s}_{\text{DIQKD}} \leqslant 2\epsilon_{\text{EC}} + \epsilon_s + \epsilon_{\text{PA}}$. The expression in Eq. (18) is derived in Appendix B. Table I lists all parameters of the DIQKD protocol.

VI. RESULTS

In this section we illustrate the potential and the versatility of our method with examples. We choose $\epsilon^{c}_{\text{DIQKD}} = 10^{-2}$, $\epsilon^{s}_{\text{DIQKD}} = 10^{-5}$, $\epsilon_{\text{EC}} = 10^{-10}$ as DIQKD parameters for all the examples shown in the following section.

TABLE I. Parameters of the DIQKD protocol.

N	Number of measurement rounds in the protocol
ξ	Fraction of parameter estimation rounds for estimating the Bell violation
η	Fraction of measurement rounds for estimating the QBER
ϵ_s	Smoothing parameter
ϵ_{EC}, ϵ'_{EC}	Error probabilities of the error correction protocol
ϵ^c_{EC}	Probability of abortion of error correction protocol
δ_{est}	Width of the statistical interval for the Bell violation hypothesis test
ϵ_{est}	Error probability of the Bell violation hypothesis test
δ_{con}	Confidence interval for the Bell test
ϵ_{con}	Error probability of the Bell violation estimation
γ_{est}	Width of the statistical interval for the QBER estimation
$\epsilon^{\gamma}_{\text{est}}$	Error probability of the QBER estimation
ϵ_{PA}	Error probability of the privacy amplification protocol
$\epsilon^c_{\text{DIQKD}}$	Completeness parameter of the DIQKD protocol
$\epsilon^s_{\text{DIQKD}}$	Soundness parameter of the DIQKD protocol

A. Scenario of m measurements each, two outcomes

We present the scenario with m measurement settings for Alice and $m+1$ for Bob (where the outcomes of only m measurement settings are used in the parameter estimation). Each of those measurement settings has two possible outcomes. Let the shared state between Alice and Bob be a maximally entangled Bell state $|\psi\rangle = \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle)$, mixed with white noise of probability p, i.e.,

$$\rho = (1-p)|\psi\rangle\langle\psi| + p\frac{\mathbb{1}}{4}, \tag{19}$$

with $p \in [0, 1]$. Both parties use σ_z as key generation measurements, resulting in the maximal possible correlation between the outcomes of Alice and Bob.

In the case of $m = 2$, consider the measurement settings of Alice and Bob that maximally violate the Clauser-Horne-Shimony-Holt (CHSH) inequality [57], i.e.,

$$\begin{aligned} x = 1 &\Rightarrow \sigma_z, \qquad x = 2 \Rightarrow \sigma_x, \\ y = 1 &\Rightarrow \frac{\sigma_z + \sigma_x}{\sqrt{2}}, \quad y = 2 \Rightarrow \frac{\sigma_z - \sigma_x}{\sqrt{2}}. \end{aligned} \tag{20}$$

For the CHSH settings with different values of white noise p, we recover the stable hyperplane stated in Table II. The secret key rate as a function of the number of measurement rounds for different values of white noise p is shown in Fig. 2. The hyperplane in Table II is equivalent to the CHSH inequality and consequently the key rate generated by our method coincides with Ref. [26] that uses a predetermined standard CHSH inequality. Though our method finds a hyperplane equivalent to the CHSH inequality, we identify the facet with the maximal violation which is then used in the DIQKD protocol. The other facets (equivalent to CHSH inequality) may admit local hidden variable models which lead to zero key.

In Refs. [30,31], the authors introduced an approach of bounding the device-independent secret key rate (DISKR) directly by using the measurement data. In the asymptotic regime, this corresponds to using a Bell inequality that leads to the maximal DISKR for the precise setup. However, small changes in the parameters (e.g., imperfections on the measurement directions) or on the measured probability distribution may lead to different Bell inequalities corresponding to the optimal secret key rate. We compare our method with Refs. [30,31] in the finite key regime. We study two different Bell scenarios. For the [2,2] scenario, we consider the CHSH settings [see Eq. (20)] and the noisy Bell state of Eq. (19) with $p = 0$ [see graph (a) of Fig. 3] and $p = 0.02$ [see graph (b) of Fig. 3]. For the [3,2] scenario (three measurement settings

TABLE II. Optimized Bell inequality for the measurement settings in Eq. (20), performed on a Bell state. Here the entries of the hyperplane vector, see Eq. (5), are given in a tabular form. For their explicit ordering see Appendix D.

1	−1	1	−1
−1	1	−1	1
1	−1	−1	1
−1	1	1	−1

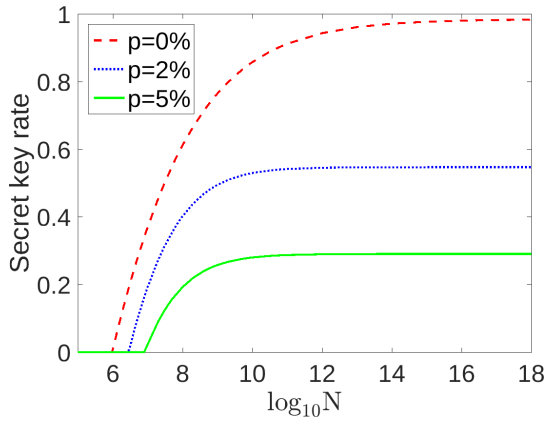

FIG. 2. Secret key rate vs logarithm of the number of rounds N using the measurement settings of Eq. (20). The state shared between two parties is the noisy Bell state [defined in Eq. (19)], where the noise is taken to be $p = 0.0$ (dashed red line), $p = 0.02$ (dotted blue line), $p = 0.05$ (green solid line).

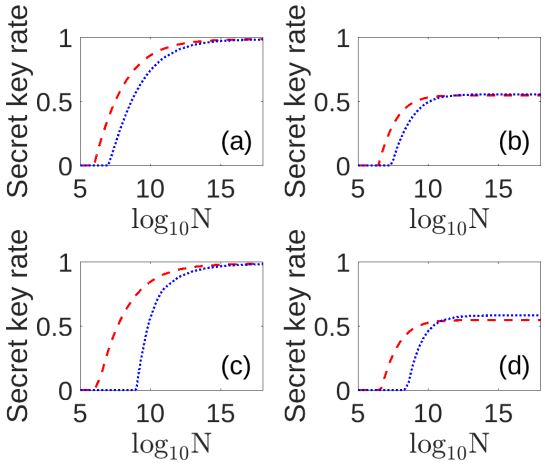

FIG. 3. Achievable secret key rate as a function of the number of measurement rounds N, comparing our method (dashed red line) and the method of Refs. [30,31] (dotted blue line) for a noisy Bell state with noise parameter p, see Eq. (19). Upper row: Measurement settings of Eq. (20) for (a) $p = 0$ and (b) $p = 0.02$. Lower row: Measurement settings of Eq. (21) for (c) $p = 0$ and (d) $p = 0.02$.

each, two outcomes per measurement), we consider the setting

$$\begin{aligned} x = 1 &\Rightarrow \sigma_z, \\ x = 2 &\Rightarrow \sin\frac{\pi}{3}\sigma_x + \cos\frac{\pi}{3}\sigma_z, \\ x = 3 &\Rightarrow \sin\frac{2\pi}{3}\sigma_x + \cos\frac{2\pi}{3}\sigma_z, \\ y = 1 &\Rightarrow \sin\frac{\pi}{6}\sigma_x + \cos\frac{\pi}{6}\sigma_z, \\ y = 2 &\Rightarrow \sigma_x, \\ y = 3 &\Rightarrow \sin\frac{5\pi}{6}\sigma_x + \cos\frac{5\pi}{6}\sigma_z, \end{aligned} \tag{21}$$

and use the noisy Bell state [Eq. (19)] with $p = 0$, $p = 0.02$ [see graphs (c) and (d) of Fig. 3]. To analyze the robustness, we incorporate fluctuations θ in the orientations in some measurement settings of Eq. (21) such that

$$\begin{aligned} x = 1 &\Rightarrow \sigma_z, \\ x = 2 &\Rightarrow \sin\left(\frac{\pi}{3} - \theta\right)\sigma_x + \cos\left(\frac{\pi}{3} - \theta\right)\sigma_z, \\ x = 3 &\Rightarrow \sin\left(\frac{2\pi}{3} + \theta\right)\sigma_x + \cos\left(\frac{2\pi}{3} + \theta\right)\sigma_z, \\ y = 1 &\Rightarrow \sin\left(\frac{\pi}{6} + \theta\right)\sigma_x + \cos\left(\frac{\pi}{6} + \theta\right)\sigma_z, \\ y = 2 &\Rightarrow \sigma_x, \\ y = 3 &\Rightarrow \sin\left(\frac{5\pi}{6} - \theta\right)\sigma_x + \cos\left(\frac{5\pi}{6} - \theta\right)\sigma_z. \end{aligned} \tag{22}$$

We use a noisy Bell state with $p = 0.02$ [see Eq. (19)] as the shared state between Alice and Bob. We use two approaches to compare the robustness of our method with Refs. [30,31]. First, we set $\theta = \frac{\pi}{60}$ [see Eq. (22)] and vary the number N of measurement rounds (see Fig. 4). Next we

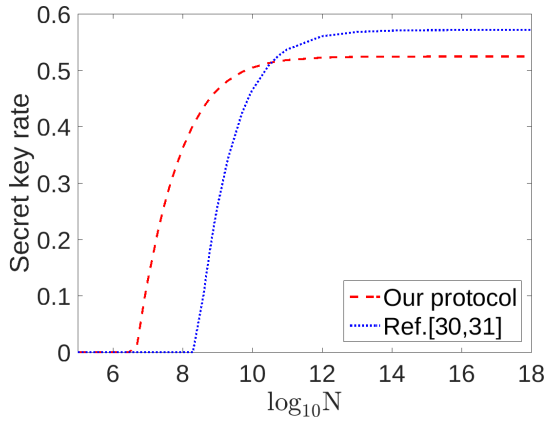

FIG. 4. Deviation of measurement direction. Secret key rate vs logarithm of the number of rounds N for our method (dashed red line) and the method of Refs. [30,31] (dotted blue line), with measurement settings of Eq. (22) where $\theta = \frac{\pi}{60}$, using a noisy Bell state with $p = 0.02$ [see Eq. (19)].

compare the methods for a range of deviations θ for $N = 10^{10}$ measurement rounds (see Fig. 5). We observe that the Bell inequality derived from our approach is stable against small fluctuations of the measurement directions or in the shared state. Our method can also generate a nonzero secret key by performing fewer measurement rounds in comparison with Refs. [30,31] (see Figs. 3–5). This is because the effect of statistical corrections in the Bell inequality violation [see Eq. (10)] is smaller in our approach. A similar behavior is also expected if the number of measurement settings per party is increased. These statistical corrections become insignificant for a high number of measurement rounds, such that the method of Refs. [30,31] yields a higher secret key in the asymptotic regime.

We point out that our method can also have advantages with respect to the CHSH scenario, when the DI secret key rate is calculated via the analytical expression from Ref. [20]: if nonoptimal measurement settings were used, we can in-

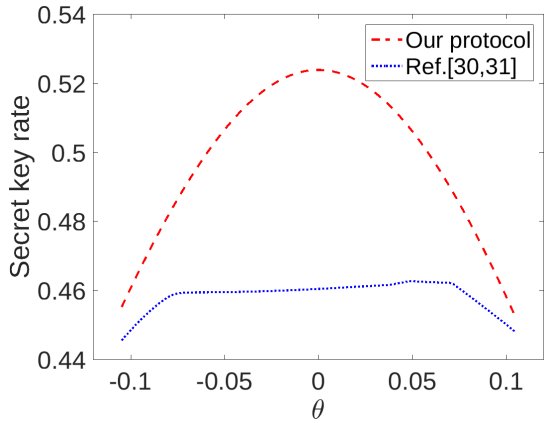

FIG. 5. Deviation of measurement direction. Secret key rate vs deviation θ of the measurement settings in Eq. (22) for our method (dashed red line) and the method of Refs. [30,31] (dotted blue line), with $N = 10^{10}$, using a noisy Bell state with $p = 0.02$ [see Eq. (19)].

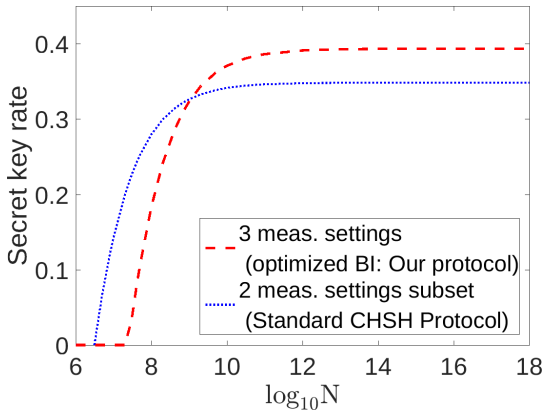

FIG. 6. Improvement for more than two measurement settings. Secret key rate vs $\log N$ for our method with measurement settings as given in Eq. (C1) (dashed red line) compared to the optimal subset of two measurement settings per party (dotted blue line). Here the secret key rate for any subset of two measurement settings per party is calculated via the analytical expression in [20] using the CHSH inequality. The shared state is a Bell state.

crease the key rate by employing additional measurement settings. As an example, we consider the observed probability distribution originating from the maximally entangled Bell state and the set of measurement settings listed explicitly in Appendix C, see Eq. (C1). With our method we can generate a higher secret key rate (for certain N) than using any subset of two measurement settings per party (and the analytical expression of [20]). See Fig. 6 for an illustration.

If the probability distribution obtained by two nonoptimal measurement settings per party does not lead to a nonzero secret key, adding another measurement setting per party and employing our strategy can be advantageous. For example, with nonoptimal measurement settings in Eq. (C2) and the maximally entangled Bell state, one cannot extract a secret key, using our method or blindly using the CHSH inequality. By adding another set of measurements for Alice and Bob, as shown in Eq. (C3), our method leads to a nonzero secret key rate.

B. Scenario of 2 measurements each, d outcomes

In this section we analyze the scenario where each party has two measurement settings in the parameter estimation rounds (Bob has an additional measurement setting which will be used in key generation rounds), and each measurement has d outcomes. The state shared between Alice and Bob is a maximally entangled state of two qudits, i.e., $|\psi\rangle = \sum_{i=0}^{d-1} \frac{1}{\sqrt{d}}|ii\rangle$, which is affected by white noise with probability p, i.e.,

$$\rho = (1-p)|\psi\rangle\langle\psi| + p\frac{\mathbb{1}}{d^2}. \tag{23}$$

We consider the measurement settings from Refs. [58,59]. The measurement is carried out in three steps. In the first step Alice applies a unitary operation on her subsystem with only nonzero terms in the diagonal equal to $e^{\iota\vec{\phi}_x(j)}$, where x denotes Alice's measurement direction, i.e., $x \in \{1,2\}$, and $j = 0,1,2,\ldots,d-1$. Similarly, Bob applies a unitary operation on his subsystem with only nonzero terms in the diagonal equal to $e^{\iota\vec{\varphi}_y(j)}$, where y denotes Bob's measurement direction, i.e., $y \in \{1,2,3\}$. These unitary operations are denoted by $U(\vec{\phi}_x)$ and $U(\vec{\varphi}_y)$ for Alice and Bob, respectively, where

$$\vec{\phi}_x \equiv [\phi_x(0), \phi_x(1), \phi_x(2), \ldots, \phi_x(d-1)],$$
$$\vec{\varphi}_y \equiv [\varphi_y(0), \varphi_y(1), \varphi_y(2), \ldots, \varphi_y(d-1)].$$

The values of these phases are chosen as

$$\begin{aligned} &\phi_1(j) = 0, \qquad \phi_2(j) = \frac{\pi}{d}j, \\ &\varphi_1(j) = \frac{\pi}{2d}j, \quad \varphi_2(j) = -\frac{\pi}{2d}j, \quad \varphi_3(j) = 0, \end{aligned} \tag{24}$$

with $j = 0,1,2,\ldots,d-1$. We use $\{x = 1, y = 3\}$ for the key generation rounds and $\{x \in (1,2), \quad y \in (1,2)\}$ for the parameter estimation rounds. The second step consists of Alice carrying out a discrete Fourier transform U_{FT} and Bob applying U_{FT}^*. The matrix elements of the Fourier transform are defined as $(U_{FT})_{jk} = \exp[(j-1)(k-1)2\pi\iota/d]$, $(U_{FT}^*)_{jk} = \exp[-(j-1)(k-1)2\pi\iota/d]$. Thus the concatenated unitaries for Alice and Bob are $V(\vec{\phi}_x) \equiv U_{FT}\,U(\vec{\phi}_x)$ and $V(\vec{\varphi}_y) \equiv U_{FT}^*\,U(\vec{\varphi}_y)$, respectively.

Finally, Alice and Bob carry out measurements in the computational basis $|i\rangle$. For $d = 3$, we find via linear optimization, see Eq. (4), the optimized Bell inequality as shown in Table III. The details of this representation of the Bell inequality are explained in Table VII of Appendix D.

TABLE III. Optimized Bell inequality for the measurement described in the text, performed on a maximally entangled state of two qutrits. Here the entries of the hyperplane vector, see Eq. (5), are given in a tabular form. For their explicit ordering see Appendix D.

1	−1	0	−1	1	0
0	1	−1	0	−1	1
−1	0	1	1	0	−1
1	0	−1	1	−1	0
−1	1	0	0	1	−1
0	−1	1	−1	0	1

The hyperplane in Table III is equivalent to the CGLMP inequality [59,60]. If the parties share the nonmaximally entangled state

$$|\Phi\rangle \equiv \frac{|00\rangle + 0.7923|11\rangle + |22\rangle}{\sqrt{2 + 0.7923^2}}, \tag{25}$$

the CGLMP inequality is maximally violated, thus resulting in a significantly higher secret key rate, as shown in Fig. 7. This trend of generating a higher secret key rate using nonmaximally entangled states is also observed for higher dimensions (i.e., $d > 3$).

Note that in this scenario with d outcomes the maximum secret key rate is $\log_2 d$. For a fair comparison, we have normalized the min-entropy [i.e., $-\log_2$ of the solution of the optimization problem of Eq. (15)] by division with $\log_2 d$ to get a rate per qubit dimension.

Comparing the DIQKD protocol with measurement settings as described around Eq. (24) for different d and the corresponding d-dimensional maximally entangled state, see

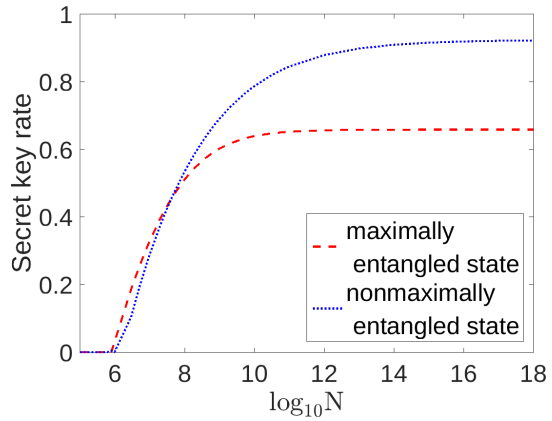

FIG. 7. Secret key rate vs $\log N$ when performing the measurement described around Eq. (24) on a maximally entangled state of two qutrits (dashed red line) and on the nonmaximally entangled state given in Eq. (25) (dotted blue line).

Eq. (23), the minimum number of measurement rounds required to have a nonzero secret key rate decreases slightly with increasing d, see Fig. 8. This follows from the fact that the minimum number of measurement rounds required to have a nonzero Bell violation decreases with increasing d. On the other side, the secret key is decreasing with increasing d (see Fig. 8) when the number of measurement rounds is sufficiently high. The nonlocality of the resultant correlation is decreasing with increasing d, which in turn results in the lower secret key.

C. Random measurement settings

In this section we analyze the case when Alice's and Bob's devices perform random measurements. We specifically focus on the fraction of events that leads to a nonzero secret key rate. First consider the $[m, 2]$ scenario, i.e., m measurement each, with two outcomes. The state shared between the parties is the noisy Bell state as in Eq. (19). We choose the raw key generation measurement operators $\{x = 1 \Rightarrow \sigma_z, y = m + 1 \Rightarrow \sigma_z\}$ in order to achieve correlated outcomes in the key measurement rounds and consequently have to exchange less error correction information. The remaining measurement operators are chosen randomly. Alice and Bob perform general unitary operators

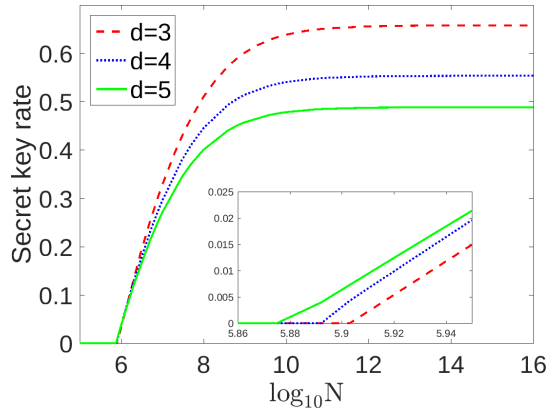

FIG. 8. Secret key rate vs $\log N$ when performing the measurement described around Eq. (24) for $d = 3$ (dashed red line), 4 (dotted blue line), and 5 (solid green line) on a maximally entangled state of two d-dimensional subsystems. The inset graph shows a zoomed-in version in the region of low number of measurement rounds, demonstrating the crossover of the curves.

TABLE IV. Approximate probability of achieving a nonzero secret key rate in the $[m, 2]$ scenario for different white noise levels p in the noisy Bell state [see Eq. (19)]. The statistics are taken over 10^5 realizations. Measurement settings of key generation rounds are fixed to be σ_z for Alice and Bob. The remaining measurement settings are performed in random orientation. For each realization, 10^{12} measurement rounds are used to compute the finite key.

	$(m_a, m_b) = 2$	$(m_a, m_b) = 3$
$p = 0\%$	∼28.6%	∼53.4%
$p = 1\%$	∼18.3%	∼46.5%
$p = 2\%$	∼10.8%	∼36.8%
$p = 3\%$	∼6.4%	∼28.2%
$p = 4\%$	∼3.9%	∼18.5%
$p = 5\%$	∼2.2%	∼11.3%

$$U(\phi, \psi, \chi) = \begin{bmatrix} e^{\iota\psi}\cos\phi & e^{\iota\chi}\sin\phi \\ -e^{-\iota\chi}\sin\phi & e^{-\iota\psi}\cos\phi \end{bmatrix} \tag{26}$$

with parameters $\psi, \chi \in [0, 2\pi]$ and $\phi \in [0, \frac{\pi}{2}]$ and then measure in the computational basis $\{|0\rangle, |1\rangle\}$. This strategy is equivalent to choosing a random measurement. In Table IV we show the fraction of events that leads to a nonzero secret key rate with random measurements. The statistics are based on 10^5 realizations. For the [2,2] scenario, the optimization in Eq. (4) will always lead to the CHSH inequality. Adding another measurement setting per party (i.e., the [3,2] scenario) significantly increases the probability of finding a hyperplane that produces a nonzero secret key rate. The first explanation of this fact is statistical. By increasing the number of settings, we increase the probability that some of them violate a Bell inequality even involving only two settings per party. Apart from that, the optimization in Eq. (4) also provides some hyperplanes for the [3,2] scenario that are independent of the hyperplanes for the [2,2] scenario. From the higher chance of Bell inequality violation, we obtain a higher chance of achieving a nonzero key. This result also reverberates the

TABLE V. Approximate probability of achieving a nonzero secret key rate in the $[2, d]$ scenario for white noise of different probability p added to the maximally entangled state of two qudits [see Eq. (23)]. All other details are as in Table IV.

	$d = 3$	$d = 4$
$p = 0\%$	∼6.4%	∼2.5%
$p = 1\%$	∼2.2%	∼0%
$p = 2\%$	∼0.3%	∼0%

results of the nonlocal volume[2] in [61–66], which increases for the pure bipartite entangled state when more measurement settings for each party are used. We observe the same phenomenon in our case, regarding the secret key rate. As the nonlocal volume shrinks by adding noise, it also reduces the probability of producing a nonzero key rate. Let us now analyze the $[2, d]$ scenario (i.e., d outcomes per measurement) with random measurement settings. The shared state is a noisy maximally entangled state of two qudits [see Eq. (23)]. We compute the approximate probability for achieving a nonzero secret key rate (see Table V). The statistics are based on 10^5 realizations. The measurements for key generation are in the computational basis. The remaining measurement settings are chosen randomly.

We observe that for $d \geqslant 3$, the probability to extract a nonzero secret key is smaller compared to the case with only two outcomes. This follows from the fact that the nonlocal volume shrinks by increasing the dimension of the maximally entangled state. This results in a smaller probability of generating nonlocal correlations and therefore a smaller chance of a Bell inequality violation [65] and smaller probability of a nonzero secret key.

VII. CONCLUSIONS

Several protocols for device-independent quantum key distribution (DIQKD) have the common feature that they rely on the violation of a predetermined Bell inequality. We propose a robust DIQKD procedure where a suitable Bell inequality is instead constructed from the measurement data. This constructed Bell inequality leads to the maximum Bell violation for the particular setup. Then we use the Bell inequality and its corresponding violation to bound the secret key rate via lower bounding the min-entropy.

We provide a finite-size key analysis of our proposed procedure. We bound the statistical fluctuations of the Bell inequality violation by Hoeffding's inequality. However, we do not claim that our choice of concentration inequality [67–69] is optimal for a finite number of measurement rounds. Note that our method could also be implemented for the estimation of global randomness in a device-independent randomness generation protocol.

We have illustrated our method with several examples for different numbers of measurement settings and different numbers of outcomes. Even though our procedure may identify a specific Bell inequality of a known type in some cases, a predefined version of this type of Bell inequalities would often lead to zero key. Our procedure identifies the one Bell inequality (out of all the equivalent ones) with maximal violation, which then leads to a nonzero secret key rate.

We have also shown cases when our method yields a higher secret key rate than using the standard CHSH inequality. In comparison to related approaches (Refs. [30,31]), we provide examples where our approach needs fewer numbers of measurement rounds to generate a nonzero secret key. Using our method, the typical number of measurement rounds to generate a nonzero key varies between 10^6 to 10^8 for the $[m, 2]$ Bell scenario and is of the order 10^6 for the $[2, d]$ Bell scenario. We further showed the performance of our method in the case of random measurement settings. Our method employs the observed measurement statistics, which can be affected by inefficient detectors. In case of no-detection events, one can follow our procedure by declaring no-detection as an additional outcome. One could also account for detector efficiencies by using the approaches of Refs. [70–72].

Finally, future work should address the use of more sophisticated methods of bounding the conditional von Neumann entropy [73–75], which could increase the secret key rate, in comparison to the bounds based on the min-entropy.

ACKNOWLEDGMENTS

The authors acknowledge support from the Federal Ministry of Education and Research (BMBF, Projects Q.Link.X and HQS). We also acknowledge support by the QuantERA project QuICHE via the German Ministry for Education and Research (BMBF Grant No. 16KIS1119K). We thank Gláucia Murta, Federico Grasselli, and Lucas Tendick for helpful discussions.

APPENDIX A: DEFINITIONS

We start with the definition of some quantities that will help us to derive the key rates for the DIQKD protocol.

Definition 1 (Min and max-entropy [76,77]). Let $\rho_{AB} \in \mathcal{P}(\mathcal{H}_A \otimes \mathcal{H}_B)$ and $\sigma_B \in \mathcal{P}(\mathcal{H}_B)$. $\mathcal{P}(\mathcal{H}_B)$ is the set of positive-semidefinite operators on the Hilbert space $\mathcal{H}_B$. The min-entropy of ρ_{AB} conditioned on σ_B is

$$H_{\min}(\rho_{AB}|\sigma_B) := -\log_2 \lambda, \tag{A1}$$

where λ is the minimum real number such that $\lambda.(\mathbb{I} \otimes \sigma_B) - \rho_{AB} \geqslant 0$. The max-entropy of ρ_{AB} conditioned on σ_B is

$$H_{\max}(\rho_{AB}|\sigma_B) := \log_2 \mathrm{Tr}\big((\mathbb{I} \otimes \sigma_B)\rho_{AB}^0\big), \tag{A2}$$

where ρ_{AB}^0 denotes the projector onto the support of ρ_{AB}.

Definition 2 (Smoothed min and max-entropy [76,78]). For a quantum state ρ_{AB} and $\epsilon \geqslant 0$, the smooth min-entropy of system A conditioned on B is defined as

$$H_{\min}^{\epsilon}(A|B) := \max_{\tilde{\rho}_{AB} \in \mathcal{B}^{\epsilon}(\rho_{AB})} H_{\min}(A|B)_{\tilde{\rho}_{AB}}, \tag{A3}$$

and the smooth max-entropy of system A conditioned on B is defined as

$$H_{\max}^{\epsilon}(A|B) := \min_{\tilde{\rho}_{AB} \in \mathcal{B}^{\epsilon}(\rho_{AB})} H_{\max}(A|B)_{\tilde{\rho}_{AB}}. \tag{A4}$$

$\mathcal{B}^{\epsilon}$ is an ϵ-ball of subnormalized operators around the state ρ_{AB} defined in terms of the purified distance.

[2]The nonlocal volume is a statistical measure of nonlocality introduced in [61]. It is defined as the probability that the correlations, generated from randomly chosen projective measurements made on a given state $|\psi\rangle$, violate any Bell inequality (a witness of nonlocality) by any extent. Generally, the nonlocal volume for a given state $|\psi\rangle$ is obtained by $\int d\Omega f(|\psi\rangle, \Omega)$, where one integrates over the measurement parameters Ω [62]. $f(|\psi\rangle, \Omega)$ is an indicator function that takes the value 1 if the resultant correlations, generated from the state and measurements, are nonlocal. Otherwise, it will take the value 0.

Now we focus on the security parameters of quantum key distribution. The security of quantum key distribution can be split into two conditions.

Definition 3 (Correctness [5,25,42]). A DIQKD protocol is ϵ_{corr}-correct if the final key $\tilde{K}_A$ of Alice differs from the final key $\tilde{K}_B$ of Bob with probability less than ϵ_{corr}, i.e.,

$$\Pr(\tilde{K}_A \neq \tilde{K}_B) \leqslant \epsilon_{\text{corr}}. \tag{A5}$$

Definition 4 (Secrecy [5,25,42]). For any $\epsilon_{\text{sec}} \geqslant 0$, a DIQKD protocol is ϵ_{sec} with respect to the adversary E if the joint state satisfies

$$p(\Omega)\tfrac{1}{2}\|\rho_{\tilde{K}_A E|\Omega} - \tau_{\tilde{K}_A} \otimes \rho_E\|_1 \leqslant \epsilon_{\text{sec}}, \tag{A6}$$

where $\tau_{\tilde{K}_A}$ is the maximally mixed state on $\tilde{K}_A$ of the protocol. Here $p(\Omega)$ is the probability of not aborting the protocol.

If a protocol is ϵ_{corr}-correct and ϵ_{sec}-secret, then it is $\epsilon^s_{\text{DIQKD}}$-correct and secret for any $\epsilon^s_{\text{DIQKD}} \geqslant \epsilon_{\text{corr}} + \epsilon_{\text{sec}}$. The correctness (see Def. 3) of the final key is ensured by the error correction step. During error correction, Alice sends a sufficient amount of information to Bob so that he can correct his raw key. If Alice and Bob do not abort in this step, then the probability that they end up with different raw keys is guaranteed to be very small (below ϵ_{EC}). For the secrecy of the protocol (see Def. 4) one needs to estimate how far the final state describing Alice's key and the eavesdropper's system is from the ideal one.

Definition 5 (Secret key rate [25,42]). If a protocol generates a correct and secret key of length l after n rounds, the secret key rate is defined as

$$r = \frac{l}{n}. \tag{A7}$$

Any useful DIQKD protocol should not abort almost all the time. This is apprehended by the concept of completeness.

Definition 6 (Security [25,42]). A DIQKD protocol is $(\epsilon^s_{\text{DIQKD}}, \epsilon^c_{\text{DIQKD}}, l)$-secure if

(1) (Soundness) For any implementation of the protocol, either it aborts with probability greater than $1 - \epsilon^s_{\text{DIQKD}}$ or an $\epsilon^s_{\text{DIQKD}}$-correct and secret key of length l is obtained.

(2) (Completeness) There exists an honest implementation of the protocol such that the probability of not aborting, $p(\Omega)$, is greater than $1 - \epsilon^c_{\text{DIQKD}}$.

In the privacy amplification step, Alice and Bob want to turn their equal string of bits, which may be partially known to an eavesdropper, into a shorter completely secure string of bits. For this step, a 2-universal family of hash functions is needed.

Definition 7 (2-universal hash function). A family of hash functions $\mathcal{F} = \{f : \{0,1\}^n \to \{0,1\}^\ell\}$ is called 2-universal if for every two strings $x, x' \in \{0,1\}^n$ with $x \neq x'$ then

$$\Pr_{f\in\mathcal{F}}(f(x) = f(x')) = \frac{1}{2^\ell}, \tag{A8}$$

where f is chosen uniformly at random in $\mathcal{F}$. The property of 2-universality ensures a good distribution of the outputs. For $\ell \leqslant n$ there always exists a 2-universal family of hash functions [79].

Now we will state the quantum leftover hashing lemma [77,80]. It quantifies the secrecy of a protocol as a function of a conditional entropy of the state before privacy amplification and the length of the final key.

Theorem 1 (Leftover hashing lemma with smooth min-entropy [25,42,80]). Let $\rho_{A^n E}$ be a classical quantum state. Let $\mathcal{F}$ be a 2-universal family of hash functions, from $\{0,1\}^n$ to $\{0,1\}^l$, that maps the classical n-bit string A^n into K_A. Then

$$\|\rho_{K_A F E} - \tau_{K_A} \otimes \rho_{F E}\| \leqslant 2^{-\frac{1}{2}(H^\epsilon_{\min}(A^n|E)_\rho - l)} + 2\epsilon,$$

where F is a classical register that stores the hash function f.

With the leftover hash lemma and the definition of secrecy (see Def. 4), we express the length of a secure key as a function of the entropy of Alice's raw key conditioned on Eve's information before privacy amplification.

Theorem 2 (Key length [25,42]). Let $P(\Omega)$ be the probability that the DIQKD protocol does not abort for a particular implementation. If the length of the key generated after privacy amplification is given by

$$l \leqslant H^{\epsilon_s/P(\Omega)}_{\min}(A^n|E)_{\rho_{|\Omega}} - 2\log\frac{1}{2\epsilon_{\text{PA}}},$$

then the DIQKD protocol is $\epsilon_{\text{PA}} + \epsilon_s$ secret.

In this paper we have considered the II D scenario (collective attacks). In the assumption of collective attacks, the distributed state and the behavior of Alice's and Bob's devices are the same in every round of the protocol. Eve can carry out arbitrary operations in her quantum side information. This assumption implies that after n rounds of the protocol, the state shared by Alice, Bob, and Eve is $\rho_{A^n B^n E} = \rho^{\otimes n}_{ABE}$. The quantum asymptotic equipartition property [76,81] allows us to bound the conditional smooth min-entropy of state $\rho^{\otimes}_{AE}$ by the conditional von Neumann entropy of the state ρ_{AE}.

Theorem 3 (Asymptotic equipartition property [81]). Let $\rho = \rho^{\otimes n}_{AE}$ be an IID state. Then for $n \geqslant \frac{8}{5}\log\frac{2}{\epsilon^2}$,

$$H^\epsilon_{\min}(A^n|E^n)_{\rho^{\otimes n}_{AE}} > nH(A|E)_{\rho_{AE}} - \sqrt{n}\delta(\epsilon,\chi),$$

and similarly,

$$H^\epsilon_{\max}(A^n|E^n)_{\rho^{\otimes n}_{AE}} < nH(A|E)_{\rho_{AE}} + \sqrt{n}\delta(\epsilon,\chi),$$

where $\delta(\epsilon,\chi) = 4\log(\chi)\sqrt{\log\frac{2}{\epsilon^2}}$ and $\chi = \sqrt{2^{-H_{\min}(A|E)_{\rho_{AE}}}} + \sqrt{2^{H_{\max}(A|E)_{\rho_{AE}}}} + 1$.

Lemma 1 [82,83]. Let $\mathcal{X}_{n+k}$ be a random binary string of $n+k$ bits, $\mathcal{X}_k$ be a random sample (without replacement) of m entries from the string $\mathcal{X}_{n+k}$, and $\mathcal{X}_n$ be the remaining bit string. Λ_k and Λ_n are the frequencies of bit value 1 in string $\mathcal{X}_k$ and $\mathcal{X}_n$, respectively. For any $\varepsilon_1 > 0$, it holds the upper tail inequality:

$$\Pr[\Lambda_n \geqslant \Lambda_k + \gamma_1(n,k,\Lambda_k,\varepsilon_1)] > \varepsilon_1, \tag{A9}$$

where $\gamma_1(a,b,c,d)$ is the positive root of

$$\ln\binom{bc}{b} + \ln\binom{ac + a\gamma_1(a,b,c,d)}{a} = \ln\binom{(a+b)c + a\gamma_1(a,b,c,d)}{a+b} + \ln d.$$

For $\varepsilon_2 > 0$, we have the lower tail inequality:

$$\Pr[\Lambda_n \leqslant \Lambda_k - \gamma_2(n,k,\Lambda_k,\varepsilon_2)] > \varepsilon_2, \tag{A10}$$

where $\gamma_2(a,b,c,d)$ is the positive root of

$$\ln\binom{bc}{b}+\ln\binom{ac-a\gamma_2(a,b,c,d)}{a}=\ln\binom{(a+b)c-a\gamma_2(a,b,c,d)}{a+b}+\ln d.$$

Lemma 2 [37]. Let $X_1, X_2, \ldots, X_n$ be independent random variables strictly bounded by the intervals $[a_i, b_i]$, i.e., $a_i \leqslant X_i \leqslant b_i$. We define

$$\bar{X}=\frac{1}{n}(X_1+X_2+\cdots+X_n).$$

Then Hoeffding's inequality reads

$$\Pr(\bar{X}-E[\bar{X}]\geqslant t)\leqslant \exp\left(-\frac{2n^2t^2}{\sum_{i=1}^n(b_i-a_i)^2}\right).$$

Let $c_i := b_i - a_i$ and $c_i \leqslant C\ \forall\ i$. Then Hoeffding's inequality reads

$$\Pr(\bar{X}-E[\bar{X}]\geqslant t)\leqslant \exp\left(-\frac{2n^2t^2}{nC^2}\right)=\exp\left(-\frac{2nt^2}{C^2}\right).$$

APPENDIX B: SECRET KEY ANALYSIS

Theorem 4 (Completeness). The DIQKD protocol stated in Sec. IV is $\epsilon_{\text{est}}+\epsilon_{\text{est}}^{\gamma}$ complete.

Proof. The protocol can abort in two instances. Either it will abort if the error correction failed or if the estimated Bell violation $B[\hat{\mathbf{P}}_3]$ is not high enough. The probability that the error correction fails can only happen if the real QBER Q is larger than $\hat{Q}+\gamma_{\text{est}}$, which happens with probability $\epsilon_{\text{est}}^{\gamma}$, see Sec. IV for details. The protocol also aborts if the estimated Bell violation $B[\hat{\mathbf{P}}_3]$ is smaller $B[\hat{\mathbf{P}}_2]-\delta_{\text{est}})$, see Sec. IV for details. Thus, considering an honest implementation consisting of IID rounds, we can bound the probability of abortion of the protocol:

$$\begin{aligned}\text{p(abort)}&=\text{p((EC aborts) or (Bell test fails))}\\&\leqslant \text{p(EC aborts)}+\text{p(Bell test fails)}\\&\leqslant \text{p(QBER test fails)}+\text{p(Bell test fails)}\\&=p(Q>\hat{Q}+\gamma_{\text{est}})+p(B[\hat{\mathbf{P}}_3]<B[\hat{\mathbf{P}}_2]-\delta_{\text{est}})\\&=\epsilon_{\text{est}}^{\gamma}+\epsilon_{\text{est}},\end{aligned}\tag{B1}$$

where ϵ_{est} is defined in Eq. (10), and $\epsilon_{\text{est}}^{\gamma}$ is defined in Eq. (11). Thus we get $\epsilon_{\text{DIQKD}}^{c}\leqslant\epsilon_{\text{est}}+\epsilon_{\text{est}}^{\gamma}$.

For the **soundness**, we have to evaluate the correctness and secrecy, defined in Def. 3 and Def. 4, respectively. In case of correctness, if we have an error correction protocol that does not abort, then Alice (Bob) will have the raw key K_A (K_B) after the protocol. The string K_B differs from K_A with probability less than ϵ_{EC}, and as the final keys $\tilde{K}_A$ and $\tilde{K}_B$ are equal if the raw keys are equal, it follows that

$$P(\tilde{K}_A\neq\tilde{K}_B)\leqslant P(K_A\neq K_B)\leqslant\epsilon_{\text{EC}}.$$

For secrecy, let us recall that Ω is defined as the event when the protocol does not abort. This happens when the error correction protocol does not abort and achieved the required Bell violation according to Alice's and Bob's outputs (and inputs). Now define the event $\hat{\Omega}$ as the event Ω (protocol not aborting) and the error correction being successful, i.e., $K_A = K_B$. Thus

$$\begin{aligned}\|\rho_{\tilde{K}_AE_{|\Omega}}-\tau_{\tilde{K}_A}\otimes\rho_E\|_1&\leqslant\|\rho_{\tilde{K}_AE_{|\Omega}}-\rho_{\tilde{K}_AE_{|\hat{\Omega}}}\|_1\\&+\|\rho_{\tilde{K}_AE_{|\hat{\Omega}}}-\tau_{\tilde{K}_A}\otimes\rho_E\|_1\\&\leqslant\epsilon_{\text{EC}}+\|\rho_{\tilde{K}_AE_{|\hat{\Omega}}}-\tau_{\tilde{K}_A}\otimes\rho_E\|_1.\end{aligned}\tag{B2}$$

The first inequality follows from the triangular inequality of the trace distance [84]. $\rho_{\tilde{K}_AE_{|\Omega}}$ is the joint classical quantum state of Alice and Eve if the protocol does not abort. $\rho_{\tilde{K}_AE_{|\hat{\Omega}}}$ is the joint classical quantum state of Alice and Eve if the protocol does not abort and the error correction is successful. When error correction succeeds, the probability of $K_A = K_B$ is higher than $(1-\epsilon_{\text{EC}})$. Conversely, the probability $K_A \neq K_B$ is less than ϵ_{EC}. Thus the second inequality of Eq. (B2) comes from

$$\begin{aligned}\|\rho_{\tilde{K}_AE_{|\Omega}}-\rho_{\tilde{K}_AE_{|\hat{\Omega}}}\|_1&\leqslant(1-\epsilon_{\text{EC}})\|\rho_{\tilde{K}_AE_{|\hat{\Omega}}}-\rho_{\tilde{K}_AE_{|\hat{\Omega}}}\|_1\\&+\epsilon_{\text{EC}}\|\rho_{\tilde{K}_AE_{|\hat{\Omega}}}-\rho_{\tilde{K}_AE_{|\hat{\Omega}^c}}\|_1\leqslant\epsilon_{\text{EC}},\end{aligned}\tag{B3}$$

where $\hat{\Omega}^c$ is defined as the event when the protocol does not abort but error correction is not successful, i.e., $K_A \neq K_B$.

Now we proceed to evaluate the term $\|\rho_{\tilde{K}_AE_{|\hat{\Omega}}}-\tau_{\tilde{K}_A}\otimes\rho_E\|_1$ of Eq. (B2). We will follow the path shown in [25,42]. Given that the protocol did not abort, the maximal length of a secure key is determined by the smooth min-entropy of Alice's raw key conditioned on all information available to the eavesdropper (see the leftover hashing lemma in Theorem. 1). In our protocol (see Sec. IV), it is given by $H_{\min}^{\epsilon_s}(A^N|X^NY^NT^NEO_{\text{EC}})_{\rho_{|\hat{\Omega}}}$. Here we recall that O_{EC} is the information exchanged by Alice and Bob during the error correction protocol. X^N and Y^N are the input bit strings (measurement settings) for Alice and Bob, respectively. A^N is the output bit string of Alice. T^N is the shared random key that determines whether the round is a test or a key generation round. $\hat{\Omega}$ is the event that the protocol does not abort and error correction succeeds.

In order to bypass the conditioned state of $H_{\min}^{\epsilon_s}(A^N|X^NY^NT^NEO_{\text{EC}})_{\rho_{|\hat{\Omega}}}$, we can start from the definition of secrecy (see Def. 4). Then we have to bound the term

$$p(\Omega)\|\rho_{\tilde{K}_AFE_{|\Omega}}-\tau_{\tilde{K}_A}\otimes\rho_{FE}\|_1=\|\rho_{\tilde{K}_AFE\wedge\Omega}-\tau_{\tilde{K}_A}\otimes\rho_{FE\wedge\Omega}\|_1,\tag{B4}$$

where $\rho_{\tilde{K}_AFE\wedge\Omega}=p(\Omega)\rho_{\tilde{K}_AFE_{|\Omega}}$ is a subnormalized state. Here we recall that F is the classical register that stores the hash function f (see Def. 7).

Now using the leftover hashing lemma in Theorem 1, we can generate an $(\epsilon_s+\epsilon_{\text{PA}})$-secret key of length [42]

$$l\leqslant H_{\min}^{\epsilon_s}(A^N|E)_{\rho\wedge\Omega}-2\log\frac{1}{2\epsilon_{\text{PA}}}.\tag{B5}$$

In Ref. [85] it is proved that

$$H_{\min}^{\epsilon_s}(A^N|E)_{\rho\wedge\Omega} \geqslant H_{\min}^{\epsilon_s}(A^N|E)_{\rho}. \tag{B6}$$

Thus we proceed to estimate the quantity $H_{\min}^{\epsilon_s}(A^N|X^NY^NT^NEO_{\mathrm{EC}})_\rho$ in order to bound the achievable secret key of length l.

Using the chain rule relation for the smooth min-entropy conditioned on classical information [76], we can write

$$H_{\min}^{\epsilon_s}(A^N|X^NY^NT^NEO_{\mathrm{EC}})_\rho = H_{\min}^{\epsilon_s}(A^N|X^NY^NT^NE)_\rho - \mathrm{leak}_{\mathrm{EC}}. \tag{B7}$$

Thus, in order to bound $H_{\min}^{\epsilon_s}(A^N|X^NY^NT^NEO_{\mathrm{EC}})_\rho$, we have to lower bound $H_{\min}^{\epsilon_s}(A^N|X^NY^NT^NE)_\rho$ and upper bound $\mathrm{leak}_{\mathrm{EC}}$ (the leakage due to the error correction).

1. Estimation of $\mathrm{leak}_{\mathrm{EC}}$

Alice and Bob perform an EC procedure so that Bob can compute a guess of Alice's raw key A^N. In order to verify if EC is successful, Alice chooses a two-universal hash function (uniformly at random) from the family of hash functions and computes a hash of length $\log(\frac{1}{\epsilon_{\mathrm{EC}}})$ from her raw keys A^N. Then she sends the chosen hash function and the hashed value of her bits to Bob via a public channel. We denote all the classical communication (information leaked during EC, hash function, and the hashed value for verification) by O_{EC}. Bob computes the hash function on his key. If the hashed values are equal, then Alice's and Bob's keys are the same with high probability. If the hashed values are different, the parties will abort the protocol. During this whole process the amount of information about the key exposing to the adversary Eve is termed as $\mathrm{leak}_{\mathrm{EC}}$. In Ref. [77] the $\mathrm{leak}_{\mathrm{EC}}$ is bounded by

$$\mathrm{leak}_{\mathrm{EC}} \leqslant H_0^{\epsilon'_{\mathrm{EC}}}(A^N|B^NX^NY^NT^N) + \log\frac{1}{\epsilon_{\mathrm{EC}}}, \tag{B8}$$

where $\epsilon_{\mathrm{EC}}^c = \epsilon_{\mathrm{EC}} + \epsilon'_{\mathrm{EC}}$ (see Table I). H_0 is the Rényi entropy introduced in Ref. [77]. In Ref. [76], it is denoted as $\bar{H}_0^{\uparrow}$. If Alice and Bob do not abort, then their resultant bit string is identical ($K_A = K_B$) with at least $1-\epsilon_{\mathrm{EC}}$ probability. We can bound the entropy $H_0^{\epsilon'_{\mathrm{EC}}}(A^N|B^NX^NY^NT^N)$ in the following way:

$$\begin{aligned}
&H_0^{\epsilon'_{\mathrm{EC}}}(A^N|B^NX^NY^NT^N)\\
&\leqslant H_{\max}^{\frac{\epsilon'_{\mathrm{EC}}}{2}}(A^N|B^NX^NY^NT^N) + \log\left(\frac{8}{\epsilon'^2_{\mathrm{EC}}} + \frac{2}{2-\epsilon'_{\mathrm{EC}}}\right)\\
&\leqslant NH(A|BXYT) + 4\sqrt{N}\log(2\sqrt{2^{\log_2 d}}+1)\sqrt{\log\frac{8}{{\epsilon'_{\mathrm{EC}}}^2}}\\
&\quad + \log\left(\frac{8}{\epsilon'^2_{\mathrm{EC}}} + \frac{2}{2-\epsilon'_{\mathrm{EC}}}\right).
\end{aligned} \tag{B9}$$

For the definition of $H_0^{\epsilon}(A|B)$, see Ref. [77]. The first inequality of Eq. (B9) comes from Ref. [80] and Eq. (B11) of Ref. [42]. The last inequality comes from the asymptotic equipartition property (see Theorem 3), where we used the relations

$$\begin{aligned}
\delta(\epsilon,\chi) &= 4\log(\chi)\sqrt{\log\frac{2}{\epsilon^2}}\\
&\leqslant 4\log(2\sqrt{2^{\log_2 d}}+1)\sqrt{\log\left(\frac{2}{\epsilon^2}\right)}.
\end{aligned} \tag{B10}$$

Here we have used $\chi \leqslant 2\sqrt{2^{\log_2 d}}+1$, which comes from

$$\begin{aligned}
\chi &= \sqrt{2^{-H_{\min}(A|E)_{\rho_{AE}}}} + \sqrt{2^{H_{\max}(A|E)_{\rho_{AE}}}} + 1\\
&\leqslant 2\sqrt{2^{H_{\max}(A|XYTE)_\rho}} + 1\\
&\leqslant 2\sqrt{2^{\log_2 d}} + 1.
\end{aligned} \tag{B11}$$

The first inequality of Eq. (B11) follows from the fact that A is a classical register and therefore has positive conditional min-entropy, which implies $-H_{\min}(A|XYTE) \leqslant H_{\min}(A|XYTE) \leqslant H_{\max}(A|XYTE)$. For the second inequality of Eq. (B11), we use $H_{\max}(A|XYTE) \leqslant \log_2 d$.

Therefore, from Eqs. (B8) and (B9), we can bound the leakage in the following way:

$$\begin{aligned}
\mathrm{leak}_{\mathrm{EC}} &\leqslant NH(A|BXYT)\\
&\quad + \sqrt{n}(4\log(2\sqrt{2^{\log_2 d}}+1))\sqrt{\log\frac{8}{{\epsilon'}^2_{\mathrm{EC}}}}\\
&\quad + \log\left(\frac{8}{\epsilon'^2_{\mathrm{EC}}} + \frac{2}{2-\epsilon'_{\mathrm{EC}}}\right) + \log\frac{1}{\epsilon_{\mathrm{EC}}}.
\end{aligned} \tag{B12}$$

Now we bound the single-round von Neumann entropy $H(A|BXYT)$ as

$$\begin{aligned}
H(A|BXYT) &= p(T=0)H(A|BXYT=0)\\
&\quad + p(T=1)H(A|BXYT=1)\\
&\leqslant (1-\xi)H(A|BXYT=0) + \xi\log_2 d\\
&\leqslant (1-\xi-\eta)H(A|BXYT=0) + (\xi+\eta)\log_2 d.
\end{aligned} \tag{B13}$$

See Table I for the details of ξ, η, and γ_{est}. For the first equality, we have used that for the conditional von Neumann entropy it holds that $H(A|BX)_\rho = \sum_x p(X=x)H(A|BX=x)$. We divide the measurement rounds into key generation (specified by $T=0$) and parameter estimation (specified by $T=1$), for details see Sec. IV. The first inequality comes from the fact that parameter estimation round's measurements were publicly communicated to estimate the Bell inequality and the corresponding violation. η rounds of the raw key generation measurement were communicated through a public channel to estimate the QBER, which leads to the last inequality.

Now our goal is to estimate $H(A|BXYT=0)$. For dichotomic observables and uniform marginals, $H(A|B)$ can be expressed as $h(Q)$ [20], where h is the binary entropy function, $h(p) := -p\log_2 p - (1-p)\log_2(1-p)$. Similarly for the $[(m_a, m_b), d]$ Bell scenario, $H(A|B)$ can be expressed as a function of the QBER, $H(A|B) = -Q\log_2 Q - (1-Q)\log_2(1-Q) + Q\log_2(d-1)$ [56].

For our specific protocol (see Sec. IV), we bound $H(A|BXYT=0)$ by a function of $\hat{Q}_1 + \gamma_{\mathrm{est}}$ (observed QBER

TABLE VI. Bell inequality table for the [2,2] scenario.

$h^{11}_{A_1B_1}$	$h^{12}_{A_1B_1}$	$h^{11}_{A_1B_2}$	$h^{12}_{A_1B_2}$
$h^{21}_{A_1B_1}$	$h^{22}_{A_1B_1}$	$h^{21}_{A_1B_2}$	$h^{22}_{A_1B_2}$
$h^{11}_{A_2B_1}$	$h^{12}_{A_2B_1}$	$h^{11}_{A_2B_2}$	$h^{12}_{A_2B_2}$
$h^{21}_{A_2B_1}$	$h^{22}_{A_2B_1}$	$h^{21}_{A_2B_2}$	$h^{22}_{A_2B_2}$

+ estimated statistical error), see Sec. V for details:

$$H(A|BXY, T=0) \leqslant f(\hat{Q}+\gamma_{\text{est}}), \tag{B14}$$

where $f(x) = h(x) + x\log_2(d-1)$ (d is the number of outcomes per measurement in the Bell scenario) and h is the binary entropy function. From Eqs. (B13) and (B14), it then follows that

$$H(A|BXYT) \leqslant (1-\xi-\eta)f(\hat{Q}+\gamma_{\text{est}}) + (\xi+\eta)\log_2 d. \tag{B15}$$

The leakage due to error correction is given by [from Eqs. (B12) and (B15)]

$$\begin{aligned}\text{leak}_{\text{EC}} \leqslant{}& N[(1-\xi-\eta)f(\hat{Q}+\gamma_{\text{est}}) + (\xi+\eta)\log_2 d] \\ &+ \sqrt{N}\left(4\log(2\sqrt{2^{\log_2 d}}+1)\right)\sqrt{\log\frac{8}{\epsilon'^2_{\text{EC}}}} \\ &+ \log\left(\frac{8}{\epsilon'^2_{\text{EC}}} + \frac{2}{2-\epsilon'_{\text{EC}}}\right) + \log\frac{1}{\epsilon_{\text{EC}}}.\end{aligned} \tag{B16}$$

2. Estimation of min-entropy $H^{\epsilon_s}_{\min}(A^N|X^NY^NT^NE)_\rho$

Finally, we lower bound $H^{\epsilon_s}_{\min}(A^N|X^NY^NT^NE)_\rho$. We use the asymptotic equipartition property (see Theorem 3) to lower bound the min-entropy of N rounds by the von Neumann entropy of single rounds:

$$\begin{aligned}&H^{\epsilon_s}_{\min}(A^N|X^NY^NT^NE)_\rho \\ &\geqslant NH(A|XYTE)_\rho - 4\sqrt{N}\log(2\sqrt{2^{\log_2 d}}+1)\sqrt{\log\frac{2}{\epsilon_s^2}}.\end{aligned} \tag{B17}$$

Since Alice's actions (and her device's) are independent of Bob's choice of input, adding information about Y (Bob's input) does not increase (or decrease) the conditional von Neumann entropy $H(A|X,E)_\rho$. Since $H(A|X,E)_\rho$ and $H(A|XYE, T=1)_\rho$ are equivalent in our setup, we will use both terms interchangeably. In the general scenario, the conditional von Neumann entropy is hard to calculate analytically. But the conditional von Neumann entropy can be lower bounded by the conditional min-entropy as

$$H(A|XYT,E)_\rho \geqslant H_{\min}(A|XYT,E)_\rho. \tag{B18}$$

The advantage of looking at the conditional min-entropy is that we can express it as $H_{\min}(A|XYE, T=1)_\rho = -\log_2 P_g(A|X,E)$ [43], where $P_g(A|X,E)$ is Eve's guessing probability about Alice's X-measurement results A conditioned on her side information E. $P_g(A|X,E)$ can be upper bounded by a function G_x of the expected Bell violation $B[\mathbf{P}]$ [26] by solving a semidefinite program [44], i.e., $P_g(A|X,E) \leqslant G_x(B[\mathbf{P}])$. For our specific protocol (see Sec. IV), we will lower bound the min-entropy (via upper bounding the guessing probability $P_g(A|X,E)$ using the Bell inequality B and corresponding Bell value $B[\hat{\mathbf{P}}_2] - \delta_{\text{est}} - \delta_{\text{con}}$ (explained in Sec. V):

$$H_{\min}(A|XYE, T=1)_\rho \geqslant -\log_2 G_x(B[\hat{\mathbf{P}}_2] - \delta_{\text{est}} - \delta_{\text{con}}). \tag{B19}$$

Finally, putting Eqs. (B16) and (B19) together, we have either the protocol mentioned in Sec. IV aborts with probability higher than $1-(\epsilon_{\text{con}}+\epsilon^c_{\text{EC}})$ or a $(2\epsilon_{\text{EC}}+\epsilon_s+\epsilon_{\text{PA}})$-correct and secret key can be generated of length l. The length l is bounded by

$$\begin{aligned}l \leqslant{}& N[-\log_2 G_x(B[\hat{\mathbf{P}}_2] - \delta_{\text{est}} - \delta_{\text{con}}) \\ &- (1-\xi-\eta)f(\hat{Q}+\gamma_{\text{est}}) - (\xi+\eta)\log_2 d] \\ &- \sqrt{N}\left(4\log(2\sqrt{2^{\log_2 d}}+1)\left(\sqrt{\log\frac{8}{\epsilon'^2_{\text{EC}}}} + \sqrt{\log\frac{2}{\epsilon_s^2}}\right)\right) \\ &- \log\left(\frac{8}{\epsilon'^2_{\text{EC}}} + \frac{2}{2-\epsilon'_{\text{EC}}}\right) - \log\frac{1}{\epsilon_{\text{EC}}} - 2\log\frac{1}{2\epsilon_{\text{PA}}}.\end{aligned} \tag{B20}$$

TABLE VII. Bell inequality table for the [2,3] scenario.

$h^{11}_{A_1B_1}$	$h^{12}_{A_1B_1}$	$h^{13}_{A_1B_1}$	$h^{11}_{A_1B_2}$	$h^{12}_{A_1B_2}$	$h^{13}_{A_1B_2}$
$h^{21}_{A_1B_1}$	$h^{22}_{A_1B_1}$	$h^{23}_{A_1B_1}$	$h^{21}_{A_1B_2}$	$h^{22}_{A_1B_2}$	$h^{23}_{A_1B_2}$
$h^{31}_{A_1B_1}$	$h^{32}_{A_1B_1}$	$h^{33}_{A_1B_1}$	$h^{31}_{A_1B_2}$	$h^{32}_{A_1B_2}$	$h^{33}_{A_1B_2}$
$h^{11}_{A_2B_1}$	$h^{12}_{A_2B_1}$	$h^{13}_{A_2B_1}$	$h^{11}_{A_2B_2}$	$h^{12}_{A_2B_2}$	$h^{13}_{A_2B_2}$
$h^{21}_{A_2B_1}$	$h^{22}_{A_2B_1}$	$h^{23}_{A_2B_1}$	$h^{21}_{A_2B_2}$	$h^{22}_{A_2B_2}$	$h^{23}_{A_2B_2}$
$h^{31}_{A_2B_1}$	$h^{32}_{A_2B_1}$	$h^{33}_{A_2B_1}$	$h^{31}_{A_2B_2}$	$h^{32}_{A_2B_2}$	$h^{33}_{A_2B_2}$

TABLE VIII. Bell inequality table for the $[m, d]$ scenario.

$h^{11}_{A_1B_1}$	$\cdots$	$h^{1d}_{A_1B_1}$		$\cdots$		$h^{11}_{A_1B_m}$	$\cdots$	$h^{1d}_{A_1B_m}$
$\vdots$	$\ddots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$
$h^{d1}_{A_1B_1}$	$\cdots$	$h^{dd}_{A_1B_1}$		$\cdots$		$h^{d1}_{A_1B_m}$	$\cdots$	$h^{dd}_{A_1B_m}$
	$\cdots$			$\cdots$			$\cdots$	
$\vdots$	$\ddots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$
	$\cdots$			$\cdots$			$\cdots$	
$h^{11}_{A_mB_1}$	$\cdots$	$h^{1d}_{A_mB_1}$		$\cdots$		$h^{11}_{A_mB_m}$	$\cdots$	$h^{1d}_{A_mB_m}$
$\vdots$	$\ddots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$
$h^{d1}_{A_mB_1}$	$\cdots$	$h^{dd}_{A_mB_1}$		$\cdots$		$h^{d1}_{A_mB_m}$	$\cdots$	$h^{dd}_{A_mB_m}$

APPENDIX C: MEASUREMENT SETTINGS

Here we list the explicit measurement settings employed in Sec. VI A.

$$
\begin{aligned}
&x = 1 \Rightarrow \sigma_z, && && y = 1 \Rightarrow \begin{bmatrix} 0.7064 & -0.6632 + 0.2473i \\ -0.6632 - 0.2473i & -0.7064 \end{bmatrix},\\
&x = 2 \Rightarrow \begin{bmatrix} -0.1817 & 0.1307 + 0.9746i \\ 0.1307 - 0.9746i & 0.1817 \end{bmatrix}, && && y = 2 \Rightarrow \begin{bmatrix} -0.6882 & -0.2128 - 0.6936i \\ -0.2128 + 0.6936i & 0.6882 \end{bmatrix},\\
&x = 3 \Rightarrow \begin{bmatrix} -0.7746 & 0.6186 - 0.1315i \\ 0.6186 + 0.1315i & 0.7746 \end{bmatrix}, && && y = 3 \Rightarrow \begin{bmatrix} 0.4046 & -0.1960 + 0.8932i \\ -0.1960 - 0.8932i & -0.4046 \end{bmatrix}.
\end{aligned} \tag{C1}
$$

Using the following set of measurement settings for Alice and Bob in Eq. (C1), one can generate a higher secret key rate employing our method than using any subset of two measurement settings per party using the standard CHSH inequality.

$$
\begin{aligned}
&x = 1 \Rightarrow \sigma_z, && && y = 1 \Rightarrow \begin{bmatrix} -0.4091 & -0.5937 + 0.6930i \\ -0.5937 - 0.6930i & 0.4091 \end{bmatrix},\\
&x = 2 \Rightarrow \begin{bmatrix} 0.7019 & 0.5167 - 0.4903i \\ 0.5167 + 0.4903i & -0.7019 \end{bmatrix}, && && y = 2 \Rightarrow \begin{bmatrix} -0.6133 & -0.2514 + 0.7488i \\ -0.2514 - 0.7488i & 0.6133 \end{bmatrix}.
\end{aligned} \tag{C2}
$$

Using the following measurement settings in Eq. (C2) and the state in Eq. (19) with no white noise, one cannot extract a secret key using our method or blindly using the CHSH inequality.

$$
x = 3 \Rightarrow \begin{bmatrix} -0.1457 & -0.9777 + 0.1513i \\ -0.9777 - 0.1513i & 0.1457 \end{bmatrix}, \quad y = 3 \Rightarrow \begin{bmatrix} -0.9020 & -0.3795 - 0.2056i \\ -0.3795 + 0.2056i & 0.9020 \end{bmatrix}. \tag{C3}
$$

However, by adding another set of measurements for Alice and Bob mentioned in Eq. (C3), it is possible to achieve a nonzero secret key rate using our method.

APPENDIX D: TABULAR REPRESENTATION OF BELL INEQUALITY

Here we introduce an alternative representation of the hyperplane vector [see Eq. (5)]. We rearrange the entries (coefficients of the Bell inequality) in a tabular construction. For the [2,2] Bell scenario, it is represented in Table VI.

This representation is used in Table II. Similarly, we reorder the elements of the hyperplane vector for the [2,3] Bell scenario in the following way (see Table VII):

This tabular representation is used to describe the Bell inequality in Table III. For the generalized $[m, d]$ scenario, the reordered hyperplane vector is represented in Table VIII.

[1] C. H. Bennett and G. Brassard, *Proceedings of the IEEE International Conference on Computers, Systems and Signal Processing* (IEEE, New York, 1984).

[2] A. K. Ekert, Quantum Cryptography Based on Bell's Theorem, Phys. Rev. Lett. **67**, 661 (1991).

[3] C. H. Bennett, Quantum Cryptography Using any Two Nonorthogonal States, Phys. Rev. Lett. **68**, 3121 (1992).

[4] D. Bruß, Optimal Eavesdropping in Quantum Cryptography with Six States, Phys. Rev. Lett. **81**, 3018 (1998).

[5] R. Renner, Security of quantum key distribution, Int. J. Quantum Inf. **06**, 1 (2008).

[6] H.-K. Lo, X. Ma, and K. Chen, Decoy State Quantum Key Distribution, Phys. Rev. Lett. **94**, 230504 (2005).

[7] D. Gottesman, H.-K. Lo, N. Lutkenhaus, and J. Preskill, Security of quantum key distribution with imperfect devices, in *International Symposium on Information Theory, ISIT 2004* (IEEE, New York, 2004), p. 136.

[8] P. W. Shor and J. Preskill, Simple Proof of Security of the BB84 Quantum Key Distribution Protocol, Phys. Rev. Lett. **85**, 441 (2000).

[9] V. Scarani, H. Bechmann-Pasquinucci, N. J. Cerf, M. Dušek, N. Lütkenhaus, and M. Peev, The security of practical quantum key distribution, Rev. Mod. Phys. **81**, 1301 (2009).

[10] X. Ma, B. Qi, Y. Zhao, and H.-K. Lo, Practical decoy state for quantum key distribution, Phys. Rev. A **72**, 012326 (2005).

[11] H.-K. Lo, M. Curty, and K. Tamaki, Secure quantum key distribution, Nat. Photon. **8**, 595 (2014).

[12] M. Tomamichel, C. Lim, N. Gisin, and R. Renner, Tight finite-key analysis for quantum cryptography, Nat. Commun. **3**, 634 (2012).

[13] L. Lydersen, C. Wiechers, C. Wittmann, D. Elser, J. Skaar, and V. Makarov, Hacking commercial quantum cryptography systems by tailored bright illumination, Nat. Photon. **4**, 686 (2010).

[14] I. Gerhardt, Q. Liu, A. Lamas-Linares, J. Skaar, C. Kurtsiefer, and V. Makarov, Full-field implementation of a perfect eavesdropper on a quantum cryptography system, Nat. Commun. **2**, 349 (2011).

[15] Y. Zhao, C. H. F. Fung, B. Qi, C. Chen, and H. K. Lo, Quantum hacking: Experimental demonstration of time-shift attack against practical quantum-key-distribution systems, Phys. Rev. A **78**, 042333 (2008).

[16] D. Mayers and A. Yao, Quantum cryptography with imperfect apparatus, in *Proceedings of the 39th Annual Symposium on Foundations of Computer Science* (IEEE, Piscataway, NJ, 1998), pp. 503–509.

[17] J. Barrett, L. Hardy, and A. Kent, No Signaling and Quantum Key Distribution, Phys. Rev. Lett. **95**, 010503 (2005).

[18] A. Acín, N. Brunner, N. Gisin, S. Massar, S. Pironio, and V. Scarani, Device-Independent Security of Quantum Cryptography against Collective Attacks, Phys. Rev. Lett. **98**, 230501 (2007).

[19] A. Acin, N. Gisin, and L. Masanes, From Bell's Theorem to Secure Quantum Key Distribution, Phys. Rev. Lett. **97**, 120405 (2006).

[20] S. Pironio, A. Acin, N. Brunner, N. Gisin, S. Massar, and V. Scarani, Device-independent quantum key distribution secure against collective attacks, New J. Phys. **11**, 045021 (2009).

[21] E. Hänggi and R. Renner, Device-independent quantum key distribution with commuting measurements, arXiv:1009.1833.

[22] E. Hänggi, R. Renner, and S. Wolf, Efficient device-independent quantum key distribution, in *Annual International Conference on the Theory and Applications of Cryptographic Techniques* (Springer, New York, 2010), pp. 216–234.

[23] L. Masanes, R. Renner, M. Christandl, A. Winter, and J. Barrett, Full security of quantum key distribution from no-signaling constraints, IEEE Trans. Inf. Theory **60**, 4973 (2014).

[24] R. Arnon-Friedman, F. Dupuis, O. Fawzi, R. Renner, and T. Vidick, Practical device-independent quantum cryptography via entropy accumulation, Nat. Commun. **9**, 459 (2018).

[25] R. Arnon-Friedman, R. Renner, and T. Vidick, Simple and tight device-independent security proofs, SIAM J. Comput. **48**, 181 (2019).

[26] L. Masanes, S. Pironio, and A. Acín, Secure device-independent quantum key distribution with causally independent measurement devices, Nat. Commun. **2**, 238 (2011).

[27] J. Ribeiro, G. Murta, and S. Wehner, Fully device-independent conference key agreement, Phys. Rev. A **97**, 022307 (2018).

[28] T. Holz, H. Kampermann, and D. Bruß, A genuine multipartite Bell inequality for device-independent conference key agreement, Phys. Rev. Res. **2**, 023251 (2020).

[29] U. Vazirani and T. Vidick, Fully Device-Independent Quantum Key Distribution, Phys. Rev. Lett. **113**, 140501 (2014).

[30] O. Nieto-Silleras, S. Pironio, and J. Silman, Using complete measurement statistics for optimal device-independent randomness evaluation, New J. Phys. **16**, 013035 (2014).

[31] J.-D. Bancal, L. Sheridan, and V. Scarani, More randomness from the same data, New J. Phys. **16**, 033011 (2014).

[32] I. Pitowski, *Quantum Probability–Quantum Logic*, Lecture Notes in Physics (Springer Nature Switzerland AG, Cham, Switzerland, 1989).

[33] A. Fine, Hidden Variables, Joint Probability, and the Bell Inequalities, Phys. Rev. Lett. **48**, 291 (1982).

[34] I. Pitowsky, Correlation polytopes: Their geometry and complexity, Math. Program. **50**, 395 (1991).

[35] J. S. Bell, On the Einstein Podolsky Rosen paradox, Phys. Phys. Fiz. **1**, 195 (1964).

[36] J. Szangolies, H. Kampermann, and D. Bruß, Device-Independent Bounds on Detection Efficiency, Phys. Rev. Lett. **118**, 260401 (2017).

[37] W. Hoeffding, Probability inequalities for sums of bounded random variables, in *The Collected Works of Wassily Hoeffding*, edited by N. I. Fisher and P. K. Sen, Springer Series in Statistics (Springer Nature Switzerland AG, Cham, Switzerland, 1994), pp. 409–426.

[38] F. Grasselli, *Quantum Cryptography: From Key Distribution to Conference Key Agreement, Quantum Science and Technology Series* (Springer Nature Switzerland AG, Cham, Switzerland, 2020).

[39] N. J. Beaudry, Assumptions in quantum cryptography, arXiv:1505.02792.

[40] V. Scarani and R. Renner, Security bounds for quantum cryptography with finite resources, in *Workshop on Quantum Computation, Communication, and Cryptography* (Springer, 2008), pp. 83–95.

[41] I. Devetak and A. Winter, Distillation of secret key and entanglement from quantum states, Proc. R. Soc. A **461**, 207 (2005).

[42] G. Murta, S. B. van Dam, J. Ribeiro, R. Hanson, and S. Wehner, Towards a realization of device-independent quantum key distribution, Quantum Sci. Technol. **4**, 035011 (2019).

[43] R. Konig, R. Renner, and C. Schaffner, The operational meaning of min- and max-entropy, IEEE Trans. Inf. Theory **55**, 4337 (2009).

[44] N. Johnston, GETLAB: A Matlab toolbox for quantum entanglement, version 0.9, getlab.com, 2016.

[45] M. Navascués, S. Pironio, and A. Acín, Bounding the Set of Quantum Correlations, Phys. Rev. Lett. **98**, 010401 (2007).

[46] M. Navascués, S. Pironio, and A. Acín, A convergent hierarchy of semidefinite programs characterizing the set of quantum correlations, New J. Phys. **10**, 073013 (2008).

[47] J. Löfberg, YALMIP: A toolbox for modeling and optimization in Matlab, in *Proceedings of the IEEE CACSD Conference* (IEEE, New York, 2004).

[48] M. Grant and S. Boyd, CVX: Matlab software for disciplined convex programming, version 2.1. 2014, http://cvxr.com/cvx.

[49] S. Boyd, S. P. Boyd, and L. Vandenberghe, *Convex Optimization* (Cambridge University Press, Cambridge, England, 2004).

[50] M. Grant and S. Boyd, Graph implementations for nonsmooth convex programs, in *Recent Advances in Learning and Control*, edited by V. Blondel, S. Boyd, and H. Kimura, Lecture Notes in Control and Information Sciences (Springer-Verlag, Berlin, 2008), pp. 95–110, http://stanford.edu/~boyd/graph_dcp.html.

[51] P. Wittek, Algorithm 950: Ncpol2sdpa—Sparse semidefinite programming relaxations for polynomial optimization problems of noncommuting variables, ACM Trans. Math. Softw. **41**, 1 (2015).

[52] N. Johnston, QETLAB: A Matlab toolbox for quantum entanglement, version 0.9, http://qetlab.com, Jan. 2016.

[53] K.-C. Toh, M. J. Todd, and R. H. Tütüncü, SDPT3—A Matlab software package for semidefinite programming, Version 1.3, Optim. Method. Softw. **11**, 545 (1999).

[54] J. F. Sturm, Using SeDuMi 1.02, A Matlab toolbox for optimization over symmetric cones, Optim. Method. Softw. **11**, 625 (1999).

[55] M. ApS, MOSEK optimization toolbox for MATLAB, *User's Guide and Reference Manual, Version*, 4 (2019).

[56] K. Brádler, M. Mirhosseini, R. Fickler, A. Broadbent, and R. Boyd, Finite-key security analysis for multilevel quantum key distribution, New J. Phys. **18**, 073030 (2016).

[57] J. F. Clauser, M. A. Horne, A. Shimony, and R. A. Holt, Proposed Experiment to Test Local Hidden-Variable Theories, Phys. Rev. Lett. **23**, 880 (1969).

[58] A. Acin, T. Durt, N. Gisin, and J. I. Latorre, Quantum nonlocality in two three-level systems, Phys. Rev. A **65**, 052325 (2002).

[59] D. Collins, N. Gisin, N. Linden, S. Massar, and S. Popescu, Bell Inequalities for Arbitrarily High-Dimensional Systems, Phys. Rev. Lett. **88**, 040404 (2002).

[60] D. Collins and N. Gisin, A relevant two qubit Bell inequality inequivalent to the ChSh inequality, J. Phys. A: Math. Gen. **37**, 1775 (2004).

[61] E. A. Fonseca and F. Parisio, Measure of nonlocality which is maximal for maximally entangled qutrits, Phys. Rev. A **92**, 030101(R) (2015).

[62] V. Lipinska, F. J. Curchod, A. Máttar, and A. Acín, Towards an equivalence between maximal entanglement and maximal quantum nonlocality, New J. Phys. **20**, 063043 (2018).

[63] A. de Rosier, J. Gruca, F. Parisio, T. Vértesi, and W. Laskowski, Multipartite nonlocality and random measurements, Phys. Rev. A **96**, 012101 (2017).

[64] A. de Rosier, J. Gruca, F. Parisio, T. Vértesi, and W. Laskowski, Strength and typicality of nonlocality in multisetting and multipartite Bell scenarios, Phys. Rev. A **101**, 012116 (2020).

[65] A. Fonseca, A. de Rosier, T. Vértesi, W. Laskowski, and F. Parisio, Survey on the Bell nonlocality of a pair of entangled qudits, Phys. Rev. A **98**, 042105 (2018).

[66] A. Barasiński and M. Nowotarski, Volume of violation of Bell-type inequalities as a measure of nonlocality, Phys. Rev. A **98**, 022132 (2018).

[67] P. Massart, *Concentration Inequalities and Model Selection*, Lecture Notes in Mathematics Vol. 6. (Springer, Berlin, 2007).

[68] S. Boucheron, G. Lugosi, and P. Massart, *Concentration Inequalities: A Nonasymptotic Theory of Independence* (Oxford University Press, Oxford, 2013).

[69] F. Chung and L. Lu, Concentration inequalities and martingale inequalities: A survey, Internet Math. **3**, 79 (2006).

[70] F. Xu, Y.-Z. Zhang, Q. Zhang, and J.-W. Pan, Device-independent quantum key distribution with random post selection, arXiv:2110.02701.

[71] W.-Z. Liu, Y.-Z. Zhang, Y.-Z. Zhen, M.-H. Li, Y. Liu, J. Fan, F. Xu, Q. Zhang, and J.-W. Pan, High-speed device-independent quantum key distribution against collective attacks, arXiv:2110.01480.

[72] E. Y.-Z. Tan, C. C.-W. Lim, and R. Renner, Advantage Distillation for Device-Independent Quantum Key Distribution, Phys. Rev. Lett. **124** (2), 020502 (2020).

[73] 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, Nat. Commun. **12**, 2880 (2021).

[74] P. Brown, H. Fawzi, and O. Fawzi, Computing conditional entropies for quantum correlations, Nat. Commun. **12**, 1 (2021).

[75] P. Brown, H. Fawzi, and O. Fawzi, Device-independent lower bounds on the conditional von neumann entropy, arXiv:2106.13692.

[76] M. Tomamichel, *Quantum Information Processing with Finite Resources: Mathematical Foundations* (Springer, New York, 2015), Vol. 5.

[77] R. Renner and S. Wolf, Simple and tight bounds for information reconciliation and privacy amplification, in *Advances in Cryptology–ASIACRYPT 2005*, edited by B. Roy, Lecture Notes in Computer Science Vol. 3788 (Springer, Berlin, Heidelberg, 2005).

[78] A. Vitanov, F. Dupuis, M. Tomamichel, and R. Renner, Chain rules for smooth min-and max-entropies, IEEE Trans. Inf. Theory **59**, 2603 (2013).

[79] J. L. Carter and M. N. Wegman, Universal classes of hash functions, J. Comput. Syst. Sci. **18**, 143 (1979).

[80] M. Tomamichel, C. Schaffner, A. Smith, and R. Renner, Leftover hashing against quantum side information, IEEE Trans. Inf. Theory **57**, 5524 (2011).

[81] M. Tomamichel, R. Colbeck, and R. Renner, A fully quantum asymptotic equipartition property, IEEE Trans. Inf. Theory **55**, 5840 (2009).

[82] F. Grasselli, H. Kampermann, and D. Bruß, Conference key agreement with single-photon interference, New J. Phys. **21**, 123002 (2019).

[83] H.-L. Yin and Z.-B. Chen, Finite-key analysis for twin-field quantum key distribution with composable security, Sci. Rep. **9**, 17113 (2019).

[84] M. A. Nielsen and I. Chuang, *Quantum Computation and Quantum Information*, 1st ed (Cambridge University Press, Cambridge, 2011).

[85] M. Tomamichel and A. Leverrier, A largely self-contained and complete security proof for quantum key distribution, Quantum **1**, 14 (2017).

B

Upper bound on the Guessing probability using Machine Learning

Title:	Upper bound on the Guessing probability using Machine Learning
Authors:	Sarnava Datta, Hermann Kampermann and Dagmar Bruß
Journal:	Physical Review A
Impact factor:	2.971 (2021)
Date of submission:	19 December 2022
Publication status:	Submitted
Contribution by S.D:	First author (Approx 90 %)

This publication corresponds to the reference [DKB22b]. A summary of the results is presented in Chap. 7. The general framework and the research objective were worked out in collaboration with my co-authors. My co-authors and I spoke about the project frequently. They consistently provide me with thoughtful insights. I suggested the use of deep learning to estimate the guessing probability. Together with HK, I designed the pathway for establishing deep learning models. HK suggested implementing deep neural networks to estimate the optimal Bell inequality, which can then be utilized to compute the upper bound of the guessing probability. I wrote the python code of the neural network architecture on which we train our data. I followed HK and DB's suggestions on what simulations to perform and how to present them. I wrote the entire manuscript, which was proofread and improved by my co-authors.

Upper bound on the Guessing probability using Machine Learning

Sarnava Datta,* Hermann Kampermann, and Dagmar Bruß
Institut für Theoretische Physik III
Heinrich-Heine-Universität Düsseldorf
(Dated: December 19, 2022)

The estimation of the guessing probability has paramount importance in quantum cryptographic processes. It can also be used as a witness for nonlocal correlations. In most of the studied scenarios, estimating the guessing probability amounts to solving a semi-definite programme, for which potent algorithms exist. However, the size of those programs grows exponentially with the system size, becoming infeasible even for small numbers of inputs and outputs. We have implemented deep learning approaches for some relevant Bell scenarios to confront this problem. Our results show the capabilities of machine learning for estimating the guessing probability and for understanding nonlocality.

I. INTRODUCTION

Whenever the statistics of a measurement on a composite quantum state contradict the assumptions of local realism, thus violating a Bell-type inequality, the correlations are referred to as nonlocal [1]. These nonlocal correlations are used to certify private randomness in device-independent quantum key distribution (DIQKD) [2–13] and device-independent randomness generation (DIRNG) [14–22]. For quantifying randomness, estimating the guessing probability is often an important task. The guessing probability is the probability with which an adversary can guess an outcome of another party's measurement. If the guessing probability is less than 1, the adversary cannot predict the outcome with certainty. This implies the presence of intrinsic randomness in the system. However, bounding the guessing probability is not an easy task. Typically it is not possible to explicitly compute the guessing probability, but one can only provide an upper bound by solving a semi-definite optimization problem. Usually, one bounds the guessing probability from a given Bell inequality, and the corresponding quantum violation [7, 14]. Here, one needs to use the hierarchical structure of the quantum correlations [23, 24] to solve the semi-definite optimization problem. The complexity of this optimization problem is increasing and becoming computationally demanding with the number of settings and outcomes.

In this paper, motivated by the outstanding recent progress in utilizing machine learning in the field of quantum information [25–34], we develop deep learning (DL) models that predict the guessing probability along with the optimal Bell inequalities (used to upper bound the guessing probability) from an observed probability distribution using supervised machine learning. A crucial element of supervised machine learning is to generate sample data input and output to train the model. Here, we sample random quantum probability distributions and use them as the input of the training data. With this data, using the two-step method of Ref. [35], we estimate the upper bound of the guessing probability and the optimal Bell inequality, and use it as the output of the training data. After sufficient training, our DL approach can recognize the pattern and predict the guessing probability and the optimal Bell inequality with high accuracy and low average statistical error.

We organize this work as follows. We start in Sec. II by explaining the generalized Bell set-up, types of correlations and Bell inequalities. We introduce the guessing probability and show how to estimate it by solving a semi-definite programme in Sec. III. We introduce our deep learning approach in Sec. IV. We discuss how to sample quantum probability distributions from the quantum correlation space, which are then used as input for supervised learning. We build several deep learning models for predicting the guessing probability and the Bell inequality for various Bell scenarios and measure their efficiency to show the model's utility.

II. GENERALIZED BELL SET-UP

In this section, we introduce a generalized Bell set-up. In each measurement round, two parties, Alice and Bob, share a quantum state ρ_{AB} acting on $\mathcal{H}_A \otimes \mathcal{H}_B$. In the presence of an eavesdropper Eve, her side information E is described via the purification of the joint system ρ_{ABE} acting on $\mathcal{H}_A \otimes \mathcal{H}_B \otimes \mathcal{H}_E$ where $\mathrm{Tr}_E(\rho_{ABE}) = \rho_{AB}$. Each party selects locally an input (a measurement setting) which produces an output (a measurement outcome). We refer to this scenario as a Bell scenario. Alice performs measurements specified by her input $x \in X = \{1, \cdots, m\}$, where each input has k possible outcomes $a \in A = \{1, \cdots, k\}$. Similarly, Bob performs measurements specified by his input $y \in Y = \{1, \cdots, m\}$ and produces the outputs $b \in B = \{1, \cdots, k\}$. We denote this scenario as $[m, k]$ Bell scenario, i.e. m measurement settings with k outcomes each; see Fig. 1 for visualization. After many repetitions, the conditional probability $P(ab|xy)$ can be estimated. The Bell scenario is completely characterized by the set $\mathbf{P} := \{P(ab|xy)\} \subset \mathbb{R}^{m^2k^2}$ of all joint

* Sarnava.Datta@hhu.de

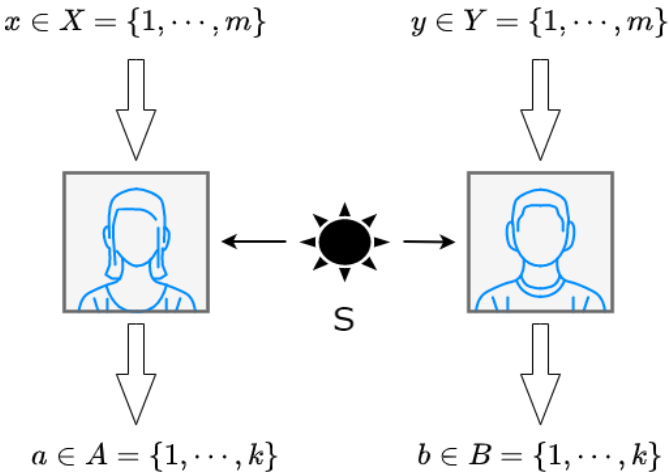

FIG. 1: Schematic description of a Bell scenario consisting of two parties, Alice and Bob. For further explanation, see the main text.

conditional probabilities which we refer to as a behavior [36]. Thus, the following constraints are imposed: positivity $P(ab|xy) \geq 0 \ \forall \ a, b, x, y$ and the normalization $\sum_{a,b=1}^{k} P(ab|xy) = 1$ for all x and y. We say the behavior is no-signaling if the input-output correlation obeys

$$\begin{aligned} \sum_{b=1}^{k} P(ab|xy) &= P(a|x) \qquad \forall a, x, y \quad \text{and} \\ \sum_{a=1}^{k} P(ab|xy) &= P(b|y) \qquad \forall b, x, y \,. \end{aligned} \tag{1}$$

The set of all correlations satisfying the no-signaling constraints forms a convex polytope $\mathcal{NS}$. A behavior is said to be local if it can be written as a convex mixture of deterministic strategies [37, 38]. The set of all local correlations forms a convex polytope $\mathcal{P}$. There exist inequalities of the form [36]

$$\sum_{a,b,x,y} C_{abxy} P(ab|xy) \leq \mathcal{I}_L \,, \tag{2}$$

which separate the set of all local correlations (in other words, the convex polytope $\mathcal{P}$) from the nonlocal behaviors. These inequalities are called Bell inequalities. A Bell inequality is specified by the coefficients $C_{abxy} \in \mathbb{R}$. We denote a Bell inequality as B, and $\sum_{a,b,x,y} C_{abxy} P(ab|xy)$ as the Bell value $B[\mathbf{P}]$ in this paper. Here, $\mathcal{I}_L$ is the classical bound, which is the maximal value over all local behaviors. Thus, a behavior with a classical origin, i.e. $\{P(ab|xy)\} \in \mathcal{P}$, cannot violate this inequality.

The Born rule of quantum theory postulates that a behavior is quantum if there exists a quantum state ρ_{AB} acting on a joint Hilbert space $\mathcal{H}_A \otimes \mathcal{H}_B$ of arbitrary dimension and measurement operators (POVM elements) $\{M_{a|x}\}$ with $M_{a|x} \geq 0$ and $\sum_a M_{a|x} = \mathbb{1} \ \forall x$, and $\{M_{b|y}\}$ with analogous properties such that

$$P(ab|xy) = \mathrm{Tr}(\rho_{AB} M_{a|x} \otimes M_{b|y}) \,. \tag{3}$$

The set of all quantum correlations forms a convex set $\mathcal{Q}$. If a behavior $\{P(ab|xy)\} \in \mathcal{Q} \setminus \mathcal{P}$, it violates at least one Bell inequality of the form in Eq. (2). The sets $\mathcal{P}$, $\mathcal{Q}$ and $\mathcal{NS}$ obey the following relation: $\mathcal{P} \subsetneq \mathcal{Q} \subsetneq \mathcal{NS}$; see Fig. 2 for a pictorial representation.

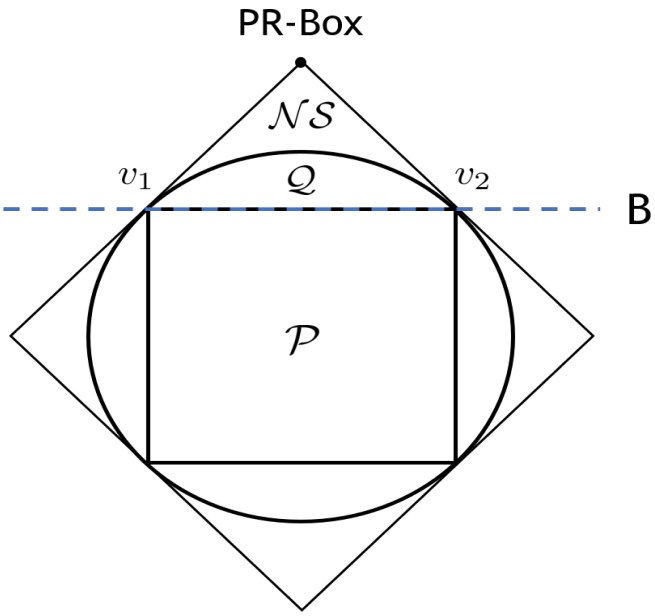

FIG. 2: A pictorial representation for the set of correlations. All classical probabilities form a convex polytope $\mathcal{P}$, which is embedded in the set $\mathcal{Q}$ of quantum correlations, which in turn is a subset of the no-signaling polytope $\mathcal{NS}$. v_1 and v_2 are vertices of the local polytope. B (blue dashed line) represents the Bell inequality which separates the classical polytope from the quantum and no-signaling set.

III. GUESSING PROBABILITY

In an adversarial black box scenario framework, the adversary Eve tries to guess some outcomes obtained by Alice and Bob. The probability that Eve can correctly guess the outcome is called the guessing probability. Here, we denote the guessing probability as $P_g(a|x, E)$, which is the guessing probability of Eve about Alice's outcome a corresponding to her measurement setting x. In Ref.[7], it is shown that $P_g(a|x, E)$ can be upper bounded by a function G_x of the observed Bell value $B[\mathbf{P}]$ of a particular Bell inequality B by semi-definite programming, i.e.

$P_g(a|x,E) \leq G_x(B[\mathbf{P}])$. One crucial element to bound the guessing probability $P_g(a|x,E)$ is to choose a suitable Bell inequality. We follow the two-step procedure of [35] where the Bell inequality is constructed from the input-output probability distribution $\mathbf{P}$ that leads to the maximum Bell violation for that particular measurement statistics.

This is achieved by solving the linear program:

$$\begin{aligned} \max_{\mathbf{h},c} \quad & \mathbf{h}^T\mathbf{P} - c\,, \\ \text{subject to} \quad & \mathbf{h}^T\mathbf{v}_p \leq c \quad \forall \quad p \in \{1,\cdots,k^{2m}\}\,, \\ & \mathbf{h}^T\mathbf{P} > c\,, \\ & -1 \leq h_i \leq 1 \quad \forall \quad i \in \{1,\cdots,m^2k^2\}\,. \end{aligned} \tag{4}$$

Here $\mathbf{h}$ is the hyperplane specifying the Bell inequality B, $\mathbf{P}$ denotes the measurement data, $\mathbf{v}_p$ corresponds to the p vertices of the classical polytope $\mathcal{P}$ and c is the classical bound. Thus the Bell inequality B found by the optimization of Eq. (4) and specified by the hyperplane vector $\mathbf{h}$ is given as:

$$\sum_{a,b,x,y} h_{abxy} P(ab|xy) \leq c\,, \tag{5}$$

where $a \in A$, $b \in B$, $x \in X$, $y \in Y$. We will use the Bell inequality B and corresponding Bell value $B[\mathbf{P}] = \sum_{a,b,x,y} h_{abxy} P(ab|xy)$ to upper bound the guessing probability $P_g(a|x,E)$ by solving the following semidefinite program [7]:

$$\begin{aligned} \max_{\rho_{AB},\{A(a|x)\},\{B(b|y)\}} \quad & P_g(a|x,E) \\ \text{subject to:} \quad & \mathrm{Tr}(\rho_{AB}\mathcal{G}) = B[\mathbf{P}]\,. \end{aligned} \tag{6}$$

In the optimization problem of Eq. (6), the guessing probability is bounded using the NPA-hierarchy [23, 24] up to level 2. The optimization is performed using standard tools YALMIP [39], CVX [40, 41] and QETLAB [42]. Note that, $A(a|x)$ and $B(b|y)$ are the measurement operators of Alice and Bob, respectively, and ρ_{AB} is the state shared between them. $\mathcal{G}$ is the Bell operator defined as:

$$\mathcal{G} = \sum_{a,b,x,y} h_{abxy} A(a|x) B(b|y)\,. \tag{7}$$

Let us denote $P_g^*(a|x,E)$ as the upper bound of the guessing probability, which is the solution to the optimization problem of Eq. (6).

IV. MACHINE LEARNING APPROACH

Providing an upper bound for the guessing probability by solving a semi-definite program is a computationally demanding task. It becomes arduous when the Bell scenario raises its complexity, i.e. for an increased number of measurement inputs and/or outputs in the Bell scenario.

Thus, in this paper, we approach solving the problem via machine learning (ML) (see Ref. [43] for detailed discussions on the concepts of machine learning) so that the trained model can estimate the guessing probability $P_g^*(a|x,E)$ from the input-output probability distribution $\{P(ab|xy)\}$. We are going to use the supervised learning technique. In a supervised ML approach, the first step is generating the training points. We use random bipartite quantum probability distributions as the supervised ML model's input (features), after generating them from facet Bell inequalities using the *weighted vertex sampling* method [44]. Since the guessing probability for local behaviors is always 1 (i.e. Eve can guess the right outcome with probability 1), we do not need to train the machine to perform well on those. Thus we only take samples from the nonlocal part of the no-signaling set, i.e. $\mathcal{NS} \setminus \mathcal{P}$. To single out the input-output correlation with a quantum realization, we reduce the samples using the NPA hierarchy to approximate the quantum realizable probability distribution.

Explicitly, we generate samples from the quantum set $\mathcal{Q}$ as follows. For the $[m,k]$ Bell scenario (i.e. m measurements, k outcomes each), the classical polytope $\mathcal{P}$ is specified by k^{2m} local vertices. The classical polytope can also be described by its facets, which represent the hyperplanes (or Bell inequalities) that separate any nonclassical (quantum and no-signaling) behavior from the classical ones. These facets are called facet Bell inequalities or tight Bell inequalities [36]; see Fig. 2 for a pictorial representation. For the $[2,2]$ scenario, eight facet Bell inequalities exist, all equivalent to the CHSH inequality [45]. For the $[3,2]$ Bell scenario, there are 648 facet Bell inequalities. These facet Bell inequalities are found using the formulation of Ref. [46] [1]. Note that all the 648 facet Bell inequalities correspond to two classes of independent facet Bell inequalities, i.e. the CHSH inequalities and the $I3322$ inequalities [47, 48]. We consider all facet Bell inequalities for the [2,2] and [3,2] Bell scenario while generating training points for the supervised machine learning problem. For the $[4,2]$ Bell scenario, there are 174 independent facet Bell inequalities [49]. Since there are many (>10000) equivalent facets [50], we will only consider the independent ones. These facet Bell inequalities are spanned by some of the local vertices of the classical polytope [2]. These vertices provide the maximum

[1] Using Ref. [46], one can calculate all the facets of a convex polytope given its vertices. The transformation from the vertex representation to the facet representation of a polytope is known as facet enumeration or convex hull problem, which uses Gaussian and Fourier-Motzkin elimination. The list of facets consists of positivity constraints and the facet Bell inequalities. Here, we only focus on the facet Bell inequalities alone.

[2] For the $[2,2]$ Bell scenario, all the facet inequalities are spanned by eight local vertices. For the $[3,2]$ Bell scenario, facet Bell inequalities, equivalent to the CHSH inequality, are spanned by thirty-two vertices. Twenty vertices span the inequalities equivalent to the $I3322$ inequality.

classical bound of the corresponding facet Bell inequality. Consider the case that n local vertices span a facet Bell inequality, where we denote the set of n vertices as $\left\{P_i^{\mathcal{L}}(ab|xy)\right\}_{i=1}^{n}$. We denote the PR-box of the corresponding facet Bell inequality as $P^{\mathrm{PR}}(ab|xy)$, see Fig. 2 for visualization. The PR-Box $P^{\mathrm{PR}}(ab|xy)$ can be defined as the probability distribution that provides the maximal no-signaling bound of the corresponding facet Bell inequality [51, 52]. We take uniformly random weighted mixtures of the $n+1$ vertices (n vertices that span the facet Bell inequality and the corresponding PR-box) with an n-fold weight on the PR-box. Formally, the sample behavior from the set $\mathcal{NS} \setminus \mathcal{P}$ can be generated as:

$$\mathbf{P} := P(ab|xy) = \frac{nw_0 P^{\mathrm{PR}}(ab|xy) + \sum_{i=1}^{n} w_i P_i^{\mathcal{L}}(ab|xy)}{nw_0 + \sum_{i=1}^{n} w_i} \tag{8}$$

where the $w_i \in [0,1]$ with $i = 0, 1, \cdots, n$, are uniformly drawn random numbers. This process is done for all facet inequalities. From this set of samples, we only select the ones with a $\mathcal{Q}_2$ realization (the second level of NPA hierarchy [23, 24]). Here we work under the assumption that $\mathcal{Q}_2$ provides a good approximation for the original quantum set $\mathcal{Q}$.

We store the probability distribution $\{P(ab|xy)\}$ and use it as the input (features) of the supervised machine learning problem, i.e.

$$\mathbf{X} := \{P(ab|xy)\}_{x,y=1,\cdots,m}^{a,b=1,\cdots,k} . \tag{9}$$

We calculate the guessing probability of each input $\mathbf{P}$ using the two-step method (see Sec. III for details), and use it as the output (target), i.e.

$$y = P_g^*(a|x, E) . \tag{10}$$

Without loss of generality, we have always calculated the guessing probability of Alice's first measurement setting. We use a deep neural network to assess the dataset and make predictions. We fed the input-output pair $\{\mathbf{X}, y\}$ (see Eq. (9) and Eq. (10)) into an artificial neural network (ANN) to learn the best possible fit. For an elaborate explanation of an artificial neural network, see Ref. [43]. Following the standard approach, we divide the dataset into two parts. The first part of the dataset is for training and validation (80%), and the second is for testing (20%). We choose a 'linear' [3] ANN with several layers as our model; see Fig. 3 for visualization.

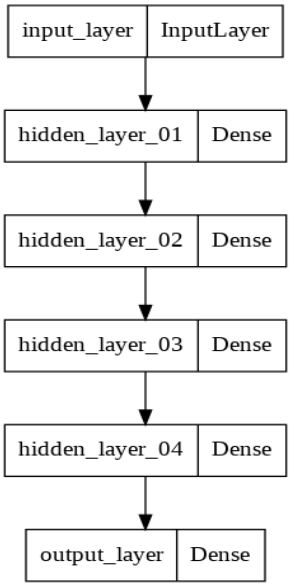

FIG. 3: Schematic description of a 'linear' neural network. It consists of an input layer, several hidden layers and an output layer without branching. The hidden layers and the output layer are dense layers, meaning that the neurons of the layer are connected to every neuron of its preceding layer.

The input layer has m^2k^2 neurons corresponding to the elements in $\{P(ab|xy)\}$. The output (last) layer has only one neuron since we only have to predict one element: the guessing probability $P_g^*(a|x, E)$. We perform 100 rounds of training using the optimizer ADAM [53], of which the first 50 rounds have a fixed learning rate of 0.001. For the next 50 rounds, we reduce the learning rate by 90% in every tenth round. We choose the activation function ReLu (Rectified linear unit) [4] in the input and the hidden layers while using the sigmoid activation function[5] in the output layer. The ReLu activation function introduces non-linearity and the sigmoid activation function keeps the output within the range of [0,1]. The mean squared error (MSE) [6] is used as our loss function, which is minimized during the training process. The trained model

[3] Here, linear means that there is no branching in the hidden layers of the neural network architecture.

[4] Relu activation function: $\phi(x) = max(0, x)$

[5] sigmoid activation function: $\phi(x) = \frac{1}{1+e^{-x}}$

[6] $\mathrm{MSE}(y, \hat{y}) = \frac{1}{N}\sum_{i=0}^{N}(y_i - \hat{y}_i)^2$, where y is the original output and $\hat{y}$ is the estimated output of the model.

Metrics \ Bell Scenario	[2,2]	[3,2]	[4,2]
MSE	0.00007	0.007	0.009
MAE	0.0003	0.01	0.025

TABLE I: Performance measures for different Bell scenarios when estimating $P_g^*(a|x,E)$ from the probability distribution $P(ab|xy)$. MAE: Mean Absolute Error (see Eq. (11)), MSE: Mean Squared Error (see Eq. (12))

generates the predicted value of the guessing probability $P_g^{\text{pred}}(a|x,E)$. To check the efficiency of our approach, we have used the mean absolute error (MAE)

$$\begin{aligned}&\text{MAE}\left[P_g^*(a|x,E), P_g^{\text{pred}}(a|x,E)\right]\\&= \frac{1}{N_{\text{test}}}\sum_{i=1}^{N_{\text{test}}}\left|P_g^*(a|x,E)_i - P_g^{\text{pred}}(a|x,E)_i\right|,\end{aligned} \tag{11}$$

and the mean squared error (MSE)

$$\begin{aligned}&\text{MSE}\left[P_g^*(a|x,E), P_g^{\text{pred}}(a|x,E)\right]\\&= \frac{1}{N_{\text{test}}}\sum_{i=1}^{N_{\text{test}}}\left(P_g^*(a|x,E)_i - P_g^{\text{pred}}(a|x,E)_i\right)^2,\end{aligned} \tag{12}$$

as a performance measure. N_{test} is the number of data points in the test set. We analyze the results for different bipartite Bell scenarios and list the errors in Table I. The average error is in the order of 10^{-4} to 10^{-2}. Such high accuracy and small error without knowing the Bell inequality are truly remarkable. We also compare the

Time per sample \ Bell Scenario	[2,2]	[3,2]	[4,2]
Mosek	95ms	496ms	1568 ms
Neural Network	27μs	35μs	49μs

TABLE II: Runtime per sample comparison for SDP solver Mosek and the neural network method for estimating the guessing probability for different Bell scenarios.

runtime performance of the neural network model with the frequently used SDP solver Mosek [54] (that can be used to upper bound the guessing probability by solving the optimization problem of Eq. (6)) in Table II. The Mosek task is generated and solved using the Ncpol2sdpa [55]. The results are evaluated over 10000 unknown samples and performed on a personal computer [7] under comparable conditions. Once the neural network is trained, we get a speed-up of $10^3 - 10^5$ for obtaining a prediction about a new instance, compared to the runtime of the usual method for solving the optimization problem; see Table II. This follows from the fact that the number of variables in the optimization process of Eq. (6) increases exponentially with the number of measurement settings (or outcomes per measurement) in the Bell scenario. Thus, it takes more computational time to perform the SDP using a classical solver like Mosek. A trained neural network only calculates the functional output using the optimized weights and biases. Only the neural network size affects the computational time needed to complete the prediction task.

However, the upper bound on the guessing probability calculated from a trained machine learning model only provides an estimate of its real value. Thus, we cannot use this estimate to bound the secret key rate. The predicted Bell inequality on the other side that generates a non-zero Bell violation (for a particular measurement statistics) can be used to bound the guessing probability (see Sec. III for details) and the secret key rate. That's why in the next step, we use deep learning to predict the associated optimal Bell inequality B, which is then used to upper bound the guessing probability (see Sec. III for details). For this purpose, we again use the neural network architecture where supervised learning is incorporated. We start by preparing the dataset where our input features are

$$\mathbf{X} := \{P(ab|xy)\}_{x,y=1,\cdots,m}^{a,b=1,\cdots,k}. \tag{13}$$

Note that, the input is identical to Eq. (9). The outputs are now the coefficients of the optimal Bell inequality B (specified by $\{h_{abxy}\}_{x,y=1,\cdots,m}^{a,b=1,\cdots,k}$, see Eq. (4)) and the guessing probability $P_g^*(a|x,E)$, i.e.

$$\mathbf{y} := \left[\{h_{abxy}\}_{x,y=1,\cdots,m}^{a,b=1,\cdots,k}, P_g^*(a|x,E)\right]. \tag{14}$$

Here we use two types of neural network architectures. The first neural network is a usual 'linear' feed forward neural network (see [43] for details, schematically represented in Fig. 3). For $[m,k]$ Bell scenario, the input layer has m^2k^2 neurons (corresponds to the elements of $\{P(ab|xy)\}$). The input layer is followed by several hidden layers. Unlike in the previous scenario, the output layer has m^2k^2+1 neurons in this case, where m^2k^2 neurons correspond to the coefficients of the Bell inequality h_{abxy}, and one neuron corresponds to the guessing probability $P_g^*(a|x,E)$. In this paper, we denote this construction of the 'linear' deep neural network as NN_1. Following the standard approach, we divide the dataset $\{\mathbf{X},\mathbf{y}\}$ (see Eq. (13) and Eq. (14)) into two sets; the first part of the dataset is for training and validation (80%), and the second part is for testing (20%). Similar to the training of the previous network, we perform 100 rounds (first 50 rounds with a 0.001 learning rate and then reduce the learning rate by 90% in every tenth round) of

[7] Specifications: Intel(R) Core(TM) i7-10510U Processor, 2.30GHz Frequency, 16.0 GB RAM

training using the gradient solver ADAM. Similar to the previous scenario, we use the activation function ReLu in the input and the hidden layers. In the output layer, the linear activation function [8] is used for m^2k^2 neurons that correspond to the optimal Bell inequality and the sigmoid activation function is incorporated for the neuron that corresponds to the guessing probability. As the cost function, we use the Mean Squared Error (MSE) which is minimized during the training process.

In addition, we use another neural network architecture with two parallel sub-models (by using branching) to interpret parts of the output that share the same input. In this construction, the input layer has m^2k^2 neurons corresponding to the elements of the probability distribution $\{P(ab|xy)\}$ of the $[m, k]$ Bell scenario. The input layer is followed by hidden layers consisting of multiple neurons. Then we bifurcate one hidden layer to create two branches. Several hidden layers then follow both branches; see Fig. 4 for visualization. The first branch of the network is for predicting the coefficients of the optimal Bell inequality $\{h_{abxy}\}_{x,y=1,\cdots,m}^{a,b=1,\cdots,k}$ and thus has m^2k^2 neurons. The second branch of the network is for predicting the guessing probability. Thus, the output layer will have only one neuron corresponding to $P_g^*(a|x, E)$. In this paper, we refer to this neural network as NN_2 which is built using the Keras functional API [56]. In NN_2, we use the ReLu activation function in the input and all the hidden layers. The linear activation function is used in the output layer of the first branch (which predicts the coefficients of the Bell inequality) while the sigmoid activation function is used in the second branch (which predicts the guessing probability). The other details of the training steps are the same as for the NN_1 neural network stated previously. Both NN_1 and NN_2 predict the Bell inequality B^{pred} (specified by the predicted coefficients $\{h_{abxy}^{\mathrm{pred}}\}_{x,y=1,\cdots,m}^{a,b=1,\cdots,k}$) and the guessing probability $P_g^{\mathrm{pred}}(a|x, E)$.

Since the neural networks NN_1 and NN_2 predict two separate entities (the optimal Bell inequality and the guessing probability), we evaluate their performance separately. We use the mean absolute error (see Eq. (11)) and mean squared error (see Eq. (12)) as our performance measure of predicting the guessing probability. The errors for different bipartite Bell scenarios are listed in Table III.

[8] Linear activation function: $\phi(x) = x$

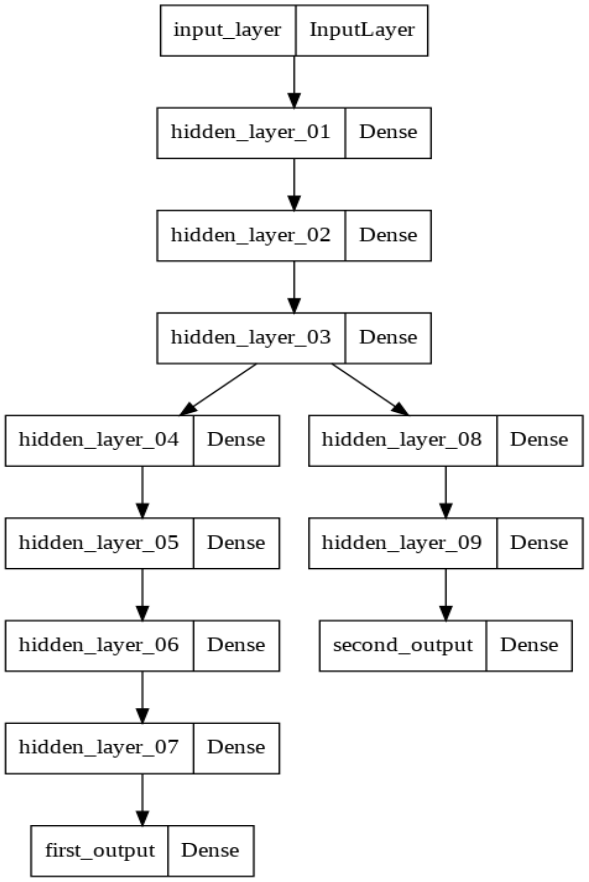

FIG. 4: Schematic diagram of a neural network where a hidden layer is bifurcated into two different arms which predict different parts of the output separately. In our scenario, input layer: $\{P(ab|xy)\}$ (m^2k^2 neurons), first output: $\{h_{abxy}\}$ (m^2k^2 neurons) and second output: $P_g^*(a|x, E)$ (1 neuron).

ANN	Metrics	[2,2]	[3,2]	[4,2]
NN_1	MSE	8.2×10^{-6}	0.002	0.009
	MAE	0.001	0.02	0.07
NN_2	MSE	1.9×10^{-7}	0.001	0.002
	MAE	0.0003	0.013	0.027

TABLE III: Statistical errors of the predicted guessing probability $P_g^{\mathrm{pred}}(a|x, E)$ with respect to the guessing probability $P_g^*(a|x, E)$ for different Bell scenarios for NN_1 and NN_2. Here the neural network is trained for predicting the guessing probability $P_g^*(a|x, E)$ and the Bell inequality B from the probability distribution $P(ab|xy)$.

Note that, for estimating the guessing probability, NN_2 yields lower statistical errors than NN_1. The reason lies in the structure of the neural network architectures. Since we create a branch in the neural network only to estimate the guessing probability, the NN_2 neural network assigns more nodes to only estimate the guessing probability than NN_1. In the case of predicting the

optimal Bell inequality B (characterized by its coefficients $\{h_{abxy}\}_{x,y=1,\cdots,m}^{a,b=1,\cdots,k}$), we use the performance measure MSE, which reads:

$$\begin{aligned}&\mathrm{MSE}\left[B, B^{\mathrm{pred}}\right]\\&= \frac{1}{m^2k^2}\frac{1}{N_{\mathrm{test}}}\sum_{i=1}^{N_{\mathrm{test}}}\sum_{a,b=1}^{k}\sum_{x,y=1}^{m}\left((h_{abxy})_i - (h^{\mathrm{pred}}_{abxy})_i\right)^2,\end{aligned} \tag{15}$$

and MAE, which reads:

$$\begin{aligned}&\mathrm{MAE}\left[B, B^{\mathrm{pred}}\right]\\&= \frac{1}{m^2k^2}\frac{1}{N_{\mathrm{test}}}\sum_{i=1}^{N_{\mathrm{test}}}\sum_{a,b=1}^{k}\sum_{x,y=1}^{m}\left|(h_{abxy})_i - (h^{\mathrm{pred}}_{abxy})_i\right|.\end{aligned} \tag{16}$$

The errors are listed in Table IV.

ANN	Metrics	[2,2]	[3,2]	[4,2]
NN_1	MSE	0.0004	0.0007	0.014
	MAE	0.001	0.002	0.067
NN_2	MSE	0.0003	0.0005	0.015
	MAE	0.0009	0.002	0.069

TABLE IV: Statistical errors of the coefficients of the predicted Bell inequality B^{pred} (predicted by the trained deep learning models NN_1 and NN_2), $\{h^{\mathrm{pred}}_{abxy}\}_{x,y=1,\cdots,m}^{a,b=1,\cdots,k}$ with respect to the original coefficients $\{h_{abxy}\}_{x,y=1,\cdots,m}^{a,b=1,\cdots,k}$ for different Bell scenarios.

Another way to evaluate the quality of the predicted Bell inequality is to use it for upper bounding the guessing probability problem (see Eq. (6)). First, we estimate the probability of $P_g^*(a|x,E) < 1$, where $P_g^*(a|x,E)$ is calculated from the predicted Bell inequality B^{pred} and the input-output probability distribution $\{P(ab|xy)\}$ of the test set. We present the results in Table V. We

ANN	[2,2]	[3,2]	[4,2]
NN_1	99.5%	98.7%	93.4%
NN_2	99.6%	99.4%	94.6%

TABLE V: Probability of $P_g(a|x,E) < 1$ when using the Bell inequality B^{pred} (predicted by the trained deep learning models NN_1 and NN_2).

also look into the statistical errors between the original guessing probability $P_g^*(a|x,E)$ from the test set and the guessing probability calculated from the predicted Bell inequality B^{pred}. We use MAE and MSE as the performance measures listed in Table VI. The high probability of generating $P_g^*(a|x,E) < 1$ with the predicted Bell inequalities (see Table V) and the small statistical errors (see Table VI) demonstrate the quality and accuracy of the predicted Bell inequality.

We again compare the computational runtime of predicting the optimal Bell inequality using the standard linear optimization of Eq. (4) with the neural network

ANN	Metrics	[2,2]	[3,2]	[4,2]
NN_1	MSE	1.7×10^{-8}	0.002	0.006
	MAE	6.3×10^{-5}	0.014	0.038
NN_2	MSE	1.5×10^{-8}	0.002	0.004
	MAE	4.1×10^{-5}	0.014	0.031

TABLE VI: Statistical errors between the guessing probability calculated from the predicted Bell inequality and the original guessing probability from the test set for various Bell scenarios and neural network constructions.

NN_1 and NN_2. The runtime for different methods is shown in Table VII. Similar to the previous scenario,

Time per sample \ Bell Scenario	[2,2]	[3,2]	[4,2]
LP	95 ms	159 ms	379 ms
NN_1	29 μs	40 μs	57 μs
NN_2	31 μs	41 μs	55 μs

TABLE VII: Runtime per sample comparison for estimation of the optimal Bell inequality using linear programming of Eq. (4) and the neural network method for different Bell scenarios. LP stands for linear programming which is performed using the Mosek solver.

the runtimes are evaluated over 10000 unknown samples and performed on the same personal computer in the same condition. The linear programming of Eq. (4) is performed with the Mosek solver using PICOS [57] python interface. We notice a significant speed-up when using the trained neural network models compared to the Mosek solver. This again follows from the fact that the number of variables in the optimization process of Eq. (4) increases with the number of measurement settings (or outcomes per measurement) in the Bell scenario while the computational time for the neural networks only depends on its size.

V. DISCUSSION & CONCLUSION

Estimating the guessing probability is a cornerstone for device-independent quantum key distribution and device-independent randomness generation. This paper introduces a novel method to estimate the guessing probability using trained deep learning models to bypass the computationally complex and cumbersome semi-definite optimization process. Computation with the trained deep learning models is significantly faster than using a conventional solver. With current technology, Bell test event rates are around 100 kHz, which results in new data every 10μs [58]. This frequency is too high for conventional SDP solvers on a single CPU. For those cases, our

deep learning approach improves the computation significantly. In principle, optimizing the size of a deep neural network that can process each event as the experiment is being conducted is possible.

The deep learning model only provides an estimation of the upper bound of the guessing probability. But it will not provide a certification. Thus, additionally, our DL model provides an estimation of the optimal Bell inequality for which the Bell violation using the measurement statistics certifies the nonlocality of input-output correlations and guarantees that the guessing probability will be less than one. Our trained deep learning models, which significantly speed up the prediction of the Bell inequality compared to a conventional linear program solver, predict a Bell inequality that can generate $P_g(a|x, E) < 1$ with a very high probability. The mean average error between the guessing probability calculated from the predicted Bell inequality and the optimal Bell inequality (calculated using Eq. (4)) is in the order of $10^{-5} - 10^{-2}$ (mean squared error is in the order of $10^{-8} - 10^{-3}$) which shows the quality of this approach such that it can efficiently be used in a DIQKD or DIRNG protocol.

We also demonstrate a method for sampling random quantum correlations (correlations which have a realization of NPA hierarchy level of 2) using the facet Bell inequalities, which is then used as input in the supervised machine learning process. Note that, while generating probability distributions, we consider all facet Bell inequalities for the $[2, 2]$ and $[3, 2]$ Bell scenario. However, since there are more than 10000 facet Bell inequalities for the $[4, 2]$ Bell scenario, we only restrict ourselves to generating probability distributions using the independent facet Bell inequalities.

To illustrate the benefits of our method, we have applied it to several relevant Bell scenarios. Note that we design and train our neural networks to minimize statistical errors. However, we do not claim that our choice of the trained neural network is optimal for estimating the guessing probability and the associated optimal Bell inequality from the measurement statistics. Other constructions of neural networks will lead to different results.

We observed that the statistical errors in the estimation of the guessing probability and the optimal Bell inequality increase with the complexity of the Bell scenario (i.e., the increase in the number of measurements per party). Since there are more inputs and outputs, our neural network architecture might not be able to generalize the extensive system with a limited number of hidden layers and nodes in each layer. To decrease the errors, one can take two steps. First, one can generate a larger dataset to train the model. Second, one can build a more extensive neural network architecture (i.e., more hidden layers or nodes in every layer). However, using a larger dataset for training or/and training a more extensive neural network will result in significantly more computational time. There is also the possibility of overfitting in an extensive network. A larger neural network architecture will also take more time to predict new instances. Therefore, one has to change the network architecture to optimize the speed and precision of a specific scenario.

Note that while comparing the runtime for the Mosek optimization solver with the trained neural network for the estimation of the guessing probability (see Table II), we implement the NPA hierarchy of level 2. The difference in computational runtime between the methods will be much more pronounced with increasing hierarchy.

Our research demonstrates the applicability of deep learning techniques for Bell nonlocality and upper bounding the guessing probability. We believe that this strategy will create several research lines. The logical next step is to apply our approach to Bell scenarios with a higher number of measurement settings and outcomes. It is also possible to expand our framework to a multipartite scenario. Another direction worth exploring for future work is investigating other neural network constructions. Beyond the advantage in speed, one could use neural network architectures to search for new Bell inequalities. Also, recall that our methodology does not account for uncertainty or offers certification of the output. It remains for future work to use techniques like probabilistic modeling [59] that can certify the correctness of the model's output.

VI. ACKNOWLEDGEMENT

The authors acknowledge support from the Federal Ministry of Education and Research BMBF (Project Q.Link.X). We thank Lucas Tendick for helpful discussions.

[1] J. S. Bell, "On the Einstein Podolsky Rosen paradox," *Physics Physique Fizika*, vol. 1, no. 3, p. 195, 1964.

[2] A. Acín, N. Brunner, N. Gisin, S. Massar, S. Pironio, and V. Scarani, "Device-independent security of quantum cryptography against collective attacks," *Physical Review Letters*, vol. 98, no. 23, p. 230501, 2007.

[3] S. Pironio, A. Acin, 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. 045021, 2009.

[4] R. Arnon-Friedman, R. Renner, and T. Vidick, "Simple and tight device-independent security proofs," *SIAM Journal on Computing*, vol. 48, no. 1, pp. 181–225, 2019.

[5] 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.

[6] J. Barrett, L. Hardy, and A. Kent, "No signaling and quantum key distribution," *Physical Review Letters*, vol. 95, no. 1, p. 010503, 2005.

[7] L. Masanes, S. Pironio, and A. Acín, "Secure device-independent quantum key distribution with causally independent measurement devices," *Nature Communications*, vol. 2, p. 238, 2011.

[8] U. Vazirani and T. Vidick, "Fully device-independent quantum key distribution," *Physical Review Letters*, vol. 113, no. 14, p. 140501, 2014.

[9] L. Masanes, R. Renner, M. Christandl, A. Winter, and J. Barrett, "Full security of quantum key distribution from no-signaling constraints," *IEEE Transactions on Information Theory*, vol. 60, no. 8, pp. 4973–4986, 2014.

[10] A. Acin, S. Massar, and S. Pironio, "Efficient quantum key distribution secure against no-signalling eavesdroppers," *New Journal of Physics*, vol. 8, no. 8, p. 126, 2006.

[11] G. Murta, S. B. van Dam, J. Ribeiro, R. Hanson, and S. Wehner, "Towards a realization of device-independent quantum key distribution," *Quantum Science and Technology*, vol. 4, no. 3, p. 035011, 2019.

[12] T. Holz, H. Kampermann, and D. Bruß, "Genuine multipartite bell inequality for device-independent conference key agreement," *Physical Review Research*, vol. 2, no. 2, p. 023251, 2020.

[13] E. Hänggi and R. Renner, "Device-independent quantum key distribution with commuting measurements," *arXiv preprint arXiv:1009.1833*, 2010.

[14] 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.

[15] O. Nieto-Silleras, C. Bamps, J. Silman, and S. Pironio, "Device-independent randomness generation from several bell estimators," *New Journal of Physics*, vol. 20, no. 2, p. 023049, 2018.

[16] S. Pironio and S. Massar, "Security of practical private randomness generation," *Physical Review A*, vol. 87, no. 1, p. 012336, 2013.

[17] J.-D. Bancal, L. Sheridan, and V. Scarani, "More randomness from the same data," *New Journal of Physics*, vol. 16, no. 3, p. 033011, 2014.

[18] O. Nieto-Silleras, S. Pironio, and J. Silman, "Using complete measurement statistics for optimal device-independent randomness evaluation," *New Journal of Physics*, vol. 16, no. 1, p. 013035, 2014.

[19] F. Bischof, H. Kampermann, and D. Bruß, "Measurement-device-independent randomness generation with arbitrary quantum states," *Physical Review A*, vol. 95, no. 6, p. 062305, 2017.

[20] A. Acín, S. Massar, and S. Pironio, "Randomness versus nonlocality and entanglement," *Physical Review Letters*, vol. 108, no. 10, p. 100402, 2012.

[21] A. Acín and L. Masanes, "Certified randomness in quantum physics," *Nature*, vol. 540, no. 7632, pp. 213–219, 2016.

[22] P. Skrzypczyk and D. Cavalcanti, "Maximal randomness generation from steering inequality violations using qudits," *Physical Review Letters*, vol. 120, no. 26, p. 260401, 2018.

[23] M. Navascués, S. Pironio, and A. Acín, "Bounding the set of quantum correlations," *Physical Review Letters*, vol. 98, no. 1, p. 010401, 2007.

[24] M. Navascués, S. Pironio, and A. Acín, "A convergent hierarchy of semidefinite programs characterizing the set of quantum correlations," *New Journal of Physics*, vol. 10, no. 7, p. 073013, 2008.

[25] J. Carrasquilla and R. G. Melko, "Machine learning phases of matter," *Nature Physics*, vol. 13, no. 5, pp. 431–434, 2017.

[26] P. Broecker, J. Carrasquilla, R. G. Melko, and S. Trebst, "Machine learning quantum phases of matter beyond the fermion sign problem," *Scientific Reports*, vol. 7, no. 1, pp. 1–10, 2017.

[27] A. Canabarro, S. Brito, and R. Chaves, "Machine learning nonlocal correlations," *Physical Review Letters*, vol. 122, no. 20, p. 200401, 2019.

[28] G. Carleo and M. Troyer, "Solving the quantum many-body problem with artificial neural networks," *Science*, vol. 355, no. 6325, pp. 602–606, 2017.

[29] D.-L. Deng, "Machine learning detection of bell nonlocality in quantum many-body systems," *Physical Review Letters*, vol. 120, no. 24, p. 240402, 2018.

[30] X. Gao and L.-M. Duan, "Efficient representation of quantum many-body states with deep neural networks," *Nature Communications*, vol. 8, no. 1, pp. 1–6, 2017.

[31] Y.-C. Ma and M.-H. Yung, "Transforming bell's inequalities into state classifiers with machine learning," *NPJ Quantum Information*, vol. 4, no. 1, pp. 1–10, 2018.

[32] P. Mehta, M. Bukov, C.-H. Wang, A. G. Day, C. Richardson, C. K. Fisher, and D. J. Schwab, "A high-bias, low-variance introduction to machine learning for physicists," *Physics Reports*, 2019.

[33] G. Torlai, G. Mazzola, J. Carrasquilla, M. Troyer, R. Melko, and G. Carleo, "Neural-network quantum state tomography," *Nature Physics*, vol. 14, no. 5, pp. 447–450, 2018.

[34] J. Venderley, V. Khemani, and E.-A. Kim, "Machine learning out-of-equilibrium phases of matter," *Physical Review Letters*, vol. 120, no. 25, p. 257204, 2018.

[35] S. Datta, H. Kampermann, and D. Bruß, "Device-independent secret key rates via a postselected bell inequality," *Phys. Rev. A*, vol. 105, p. 032451, Mar 2022.

[36] N. Brunner, D. Cavalcanti, S. Pironio, V. Scarani, and S. Wehner, "Bell nonlocality," *Reviews of Modern Physics*, vol. 86, no. 2, p. 419, 2014.

[37] I. Pitowsky, "Resolution of the einstein-podolsky-rosen and Bell paradoxes," *Physical Review Letters*, vol. 48, no. 19, p. 1299, 1982.

[38] I. Pitowsky, "Correlation polytopes: their geometry and complexity," *Mathematical Programming*, vol. 50, no. 1-3, pp. 395–414, 1991.

[39] J. Löfberg, "Yalmip: A toolbox for modeling and optimization in matlab," in *In Proceedings of the CACSD Conference*, (Taipei, Taiwan), 2004.

[40] M. Grant and S. Boyd, "Graph implementations for nonsmooth convex programs," in *Recent Advances in Learning and Control* (V. Blondel, S. Boyd, and H. Kimura, eds.), Lecture Notes in Control and Information Sciences, pp. 95–110, Springer-Verlag Limited, 2008. `http://stanford.edu/~boyd/graph_dcp.html`.

[41] M. Grant and S. Boyd, "CVX: Matlab software for disciplined convex programming, version 2.1." `http://cvxr.com/cvx`, Mar. 2014.

[42] N. Johnston, "QETLAB: A MATLAB toolbox for quantum entanglement, version 0.9." `http://qetlab.com`, Jan. 2016.

[43] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. MIT Press, 2016. `http://www.deeplearningbook.org`.

[44] T. Kriváchy, Y. Cai, J. Bowles, D. Cavalcanti, and N. Brunner, "High-speed batch processing of semidefinite programs with feedforward neural networks," *New Journal of Physics*, vol. 23, no. 10, p. 103034, 2021.

[45] J. F. Clauser, M. A. Horne, A. Shimony, and R. A. Holt, "Proposed experiment to test local hidden-variable theories," *Physical Review Letters*, vol. 23, no. 15, p. 880, 1969.

[46] K. Fukuda, "Cddlib reference manual," *Report version 093a, McGill University, Montréal, Quebec, Canada*, 2003.

[47] D. Collins and N. Gisin, "A relevant two qubit bell inequality inequivalent to the chsh inequality," *Journal of Physics A: Mathematical and General*, vol. 37, no. 5, p. 1775, 2004.

[48] K. F. Pál and T. Vértesi, "Maximal violation of a bipartite three-setting, two-outcome bell inequality using infinite-dimensional quantum systems," *Physical Review A*, vol. 82, no. 2, p. 022116, 2010.

[49] E. Z. Cruzeiro and N. Gisin, "Complete list of tight bell inequalities for two parties with four binary settings," *Physical Review A*, vol. 99, no. 2, p. 022104, 2019.

[50] T. Cope and R. Colbeck, "Bell inequalities from no-signaling distributions," *Phys. Rev. A*, vol. 100, p. 022114, Aug 2019.

[51] J. Barrett, N. Linden, S. Massar, S. Pironio, S. Popescu, and D. Roberts, "Nonlocal correlations as an information-theoretic resource," *Physical review A*, vol. 71, no. 2, p. 022101, 2005.

[52] S. Popescu and D. Rohrlich, "Quantum nonlocality as an axiom," *Foundations of Physics*, vol. 24, no. 3, pp. 379–385, 1994.

[53] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.

[54] M. ApS, "Mosek optimization toolbox for matlab," *User's Guide and Reference Manual, Version*, vol. 4, 2019.

[55] P. Wittek, "Algorithm 950: Ncpol2sdpa—sparse semidefinite programming relaxations for polynomial optimization problems of noncommuting variables," *ACM Transactions on Mathematical Software (TOMS)*, vol. 41, no. 3, pp. 1–12, 2015.

[56] F. Chollet, "keras." `https://github.com/fchollet/keras`, 2015.

[57] G. Sagnol and M. Stahlberg, "Picos: A python interface to conic optimization solvers," *Journal of Open Source Software*, vol. 7, no. 70, p. 3915, 2022.

[58] P. Bierhorst, E. Knill, S. Glancy, Y. Zhang, A. Mink, S. Jordan, A. Rommal, Y.-K. Liu, B. Christensen, S. W. Nam, *et al.*, "Experimentally generated randomness certified by the impossibility of superluminal signals," *Nature*, vol. 556, no. 7700, pp. 223–226, 2018.

[59] Z. Ghahramani, "Probabilistic machine learning and artificial intelligence," *Nature*, vol. 521, no. 7553, pp. 452–459, 2015.