

# Quantum state discrimination with applications in contextuality and randomness certification

**Carceller, Carles Roch i**

*Publication date:*
2023

*Document Version*
Publisher's PDF, also known as Version of record

# Quantum state discrimination with applications in contextuality and randomness certification

**Carles Roch i Carceller**

A thesis presented for the degree of
Doctor of Philosophy

Technical University of Denmark
Department of Physics
Denmark
August 2023

| | |
|---|---|
| Title: | Quantum state discrimination with applications in contextuality and randomness certification |
| Author: | Carles Roch i Carceller |
| Supervisors: | Assistant Professor Jonatan Bohr Brask<br>Assistant Professor Jonas Schou Neergaard-Nielsen<br>Associate Professor Joonwoo Bae |
| Period: | September 2020 - August 2023 |
| University: | Technical University of Denmark |
| Department: | DTU Physics, Department of Physics<br>Section of Quantum Physics and Information Technology, QPIT<br>Center for Macroscopic Quantum States, bigQ |

*Pel papa.*

# Abstract

The theory of quantum mechanics establishes that some physical behaviors depart drastically from the classical world we are all used to. For instance, two different state preparations on the same physical system, in general, cannot be perfectly distinguishable. This statement motivates the question of how well quantum states can be discriminated, which is the main study of quantum state discrimination. In this thesis, I investigate two important branches of quantum information science from the perspective of quantum state discrimination.

The first is quantum randomness. Randomness plays an important role in cryptography, where unpredictability is central. Here, I present a surprisingly simple protocol to generate more than one bit of certified randomness per round in a qubit prepare-and-measure scenario. The protocol is also implemented and experimentally tested in an optical platform.

The second is quantum contextuality. Contextuality is a fundamental property of quantum mechanics which states that the distribution of measurement outcomes of a physical system depends not only on its state, but also on the context in which it is measured. In this thesis I propose a noise-robust contextuality witness based on optimal two-state discrimination. Moreover, I explore contextual advantages in state discrimination tasks versus noncontextual models, which can be seen as representing classical physics.

Throughout this thesis, I center on maximum confidence discrimination, a state discrimination protocol with the goal of maximizing the confidence. That is, the probability that the state preparation is indeed the one indicated by the measurement outcome. From this perspective, I compare the power of quantum and noncontextual models in terms of randomness certification. Our results report that the certified randomness in a quantum framework is greater than in the noncontextual model, as long as the adversary is quantum in both cases.

# Resumé (Danish)

Kvantemekanikken fastslår at visse fysiske fænomener adskiller sig markant fra den klassiske forståelse. For eksempel kan to forskellige tilstande af det samme fysiske system generelt ikke skelnes perfekt. Denne påstand motiverer spørgsmålet om hvor godt kvantetilstande kan skelnes, hvilket er hovedstudiet i kvantetilstandsskelnen. I denne afhandling undersøger jeg to vigtige grene af kvanteinformationsvidenskaben set gennem kvantetilstandsskelnens perspektiv.

Den første er kvantemekanisk tilfældighed. Tilfældighed spiller en vigtig rolle inden for kryptografi, hvor uforudsigelighed er centralt. Her præsenterer jeg en overraskende simpel protokol til at generere mere end én bit certificeret tilfældighed pr. runde i et qubit forbered-og-mål-scenarie. Protokollen implementeres også og testes eksperimentelt på en optisk platform.

Den anden er kvantekontekstualitet. Kontekstualitet er en grundlæggende egenskab ved kvantemekanikken, som siger at fordelingen af måleresultater for et fysisk system afhænger ikke kun af dets tilstand, men også af den kontekst hvori der måles. I denne afhandling foreslår jeg en støjrobust kontekstualitetsindikator baseret på optimal skelnen mellem to tilstande. Derudover udforsker jeg kontekstuelle fordele i forhold til tilstandsskelnen, kontra ikke-kontekstuelle modeller, der kan ses som repræsentationer af klassisk fysik.

I hele denne afhandling fokuserer jeg på skelnen med maksimal pålidelighed, dvs. protokoller for tilstandsskelnen med målet om at maksimere tilliden til at tilstandspræpareringen faktisk er den, der angives af måleresultatet. Fra denne synsvinkel sammenligner jeg kvante- og ikke-kontekstuelle modeller i termer af tilfældighedscertificering. Vores resultater viser, at den certificerede tilfældighed i en kvantemæssig ramme er større end i den ikke-kontekstuelle model, så længe modstanderen er kvantemekanisk i begge tilfælde.

# Acknowledgements

First and foremost, I would like to thank my main supervisor Jonatan Bohr Brask for all the time and trust he put on me since I started on this long road. Your good mood, empathy and unconditional support have been of great help. I am really thankful for all our fruitful discussions, for your advise, for correcting me when I was wrong, inspire me with bright new ideas and support me when I was right.

I would also like to thank the rest of the QPIT group at the Technical University of Denmark, specially to the members of the Quantum Information Theory group with whom I overlapped over these years: Mathias Rønnow Jørgensen, Michael Jabbour, Bradley Longstaff, Anders Bjerrum, Kristian Toccacelo, Kieran Wilkinson, Esben Karlund, Paula Ruiz, Ivan Derkach, Olga Solodovnikova and Niklas Budinger. It is comfortable to see how the small theory group has been slowly growing since I started three years ago during a covid pandemic. I really appreciated the transition to a less remote working environment, being able to interact with brilliant people with whom to chat about physics or other things of interest. Furthermore, I would like to thank my co-supervisor Jonas Neegard Nielsen together with Lucas Nunes Faria and Zhenghao Liu for their work and patience in implementing in the lab a randomness certification protocol presented in this thesis. I have seen the experimental side of research, it is fascinating but very arduous, making me realise of their incredible work being even more valuable.

Moreover, I would like to thank my co-supervisor Joonwoo Bae, together with Kieran Flatt and Hanwool Lee, our close collaborators from KAIST, South Korea. I am more than grateful for the warm welcome I received when I visited their Quantum Information group, led by Joonwoo. Specifically, thank you Ashutosh Rai, Karthik Mohan, Jiyoung Yun, Hyeokjea Kwon, Seungchan Seo and Jaemin Kim for making me feel like I was home. Special thanks to Jiheon Seong, with whom I travelled around the country, for showing me a fascinating culture and teaching me the bit of Korean I needed to survive on

my own.

I feel very fortunate for all new and interesting people I met during these years. Concretely, I am grateful for the discussions with Yelena Guryanova about the NPA hierarchy and for helping me develop my ideas. Thanks to Morten Kjaergaard and Jacob Hastrup for the very interesting discussions on a QRNG implementation in a superconducting platform, a proposal that branched out in many interesting ideas. Thanks also to Armin Tavakoli, for his friendly welcome when I visited him at Lund University in his new group, and for his guidance in potential side projects. In addition, thank you Antonio Acín, for welcoming me in your group as a visitor during part of my time in Barcelona.

Last but not least, I would like to thank Gemma, my love and light, for encouraging me to pursue my personal goals, even if that requires me to stay far from home for a long time. Thanks to my family, for their support and understanding that doing research is also considered a job. Finally, thanks to my companions and friends, Agustina and Facundo, for giving me the daily support I needed in Copenhagen, for cheering me up during the bad days and making good moments even better.

# List of publications

### Publications included in this thesis

**Carles Roch i Carceller**, Kieran Flatt, Hanwool Lee, Joonwoo Bae and Jonatan Bohr Brask "*Quantum vs Noncontextual Semi-Device-Independent Randomness Certification*", Phys. Rev. Lett. **129** 050501 (2022).

Kieran Flatt, Hanwool Lee, **Carles Roch i Carceller**, Jonatan Bohr Brask and Joonwoo Bae, "*Contextual Advantages and Certification for Maximum-Confidence Discrimination*", PRX Quantum **3** 030337 (2022).

Hanwool Lee, Kieran Flatt, **Carles Roch i Carceller**, Jonatan Bohr Brask and Joonwoo Bae, "*Maximum-confidence measurement for qubit states*", Phys. Rev. A **106** 032422 (2022).

**Carles Roch i Carceller**, Lucas Nunes Faria, Zheng-Hao Liu, Ulrik Lund Andersen, Jonas Schou Neergaard-Nielsen and Jonatan Bohr Brask, "*More than one bit of semi-device- independent randomness from a single qubit*", (In preparation).

**Carles Roch i Carceller** and Jonatan Bohr Brask, "*A contextuality witness inspired by optimal state discrimination*", (In preparation).

### Publications beyond the scope of this thesis

Yu-Xiang Zhang, **Carles Roch i Carceller**, Morten Kjaergaard and Anders Søndberg Sørensen, "*Charge-Noise Insensitive Chiral Photonic Interface for Waveguide Circuit QED*", Phys. Rev. Lett. **127**, 233601 (2021).

# Contents

# Chapter 1

# Introduction

In this chapter we introduce the two main blocks that build this thesis: quantum random number generation and quantum contextuality.

## 1.1 Quantum random number generation

Randomness is the quality that governs the outcome of events or the values taken by a variable of being unpredictable and lacking a pattern. It is a fundamental aspect of many physical and mathematical systems, and it is often used to model uncertainty or to generate unique values in a wide variety of applications. Randomness can be a powerful tool in algorithms for several reasons. It can help avoiding bias that might arise if a deterministic algorithm always makes the same choices. For example, if randomly shuffling a deck of cards, every possible ordering has equal chance of occurring. In computer science for instance, random numbers are used in cryptography to generate unpredictable keys [1]. Randomness is essential in cryptographic algorithms to generate secure keys and prevent unauthorized access to data. Without randomness, an attacker could potentially predict the key or guess through brute force. Randomness can also be used to simulate complex processes that involve unpredictable or stochastic behavior with Monte Carlo simulations [2] and statistical samplings [3, 4]. Random samplings are used to model real-world phenomena and make predictions based on observed data, like in weather forecast [5]. Beyond scientific uses, random sources of data can also play an important role in political issues as for achieving fairness in non-biased selections or sortition [6], as well as in gaming and gambling [7].

Let us illustrate a simple example where randomness can improve algorithms. Imagine you want to determine whether a certain number $n$ is prime

or composite. One could simply use brute force and try to divide it by all possible factors. This "trial and error" method however gets disturbingly slow, scaling exponentially bad in the number of digits. In terms of the number of bits $a$, it would take around $O(2^{||a||/2})$ computations. Also, in case the number $n$ turns out to be composite, such method would tell you the values of its divisors, which is more information than what we asked for. A more simple and efficient algorithm can be implemented using Fermat's *little theorem* [8–10]. The basic idea goes back to a result from Pierre de Ferman during the $17^{\text{th}}$ century. Consider two integers $n$ and $x$. Fermat proved that, if $n$ is a prime number, then $x^n - x$ is always a multiple of $n$, regardless of the value of $x$. In other words, the remainder when $x^{n-1}$ is divided by $n$ is always 1. One can turn Fermat's little theorem into a primality test: take a number $n$ of interest and pick $x$ at random. If $x^n - x$ is not a multiple of $n$, we know that $n$ must be definitely composite. Otherwise, $n$ is probably prime. Each computation of $x^n$ is $\sim O(||n||^2)$, turning the scalability polynomial.

Another example of how randomness can be used to improve algorithms is with random samplings. Monte Carlo simulations use random sampling to simulate complex processes that are difficult or impossible to solve analytically. Suppose that you wanted to estimate the value of $\pi$. You could do this by simulating the process of throwing darts at a circular target. Each dart would hit the target at a random point within the unit square. The ratio of darts that landed inside the circle to the total number of darts would be an estimate of the area of the circle divided by the area of the square, which is equal to $\pi/4$. If the darts are thrown uniformly random with respect to the area of the target, the estimated value of $\pi$ gets more accurate. However, if the quality of randomness is poor, i.e. the direction of the darts is biased towards a particular direction, the estimated value gets worse.

Thus, indeed, fast sources of randomness play an important role in many tasks in the modern world. Historically, the process of generating pseudo-random bits of information [11, 12] was carried out through deterministic formulas (random number generators) which relayed on an input, or *seed*, implying some correlation of these numbers and hence their predictability. We say apparently as for a particular seed, the outputted string of random bits is always the same. Pseudo-random number generators provide a poor level of security in that regard, since they are completely predictable if the seed is revealed. Another option is to use hardware-based random number generators. These generators use physical processes, such as electronic or radioactive noise, to generate random numbers. These generators are more secure that pseudo-random number generators because they rely on highly

unpredictable physical phenomena.

It was not until the arrival of quantum mechanics that true randomness was unveiled [13, 14]. Indeed, quantum mechanics provides a probabilistic description of nature. One of the main consequences of quantum mechanics is that certain physical phenomena are inherently unpredictable. For instance, the spin of a particle in a superposition state cannot be predicted with certainty before it is measured, and even then, the measurement is probabilistic. This unpredictability can be harnessed to generate random numbers. Quantum random number generators (QRNG) use quantum phenomena to generate true random numbers. These phenomena could be the measurement of photon polarisation or the detection event of atom decay. QRNGs work by measuring the state of a quantum system and using the randomness of the measurement to generate a random number. The hardware used to implement the generator must be carefully designed and calibrated to ensure that the measurement outcomes are truly random and unbiased. That specific setting, where QRNG relies on the specific hardware used in the implementation, is called device-dependent (DD).

Beyond DD-QRNG, one can generate truly random numbers without characterising any of the involved devices. That setting (called device-independent, DI) can be done in a way that is completely independent of the specific hardware used in the implementation. The idea behind DI-QRNGs is to harness nonlocality to generate completely unpredictable random numbers. Unlike traditional DD-QRNGs, DI-QRNGs are designed to be completely independent of the specifics of the quantum hardware used to implement them. This can be done considering setups with multiple, separate parties. The nonlocal behavior can be harnessed to certify true random measurement outcomes in all parties, even if their measurement devices are treated as completely uncharacterised black boxes. The setup is required to violate a Bell inequality [15]. This is, however, technologically very demanding, as the violation must be loophole free. Further approaches focus on semi-DI settings where, while some of the the involved devices are still black boxes, others are complemented with a few general assumptions, turning them into "gray" boxes. In semi-DI settings, nonlocality is not necessary and thus, more technologically feasible.

In this thesis we present two works on semi-DI randomness certification. In Ref. [16], we design a surprisingly simple protocol to generate more than one bit of certified randomness per round in a prepare-and-measure scenario by only measuring qubit states. In Ref. [17], we compare the randomness generation power of quantum mechanics and the so-called noncontextual

models, which can be used as a reference of classicality. In the following section we elaborate further on the concept of quantum contextuality and what we refer as noncontextual models.

## 1.2 Contextuality

Contextuality is a fundamental property of quantum mechanics that states that the measurement outcome of a quantum system depends not only on its state, but also on the context in which it is measured. This means that the measurement of a quantum system can produce different results depending on the set of compatible observables that are being measured simultaneously, or the set of quantum states being prepared prior to measurement. The notion of contextuality firstly emerged from a theorem established by John Bell [18], Simon Kochen and Ernst Specker [19]. There, it is established that it is impossible to fully describe reality with a *noncontextual* hidden-variable model. A context is defined, according to the Bell-Kochen-Specker theorem, as a commutative relation between the Hermitian operators describing observable quantities. Imagine you dispose of a physical system in a particular quantum state described by the density matrix $\rho$. Also, you dispose of three different measurements identified with the Hermitian operators $A$, $B$ and $C$. Observable $A$ is jointly measurable with $B$ and $C$, which means that $[A, B] = 0$ and $[A, C] = 0$. However, $B$ and $C$ are not jointly measurable, $[B, C] \neq 0$. Thus, it is impossible to measure all three observables together. You can, however, measure $A$ and $B$, or $A$ and $C$ together. In this sense we say that $B$ and $C$ are two distinct measurement contexts for the measurement $A$. This commutative property plays an important role in the repeatability of outcomes. According to the state-update rule, if you measure first with $A$, the state left after the measurement is $A\rho A$. Then, a subsequent measurement of $B$ will not affect the result of a latter measurement of $A$, such as in the sequence $ABA$. That is not true replacing $A$ with $C$, as in that case the result of a latter measurement of $C$ will depend whether $B$ was measured subsequently, such as $CBC$.

Let us see how, in a contextual theory such as quantum mechanics, one can distinguish between preparations and measurements with different contexts. Measuring $C$ yields essentially distinct observable statistics in the context whether if $B$ was measured ($CB\rho BC$) or not ($C\rho C$). In this framework, one can tentatively define a generalised notion of contextuality, or noncontextuality, in an operational manner. A model that reproduces a noncontextual behavior in state preparations, for instance, can be viewed as one in which

these are operationally equivalent if they cannot be distinguished by any possible measurement. In Ref. [20] an ontological operational model is introduced, which reproduces some observations obtained according to the quantum theory. In this model, all properties of a physical system in question are deterministically collected in an ontic state. However, preparations and measurements are described as probabilistic samplings over the ontic space. These are called *epistemic states* and *response functions* respectively. This probabilistic description is attributed solely to ignorance. Noncontextuality is implemented through the assumption that all indistinguishable preparations and measurements are operationally equivalent. This means that all preparations (measurements) which are indistinguishable by any measurement (preparation) must be described by the same epistemic state (response function).

Apart from Kochen-Spekker and generalised contextuality, other notions emerged during the recent years. The most well-known example are the so-called *nullifications of the Kochen-Specker theorem* by Meyer [21], Kent [22], Cliffton and Kent [23] and Barrett and Kent [24]. Their work holds great significance as it has served as a source of inspiration for the development of novel Kocken-Specker inequalities. Nonetheless, in keeping with the focus and aims of this thesis, we will refrain from providing a detailed analysis of their contributions.

Here, we center our study in generalised contextuality for preparations and measurements. We use an operational noncontextual model as a reference of classicality to find contextual advantages in state discrimination in Ref. [25] and randomness certification in Ref. [17]. Also, in Ref. [26], we elaborate a contextuality witness based on optimal two state discrimination.

## 1.3 Thesis structure

The present thesis consists of a collection of five works, each presented in a different self-contained chapter, with their own introduction, main body, conclusions and supplemental material (or appendices). These are preceded by an introductory chapter in qubit state discrimination. Specifically, the remainder of this thesis is organised as follows:

In Chapter 2 we introduce the main basics in quantum state discrimination for qubit states. This includes prepare-and-measure scenarios, the Bloch representation of qubit states, the (semi)-device independent treatment

and the principles of convex optimisation. In addition, we introduce minimum error state discrimination and unambiguous state discrimination, two well-known state discrimination protocols. All these tools are indispensable throughout all works presented in this thesis.

In Chapter 3 we present and study the protocol of maximum confidence state discrimination. There, we introduce the main goal of the protocol: the confidence. Roughly speaking, that is the probability that the prepared state was indeed the one indicated by the measurement device, given the measurement outcome. We continue arguing that maximum confidence discrimination can be reduced to minimum error and unambiguous state discrimination. A geometrical interpretation in terms of the Bloch sphere is given with various examples. We finish differentiating the behaviours of the so-called maximum confidence measurement in terms of the rates of observed events.

In Chapter 4 we study contextual advantages in state discrimination. Concretely, we center in maximum confidence state discrimination, which was already presented in detail in Chapter 3. We start reviewing the contextual advantages in minimum error discrimination, previously covered in Ref. [27]. We extend this results to unambiguous state discrimination, and later on to maximum confidence discrimination which we study in much more detail.

In Chapter 5 we continue studying prepare-and-measure scenarios to derive a contextuality witness which is based on optimal state discrimination. Here, we optimise the measurement based on a witness which is defined at the level of probabilities. We sample accordingly the correlation space parameterized by error and success probabilities, seeing that a maximum confidence measurement yields correlations on the boundary. Based on this observation, we claim that the most optimal measure coincides with maximum confidence state discrimination for the whole ensemble.

In Chapter 6 we compare quantum and noncontextual models in terms of randomness certification. We center in maximum confidence state discrimination for a particular state of the prepared ensemble. Our observations suggest that, within the quantum model we are able to certify more randomness than in a noncontextual one, as long as the adversary in both cases is quantum. There is an interesting result if one compares two different worlds instead. By a quantum or noncontextual world, we mean a scenario where both the prepare-and-measure model and the adversary are quantum or noncontextual, respectively. In our results, we see that a noncontextual world can certify more randomness than a quantum world.

In Chapter 7 we center in practical quantum randomness certification. We design an extremely simple quantum randomness certification protocol, based on semi-device independent state discrimination. Additionally, the protocol is implemented in the lab through an optical platform. Our experimental results report more than one bit of randomness per round where a single qubit is measured.

We end this thesis in Chapter 8, concluding all presented works with some remarks and pointing out potential future routes of investigation.

# Chapter 2

# Quantum state discrimination for qubit states

In this chapter we present the quantum mechanical background and formalism used throughout this thesis, which is central for quantum state discrimination. It is quite impossible to give a full treatment of this subject in one chapter, as a whole book would be needed for this task. For more detailed reviews see Refs. [28–33].

## 2.1 Fundamentals in quantum state discrimination

In the following we will see some fundamental aspects which are key in quantum state discrimination. First, we will investigate how we mathematically treat a state and a measurement in a so-called prepare-and-measure setup. Then, we will show how we can view all states and measurements in the useful representation of the Bloch sphere in a two-dimensional Hilbert space. We will continue presenting device and semi-device independent scenarios which are widely considered in related randomness certification and quantum cryptography protocols. We end this section by presenting the basic ideas in convex optimisation and semidefinite programming, a tool which is central in many of the works presented in this thesis.

### 2.1.1 Prepare-and-measure scenarios

#### Quantum states

Every physical theory attempts to explain a concrete spectrum of reality from observations and consequent interpretations. Quantum mechanics is a branch of physics that describes the behaviour of matter and energy at the atomic and sub-atomic level [1]. Physical systems can be prepared in states which can be then detected and measured. From textbook quantum mechanics, quantum states are represented as density matrices $\rho$. Let $\mathcal{B}(\mathcal{H})$ denote the algebra of linear operators acting on the Hilbert space $\mathcal{H}$, and let $\mathcal{D}(\mathcal{H}) \subset \mathcal{B}(\mathcal{H})$ denote the set of states. Then $\rho \in \mathcal{D}(\mathcal{H})$. These must satisfy a set of requirements to be considered valid. Any mathematically acceptable state operator must satisfy the following three conditions:

- Positive Semi-Definiteness (PSD): $\rho \geq 0$
- Normalisation: $\mathrm{Tr}\left[\rho\right] = 1$
- Hermiticity: $\rho = \rho^{\dagger}$,

for $\rho^{\dagger}$ denoting the complex conjugate and transpose operation. Being a self-adjoint operator, $\rho$ has a spectral representation in terms of its eigenvalues and orthonormal eigenvectors. The set of acceptable mathematical state operators forms a *convex set*. This means that if two or more density matrices satisfy the above conditions, then so does a convex combination of those. Convexity is a very useful property, as we will later see, in various optimisation problems which are essential in every state discrimination problem. Within the set of all possible states we find a very special class: *pure states*. A pure state operator can be represented as $\rho = |\psi\rangle\langle\psi|$, where the unit vector $|\psi\rangle$ is called *state vector*. A common feature in all pure states is that $\mathrm{Tr}\left[\rho^2\right] = 1$. Another remarkable property is that a pure state cannot be expressed as a nontrivial convex combination of other states. However, we can find states that can be described by statistical mixtures of pure states. These are called *mixed states*. The representation of a mixed state as a convex combination of pure states is never unique. One can see that by considering the following state $\rho_a = a\,|0\rangle\langle 0| + (1-a)\,|1\rangle\langle 1|$. By defining the two vectors $|\pm\rangle = \sqrt{a}\,|0\rangle \pm \sqrt{1-a}\,|1\rangle$ one re-write the state $\rho_a$ as

$$\rho_a = \frac{1}{2}\,|+\rangle\langle +| + \frac{1}{2}\,|-\rangle\langle -| \ . \tag{2.1}$$

[1] Recently, there has been experimental evidence of macroscopic objects showing quantum behaviors [34]. Thus, quantum mechanics might then not be uniquely describing microscopic phenomena.

This is of particular interest if $|0\rangle$ and $|1\rangle$ are considered two orthonormal states forming a basis on the Hilbert space. Then, $\rho_a$ is called *maximally-mixed state* if $a = 1/2$. Formally, in a D-dimensional Hilbert space, this is represented by

$$\frac{1}{D}\mathbb{1} := \frac{1}{D}\sum_{n=1}^{D} |n\rangle \langle n| \ , \tag{2.2}$$

for $\mathbb{1}$ being the *identity operator* and $|n\rangle$ forming an orthonormal basis.

### Quantum observables

In a real experiment, what we observe are not states but the outcomes of any measurement we perform on those states. In general, a pair of quantum states are not always perfectly distinguishable from each other. This motivates the question on how well can we discriminate such states with a particular measurement. In quantum mechanics, every observable corresponds to an operator $A \in \mathcal{B}(\mathcal{H})$ which is Hermitian, $A = A^\dagger$, and outcomes that are identified with its eigenvalues, $A = \sum_i \lambda_i P_i$. These projectors $P_i$ are identified as measurement effects which are related to the probability of observation through the Born rule [35]: $p(\lambda_i) = \mathrm{Tr}\,[\rho P_i]$.

The general scenarios to study quantum state discrimination are so called *prepare-and-measure* scenarios. These involve two devices: a *preparation device* and a *measurement device*. The first device receives an input $x \in X$ which determines the procedure for preparing a physical state represented by the density matrix $\rho_x$. This state is then sent to the second device, which will perform a measurement $M$ on it. From that measurement, an outcome event $b \in B$ will occur. We denote by $\hat{\pi}_b$ the probability operator associated to each measurement outcome $b$. To be considered valid probability operators, these must obey a list of constraints:

- Positive Semi-Definiteness (PSD): $\hat{\pi}_b \geq 0$, $\forall\ b \in B$.
- Normalisation: $\sum_{b\in B} \hat{\pi}_b = \mathbb{1}$.
- Hermiticity: $(\hat{\pi}_b)^\dagger = \hat{\pi}_b$, $\forall\ b \in B$.

The collection of probability operators $\hat{\pi}_b$ that meet these criteria are said to form a generalised measurement or a positive operator-valued measure (POVM). Roughly speaking, a POVM is a mathematical tool to express the

probabilities of the outcomes of a particular measurement strategy. In general, one can consider multiple measurement strategies available to the measurement device specified by an input $y$. Throughout this thesis, however, we consider protocols where the user can only choose a single measurement strategy and thus, we ignore any input the measurement device can receive. The goal in quantum state discrimination is to determine the optimal measurement strategy, which can be expressed mathematically as finding the best POVM. The optimality is defined based on a predetermined goal, usually represented as a linear combination of the observed probabilities $p(b|x)$. The mathematical tools used to describe the prepared states $\rho_x$ and measurements $\hat{\pi}_b$ allow for the calculation of the observed probabilities using the Born rule [35] as

$$p(b|x) = \mathrm{Tr}\left[\hat{\pi}_b \rho_x\right] \ . \tag{2.3}$$

The previous conditions applied to valid quantum states and POVMs ensure that $p(b|x)$ are valid probability distributions. Explicitly, PSD condition implies $p(b|x) \geq 0$, normalisation implies $\sum_b p(b|x) = 1$, $\forall x$ and Hermicity implies that $p(b|x)$ are real valued quantities. The collection of conditional probabilities are a very effective tool to represent the correlations reproducible by the quantum theory. Also, the goal and constraints in any state discrimination protocol is represented through these correlations, making them central in this thesis.

### 2.1.2 Qubit states and the Bloch sphere

Every physical system can be prepared in a particular state taken from an ensemble. The simplest example is a physical variable that can take two possible values, usually labeled as 0 or 1. The information carried by this binary variable is called a *bit*. A bit of information can be stored in any 2-level quantum system, whose basis states may be denoted as orthogonal vectors in the Hilbert space: $|0\rangle$ and $|1\rangle$. In addition, the superposition principle allows for a continuum of pure states of the form

$$|\psi\rangle = c_0 |0\rangle + c_1 |1\rangle \, , \tag{2.4}$$

where the complex coefficients must satisfy normalisation through $|c_0|^2 + |c_1|^2 = 1$. The state $|\psi\rangle$ is called *qubit*, since it carries information within the two amplitudes $c_0$ and $c_1$. This information is quantum in the sense that it is encoded in a state under the quantum superposition principle. All 2-dimensional Hilbert spaces are isomorphic, which means that we can describe any two-level quantum system through the qubit state (2.4). This is a powerful tool since this includes $\frac{1}{2}$ spin systems, polarized photons and

superconducting anharmonic two-energy levels among many other examples, which can be treated with the same mathematical formalism.

A qubit state can be associated with a three-dimensional vector $\vec{v} = (x, y, z)$ such that $0 \leq ||\vec{v}|| \leq 1$. Then, any qubit state operator can be expressed as

$$\rho = \frac{1}{2}\left(\mathbb{1} + \vec{v} \cdot \vec{\sigma}\right) \ , \tag{2.5}$$

where $\vec{\sigma} := (X, Y, Z)$ is the vector collecting all three Pauli matrices

$$X = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \quad Y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} \quad Z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \ . \tag{2.6}$$

One can use the association in (2.5) to draw a qubit state space in a three-dimensional space. All vectors in this space are confined in a unit-radius fictitious sphere, which is called the *Bloch sphere*. Any point on the surface of the Bloch sphere is associated to a pure qubit state. One can see that by computing

$$\mathrm{Tr}\left[\rho^2\right] = \frac{1}{2}\left(1 + ||\vec{v}||^2\right) \ . \tag{2.7}$$

Observe that $\mathrm{Tr}\left[\rho^2\right] = 1 \iff ||\vec{v}|| = 1$. On the other hand, a maximally mixed state is associated with a representation (2.5) with a zero vector $\vec{v} = 0$, which is located at the center of the sphere. The purity of a qubit state is related to the length of its *Bloch vector* $\vec{v}$.

We can further relate the Bloch sphere representation of pure states with their state vector representation. Consider the state

$$|\psi\rangle = \cos\frac{\theta}{2}|0\rangle + e^{i\varphi}\sin\frac{\theta}{2}|1\rangle \ . \tag{2.8}$$

The representation of this state in terms of the Bloch vector (see Fig. 2.1) is

$$\vec{v} = (\sin\theta\cos\varphi, \sin\theta\sin\varphi, \cos\theta) \ . \tag{2.9}$$

This representation, for pure states, is always unique. This means that every pure state has a single associated point on the surface of the Bloch sphere. This reflects the impossibility of expressing a pure state as a non-trivial convex combination of other states. On the Bloch sphere, a convex combination of two pure states can be interpreted as picking a spot on the

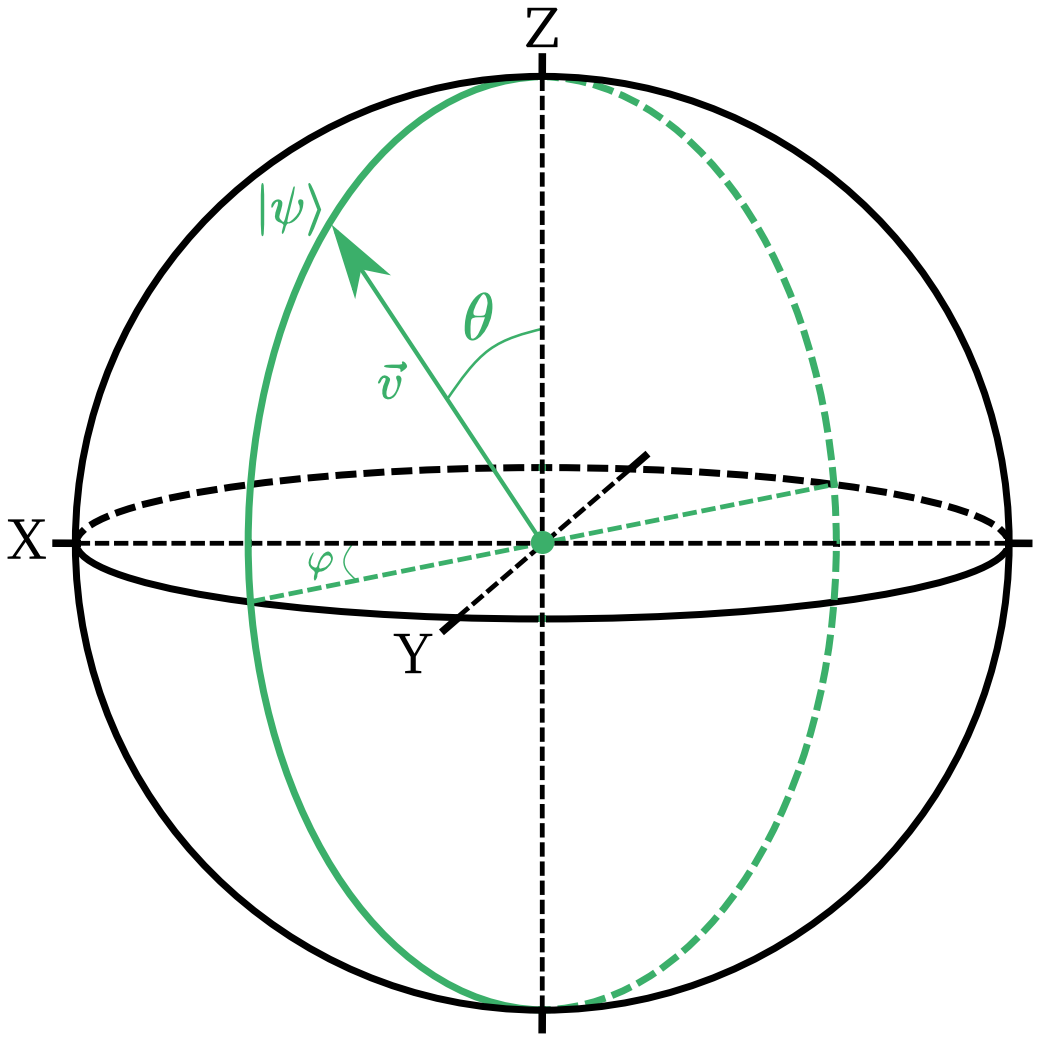



Figure 2.1: Representation of the generic pure qubit state in (2.8) on the Bloch sphere.

straight path connecting the two places on the sphere's surface that represent those states. The length between these points represent the weights of the convex combination. Then, every mixed state is associated with a single point inside the Bloch sphere. Due to the convexity of the set, any point on a line crossing the sphere will be inside the sphere. However, one can consider infinite lines crossing a single point inside the Bloch sphere. This reflects the fact that there are infinite different convex combinations of pure states to represent a single mixed state.

The Bloch sphere representation can be used to facilitate the interpretation of states, transformations and measurements in quantum state discrimination. Consider for instance, without loss of generality, the pair of pure states laying on the real plane of the Bloch sphere (that is the X-Z plane):

$$|\psi_0\rangle = \cos\frac{\theta}{2}|0\rangle + \sin\frac{\theta}{2}|1\rangle \quad \text{and} \quad |\psi_1\rangle = \cos\frac{\theta}{2}|0\rangle - \sin\frac{\theta}{2}|1\rangle \ . \tag{2.10}$$

The overlap of these two states is essentially characterised by the inner product of their Bloch vectors:

$$|\langle\psi_0|\psi_1\rangle|^2 = \cos^2\theta = \frac{1+\vec{v}_0\cdot\vec{v}_1}{2} \ . \tag{2.11}$$

A part from states, measurements can also be represented on the Bloch sphere. Consider a POVM $\{\hat{\pi}_b\}_{b=1}^n$ representing a measurement with $n$ different outcomes. Each POVM element has the following Bloch representation:

$$\hat{\pi}_b = \frac{R_b}{2}\left(\mathbb{1} + \vec{u}_b\cdot\vec{\sigma}\right) \ . \tag{2.12}$$

To be considered a valid POVM, the parameter $R_b$ and Bloch vectors $\vec{u}_b$ must satisfy

$$R_b \geq 0, \quad \sum_b R_b = 2 \quad \text{and} \quad \sum_b R_b\vec{u}_b = \vec{0} \ , \tag{2.13}$$

for $\vec{0}$ being a vector will zeros in all its entries. The Bloch representation of measurements is also a useful geometrical interpretation. Analogously to quantum states, POVM elements lying on the surface of the Bloch sphere (that is, with a unitary Bloch vector $||\vec{u}_b|| = 1$) represent rank-1 measurements. If, additionally, $R_b = 1$, the POVM element represents a projector onto a particular state.

The Bloch sphere representation turns out to be quite useful to represent states and measurements in a qubit space. Any state discrimination can be represented and better understood as such. It is thus worth it to mention the analogous Bloch representation for protocols involving states with higher dimension. For qutrit states for instance, the Pauli matrices for $SU(2)$ are naturally generalized by the Gell-Mann matrices in $SU(3)$ [36]. This representation has eight dimensions, making it difficult to imagine. In this thesis however, we will use only the Bloch sphere representation for qubit states, without the need of extending to higher dimensions.

### 2.1.3 Device independence and semi-device independence

Before getting into specific state discrimination protocols, we need to specify what assumptions we place on the involved devices. Considering the amount of trust we put in the devices involved in a prepare-and-measure scenario is a central aspect in quantum cryptography or randomness certification protocols. There, the aim is to generate a list of values to be kept in secret from any external agent, whom can indeed be manipulating our devices without us even being able to detect it. Assuming the involved devices to be untrusted empowers the security in quantum cryptography and the unpredictability in randomness certification protocols.

Let us start with the most straight-forward choice: a simple prepare-and-measure scenario where both preparation and measurement devices are fully characterized. This means that all state preparations and measurements are known, i.e. the state operators $\rho_x$ and POVM elements $\{\hat{\pi}_b\}$ are fixed. These schemes are often known as device-dependent (DD) protocols. DD protocols can be tailored to take advantage of the specific properties and capabilities of the devices involved. This flexibility allows for protocols well-suited to the available resources, potentially improving the efficiency and accuracy of the state discrimination task. Also, by utilising specific features of the devices, DD protocols can optimise the use of resources. However, the heavy reliability on the capabilities and characteristics of the devices places limitations in concrete implementations. Also, these are vulnerable to attacks that exploit weaknesses in the devices such as Trojan horse attacks or untrusted providers. In that regard, quantum behaviours such as nonlocality and superposition permit the design of protocols where the devices can be only partially characterized. Measurement-device independent protocols, for instance, are those where the measurement is completely unknown, but is

required to reproduce a set of observable statistics [37–42]. On the other hand, if the state preparations are the completely unknown part, with a fully revealed measurement, the protocol is said to be source-device independent [43–46]. One can draw a richer picture if both state preparation and measurement parts are fully unknown. These are called device independent (DI) protocols [47–53]. As a matter of fact, however, DI protocols can only be staged in mulipartite scenarios, where nonlocality enters into play as an essential ingredient. This is a very restrictive condition, since it is extremely difficult to experimentally build and maintain coherently entangled systems in the lab and break a Bell inequality in a loop-hole free manner [54–59]. What one can do in single-party prepare-and-measure scenarios is to partially uncover the preparation by bounding a concrete particularity of the prepared states. In that semi-DI setting, for example, one can bound the amount of energy that can be transmitted from preparation to measurement devices[60–64]. One can observe quantum correlations that depart form classical models with a rich structure [65]. Other semi-DI approaches make assumptions on the dimension of the Hilbert space [66, 67] and the overlap of the prepared states [17, 68]. In this thesis we will focus on the latter, although all three semi-DI approaches are inherently related.

### 2.1.4 Convex optimisation

The basic structure in every state discrimination problem can be framed as a convex optimisation problem [69]. Consider a vector space denoted by $X$. A convex subset $\mathcal{X} \in X$ is one where, for any two vectors $\vec{x}, \vec{y} \in \mathcal{X}$ and any scalar $t$ between 0 and 1, the vector $(1-t)\vec{x} + t\vec{y}$ is also contained in $\mathcal{X}$. Geometrically, that means that any line segment connecting two points in $\mathcal{X}$ lies entirely within $\mathcal{X}$. The goal of convex optimisation is to solve any problem of the form

$$\begin{aligned} \underset{\vec{x}}{\text{minimize}} \quad & f(\vec{x}) \\ \text{such that} \quad & \vec{x} \in \mathcal{X} \ , \end{aligned} \tag{2.14}$$

for $f$ being a convex function, i.e., $f\left((1-t)\vec{x} + t\vec{y}\right) \leq (1-t)f(\vec{x}) + tf(\vec{y})$, for $\vec{x}, \vec{y} \in \mathcal{X}$ and $0 \leq t \leq 1$. Any $\vec{x} \in X$ satisfying $\vec{x} \in \mathcal{X}$ is considered a *feasible point*. Convex optimisation problems are much easier to solve than general optimisation problems. Convex optimisation algorithms can find the global minimum of a convex objective function in polynomial time, which makes them very useful in many real-word applications. The set of feasible correlations involved in every quantum state discrimination scenario is a convex sub-space, making convex optimisation a central tool. One can go

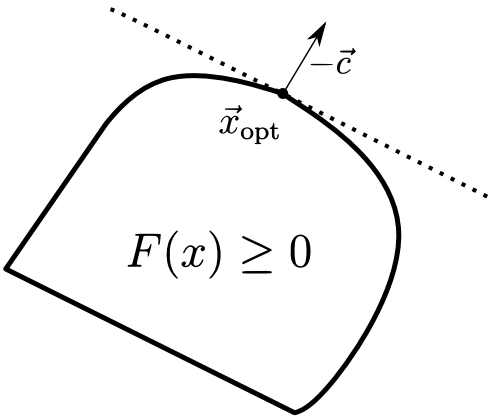



Figure 2.2: Illustration of a simple semidefinite program.

further and consider even simpler situations with optimisation problems with linear object functions and constraints. *Linear programming* is a branch of convex optimisation where $\mathcal{X}$ is a polytope (that is, a convex set defined by a finite number of inequalities), and $f$ is a linear function on the variables $\vec{x} \in X$ of the optimisation problem. The problems in linear programming are of the kind of

$$\begin{aligned} \underset{\vec{x}}{\text{minimize}} \quad & \vec{a} \cdot \vec{x} + \vec{b} \\ \text{such that} \quad & C\vec{x} \geq \vec{d}\,, \end{aligned} \tag{2.15}$$

for $\vec{a} \in X$, $\vec{b} \in X$, $\vec{d} \in X$ and the matrix $C$ being inputs to the problem. Linear programming is an essential instrumental tool to optimise hidden-variable correlations such as in local or noncontextual models [20]. These can serve in contextuality witnessing or nonlocality detection [53, 70].

In order to formally deal with discrimination tasks involving quantum states with quantum measurements, we need to use a more sophisticated tool, namely, *semidefinite programming* [30]. A semidefinite program (SDP) is a linear convex optimisation problem of the form

$$\begin{aligned} \underset{\vec{x}}{\text{minimize}} \quad & \vec{c}^{\,T} \cdot \vec{x} \\ \text{such that} \quad & F_0 + \sum_{i=1}^{m} x_i F_i \geq 0\ . \end{aligned} \tag{2.16}$$

The given data of the problem are $c \in \mathbb{R}^m$ and the $m+1$ symmetric matrices $F_0, \ldots, F_m \in \mathbb{R}^{n \times n}$. Beware the change of notation: if $M$ is a square

matrix, then $M \geq 0$ implies that $M$ is positive semidefinite (PSD). In fact, throughout this thesis any inequality relating two matrices implies a PSD condition. In Fig. 2.2 we depict a very simple semidefinite program. The boundary of the feasible region where available solutions exsist is delimited with a black curve. Roughly speaking, the goal in a semi-definite program is to move as far as possible in the direction $-\vec{c}$ while staying on the feasible region.

Semidefinite programming can be thought as an extension of linear programming where the element-wise inequalities between vectors are replaced with PSD constraints, i.e., matrix inequalities. These kind of problems fit quite well with state discrimination scenarios. States and measurements are also represented through square PSD matrices. Also, quantum correlations are computed through the Born rule (2.3) which is a linear function on the state or measurement whenever the other is given, making SDPs a very convenient tool.

There exist free solvers available to solve SDPs. These exploit the convex architecture of the problem to provide an approximate solution of the problem and also rigorous bounds on the precision of the solution with respect to its exact value. Throughout this thesis we make use of MOSEK [71]. We also use SCS [72] to find approximate solutions, with much less precision, only in the cases where MOSEK cannot find a solution. We do not work directly with these solvers, but through a general convex optimisation package provided in PYTHON [73]: CVXPY [74, 75]. The advantage of using CVXPY is that the user does not need to write the programs in the standard form (2.16). One can just indicate which are the semidefinite constraints, object function and declare the variables of the problem.

### Duality in semidefinite programming

The semidefinite program we introduced in (2.16) (which we will call from now on *primal*) can be reformulated into its *dual* version. To do so, let us first write the *Lagrangian* $\mathcal{L}$ associated with the primal problem

$$\begin{aligned} \mathcal{L} &= \sum_i c_i x_i - \mathrm{Tr}\left[\left(F_0 + \sum_i x_i F_i\right) Z\right] \\ &= \sum_i x_i \left(c_i - \mathrm{Tr}\left[F_i Z\right]\right) - \mathrm{Tr}\left[F_0 Z\right] \ , \end{aligned} \tag{2.17}$$

where $Z \in \mathbb{R}^{n\times n}$ is introduced as a Lagrange multiplier. The stationary points of $\mathcal{L}$ are obtained by nullifying the partial derivatives $\partial_{x_i}\mathcal{L} = 0$ which means that $c_i - \mathrm{Tr}\left[F_i Z\right] = 0$. The solution of the original constrained problem is

always a saddle point of $\mathcal{L}$, identified among the stationary points. We then introduce the associated dual problem as the following maximisation

$$\begin{aligned} \underset{Z}{\text{maximize}} \quad & -\operatorname{Tr}[F_0 Z] \\ \text{subject to} \quad & \operatorname{Tr}[F_i Z] = c_i, \ \forall i \\ & Z \geq 0 \ . \end{aligned} \tag{2.18}$$

The key property of the dual SDP is that it yields bounds on the optimal value of the primal SDP. Consider the difference between both object functions in the primal and dual formulations

$$\begin{aligned} \sum_i x_i c_i + \operatorname{Tr}[F_0 Z] &= \sum_i x_i \operatorname{Tr}[F_i Z] + \operatorname{Tr}[F_0 Z] \\ &= \operatorname{Tr}\left[\left(\sum_i x_i F_i + F_0\right) Z\right] \geq 0 \ , \end{aligned} \tag{2.19}$$

where in the last step we used the fact that $\operatorname{Tr}[AB] \geq 0$ if $A = A^T \geq 0$ and $B = B^T \geq 0$ [30]. Thus, $-\operatorname{Tr}[F_0 Z] \leq \vec{c}^{\,T} \cdot \vec{x}$, which means that the value of the object function of the dual SDP for any feasible $Z$ is smaller than or equal to the object function of the primal SDP.

Let $p^*$ be the optimal value of the primal SDP, and $d^*$ the optimal value of the dual SDP. We have that $p^* = d^*$ if the primal or dual SDPs is strictly feasible, i.e. there exist a set of variables satisfying the optimisation constraints [76]. Roughly speaking, the dual formulation of the SDP provides an analogous problem to reach the optimal solution from the primal SDP from the opposite direction. In other words, both primal and dual SDPs yield the exact same optimal solution (if strictly feasible) from above and below.

Other than a different perspective, the dual problem can also help providing with useful optimality conditions. Consider non-empty optimal sets of feasible solutions $\vec{x}$ and $Z$. From (2.19) we have that $\operatorname{Tr}[(\sum_i x_i F_i + F_0) Z] = 0$. Since $\sum_i x_i F_i + F_0 \geq 0$ and $Z \geq 0$, one has that both symmetric matrices are orthogonal. This is called the *complementary slackness* condition. Another optimality condition we already exposed as a constraint in the dual SDP is $\operatorname{Tr}[F_i Z] = c_i$, which is known as *Lagrange stability*.

Throughout most of all works presented in this thesis, duality in a semidefinite program is exploited in order to derive the optimality conditions, as well as many other features from which we can benefit. The dual formulation

in many cases presents a simpler problem computationally speaking, with dual variables smaller in size. We harness this fact, and use it to speed-up computations. Finally, in most optimisation problems derived in state discrimination, correlations enter in the constraints on the primal SDP, which in turn appear as an affine combination on the dual object function. We use this in Ref. [16] to account for deviations provoked by finite-size effects on the collected data from an experiment, directly plugging in the corrected correlations on the solution of the dual SDP.

In the following sections we will review two of the most studied quantum state discrimination protocols. We will look at the particular case the discrimination of two qubit states, show the POVM representation of the optimal measurements and the form of the observed conditional probabilities.

## 2.2 Minimum error state discrimination

Minimum error state discrimination (MESD) is a fundamental problem in quantum information science that deals with the task of identifying an unknown quantum state from a set of possible states. In this process, an observer (receiver) receives a quantum system in one of several known states and must determine which state it is, based on the measurement outcome. The challenge in MESD is to determine the optimal measurement strategy that minimizes the probability of making an error in the state discrimination task.

### 2.2.1 Two state MESD

Let us have a look at the special case where the prepared state is known to be one of the two pure states $|\psi_x\rangle \in \{|\psi_0\rangle, |\psi_1\rangle\}$, with a density matrix representation $\rho_x = |\psi_x\rangle\langle\psi_x|$, according to the associated prior probabilities $\{p_0, p_1 = 1 - p_0\}$. The outcomes of the measurement will be described by the POVM $\{\hat{\pi}_b\}$ for $b = 0, 1$. The averaged probability of successfully guessing the prepared state will then be expressed as

$$p_{\text{suc}} := p_0 p(0|0) + p_1 p(1|1) = p_0 \operatorname{Tr}[\hat{\pi}_0 \rho_0] + p_1 \operatorname{Tr}[\hat{\pi}_1 \rho_1] \ , \tag{2.20}$$

for $p(b|x)$ following the Born rule in described in (2.3).

On the other hand, the averaged probability of miss-identifying the prepared state, or error probability, is defined by

$$p_{\text{err}} := p_0 p(1|0) + p_1 p(0|1) = p_0 \operatorname{Tr}[\hat{\pi}_1 \rho_0] + p_1 \operatorname{Tr}[\hat{\pi}_0 \rho_1] \ . \tag{2.21}$$

Normalisation and positivity apply in the probabilities defined in (2.20) and (2.21). One can easily check that, if $\{\hat{\pi}_0, \hat{\pi}_1\}$ form a valid POVM and $\{\rho_0, \rho_1\}$ are valid density matrices, $p_{\text{suc}} + p_{\text{err}} = 1$, $p_{\text{suc}} \geq 0$ and $p_{\text{err}} \geq 0$. We want to find now the optimal pair $\{\hat{\pi}_0, \hat{\pi}_1\}$ that minimizes $p_{\text{err}}$. We can formally write down the problem as the following SDP:

$$\begin{aligned} \underset{\{\hat{\pi}_b\}}{\text{minimize}} \quad & p_{\text{err}} = p_0 \operatorname{Tr}\left[\hat{\pi}_1 \rho_0\right] + p_1 \operatorname{Tr}\left[\hat{\pi}_0 \rho_1\right] \\ \text{subject to:} \quad & \hat{\pi}_0 \geq 0, \ \hat{\pi}_1 \geq 0 \\ & \hat{\pi}_0 + \hat{\pi}_1 = \mathbb{1} \ . \end{aligned} \tag{2.22}$$

Thanks to the normalisation of the POVM, we can write the error probability in terms of a single element of the POVM as $p_{\text{err}} = p_0 - \operatorname{Tr}\left[\hat{\pi}_0 \left(p_0\rho_0 - p_1\rho_1\right)\right]$. The trace of any operator is invariant under unitary transformations. So, switching to the eigenbasis of the operator $p_0\rho_0 - p_1\rho_1$, the error probability will reach its minimum whenever $\hat{\pi}_0$ is a projector onto the eigenvector corresponding to the largest eigenvalue of $p_0\rho_0 - p_1\rho_1$. To see that, we can express both qubit states in terms of the Bloch vector $\vec{u}_x$ as $\rho_x = \frac{1}{2}\left[\mathbb{1} + \vec{u}_x\vec{\sigma}\right]$, where $\vec{\sigma} = (X, Y, Z)$ is a vector containing the Pauli matrices in (2.6). Let us consider the pair of states defined by the Bloch vectors $\vec{u}_0 = (\sin\phi, 0, \cos\phi)$ and $\vec{u}_1 = (-\sin\phi, 0, \cos\phi)$, where the angle $\phi$ determines the overlap $\langle\psi_0|\psi_1\rangle = \cos\phi$. The eigenvalues of the operator $p_0\rho_0 - p_1\rho_1$ in terms of the prior probabilities $p_x$ and the overlap between both states are

$$\lambda_\pm = \frac{1}{2}\left[p_0 - p_1 \pm \sqrt{1 - |\langle\psi_0|\psi_1\rangle|^2 \left(1 - (p_0 - p_1)^2\right)}\right] \ . \tag{2.23}$$

Obviously, the largest of these values is given by the one with the plus sign $\lambda_+$. Thus, at the end of the day the minimal attainable error probability, in the framework of two-state discrimination, yields

$$p_{\text{err}} \geq \frac{1}{2}\left[1 - \sqrt{1 - 4p_0p_1|\langle\psi_0|\psi_1\rangle|^2}\right] \ . \tag{2.24}$$

The lower bound in the error probability is commonly known as the Helstrom bound, named after Carl W. Helstrom, who first derived it in [77].

### 2.2.2 Example of implementation

The Helstrom measurement was first implemented in the laboratory in 1997 [78], using highly attenuated laser light where the two states are non-orthogonal polarisation states. Here we will show a very simple scheme to perform a Helstrom measurement using polarized photons to encode qubit

states.

Let $|\updownarrow\rangle$ and $|\leftrightarrow\rangle$ denote the mutually orthogonal vertical and horizontal polarization states, respectively. Consider a pair of polarized photon states parametrerised by the polarisation angle $\theta$:

$$|\psi_0\rangle = \cos\theta\,|\updownarrow\rangle + \sin\theta\,|\leftrightarrow\rangle \tag{2.25}$$
$$|\psi_1\rangle = \cos\theta\,|\updownarrow\rangle - \sin\theta\,|\leftrightarrow\rangle \ . \tag{2.26}$$

Each state is prepared with fixed equal probability $p_0 = p_1 = 1/2$. The overlap is fixed by the polarisation angle $\langle\psi_0|\psi_1\rangle = \cos 2\theta$. A Helstrom measurement can be implemented driving these states through a diagonal-polarisation beamsplitter (PBS) and putting photo-detectors at each end. Assuming that the PBS is set to filter diagonally polarised photons $(|\updownarrow\rangle + |\leftrightarrow\rangle)/\sqrt{2}$, the probability that the photo-detector at the other end of the PBS clicks is

$$\left|\left(\frac{\langle\updownarrow| + \langle\leftrightarrow|}{\sqrt{2}}\right)|\psi_0\rangle\right|^2 = \frac{1}{2}\left(1 + \sin 2\theta\right) \tag{2.27}$$
$$\left|\left(\frac{\langle\updownarrow| + \langle\leftrightarrow|}{\sqrt{2}}\right)|\psi_1\rangle\right|^2 = \frac{1}{2}\left(1 - \sin 2\theta\right) \ . \tag{2.28}$$

On the other hand, the probability that the other photo-detector clicks is

$$\left|\left(\frac{\langle\updownarrow| - \langle\leftrightarrow|}{\sqrt{2}}\right)|\psi_0\rangle\right|^2 = \frac{1}{2}\left(1 - \sin 2\theta\right) \tag{2.29}$$
$$\left|\left(\frac{\langle\updownarrow| - \langle\leftrightarrow|}{\sqrt{2}}\right)|\psi_1\rangle\right|^2 = \frac{1}{2}\left(1 + \sin 2\theta\right) \ . \tag{2.30}$$

Thus, by assigning the outcome $b = 0$ to the event that the photo-detector at the other end of the PBS clicks and $b = 1$ otherwise, one reaches the Helstrom bound.

## 2.3 Unambiguous state discrimination

Unambiguous state discrimination (USD) is a problem in quantum information science that deals with the task of identifying an unknown quantum state from an ensemble of possible states, without making any errors [79–81]. In contrast to the MESD, where some errors are allowed, the goal in USD is to find a measurement strategy such that the probability of a false positive or false negative is zero. The zero error condition introduces the possibility that

some measurement outcomes turn inconclusive. In other words, the observer must be able to determine with certainty the state of the quantum system, or else reject it if the state cannot be determined. The problem of USD was first introduced by Ivanovic in 1987 [79], where he provided an example of two non-orthogonal quantum states that can be perfectly distinguished without making an error, a task that was previously thought to be impossible.

### 2.3.1 Two state USD

Let us have a look at the case previously studied of the discrimination of two pure states $|\psi_x\rangle \in \{|\psi_0\rangle, |\psi_1\rangle\}$. Consider the success and error probabilities defined in (2.20) and (2.21) respectively. We aim to find a measurement able to determine the state of the prepared physical system without making any errors, i.e. $p_{\text{err}} = 0$. Normalisation constraints in this particular case imply $p_{\text{suc}} = 1$, which goes against the superposition principle of quantum mechanics for states with finite non-zero overlaps. This issue is addressed by including an additional possible outcome which we will label $b = \emptyset$, with its corresponding POVM element $\hat{\pi}_\emptyset$. This outcome will not provide any information about which physical state was prepared. Thus, we shall call it *inconclusive*. The probability of having an inconclusive outcome reads

$$p_\emptyset := p_0 p(\emptyset|0) + p_1 p(\emptyset|1) = p_0 \operatorname{Tr}\left[\hat{\pi}_\emptyset \rho_0\right] + p_1 \operatorname{Tr}\left[\hat{\pi}_\emptyset \rho_1\right] \ , \tag{2.31}$$

for $p(b|x)$ also following the Born rule in described in (2.3).

Normalisation and positivity are still applicable to the defined probabilities. It is easy to verify that, provided $\{\hat{\pi}_0, \hat{\pi}_1, \hat{\pi}_\emptyset\}$ constitute a legitimate POVM and $\{\rho_0, \rho_1\}$ are valid density matrices, the following conditions hold: $p_{\text{suc}} + p_{\text{err}} + p_\emptyset = 1$, $p_{\text{suc}} \geq 0$, $p_{\text{err}} \geq 0$, and $p_\emptyset \geq 0$. We now aim to find the optimal USD measurement which minimizes the probability $p_\emptyset$. We can formally write down the problem as the following SDP:

$$\begin{aligned} \underset{\{\hat{\pi}_b\}}{\text{minimize}} \quad & p_\emptyset = \operatorname{Tr}\left[\hat{\pi}_\emptyset \rho_0\right] + p_1 \operatorname{Tr}\left[\hat{\pi}_\emptyset \rho_1\right] \\ \text{subject to:} \quad & \hat{\pi}_b \geq 0 \ \forall b \\ & \sum_b \hat{\pi}_b = \mathbb{1} \\ & p_0 \operatorname{Tr}\left[\hat{\pi}_1 \rho_0\right] + p_1 \operatorname{Tr}\left[\hat{\pi}_0 \rho_1\right] = 0 \ . \end{aligned} \tag{2.32}$$

The main constraint in USD applies to POVM elements $\hat{\pi}_0$ and $\hat{\pi}_1$, such that $\operatorname{Tr}\left[\hat{\pi}_0 \rho_1\right] = \operatorname{Tr}\left[\hat{\pi}_1 \rho_0\right] = 0$. This condition can only be satisfied if $\hat{\pi}_0$ and $\hat{\pi}_1$ are rank-1, together with $\rho_0$ and $\rho_1$ being linearly independent pure states (also rank-1). Let $|\psi_x^\perp\rangle$ denote the orthogonal state to $|\psi_x\rangle$, i.e. $\langle\psi_x|\psi_x^\perp\rangle = 0$. In a

two-dimensional space spanned by the pure states $\{|\psi_0\rangle, |\psi_1\rangle\}$ this condition translates to $\hat{\pi}_0 = c_0 |\psi_1^\perp\rangle \langle\psi_1^\perp|$ and $\hat{\pi}_1 = c_1 |\psi_0^\perp\rangle \langle\psi_0^\perp|$, for $c_0$ and $c_1$ being real non-negative constants which we now aim to determine. Normalisation leaves the POVM element corresponding to the inconclusive outcome as $\hat{\pi}_\varnothing = \mathbb{1} - c_0 |\psi_1^\perp\rangle \langle\psi_1^\perp| - c_1 |\psi_0^\perp\rangle \langle\psi_0^\perp|$. With that, the probability of inconclusive events is

$$p_\varnothing = 1 - (1 - |\langle\psi_0|\psi_1\rangle|^2)(c_0 p_0 + c_1 p_1) \ . \tag{2.33}$$

To minimize $p_\varnothing$ it is clear that we seek maximal $c_0$ and $c_1$. However, the PSD constraint $\hat{\pi}_\varnothing \geq 0$ leaves the constants limited to fulfil the relation

$$c_0 c_1 (1 - |\langle\psi_0|\psi_1\rangle|^2) \geq c_0 + c_1 - 1 \ . \tag{2.34}$$

That places a tight bound on $c_0$ and $c_1$ that constitute the optimal USD measurement and thus, for that, we will only consider the equality. The following steps are in order: from (2.34) isolate $c_0$, substitute it in (2.33) and find a minimum by solving

$$\left.\frac{\mathrm{d}p_\varnothing}{\mathrm{d}c_1}\right|_{c_1^*} = (1 - |\langle\psi_0|\psi_1\rangle|^2) \left( \frac{p_0 |\langle\psi_0|\psi_1\rangle|^2}{\left[1 - c_1^*(1 - |\langle\psi_0|\psi_1\rangle|^2)\right]^2} + p_1 \right) = 0 \ . \tag{2.35}$$

This results in

$$c_1 \leq c_1^* = \frac{p_1 - |\langle\psi_0|\psi_1\rangle| \sqrt{p_0 p_1}}{p_1(1 - |\langle\psi_0|\psi_1\rangle|^2)} \underset{(2.34)}{\Longrightarrow} c_0 \leq \frac{p_0 - |\langle\psi_0|\psi_1\rangle| \sqrt{p_0 p_1}}{p_0(1 - |\langle\psi_0|\psi_1\rangle|^2)} \ . \tag{2.36}$$

One then can directly plug this bounds in (2.33) and see that the minimal attainable rate of inconclusive events, in the framework of USD, is

$$p_\varnothing \geq 2\sqrt{p_0 p_1} |\langle\psi_0|\psi_1\rangle| \ . \tag{2.37}$$

These results show that, in some cases, a POVM that cannot be reduced into standard projectors of the initial state onto orthogonal states may be preferable to a rank-1 POVM.

### 2.3.2 Example of implementation

The first experimental implementation of USD in the lab was in 2000, through a so called Ivanovic-Dieks-Peres (IDP) measurement [82]. There, two nonorthogonal linear polarization states of lights were discriminated using an optical interferometer to separate the appropriate components of light, manipulate them and recombine them in order to perform the desired

measurement. Here, let us consider an alternative implementation presented in Ref. [68], which employs an optical setup with time-bin encoded coherent states.

Consider a source of light which we can block with a black cover. This source is tuned to produce a beam with a coherent amplitude $\alpha$, so that a coherent state

$$|\alpha\rangle = e^{-|\alpha|^2/2} \sum_{n=0}^{\infty} \frac{\alpha^n}{\sqrt{n!}} |n\rangle \tag{2.38}$$

is prepared. If we block the light instead, the prepared state would be nothing else than vacuum $|0\rangle$. At each round, we prepare two pulses of light separated in two different time-bins which we call *early* and *late*. At each pulse we can choose whether we block or not the light. With this toolbox, let us consider the two possible prepared states $|\psi_x\rangle$:

$$|\psi_0\rangle = |\alpha\rangle_E \otimes |0\rangle_L \qquad |\psi_1\rangle = |0\rangle_E \otimes |\alpha\rangle_L \ . \tag{2.39}$$

We use the tensor product notation "$\otimes$" to differentiate between states prepared on the early "$|\cdot\rangle_E$" or late "$|\cdot\rangle_L$" time-bins. For $x = 0$, we let light pass through during the early pulse but block it during the late pulse. For $x = 1$ we do the opposite. The distinguishability of these two states is characterized by their overlap: $\langle\psi_0|\psi_1\rangle = e^{-|\alpha|^2}$. To perform USD, we consider the measurement which simply consists of a single photo-detector at the end of the light-beam. Whenever the photo-detector receives a photon, it will emit a signal which will be registered in a set of collected data. We call this event a "click" of the photo detector. To see that this simple measurement unambiguously discriminates the prepared states, note that the only preparation that could trigger a click on the first time-bin is $|\psi_0\rangle$. The same occurs for a click on the second time-bin which is only possible if $|\psi_1\rangle$ is prepared. In that regard, let us label the measurement events $b$ with $b = 0$ corresponding to the outcome when the detector clicks during the early time-bin, and $b = 1$ when it clicks during the late time-bin. Thus, the error conditional probabilities $p(b = 0|x = 1) = p(b = 1|x = 0) = 0$ are null, a necessary condition in USD. Moreover, coherent states have a non-zero support on the vacuum, which implies that in some rounds we will not see clicks on the detector. Whenever the detector does not click, we cannot certify which state was prepared. We label the no-click events with $b = \emptyset$, implying that we denote them as inconclusive.

To calculate the probabilities of each event we need to operationally describe the measurement events. Let us use the decomposition of the identity on the photon-number basis to identify each possible event. Also, let us separate each time-bin event in two different Hilbert spaces.

$$\begin{aligned}
\mathbb{1}_E \otimes \mathbb{1}_L =& \underbrace{\left(\sum_{n=0}^{\infty} |n\rangle\langle n|\right)}_{\text{Early bin}} \otimes \underbrace{\left(\sum_{m=0}^{\infty} |m\rangle\langle m|\right)}_{\text{Late bin}} \\
=& \left(|0\rangle\langle 0| + \sum_{n=1}^{\infty} |n\rangle\langle n|\right) \otimes \left(|0\rangle\langle 0| + \sum_{m=1}^{\infty} |m\rangle\langle m|\right) \\
=& \underbrace{|0\rangle\langle 0| \otimes |0\rangle\langle 0|}_{\text{Detectors don't click}} + \underbrace{|0\rangle\langle 0| \otimes \sum_{m=1}^{\infty} |m\rangle\langle m|}_{\text{Click on late bin}} + \underbrace{\sum_{n=1}^{\infty} |n\rangle\langle n| \otimes |0\rangle\langle 0|}_{\text{Click on early bin}} \\
&+ \underbrace{\sum_{n=1}^{\infty} |n\rangle\langle n| \otimes \sum_{m=1}^{\infty} |m\rangle\langle m|}_{\text{Click on both time-bins}} .
\end{aligned} \tag{2.40}$$

The last event we identity from the decomposition is the one corresponding to the detector clicking at both time-bins. This event is orthogonal to both states $|\psi_x\rangle$ in (2.39). Thus, we will not consider it. The probabilities $p(b|x)$ corresponding to the other events can be computed as follows. The success probabilities are

$$\begin{aligned}
p(0|0) =& \langle\psi_0| \left(\sum_{n=1}^{\infty} |n\rangle\langle n| \otimes |0\rangle\langle 0|\right) |\psi_0\rangle = \sum_{n=1}^{\infty} |\langle\alpha|n\rangle|^2 \\
=& \sum_{n=1}^{\infty} |e^{-|\alpha|^2/2} \sum_{n'=0}^{\infty} \frac{\alpha^{n'}}{\sqrt{n'!}} \langle n'|n\rangle|^2 \\
=& e^{-|\alpha|^2} \sum_{n=1}^{\infty} \frac{\alpha^{2n}}{n!} = e^{-|\alpha|^2} \left(e^{|\alpha|^2} - 1\right) = 1 - e^{-|\alpha|^2}
\end{aligned} \tag{2.41}$$

and

$$\begin{aligned} p(1|1) = &\langle\psi_1| \left( |0\rangle\langle 0| \otimes \sum_{m=1}^{\infty} |m\rangle\langle m| \right) |\psi_1\rangle = \sum_{m=1}^{\infty} |\langle\alpha|m\rangle|^2 \\ = &\sum_{m=1}^{\infty} |e^{-|\alpha|^2/2} \sum_{n'=0}^{\infty} \frac{\alpha^{n'}}{\sqrt{n'!}} \langle n'|m\rangle|^2 \\ = &e^{-|\alpha|^2} \sum_{m=1}^{\infty} \frac{\alpha^{2m}}{m!} = e^{-|\alpha|^2} \left( e^{|\alpha|^2} - 1 \right) = 1 - e^{-|\alpha|^2} \ . \end{aligned} \tag{2.42}$$

The probabilities of having an inconclusive event are

$$p(\emptyset|x) = \langle\psi_x| \left( |0\rangle\langle 0| \otimes |0\rangle\langle 0| \right) |\psi_x\rangle = |\langle\alpha|0\rangle|^2 = e^{-|\alpha|^2} \ . \tag{2.43}$$

We see that with this implementation we match the minimal attainable rate of inconclusive events according to optimal USD: $p(\emptyset|x) = |\langle\psi_0|\psi_1\rangle|$, which we derived in (2.37).

# Chapter 3

# Maximum confidence measurement for qubit states

In this chapter we present the results in "Maximum confidence measurement for qubit states" [83], authored by Hanwool Lee, Kieran Flatt, Carles Roch i Carceller, Jonatan Bohr Brask and Joonwoo Bae. This work was published in Physical Review A.

## 3.1 Abstract

In quantum state discrimination, one aims to identify unknown states from a given ensemble by performing measurements. Different strategies such as minimum error state discrimination or unambiguous state identification find different optimal measurements. Maximum confidence measurements (MCMs) maximise the confidence with which inputs can be identified given the measurement outcomes. This unifies a range of discrimination strategies including minimum error and unambiguous state identification, which can be understood as limiting cases of MCM. In this work, we investigate MCMs for general ensembles of qubit states. We present a method for finding MCMs for qubit-state ensembles by exploiting their geometry, and apply it to several interesting cases, including ensembles of two and four mixed states and ensembles of an arbitrary number of pure states. We also compare MCMs to minimum error and unambiguous discrimination for qubits. Our results provide interpretations of various qubit measurements in terms of MCMs and can be used to devise qubit protocols.

## 3.2 Introduction

One fundamental difference between classical and quantum physics is that, while all information about the physical state of a quantum system is captured by its quantum state, such states are in general not perfectly distinguishable. Specifically, no measurement can perfectly discriminate nonorthogonal quantum states. This is closely related to other fundamental results in quantum mechanics such as the impossibility of perfectly copying quantum states [84] and of faster-than-light signaling [85]. The limits to discriminating between quantum states have numerous applications in quantum information science. Such limits are key to the security of quantum key distribution [86, 87]; near-optimal state discrimination enables approximate quantum error correction [88]. They are also useful for operationally interpreting the differences between separable and entangled states [89, 90] (see also [91, 92]). For further examples of the wide impact of quantum state discrimination, see the related reviews Refs. [29, 92–97].

If it is impossible to perfectly discriminate quantum states, the natural thing to ask is precisely how well one can. This in turn introduces the need for different figures of merit, corresponding to variations of the discrimination task. In general, the task consists in identifying states drawn from some ensemble, given a single copy of the state and prior knowledge of the possible states. Two well-studied cases are minimum error state discrimination (MESD) and unambiguous state discrimination (USD). In MESD, one aims to minimize the probability that the state is misidentified while forbidding inconclusive outcomes [77, 98, 99]. In USD, one instead enforces that the state is never misidentified, at the price of allowing for a nonzero inconclusive-outcome rate, which one then aims to minimize [79–81]. Both MESD and USD are naturally formulated as statements about the conditional probabilities for observing certain outcomes, given that particular states were prepared.

Interestingly, distinct figures of merits in quantum state discrimination can be rephrased in terms of predictive and retrodictive formulations of quantum probabilities [100]. Predictive probabilities are probabilities of future events conditioned on past events, which, in this context, are the probabilities of the outcomes conditioned on the input states. Retrodictive probabilities are probabilities of past events conditioned on future events occurring; here this means the probabilities, conditioned on the observed outcomes, that particular input states were prepared. Predictive and retrodictive probabilities can be linked via Bayes' theorem.

In this work we focus on maximum confidence discrimination, which is most naturally formulated in the retrodictive picture. The figure of merit here is the confidence, defined as the conditional probability that an input was prepared given that the corresponding outcome was observed. A maximum confidence measurement (MCM) is a measurement strategy which achieves the best possible confidence. Maximum confidence measurements were introduced in Ref. [101]. They unify the MESD and USD settings of state discrimination. In particular, MCMs implement USD whenever USD is possible for the given ensemble and MESD if a zero inconclusive rate is enforced and the maximum confidence considers an ensemble itself. In general, they make optimal use of detection events for guessing which states were prepared in the past [102–106].

We investigate MCMs for qubit states and determine general relations between a given ensemble and its MCM. We present a method for finding MCMs by exploiting the geometry of the Bloch sphere directly, without reference to the algebraic optimization problem, in a similar manner to geometric schemes for MESD of $n$ qubit states [107–109]. We then consider several particular ensembles of qubit states, derive their MCMs, and also compare to MESD and USD.

The article is structured as follows. In Sec. 3.3 we start by briefly recalling the state discrimination problem in the simplest case of two pure states and results for optimal MESD and USD. In Sec. 3.4 we summarize MCMs. In Sec. 3.5 we formulate the problem of identifying an optimal MCM for qubits as a semidefinite program and present optimality conditions. The relations between state ensembles and MCMs are found by exploiting the Bloch sphere geometry. In Sec. 3.6 various ensembles of qubit-state ensembles are considered and their MCMs are explicitly derived. We consider two mixed states, geometrically uniform states, tetrahedron states, and asymmetric states. In Sec. 3.7 we derive the bounds on the observed rates of events which delimit different behaviours for MCMs. In Sec. 3.8 we summarize.

## 3.3 MESD and USD for two pure states

Let us consider the simplest non-trivial ensemble, consisting of two pure states, $|\psi_0\rangle$ and $|\psi_1\rangle$, generated with *a priori* probabilities $p_0$ and $p_1$, respectively. A measurement device receives state $|\psi_x\rangle$ with $x \in \{0, 1\}$, drawn from this ensemble, and provides an output $b \in \{0, 1, \emptyset\}$. The output can be understood as a guess for what input was prepared, i.e. for the value of $x$, with $b = \emptyset$ denoting inconclusive outcomes. One can thus define an average error rate

and an inconclusive rate, respectively, as

$$p_{\text{err}} = p_0 p(b = 1|x = 0) + p_1 p(b = 0|x = 1), \tag{3.1}$$

and

$$p_{\text{ø}} = p_0 p(b = \text{ø}|x = 0) + p_1 p(b = \text{ø}|x = 1), \tag{3.2}$$

where $p(b|x)$ denotes the conditional probability of observing outcome $b$ given input $x$.

In MESD, the goal is to minimise $p_{\text{err}}$ under the constraint that no inconclusive outcomes occur, i.e., $p(b = \text{ø}|x = 0) = p(b = \text{ø}|x = 1) = 0$. In this case, the minimal error rate is known as the Helstrom bound [77, 98, 99]

$$p_{\text{err}} = \frac{1}{2} - \frac{1}{2}\|p_0 |\psi_0\rangle \langle\psi_0| - p_1 |\psi_1\rangle \langle\psi_1| \|_1, \tag{3.3}$$

where $\| \cdot \|_1$ denotes the trace distance.

This result applies to an arbitrary pair of quantum states and is found by a measurement with a construction as follows. As it is shown in (3.3), the optimal measurement can be found in the support of given states $|\psi_0\rangle$ and $|\psi_1\rangle$. Then, two optimal positive-positive-operator-valued-measure (POVM) elements $\hat{\pi}_0$ and $\hat{\pi}_1$ are found as projectors with positive and negative eigenvalues of the operator $(p_0 |\psi_0\rangle \langle\psi_0| - p_1 |\psi_1\rangle \langle\psi_1|)$.

One can also notice that, independently to a dimension of a Hilbert space where two states can be described, the two-state discrimination problem can be reduced to a two-dimensional space spanned by $|\psi_0\rangle$ and $|\psi_1\rangle$. In this sense, the two-state problem is equivalent to discrimination of two qubit states. Then, by referring to a Bloch sphere, an optimal measurement with POVM elements $\hat{\pi}_0$ and $\hat{\pi}_1$ can be found in a diameter of a half-plane due to the completeness, i.e., $\hat{\pi}_0 + \hat{\pi}_1 = \mathbb{1}$. The Helstrom bound in (3.3) clarifies that the diameter should be parallel to the difference $(p_0 |\psi_0\rangle \langle\psi_0| - p_1 |\psi_1\rangle \langle\psi_1|)$.

In USD, on the other hand, the goal is to minimise $p_{\text{ø}}$ under the constraint that no errors occur, i.e., $p(b = 1|x = 0) = p(b = 0|x = 1) = 0$. In this case, the minimal inconclusive rate is

$$p_{\text{ø}} = 2\sqrt{p_0 p_1}| \langle\psi_0|\psi_1\rangle |. \tag{3.4}$$

If one hopes to be certain about which state was prepared, it suffices to rule out the other option. If one measurement outcome is $|\psi_0^{\perp}\rangle$, such that

$\langle\psi_0|\psi_0^\perp\rangle = 0$, then that outcome can never occur when $|\psi_0\rangle$ is measured. This means that the prepared state must have been $|\psi_1\rangle$. The same holds for the other state, so that the POVM must include among its elements the two states orthogonal to those in the ensemble. A measurement consisting of just those outcomes, however, will not be complete, and so the POVM must be completed by a third element, which is the inconclusive one. Each of the elements must be weighted by constant factors and that associated with the third outcome determines the inconclusive rate. It is thus minimised. In this manner, the rate (3.4) is attained [29].

## 3.4 Maximum confidence measurement

We now turn to the more general case of discriminating between an arbitrary number of states. Let $S$ denote an ensemble of quantum states in which the states $\rho_x$ are generated with a priori probabilities $p_x$:

$$S = \{p_x, \rho_x\}_{x=0}^{n-1}, \quad \text{and} \quad \rho = \sum_{x=0}^{n-1} p_x \rho_x. \tag{3.5}$$

The most general measurement corresponds to an $n+1$-outcome POVM, denoted by $\{\hat{\pi}_b\}_{b=0}^{n-1,\varnothing}$, where outcome $b = \varnothing$ collects inconclusive events, and the rest $b = 0, \cdots, n-1$ denotes the guess that the input $\rho_{x=b}$ was prepared.

Let $\rho_x$ denote the state of particular interest in the ensemble. The probability that the correct state is identified is the confidence associated with the measurement [101],

$$C_x := p_{x|b}(x|x) = \frac{p_x p_{b|x}(x|x)}{\eta_x} = \frac{p_x \operatorname{Tr}[\rho_x \hat{\pi}_x]}{\operatorname{Tr}[\rho \hat{\pi}_x]}, \tag{3.6}$$

where Bayes' theorem is applied, and $\eta_b = \operatorname{Tr}[\rho\hat{\pi}_b]$ is the rate with which the outcome $b$ is triggered. For example, $C_x = 1$ for some $x$ signifies unambiguous identification of the state $\rho_x$ by a detection event on $\hat{\pi}_x$. Given a detection event, a state $\rho_x$ is verified with certainty. Unambiguous discrimination of quantum states is achieved when $C_x = 1$ for all $x$. In (3.6) we explicitly show the sub-indices in the conditional probabilities for clarity, but we will drop them from now on.

The confidence in (3.6) can be maximised by optimising over each POVM element according to

$$\max C_x = \max_{\hat{\pi}_x} \frac{p_x \operatorname{Tr}[\rho_x \hat{\pi}_x]}{\operatorname{Tr}[\rho \hat{\pi}_x]}, \tag{3.7}$$

where $0 \leq \hat{\pi}_x \leq 1$. A valid POVM, which attains the optimum for all $x$, can always be obtained by re-scaling the $\hat{\pi}_x$ and including one additional element $\hat{\pi}_\emptyset$ which collects inconclusive outcomes. Such a measurement is called an MCM. In general, we have $\hat{\pi}_\emptyset \neq 0$.

As mentioned, when unambiguous discrimination is possible for an ensemble, the MCM is identical to the measurement giving unambiguous discrimination. An MCM for an ensemble of two pure states, for instance, will identify each state with perfect confidence. Note, however, that an MCM can be introduced for ensembles for which unambiguous discrimination is impossible, such as three-qubit states.

One may consider the maximum confidence for an ensemble itself: denoting by $\eta_x$ the probability that a detector $\hat{\pi}_b$ shows a detection event, i.e., $\eta_b = \mathrm{Tr}\left[\rho\hat{\pi}_b\right]$, the maximization

$$\max \sum_{x=0}^{n-1} \eta_x C_x \tag{3.8}$$

over a complete measurement equals the highest success probability in minimum error state discrimination [29]. We remark that an MCM provides a unifying picture of different figures of merits in quantum state discrimination.

The maximisation of $C_x$ in (3.6), which is computationally feasible, is greatly facilitated by the ansatz [101]

$$\hat{\pi}_b = \eta_b \rho^{-1/2} \hat{Q}_b \rho^{-1/2} \quad \text{for} \quad \hat{Q}_b \geq 0,\ \mathrm{Tr}\left[\hat{Q}_b\right] = 1\ . \tag{3.9}$$

We write explicitly

$$\max_{\{\hat{\pi}_b\}_{b=0}^{n-1,\emptyset}} C_x = \max_{\left\{\hat{Q}_b\right\}_{b=0}^{n}} \mathrm{Tr}\left[\tilde{\rho}_x \hat{Q}_x\right]\ , \tag{3.10}$$

for $\tilde{\rho}_x = \rho^{-1/2} p_x \rho_x \rho^{-1/2}$. The confidence $C_x$ is therefore maximised if $\hat{Q}_x$ is a projector onto the eigenstate of $\tilde{\rho}_x$ with largest eigenvalue, i.e.

$$\max C_x = ||\rho^{-1/2} p_x \rho_x \rho^{-1/2}||_{\mathrm{op}}\ , \tag{3.11}$$

where $||\cdot||_{\mathrm{op}}$ denotes the operator norm $||A||_{\mathrm{op}} = \sup_{||v||=1} ||Av||$. Once an optimal operator in (3.10), denoted by $Q_x^*$, is obtained, an optimal POVM element $\hat{\pi}_x^*$ is found as (3.9).

## 3.5 MCM for qubit states

In this section we approach the maximum confidence in (3.7) from the point of view of convex optimisation. We first show a semidefinite program (SDP) for the optimisation problem and then analyze the optimality conditions in order to show that a general structure relates the states to their MCM.

### 3.5.1 Convex optimisation

We begin with the maximisation problem in (3.10) which is linear with respect to a state of interest. The optimization problem can be written as an SDP as

$$p^* = \underset{\hat{Q}_x}{\text{maximize}} \quad \text{Tr}\left[\tilde{\rho}_x \hat{Q}_x\right] \tag{3.12}$$

$$\text{such that} \quad \hat{Q}_x \geq 0,\ \text{Tr}\left[\hat{Q}_x\right] = 1. \tag{3.13}$$

Its dual problem is found by constructing the Lagrangian

$$\mathcal{L}\left(Q_b, \lambda_b, Z_b\right) = \text{Tr}\left[\tilde{\rho}_x \hat{Q}_b\right] \delta_{b,x} + \lambda_b \left(1 - \text{Tr}\left[\hat{Q}_b\right]\right) + \text{Tr}\left[\hat{Q}_b Z_b\right] \ . \tag{3.14}$$

We introduced the dual variables $\lambda_b$ and $Z_b \geq 0$ for $b = 0, 1, \ldots n$, corresponding to the trace-one and PSD constraints for $\hat{Q}_b$, respectively. Finding the supremum of this Lagrangian gives the dual function

$$\begin{aligned} \mathcal{S}\left(\lambda_b, Z_b\right) &= \sup_{\hat{Q}_b} \mathcal{L}\left(\hat{Q}_b, \lambda_b, Z_b\right) \\ &= \lambda_b + \sup_{\hat{Q}_b} \text{Tr}\left[\left(\tilde{\rho}_x \delta_{b,x} - \lambda_b \mathbb{1} + Z_b\right) \hat{Q}_b\right] \ . \end{aligned} \tag{3.15}$$

The supremum (3.15) will diverge unless the last term vanishes. Thus, the dual parameters must satisfy the condition $\tilde{\rho}_x \delta_{b,x} - \lambda_b^* \mathbb{1} + Z_b^* = 0$, which is called Lagrangian Stability (LS) [110]. Another condition that optimal parameters satisfy is the Complementary Slackness (CS) [111], given as $\text{Tr}\left[Z_b^* \hat{Q}_b^*\right] = 0$. With the CS condition, LS enters as an inequality constraint in the dual problem. We are now ready to formally state the dual problem as the following minimisation problem:

$$d^* = \underset{\lambda_b}{\text{minimize}} \quad \lambda_b \tag{3.16}$$

$$\text{subject to} \quad \lambda_b \rho - \delta_{b,x} p_x \rho_x \geq 0 \ .$$

To this point, since the confidence is maximised for a particular state of interest $\rho_x$, we will only focus on the outcome $b = x$ identifying that particular

state. In general, it holds that the optimal primal solution is greater or equal than the dual. The equality holds when the problem is strictly feasible. In the present case, both primal (3.10) and dual (3.16) are simultaneously feasible, and one can find the maximum confidence $C_x$ from both ($p^* = d^* = \max C_x$).

Both the LS and CS conditions must be fulfilled by the parameters that yield the optimal solution of the problem. One can see that by using a Linear Complementarity Problem (LCP) [112] which is commonly used to understand better the convex optimisation structure of the problem. Technically speaking, while the optimisation problem, either primal or dual, includes inequality constraints, an LCP directly analyses the optimality conditions, which are in form of equalities. The LCP is given by the LS and CS conditions, and those parameters satisfying these equalities will automatically find an optimal solution.

We write again the LCP conditions but first, we introduce a new parameter $r_x \geq 0$ and a complementary state $\sigma_x$ such that $r_x\sigma_x = \sqrt{\rho}Z_x\sqrt{\rho}$. The reason why we call $\sigma_x$ complementary will be apparent later. With these new parameters, the optimality conditions turn

$$\text{Lagrange stability:} \quad \lambda_x\rho = q_x\rho_x + r_x\sigma_x \tag{3.17}$$

$$\text{Complementary slackness:} \quad r_x \operatorname{Tr}[\sigma_x\hat{\pi}_x] = 0\ . \tag{3.18}$$

Since both primal and dual problems are feasible, those primal and dual parameters satisfying (4.62) and (4.63) automatically pinpoint the optimization problem's solution. Once dual parameters are found from (4.62), the optimal POVM element is characterized by (4.63). Note that an optimal POVM element is found by the equalities given in the optimality conditions.

### 3.5.2 MCM for qubit states

We now investigate the optimality conditions for qubit states and show how one can solve the optimization problem directly. Both the maximum confidence and optimal POVM elements can be found. Let us begin with the condition in (4.63). The product of a complementary state $\sigma_x$ and an optimal POVM must be zero. Since the optimal measurement satisfies $\hat{\pi}_x \neq 0$ for $\forall x = 0, \cdots, n-1$, it holds that both $\sigma_x$ and $\hat{\pi}_x$ must be rank-1 and orthogonal with each other.

Let us consider the Lagrangian stability in (4.62), which can be rewritten

for all $x = 0, \ldots, n-1$ as

$$\rho = \mu_x \rho_x + (1 - \mu_x)\sigma_x \ , \tag{3.19}$$

where $\mu_x = \frac{q_x}{\lambda_x}$. Note that decompositions above for qubit states have been also obtained in Ref. [105]. MCMs can be computed analytically from this relation, which implies

$$\mathrm{Tr}\left[\sigma_x^2\right] = \frac{\mathrm{Tr}\left[(\rho - \mu_x\rho_x)^2\right]}{(1-\mu_x)^2} \ . \tag{3.20}$$

For qubit states, the complementary state that fulfills the CS condition in (3.18) is rank-1. This leaves the left-hand side equal to 1 and then, one can analytically find $\mu_x$ knowing the whole ensemble $\rho$ and the state of interest $\rho_x$. Suppose that the state of interest $\rho_x$ is pure, i.e., $\mathrm{Tr}\left[\rho_x^2\right] = 1$. We then have

$$\mu_x = \frac{1 - \mathrm{Tr}\left[\rho^2\right]}{2\left[1 - \mathrm{Tr}\left[\rho\rho_x\right]\right]} \ , \tag{3.21}$$

which can be computed from an ensemble $\rho$ and a state of interest $\rho_x$.

When a state of interest is not pure, we have

$$\mu_x = \frac{1 - \mathrm{Tr}\left[\rho\rho_x\right] - \mathrm{Det}\left(\rho, \rho_x\right)}{1 - \mathrm{Tr}\left[\rho_x^2\right]} \ , \tag{3.22}$$

where

$$\mathrm{Det}\left(\rho, \rho_x\right) = \sqrt{\left(1 - \mathrm{Tr}\left[\rho\rho_x\right]\right)^2 - \left(1 - \mathrm{Tr}\left[\rho^2\right]\right)\left(1 - \mathrm{Tr}\left[\rho_x^2\right]\right)} \ . \tag{3.23}$$

The maximum confidence is obtained as

$$\max C_x = \lambda_x^* = \frac{p_x}{\mu_x} \ . \tag{3.24}$$

when the state of interest $\rho_x$ is prepared a priori with probability $p_x$. We have therefore shown how to compute the maximum confidence for a state of interest. Once $\mu_x$ is found as above, one can find the complementary state $\sigma_x$ in (3.19), from which the optimal POVM element is also found.

**Example: N qubit pure states**

To illustrate our approach, let us consider an ensemble of $n$ arbitrary pure states:

$$\begin{aligned} &\{|\psi_j\rangle\}_{j=0}^{n-1}, \quad \text{where } |\psi_0\rangle = |0\rangle \quad \text{and} \\ &|\psi_j\rangle = \cos\frac{\theta_j}{2}|0\rangle + e^{i\phi_j}\sin\frac{\theta_j}{2}|1\rangle \text{ for } j = 1, \cdots, n-1. \end{aligned} \tag{3.25}$$

Note that the angles $(\theta_j, \phi_j)$ are arbitrary and the state of interest is denoted by $|\psi_0\rangle$. One can compute the maximum confidence as

$$\max C_0 = \frac{2\left(1 - \mathrm{Tr}\left[\rho_0 \rho_M\right]\right)}{n + 1 - (n-1)\,\mathrm{Tr}\left[\rho_M^2\right] - 2\,\mathrm{Tr}\left[\rho_0 \rho_M\right]}, \tag{3.26}$$

where $\rho_M$ is an equally weighted mixture of $n-1$ states $|\psi_j\rangle$ for $j = 1, \cdots, n-1$. It is seen that the maximum confidence depends on two parameters: the purity of an ensemble $\rho_M$ and the fidelity between $\rho_M$ and $\rho_0$.

In addition, as shown in Refs. [17, 25], the maximum confidence is closely related to the outcome rate, the probability that a detection event occurs, denoted by $\eta_0 = \mathrm{Tr}[\rho\hat{\pi}_0]$. Here, the outcome rate is upper-bounded by

$$\eta_+ = 1 + \frac{\mu_0\,\mathrm{Tr}[\rho\rho_0] - \mathrm{Tr}[\rho^2]}{1 - \mu_0}, \tag{3.27}$$

where $\mu_0 = (1 - \mathrm{Tr}[\rho^2])/(2(1 - \mathrm{Tr}[\rho\rho_0]))$.

It's worth emphasizing that any $n$ state discrimination problem within the MCM framework can be turned into a two-state discrimination problem. Since an MCM can only focus on one state of interest ($\rho_0$), the rest can be collected in a mixture $\rho_M$. Maximum confidence can be straight computed with (3.26), which is equivalent to (3.24) for equiprobable preparations.

### 3.5.3 Geometry of the MCM

The general structure of qubit states and MCMs can be depicted on the Bloch sphere. We analyze here the optimality condition geometrically and present the structure. We also show forms of the maximum confidence different from (3.24).

Let us refer to Fig. 3.1. Note that the natural distance measure in the Bloch sphere is given by the Hilbert-Schmidt norm, which turns out to be proportional to the trace norm for qubit cases [108], i.e.,

$$\sqrt{2} d_{\mathrm{HS}}\left(\rho, \sigma\right) = ||\rho - \sigma||_1, \tag{3.28}$$

where $d_{\mathrm{HS}}\left(\rho, \sigma\right) = \sqrt{\mathrm{Tr}\left[(\rho - \sigma)^2\right]}$. For instance, the trace notm between two orthogonal qubit states equals 2, and the Hilbert-Schmidt distance is $\sqrt{2}$. Thus, one can consider two measures interchangeably in the Bloch sphere

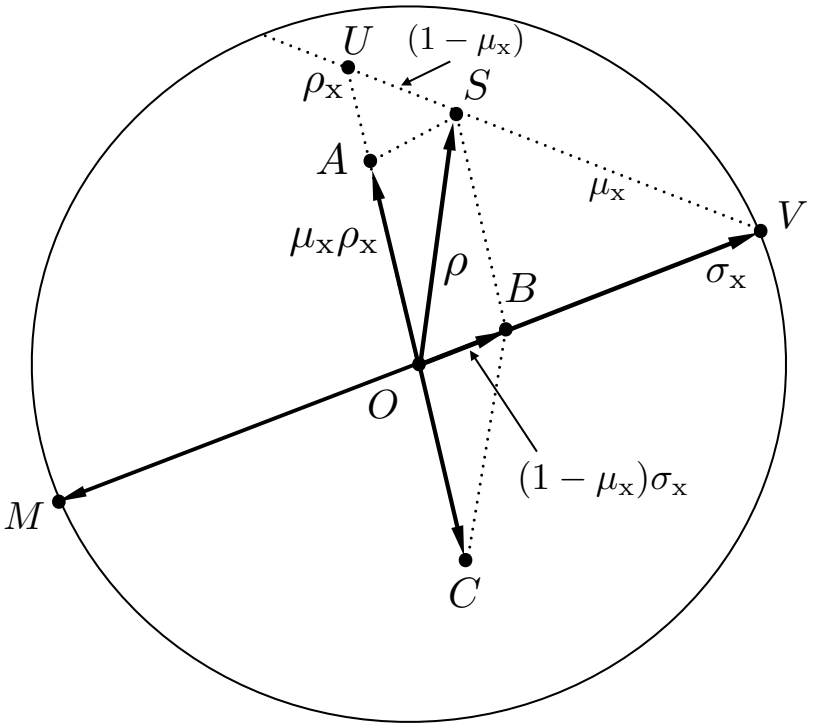



Figure 3.1: Figure extracted from [83]. Geometry of an MCM for qubit states on the Bloch sphere. The arrows represent Bloch vectors. For instance, Bloch vectors $OM$ and $OV$ reach opposite points on the surface of the Bloch sphere, and represent pure orthogonal states. An ensemble $\rho$ and a state of interest $\rho_x$ correspond to $OS$ and $OU$ respectively. Since $\rho$ is a convex combination of $\rho_x$ and a complementary state $\sigma_x$, the state $\sigma_x$ is immediately obtained as $OV$ by extending $US$ until it reaches the surface. An optimal POVM element (MCM) corresponds to $OM$. It holds that $OA + OB = OS$ and $OS + OC = OB \propto OV$.

and relate them by a factor of $\sqrt{2}$.

We begin interpreting (3.19). An ensemble $\rho$ is given as a convex mixture of a state of interest $\rho_x$ and its complementary one $\sigma_x$. This means that the Bloch vector of a state $\rho$ lies on a line connecting two Bloch vectors of two states $\rho_x$ and $\sigma_x$. It also implies that the Bloch vector of a state $\sigma_x$ can be found on a line connecting those of two states $\rho$ and $\rho_x$. Let us recall from the complementary slackness optimality condition in (3.18) that a complementary state $\sigma_x$ must be rank-1. Therefore, one can find a complementary state $\sigma_x$ on the surface at which the line connecting two known states $\rho$ and $\rho_x$ meet (see Fig. 3.1). Once a complementary states is found, an optimal POVM element is obtained as the orthogonal complement: $\hat{\pi}_x^* \propto \sigma_x^\perp$. Both operators $\hat{\pi}_x^*$ and $\sigma_x^\perp$ are rank-1.

Let us also explain the relations between the states and the MCM, as shown in Fig. 3.1. Given states $\rho$ and $\rho_x$, displayed as $OS$ and $OU$ respectively, an optimal measurement is found as $OM$ that is orthogonal to $OV$ obtained on the sphere by extending $US$. The Bloch vector of a complementary state that corresponds to $OV$ can be found as follows. Throughout, let $\vec{r}(\rho)$ denote the Bloch vector of a qubit state $\rho$. A vector $\overrightarrow{US}$ lying on a line defined by points $U$ and $S$ is given by

$$\overrightarrow{US} = (\vec{r}(\rho) - \vec{r}(\rho_x))t_x + \vec{r}(\rho) \ , \tag{3.29}$$

for some $t_x \geq 0$. The complementary state's Bloch vector $\vec{r}(\sigma_x)$ is found when $t_x$ is fixed such that $||\overrightarrow{US}|| = 1$.

From the convex ombination in (3.19) it holds that

$$\frac{||\rho - \rho_x||_1}{||\rho - \sigma_x||_1} = \frac{1 - \mu_x}{\mu_x} \ . \tag{3.30}$$

From the relation above, it is straightforward to find $\mu_x = ||\rho - \sigma_x||_1 / ||\rho_x - \sigma_x||_1$, so that

$$\max C_x = \frac{p_x||\rho_x - \sigma_x||_1}{||\rho - \sigma_x||_1} = p_x \left(1 + \frac{1}{t_x}\right) \ , \tag{3.31}$$

for $t_x$ fulfilling $||\overrightarrow{US}|| = 1$.

We have therefore shown that a complementary state can be directly found by exploiting the qubit state geometry, as well as an MCM. In summary, the maximum confidence for qubit states can be written in the various forms in (3.24) and (3.31).

### 3.5.4 Minimum probability of inconclusive events

Having found POVM elements for an MCM, let us consider the probability inconclusive outcomes. We previously mentioned that a POVM element in an MCM is rank-1. For a general ensemble (3.5), let $\hat{\pi}_x = c_x \left|\varphi_x^\perp\right\rangle \left\langle\varphi_x^\perp\right|$ denote a POVM element for each state where $c_x$ is a non-negative constant and $\left|\varphi_x^\perp\right\rangle \left\langle\varphi_x^\perp\right|$ a rank-1 projector onto a pure state orthogonal to the complementary state $\sigma_x = \left|\varphi_x\right\rangle \left\langle\varphi_x\right|$. These projectors are immediatly obtained such that they perform an MCM. Then, a set of is constants $\{c_x\}_{x=0}^{N-1}$ is chosen to find the probability of inconclusive outcomes, for which the POVM element is denoted by

$$\hat{\pi}_\emptyset = \mathbb{1} - \sum_{x=0}^{N-1} c_x \left|\varphi_x^\perp\right\rangle \left\langle\varphi_x^\perp\right| \ , \tag{3.32}$$

such that $\eta_\emptyset = \mathrm{Tr}\left[\rho\hat{\pi}_\emptyset\right]$.

Some remarks are in order. Firstly, an MCM for an ensemble (3.5) varies by choosing different values of $\{c_x\}_{x=0}^{n-1}$, for all of which an MCM holds true. This immediately concludes that an MCM for an ensemble is not unique. Secondly, if the convex hull of POVM elements performing an MCM contains the identity, so that $\{c_x\}_{x=0}^{n-1}$ can be chosen such that $\sum_x \hat{\pi}_{x=0}^{n-1} = \mathbb{1}$, one can find an MCM that is also complete. Consequently, an inconclusive outcome does not occur, since $\hat{\pi}_\emptyset = 0$ and then $\eta_\emptyset = 0$. Thirdly, if the convex hull of POVM elements does not contain the identity, the probability of inconclusive outcomes is non-zero. We thus introduce an optimisation problem to minimize $\eta_\emptyset$. The problem is defined as follows

$$\underset{c_x}{\text{minimize}} \ \ \mathrm{Tr}\left[\rho\hat{\pi}_\emptyset\right] \tag{3.33}$$

$$\text{subject to } c_x \geq 0, \ \hat{\pi}_\emptyset \geq 0 \ . \tag{3.34}$$

The optimisation problem can be approached from the following Lagrangian,

$$\mathcal{L} = \mathrm{Tr}\left[\rho\hat{\pi}_\emptyset\right] - \sum_x \nu_x c_x - \mathrm{Tr}\left[K\hat{\pi}_\emptyset\right] \ , \tag{3.35}$$

where $K \geq 0$ and $\nu_x \geq 0$ are dual parameters. The optimality conditions contain the Lagrangian stability,

$$\left\langle\varphi_x^\perp\right| K \left|\varphi_x^\perp\right\rangle + \nu_x - \left\langle\varphi_x^\perp\right| \rho \left|\varphi_x^\perp\right\rangle = 0 \ , \tag{3.36}$$

and the complementary slackness

$$\nu_x c_x = 0, \ \text{and} \ \ \mathrm{Tr}\left[K\hat{\pi}_\emptyset\right] = 0 \ . \tag{3.37}$$

This optimization problem works for an arbitrary ensemble of quantum states. In what follows, let us rewrite the problem specifically for qubit states.

One finds that complementary slackness condition (3.37) implies that $\hat{\pi}_\emptyset$ is rank-1 for qubit states. Hence, it holds that

$$\left(\frac{\hat{\pi}_\emptyset}{\mathrm{Tr}\left[\hat{\pi}_\emptyset\right]}\right)^2 = \frac{\hat{\pi}_\emptyset}{\mathrm{Tr}\left[\hat{\pi}_\emptyset\right]} \ , \tag{3.38}$$

which, according to (3.32), is equivalent to

$$1 - \sum_x c_x + \frac{1}{2}\sum_{x,x'} \left(1 - |\left\langle \varphi_x^\perp | \varphi_{x'}^\perp \right\rangle|^2\right) c_x c_{x'} = 0 \ . \tag{3.39}$$

In addition, also according to (3.32), we have that $\mathrm{Tr}\left[\hat{\pi}_\emptyset\right] \geq 0$, which means that $2 - \sum_x c_x \geq 0$. With these relations as constraints now, the optimisation problem in (3.33) can be rewritten as

$$\begin{aligned} \underset{c_x}{\text{minimize}} \quad & \mathrm{Tr}\left[\rho\left(1 - \sum_x c_x \left|\varphi_x^\perp\right\rangle\left\langle\varphi_x^\perp\right|\right)\right] \\ \text{subject to} \quad & c_x \geq 0, \ 2 - \sum_x c_x \geq 0 \\ & 1 - \sum_x c_x + \frac{1}{2}\sum_{x,x'} \left(1 - |\left\langle \varphi_x^\perp | \varphi_{x'}^\perp \right\rangle|^2\right) c_x c_{x'} = 0 \ . \end{aligned} \tag{3.40}$$

The probability of inconclusive outcomes for an MCM of qubit states can be generally obtained by solving this optimisation problem. We reiterate that, once a set of projectors for an MCM is obtained, the optimisation problem above finds a set of optimal coefficients $\{c_x\}_{x=0}^{n-1}$ to minimize the probability of inconclusive outcomes.

## 3.6 Various qubit states

In the following, let us apply the geometric structure of MCM to various ensembles of qubit states. We show how states and their MCM are related to each other. We also compare MCMs for qubit states to measurements for unambiguous and minimum error state discrimination.

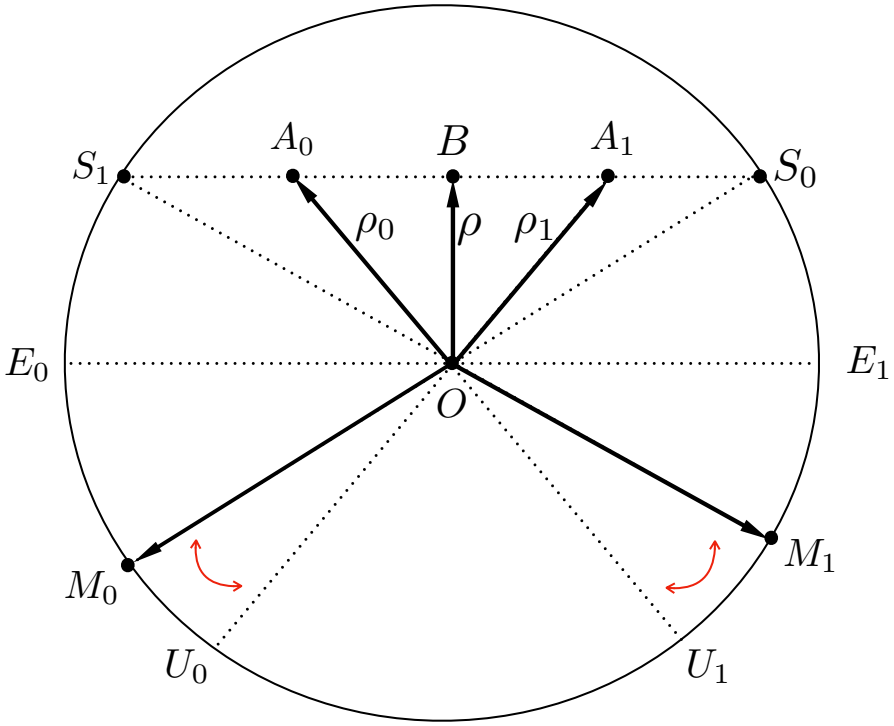



Figure 3.2: Figure extracted from [83]. The couple of states considered in (3.41) are shown in the Bloch sphere. $OA$ and $OB$ denote the Bloch vectors of the states $\rho_0$ and $\rho_1$ respectively. Complementary states $\sigma_0$ and $\sigma_1$ are found by extending the lines connecting $A_0B$ and $A_1B$. These are represented by $OS_1$ and $OS_2$ respectively. The MCM is obtained by flipping $OS_1$ and $OS_2$ to their opposite directions, and are represented by the vectors $\overrightarrow{OM_0}$ and $\overrightarrow{OM_1}$. These align with $OU_0$ and $OU_1$ which are the directions of unambiguous measurements in the noiseless case. When the POVMs are aligned with $OE_0$ and $OE_1$, one recovers the optimal measurement for MESD. Then, we see that MCM is found between USD and MESD.

### 3.6.1 Two qubit states

The first example is two qubit states, each prepared with equal *a priori* probabilities (see Fig. 3.2)

$$\rho_x = r\left|\psi_x\right\rangle\left\langle\psi_x\right| + (1-r)\frac{\mathbb{1}}{2}, \quad x = 0,1 \quad \text{and} \tag{3.41}$$
$$\rho = \frac{r}{2}(\left|\psi_0\right\rangle\left\langle\psi_0\right| + \left|\psi_1\right\rangle\left\langle\psi_1\right|) + (1-r)\frac{\mathbb{1}}{2}.$$

Unambiguous discrimination is not possible for this ensemble states if $r < 1$. Two pure states may be parameterized by $\cos\theta = \langle\psi_0|\psi_1\rangle$, so we can without loss of generality write

$$\left|\psi_x\right\rangle = \cos\frac{\theta}{2}\left|0\right\rangle + (-1)^x \sin\frac{\theta}{2}\left|1\right\rangle. \tag{3.42}$$

The maximum confidence for each state is computed as

$$\max C_x = \frac{1}{2}\left(1 + \frac{r\sqrt{1-\cos^2\theta}}{\sqrt{1-r^2\cos^2\theta}}\right). \tag{3.43}$$

The MCM can be obtained from the Bloch vectors of the states:

$$\vec{r}(\rho_x) = \{(-1)^x r\sin\theta, 0, r\cos\theta\} \tag{3.44}$$
$$\vec{r}(\rho) = \{0, 0, r\cos\theta\}\ . \tag{3.45}$$

From these, the Bloch vectors of the complementary states are found, using (3.31), to be

$$\vec{r}\,(\sigma_x) = \left\{t_x(-1)^{x+1} r\sin\theta, 0, r\cos\theta\right\} \tag{3.46}$$
$$\text{with:} \quad t_x = \frac{\sqrt{1-r^2\cos^2\theta}}{r\sqrt{1-\cos^2\theta}}\ ,$$

where we note that $||\vec{r}(\sigma_x)|| = 1$. An optimal POVM is rank-1 and can be described by the unit Bloch vector $\vec{r}(\hat{\pi}_x) = -\vec{r}(\sigma_x)$. That is, an optimal POVM element for state $\rho_x$ is given by $\hat{\pi}_x \propto (\mathbb{1} + \vec{r}(\hat{\pi}_x)\cdot\vec{\sigma})/2$, for $\vec{\sigma} = \{X, Y, Z\}$ is a vector containing the Pauli matrices in (2.6). We show the geometric picture in Fig. 3.2.

Remarks are in order. Firstly, suppose that pure states are given, i.e. $r = 1$. Then, we have that $\vec{r}(\hat{\pi}_x) = \vec{r}(\rho_{x+1})$, meaning $\hat{\pi}_x \perp \rho_{x+1}$. In this case, an MCM coincides with USD. Secondly, the MCM varies according to the noise parameter $r$. Thirdly, for all values of $r \in (0, 1]$, an MCM is never a null

measurement. The same holds true even for different *a priori* probabilities $p_x$. That is, the act of not measuring can never give the maximal confidence. This contrasts with certain cases of MESD, in which a null measurement is optimal whenever $|p_0 - p_1| > ||p_0\rho_0 - p_1\rho_1||_1$.

The probability of inconclusive outcomes can be minimized according to (3.40). Since we consider equal *a priori* probabilities, the problem present the following symmetry $c_0 = c_1$. Then, it is straightforward to solve the optimisation problem. The minimal probability of inconclusive events is then $\eta_\emptyset = p\cos\theta$. Note that in the noiseless case ($r = 1$) one recovers the USD bound in (2.37).

### 3.6.2 Geometrically uniform states

A set of $n$ states $\{\rho_x\}_{x=0}^{n-1}$ are geometrically uniform when there exists a unitary transformation $U$ such that $U\rho_x U^\dagger = \rho_{x+1}$, $\forall x$, and $U^n = \mathbb{1}$ [113]. As one example (see Fig. 3.3), geometrically uniform qubit pure states can be written as

$$|\psi_x\rangle = \cos\frac{\theta}{2}|0\rangle + e^{x\frac{2\pi}{n}i}\sin\frac{\theta}{2}|1\rangle \ . \tag{3.47}$$

Note that a set of $n$ states $\{\rho_x\}_{x=0}^{n-1}$ generalized the three qubit states considered in Ref. [101]. Assume that the states are given with equal *a priori* probabilities $p_x = 1/n$, $\forall x$. Then, the ensemble is given by

$$\rho = \frac{1}{2}\left(\mathbb{1} + \cos\theta Z\right) \ . \tag{3.48}$$

Since we consider pure states, we have $\mu_x = 1/2$, $\forall x$. The maximum confidence is then given by

$$\max C_x = \frac{2}{n} \ . \tag{3.49}$$

In this case, the maximum confidence concerning a particular state of interest only depends on the cardinality of an ensemble: it is less confident to detect a particular state of a larger set, and vice versa.

An optimal measurement can be found as follows. The Bloch vectors of the states are

$$\vec{r}(\rho_x) = \left\{\cos\frac{2\pi}{n}x\sin\theta, \sin\frac{2\pi}{n}x\sin\theta, \cos\theta\right\} \tag{3.50}$$

$$\vec{r}(\rho) = \{0, 0, \cos\theta\} \ . \tag{3.51}$$

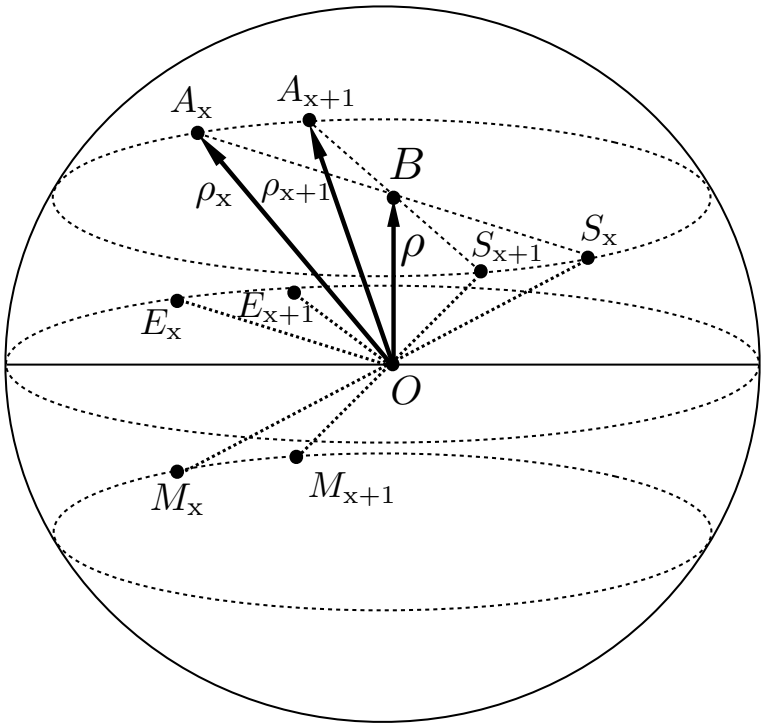



Figure 3.3: Figure extracted from [83]. Geometrically uniform pure states $OA_x$ form a circle with radius $BA_x$, where $OB$ denotes the ensemble of the states with equal *a priori* probabilities. For every particular state $OA_x$, its complementary state is found in $OS_x$ by extending $BA_x$ to the surface of the Bloch sphere. Then, one finds the corresponding MCM $OM_x$ flipping $OS_x$. We highlight the half-plane defined by the collection of midpoints $A_xM_x$. The MESD measurement is formed by vectors aligned with $OE_x$, which are projections of $OA_x$ onto the half-plane. For states lying on the half-plane, MCM coincides with a measurement in MESD.

Then, the Bloch vectors of the complementary states are found, using (3.31), to be

$$\vec{r}(\sigma_x) = \left\{ -\cos\left(\frac{2\pi}{n}x\right)\sin\theta, -\sin\left(\frac{2\pi}{n}x\right)\sin\theta, \cos\theta \right\} \ , \tag{3.52}$$

where $\{-\vec{r}(\sigma_x)\}_x$ denote the vectors of optimal POVM elements.

The minimal probability of inconclusive events can be obtained solving the optimisation problem in (3.40). Since we consider equal *a priori* probabilities, the problem presents the following symmetry $c_x = c_{x'}$, $\forall x$. Then, the minimal probability of inconclusive events is $\eta_\emptyset = |\cos\theta|$. Note that if $n = 3$, one recovers the same results as in Ref. [101].

One cannot recover the same results from USD for $n > 2$. The measurement for MESD is found on the half-plane [fig ref], and the success probability is $p_{\text{suc}} = (1+\sin\theta)/n$. That is, the geometrically uniform states with $\theta = \pi/2$, the MCM recovers the MESD bound.

### 3.6.3 Tetrahedron states

For the next example we consider an ensemble of tetrahedral states,

$$|\psi_0\rangle = |0\rangle \ , \ |\psi_x\rangle = \sqrt{\frac{1}{3}}\,|0\rangle + e^{x\frac{2\pi}{3}i}\sqrt{\frac{2}{3}}\,|1\rangle \ , \ \text{for: } x = 1,2,3 \ . \tag{3.53}$$

These states form a tetrahedrom in the Bloch sphere, as shown in Fig. 3.4, hence the name. These are symmetric, informationally complete (SIC) states, since $|\langle\psi_x|\psi_{x'}\rangle| = 1/3$ for $x \neq x'$ [114].

To be more general, let us consider noisy tetrahedron states

$$\rho_x = r\,|\psi_x\rangle\langle\psi_x| + (1-r)\frac{\mathbb{1}}{2}, \ \text{for: } x = 0,1,2,3 \ , \tag{3.54}$$

with equal *a priori* probabilities, hence $\rho = \mathbb{1}/2$. From (3.24), it follows that

$$\mu_x = \frac{1}{1+r} \ \text{with} \ \max C_x = \frac{1+r}{4}, \ \text{for: } x = 0,1,2,3 \ . \tag{3.55}$$

Note that, for pure noiseless states, the maximum confidence is 1/2. Since

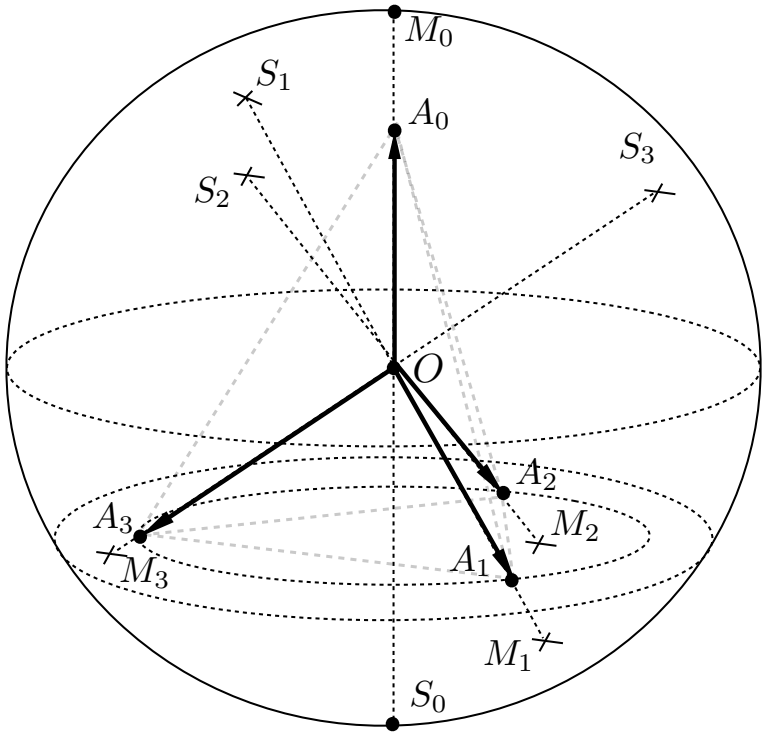



Figure 3.4: Figure extracted from [83]. Collection of noisy tetrahedron states $OA_x$, the ensemble of which is $\mathbb{1}/2$, represented as the origin point $O$ on the Bloch sphere. The complementary states are then directly on the opposite directions of the given states: $OS_x$. This means that the optimal POVM yielding an MCM ($OM_x$) are projectors onto the same directions of the given states $OA_x$.

the Bloch vectors of the tetrahedron states are given by

$$\vec{r}(\rho_0) = \{0, 0, r\} \tag{3.56}$$

$$\vec{r}(\rho_x) = \left\{ \frac{2\sqrt{2}}{3} r \cos\left(\frac{2\pi}{3}x\right), \frac{2\sqrt{2}}{3} r \sin\left(\frac{2\pi}{3}x\right), -\frac{1}{3}r \right\} \tag{3.57}$$

$$\vec{r}(\rho) = \{0, 0, 0\} \ , \tag{3.58}$$

for $x = 0, 1, 2, 3$, one can find the Bloch vectors of the complementary states,

$$\vec{r}(\sigma_x) = -\frac{1}{r}\vec{r}(\rho_x), \text{ for: } x = 0, 1, 2, 3 \ . \tag{3.59}$$

We show the corresponding MCM for tetrahedron states in Fig. 3.4. This MCM coincides with the measurement that minimizes the error, i.e., MCM and a MESD measurement are equivalent for the tetrahedron states [108, 109]. We also remark that, as shown in Sec. 3.5.4, since the convex hull of projectors contains the identity, one can always find an MCM with zero inconclusive events.

### 3.6.4 Asymmetric states I

Consider the ensemble of three asymmetric states, which is constructed by slightly modifying one of the three geometrically uniform states. We look at the three states

$$|\psi_0\rangle = \cos\frac{\theta}{2}|0\rangle + \sin\frac{\theta}{2}|1\rangle \ , \quad \text{and}$$

$$|\psi_1\rangle = \frac{1}{2}|0\rangle + \frac{\sqrt{3}}{2}|1\rangle), \quad |\psi_2\rangle = \frac{1}{2}(|0\rangle - \sqrt{3}|1\rangle). \tag{3.60}$$

That is, two states $|\psi_1\rangle$ and $|\psi_2\rangle$ are fixed and a state $|\psi_0\rangle$ is varied by an angle $\theta$. The Bloch vectors are

$$\vec{n}(\psi_0) = (\sin\theta, 0, \cos\theta), \quad \vec{n}(\psi_1) = (\frac{\sqrt{3}}{2}, 0, -\frac{1}{2}) \tag{3.61}$$

$$\vec{n}(\psi_2) = (\frac{-\sqrt{3}}{2}, 0, -\frac{1}{2}), \quad \vec{n}(\rho) = \frac{1}{3}(\sin\theta, 0, -1 + \cos\theta).$$

It turns out that an MCM for them does not contain any symmetry, as can be seen in Fig. 3.5. We make use of the expression in (3.31) to get

$$t_0 = \frac{9 - 2(1 - \cos\theta)}{9 - 4(1 - \cos\theta)} \ \text{ and} \tag{3.62}$$

$$t_x = \frac{9 - 2(1 - \cos\theta)}{9 - (1 - \cos\theta) - 3\sqrt{3}(-1)^x \sin\theta}, \ x = 1, 2. \tag{3.63}$$

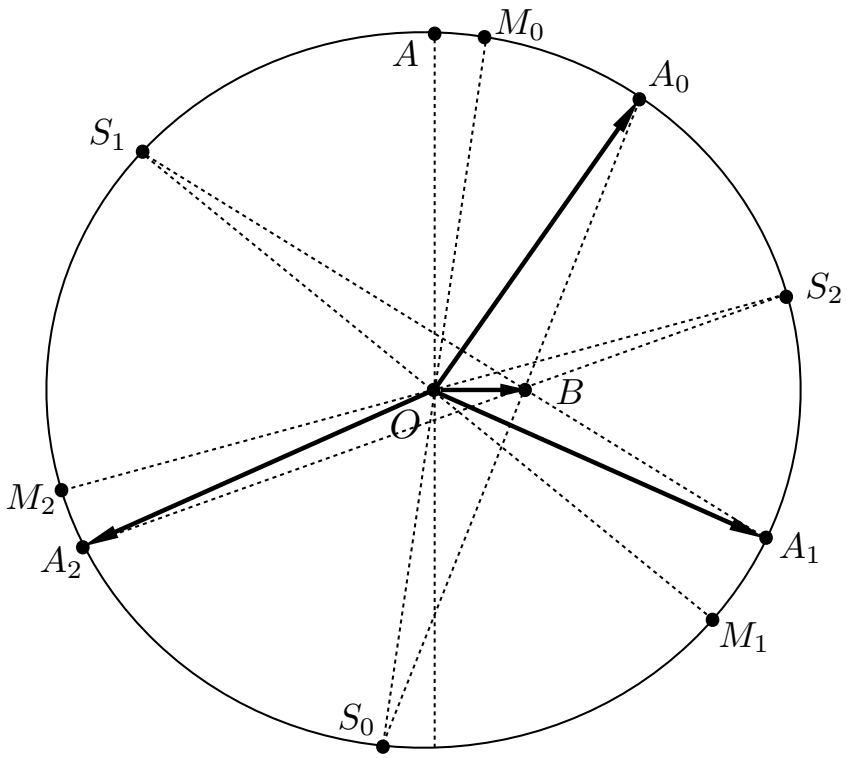



Figure 3.5: Three states $OA$, $OA_1$ and $OA_2$ are geometrically uniform. The first state is slightly tilted so that an ensemble of three states $OA_x$ for $x = 0, 1, 2$ is considered. The ensemble is denoted by $OB$. Complementary states $OS_x$ are found by extending $A_xB$, and optimal POVM elements are found by inverting $OS_x$ with respect to $O$. None of the elements $OM_x$ are identical to states $OA_x$.

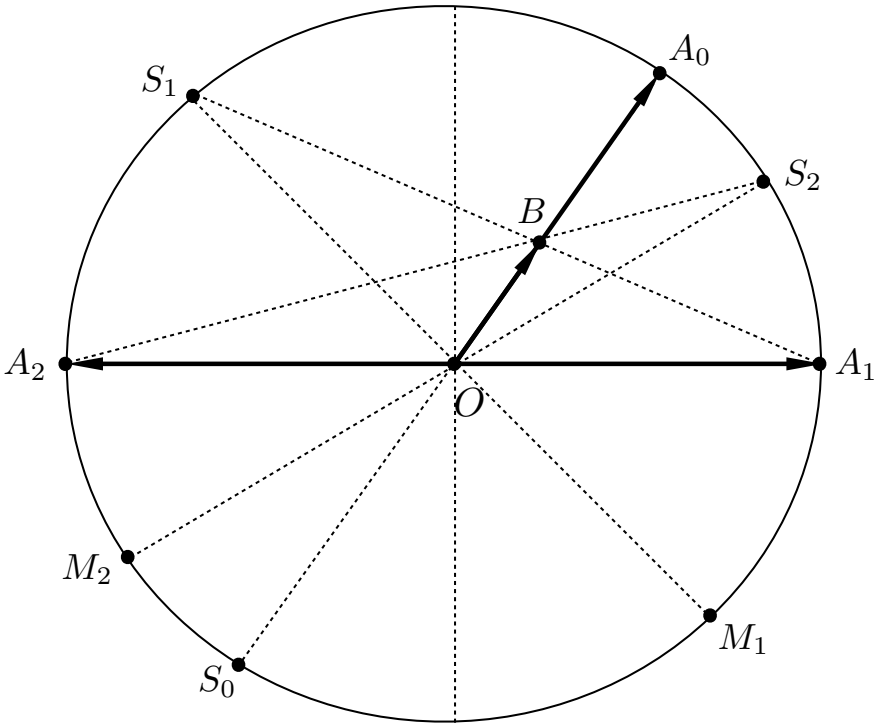



Figure 3.6: Three states $OA_0$, $OA_1$ and $OA_2$ are considered where $OA_1$ and $OA_2$ are orthogonal. Complementary states $OS_x$ are found on the sphere by extending $A_xB$. An optimal POVM consists of $OA_0$, $OM_1$, and $OM_2$. A measurement for minimum error state discrimination contains two POVM elements $OA_1$ and $OA_2$.

The maximum confidence is found to be $(1+1/t_x)/3$. It is seen that an MCM for the asymmetric states does not contain any symmetry.

### 3.6.5 Asymmetric states II

The next example of asymmetric states considered is the following

$$|\psi_0\rangle = \cos\frac{\theta}{2}|0\rangle + \sin\frac{\theta}{2}|1\rangle\,, \quad |\psi_1\rangle = |+\rangle\,, \text{ and } |\psi_2\rangle = |-\rangle\ ,$$

where each state is prepared with equal *a priori* probability (see Fig. 3.6). Similarly to the previous case, two states $|\psi_1\rangle$ and $|\psi_2\rangle$ are fixed and a state $|\psi_0\rangle$ varies by an angle $\theta$. In contrast with the ensemble in (3.60), the pair of states $|\pm\rangle$ contains a symmetry: they are invariant under a rotation about

the $x$-axis. Their Bloch vectors are

$$\vec{n}(\psi_0) = (\sin\theta, 0, \cos\theta), \quad \vec{n}(\psi_1) = (1,0,0)$$
$$\vec{n}(\psi_2) = (-1,0,0), \quad \vec{n}(\rho) = \frac{1}{3}(\sin\theta, 0, \cos\theta). \tag{3.64}$$

We again exploit the expression in (3.31) to find

$$t_0 = 2, \quad t_1 = \frac{4}{5 - 3\sin\theta}, \quad t_2 = \frac{4}{5 + 3\sin\theta}. \tag{3.65}$$

It follows that

$$\max C(x) = \frac{1}{3}(1 + \frac{1}{t_x}). \tag{3.66}$$

Interestingly, the maximum confidence for the state $|\psi_0\rangle$, which is parameterized by $\theta$, does not depend on the angle. The maximum confidence for the other two states depends upon the angle $\theta$ from the other state $|\psi_0\rangle$.

In contrast to the three states in the case of the ensemble in (3.60), the MCM contains a symmetry, seen from the Bloch vectors of the complementary states which are

$$\hat{r}_0 = -\vec{n}(\psi_0) \tag{3.67}$$
$$\hat{r}_1 = (\frac{1}{3}\sin\theta - 1, 0, \frac{1}{3}\cos\theta)t_1 + \frac{1}{3}(\sin\theta, 0, \cos\theta)$$
$$\hat{r}_2 = (\frac{1}{3}\sin\theta + 1, 0, \frac{1}{3}\cos\theta)t_2 + \frac{1}{3}(\sin\theta, 0, \cos\theta).$$

That is, an optimal POVM element for the state $|\psi_0\rangle$ shares its Bloch vector with the state $\vec{n}(\psi_0)$. An MCM for two states $|\pm\rangle$ depends on the angle $\theta$ of the other state $|\psi_0\rangle$.

In the case of minimum error state discrimination for the ensemble, an optimal measurement does not aim to detect a state $|\psi_0\rangle$. It contains two POVM elements having Bloch vectors $\vec{n}(\psi_1)$ and $\vec{n}(\psi_2)$. Then, a detection event on the first (second) POVM element characterized by $\vec{n}(\psi_1)$ ($\vec{n}(\psi_2)$) concludes a state $|\psi_1\rangle$ ($|\psi_2\rangle$). In this way, the guessing probability is given as 2/3 [108].

## 3.7 Bounds on the rate of observed events

We add this section, which is not included in [83], for completeness in the framework of presenting MCMs. The results presented in this section turn

useful for understanding future chapters.

As previously discussed, the maximum attainable confidence has an inherent dependence on the rate of observed events $\eta_x$. This dependence manifests as a bound, limited by the geometry of the MCM. Also, depending whereas the confidence of a single state of interest or the confidence of the whole ensemble is maximised, the achievable rates will differ. Hence, we will particularly look at both cases, and specify the analytical values of these bounds. We will consider the task of discriminating the states in the ensemble introduced in (3.41).

### 3.7.1 MCM of the whole ensemble

The maximum confidence $C$ in (3.43) can be simultaneously reached for the whole ensemble $\rho$ if both POVM elements $\hat{\pi}_0$ and $\hat{\pi}_1$ are proportional to the projectors onto the orthogonal complementary states. Since no state has a preference, we rather not specify a particular bound for $\eta_0$ or $\eta_1$, but for the rate of inconclusive events. The minimum value of $\eta_\varnothing$ at which both confidences $C_0 = C_1 = C$ are maximum can be obtained by studying the positive semi-definiteness constraint on the POVM element $\hat{\pi}_\varnothing$. The process is the same as we did when finding a lower bound on $\eta_\varnothing$ but in USD (see Sec. 2.3.1). First of all, we write the rate of inconclusive events as $\eta_\varnothing = 1 - (1 - r^2\cos^2\theta)(c_0 + c_1)/2$. Then, one finds that $\hat{\pi}_\varnothing \geq 0$ if and only if $c_0 c_1(1 - r^2\cos^2\theta) \geq c_0 + c_1 - 1$. Any values of $c_0$ and $c_1$ satisfying this inequality constitute a valid MCM and thus, place a bound on the rates $\eta_x$. The following steps are in order: from the previous inequality isolate $c_i$, substitute it in $\eta_\varnothing$ and find a minimum by solving $\left.\frac{d\eta_\varnothing}{dc_i}\right|_{c_i^*} = 0$. At the end of the day, the minimum rate $\eta_\varnothing$ for which an MCM is possible is

$$\eta_\varnothing \geq r\cos\theta \ . \tag{3.68}$$

Again, in the noiseless case one recovers the exact same result in USD (2.37). For smaller rates of inconclusive events the maximum confidence is no longer achievable. However, we can still find a bound which is strictly dependent on $\eta_\varnothing$. Let us use the geometrical interpretation on the Bloch sphere. The maximum confidence for rates $1 \geq \eta_\varnothing \geq r\cos\theta$ is achieved by rank-1 POVM elements. As $\eta_\varnothing$ becomes smaller, the Bloch vector of $\hat{\pi}_x$ increases in length until it reaches the surface of the Bloch sphere for $\eta_\varnothing = r\cos\theta$. If one now wants to see what happens if $\eta_\varnothing$ keeps decreasing, the Bloch vector of the POVM elements $\hat{\pi}_x$ must rotate in a particular direction. This rotation will increase the error probability (i.e. $\mathrm{Tr}\left[\rho_x\hat{\pi}_b\right]$ for $b \neq x$ and $b \neq \varnothing$), which can

be mitigated by decreasing the value of the proportionality coefficients $c_x$. Observe that we converted our maximum confidence problem into a minimum error problem with a fixed rate of inconclusive events. Let us introduce the success $p_{\text{suc}}$ and error $p_{\text{err}}$ probabilities as

$$p_{\text{suc}} = \frac{1}{2}\left(p(0|0) + p(1|1)\right) \qquad p_{\text{err}} = \frac{1}{2}\left(p(1|0) + p(0|1)\right) \ . \tag{3.69}$$

Without loss of generality, one can consider the symmetric case ($p(0|0) = p(1|1)$ and $p(0|1) = p(1|0)$) and re-write the confidence as $C = p_{\text{suc}}/(p_{\text{suc}}+p_{\text{err}})$. Roughly speaking, here $p_{\text{suc}}$ and $p_{\text{err}}$ correspond to the success and error probabilities of measuring $\rho_x$ using the best possible measure to discriminate the complementary states $\sigma_x$. The inconclusive rate is then $1-\eta_\emptyset = p_{\text{suc}}+p_{\text{err}}$. Finding the maximum confidence is equivalent to maximising the difference between the success and error probabilities, fixing the rate of inconclusive events. In Chapter 5 we find the explicit analytic solution for the maximal success and minimum error. Using that solution, at the end of the day, one ends up with the maximum confidence

$$C = \frac{1}{2}\left(1 + \frac{r\sqrt{(1-\cos^2\theta)(1+r\cos\theta-2\eta_\emptyset)}}{\sqrt{1+r\cos\theta}(1-\eta_\emptyset)}\right) \quad \text{for} \quad \eta_\emptyset \leq r\cos\theta \ . \tag{3.70}$$

The values of the maximum confidence for the whole ensemble are shown in Fig. 3.7 on the left. We thus found an upper bound on the confidence for any rate of inconclusive events. This generalises the results from USD and MESD due to the flexibility of accepting any value of inconclusive events.

### 3.7.2 MCM for a state of interest

One can only be interested in a particular state $\rho_x$ of the whole ensemble $\rho$. Then, the only object of interest is the individual confidence $C_x$ in correctly measuring state $\rho_x$. In our present case, we will focus in maximising the confidence $C_0$ of $\rho_0$ (although results are valid for any state $\rho_x$) from the ensemble formed by the states in (3.41). The rate of events of interest is then $\eta_0$. For small enough values of $\eta_0$, the maximum confidence in (3.43) can be reached. That is, when the POVM element corresponding to the outcome of interest (i.e. $\hat{\pi}_0$) is rank-1, proportional to the projector onto the orthogonal complementary state $|\varphi_0^\perp\rangle$. The positive semi-definiteness condition $\hat{\pi}_\emptyset \geq 0$ applied to the coefficients $c_x$ states that $c_0c_1(1-r^2\cos^2\theta) \geq c_0+c_1-1$. Since we are only interested in the case of $x=0$, one sees that $c_0 \leq 1$. This means that, since $c_x = 2\eta_x/(1-r^2\cos^2\theta)$, $\forall x$, the possible rates $\eta_0$ at which the

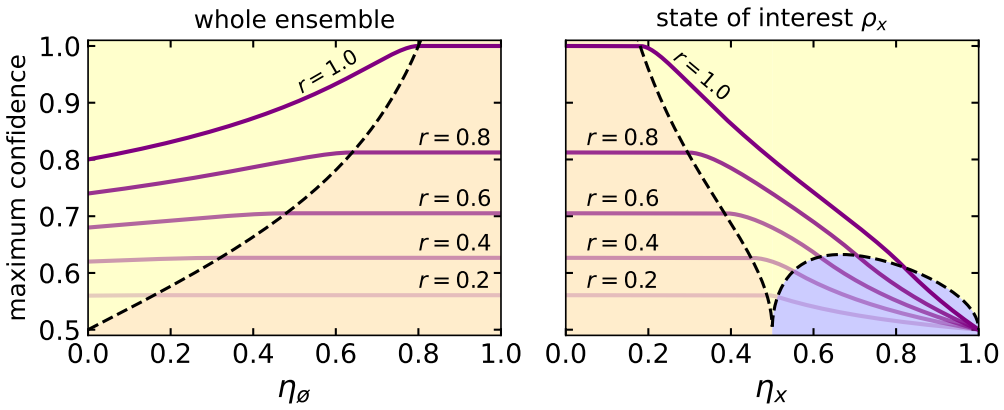



Figure 3.7: Maximum confidence for the whole ensemble $\rho$ (left figure) and for a particular state of interest $\rho_x$ (right figure) for different noise parameters $r$. The discrimination task is done between the pair of states in (3.41). We show the particular case of two-state discrimination with a fixed overlap $\cos\theta = 0.8$. In the orange regions the maximal confidence in (3.43) is reached. On the left figure, the yellow region covers inconclusive rates $\eta_\varnothing \geq r\cos\theta$. On the right figure, the yellow region covers rates within the range $(1 - r^2\cos^2\theta) \leq 2\eta_x \leq (1 + r^2\cos^2\theta)$, whilst the blue region shows $2\eta_x \geq (1 + r^2\cos^2\theta)$ cases.

maximum confidence is achieved are

$$0 \leq \eta_0 \leq \frac{1}{2}\left(1 - r^2\cos^2\theta\right) \ . \tag{3.71}$$

Note that here the maximum confidence is achieved for inconclusive rates smaller than the usual USD bound in (3.68). Explicitly, one sees that $\eta_\varnothing = 1 - \eta_0 - \eta_1 \geq r\cos\theta \geq r^2\cos^2\theta$ according to (3.71). The interpretation is the following: for inconclusive rates $\eta_\varnothing \geq r\cos\theta$, both states $\rho_x$ can be discriminated with maximum confidence, as shown in the previous section, but if $r\cos\theta \geq \eta_\varnothing \geq r^2\cos^2\theta$, only one POVM can have rank-1, in this case $\hat{\pi}_0$. In the noiseless case ($r = 1$), if $1 \geq \eta_\varnothing \geq r\cos\theta$ the problem reduces to USD, whereas if $r\cos\theta \geq \eta_\varnothing \geq r^2\cos^2\theta$ only $\rho_0$ can be unambiguously identified.

If the experiment reaches greater values of $\eta_0$, the maximum confidence in (3.43) cannot be achieved. However, we can still place a tight bound on the confidence $C_0$. To do that, we need to allow $\hat{\pi}_0$ to rotate towards the state $\rho_0$ in the Bloch sphere representation. Unlike the maximum confidence for the whole ensemble, we only need to minimise the error probability $\mathrm{Tr}\left[\rho_1\hat{\pi}_0\right]$ which appears in the denominator of $C_0$. This means that, for $\eta_0 \geq (1 - r^2\cos^2\theta)/2$, we only care about the rotation of $\hat{\pi}_0$ towards $\rho_0$, while $c_0$ is kept at maximum ($c_0 = 1$). We denote by $\phi_0$ the Bloch angle of $\hat{\pi}_0$ with $Z$ (for $Z$ being the Pauli matrix in (2.6)), which will run within the range $\arccos\left[r\cos\left(\pi - \theta\right)\right] \geq \phi_0 \geq \arccos\left(r\cos\theta\right)$. The rate $\eta_0$ can be expressed in terms of the angle $\phi_0$ as $\eta_0 = (1 - r\cos\theta\cos\phi_0)/2$, which means that the range of rates $\eta_0$ at which this behavior is attainable is

$$\frac{1}{2}\left(1 - r^2\cos^2\theta\right) \leq \eta_0 \leq \frac{1}{2}\left(1 + r^2\cos^2\theta\right) \ . \tag{3.72}$$

The maximum confidence in that range of rates can be directly computed and is given by

$$C_0 = \frac{1}{2}\left(1 + \frac{\sqrt{(1 - \cos^2\theta)(r^2\cos^2\theta - (1 - 2\eta_0)^2)}}{2\cos\theta\eta_0}\right) \ . \tag{3.73}$$

For greater values of $\eta_0$, the measurement cannot be longer represented by a POVM with a rank-1 element $\hat{\pi}_0$. The only direction to go, while keeping the error probability $\mathrm{Tr}\left[\rho_1\hat{\pi}_0\right]$ at minimum, is to increase $\mathrm{Tr}\left[\hat{\pi}_0\right]$ while paying the price of decreasing the length of the Bloch vector of the operator $\hat{\pi}_0$. The rate $\eta_0$ (now corresponding to a POVM $\hat{\pi}_0$ with a fixed angle $\phi_0 = \arccos\left(r\cos\theta\right)$ but a variable vector length $r_{\hat{\pi}_0}$) is then expressed

by $\eta_0 = \mathrm{Tr}\left[\hat{\pi}_0\right](1 + r r_{\hat{\pi}_0} \cos^2\theta)/2$. The rate $\eta_0$ is saturated when $\hat{\pi}_0 \to \mathbb{1}$, i.e. $r_{\hat{\pi}_0} = 0$ and $\mathrm{Tr}\left[\hat{\pi}_0\right] = 2$. The rate of events $\eta_0$ is then bounded by

$$\frac{1}{2}\left(1 + r^2 \cos^2\theta\right) \leq \eta_0 \leq 1 \ . \tag{3.74}$$

In terms of the maximum attainable confidence for this particular rate of observed events, this means

$$C_0 = \frac{1}{2}\left(1 + \frac{r\sqrt{1-\cos^2\theta}}{\sqrt{1 - r^2\cos^2\theta}} \frac{1-\eta_0}{\eta_0}\right) \ . \tag{3.75}$$

In Fig. 3.7, on the right, we show the bounds on the maximum confidence for a state of interest. As a safe check, the analytical results are compared with numerical solutions from semidefinite programming, and match perfectly.

## 3.8 Conclusion

In summary, we have investigated MCMs for qubit states. We have presented a simple scheme to find MCMs for qubit states when an ensemble and a state of interest are given. The scheme exploits the geometry in a Bloch sphere without resorting to the computational optimization problem. We then considered various qubit states. From the cases of two qubit states, it is shown that an MCM lies between two strategies, minimum error and unambiguous discrimination. An MCM for geometrically uniform states generalizes an example from Ref. [101]. An MCM for tetrahedron states is identical to a measurement for minimum error state discrimination. Otherwise, when an ensemble does not contain any symmetry, it was seen that MCMs highly depends on the particular state of interest.

Our results elucidate the meanings of different qubit measurements, each of which may aim to maximize different figures of merit. Measurements for various qubit ensembles may also be used to devise quantum protocols to certify the properties of qubit states.

# Chapter 4

# Contextual advantages and certification for maximum confidence discrimination

In this chapter we present the results in "Contextual advantages and certification for maximum confidence discrimination" [25], authored by Kieran Flatt, Hanwool Lee, Carles Roch i Carceller, Jonatan Bohr Brask and Joonwoo Bae. This work was published in PRX Quantum.

## 4.1 Abstract

One of the most fundamental results in quantum information theory is that no measurement can perfectly discriminate between non-orthogonal quantum states. In this work, we investigate quantum advantages for discrimination tasks over noncontextual theories by considering a maximum confidence measurement that unifies different strategies of quantum state discrimination, including minimum- error and unambiguous discrimination. We first show that maximum confidence discrimination, as well as unambiguous discrimination, contains contextual advantages. We then consider a semi-device independent scenario of certifying maximum confidence measurement. The scenario naturally contains undetected events, making it a natural setting to explore maximum confidence measurements. We show that the certified maximum confidence in quantum theory also contains contextual advantages. Our results establish how the advantages of quantum theory over a classical model may appear in a realistic scenario of a discrimination task.

## 4.2 Introduction

Quantum information processing displays advantages over its classical counterpart. These gaps have their origins in fundamental results that show how the two types of theories differ. Quantum key distribution protocols, for example, exploit the indistinguishability of non-orthogonal states to establish security without any assumptions on the computational capabilities [87]. Likewise, measurements on entangled states may give rise to nonlocal correlations, which cannot be produced from classical systems [15, 53]. Nonlocal correlations lead to various practical quantum information applications, in particular device independent quantum information processing, including secure communication [49, 50, 115] and randomness generation [116, 117]. These correlations are used, on the other hand, as a useful tool in the certification of quantum resources such as entanglement, which enables the aforementioned advantages for quantum information processing [118].

The fact that two nonorthogonal states cannot be perfectly discriminated is among the most fundamental results in quantum information theory [77]. This is closely connected to other key results, such as the quantum no-cloning theorem [84] and no-signaling condition [85]. If perfect clones of a pair of non-orthogonal states could be obtained, it would be possible to perfectly discriminate the states. Conversely, perfect discrimination between non-orthogonal states makes it possible to prepare copies of the states. Quantum cloning converges, in the asymptotic limit, to quantum state discrimination [119], which is then limited by the no-signaling condition [120, 121]. The results for a pair of non-orthogonal states have been applied to quantum cryptographic protocols [87] and various other tasks in quantum information theory [29, 93–95, 97].

The distinction between quantum and noncontextual theories in the task of two-state discrimination has recently been shown [27]. In a noncontextual ontological model, operationally equivalent experimental procedures have the same representation. This feature does not hold for quantum theories, so noncontextuality can be understood as one form of classicality. In the aforementioned work, the maximal success probability in two-state, minimum error state discrimination (MESD) is characterised in a noncontextual ontological model. It turns out that two-state MESD in quantum theory is more successful than the derived limitation, showing contextual advantages for quantum state discrimination.

From the point of view of realising these quantum advantages, a general

difficulty lies in the inherent noise of quantum measurements. Even if a state has been prepared, it will sometimes not be detected due to, for example, photon losses. This is treated in MESD by binning such cases among the possible outcomes at the cost of increasing the error rate.

Another form of quantum state discrimination may be considered. In unambiguous state discrimination (USD), a conclusion from certain detection events is never wrong but there is a possibility that no guess is returned [79–81]. An additional arm that collects all inconclusive outcomes is included. The possibility of realising USD, however, highly depends on parameters such as the Hilbert space dimension and the number of states. For instance, for qubit states it cannot be realised for cases other than two pure states.

A figure of merit that operationally unifies the different senses of quantum state discrimination is the confidence [101]. The confidence is defined as the probability that, given a detection event, a detector correctly concludes that a state, chosen among an ensemble, has been prepared. In a maximum confidence measurement (MCM) this figure of merit is maximised. Detectors in USD have certainty as the maximum confidence since a detection event never leads to a wrong conclusion. An MCM performs MESD if the confidence over the whole ensemble of states is considered.

It should be noted that MCMs are concerned with detected events only. The consequence is that MCMs do not suffer from the same weaknesses as MESD or USD. This is closely connected to a retrodictive view of quantum theory, whereby detected events in the present assert statements about state preparation in the past, as discussed in a recent review [100]. One may therefore exploit MCMs to pave a way to gain contextual advantages with imperfect measurement devices in a realistic setting. It is also possible, taking a different point of view of retrodictive quantum theory, to certify the maximum confidence one can have in uncharacterised detectors used for state discrimination. This may be interpreted as a semi-device independent scenario, [65, 122] under the assumption that states are well-characterised but the measurements not at all.

In this work, we establish contextual advantages for both state discrimination and its certification in a realistic scenario where undetected events may appear. We first present contextual advantages for USD by showing that the minimal rate of inconclusive outcomes in quantum theory is strictly lower than that in a noncontextual ontological model. Then, the contextual advantages are shown for maximum confidence discrimination: an MCM in quantum

theory gives rise to a higher maximum confidence over a noncontextual theory. We next consider a semi-device independent scenario with uncharacterised detectors. We develop the framework of certifying the maximum confidence in the scenario given a preparation of states and detected events. It is shown that the certifiable maximum confidence in quantum theory contains contextual advantages in the realistic scenario that may contain undetected events. Our results provide the unifying framework for the existence and the certification of contextual advantages in a realistic quantum state discrimination scenario.

The paper is organised as follows. In Sec. 4.3, we begin with a summary of different figures of merits in quantum state discrimination. The contextual advantages for minimum error, unambiguous and maximum confidence quantum state discrimination are then shown in Sec. 4.4. We then present the certification of an MCM in Sec. 4.5. Two-input and three-outcome scenarios, with one arm containing the undetected events only, are considered. In Sec. 4.6, we compare quantum and noncontextual theories in the certification of an MCM, then include noise in our model in Sec. 4.7. Finally, we summarise the results and discuss related questions in Sec. 4.8.

## 4.3 Background

Let us begin by collecting the terminology and notation to be used throughout the manuscript. We also summarise different figures of merits in quantum state discrimination.

For convenience, state discrimination can be framed as a communication protocol for two parties, named Alice for preparation and Bob for measurement. Alice prepares her quantum system in one of the states in an ensemble of $n$ states, denoted by

$$\text{ensemble: } S_n = \{p_x, \rho_x\}_{x=0}^{n-1}, \tag{4.1}$$

which describes a state $\rho_x$ is generated with prior probabilities $p_x$ for $x = 1, \cdots, n$. Bob then performs an $n$ outcome measurement described by positive-operator-valued-measure (POVM) elements, denoted by

$$\text{measurement: } M_n = \{\hat{\pi}_b\}_{b=0}^{n-1}, \tag{4.2}$$

each of which may be optimised to give a correct guess about a state that has been prepared. For completeness, the condition $\sum_b \hat{\pi}_b = \mathbb{1}$ must be satisfied.

### 4.3.1 Minimum error and unambiguous state discrimination

In MESD, the figure of merit, called the success probability $p_{\text{suc}}$, is the highest probability of detecting $x$ correctly on average:

$$p_{\text{suc}} = \max \sum_{x=1}^{n} p_x \operatorname{Tr}\left[\rho_x \hat{\pi}_x\right], \tag{4.3}$$

where the maximisation runs over a complete measurement. A closed form of the maximal success probability is known for two states in general,

$$p_{\text{suc}} = \frac{1}{2} + \frac{1}{2}\|p_0\rho_0 - p_1\rho_1\|_1, \tag{4.4}$$

where $\|X\|_1 = \operatorname{Tr}\sqrt{X^\dagger X}$. Otherwise, a closed form has been found in some specific cases only [108, 109, 123, 124]. While the error, averaged over the states in $S_n$, is minimised, not all detection events lead to a correct guess.

Detection events in USD measurements identify states with certainty. This is possible if the probability of outcome $b$ given a state $\rho_x$ is given by

$$P_{\text{M|P}}(b|x) := \operatorname{Tr}[\rho_x \hat{\pi}_b] \propto \delta_{x,b}\ , \tag{4.5}$$

where M and P denote a measurement and a preparation, respectively. This shows that the detector described by $\hat{\pi}_b$ responds to $\rho_x$ but not the other states. Under this condition, it may, however, appear that a measurement is not complete, i.e., $\sum_b \hat{\pi}_b < \mathbb{1}$. An additional outcome $\hat{\pi}_\emptyset$ is included to fulfill the completeness condition:

$$\sum_{b=0}^{n-1} \hat{\pi}_b + \hat{\pi}_\emptyset = \mathbb{1}. \tag{4.6}$$

The arm described by $\hat{\pi}_\emptyset$ collects those detection events which give ambiguous conclusions. Then, in the case of USD, a conclusion from a detection event is completely unambiguous since no error in the legitimate arms is permitted. For qubit states, this is possible only when two pure states can be prepared. Preparation of pure states with certainty would not be feasible in a realistic setting either. We can say that it is not practical to meet the conditions in (4.4) and (4.5) in MESD and USD, respectively.

### 4.3.2 Maximum confidence discrimination

The notion of confidence for a detection event in a discrimination task has been defined as the conditional probability [101]:

$$\text{confidence}: \; C_x = P_{\mathrm{P|M}}(x|x) \; , \tag{4.7}$$

i.e., the probability that a detection event corresponding to $\hat{\pi}_x$ correctly indicates that a preparation was $\rho_x$. One can interpret this retrodictively, as a detected event implying a conclusion about state preparation done in the past.

The confidence may be computed with quantum probabilities by using Bayes' rule,

$$C_x^Q = \frac{P_{\mathrm{P}}(x)P_{\mathrm{M|P}}(x|x)}{P_{\mathrm{M}}(x)} = \frac{p_x \,\mathrm{Tr}[\hat{\pi}_x\rho_x]}{\mathrm{Tr}[\hat{\pi}_x\rho]}, \tag{4.8}$$

where $P_{\mathrm{M}}(x)$ is the probability of a detection event on the detector $\hat{\pi}_b$ for an ensemble $\rho = \sum_x p_x\rho_x$ and $P_{\mathrm{P}}(x) = p_x$ the *a priori* probability. Hence, an MCM aims to maximises the confidence above. Throughout, an MCM in quantum theory is denoted by

$$\max \; C_x^Q \; , \tag{4.9}$$

where the maximisation runs over all measurements. It should be added that it does not matter whether a particular $x$ is optimized or all $x$ are optimized simultaneously. The confidence for a particular state will depend upon a single POVM element only, so the measurement can be completed in any matter. The only difference between individual and ensemble optimisation is the possibility of different inconclusive outcome rates. In the present work, our focus is on the maximal confidence that we can have in a *particular* state and so the distinction can be ignored. Note that an MCM can be defined for any ensemble in (4.1).

We remark that maximum confidence state discrimination is well-fitted to a realistic scenario including imperfect preparations and measurements. Firstly, it can be adapted to cases where the detected measurement statistics are not complete whereas MESD can only find the optimal guessing probability whenever a measurement is complete. As MCM is concerned with detected events only, undetected ones can be counted as ambiguous outcomes. Secondly, an MCM can be considered for ensembles for which unambiguous measurement outcomes can not be obtained. MCM presents, for these

reasons, a more realistic setting for identifying a state among a given ensemble.

Maximum confidence state discrimination also provides a unifying framework of the aforementioned figures of merits in state discrimination. An MCM coincides with USD if $C_x = 1$ for all $x$. In this sense, whenever USD is possible for an ensemble, it will emerge as the MCM. One can also apply an MCM to maximise the success probability over an ensemble or a sub-ensemble by taking into account in the possibility of undetected events occurring:

$$\max \sum_x P_M(x) C_x = \max \sum_x p_x \operatorname{Tr}[\hat{\pi}_x \rho_x] \tag{4.10}$$

where the maximisation runs over a complete measurement. An MCM as defined above reproduces MESD if the inconclusive outcome rate is zero. It is also worth noting that optimal measurements for MESD, USD, and maximum confidence discrimination are generally not identical [101].

## 4.4 Contextual advantages for quantum state discrimination

It is central to the field of quantum information theory to find circumstances in which quantum experiments perform differently to their classical equivalents. This distinction will depend upon the chosen notion of (non)classicality. In many cases, the most relevant definition is that classical theories do not violate Bell-like inequalities, whereas nonclassical ones do. Space-like separated correlations, however, are required in order for this definition to be useful but do not appear in all experimental applications. Bell violations, in addition, will occur only for the particular class of nonseparable states.

Noncontextuality, the independence of experimental statistics from the context in which they are gathered, is a more widely applicable notion of classicality [125]. It is defined independently of the choice of state and is valid for experiments, such as state discrimination, that are spacelike local. The concept has been introduced by Kochen and Specker [19] and more recently generalized by Spekkens [20].

In all forms, noncontexuality acts as a constraint on various objects within the theory. By "objects", we mean states, channels or transformations. In the original proof of the Kochen-Specker theorem, which states that quantum theory is contextual, a set of 117 projective measurements are shown to violate

noncontextuality. This means that we must constrain any classical theory to contain less than this number of possible measurements. In a similar manner, Spekkens has shown that the "trine ensemble," the set of three symmetric states on the Bloch sphere, cannot be prepared in a noncontextual theory. It follows that quantum theory is able to recreate a greater range of experimental statistics.

A number of "noncontextual inequalities" have been derived [27, 126, 127]. In these, a particular figure of merit is optimized within a theory constrained by noncontextuality. The maximal value of some figures of merit, for the aforementioned reasons, is less than what can be obtained by quantum theories and the latter consequently violate such inequalities. In such a case, it is said that quantum theory contains "contextual advantages". This question has been addressed in Ref. [27], where it has been shown that the MESD of quantum states contains contextual advantages.

In this section, we consider USD and MCM and show contextual advantages. For the latter case, a pair of mixed states for which USD cannot be achieved are considered. Thus, we show contextual advantages for quantum state discrimination in general. We begin with a review of noncontextual ontological models and then consider MESD, USD and MCM.

### 4.4.1 Noncontextual ontological model

An operational theory contains descriptions of possible operations, such as preparations and measurements, and a prescription for calculating probabilities of measurement outcomes. Here, let us review noncontextual ontological models [20, 27, 128] and characterise preparation noncontextuality.

Let $\Lambda$ denote an ontic state space so that an element $\lambda \in \Lambda$ fully characterises the physical properties of a given system. A state preparation $x$ corresponds to a sample of the ontic state space up to a probability distribution $\mu_x(\lambda)$, which is called an epistemic state. A measurement $M$ contains a set of possible outcomes that occur with a dependence on the ontic state space. An outcome denoted by $b$ is represented by a response function $\xi_{b|M}(\lambda)$ that satisfies

$$\begin{aligned} &\text{positivity :} && \xi_{b|M}(\lambda) \geq 0,\ \forall b,\ \forall \lambda \\ &\text{completeness :} && \sum_b \xi_{b|M}(\lambda) = 1,\ \forall \lambda\ , \end{aligned} \tag{4.11}$$

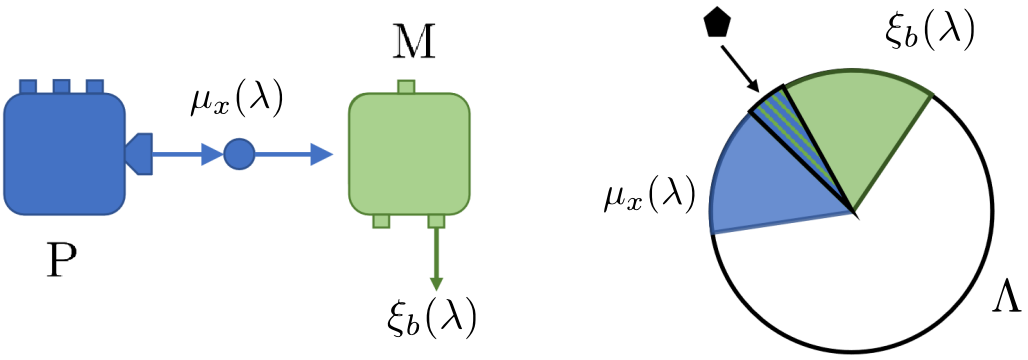



Figure 4.1: Figure extracted from Ref. [25]. A prepare-and-measure scenario in an ontological model (left) is modelled by the ontic state space $\Lambda$ (right). Note that P denotes a preparation represented by an epistemic state $\mu_x(\lambda)$ and M a measurement by a response function $\xi_b(\lambda)$. The probabilities extracted from the theory are given by integrals over the overlap marked by a black pentagon ( see (4.13) ).

so that it is interpreted probabilistically. Probabilities extracted from the ontological model with a preparation $x$ and a measurement $M$ are given by

$$p(b|x,M) = \int_\Lambda d\lambda\ \mu_x(\lambda)\xi_{b|M}(\lambda). \tag{4.12}$$

The preparation noncontextuality criterion is then identified as follows. Consider two preparations $\mu_x(\lambda)$ and $\mu_{x'}(\lambda)$ that cannot be distinguished by any measurement, i.e., no response function provides different probabilities for the preparations. These preparations are called *operationally equivalent*. A model is then *preparation noncontextual* if the operational equivalence of a set of preparatory processes implies that they are represented by the same epistemic state:

$$p(b|x,M) = p(b|x',M), \quad \forall\{b|M\} \implies \mu_x(\lambda) = \mu_{x'}(\lambda) \quad \forall\lambda \in \Lambda. \tag{4.13}$$

Measurement noncontextuality can also be defined in a similar manner.

Having introduced an operational framework above, we use definitions and notations in the following manner. For an epistemic state $\mu_x(\lambda)$, a non-overlapping state is denoted by $\mu_x^\perp(\lambda)$ such that

$$\mu_x(\lambda)\mu_x^\perp(\lambda) = 0, \quad \forall\lambda \in \Lambda. \tag{4.14}$$

In preparation-noncontextual theories, an epistemic state $\mu_x(\lambda)$ uniquely defines its non-overlapping state $\mu_x^\perp(\lambda)$. This does not necessarily hold for

other theories. The support of an epistemic state $\mu_x$ is defined as follows:

$$\text{supp}[\mu_x(\lambda)] = \{\lambda \in \Lambda \ : \ \mu_x(\lambda) \neq 0\}. \tag{4.15}$$

For instance, we have $\text{supp}[\mu_x(\lambda)] \cap \text{supp}[\mu_x^{\perp}(\lambda)] = \emptyset$.

An important set of response functions is the set representing projectors from quantum theory. In quantum theory, each POVM element $E_b = |b\rangle\langle b|$ of a projective measurement satisfies $\text{Tr}[E_b|x\rangle\langle x|] = \delta_{x,b}$ for an ensemble $\rho_x = |x\rangle\langle x|$ for $x = 1, ..., N$. In an operational theory, $E_b$ is represented by $\xi_b(\lambda)$ and $\rho_x$ by $\mu_x(\lambda)$. It has been shown that, in noncontextual theories, the corresponding response functions take the form [27]

$$\xi_b(\lambda) = \begin{cases} 1 & \text{if } \lambda \in \text{supp}[\mu_b(\lambda)] \\ 0 & \text{otherwise ,} \end{cases} \tag{4.16}$$

in which $\mu_b(\lambda)$ are epistemic states representing pure-state preparations. In this sense, they are *outcome deterministic*.

As for two-state discrimination in a noncontextual model, a useful quantity is the confusability, which is the probability of finding one state $\mu_x(\lambda)$ given a measurement on a different state $\mu_y(\lambda)$ [27, 129]. In a preparation noncontextual model, the confusability for a pair of states $\mu_x(\lambda)$ and $\mu_y(\lambda)$ can be defined as follows:

$$c_{x,y} = \int_{\text{supp}[\mu_x(\lambda)]} d\lambda \mu_y(\lambda). \tag{4.17}$$

In quantum theory, the confusability for two pure states can be identified with the state overlap

$$c_{x,y} = \text{Tr}[|\psi_x\rangle\langle\psi_x| \ |\psi_y\rangle\langle\psi_y|] = |\langle\psi_x|\psi_y\rangle|^2. \tag{4.18}$$

It is clear that that the confusability is symmetric, i.e., $c_{x,y} = c_{y,x}$.

### 4.4.2 Contextual advantages for MESD

The first instance of contextual advantages for quantum state discrimination has been shown for MESD. In Ref. [27], MESD for two states in a noncontextual model was considered, and contextual advantages for MESD of quantum states were shown.

Suppose that two quantum states $|\psi_0\rangle$ and $|\psi_1\rangle$ are provided, for which an optimal measurement for MESD is denoted by $M = \{\hat{\pi}_0, \hat{\pi}_1\}$. Two states can be characterised by the angle between them,

$$\cos\theta = |\langle\psi_0|\psi_1\rangle| = \sqrt{c_{0,1}} \tag{4.19}$$

where $c_{0,1}$ is the confusability and we take the positive-valued square root. It suffices to consider a two dimensional Hilbert space. It is clear that one can find the statistics of measurement outcomes from the states and the measurement. The guessing probability for two quantum states in (4.4) can be straightforwardly computed.

The ensemble consisting of the states $|\psi_0\rangle$ and $|\psi_1\rangle$ only, however, does not imply any equivalence relations so that noncontextuality cannot yet be used to constrain the model. Another pair of states, $|\psi_0^\perp\rangle$ and $|\psi_1^\perp\rangle$, must be used. The overlap and optimal guessing probability of this ensemble are identical to those of the former. The two pairs of states are related by

$$\frac{1}{2}\left(|\psi_0\rangle\langle\psi_0| + |\psi_0^\perp\rangle\langle\psi_0^\perp|\right) = \frac{1}{2}\left(|\psi_1\rangle\langle\psi_1| + |\psi_1^\perp\rangle\langle\psi_1^\perp|\right) = \frac{\mathbb{1}}{2}\ , \tag{4.20}$$

where $\mathbb{1}$ is the identity operator for the subspace spanned by the two states. This provides an equivalence relation between the two quantum ensembles which can be used to derive relations between epistemic states.

A noncontextual model is then constructed such that it is consistent with this equivalence relation. Two epistemic states, denoted by $\mu_0(\lambda)$ and $\mu_1(\lambda)$, can be introduced so that they have the same confusability $c_{0,1}$ with the quantum states (4.19), see also (4.17) and (4.18). The state space in a noncontextual model should also satisfy the equivalence relation. This implies that there exist mirrored states $\mu_0^\perp(\lambda)$ and $\mu_1^\perp(\lambda)$ such that

$$\frac{1}{2}\mu_0(\lambda) + \frac{1}{2}\mu_0^\perp(\lambda) = \frac{1}{2}\mu_1(\lambda) + \frac{1}{2}\mu_1^\perp(\lambda) = \mu_{\mathbb{1}/2}(\lambda), \tag{4.21}$$

where $\mu_{\mathbb{1}/2}(\lambda)$ is the *maximally mixed state* in a noncontextual model analogous to $\mathbb{1}/2$ in the quantum theory. This is consistent with (4.20). Note also that the mirrored states share the same confusability with the original pair.

In Ref. [27], it is shown that the preparation nontextuality constrains the statistics in terms of various sharp measurements (see (4.16) ) and finds the success probability as follows:

$$p_{\text{suc}}^{\text{NC}} = 1 - \frac{1}{2}c_{0,1} \tag{4.22}$$

which is strictly less than the quantum bound in (4.4), i.e.,

$$p_{\text{suc}}^{\text{Q}} = \frac{1}{2} + \frac{1}{2}\sqrt{1 - c_{0,1}}. \tag{4.23}$$

This result is significant in that it shows that the predictions of noncontextual theories differ quantitatively from those of quantum theory. The results can also apply to mixed states when noise is present.

### 4.4.3 Contextual advantages for USD

Another scenario in state discrimination is USD, where, rather than finding the highest success probability over an ensemble, each state is identified with certainty. As with MESD, a noncontextual model of USD can be constructed. Given the constraint of USD, the aim is to minimise the probability of having inconclusive outcomes. In what follows, we show USD in a noncontextual theory and derive a noncontextual inequality associated with the rate of inconclusive outcomes, from which contextual advantages for quantum USD are shown.

#### Quantum states

Let us first consider two pure quantum states $|\psi_0\rangle$ and $|\psi_1\rangle$ for which USD can be performed. The POVM elements may be given as

$$\hat{\pi}_0 \propto |\psi_1^{\perp}\rangle\langle\psi_1^{\perp}| \text{ and } \hat{\pi}_1 \propto |\psi_0^{\perp}\rangle\langle\psi_0^{\perp}| \tag{4.24}$$

where $\langle\psi_0^{\perp}|\psi_0\rangle = \langle\psi_1^{\perp}|\psi_1\rangle = 0$. An additional POVM element $\hat{\pi}_{\text{ø}}$ is needed to collect inconclusive outcomes. The probability of inclusive outcomes for the quantum states denoted by $\eta_{\text{ø}}^{Q}$ is known to be [79–81],

$$\min \eta_{\text{ø}}^{Q} = |\langle\psi_0|\psi_1\rangle| = \sqrt{c_{0,1}} \tag{4.25}$$

where the minimization runs over complete measurements and $c_{0,1}$ is the confusability in (4.18). In Fig. 4.2, the probability in (4.25) is plotted.

#### States with preparation noncontextuality

We then consider USD for states with preparation noncontextuality. Let us first investigate constraints on response functions $\xi_b(\lambda)$. Note that a response function corresponding to a sharp measurement can be expressed in the form

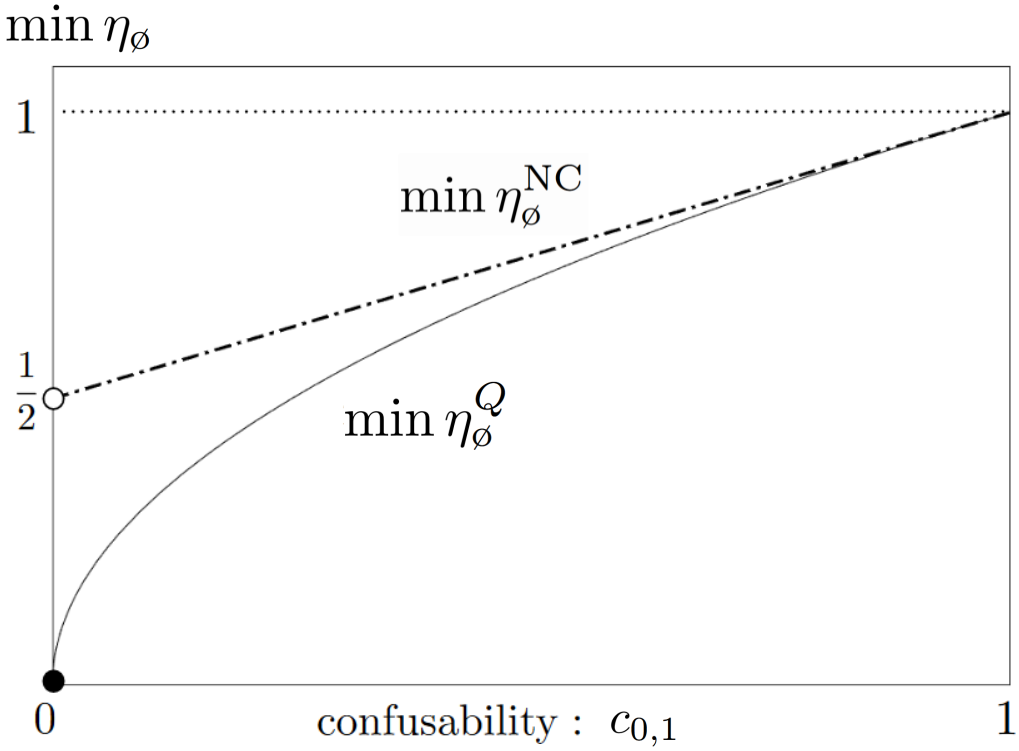



Figure 4.2: Figure extracted from Ref. [25]. The minimum value of the inconclusive error rate $\eta_\emptyset$ is plotted against the confusability $c_{0,1} > 0$ for preparation noncontextual theories (dashed) and quantum theory (solid). It is shown that the former is strictly greater than the latter, meaning that USD for quantum states contains advantages over a noncontextual model. In a noncontextual theory, the probability of inconclusive outcomes is 0 for $c_{0,1} = 0$. There is a sharp discontinuity, see the main text.

of (4.16). This can be generalised by including a probabilistic mixture of measurement outcomes. Hence, the most general form of a response function $\xi_b(\lambda)$ unambiguously identifying an outcome $b = x$ takes the form

$$\xi_x(\lambda) = \begin{cases} q & \text{if } \lambda \in \text{supp}[\mu_x(\lambda)] \\ 0 & \text{if } \lambda \in \text{supp}[\mu_x^{\perp}(\lambda)] \end{cases} \tag{4.26}$$

for an epistemic state $\mu_x(\lambda)$, which can be freely chosen, and $0 \leq q \leq 1$. A response function with the structure above may represent a POVM element in the form $q|\psi_b\rangle\langle\psi_b|$ in quantum theory.

The condition that a measurement outcome gives an unambiguous conclusion is

$$\text{P}_{\text{M|P}}(\xi_b|\mu_x) \propto \delta_{x,b} \ , \tag{4.27}$$

for $\delta_{i,j}$ being the Kronecker delta. The condition, applied to a response function in the form of (4.26), identifies the following response function for two-state USD:

$$\xi_0(\lambda) = \begin{cases} q & \text{if } \lambda \in \text{supp}[\mu_1^{\perp}(\lambda)] \\ 0 & \text{if } \lambda \in \text{supp}[\mu_1(\lambda)]. \end{cases} \tag{4.28}$$

The same argument applies to the other response function $\xi_1(\lambda)$. Note that two states are given with an equal *a priori* probability. Because the confusability of $\mu_0(\lambda)$ with $\mu_1^{\perp}(\lambda)$ is the same as that for $\mu_1(\lambda)$ with $\mu_0^{\perp}(\lambda)$, we can, without loss of generality, take the weighting parameter $q \in [0, 1]$ to be the same for both response functions $\xi_0(\lambda)$ and $\xi_1(\lambda)$. The probability of unambiguous outcomes is thus proportional to $q$, which we hence aim to maximise. Equivalently, the probability of inconclusive outcomes is to be minimised.

In fact, two response functions $\xi_0(\lambda)$ and $\xi_1(\lambda)$ do not form a complete measurement for the same reason as in USD for quantum states: completeness enforces that $\sum_b \xi_b(\lambda) = 1$ for all $\lambda \in \Lambda$. It is necessary to have an additional response function denoted by $\xi_{\text{ø}}(\lambda)$ that collects all inconclusive outcomes

$$\xi_{\text{ø}}(\lambda) = 1 - \xi_0(\lambda) - \xi_1(\lambda). \tag{4.29}$$

Note also that $\xi_{\text{ø}}(\lambda) \geq 0$ for all $\lambda$. The region in which the probability of inconclusive outcomes is minimal can be characterised by the subset

$$\{\lambda \in \Lambda \ : \ \lambda \in \text{supp}[\mu_0(\lambda)] \cap \text{supp}[\mu_1(\lambda)]\} \tag{4.30}$$

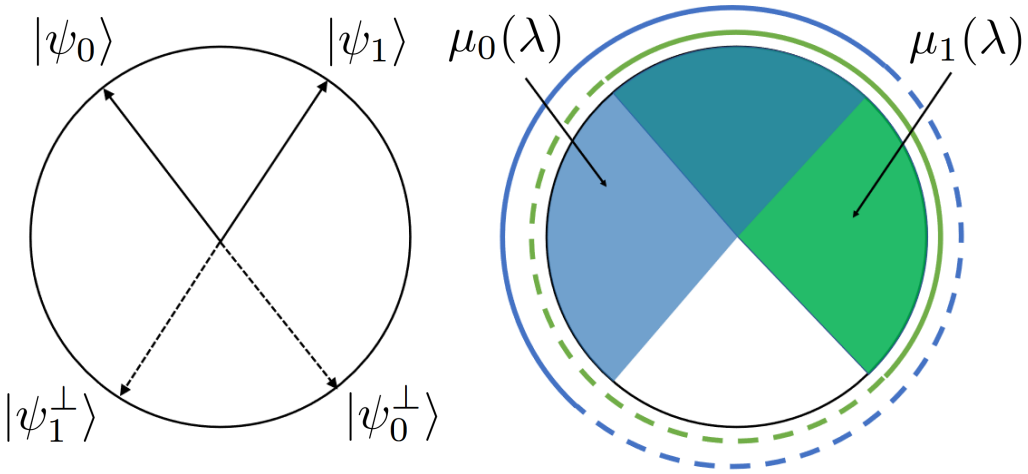



Figure 4.3: Figure extracted from Ref. [25]. Two pure states $|\psi_0\rangle$ and $|\psi_1\rangle$ are considered in USD, for which an optimal measurement corresponds to their orthogonal states $|\psi_0^\perp\rangle$ and $|\psi_1^\perp\rangle$, respectively. The inconclusive outcomes are collected by a POVM element constructed by an equal mixture of given states. The structure is shared with USD of two states $\mu_0(\lambda)$ and $\mu_1(\lambda)$, diagrammed by blue and green regions respectively. The outer lines signify the supports of reponse functions. $\xi_0(\lambda)$ has the same support as $\mu_1^\perp(\lambda)$ (green-dashed line) and $\xi_1(\lambda)$ has the same support as $\mu_0^\perp(\lambda)$ (blue-dashed line).

where both $\xi_0(\lambda)$ and $\xi_1(\lambda)$ are nonzero. Using the response functions above, it holds that $\xi_\varnothing(\lambda) = 1 - 2q$ in the region. The maximization of $q$ thus corresponds to minimising the response function $\xi_\varnothing(\lambda) \geq 0$: one can find $q = 1/2$.

The response functions for USD in (4.26) are thus given by

$$\xi_0(\lambda) = \begin{cases} \frac{1}{2} & \text{if } \lambda \in \text{supp}[\mu_1^\perp(\lambda)] \\ 0 & \text{if } \lambda \in \text{supp}[\mu_1(\lambda)] \end{cases}, \quad \xi_1(\lambda) = \begin{cases} \frac{1}{2} & \text{if } \lambda \in \text{supp}[\mu_0^\perp(\lambda)] \\ 0 & \text{if } \lambda \in \text{supp}[\mu_0(\lambda)] \end{cases}. \tag{4.31}$$

From above and (4.29), the response function giving inconclusive outcomes can be obtained :

$$\xi_\varnothing(\lambda) = \frac{1}{2}\left(1 - 2\xi_0(\lambda)\right) + \frac{1}{2}\left(1 - 2\xi_1(\lambda)\right) = \frac{1}{2}\overline{\xi}_0(\lambda) + \frac{1}{2}\overline{\xi}_1(\lambda), \tag{4.32}$$

with $\overline{\xi}_b(\lambda) = 1 - 2\xi_b(\lambda)$ for $b = 0, 1$. Let us write this as

$$\overline{\xi}_0(\lambda) := \xi_{\mu_1}(\lambda) = \begin{cases} 1 & \text{if } \lambda \in \text{supp}[\mu_1(\lambda)] \\ 0 & \text{if } \lambda \in \text{supp}[\mu_1^\perp(\lambda)], \end{cases} \tag{4.33}$$

with $\xi_{\mu_1}(\lambda)$ corresponding to a sharp measurement for the epistemic state $\mu_1(\lambda)$. The same argument also applies to the response function $\overline{\xi}_1(\lambda) = \xi_{\mu_0}(\lambda)$. Bringing all of these together, we have the response function for inconclusive outcomes as follows,

$$\xi_{\emptyset}(\lambda) = \frac{1}{2}\xi_{\mu_0}(\lambda) + \frac{1}{2}\xi_{\mu_1}(\lambda). \tag{4.34}$$

It is therefore shown that the response function is given by a convex combination of two response functions which correspond to sharp measurements for the states in the ensemble.

In fact, the measurement can be operationally realized by applying two complete sets

$$\{\xi_{\mu_0}(\lambda), \xi_{\mu_0^\perp}(\lambda)\} \text{ and } \{\xi_{\mu_1}(\lambda), \xi_{\mu_1^\perp}(\lambda)\}\ , \tag{4.35}$$

with probability 1/2, respectively. Outcomes with $\xi_{\mu_b^\perp}(\lambda)$ for $b = 0, 1$ collect inconclusive outcomes and the others lead to unambiguous conclusions. The relevant epistemic states are also depicted in Fig. 4.3 alongside the analogous quantum states.

It is clear that the measurement leads to USD in the following sense. In quantum theory, a measurement strategy of USD consists of three outcomes, two of which show unambiguous detection events and the third of which gives an inconclusive result. In the case of the response functions obtained in a noncontextual theory, there is no chance that the outcome $\xi_{\mu_1^\perp}(\lambda)$ occurs if $\mu_1(\lambda)$ is prepared. The epistemic state $\mu_0(\lambda)$ will likewise never result in the detector associated with the response function $\xi_{\mu_0^\perp}(\lambda)$ being triggered. These results are, therefore, unambiguous. The remaining outcomes are $\xi_{\mu_0}(\lambda)$ and $\xi_{\mu_1}(\lambda)$ and could be triggered by either of the possible epistemic state. These outcomes are collected into the inconclusive outcomes.

Having characterised the optimal measurement, we are now in a position to compute the rate of inconclusive outcomes in noncontextual theories. Given the measurement shown above, the probability of inconclusive outcomes is given as

$$\min_{\xi_{\emptyset|M}} \eta_{\emptyset}^{\mathrm{NC}} = \int_\Lambda d\lambda \frac{1}{2}\left(\mu_0(\lambda) + \mu_1(\lambda)\right)\xi_{\emptyset|M}(\lambda) = \frac{1}{2}\left(1 + c_{0,1}\right). \tag{4.36}$$

In Fig. 4.2, the probabilities of inconclusive outcomes in quantum theory and a contextual model are compared. Hence, contextual advantages for USD of

quantum states are shown.

The caveat is the case when $c_{0,1} = 0$, where the rate of inconclusive outcomes in a noncontextual theory is in fact given by 0. By definition, USD is possible with no inconclusive outcomes. It should be noted that the parameter $q$ in the sharp measurement in (4.26) can be made equal to one when there is no overlap between the desired response functions. As soon as their supports have some non-zero overlap, no matter how small that region is, the framework enforces that $q \leq 1/2$. There is a discontinuity in the probability of inconclusive outcomes in a noncontextual theory. Therefore, the probability in (4.36) is valid for $c_{0,1} > 0$ only.

Finally, it is worth mentioning a physical reason that the aforementioned measurement is a form of USD in a noncontextual theory. There are two classes of measurement possible in quantum theory. Most simply, we can perform projective measurements and probalistically mix the outcomes. Outside of this, we can access a greater set of measurements by entangling the system with an ancilla and then projectively measuring the latter, following the Neumark dilation theorem. An example of this type would be a measurement of the three symmetric qubit states, which requires entanglement with a qutrit. However, as this resource is not available in a noncontextual theory, only the first class can be implemented. Indeed, it has been previously shown that the correlations available to a preparation noncontextual model must be local [27]. This prevents access to the wider class of POVM elements and we are restricted to the form which was just found.

### 4.4.4 Contextual advantages for MCM

In this subsection, we consider two mixed quantum states for which USD cannot be achieved. Maximum confidence discrimination can be, however, defined, for which we show contextual advantages over a noncontextual model.

#### MCM in quantum theory

We consider a pair of mixed quantum states given with equal *a priori* probabilities ($p_x = 1/2$),

$$\rho_0 = r\left|\psi_1\right\rangle\left\langle\psi_1\right| + (1-r)\frac{\mathbb{1}}{2}\,, \qquad \rho_1 = r\left|\psi_1\right\rangle\left\langle\psi_1\right| + (1-r)\frac{\mathbb{1}}{2}\,, \tag{4.37}$$

The confusability for two pure states is denoted by $|\left\langle\psi_0|\psi_1\right\rangle| = \sqrt{c_{0,1}}$. It is straightforward to find an MCM for a quantum state. Following (4.8), we

must evaluate

$$\max C_0^Q = \max \frac{p_0 \operatorname{Tr}[\hat{\pi}_0 \rho_0]}{\operatorname{Tr}[\hat{\pi}_0 \rho]} \,, \tag{4.38}$$

where the maximisation runs over POVM elements and $\rho = (\rho_0 + \rho_1)/2$ denotes the ensemble of given states. The maximisation above can be solved as [101],

$$\max C_0^Q = ||\sqrt{\rho}^{-1} p_0 \rho_0 \sqrt{\rho}^{-1}||_\infty \tag{4.39}$$

where $||\cdot||_\infty$ denotes an operator norm. One can find the maximum confidence and write it in terms of the confusability as follows,

$$\max C_0^Q = \frac{1}{2}\left(1 + \frac{r\sqrt{1-c_{0,1}}}{\sqrt{1-r^2 c_{0,1}}}\right). \tag{4.40}$$

Note that the noiseless case $r = 1$ considering two pure states reproduces USD where the maximum confidence is 1.

### MCM in a noncontextual model

In Sec. 4.4.3, epistemic states $\mu_x(\lambda)$ are associated with pure states $|\psi_x\rangle$ for $x = 0, 1$. Here, we consider a noisy preparation in a noncontextual model in the following:

$$\begin{aligned} \widetilde{\mu}_0(\lambda) &= r\mu_0(\lambda) + (1-r)\mu_{\mathbb{1}/2}(\lambda), \\ \widetilde{\mu}_1(\lambda) &= r\mu_1(\lambda) + (1-r)\mu_{\mathbb{1}/2}(\lambda). \end{aligned} \tag{4.41}$$

The overall ensemble is then given by

$$\begin{aligned} \mu_P(\lambda) &= \frac{1}{2}\widetilde{\mu}_0(\lambda) + \frac{1}{2}\widetilde{\mu}_1(\lambda) \\ &= (1-r)\mu_{\mathbb{1}/2}(\lambda) + r\left(\frac{1}{2}\mu_0(\lambda) + \frac{1}{2}\mu_1(\lambda)\right). \end{aligned} \tag{4.42}$$

The goal is now to compute the maximum confidence, denoted by $\max C_1^{NC}$, for the state $\widetilde{\mu}_1(\lambda)$ above, and compare it with the quantum counterpart in (4.40).

The confidence is defined as

$$C_0^{NC} = \frac{1}{2\eta_0}\int_\Lambda d\lambda \mu_0(\lambda)\xi_b(\lambda) \tag{4.43}$$

where $\eta_0$ denotes the outcome rate defined by the ensemble and the response function:

$$\eta_0 = \int_\Lambda d\lambda \mu_P(\lambda)\xi_b(\lambda). \tag{4.44}$$

Since our aim is to determine the support of our epistemic state, it is natural to consider a sharp measurement. This can be justified by considering a response function as a mixture of sharp measurements, $\xi_i(\lambda) = \alpha\xi_j(\lambda)+\beta\xi_k(\lambda)$, with $\alpha, \beta \in [0, 1]$ such that $\xi_i(\lambda)$ is a proper response function. A simple calculation reveals that the confidence given by this measurement will take the form

$$C_0^{NC} = 1 - \alpha\frac{p\left(\xi_j|\mu_0\right)}{\eta_0} - \beta\frac{p\left(\xi_k|\mu_0\right)}{\eta_0}\ . \tag{4.45}$$

It is readily seen that a nonsharp measurement will always be less confident than a sharp one. Note that the special case $\alpha = \beta = 0$ is ruled out, as it would give $\eta_0 = 0$. without loss of generality, then, we can represent our maximally confident response function $\xi_0(\lambda)$ as a sharp measurement of the placeholder epistemic state $\mu_x(\lambda)$:

$$\overline{\xi}_b(\lambda) = \begin{cases} 1 & \text{if } \lambda \in \text{supp}[\mu_b(\lambda)] \\ 0 & \text{if } \lambda \in \text{supp}[\mu_1^\perp(\lambda)]. \end{cases} \tag{4.46}$$

the probability of outcome $\xi_b(\lambda)$ given that the state $\mu_i(\lambda)$ is prepared will be given by the confusability $c_{i,b}$, which simplifies our notation in what follows.

The outcome rate can be rewritten by using (4.42),

$$\eta_0 = (1-r)\int_\Lambda d\lambda \mu_{\mathbb{1}/2}(\lambda)\xi_b(\lambda) + \frac{r}{2}\int_\Lambda d\lambda(\mu_0(\lambda)+\mu_1(\lambda))\xi_b(\lambda). \tag{4.47}$$

In a noncontextual theory, it holds that for all $x$, the maximally mixed epistemic state can be written as

$$\mu_{\mathbb{1}/2} = \frac{1}{2}\left(\mu_x(\lambda) + \mu_x^\perp(\lambda)\right) \tag{4.48}$$

where $\mu_x^\perp(\lambda)$ is the non-overlapping epistemic state to $\mu_x(\lambda)$. This means that the first integral above is equal to $1/2$. The other integral can be expressed in terms of the confusability so that the outcome rate can be written as

$$\eta_0 = \frac{1}{2}\left((1-r) + r(c_{0,b} + c_{1,b})\right). \tag{4.49}$$

The same argument applies to evaluating the numerator in (4.43). After all these steps, we obtain

$$C_0^{NC} = \frac{1}{2}\left(1 + \frac{r(c_{0,b} - c_{1,b})}{(1-r) + r(c_{0,b} + c_{1,b})}\right) , \tag{4.50}$$

which is characterised in terms of the noise parameter $r$ and the confusabilities $c_{0,b}$ and $c_{1,b}$.

It remains to maximise the confidence over response functions. That is, one should maximise the difference between the confusabilities $c_{0,b}$ and $c_{1,b}$ while minimising their sum. On one hand, we recall from MESD that the following relation holds

$$\int_\Lambda d\lambda \left(\frac{1}{2}\mu_0(\lambda)\xi_b(\lambda) + \frac{1}{2}\mu_1(\lambda)\xi_{\bar{b}}(\lambda)\right) \leq 1 - \frac{c_{0,1}}{2} . \tag{4.51}$$

Note that $\xi_{\bar{b}}(\lambda) = 1 - \xi_b(\lambda)$ for all $\lambda$. Substituting in this and rearranging then gives us

$$c_{0,b} - c_{1,b} \leq (1 - c_{0,1}) , \tag{4.52}$$

with equality if and only if $b = \bar{1}$, i.e., the response function is a sharp measurement of the epistemic state that has no overlap with $\mu_1(\lambda)$. Thus, the maximum of the difference $c_{0,b} - c_{1,b}$ is shown.

On the other hand, the sum $c_{0,b} + c_{1,b}$ can be constrained in the following way. It is bounded from above as follows,

$$c_{0,b} + c_{1,b} = 1 + c_{0,b} - c_{\bar{1},b} \leq 1 + (1 - c_{0,\bar{1}}) \leq 1 + c_{0,1} . \tag{4.53}$$

It is also bounded from below by

$$c_{0,b} + c_{1,b} = 2 - c_{\bar{0},b} - c_{\bar{1},b} \geq 2 - (1 + c_{\bar{0},\bar{1}}) \geq 1 - c_{0,1} . \tag{4.54}$$

To summarise, we show the upper and lower bounds

$$(1 - c_{0,1}) \leq c_{0,b} + c_{1,b} \leq (1 + c_{0,1}) . \tag{4.55}$$

Thus, the optimal choice by which the sum $c_{0,b}+c_{1,b}$ is minimised and also at the same time the difference $c_{0,b} - c_{1,b}$ is maximised is given by by $b = \bar{1}$. We can thus conclude that the maximum confidence in (4.50) is given by the response function $\xi_{\bar{1}}(\lambda)$. Note that the measurement is identical to that in USD. The maximum confidence is then given by

$$\max C_1^{NC} = \frac{1}{2}\left(1 + \frac{r(1 - c_{0,1})}{1 - rc_{0,1}}\right) \tag{4.56}$$

which is now determined by the noise parameter $r$ and the confusability $c_{0,1}$ only. The case of USD is reproduced by noiseless cases $r = 1$.

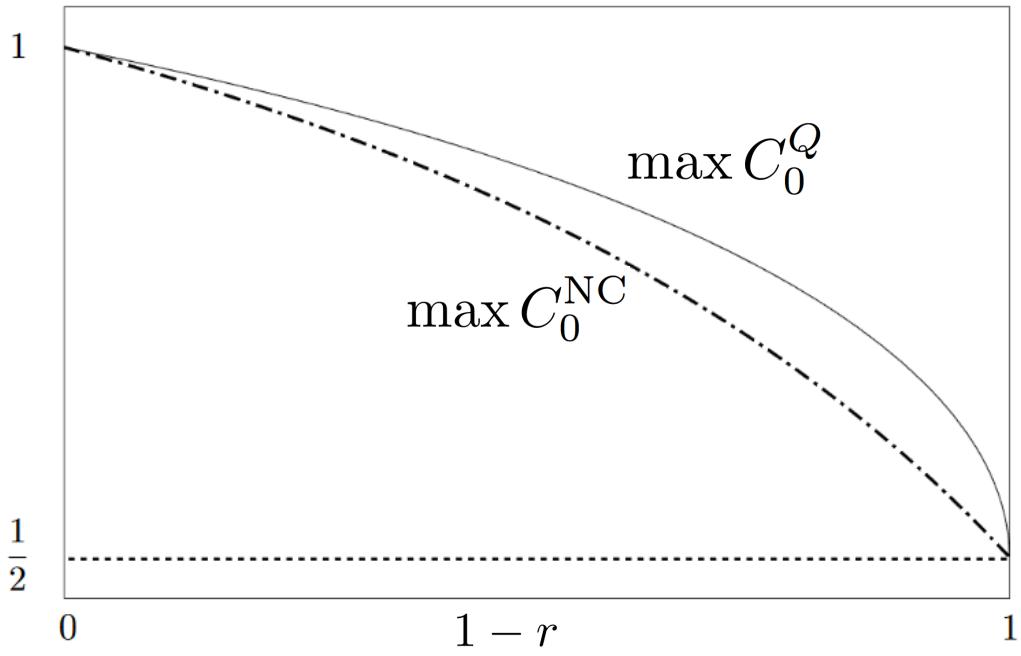



Figure 4.4: Figure extracted from Ref. [25]. The maximum confidence is computed for a pair of quantum states (solid) and also in a noncontextual model (dashed). The noise parameter is denoted by $r \in [0, 1]$, see (4.37) and (4.41): the case $r = 1$ reproduces USD. Contextual advantages for an MCM of noisy quantum states are shown for $r \in (0, 1)$.

**Comparison**

We compute the maximum confidence for quantum states in (4.40) and in a noncontextual model in (4.56). For $r \in (0, 1)$,

$$\max C_0^Q > \max C_0^{\mathrm{NC}} \tag{4.57}$$

holds, which shows contextual advantages for MCMs of quantum states, as seen in Fig. 4.4.

One can investigate optimal measurements for an MCM in quantum and noncontextual theories. In a noncontextual model, an MCM for the noisy states is identical to the measurement used in USD. This shows that an MCM does not rely on the noise parameter in (4.41). That is, the measurement that realises USD is also an MCM for noisy states in (4.41).

Interestingly, an optimal measurement that realises USD for two quantum states cannot be extended to noisy states in (4.37). Suppose that, for the POVM element that performs USD for a state $|\psi_0\rangle$, is given as $\hat{\pi}_0 \propto |\psi_1^\perp\rangle\langle\psi_1^\perp|$. If the measurement is performed on a noisy state $\rho_0$ in (4.37), it is not difficult to see that the maximum confidence is equal to (4.56) in a noncontextual model. No quantum advantage is concluded. In other words, the noncontextual bound in (4.56) can be reproduced in quantum theory by applying the USD measurement of the original states to the noisy states.

In fact, an MCM for the noisy states relies on the noise parameter $r$. To be explicit, an MCM is given by $\hat{\pi}_b \propto |\phi_b\rangle\langle\phi_b|$ for $b = 0, 1$ where

$$|\phi_b\rangle = \sqrt{\frac{1 - r\sqrt{c_{0,1}}}{2}}|0\rangle + (-1)^b \sqrt{\frac{1 + r\sqrt{c_{0,1}}}{2}}|1\rangle \tag{4.58}$$

for states $\rho_b$, respectively. With the measurement above, the maximum confidence for quantum states in (4.40) can be obtained, see also Fig. 4.5.

## 4.5 Certifying maximum confidence

So far, we have shown that quantum state discrimination in the forms of MESD, USD and MCM generally contains contextual advantages. However, a measurement in a realistic scenario consists of imperfections: it may be neither complete nor sharp. One can therefore ask if the quantum advantages for state discrimination can be obtained in practice when, in particular, undetected events are present.

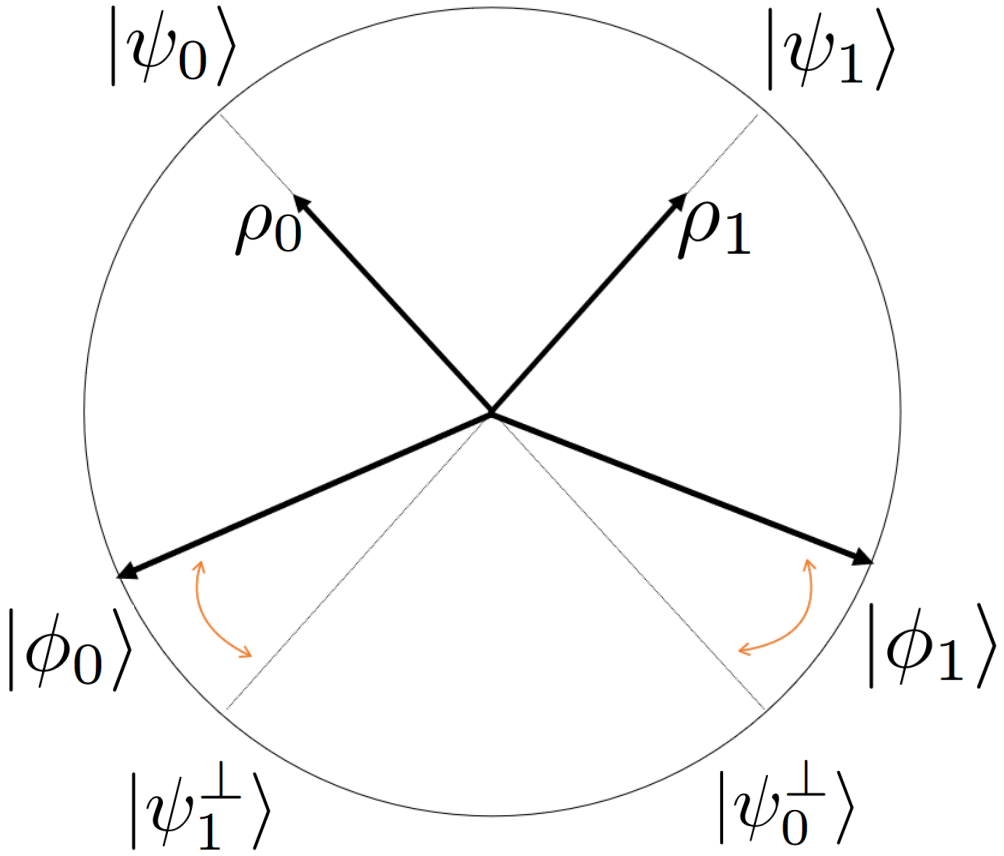



Figure 4.5: Figure extracted from Ref. [25]. An MCM for two states $\rho_0$ and $\rho_1$ in (4.37) is shown. An optimal POVM element $|\phi_b\rangle\langle\phi_b|$ for state $\rho_b$ for $b = 0, 1$ relies on the noise parameter $r$ ( see (4.58) ).

The aforementioned list of imperfections in a quantum measurement can be dealt with in a black-box scenario where the statistics of inputs, outputs, and their relations are available but one has no knowledge of the functioning of the measurement apparatus. Given the input and output statistics of a black-box measurement, one may be able to characterize the worst-case scenario. The most well-known instance of this approach is device independent quantum information processing, in which no assumptions are placed upon the measurement but quantum theory is assumed as a working principle [49, 50]. It often requires loophole-free Bell tests, which are difficult to implement in practice. In such a case, some assumptions may be relaxed. This is the case for quantum steering, in which two-party quantum correlations are characterized in a one-sided device independent manner [130]. One can then see a relation between the number of assumptions made on devices and the level of certification: the more one assumes, the weaker is the certification. In a semi-device independent (semi-DI) scenario, one considers all possible assumptions and then adjusts so that the scenario is realistic [65]. Quantum state discrimination in an semi-DI scenario is hence the most practical setting for finding which state among an ensemble is detected given statistics gathered with realistic quantum detectors.

In this section, we consider the realistic scenario of quantum state discrimination in an semi-DI scenario. Namely, a measurement is not yet characterised for an ensemble of quantum states and may also be incomplete. We present a framework for certifying the maximum confidence in the semi-DI scenario.

### 4.5.1 Semi-device independent scenario

Let us begin by presenting the semi-DI scenario to consider. A set of well-characterised $n$ states, as in (4.1), is assumed and detected events are provided. By repeating a prepare-and-measure experiment, the rates of detection events on the $n$ arms are collected. It is also assumed that states are prepared in an *independently and identically distributed* manner. The observed probabilities from detectors are denoted by

$$\text{outcome rate}: \ \eta_{\text{obs}} = \{\eta_b\}_{b=0}^{n-1,\emptyset} \tag{4.59}$$

where $\eta_b = \text{Tr}[\hat{\pi}_b\rho]$ for an ensemble $\rho = \sum_x p_x\rho_x$ and some POVM element $\hat{\pi}_b$. Note that $\eta_\emptyset$ denotes the collection of undetected events. The probability $\eta_b$ is called an *outcome rate* throughout. The certification scheme is desplayed in Fig. 4.6.

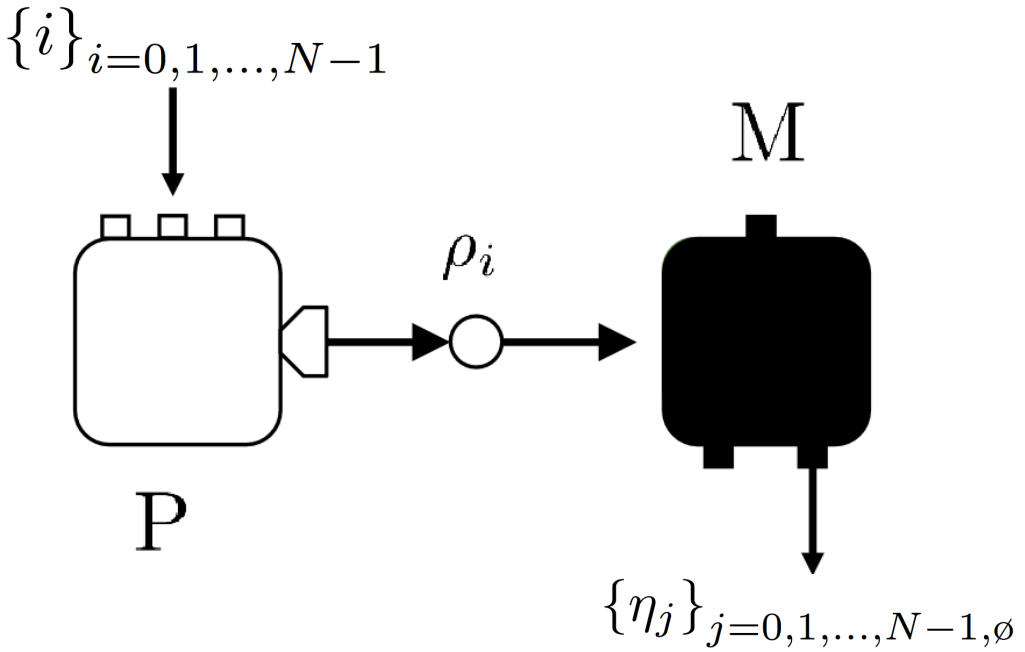



Figure 4.6: Figure extracted from Ref. [25]. The prepared quantum states are well-characterised (white). Detectors are arranged to determine which state has been sent. The certification of maximum confidence of a measurement can be obtained from outcome rates from untrusted detectors (black).

### 4.5.2 Certification of maximum confidence for quantum states

**The framework in quantum theory**

For full generality, we consider an MCM with a predetermined weight $\{\alpha_x\}_{b=0}^{n-1}$ denoted by

$$\langle C^Q \rangle_\alpha = \sum_{x=0}^{n-1} \alpha_x C_x^Q. \tag{4.60}$$

The parameters $\{\alpha_x\}$ may define a figure of merit in state discrimination. For instance, if they are identical to the outcome rates, i.e., $\alpha_x = \eta_x$ for $\forall x \in \{0, \cdots, n-1\}$, the MCM maximises a success probability in the absence of undetected events. This can be seen in the relation in (4.10). When considering an MCM for the $k$-th single detector only, one can put $\alpha_x = \delta_{x,k}$.

Given an ensemble $S_n$ in (4.1) and detected probabilities $\eta_{\rm obs}$ in (4.59), the certification of the maximum confidence is formulated as an optimization problem,

$$\begin{aligned} \text{maximise} \quad & \langle C^Q \rangle_\alpha \\ \text{subject to} \quad & \hat{\pi}_b \geq 0, \ \sum_{b=0}^{n-1} \hat{\pi}_b + \hat{\pi}_\emptyset = \mathbb{1} \\ & \mathrm{Tr}[\hat{\pi}_b \rho] = \eta_b, \quad b = 0, 1, \cdots, n-1, \emptyset \end{aligned} \tag{4.61}$$

where $\eta_\emptyset$ is the collection of undetected events. The optimisation problem can be solved by a semidefinite program (SDP). This SDP is computationally feasible. Note also that, as it is shown the above, the optimisation problem is equivalent to MESD of the $n$ states with prior probabilities $\{\alpha_x p_x/\eta_x\}_{x=0}^{n-1}$ where a measurement may be incomplete, i.e., $\sum_{x=0}^{n-1} \eta_x < 1$.

Note that the SDP in (4.61) is numerically feasible: one can obtain a solution by realizing the optimization numerically [69]. On the other hand, an analytic solution to the optimization problem in (4.61) can be attempted: the strategy is known as the linear complementarity problem (LCP), which directly considers the optimality conditions [112]. For instance, the case of MESD for qubit states has been approached by the LCP and a geometric method of finding an optimal measurement has been presented [108]. Technically, the LCP may be considered more difficult in that a larger set of parameters is taken into account. Its usefulness, however, lies in the fact that

the general structure of an optimization problem can be exploited so that analytic solutions can be achieved. The key fact in exploiting the LCP is the optimality conditions characterized by the Karush-Kuhn-Tucker (KKT) conditions [69, 112]. In this way, an optimization problem, either maximization and minimization, can be solved by equalities [112].

By following the techniques in convex optimization in Ref. [69], the optimality conditions are the Lagrangian stability and the complementary slackness are as follows:

$$\text{Lagrangian stability}: \ K = \alpha_x \frac{p_x}{\eta_x}\rho_x + r_x\sigma_x - s_x\rho, \ \text{and} \ \ K = r_\emptyset\sigma_\emptyset \tag{4.62}$$

$$\text{Complementary slackness}: \ r_x \operatorname{Tr}[\hat{\pi}_x\sigma_x] = 0, \ \forall x \tag{4.63}$$

with dual parameters $K$ and $\{s_x, r_x, \sigma_x\}$, in which $\{s_x\}$ and $\{r_x \geq 0\}$ are constants and $\{\sigma_x\}$ quantum states. Once these parameters satisfy the optimality conditions, they are automatically optimal parameters and therefore solve the problem in (4.61). With the optimal parameters $K^*$ and $\{s_x^*, r_x^*, \sigma_x^*\}$ that satisfy the conditions above, the maximum confidence is given as

$$\max\langle C^Q\rangle_\alpha = \operatorname{Tr}[K^*] + \sum_{x=0}^{n-1} s_x^*\eta_x^*. \tag{4.64}$$

A detailed derivation of the optimality conditions is shown in Sec. 4.9. Our SDP is universal in the sense that, as long as we are in the certification scenario, it can be applied to ensembles including any possible set of quantum states prepared with any probabilities.

### Certification of an MCM for a two-state ensemble

To illustrate the certification scenario, we consider two equally probable states

$$|0\rangle \text{ and } |+\rangle = \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle). \tag{4.65}$$

Let $C_0^Q$ denote the confidence for the first detector to conclude the state $|0\rangle$. The outcome rate in the first detector is also obtained as $\eta_0$. We implement a numerical package of an SDP for the optimisation problem in (4.61), from which the certifiable maximum confidence on the first detector is obtained as follows. When the outcome rate is in the range such that $\eta_0 < 1/4$, the SDP returns $\max C_0^Q = 1$. For the range $\eta_0 > 1/4$, it is numerically shown that the maximum confidence is decreasing from 1 to 1/2 as the rate $\eta_0$ increases. The results are obtained by implementing an SDP numerically. Note that as

an approach to the LCP, one can find an analytic solution for two arbitrary states, one can find an analytic solution for two arbitrary states, which is presented in Sec. 4.6.

The maximum confidence above is interpreted as follows. If the outcome rate is low such that $\eta_0 \leq 1/4$, one cannot rule out the possibility that the first detector performs USD. When the outcome rate is more frequent, with $\eta_0 > 1/4$, it is clear that the detector cannot perform USD since the maximum confidence is strictly less than 1. As the outcome rate increases, the maximum confidence on the first arm becomes lower. The example shows a trade-off relation between the maximum confidence and the outcome rate.

## 4.6 Contextual advantages for certifiable maximum confidence

Let us now consider a realistic two-state discrimination scenario in which two states are prepared but three outcomes, including an additional one that collects undetected events, are provided. The certification of the maximum confidence in a detector is investigated and its contextual advantage is analysed.

### 4.6.1 Quantum state discrimination in practice

Here, we investigate the certifiable maximum confidence in a realistic two-state discrimination in detail. The framework developed in Sec. 4.5.2 is applied to certify the maximum confidence on a single detector. We recall that a pair of two pure states can always be identified by a single parameter $\theta$ such that

$$\cos\theta = \langle\psi_0|\psi_1\rangle = \sqrt{c_{0,1}}\ , \tag{4.66}$$

with the confusability $c_{0,1}$. This also means that any two-state discrimination problem can be mapped onto a two-dimensional plane spanned by the two states, i.e., $\mathrm{span}\{|\psi_0\rangle, |\psi_0^\perp\rangle\} = \mathrm{span}\{|\psi_1\rangle, |\psi_1^\perp\rangle\}$. Hence, without loss of generality, a two-state discrimination problem can be safely restricted to a qubit space. Let us write down the ensemble as

$$|\psi_0\rangle = \cos\frac{\theta}{2}|0\rangle + \sin\frac{\theta}{2}|1\rangle \quad \text{and} \quad |\psi_1\rangle = \cos\frac{\theta}{2}|0\rangle - \sin\frac{\theta}{2}|1\rangle\,, \tag{4.67}$$

which may be prepared with *a priori* probabilities $p_0$ and $p_1$, respectively.

Two detectors are arranged to find which of the states has been sent. A "click" in the first detector concludes that the state $|\psi_0\rangle$ was prepared and a detection event in the second one is for the state $|\psi_1\rangle$. The experiment is performed repeatedly so that one finds the rate of detection events in each arm. There are also cases where no detections are reported due to either the loss of prepared quantum systems during transmission or the failure of detectors to respond. After a measurement is repeated, outcome rates are found to be

$$\eta_{\text{obs}} = \{\eta_0, \eta_1, \eta_\varnothing\}, \tag{4.68}$$

where $\eta_\varnothing$ is the rate of undetected events.

For outcome rates compatible with quantum theory, there exist POVM elements $\{\hat{\pi}_b\}_{b=0}^{1,\varnothing}$ for the ensemble $\rho$ such that

$$\eta_b = \text{Tr}[\hat{\pi}_b \rho] \quad \text{where} \quad \rho = \sum_{x=0,1} p_x \left|\psi_x\right\rangle \left\langle\psi_x\right| \tag{4.69}$$

where the measurement fulfills the condition, $\hat{\pi}_0 + \hat{\pi}_1 + \hat{\pi}_\varnothing = \mathbb{1}$. In general, the figure of merit can be written for predetermined parameters $\alpha = \{\alpha_0, \alpha_1\}$:

$$\max \langle C^Q \rangle_\alpha = \max \left( \alpha_0 C_0^Q + \alpha_1 C_1^Q \right) , \tag{4.70}$$

where the maximization runs over POVM elements. The optimisation problem can be solved analytically with the optimality conditions in (4.62) and (4.63).

### 4.6.2 Maximum confidence on a quantum state

The maximum confidence in the realistic two-state discrimination scenario above can be certified as follows. For simplicity, let us assume the preparation of equiprobable states, i.e., $p_0 = p_1 = 1/2$ and show the certification for the first detector. The detailed derivation is shown in Sec. 4.10.

In the certification scenario, a detector in a two-state discrimination scenario shows an outcome rate $\eta_0$ when the measurement is repeated. Using our KKT conditions, it can be shown that the certifiable maximum confidence on such a measurement is given by

$$\max C_0^Q = \begin{cases} 1, & \text{for } \eta_0 \in [0, c_-] \\ 1/2 + f(\eta_0, c_{0,1}), & \text{for } \eta_0 \in [c_-, c_+] \\ 1/(2\eta_0), & \text{for } [c_+, 1] \end{cases} \tag{4.71}$$

where

$$c_{\pm} = \frac{1}{2}(1 \pm c_{0,1}) \tag{4.72}$$

$$f(\eta_0, c_{0,1}) = \frac{1}{4\eta_0}\sqrt{\left(\frac{1-c_{0,1}}{c_{0,1}}\right)\left(c_{0,1} - (1-2\eta_0)^2\right)}\ . \tag{4.73}$$

Note that certification depends upon the outcome rate $\eta_0$ of detected events only for a given ensemble of states.

An optimal measurement for maximum confidence state discrimination can be characterized according to the outcome rate. For an outcome rate $\eta_0 \leq c_+$, an optimal measurement is given by rank-one POVM elements. For $\eta_0 > c_+$ where the outcome rate is relatively higher, the maximum confidence is obtained from a rank-2 POVM element. One can find that that too frequent detection events, i.e., $\eta_0 \geq c_+$, rule out a rank-1 measurement for maximum confidence discrimination: thus, a rank-2 measurement is also certified.

### 4.6.3 Contextual advantage

We now investigate the certification of an MCM in a noncontextual theory and compare it with the quantum case. To this end, the main task is to optimise a measurement in a noncontextual ontological model, i.e., a response function $\xi_0(\lambda)$ in the first arm, given the extra constraint with a fixed outcome rate $\eta_0$. We write two epistemic states as $\mu_0(\lambda)$ and $\mu_1(\lambda)$, showing the confusability $c_{0,1}$ that is the same as that of quantum states defined in (4.67). $\xi_0(\lambda)$ identifies $\mu_0(\lambda)$ and, for the same reason as discussed above (4.67), is taken to be a sharp measurement of the state $\mu_b(\lambda)$. As elsewhere, the ensemble is prepared equiprobably, that is, with $p_0 = p_1 = 1/2$.

The fixed outcome rate must first be addressed. The outcome rate can be expressed in terms of the confusabilities as

$$\eta_0 = \frac{1}{2}c_{0,b} + \frac{1}{2}c_{1,b}. \tag{4.74}$$

where $b$ labels the sharp response function for the epistemic state $\mu_b(\lambda)$. We can see, following (4.55), that a sharp measurement will only be able to attain outcome rates in the range

$$c_- \leq \eta_0 \leq c_+. \tag{4.75}$$

For rates less than the lower bound, we must use a sharp measurement weighted by a probability. Such response functions were seen in (4.26). For

rates above this bound, a "rank-2" response function (i.e., one consisting of mixing multiple outcomes) is required. We note that these boundaries are exactly the same as those from the quantum case ( see (4.71) ). Each region of our piecewise confidence function will be addressed in what follows.

Let us begin with the infrequent detection region where $\eta_0 \leq c_-$. Here we must again use a response function of the form

$$\xi_0(\lambda) = \begin{cases} q & \text{if } \lambda \in \text{supp}[\mu_b(\lambda)] \\ 0 & \text{if } \lambda \in \text{supp}[\mu_b^{\perp}(\lambda)]. \end{cases} \tag{4.76}$$

With this function we can express the confidence as

$$C_1^{NC} = \frac{q}{2\eta_0} \int_\Lambda d\lambda \mu_1(\lambda)\xi_b(\lambda) = \frac{q}{2\eta_0} c_{0,b}. \tag{4.77}$$

The goal is to maximise the confusability over a constant outcome rate. To take the latter into account, we use the $\ell_1$ distance, which is shown in Ref. [131] as follows:

$$c_{x,b} = 1 - \frac{1}{2}||\mu_x - \mu_b||_1, \text{ where } ||\mu_x - \mu_b||_1 = \int_\Lambda d\lambda |\mu_x(\lambda) - \mu_b(\lambda)|. \tag{4.78}$$

We now express $\eta_0$ in terms of the $\ell_1$ distance:

$$\eta_0 = \frac{q}{2}\left(c_{b,0} + c_{b,1}\right) = \frac{q}{2}\left(2 - \frac{1}{2}\|\mu_b - \mu_0\|_1 - \frac{1}{2}\|\mu_b - \mu_1\|_1\right). \tag{4.79}$$

The triangle inequality allows us to exploit the relation,

$$\|\mu_b - \mu_1\|_1 \leq \|\mu_b - \mu_0\|_1 + \|\mu_0 - \mu_1\|_1. \tag{4.80}$$

Combining this result with (4.78) above and writing in terms of $c_{0,1}$, we obtain

$$\|\mu_b - \mu_0\|_1 \geq 1 - 2\frac{\eta_0}{q} + c_{0,1}, \tag{4.81}$$

or, in a more convenient form using the confusabilities,

$$c_{0,b} \leq \frac{\eta_0}{q} + c_-, \tag{4.82}$$

as the upper bound on the confusability $c_{1,b}$ consistent with a fixed outcome rate and weighted by the probabilistic parameter $q$, under the assumption

that a "rank-1" operator is able to recreate the desired $\eta_0$. Bringing all of these together, the maximum confidence can be expressed as

$$\max C_0^{NC} = \max_q \left( \frac{1}{2} + \frac{c_-}{2\eta_0} q \right), \tag{4.83}$$

where the maximisation runs over the variable $q \in [0, 1]$.

Let us now find the certified maximum confidence given an outcome rate $\eta_0$. For the range of the outcome rate where $\eta_0 < c_-$, the optimal parameter can be chosen as $q = 2\eta_0/c_-$. Thus, the certifiable maximum confidence is given as $C_0^{NC} = 1$. These cases can be interpreted as USD, except that the confidence of the other arm of the detector is not yet specified. Therefore, a distinction between the quantum and noncontextual theories is not found in terms of the maximum confidence of a given state. Of course, as it has been shown above, there is a distinction in terms of a different figure of merit, the rate of ambiguous outcomes.

The next range to consider is when the outcome rate is within the bounds, $\eta_0 \in [c_-, c_+]$ where we recall $c_\pm = (1 \pm c_{0,1})/2$. Here, sharp measurements give the desired outcome rate and, therefore, are treated simply by letting $q = 1$ in the above calculation. This gives a maximum confidence,

$$\max C_1^{NC} = \frac{1}{2} \left( 1 + \frac{1 - c_{0,1}}{2\eta_0} \right) < \max C_0^Q, \tag{4.84}$$

with the maximum confidence in quantum theory in (4.71). Thus, a quantum advantage is shown in the range (see Fig. 4.7).

For the high-outcome-rate range where $\eta_0 \geq c_+$, we deduce the response function by considering the behaviour at two values of $\eta_0$. The confidence must be continuous at the point $\eta_0 = c_+$ and the response function at this point is a sharp measurement of $\mu_0(\lambda)$. The response function for higher values of $\eta_0$ must consist of binning together multiple measurement outcomes due to the bounds on sharp measurements. At $\eta_0 = 1$, the response function will be equal to one across the whole ontic state space, which can be decomposed into a sum of two non-overlapping sharp measurements. We can see that the response function will take the form

$$\xi_0(\lambda) = \xi_{\mu_0}(\lambda) + a\xi_{\mu_0^\perp}(\lambda) \ , \tag{4.85}$$

where $a$ is some constant that can be determined by evaluating the associated

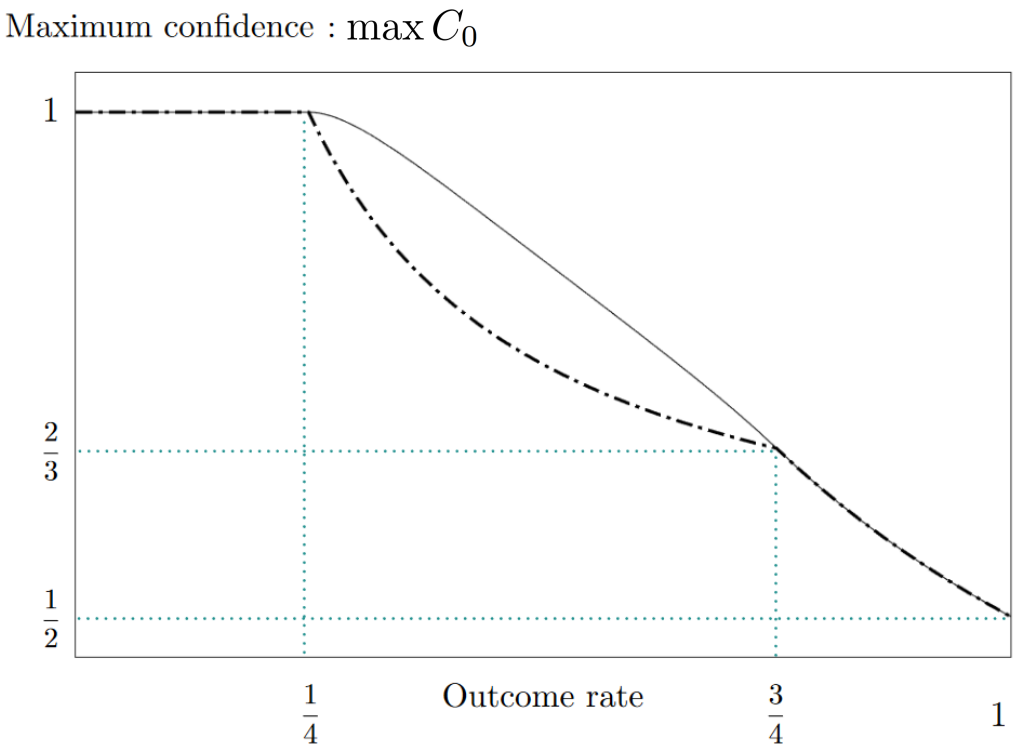



Figure 4.7: Figure extracted from Ref. [25]. In two-state discrimination between $|0\rangle$ and $(|0\rangle + |1\rangle)/\sqrt{2}$, the certifiable MCM in the first detector, denoted by $\max C_0$, is plotted with respect to outcome rate $\eta_0$. The certifiable MCM is shown in both quantum theory (solid line) and in a noncontextual ontological model (dotted line). A low detection rate $\eta_0 \leq 1/4$ is compatible with USD. The contextual advantages exist whenever an outcome rate is within the range $\eta_0 \leq 3/4$. However, no contextual advantage can be obtained if an outcome rate is too high for $\eta_0 > 3/4$.

outcome rate. Doing this gives

$$\xi_0(\lambda) = \xi_{\mu_0}(\lambda) + \frac{\eta_0 - c_+}{1 - c_+}\xi_{\mu_0^\perp}(\lambda). \tag{4.86}$$

This function gives a outcome rate $\eta_0$ and a maximum confidence

$$\max C_0^{NC} = \frac{1}{2\eta_0} \tag{4.87}$$

which is again the same as the behaviour in the quantum case.

Let us summarise the key features of the response function, which is optimized according to the outcome rate $\eta_0 \in [0, 1]$. In the low outcome region with $\eta_0 \leq c_-$, the optimal response function has the same support as the state $\mu_1(\lambda)$, on which it linearly increases from zero to one as the outcome rate goes from zero to $c_-$. In the central region with $\eta_0 \in [c_-, c_+]$, the optimal response function corresponds to a projective measurement which slightly shifts its support away from $\text{supp}[\mu_1(\lambda)]$ and towards $\text{supp}[\mu_0(\lambda)]$, to which it coincides when $\eta_0 = c_+$. Finally, when the outcome rate is even higher for $\eta_0 \geq c_+$, the support includes the rest of the ontic state space. The response function increases linearly on the region of two supports $\text{supp}[\xi_0(\lambda)]$ and $\text{supp}[\mu_0^\perp(\lambda)]$. When $\eta_0 = 1$, the response function will be equal to one for all ontic states.

Interestingly, the three ranges showing distinct forms of the response functions in a noncontextual model and an optimal measurement in quantum theory precisely coincide with each other. Contextual advantages in terms of a higher maximum confidence are shown in the central region only, where a sharp measurement turns out to be optimal (see also Fig. 4.7). In the next section, noisy preparations are considered where the aforementioned properties do not hold in general. Contextual advantages in terms of a higher maximum confidence appear over the whole range of outcome rates. The ranges giving distinct forms of a measurement in quantum and noncontextual theories no longer coincide with each other.

## 4.7 Certifiable maximum confidence on noisy preparation

We consider a noisy preparation and investigate contextual advantages in the certification of an MCM. We first recall the result in Sec. 4.4.4 that the

contextual advantages for the MCM hold true for noisy quantum states. We here extend the contextual advantage to the certification scenario. Again, let us consider a pair of mixed states given with equal *a priori* probabilities

$$\rho_0 = r\left|\psi_0\right\rangle\left\langle\psi_0\right| + (1-r)\frac{\mathbb{1}}{2}, \quad \rho_1 = r\left|\psi_1\right\rangle\left\langle\psi_1\right| + (1-r)\frac{\mathbb{1}}{2}. \tag{4.88}$$

We also exploit the confusability for the pure states, $c_{0,1} = |\left\langle\psi_0|\psi_1\right\rangle|^2$. In what follows, we compute the certified maximum confidence when the outcome rate is given by $\eta_0$ in the first arm.

### 4.7.1 Quantum states

We apply the same method used in Sec. 4.5 to compute the certifiable maximum confidence. The detailed derivation is shown in in Sec. 4.10. It is fairly straightforward to obtain the results. Contrary to the noiseless case in Sec. 4.5, it is found that the ranges in which different kinds of measurements are optimal do not coincide between quantum and noncontextual theories. The certifiable maximum confidence can be summarised depending on the range of the outcome rate.

First, when the outcome rate is in the range $\eta_0 \in [0, \eta_0^{(-)}]$ where

$$\eta_0^{(\pm)} = \frac{1}{2}\left(1 \pm r^2 c_{0,1}\right), \tag{4.89}$$

the confidence is given by,

$$\max C_0^Q = \frac{1}{2}\left(1 + \frac{r\sqrt{1-c_{0,1}}}{\sqrt{1-r^2 c_{0,1}}}\right). \tag{4.90}$$

Note that the noiseless case $r = 1$ reproduces USD and also the boundary condition in the range $\eta_0^{(-)} = c_-$ in (4.71). For noisy cases with $r < 1$, it holds that $\eta_0^{(-)} > c_-$.

Second, when $\eta_0 \in [\eta_0^{(-)}, \eta_0^{(+)}]$ the certifiable maximum confidence is computed as

$$\max C_0^Q = \frac{1}{2} + g_p(\eta_0, c_{0,1}) \ , \tag{4.91}$$

where

$$g_p(\eta_0, c_{0,1}) = \frac{1}{4\eta_0}\sqrt{\left(\frac{1-c_{0,1}}{c_{0,1}}\right)(r^2 c_{0,1} - (1-2\eta_0)^2)}. \tag{4.92}$$

Note that the case $r = 1$ reproduces the certifiable maximum confidence in a noiseless case in (4.71).

Third, when the outcome rate is in the range $\eta_0 \in [\eta_0^{(+)}, 1]$, the certifiable maximum confidence is obtained as

$$\max C_0^Q = \frac{1}{2}\left(1 + \frac{r\sqrt{1-c_{0,1}}}{\sqrt{1-r^2 c_{0,1}}}\left(\frac{1}{\eta_0} - 1\right)\right). \tag{4.93}$$

Note that it holds that $\eta_0^{(+)} < c_+$ for noisy cases with $r < 1$ (see (4.71)). In addition, the noiseless case $r = 1$ also reproduces (4.71).

### 4.7.2 Noncontextual model

Similarly to what is shown in Sec. 4.4.4, we consider noisy states $\tilde{\mu}_0(\lambda)$ and $\tilde{\mu}_1(\lambda)$ as defined in (4.41) with *a priori* probabilities 1/2, respectively. In the certification scenario, it is assumed that the outcome rate in the first arm is given by $\eta_0$. We then aim to find the certifiable maximum confidence on, say, the first arm.

Sharp measurements cannot reproduce all outcome rates, as shown in (4.55). The outcome rate $\eta_0$ can be obtained using

$$\eta_0 = \int_\Lambda d\lambda \mu_P(\lambda)\xi_0(\lambda) \tag{4.94}$$

where $\mu_P(\lambda)$ denotes the ensemble in (4.42). Applying (4.55) to the integral above, one can obtain bounds on the outcome rate as follows:

$$\frac{1}{2}\left(1 - rc_{0,1})\right) \leq \eta_0 \leq \frac{1}{2}\left(1 + rc_{0,1}\right). \tag{4.95}$$

Note that a sharp measurement can produce the desired statistics in the range above. If we require an outcome rate is below the lower bound, we must use weighted sharp measurements. If the desired outcome rate is higher than the upper bound, we must use rank-2 equivalent measurements. Interestingly, the boundaries in a noncontextual theory are different from those in the quantum case in Sec. 4.6 (see also Fig. 4.8).

Once the outcome rate is in the range $\eta_0 \in [0, (1 - rc_{0,1})/2]$, the measurement for the maximum confidence must be a weighted sharp measurement, i.e., we again let $\xi_0(\lambda) = q\xi_b(\lambda)$ where $0 \leq q \leq 1$ and $\xi_b(\lambda)$ is a sharp measurement

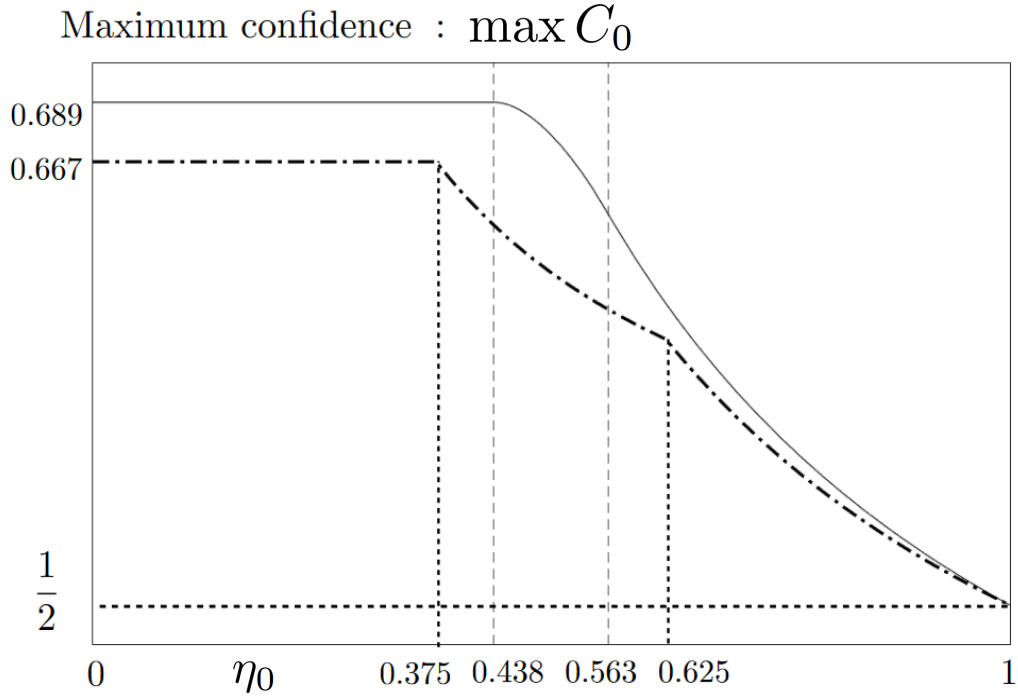



Figure 4.8: Figure extracted from Ref. [25]. The certifiable maximum confidence is shown for two noisy states $c_{0,1} = 1/2$ and $r = 1/2$ where $r$ is the noise parameter in (4.88). The certifiable maximum confidence varies depending on a outcome rate $\eta_0$. The certifiable maximum confidence for quantum states (solid) is higher than that in a noncontextual model (dotted) for all $\eta_0 \in [0, 1]$.

for a to-be-determined epistemic state.

For this response function, it holds that

$$\eta_1 = \frac{q}{2}\left((1-r) + r(c_{0,b} + c_{1,b})\right) \tag{4.96}$$

which can be written as,

$$q = \frac{2\eta_0}{(1-r) + r(c_{0,b} + c_{1,b})}. \tag{4.97}$$

Thus, the value $q$ is obtained from a given $\eta_0$. Let us express the confidence in terms of the confusabilities,

$$\begin{aligned} C_0^{NC} &= \frac{1}{2\eta_0}\int_\Lambda d\lambda \tilde{\mu}_0(\lambda)\xi_0(\lambda) = \frac{q}{2\eta_0}\int_\Lambda d\lambda \tilde{\mu}_0(\lambda)\xi_b(\lambda) \\ &= 1 - \frac{(1-r) + 2rc_{1,b}}{(1-r) + r(c_{0,b} + c_{1,b})}. \end{aligned} \tag{4.98}$$

To find the maximum confidence, one has to minimise the fraction by finding $b$ such that the numerator is minimal and the denominator is maximal. It turns out that the optimal choice is given by $b = \bar{1}$. It is obvious that $c_{1,b}$ is minimized with $b = \bar{1}$. From (4.55), the sum $c_{0,b} + c_{1,b}$ is minimal as $1 - c_{0,1}$. Therefore, we have

$$\max C_0^{NC} = 1 - \frac{1-r}{2\left(1 - rc_{0,1}\right)} \tag{4.99}$$

which also shows that the noiseless case $r = 1$ reproduces the case USD.

When the outcome rate is in the range in (4.95), the measurement must be sharp and we again use $\xi_0(\lambda) = \xi_b(\lambda)$ to avoid confusion between response functions. We apply the same technique used in Sec. 4.6.3. The key tool is the inequality,

$$1 - c_{b,1} \leq 2 - c_{0,b} - c_{0,1}, \tag{4.100}$$

Note also that

$$\eta_0 = \frac{1-r}{2} + \frac{r}{2}\left(c_{0,b} + c_{1,b}\right), \tag{4.101}$$

from which,

$$c_{1,b} = \frac{2\eta_0 - (1-r)}{r} - c_{0,b}. \tag{4.102}$$

All these imply that

$$c_{0,b} \leq \frac{1}{2}\left(1 + \frac{2\eta_0 - (1-r)}{r} - c_{0,1}\right). \tag{4.103}$$

The confidence is given by

$$C_0^{NC} = \frac{1}{2\eta_0}\left(\frac{1-r}{2} + rc_{0,x}\right), \tag{4.104}$$

which has the maximum as follows,

$$\max C_0^{NC} = \frac{1}{2} + \frac{r(1-c_{0,1})}{4\eta_0}. \tag{4.105}$$

This agrees with (4.84) when $r = 1$.

Again, in the range $\eta_0 \in [(1 + rc_{0,1})/2, 1]$ when the outcome rate is high, the response function can be directly deduced. The response function will take the form

$$\xi_0(\lambda) = \xi_{\mu_0}(\lambda) + a\xi_{\mu_0^\perp}(\lambda) \tag{4.106}$$

as in (4.85) and for the same reasons, where $\xi_{\mu_0}(\lambda)$ is the sharp measurement associated with $\mu_0(\lambda)$. Note that the value $a$ is fixed by the outcome rate and can be found by calculating the $\eta_0$ given by the response function,

$$a = \frac{2\eta_0 - 1 - rc_{0,1}}{1 - rc_{0,1}}. \tag{4.107}$$

The confidence is therefore obtained as

$$\max C_0^{NC} = \frac{1}{2\eta_0}\left(1 - \frac{(1-r)(1-\eta_0)}{1 - rc_{0,1}}\right). \tag{4.108}$$

This agrees with (4.86) for cases $r = 0, 1$.

### 4.7.3 Comparison

A comparison of the noiseless and noisy cases in Sec. 4.6 and Sec. 4.7 shows that the maximum confidence can be characterised depending upon how frequent an outcome rate is, that is, within the ranges of low, intermediate and high outcome rates. The feature commonly shared between them is that the maximum confidence does not increase as the outcome rate gets more frequent: a less frequent outcome rate implies a higher the maximum

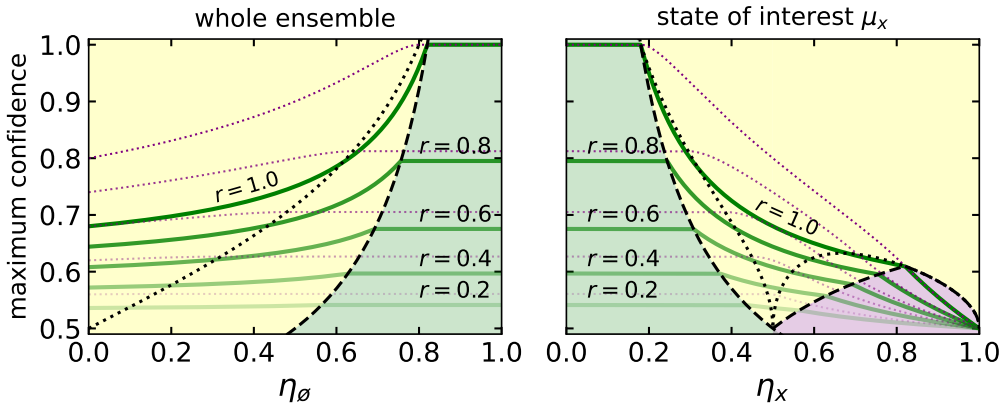



Figure 4.9: Maximum confidence in a qubit state discrimination scenario, according to quantum (purple dotted lines) and noncontextual (green straight lines) models. We directly compare with Fig. 3.7, where only the quantum case is considered, with $\sqrt{c_{0,1}} = 0.8$. Black lines correspond to the bounds on the outcome rates at which behaviours change in quantum (dotted lines) and noncontextual (dashed lines) models. On the left-hand side: the yellow area is delimited by $0 \leq \eta_\varnothing \leq (1+rc_{0,1})/2$ and the green area by $(1+rc_{0,1})/2 \leq \eta_\varnothing \leq 1$. On the right-hand side: the green area is limited by $0 \leq \eta_x \leq (1-rc_{0,1})/2$, the yellow area by $(1-rc_{0,1})/2 \leq \eta_x \leq (1+rc_{0,1})/2$ and the purple area by $(1+rc_{0,1})/2 \leq \eta_x \leq 1$.

confidence and vice versa (see Fig. 4.9.).

Contrasting the cases, it is shown that the ranges characterising the maximum confidence coincide in quantum and noncontextual theories when the preparation is noiseless. Contextual advantages are shown in the intermediate range only. In the noisy case, the ranges are distinct in quantum and noncontextual theories, where the intermediate range becomes narrower. Contextual advantages in this scenario appear in the whole range of outcomes rates.

It is observed that the contextual advantages appearing in the low and high outcome rates are related with each other. Let us consider the range of lower detection rate in a noisy case,

$$\eta_0 < \frac{1}{2}\left(1 - rc_{0,1})\right). \tag{4.109}$$

In the low-outcome-rate region, the confidence is constant. We can evaluate the difference between the two functions and denote it as follows:

$$\Delta_L := \left(\max C_0^Q - \max C_0^{NC}\right)\Big|_{\eta_0<(1-rc_{0,1})/2} . \tag{4.110}$$

One can find that the gap is strictly positive if $r < 1$ and zero for $r = 1$. Then, for a higher outcome rate where

$$\eta_0 > \frac{1}{2}\left(1 + rc_{0,1})\right) \tag{4.111}$$

it turns out that the gap between quantum and noncontextual theories can be written as,

$$\Delta_H := \left(\max C_0^Q - \max C_0^{NC}\right)\Big|_{\eta_0>(1+rc_{0,1})/2} = \left(\frac{1}{\eta_0} - 1\right)\Delta_L. \tag{4.112}$$

which is also strictly positive for $r < 1$. If no contextual advantage appears in the low-outcome-rate range, i.e., $\Delta_L = 0$, then neither does it when the outcome rate is high, i.e., $\Delta_H = 0$.

## 4.8 Conclusion

State discrimination is a fundamental tool in information processing in general [97]. Its central role in quantum information applications requires us to specify exactly when quantum theory provides an advantage compared to classical theories. When classicality is regarded as noncontextuality, the contextual

advantages for quantum state discrimination are characterized in the scenario of two-state MESD, for which noisy measurements are also considered, so that the observation of the advantages is experimentally feasible [27].

The main contribution of the present work is twofold. On one hand, we show that contextual advantages generally exist in quantum state discrimination: the advantages exist not only in MESD but also in the cases of USD and, more generally, maximum confidence state discrimination, which presents a unifying framework for state discrimination including USD and MESD. Our results hence show that quantum information applications based on state discrimination in general can be leveraged to attain quantum advantages. We also develop an optimization technique for investigating MCM. An MCM for multiple quantum states in any dimension can hence be computed (see also an analytic derivation for qubit states [83, 105]).

Furthermore, we examine and compare maximum confidence measurements in noncontextual and quantum theories. In a noncontextual theory, it turns out that the optimal measurement for an ensemble is unchanged even if the preparation is noisy. In quantum theory, however, the MCM is constructed depending upon how much noise is present in given states. Consequently, an MCM for noisy states in quantum theory shows a higher maximum confidence over a noncontextual theory. We remark that our examination of the optimal measurements for noisy states is distinct from the results of noncontextual inequalities presented in Ref. [27]. While the noncontextual bounds are typically found at a higher level, we analyze in detail the optimal measurements as well as their structure. Our techniques will be applicable to future research in the field and give us a closer look at the structure of measurements on the ontic state space.

On the other hand, our findings that contextual advantages for state discrimination exist in general, precisely, for MCMs, leverage an semi-DI scenario of certifying a quantum measurement. We develop a framework for certifying the maximum confidence on a quantum measurement in a realistic semi-DI manner, where the outcome rates for an ensemble of states are provided by an experiment while the measurement is neither fully characterized nor trusted. Note also that the undetected events that may be present in an experiment are naturally taken into account. In fact, undetected events appear so frequently in photonic quantum experiments that this directly applies to various quantum information applications: quantum key distribution, Bell nonlocality detection, photonic quantum computation, etc. It should also be pointed out that our consideration in the framework of an

semi-DI scenario goes beyond noisy measurements (e.g., Ref. [27]), which shows the resilience of contextual advantages to noise. The semi-DI scenario considered here deals with untrusted or uncharacterized measurement devices and also undetected events, which are not allowed in MESD in particular.

We then formulate an optimization problem for the certification of the maximum confidence on a quantum measurement in an semi-DI scenario. To demonstrate contextual advantages in an semi-DI scenario, we develop the certification of the maximum confidence in a noncontextual ontological model and also in a quantum theory and then compare them. It is seen that one can always find contextual advantages in the maximum confidence on quantum states in an semi-DI scenario with untrusted detectors. Our results show that quantum state discrimination in the most realistic scenario with uncharacterized and lossy detectors can achieve its advantage over a noncontextual ontological model.

Finally, it is worth mentioning that there is still much room to generalize further by considering a wider class of ensembles. In particular, three-state discrimination poses an interesting problem due to the impossibility of creating the symmetric three-state ensemble in a noncontextual theory [20]. Contextual advantages associated with discrimination of general mixed states are an intriguing extension. As such states can be decomposed in the Hilbert-space formalism in a number of different possible ways, their representation in a noncontextual operational ontological model should have rich consequences. The exploration of such areas will further our understanding of the quantum-classical boundaries.

Our results set the ground for understanding how quantum information applications that exploit quantum state discrimination can achieve advantages over a classical theory in a realistic scenario. Our results are not limited to advances in quantum information theory from the fundamental point of view but, more importantly, can be exploited to show quantum advantages in existing practical quantum information applications. Among the tasks using state discrimination, it would be interesting to investigate randomness generation (see, e.g., Ref. [17, 68]). It would also be interesting to investigate contextual advantages in quantum computing tasks, such as quantum machine learning, where state discrimination is often processed to manipulate classical data over the limitations of conventional computing [132].

## 4.9 Appendix A: Derivation of the optimality condition in the certification scenario

We here derive the optimality conditions in (4.62) and (4.63), which allow for the certification of an MCM given specified outcome statistics. That is, given outcome rates $\eta_x$, the goal is to maximise $\sum_x \alpha_x C_x$ over a measurement, where

$$C_x = \frac{p_x}{\eta_x} \operatorname{Tr}[\hat{\pi}_x \rho_x]. \tag{4.113}$$

In fact, the optimisation problem can be written as an SDP. The primal problem is the following:

$$\begin{aligned} p^* = \text{maximize } & \sum_{x=1}^{n} \alpha_x \frac{p_x}{\eta_x} \operatorname{Tr}[\hat{\pi}_x \rho_x] \\ \text{subject to } & \hat{\pi}_b \geq 0, \ \sum_{b=0}^{n-1} \hat{\pi}_b \leq \mathbb{1}, \\ & \operatorname{Tr}[\rho \hat{\pi}_x] = \eta_x \end{aligned} \tag{4.114}$$

Let $\hat{\pi}_\emptyset = \mathbb{1} - \sum_{b=0}^{n-1} \hat{\pi}_b \geq 0$ denote a slack variable that takes undetected events into account. Let us introduce dual variables $r_b \sigma_b$ for inequality constraint where $r_b \geq 0$ and $\sigma_b$ is a quantum state, and $K$ and $s_x$ to derive the Lagrangian functional in the following,

$$\begin{aligned} & \mathcal{L}(\{\hat{\pi}_b\}_{b=0}^{n-1,\emptyset}, \{r_b\}_{b=0}^{n-1,\emptyset}, \{\sigma_b\}_{b=0}^{n-1,\emptyset}, \{s_x\}_{x=0}^{n-1}, K) \\ &= \sum_{x=0}^{n-1} \alpha_x \frac{q_x}{\eta_x} \operatorname{Tr}[\hat{\pi}_x \rho_x] + \sum_{b=0}^{n-1,\emptyset} r_b \operatorname{Tr}[\hat{\pi}_b \sigma_b] + \operatorname{Tr}[K(I - \sum_{b=0}^{n-1,\emptyset} \hat{\pi}_b)] + \sum_{x=0}^{n-1} s_x(\eta_x - \operatorname{Tr}[\rho \hat{\pi}_x]) \\ &= \sum_{b=0}^{n-1} \operatorname{Tr}[\hat{\pi}_x(\alpha_x \frac{\rho_x}{\eta_x} + r_x \sigma_x - K - s_x \rho)] + \operatorname{Tr}[\hat{\pi}_\emptyset(r_\emptyset \sigma_\emptyset - K)] + \operatorname{Tr}[K] + \sum_{x=0}^{n-1} s_x \eta_x. \end{aligned} \tag{4.115}$$

The dual functional is derived as follows,

$$\begin{aligned} & g(\{r_b\}_{b=0}^{n-1,\emptyset}, \{\sigma_b\}_{b=0}^{n-1,\emptyset}, \{s_x\}_{x=0}^{n-1}, K) \\ &= \sup_{\{\hat{\pi}_b\}_{b=0}^{n-1,\emptyset}} \mathcal{L}(\{\hat{\pi}_b\}_{b=0}^{n-1,\emptyset}, \{r_b\}_{b=0}^{n-1,\emptyset}, \{\sigma_b\}_{b=0}^{n-1,\emptyset}, \{s_x\}_{x=0}^{n-1}, K) \\ &= \begin{cases} \operatorname{Tr}[K] + \sum_{b=0}^{n-1} s_x \eta_x & \text{if } \frac{\alpha_x}{\eta_x} \rho_x + r_x \sigma_x - K - s_x \rho = 0 \text{ and } r_\emptyset \sigma_\emptyset - K = 0, \\ & \text{for } x = 0, 1, \cdots, n-1 \\ +\infty & \text{otherwise.} \end{cases} . \end{aligned} \tag{4.116}$$

Since the dual functional does not diverge, we have that

$$K = r_{\emptyset}\sigma_{\emptyset}, \text{ and } K = \alpha_x \frac{\rho_x}{\eta_x} + r_x\sigma_x - s_x\rho, \quad x = 0, 1, 2, \cdots n-1. \tag{4.117}$$

This condition is called the *Lagrangian stability*. The dual problem can be written as,

$$\begin{aligned} d^* = \text{minimize} \quad & \mathrm{Tr}[K] + \sum_{b=0}^{n-1} s_x\eta_x \\ \text{subject to} \quad & K + s_x\rho \geq \frac{\alpha_x}{\eta_x}\rho_x, \text{ and} \\ & K \geq 0. \end{aligned} \tag{4.118}$$

In general, it holds that $p^* \geq d^*$. The equality holds when the problem is strictly feasible. For instance, one can choose $\hat{\pi}_b = \eta_b \mathbb{1}$ for all $b$ to show that the primal problem is strictly feasible. We thus have that $p^* = d^*$.

When the dual and primal problems give the same solution, one can also solve the optimisation problem by analyzing the optimality conditions directly. For the SDP above, the optimality conditions are listed as,

$$\text{Lagrangian stability: } K = \alpha_x \frac{q_x}{\eta_x}\rho_x + r_x\sigma_x - s_x\rho, \; K = r_{\emptyset}\sigma_{\emptyset} \; \forall x \tag{4.119}$$

$$\text{Complementary slackness:} \quad r_b \,\mathrm{Tr}[\hat{\pi}_b\sigma_b] = 0, \quad \forall b \;, \tag{4.120}$$

together with the constraints in the primal and dual problems. Although the optimality conditions contain a greater number of variables than the primal and dual problems, they are useful for exploiting the generic structure existing in an optimisation problem.

## 4.10 Appendix B: Solving the optimality conditions for certifying the maximum confidence

We here show the approach of the so-called linear complementarity problem in the certification of a maximum confidence. We consider qubit states and show how the optimality conditions can be directly analysed.

Suppose that two states $\rho_1$ and $\rho_1$ are given with prior probability 1/2, respectively,

$$\rho_0 = r\left|\psi_0\right\rangle\left\langle\psi_0\right| + (1-r)\frac{\mathbb{1}}{2}, \quad \text{and} \quad \rho_1 = r\left|\psi_1\right\rangle\left\langle\psi_1\right| + (1-r)\frac{\mathbb{1}}{2} \tag{4.121}$$

for which the outcome rates given by $\eta_0$ and $\eta_1$. The goal is now to find the certifiable maximum confidence on the first arm. Let us begin with the following primal problem:

$$\begin{aligned} p^* = \text{maximize} \quad & \frac{1}{2\eta_0}\operatorname{Tr}[\hat{\pi}_0\rho_0] \\ \text{subject to} \quad & 0 \leq \hat{\pi}_0 \leq \mathbb{1}, \\ & \operatorname{Tr}[\hat{\pi}_0\rho] = \eta_0. \end{aligned} \tag{4.122}$$

The Lagrangian function can be constructed as

$$\begin{aligned} &\mathcal{L}(\hat{\pi}_0, X_0, X_1, \lambda) \\ &= \frac{1}{2\eta_0}\operatorname{Tr}[\rho_0\hat{\pi}_0] + \operatorname{Tr}[X_0\hat{\pi}_0] + \operatorname{Tr}[(I-\hat{\pi}_0)X_1] + \lambda(\eta_0 - \operatorname{Tr}[\rho\hat{\pi}_0]) \\ &= \lambda\eta_0 + \operatorname{Tr}[X_1] + \operatorname{Tr}[(\frac{\rho_0}{2\eta_0} + X_0 - X_1 - \lambda\rho)\hat{\pi}_0] \end{aligned} \tag{4.123}$$

from which the dual problem can be obtained:

$$\begin{aligned} d^* = \text{minimize} \quad & \lambda\eta_0 + \operatorname{Tr}[X_1] \\ \text{subject to} \quad & X_1 + \lambda\rho \geq \frac{1}{2\eta_0}\rho_0, \\ & X_1 \geq 0 \end{aligned} \tag{4.124}$$

The optimality conditions can be found and listed out as follows,

$$\begin{aligned} X_0 - X_1 &= \lambda\rho - \frac{1}{2\eta_0}\rho_0 \\ X_0, X_1 &\geq 0 \\ \hat{\pi}_0 X_0 &= 0 \\ (I-\hat{\pi}_0)X_1 &= 0 \\ 0 \leq \hat{\pi}_0 &\leq I \\ \operatorname{Tr}[\rho\hat{\pi}_0] &= \eta_0. \end{aligned} \tag{4.125}$$

Since qubit measurements are considered, $X_0X_1 = 0$ holds. Since the non-negative operators $X_0$ and $X_1$ are orthogonal, they can be obtained from the

spectral decomposition in (4.125). Let $\nu_\pm$ denote the positive and negative eigenvalues $|\nu_\pm\rangle$, respectively, so that

$$X_0 - X_1 = \lambda\rho - \frac{1}{2\eta_0}\rho_0 = \nu_+ |\nu_+\rangle \langle\nu_+| + \nu_- |\nu_-\rangle \langle\nu_-| \tag{4.126}$$

where

$$\nu_\pm = \frac{\tan\theta}{4\eta_0}(\gamma \pm \sqrt{1+\gamma^2} r\cos\theta) \tag{4.127}$$

$$|\nu_\pm\rangle = \frac{1}{\sqrt{2+2\gamma^2 \mp 2\gamma\sqrt{1+\gamma^2}}}(|0\rangle + (\gamma \mp \sqrt{1+\gamma^2})|1\rangle) \tag{4.128}$$

with $\gamma = (2\eta_0\lambda - 1)\cot\theta$. It is straightforward to find the maximum confidence,

$$\begin{aligned}\max C_0^{\mathcal{Q}} &= \lambda\eta_0 + \mathrm{Tr}[X_1] = \frac{1}{2}(1 + \gamma\tan\theta) - \nu_- \\ &= \frac{1}{2} + \frac{\tan\theta}{4\eta_0}[(2\eta_0 - 1)\gamma + r\cos\theta\sqrt{1+\gamma^2}],\end{aligned} \tag{4.129}$$

where the parameter $\gamma$, relying on the dual parameter $\lambda$, needs to be further optimised. If either $X_0$ or $X_1$ is of full-rank, then the optimisation becomes trivial since $\hat{\pi}_0 = 0$ or $\mathbb{1}$. Assuming $X_0$ and $X_1$ are not full-rank, there are three possible cases for $X_0X_1 = 0$.

Firstly, we consider that $X_0 = 0$ and $X_1 > 0$. Since $X_0 = 0$, we have that $\nu_+ = 0$,

$$\gamma = \frac{-r\cos\theta}{\sqrt{1 - r^2\cos^2\theta}} \tag{4.130}$$

and

$$\begin{aligned}\max C_0^{\mathcal{Q}} &= \frac{1}{2}\left[1 + \frac{r\sin\theta}{\sqrt{1-r^2\cos^2\theta}}\left(\frac{1}{\eta_0} - 1\right)\right], \\ \text{where} \quad |\nu_-\rangle &= \frac{1}{\sqrt{2}}(\sqrt{1 + r\cos\theta}\,|0\rangle + \sqrt{1 - r\cos\theta}\,|1\rangle).\end{aligned} \tag{4.131}$$

Since $X_1 = -\nu_- |\nu_-\rangle \langle\nu_-|$ and $X_1(I - \hat{\pi}_0) = 0$,

$$\hat{\pi}_0 = \quad |\nu_-\rangle \langle\nu_-| + \alpha |\nu_+\rangle \langle\nu_+| = \alpha I + (1-\alpha)|\nu_-\rangle \langle\nu_-| \tag{4.132}$$

for some constant $0 \leq \alpha \leq 1$. That is, the optimal measurement is a convex combination of $I$ and $|\nu_-\rangle \langle\nu_-|$. To find $\alpha$, we use condition $\mathrm{Tr}[\hat{\pi}_0\rho] = \eta_0$ so that

$$\alpha = \frac{2\eta_0 - 1 - r^2\cos^2\theta}{1 - r^2\cos^2\theta}. \tag{4.133}$$

Since $\alpha \geq 0$, the outcome rate is constrained by $\eta_0 \geq \frac{1}{2}(1 + r^2 \cos^2\theta)$.

Secondly, we consider that $X_1 = 0$ and $X_0 > 0$. For $X_1 = 0$, for which it holds that $\nu_- = 0$.

$$\gamma = \frac{r\cos\theta}{\sqrt{1 - r^2\cos^2\theta}}. \tag{4.134}$$

Since $X_0$ is rank-one, the optimal measurement must be rank-one $\hat{\pi}_0 = \beta \left|\nu_-\right\rangle \left\langle\nu_-\right|$. It is straightforward to find the maximum confidence,

$$\begin{aligned} \max C^Q(1) &= \frac{1}{2}(1 + \frac{r\sin\theta}{\sqrt{1 - r^2\cos^2\theta}}) \\ \text{where} \quad \left|\nu_-\right\rangle &= \frac{1}{\sqrt{2}}(\sqrt{1 - r\cos\theta}\left|0\right\rangle + \sqrt{1 + r\cos\theta}\left|1\right\rangle). \end{aligned} \tag{4.135}$$

The optimal measurement is given by

$$\hat{\pi}_0 = \frac{2\eta_0}{1 - r^2\cos^2\theta}\left|\nu_-\right\rangle\left\langle\nu_-\right|. \tag{4.136}$$

The condition $\beta \leq 1$ is equivalent to $\eta_0 \leq \frac{1}{2}(1 - r^2\cos^2\theta)$.

Thirdly, $X_0 > 0$ and $X_1 > 0$. Since $X_0$ and $X_1$ are both rank-one, optimal POVM elements $\hat{\pi}_0$ and $\mathbb{1} - \hat{\pi}_0$ are also rank-one so that $\hat{\pi}_0 = \left|\nu_-\right\rangle\left\langle\nu_-\right|$. From the condition $\mathrm{Tr}[\hat{\pi}_0\rho] = \eta_0$, we find

$$\gamma = \frac{1 - 2\eta_0}{\sqrt{r^2\cos^2\theta - (1 - 2\eta_0)^2}}.$$

We then have,

$$\begin{aligned} \max C_0^Q &= \frac{1}{2} + \frac{\tan\theta}{4\eta_0}\sqrt{r^2\cos^2\theta - (1 - 2\eta_0)^2} \\ \text{where} \quad \left|\nu_-\right\rangle &= \frac{1}{\sqrt{2}}\left(\sqrt{1 - \frac{1 - 2\eta_0}{r\cos\theta}}\ \left|0\right\rangle + \sqrt{1 + \frac{1 - 2\eta_0}{r\cos\theta}}\ \left|1\right\rangle\right). \end{aligned}$$

The conditions $\nu_+ \geq 0$ and $\nu_- \leq 0$ are equivalent to $\frac{1}{2}(1 - r^2\cos^2\theta) \leq \eta_0 \leq \frac{1}{2}(1 + r^2\cos^2\theta)$.

# Chapter 5

# A contextuality witness inspired by optimal state discrimination

In this chapter we present the results in "A contextuality witness inspired by optimal state discrimination" [26], authored by Carles Roch i Carceller and Jonatan Bohr Brask. A final version of this work is still in preparation.

## 5.1 Abstract

Many protocols and tasks in quantum information science rely inherently on the fundamental aspect of contextuality to provide advantages over their classical counterparts, and contextuality represents one of the main differences between quantum and classical physics. In this work we present a witness for preparation and measurement contextuality inspired by optimal two-state discrimination. The main idea is based on finding the accessible averaged success and error probabilities in both classical and quantum models. We can then construct a noncontextuality inequality and associated witness, which we find to be robust against depolarising noise in both state preparation and measurements.

## 5.2 Introduction

Contextuality is a fundamental aspect of quantum mechanics which states that the result of measurements may depend on which other compatible measurements are jointly performed, in contrast with classical models, which allow no such dependence and are noncontextual. The Bell-Kochen-Specker theorem [18, 19] demonstrates that quantum theory is incompatible with noncontextual

hidden-variable models. It has been demonstrated that contextuality constitutes as a resource for various applications in quantum information including magic states [133], quantum key distribution [134], device-independent security [135] and quantum randomness certification [136, 137]. The traditional definition of contextuality requires a composite system, and its standard proof applies to Hilbert spaces of dimension three or higher [138, 139]. The notion of (non)contextualilty has been further generalised in the work of Spekkens [20], based on operational equivalences and ontological models. Similar to Kochen-Specker, generalised contextuality has also been proven to provide a resource for certain quantum information tasks. For instance parity-oblivious multiplexing [140, 141], random-access codes [142], quantum randomness certification [17], communication [143–145], and state discrimination [25, 27]. Other interesting works have studied some limitations of physical theories [144]. Quantum theory has also been shown to be less preparation contextual than the general operational theory known as box world [146].

In this work we aim to find a simple witness for generalised contextuality in the sense introduced in [20]. While a number of contextuality witnesses exist in the literature [147–152], here we benefit from a simple prepare-and-measure scenario with two preparations and a single measurement to find a good contextuality witness inspired by optimal state discrimination.

## 5.3 Basic notions in state discrimination

Any state discrimination scenario is formed by state preparations and effects [153, 154]. The former are labeled by preparation procedures $x \in X$ and the latter by answers $b \in B$ to the questions in $X$, which can be answered in an experiment. From the list of questions $X$ and answers $B$, the gathered data is usually expressed as conditional probabilities $p(b|x)$. The goal in state discrimination is to determine $x$ from the transmitted states, i.e. achieving $b = x$. Then, an optimisation problem is built, where these probabilities take a principal role. All the involved correlations are further constrained to obey a particular set of rules based on a concrete model (quantum, noncontextual,...). Furthermore, as is customarily done in state discrimination settings, we name the probabilities depending whereas the answer is correct, wrong or else. If the answer to the question $x$ is $b = x$, we define $p(b = x|x)$ as the *success* probability, whereas $p(b \neq x|x)$ is called the *error* probability. One must also consider events where the answer $b$ is not in the set of questions $X$ (i.e. $X \subseteq B$). We group answers not in $X$ and label them by $b = \emptyset$. We denote $p(b = \emptyset|x)$ the *inconclusive* probability.

Success, error and inconclusive probabilities play each a different role in the discrimination scenario [29, 95, 97]. Different state discrimination tasks can be defined by different figures of merits, which are functions of the observed conditional probabilities, and different constraints on the same probabilities. For example, the goal in minimum error state discrimination (MESD) is to maximise the success probability whilst inconclusive events do not occur [108, 155, 156] (hence converting the goal into a minimisation of the error probability due to normalisation). On the other hand, in unambiguous state discrimination (USD), the goal is also to maximise the success probability, with the main constraint that error probabilities must vanish [79–81] (thus converting the goal into a minimisation of inconclusive probabilities). Lastly, in maximum confidence state discrimination (MCSD), the goal is to maximise the confidence, that is, the probability of receiving input $x$ given the outcome $b = x$) which can be expressed as the success probability divided by the rate of events of interest [101–104, 106, 157]. Concretely, for a particular state of interest $x$, $C_x := p_x p(b = x|x)/\eta_x$, for $\eta_b = \sum_x p_x p(b|x)$, where $p_x$ are the prior probabilities for each preparation $x$. If one centers on the whole prepared ensemble instead, the total equally-weighted confidence is $C = \sum_x C_x/n$, for $n$ being the total number of state-preparations. No further constraints are applied to MCSD, making it rather a more general approach. Also, it can be reduced to MESD and USD as concrete cases. If $C_x = 1$, the input $x$ must be unambiguously identified, or if $C = 1$ the whole ensemble is unambiguously discriminated, resulting in USD. On the other hand, MESD is recovered by adopting $\sum_x \eta_x C_x$ as the figure of merit.

## 5.4 Scenario

We focus on two-state discrimination with equiprobable preparations, characterised by the sets of preparations $X = \{0, 1\}$ and outcomes $B = \{0, 1, \emptyset\}$, and $p_x = 1/2$. We also introduce the averaged success $p_{\text{suc}}$, error $p_{\text{err}}$, and inconclusive $p_\emptyset$ probabilities as

$$p_{\text{suc}} := \frac{1}{2}\left(p(0|0) + p(1|1)\right), \tag{5.1}$$

$$p_{\text{err}} := \frac{1}{2}\left(p(1|0) + p(0|1)\right), \tag{5.2}$$

$$p_\emptyset := \frac{1}{2}\left(p(\emptyset|0) + p(\emptyset|1)\right) = 1 - p_{\text{suc}} - p_{\text{err}}. \tag{5.3}$$

We will fix $p_\emptyset$ and ask the following question: which regions in correlation space, parameterized by $p_{\text{suc}}$ and $p_{\text{err}}$, are feasible in quantum mechanics or

in a noncontextual model? The answer to this question is not trivial if state preparations are not perfectly distinguishable. For fixed inconclusive rate, the sum $p_{\text{suc}} + p_{\text{err}} = 1 - p_{\emptyset}$ is fixed and we can focus on the difference. We therefore define the following witness on the level of probabilities

$$\mathcal{W} := \frac{1}{2}\left(p_{\text{suc}} - p_{\text{err}}\right) \ . \tag{5.4}$$

For each model, we will separately formulate an optimisation problem, and find a bound on $\mathcal{W}$. The feasible region is necessarily convex since, for two different measurement strategies producing different behaviours, probabilistically choosing between them (using local randomness) defines another valid measurement strategy. The corresponding behaviour will then be the convex combination of the first two behaviours. We can thus use techniques in convex optimisation to efficiently solve the maximisation problem for each model.

### 5.4.1 Quantum model

Consider an ensemble of two noisy states $\rho_x = r_s \left|\psi_x\right\rangle \left\langle\psi_x\right| + (1 - r_s)\mathbb{1}/2$ for $x = 0, 1$, characterised by the overlap $\delta = \left|\left\langle\psi_0|\psi_1\right\rangle\right|$. Let $\hat{\pi}_b$ represent a valid POVM for $b = 0, 1, \emptyset$, such that $\text{Tr}\left[\rho_x \hat{\pi}_b\right] = p(b|x)$ according to the Born rule. Our goal is to find the maximum difference between success and error probabilities, for a fixed inconclusive rate. To do so, let us introduce the following operator

$$\hat{\Delta}_x := \frac{(-1)^x}{2}\left(\hat{\pi}_0 - \hat{\pi}_1\right) \ . \tag{5.5}$$

We aim to find the maximum difference

$$\mathcal{W}^{\text{Q}} := \max \frac{1}{2}\left(p^{\text{Q}}_{\text{suc}} - p^{\text{Q}}_{\text{err}}\right) = \max \sum_x \text{Tr}\left[\hat{\Delta}_x \rho_x\right] \ , \tag{5.6}$$

where the optimisation is over all measurements forming valid POVMs $\hat{\pi}_b \geq 0$ and $\sum_b \hat{\pi}_b = \mathbb{1}$, and subject to $p_{\emptyset} = \frac{1}{2}\text{Tr}\left[(\rho_0 + \rho_1)\hat{\pi}_{\emptyset}\right]$. This maximisation can be rendered as a semidefinite program (SDP) [30].

In Sec. 5.7 we find an analytical form of the optimal measurement that solves our initial problem. The solution to (5.6) is given by

$$\begin{aligned} \mathcal{W}^{\text{Q}} &= r_s \sqrt{(1-\delta^2)\left(1 - \frac{2p_{\emptyset}}{1 + r_s\delta}\right)} && \text{for } p_{\emptyset} \leq r_s\delta \\ \mathcal{W}^{\text{Q}} &= r_s\sqrt{1-\delta^2}\sqrt{1 - r_s^2\delta^2}\,\frac{1 - p_{\emptyset}}{1 - r_s^2\delta^2} && \text{for } p_{\emptyset} \geq r_s\delta \ . \end{aligned} \tag{5.7}$$

One can also write down the optimal success and error probabilities. For $p_\emptyset \leq r_s\delta$

$$p_{\text{suc}}^{\text{Q}} = \frac{1}{2}\left(1 + \mathcal{W}^{\text{Q}} - p_\emptyset \frac{1+\delta}{1+r_s\delta}\right) \ , \tag{5.8}$$

and for $p_\emptyset \geq r_s\delta$:

$$p_{\text{suc}}^{\text{Q}} = \frac{1}{2}\left(1 + \mathcal{W}^{\text{Q}} - p_\emptyset\right) \ , \tag{5.9}$$

and $p_{\text{err}}^{\text{Q}} = 1 - p_{\text{suc}}^{\text{Q}} - p_\emptyset$.

The success and error probabilities we found are the maximal and minimal probabilities according to the quantum theory in a qubit state discrimination problem. Interestingly, one can recover the bounds from other protocols as specific cases. For instance, if the experiment only reproduces conclusive measurement outcomes ($p_\emptyset = 0$), the problem is reduced to the usual MESD. Then, one recovers the Helstrom bound as a minimum error rate $p_{\text{err}}$ [99, 108, 158]. On the other hand, if the experiment is designed with a null error rate ($p_{\text{err}} = 0$) and zero noise ($r_s = 1$), one recovers USD. In that case, the maximal success probability is $p_{\text{suc}} = 1 - \delta$, leaving a minimal rate of inconclusive events $p_\emptyset = \delta$, the minimal value for USD [159, 160]. Finally, one can directly compute the maximum confidence of the whole ensemble by writing $C = p_{\text{suc}}/(p_{\text{suc}} + p_{\text{err}})$. One recovers the maximum confidence obtained in [25] if $p_\emptyset \leq r_s\delta$. For larger values of the inconclusive rate, one can still compute the maximum confidence with the same formula since MCD and the present scheme share the exact same goal (maximise the success and minimize the error probabilities).

### 5.4.2 Noncontextual model

Let us start by presenting the model we use to contrast quantum contextuality. We take an ontological model of the prepare-and-measure scenario [20, 27, 161]. The system is associated with an ontic state space $\Lambda$ in which each point $\lambda \in \Lambda$ completely defines all physical properties, i.e. the outcomes of all possible measurements. Each state preparation $x$ samples the ontic state space according to a probability distribution $\mu_x(\lambda)$, referred to as the *epistemic state*. Each measurement is defined by a set of *response functions*, that is, non-negative functions $\xi_b(\lambda)$ over the ontic space, such that $\sum_b \xi_b(\lambda) = 1$ for all $\lambda \in \Lambda$. The probability of obtaining the outcome $b$ when state $\mu_x$ was prepared is then

$$p(b|x) = \int_\Lambda d\lambda \ \mu_x(\lambda)\xi_b(\lambda). \tag{5.10}$$

While distinct ontic states can be perfectly discriminated, epistemic states with overlapping distributions cannot. This overlap introduces the notion of *confusability* between two epistemic states $\mu_x$ and $\mu_y$:

$$c_{x,y} := \int_{\text{supp}[\mu_x(\lambda)]} d\lambda \; \mu_y(\lambda) \; . \tag{5.11}$$

It is the discrimination of epistemic states which we compare against quantum state discrimination.

Furthermore, we require the considered ontological model to be preparation-noncontextual. Two preparations are said to be operationally equivalent if they cannot be distinguished by any measurement, and an ontological model is said to be preparation-noncontextual if all operationally equivalent preparations are assigned to the same epistemic state. To impose noncontextuality on the ontological model, we imply the existence of a particular pair of states $S := \{\mu_0, \mu_1\}$ and complementary states $S^\perp := \left\{\mu_0^\perp, \mu_1^\perp\right\}$. That is, for every set of states $\mu_0(\lambda), \mu_1(\lambda) \in S$ with pairwise confusability $c_{0,1}$ , there must exist a complementary set of states $\mu_0^\perp(\lambda), \mu_1^\perp(\lambda) \in S^\perp$ with the same pairwise confusability $c_{0,1}$. Each pair of states $\mu_x$ and $\mu_x^\perp$ must have non-overlapping supports, and any convex combination of those should be operationally equivalent. By noncontextuality, they must be equal $\frac{1}{2}\mu_x + \frac{1}{2}\mu_x^\perp = \frac{1}{2}\mu_y + \frac{1}{2}\mu_y^\perp$. This statement implies that any pair of epistemic states $\mu_x(\lambda)$ and $\mu_y(\lambda)$ with overlapping supports are equivalent on the overlap, i.e.

$$\mu_x(\lambda) = \mu_y(\lambda) \; \forall \lambda \in \text{supp}\left[\mu_x(\lambda)\right] \cup \text{supp}\left[\mu_y(\lambda)\right] \; . \tag{5.12}$$

This, in turn, results in symmetric confusabilities, $c := c_{x,y} = c_{y,x}$. Quantum and noncontextual models can be then compared through $\delta^2 = c$. The noncontextual model we use in this work is to be understood as an attempt to describe the quantum theory but with noncontextual preparations and measures. One can see it relies on a deterministic ontic space but with the added probabilistic nature brought by the epistemic states. Thus, it should reproduce a subset of the spectrum of quantum probabilities in a state discrimination scenario.

We now present the main problem in a noncontextual model. The two preparations are represented by the following epistemic states affected by depolarising noise

$$\begin{aligned} \tilde{\mu}_0(\lambda) &= r_s \mu_0(\lambda) + (1 - r_s)\mu_{\mathbb{1}/2}(\lambda) \\ \tilde{\mu}_1(\lambda) &= r_s \mu_1(\lambda) + (1 - r_s)\mu_{\mathbb{1}/2}(\lambda) \; . \end{aligned} \tag{5.13}$$

These can be characterised by the confusability of the noiseless epistemic states $c$ from (5.11). We also consider a single measurement with two conclusive outcomes $b = 0, 1$ and an inconclusive result $b = \emptyset$, represented by the response functions $\xi_b(\lambda)$. Let us define the analogous observable in (5.5) for the noncontextual model

$$\Delta_x^{\mathrm{NC}}(\lambda) := \frac{(-1)^x}{2} \left(\xi_0(\lambda) - \xi_1(\lambda)\right) \ . \tag{5.14}$$

Then, we can rewrite the problem as the maximisation of

$$\begin{aligned} \mathcal{W}^{\mathrm{NC}} &:= \max \frac{1}{2} \left(p_{\mathrm{suc}}^{\mathrm{NC}} - p_{\mathrm{err}}^{\mathrm{NC}}\right) \\ &= \max \sum_x \int_\Lambda d\lambda \ \tilde{\mu}_x(\lambda) \Delta_x^{\mathrm{NC}}(\lambda) \ , \end{aligned} \tag{5.15}$$

subject to $\xi_b(\lambda) \geq 0$ and $\sum_b \xi_b(\lambda) = 1$, $\forall \lambda$ are valid response functions, and the rate of inconclusive events is given by $p_\emptyset = \frac{1}{2} \int d\lambda \left(\tilde{\mu}_0(\lambda) + \tilde{\mu}_1(\lambda)\right) \xi_\emptyset(\lambda)$ and fixed. In Sec. 5.7 we show how this maximisation can be rendered as a simple linear problem, for which we are able to find an analytical solution:

$$\begin{aligned} \mathcal{W}^{\mathrm{NC}} &= r_s \left(1 - c\right) \left(1 - \frac{p_\emptyset}{1 + r_s c}\right) \quad \text{for } p_\emptyset \leq (1 + r_s c)/2 \\ \mathcal{W}^{\mathrm{NC}} &= (1 - p_\emptyset) \frac{r_s (1 - c)}{1 - r_s c} \quad \text{for } p_\emptyset \geq (1 + r_s c)/2 \ . \end{aligned} \tag{5.16}$$

This results in the following success and error probabilities. For $p_\emptyset \leq (1 + r_s c)/2$:

$$p_{\mathrm{suc}}^{\mathrm{NC}} = \frac{1 + r_s}{2} \left(1 - \frac{p_\emptyset}{1 + r_s c}\right) - \frac{r_s c}{2} \ , \tag{5.17}$$

and for $p_\emptyset \geq (1 + r_s c)/2$:

$$p_{\mathrm{suc}}^{\mathrm{NC}} = \frac{1}{2} \left(1 + \mathcal{W}^{\mathrm{NC}} - p_\emptyset\right) \ , \tag{5.18}$$

and $p_{\mathrm{err}}^{\mathrm{NC}} = 1 - p_{\mathrm{suc}}^{\mathrm{NC}} - p_\emptyset$.

## 5.5 Results

One can see the space of probabilities drawn in Fig. 5.1. The white region delimited by the black contour shows the feasible space in the case of fully

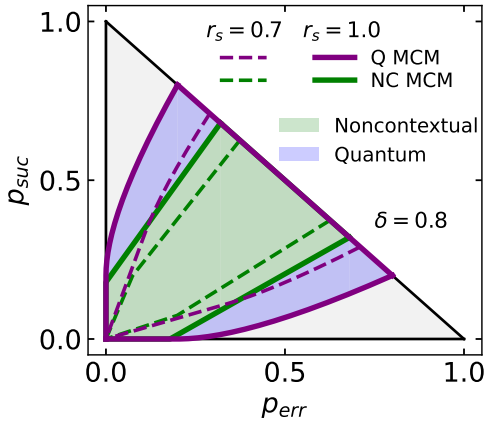



Figure 5.1: Figure extracted from Ref. [26]. Space of probabilities corresponding to a two-state discrimination setting. Continuous lines denote maximum confidence measurements in both quantum (purple) and noncontextual (green) models. Even with a bounded value of noise ($r_s = 0.7$), the MCM line according to the quantum model falls outside the noncontextual region.

distinguishable preparations. That is, when states can be directly identified with ontic states $\lambda$. The area shaded in blue shows the feasible space according to the quantum theory. In its contour we can find $p^{\mathrm{Q}}_{\mathrm{suc}}$ from (5.8) and (5.9), for $r_s = 1$. The region reproducible by the noncontextual model (green area) is included within the quantum limits. Similarly, we find $p^{\mathrm{NC}}_{\mathrm{suc}}$, from (5.17) and (5.18), in its contour, also for $r_s = 1$. The area of the quantum (and noncontextual) feasible space becomes narrower as the effects of superposition become more evident (i.e. when the overlap $\delta = \sqrt{c}$ increases). We can see then how the quantum predictions depart from classical interpretations. There, the contextual aspect of the quantum model becomes more visible.

Moreover, we can identify some extremes of the quantum region with the bounds found in each state discrimination protocol. The diagonal line that delimits the right-upper part of the feasible regions covers the state discrimination scenarios with zero inconclusive rates. The vertices of the quantum region on that line reproduce the Helstrom bound [77, 98, 99] obtained in MESD. The same applies for the vertices corresponding to the noncontextual line, which reproduce the maximal success probabilities in MESD obtained in [27]. Also, the maximal $p_{\mathrm{err}}$ and $p_{\mathrm{suc}}$ lying on the bottom and left-most sides reproduce the maximal unambiguous error and success rates for quantum [79] and noncontextual [25] models, obtained in USD. Finally, the whole purple line (called MCM lines for reasons that will be apparent later) surrounding all the quantum feasible region reproduce the maximum confidence for any value

of $p_\emptyset$ [83, 101], whilst the green line does the same but for a noncontextual model [25]. We can see that by writing the confidence $C = p_{\text{suc}}/(p_{\text{suc}} + p_{\text{err}})$, the value of which (along the lines of Fig. 5.1) coincides with the bounds found in the literature [25, 83, 101, 102, 157].

When noisy states affected by depolarizing noise are taken into play, the bounds on all protocols depart from the borders of the quantum region. The Helstrom bound from both quantum and noncontextual MESD comes closer to the center of the probability space as noise increases. Also, when noise is taken into account, USD is not possible as here we can see that the bottom and left-most borders are not reachable. The space closed by the MCM lines also turns narrower. For a given noisy ensemble, the points on the quantum region outside the MCM lines are not accessible.

During the whole discussion we have been ignoring the fact that, indeed, the POVM in (5.28) is a maximum confidence measurement (MCM)[83]. Thus, according to Fig. 5.1, an MCM (for an entire ensemble $\rho$ of qubit states) reaches the limits of the space of feasible correlations. This indicates that an MCM is the most optimal measure in any (qubit-)state discrimination scenario.

A different perspective is plotted in Fig. 5.2. Contextual behavior is manifested above the dashed-black line, in the blue shaded region. We can write this statement a the following noncontextual inequality: Given an experiment with two preparations $X \in \{0, 1\}$, with a distinguishability (which can be characterized by an overlap $|\langle\psi_0|\psi_1\rangle|$) bounded by $\delta$, the following inequality is always fulfilled in a noncontextual scenario:

$$\mathcal{W}^* := \left.\frac{p_{\text{suc}}^{\text{NC}} - p_{\text{err}}^{\text{NC}}}{2}\right|_{r_s=1} \leq (1-\delta^2)\left(1 - \frac{p_\emptyset}{1+\delta^2}\right) . \tag{5.19}$$

Whenever that inequality is violated in a two-state prepare-and-measure scenario, contextuality is solely manifested. For a concrete value of noise and nonzero inconclusive events, contextuality can still be harnessed. The quantum model (bounded by the purple line in Fig. 5.2) still reproduces contextual behavior.

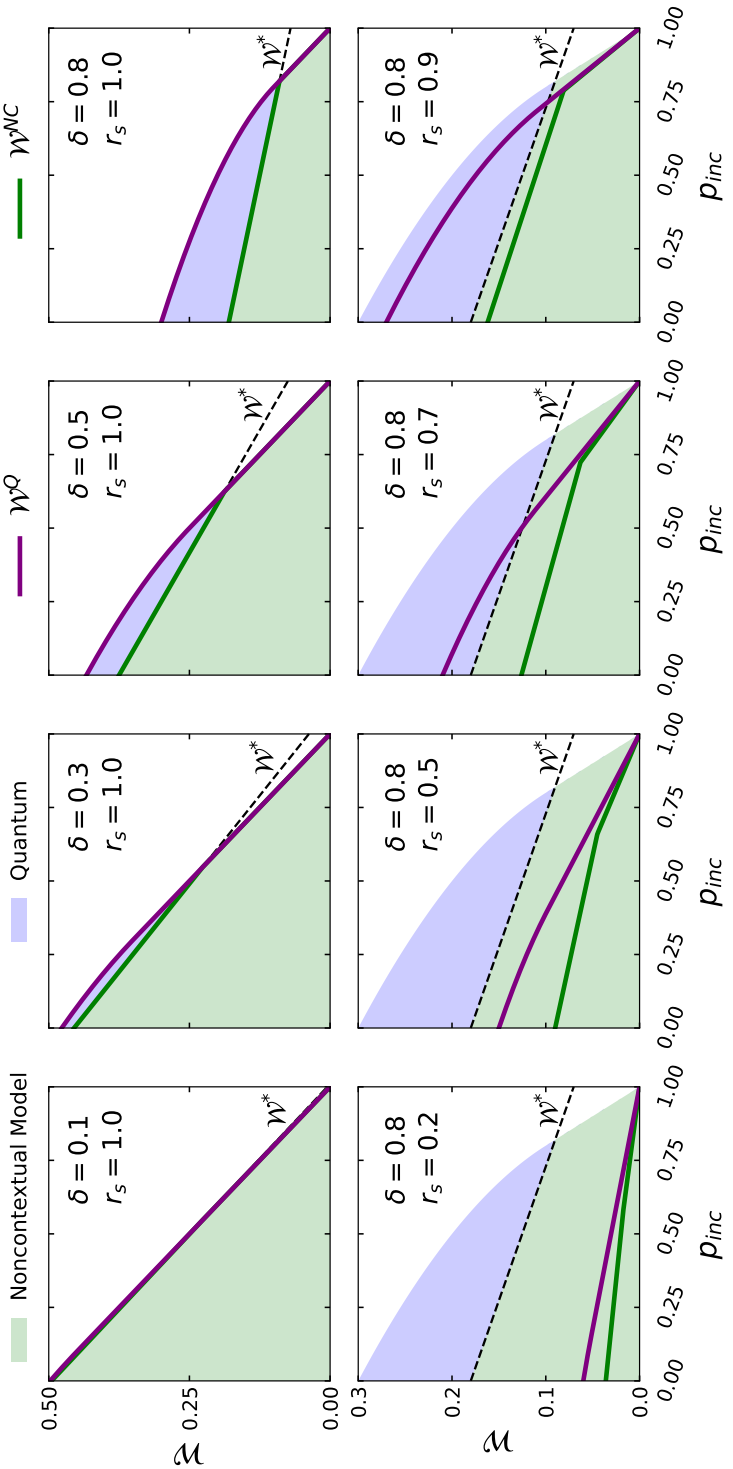



Figure 5.2: Figure extracted from Ref. [26]. Bounds on the witness $\mathcal{W}$ according to quantum and noncontextual models. On the first row we show noiseless cases with different overlaps. Below, on the second row, we fix a particular overlap and show the effects of depolarising noise on the preparation. The green area denotes the feasible values according to quantum and noncontextual models and the blue region solely for the quantum model.

Note that the distinguishability (denoted through the overlap $|\langle\psi_0|\psi_1\rangle|$) does not necessarily need to be fixed but it is enough to bound it. We can do that because the bounds we found in $\mathcal{W}$ decrease as preparations become less distinguishable. Also, note that the measurement can be completely left uncharacterised as only statistics are taken into account. These two particularities highlight the semi-device independent setting of the scenario we are discussing.

One can also introduce a measurement affected by depolarising noise, in addition to the noise in the state preparation. Depolarising noise on the measurement affects directly the observed probabilities according to the *efficiency* $r_m$ as follows

$$p(b|x) \quad \longrightarrow \quad r_m p(b|x) + (1 - r_m)\frac{1}{n_{\mathrm{B}}} \ , \tag{5.20}$$

for $n_{\mathrm{B}}$ being the number of different measurement outcomes (in our case $n_{\mathrm{B}} = 3$). This noise is reflected on the witness as a visibility parameter, which turns the maximum value of $\mathcal{W}$ to $r_m\mathcal{W}$ for both quantum and noncontextual models. We see in Fig. 5.3 how both noises in preparation and measurement affect the manifestation of contextuality in a state discrimination experiment. Observe that even noisy states can exhibit contextual behavior up to the following value of noise:

$$r_m \geq \frac{(1-\delta^2)(1-\frac{p_\varnothing}{1+\delta^2})}{r_s\sqrt{(1-\delta^2)(1-\frac{2*p_\varnothing}{1+r_s\delta})}} \ , \tag{5.21}$$

which follows the dashed contour in Fig. 5.3.

## 5.6 Conclusion

We presented a witness of contextuality in two-state discrimination scenarios. We started by formulating the problem of finding the most optimal measurement in a two state discrimination setting that allows for inconclusive events. That problem led to the definition of a quantity $\mathcal{W}$ in (5.4). Not only we solved the optimality problem for the quantum theory but also for a noncontextual ontological model. We found that $\mathcal{W}$ can be used as a witness of contextuality. Furthermore, we defined a semi-device independent noncontextuality inequality in (5.19) that delimits the frontier between noncontextual and solely contextual behaviors on the probability space feasible to the quantum theory. That inequality allows for a flexible rate of inconclusive

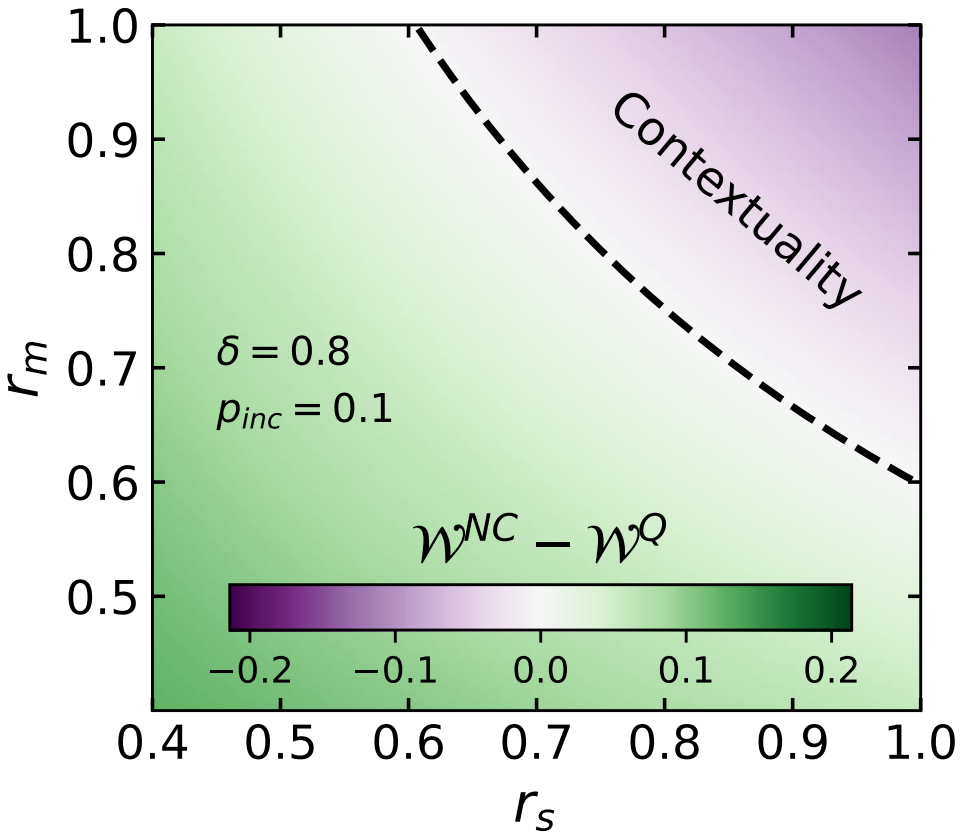



Figure 5.3: Figure extracted from Ref. [26]. Solely manifestation of contextuality for different values of depolarising noise on the state preparation ($r_s$) and on the measurement ($r_m$).

events and overlaps, and is robust against depolarising noise both in the preparation and measurement devices.

## 5.7 S1: Optimal measurements

In this part of the supplemental material, we derive the optimal measurements for both the quantum and noncontextual models considered in the main text. We call a measurement optimal when it maximises the probability success whilst the error probability is kept at minimum.

### 5.7.1 Quantum model

Let us start by writing the semidefinite program that finds the optimal measurement according to the quantum theory. We express the difference we aim to maximise as

$$\begin{aligned}\mathcal{W}^{\mathrm{Q}} = \frac{1}{2}\left(p^{\mathrm{Q}}_{\mathrm{suc}} - p^{\mathrm{Q}}_{\mathrm{err}}\right) &= \frac{1}{2}\left(\mathrm{Tr}\left[\rho_0\hat{\pi}_0\right] + \mathrm{Tr}\left[\rho_1\hat{\pi}_1\right]\right) - \frac{1}{2}\left(\mathrm{Tr}\left[\rho_0\hat{\pi}_1\right] + \mathrm{Tr}\left[\rho_1\hat{\pi}_0\right]\right) \\ &= \frac{1}{2}\,\mathrm{Tr}\left[(\rho_0-\rho_1)(\hat{\pi}_0-\hat{\pi}_1)\right] = \sum_{x=0}^{1}\mathrm{Tr}\left[\rho_x\hat{\Delta}_x\right] \ , \quad (5.22)\end{aligned}$$

where we included the operator $\hat{\Delta}_x$ introduced in the main text. One can write down the problem in the following SDP form:

$$\begin{aligned}\underset{\{\hat{\pi}_b\}}{\text{maximise}} \quad & \frac{1}{2}\,\mathrm{Tr}\left[\rho_x\hat{\Delta}_x\right] \\ \text{subject to:} \quad & \hat{\pi}_b \geq 0 \quad \sum_b \hat{\pi}_b = \mathbb{1} \\ & \frac{1}{2}\,\mathrm{Tr}\left[(\rho_0+\rho_1)\hat{\pi}_\varnothing\right] = p_\varnothing \ .\end{aligned} \quad (5.23)$$

We begin by analysing the optimality conditions of the POVM $\hat{\pi}_b$. The corresponding Lagrangian is given by

$$\begin{aligned}\mathcal{L} = &\frac{1}{2}\,\mathrm{Tr}\left[(\rho_0-\rho_1)(\hat{\pi}_0-\hat{\pi}_1)\right] + \sum_b r_b\,\mathrm{Tr}\left[\hat{\pi}_b\sigma_b\right] \quad (5.24)\\ &+ \mathrm{Tr}\left[K\left(\mathbb{1} - \sum_b \hat{\pi}_b\right)\right] + s\left(p_\varnothing - \frac{1}{2}\,\mathrm{Tr}\left[(\rho_0+\rho_1)\hat{\pi}_\varnothing\right]\right) \ ,\end{aligned}$$

where we introduced the following dual variables: $r_b\sigma_b$ for the PSD constraint, $K$ accounting for the normalisation constraint and $s$ for the constraint fixing

the inconclusive rate. The dual problem can be straight formulated through the supremum of the Lagrangian

$$\mathcal{S} = \sup_{\hat{\pi}_b} \ \mathrm{Tr}\left[K\right] + Qs + \sum_b \mathrm{Tr}\left[\hat{\pi}_b G_b\right] \ , \tag{5.25}$$

where we defined

$$G_b = \frac{1}{2}\mathrm{Tr}\left[\left(\rho_0 - \rho_1\right)\left(\delta_{b,0} - \delta_{b,1}\right)\right] + r_b\sigma_b - \frac{1}{2}\left(\rho_0 + \rho_1\right)\delta_{b,\emptyset} - K \ . \tag{5.26}$$

The supremum in (5.25) will diverge unless $G_b = 0$. This leads to the Lagrange stability optimality condition. Complementary slackness reads $r_b \,\mathrm{Tr}\left[\sigma_b\hat{\pi}_b\right] = 0$, which in a qubit space means that $\pi_b$ must be rank-1. With all that, we are ready to derive the form of the optimal POVM. We consider a pair of pure states $\rho_x$ oriented symmetrically with respect to the $Z$ pole in the Bloch sphere. That means

$$\begin{aligned} \rho_0 &= \frac{1}{2}\left[\mathbb{1} + r\sqrt{1-\delta^2}X + r_s\delta Z\right] \\ \rho_1 &= \frac{1}{2}\left[\mathbb{1} - r_s\sqrt{1-\delta^2}X + r\delta Z\right] \ . \end{aligned} \tag{5.27}$$

By fixing the orientation of the states in that way,the problem acquires a symmetry with respect to the states. This is very convenient since we can directly find the analytical form of $\hat{\pi}_\emptyset$ through the last constraint in (5.23): $\hat{\pi}_\emptyset = \frac{p_\emptyset}{1+r_s\delta}\left[\mathbb{1} + Z\right]$. The other two POVM elements ($\hat{\pi}_0$ and $\hat{\pi}_1$) can be also directly found noting that the maximum in (5.23) is reached when $\hat{\pi}_0 - \hat{\pi}_1$ is proportional to $\rho_0 - \rho_1 = r_s\sqrt{1-\delta^2}X$ (according to (5.27)). At the end of the day, this leaves us with the following optimal POVM:

$$\begin{aligned} \hat{\pi}_0 =& \frac{1}{2}\left(1 - \frac{p_\emptyset}{1+r_s\delta}\right)\left[\mathbb{1} + \frac{1+r_s\delta}{r_s\sqrt{1-\delta^2}}\frac{\mathcal{W}^{\mathrm{Q}}}{1+r_s\delta-p_\emptyset}X - \frac{p_\emptyset}{1+r_s\delta-p_\emptyset}Z\right] \\ \hat{\pi}_1 =& \frac{1}{2}\left(1 - \frac{p_\emptyset}{1+r_s\delta}\right)\left[\mathbb{1} - \frac{1+r_s\delta}{r\sqrt{1-\delta^2}}\frac{\mathcal{W}^{\mathrm{Q}}}{1+r_s\delta-p_\emptyset}X - \frac{p_\emptyset}{1+r_s\delta-p_\emptyset}Z\right] \\ \hat{\pi}_\emptyset =& \frac{2Q}{1+r_s\delta}\frac{1}{2}\left[\mathbb{1} + Z\right] \ , \end{aligned} \tag{5.28}$$

for $p_\emptyset \leq r_s\delta$, and

$$\begin{aligned} \hat{\pi}_0 =& \frac{1}{2}\frac{1-p_\emptyset}{1-r_s^2\delta^2}\left[\mathbb{1} + \sqrt{1-r_s^2\delta^2}X - r_s\delta Z\right] \\ \hat{\pi}_1 =& \frac{1}{2}\frac{1-p_\emptyset}{1-r_s^2\delta^2}\left[\mathbb{1} - \sqrt{1-r_s^2\delta^2}X - r_s\delta Z\right] \\ \hat{\pi}_\emptyset =& \frac{2(p_\emptyset - r_s^2\delta^2)}{1-r_s^2\delta^2}\frac{1}{2}\left[\mathbb{1} + \frac{1-p_\emptyset}{p_\emptyset - r_s^2\delta^2}Z\right] \ , \end{aligned} \tag{5.29}$$

for $p_\emptyset \geq r_s\delta$. That POVM yields the optimal measurement that maximises the difference between success and error probabilities for a fixed rate of inconclusive events.

### 5.7.2 Noncontextual model

In a noncontextual model we can write the problem in the following maximisation form:

$$\begin{aligned} \underset{\xi_b(\lambda)}{\text{maximise}} \quad & \frac{1}{2}\int d\lambda \left(\tilde{\mu}_0(\lambda) - \tilde{\mu}_1(\lambda)\right)\left(\xi_0(\lambda) - \xi_1(\lambda)\right) \\ \text{subject to:} \quad & \xi_b(\lambda) \geq 0,\ \sum_b \xi_b(\lambda) = 1\ \forall\lambda \\ & p_\emptyset = \frac{1}{2}\int d\lambda \left(\tilde{\mu}_0(\lambda) + \tilde{\mu}_1(\lambda)\right)\xi_\emptyset(\lambda)\ . \end{aligned} \tag{5.30}$$

We consider a pair of noisy epistemic states affected by depolarising noise:

$$\begin{aligned} \tilde{\mu}_0(\lambda) &= r_s\mu_0(\lambda) + (1-r_s)\mu_{\mathbb{1}/2}(\lambda) \\ \tilde{\mu}_1(\lambda) &= r_s\mu_1(\lambda) + (1-r_s)\mu_{\mathbb{1}/2}(\lambda)\ . \end{aligned} \tag{5.31}$$

These are characterised by the confusability of the noiseless states given by

$$c = c_{1,0} = \int_{\text{supp}[\mu_0(\lambda)]} d\lambda\mu_1(\lambda)\ . \tag{5.32}$$

For low enough rates of inconclusive events, the optimal response functions are those which unambiguously discriminate the noiseless epistemic states. These are of the following form

$$\xi_0(\lambda) = \begin{cases} q & \text{if } \lambda \in \text{supp}[\mu_1^\perp(\lambda)] \\ 0 & \text{otherwise}\ . \end{cases} \quad \xi_1(\lambda) = \begin{cases} q & \text{if } \lambda \in \text{supp}[\mu_0^\perp(\lambda)] \\ 0 & \text{otherwise}\ . \end{cases} \tag{5.33}$$

One can determine the value of $q$ in terms of the rate of inconclusive events and obtain

$$q = \frac{1-p_\emptyset}{1-r_sc}\ , \tag{5.34}$$

leaving the following extremal success and error probabilities:

$$p_{\text{suc}}^{\text{NC}} = \frac{1-p_\emptyset}{2}\left(1 + \frac{r_s(1-c)}{1-r_sc}\right) \quad p_{\text{err}}^{\text{NC}} = \frac{1-r_s}{2}\frac{1-p_\emptyset}{1-r_sc}\ . \tag{5.35}$$

Normalisation implies in (5.34) that $p_\emptyset \geq (1 + r_s c)/2$. In the noiseless case, that is the lower bound on the rate of inconclusive events in USD. Also, note that the confidence can be written as $C = p_{\text{suc}}/(1 - p_\emptyset)$, which according to (5.35) one recovers the maximum confidence in [25] according to a noncontextual model.

For smaller rates $p_\emptyset$, the support of the response functions corresponding to conclusive outcomes will shift to the support of the oposite states. In other words, we can write down these response functions as follows

$$\xi_0(\lambda) = \begin{cases} a & \text{if } \lambda \in \text{supp}[\mu_0(\lambda)] \cup \text{supp}[\mu_1^\perp(\lambda)] \\ b & \text{if } \lambda \in \text{supp}[\mu_0^\perp(\lambda)] \cup \text{supp}[\mu_1^\perp(\lambda)] \\ a - b & \text{if } \lambda \in \text{supp}[\mu_0(\lambda)] \cup \text{supp}[\mu_1(\lambda)] \\ 0 & \text{otherwise} \end{cases} \tag{5.36}$$

$$\xi_1(\lambda) = \begin{cases} a & \text{if } \lambda \in \text{supp}[\mu_0^\perp(\lambda)] \cup \text{supp}[\mu_1(\lambda)] \\ b & \text{if } \lambda \in \text{supp}[\mu_0^\perp(\lambda)] \cup \text{supp}[\mu_1^\perp(\lambda)] \\ a - b & \text{if } \lambda \in \text{supp}[\mu_0(\lambda)] \cup \text{supp}[\mu_1(\lambda)] \\ 0 & \text{otherwise} \end{cases} . \tag{5.37}$$

Note that if $a = b$ we recover the response functions in (5.33). This allows us to rewrite the initial optimisation problem (5.30) in the following form

$$\begin{aligned} \underset{a,b}{\text{maximise}} \quad & \frac{1}{2} a r_s (1 - c) \\ \text{subject to} \quad & a - b \leq 2 \ , \quad b \leq \frac{1}{2} \\ & p_\emptyset = 1 - a(1 + r_s c) + 2 b r_s c \ . \end{aligned} \tag{5.38}$$

The optimal values of the parameters $a$ and $b$ are

$$a = 1 - \frac{p_\emptyset}{1 + r_s c} \quad b = \frac{1}{2} \ . \tag{5.39}$$

Then, one can write the success and error probabilities directly as follows:

$$p_{\text{suc}}^{\text{NC}} = \frac{1 + r_s}{2} \left( 1 - \frac{p_\emptyset}{1 + r_s c} \right) - \frac{r_s c}{2} \tag{5.40}$$

$$p_{\text{err}}^{\text{NC}} = \left( r_s c + \frac{1 - r_s}{2} \right) \left( 1 - \frac{p_\emptyset}{1 + r_s c} \right) - \frac{r_s c}{2} \ . \tag{5.41}$$

One can use this result to obtain the maximum confidence for smaller inconclusive rates (i.e. for $p_\emptyset \leq (1 + r_s c)/2$), which yields

$$\max C^{\mathrm{NC}} = \frac{1}{2(1 - p_\emptyset)} \left( (1 + r_s) \left( 1 - \frac{p_\emptyset}{1 + r_s c} \right) - r_s c \right) \ . \tag{5.42}$$

We can claim that this is the maximum confidence since it is also achieved by a measurement that simultaneously minimizes the error and maximises the success.

# Chapter 6

# Quantum vs. noncontextual randomness certification

In this chapter we present the results in "Quantum vs. noncontextual semi-device independent randomness certification" [17], authored by Carles Roch i Carceller, Kieran Flatt, Hanwool Lee, Jonatan Bohr Brask and Joonwoo Bae. This work is published in Physical Review Letters.

## 6.1 Abstract

We compare the power of quantum and classical physics in terms of randomness certification from devices which are only partially characterised. We study randomness certification based on state discrimination and take noncontextuality as the notion of classicality. A contextual advantage was recently shown to exist for state discrimination. Here, we develop quantum and noncontextual semi-device independent protocols for random-number generation based on maximum-confidence discrimination, which generalises unambiguous and minimum-error state discrimination. We show that, for quantum eavesdropppers, quantum devices can certify more randomness than noncontextual ones whenever none of the input states are unambiguously identified. That is, a quantum-over-classical advantage exists.

## 6.2 Introduction

Quantum physics departs radically from everyday experience. Observations on quantum systems can defy classical notions of cause and effect and exploiting quantum effects enables advantages for a number of applications including precision sensing, computing, and information security. Understanding the

quantum-classical boundary is both of fundamental importance to the foundations of physics in general and of relevance to characterising and quantifying quantum-over-classical advantages in specific tasks and applications.

In this work, we compare the power of quantum and classical physics for randomness certification. Random numbers are needed for many tasks in science and technology [14, 162]. In particular, high-quality randomness is central to cryptographic security and thus to much of modern information technology. Due to the inherent randomness in quantum measurements, strong guarantees can be established for the extraction of randomness from quantum systems. In fact, randomness can be certified with little or no trust in the devices used to generate it. In setups with multiple, separate parties, randomness can be certified in a device-independent (DI) setting, where the devices are treated as untrusted black boxes [58, 116, 117]. In that setting, the relevant notion of classicality is locality (also known as local causality), in the sense of Bell [15, 53], and the setup is required to violate a Bell inequality to generate randomness. This is, however, technologically very demanding, as the violation must be loophole free [116, 163–167]. Here, we focus on the semi-DI setting, where the black boxes are complemented by a few, general assumptions, representing an increased level of trust in the devices. This renders implementations much more accessible, and semi-DI randomness certification can be realised in simple prepare-and-measure setups [44, 61, 66–68, 168–174]. As our notion of classicality we adopt noncontextuality [19, 175], in the form introduced by Spekkens [20], which is applicable also in scenarios which do not have the multipartite structure of Bell tests.

We consider semi-DI randomness certification based on state discrimination, where the partial trust in the devices consists in an assumption about the distinguishability of the prepared states. In particular, we consider maximum-confidence state discrimination [101]. In the context of randomness certification, a semi-DI protocol based on unambiguous state discrimination was previously demonstrated [68], and in the context of comparing quantum and noncontextual models, a quantum advantage for minimum-error state discrimination was demonstrated by Schmid and Spekkens [27]. Maximum confidence state discrimination is more general, containing minimum-error and unambiguous state discrimination as particular cases. In related work, we demonstrate a quantum-over-noncontextual advantage for maximum-confidence state discrimination [25]. In the present work, we find a rich picture. In a setting where the devices are either quantum or noncontextual, but where the eavesdropper in both cases is allowed quantum powers, quantum devices outperform noncontextual ones. However, comparing a quantum

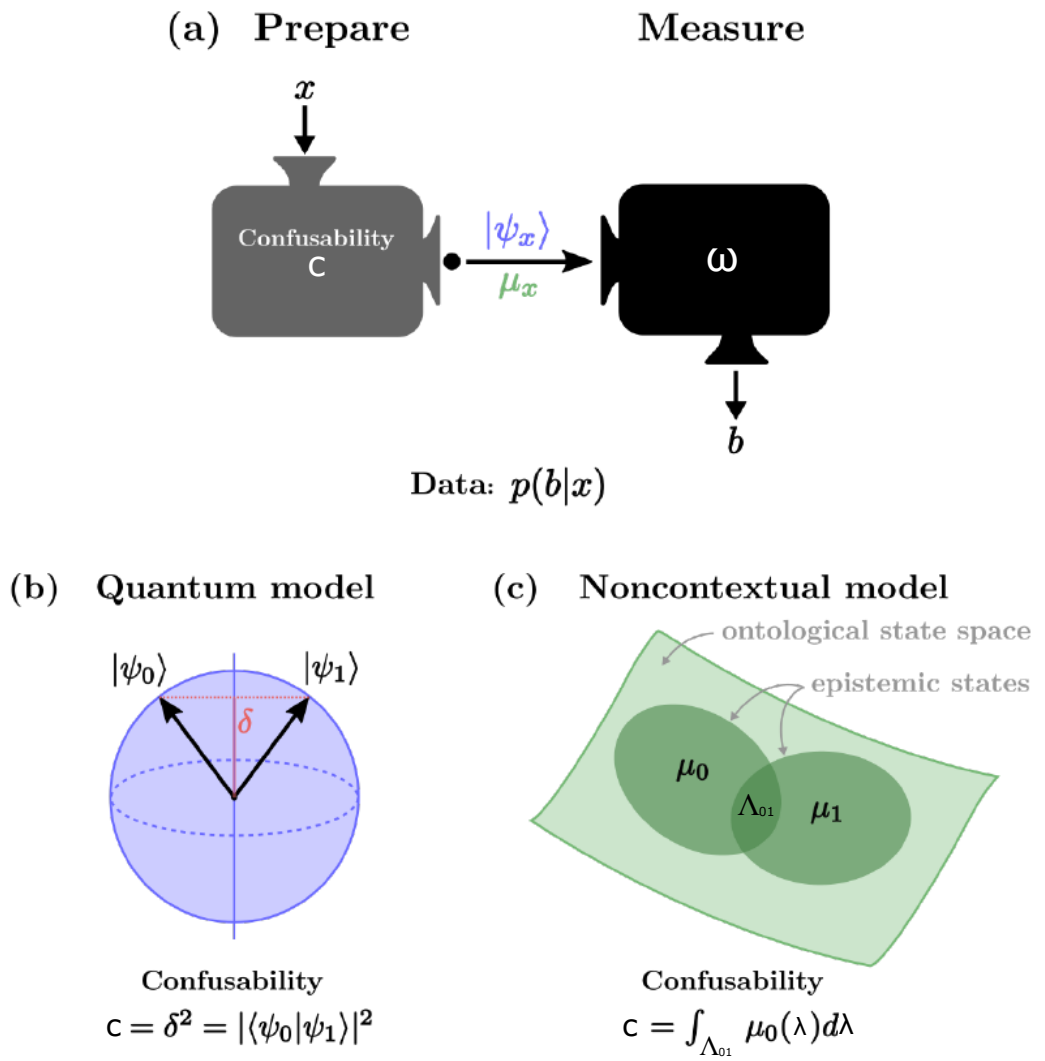



Figure 6.1: Figure extracted from Ref. [17]. **(a)** Prepare-and-measure scenario for state discrimination and randomness certification, in quantum and non-contextual settings. A preparation device takes an input and transmits states to a measurement device, which produces an output. From an assumption about the distinguishability of the states and the observed input-output correlations, the entropy in the raw output can be bounded and random numbers extracted from it. **(b)** In the quantum setting, the distinguishability is quantified by the overlap of the quantum states. For binary inputs, these can be represented by qubit states. **(c)** In the non-contextual setting, there is an ontological state space, consisting of perfectly distinguishable states. The preparation device emits epistemic states, given by probability distributions over ontological states. The distinguishability of epistemic states is quantified by the confusability, which measures the overlap of the corresponding distributions.

universe with quantum eavesdroppers against a noncontextual universe with noncontextual (hence less powerful) eavesdroppers, the amount of quantum certifiable randomness may be both larger than, smaller than or equal to the amount of noncontextual randomness, depending on the distinguishability of the states and the observed confidence of discrimination.

A prepare-and-measure setting for state discrimination and randomness certification is illustrated in Fig. 6.1(a). We will restrict our attention to binary inputs $x \in \{0, 1\}$ and ternary outputs $b \in \{0, 1, \emptyset\}$. In the case of state discrimination, $b$ represents a guess for which state was prepared, with $\emptyset$ labelling inconclusive outcomes. For randomness certification, the amount of true randomness present in the output $b$ can be lower bounded based on the observed distribution $p(b|x)$ and an assumption on the distinguishability of the prepared states. We start by considering state discrimination, first in the quantum case and then for noncontextual theories.

## 6.3 Quantum state discrimination

In quantum state discrimination, quantum states $\rho_x$ are prepared and the measurement device implements a POVM with elements $\hat{\pi}_b$, resulting in the distribution $p(b|x) = \mathrm{Tr}[\rho_x \hat{\pi}_b]$. For binary inputs, without loss of generality, the state space can be taken to be a qubit space. When the states are furthermore pure, $\rho_x = |\psi_x\rangle \langle\psi_x|$, their distinguishability can be quantified simply by their overlap $\delta = |\langle\psi_0|\psi_1\rangle|$. Its estimation will depend on the implementation. For instance, in [68] a time-bin encoding with coherent states was used. In that case, the overlap can be controlled through the amplitude of the pulses. Different quantifiers of performance can be adopted.

In minimum-error state discrimination (MESD), no inconclusive outcomes are permitted, $p(\emptyset|x) = 0$, and the figure of merit is the average error rate $p_{\mathrm{err}} = p_0 p(1|0) + p_1 p(0|1)$, where $p_x$ is the prior probability for input $x$. Optimal MESD achieves a minimal error rate given by the Helstrom bound $p_{\mathrm{err}} = \frac{1}{2}(1 - \sqrt{1 - 4p_0 p_1 \delta^2})$ [176]. Thus, errors are unavoidable for non-orthogonal states.

Errors can be suppressed at the cost of a non-zero rate of inconclusive outcomes. In unambiguous state discrimination (USD), the error probabilities are strictly zero, $p(0|1) = p(1|0) = 0$, and the average inconclusive rate $p_{\mathrm{inc}} = p_0 p(\emptyset|0) + p_1 p(\emptyset|1)$ can be taken as the figure of merit. For unbiased inputs, $p_0 = p_1 = \frac{1}{2}$, optimal USD achieves $p_{\mathrm{inc}} = \delta$ [29]. In the case of qubits,

USD is possible only for two pure states.

Maximum confidence state discrimination (MCSD) generalises the notions of MESD and USD [101]. The *confidence* $C_x$ is the probability that, given an outcome $b = x$, the input was $x$. From Bayes' theorem

$$C_x = \frac{p_x}{\eta_x} p(x|x), \tag{6.1}$$

where $\eta_b = \sum_x p(b|x)p_x$ is the rate of outcome $b$ (i.e. the marginal distribution of the output). In MCSD, the figure of merit is a given $C_x$, or any convex combination of them, and the goal is to maximise this quantity. When $C_x = 1$, the input $x$ is unambiguously identified. Hence, unambiguous discrimination is a particular case of MCSD, and if no further constraints are imposed, MCSD recovers USD whenever the latter is possible. This is the case for an arbitrary number of linearly independent pure states, and thus in particular always for two distinct pure states, as considered here. MESD can also be recovered by adopting $\eta_0 C_0 + \eta_1 C_1 = 1 - p_{\text{err}}$ as the figure of merit, when the inconclusive rates are zero [29]. In general, MCSD is flexible and can handle situations in which both error rates and inconclusive rates are nonzero.

## 6.4 Noncontextual state discrimination

We now proceed to consider noncontextual state discrimination. We start from an ontological model of the prepare-and-measure scenario [27, 161]. The system is associated with an ontic state space $\Lambda$ in which each point $\lambda$ completely defines all physical properties, i.e. the outcomes of all possible measurements. Each state preparation $x$ samples the ontic state space according to a probability distribution $\mu_x(\lambda)$, referred to as the epistemic state. Each measurement is defined by a set of response functions, that is, non-negative functions $\xi_b(\lambda)$ over the ontic space, such that $\sum_b \xi_b(\lambda) = 1$ for all $\lambda \in \Lambda$. The probability of obtaining the outcome $b$ when state $\mu_x$ was prepared is

$$p(b|x) = \int_\Lambda d\lambda \mu_x(\lambda)\xi_b(\lambda). \tag{6.2}$$

While distinct ontic states can be perfectly discriminated, epistemic states with overlapping distributions cannot. It is the discrimination of epistemic states which we compare against quantum state discrimination.

To compare the two requires a notion analogous to the quantum state overlap. Note that $\delta^2 = |\langle\psi_0|\psi_1\rangle|^2$ can be thought of as the probability

that an outcome corresponding to projection onto $|\psi_1\rangle$ occurs when $|\psi_0\rangle$ was prepared (or vice versa). Similarly, in the ontological model we define sharp outcomes as outcomes that are certain to occur for a given preparation. $\xi_b$ is a sharp outcome for $\mu_x$ if $p(b|x) = 1$. For discrimination of $\mu_0$ and $\mu_1$, the *confusability* $c_{0,1}$ is then the probability that a sharp outcome for $\mu_1$ occurs when $\mu_0$ was prepared. For preparation-noncontextual models, that we now introduce, one has the same symmetry as in the quantum case $c_{0,1} = c_{1,0} = c$, and the models can be compared for $c = \delta^2$.

Two preparation procedures are said to be operationally equivalent if they cannot be distinguished by any measurement, and the ontological model is said to be *preparation noncontextual* if all operationally equivalent preparations are represented by the same epistemic state. We take preparation noncontextuality as our notion of classicality and refer to it simply as noncontextuality. We impose two requirements on the noncontextual model. First, it reproduces the observed distribution $p(b|x)$. Second, we need an operational equivalence to which noncontextuality can be applied. We take the model to reproduce the existence of complementary states $|\psi_x^\perp\rangle$, with $|\psi_x\rangle\langle\psi_x| + |\psi_x^\perp\rangle\langle\psi_x^\perp| = \mathbb{1}$ and $|\langle\psi_0^\perp|\psi_1^\perp\rangle| = \delta$. That is, in addition to the epistemic states $\mu_0$, $\mu_1$, it must also contain two states $\mu_0^\perp$, $\mu_1^\perp$ such that their confusability is $c$, they obey $\mu_x\mu_x^\perp = 0$, and the convex combinations $\frac{1}{2}\mu_x + \frac{1}{2}\mu_x^\perp$ for $x = 0, 1$ correspond to operationally equivalent preparations. By noncontextuality they must hence be equal $\frac{1}{2}\mu_0 + \frac{1}{2}\mu_0^\perp = \frac{1}{2}\mu_1 + \frac{1}{2}\mu_1^\perp$. It was shown by Schmid and Spekkens, under similar assumptions, that quantum mechanics outperforms noncontextual theory for MESD in the sense that the Helstrom bound is lower than the minimum achievable error rate in the noncontextual model for any value of $c$ [27]. In Ref. [25], we study quantum vs. noncontextual maximum-confidence discrimination.

## 6.5 Semi-device independent randomness certification

The prepare-and-measure state-discrimination setup can be exploited for semi-DI randomness certification by taking $c$ as given while the devices are otherwise uncharacterised (the states and measurements are unknown), and then assess the randomness of $b$ based on the observed distribution $p(b|x)$. Intuitively, if $p(b|x)$ is close to optimal discrimination for the given $c$, this constrains the measurements to be close to the optimal ones, and the predictability of $b$ to someone with perfect knowledge of the states and measurements can

be estimated. More precisely, we introduce a hidden variable $\omega$, distributed according to $q(\omega)$, labelling measurement strategies. The average guessing probability for an eavesdropper with access to $\omega$ and the input $x$

$$p_g = \sum_x p_x \sum_\omega q(\omega) \max_b p(b|x,\omega), \tag{6.3}$$

with $p(b|x,\omega)$ given by $\mathrm{Tr}[\rho_x \hat{\pi}_b^\omega]$ when the eavesdropper is quantum and by (6.2) with response function $\xi_b^\omega$ if the eavesdropper is restricted to be noncontextual. Note that $\omega$ is assumed to be independent of $x$ (otherwise the discrimination problem becomes trivial). We quantify the randomness by the min-entropy $H_{min} = -\log_2 p_g$, which gives the number of (almost) uniformly random bits which can be extracted per round of the protocol [177].

### 6.5.1 Quantum guessig probability

Since the measurement strategies are unknown to the user, to certify randomness $p_g$ must be upper-bounded by optimising over all strategies compatible with the observed data. We focus on MCSD for the input $x = 0$ and impose only that the rate $\eta_0$ and the confidence $C_0$ are reproduced (as opposed to the full distribution $p(b|x)$. For a quantum eavesdropper, $p_g \leq p_g^Q$ with

$$p_g^Q = \max_{q(\omega),\Pi_b^\omega} \sum_{x,\omega} p_x q(\omega) \max_b \mathrm{Tr}[\rho_x \hat{\pi}_b^\omega], \tag{6.4}$$

subject to $q(\omega)$ and $\hat{\pi}_b^\omega$ being valid probability distributions and POVMs respectively, $\sum_{x,\omega} q(\omega) p_x \,\mathrm{Tr}[\rho_x \hat{\pi}_0^\omega] = \eta_0$ and $\sum_\omega q(\omega) p_0 \,\mathrm{Tr}[\rho_0 \hat{\pi}_0^\omega] = \eta_0 C_0$. Without loss of generality, the states can be fixed to any pair of states with overlap $\delta$. Thus $p_g^Q$ is a function only of the confusability $c$ and the distribution $p(b|x)$. The optimisation problem in (6.4) can be rendered as a semidefinite program, as we show in [178].

### 6.5.2 Noncontextual guessing probability

Similarly, the guessing probability for a noncontextual eavesdropper is bounded by $p_g \leq p_g^{NC}$ with

$$p_g^{NC} = \max_{q(\omega),M_b^\omega} \sum_{x,\omega} p_x q(\omega) \max_b \int_T d\lambda \mu_x(\lambda) \xi_b^\omega(\lambda), \tag{6.5}$$

where now $\xi_b^\omega$ must be valid response functions, and the constraints are the same as in the quantum case with the Born rule replaced by (6.2).

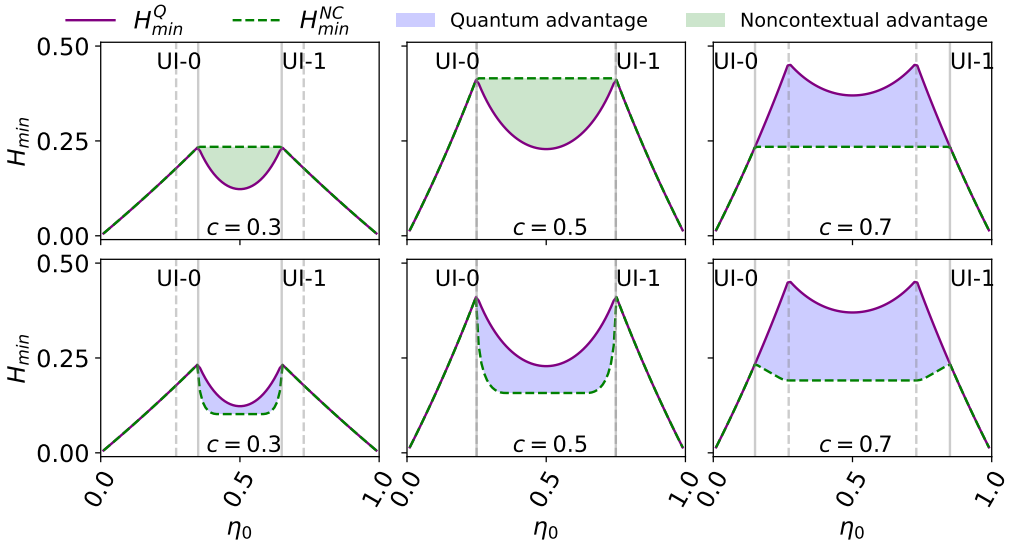



Figure 6.2: Figure extracted from Ref. [17]. Quantum $H^Q_{min}$ and noncontextual $H^{NC}_{min}$ certifiable min-entropies vs. output rate $\eta_0$, for three different confusabilities $c$, optimal confidence $C_0$ and equal prior probabilities $p_0 = p_1 = \frac{1}{2}$. Solid vertical lines delimit parameter regions in which input $x$ is unambiguously identified, labelled UI-$x$. Dashed vertical lines indicate rates at which $H^Q_{min}$ is maximal. The confidences are maximal in all plots. Top row: eavesdroppers in quantum and noncontextual models are respectively quantum and noncontextual. Bottom row: a quantum eavesdropper is considered in both cases.

In a noncontextual theory, a pair of epistemic states must be equal on the overlap of their supports [20, 27]. This allows a general response function to be decomposed into four extremal functions, corresponding to integrals over the regions defined by the overlapping supports of $\mu_0$, $\mu_1$ and their non-overlapping partners. These integrals are, furthermore, functions of the confusability $c$. Using this, in [178] we show that (6.5) can also be rendered as a semidefinite program.

## 6.6 Results

In Fig. 6.2, we compare the certifiable quantum and noncontextual min-entropies, $H^Q_{min}$ and $H^{NC}_{min}$, in two different manners, focusing on equal prior probabilities $p_0 = p_1 = \frac{1}{2}$ for simplicity. First, we compute the certifiable $H_{min}$ within each theory (top row), i.e., $H^Q_{min}$ when the device attains the

maximum quantum confidence and the eavesdropper is also quantum, and $H_{min}^{NC}$ for maximum noncontextual confidence and a noncontextual eavesdropper. This is the maximal certifiable randomness in each theory, as $H_{min}$ is maximised for optimal discrimination. Second, we consider the case in which the eavesdropper is always quantum (bottom row). That is, the minimum entropy is computed via the quantum SDP. Since quantum MCSD can reach higher confidences than noncontextual MCSD, $C_0$ is not necessarily the same in the two cases. In [178] we went beyond the study of pure states by studying the case where noisy (mixed) states $\rho'_x = (1-r)\rho_x + r\mathbb{1}/2$ are prepared. Distinguishablity is still bounded by $c$, and the eavesdropper has no access to decompositions of the mixture. The qualitative behaviour in this setting is similar and thus our main conclusions remain valid.

In the first case we find quantum-over-noncontextual as well as noncontextual-over-quantum advantages in terms of certifying randomness. Whenever any of the states is unambiguously identified by the measurement device, the quantum and noncontextual certifiable randomness are equal, $H_{min}^{Q} = H_{min}^{NC}$. Outside these regions, for confusabilities $c < 1/2$ there is a noncontextual advantage, while for $c > 1/2$ a quantum advantage appears and eventually dominates for large $c$. We interpret this as follows. A quantum eavesdropper is more powerful than a noncontextual one, but optimal quantum discrimination also imposes stronger constraints on the measurement device. For states that are easy to discriminate (low $c$), the former effect wins while for states that a hard to distinguish (high $c$), the second effect dominates. Note that a noncontextual advantage appears only in a universe where the eavesdropper is noncontextual, but does not have access to the ontic state.

In the second case, the eavesdropper is quantum in both models, i.e., we allow the eavesdropper in the noncontextual setting more power. As may be expected, quantum devices are then always at least as powerful as noncontextual ones, with a quantum-over-noncontextual advantage appearing for all values of $c$ whenever none of the inputs are unambigiously identified.

The maximal quantum advantage in terms of generating unpredictable (random) measurement outputs for a quantum eavesdropper is plotted against the confusability in Fig. 6.3. The quantum advantage is largest for nearly indistinguishable states (similar to what was found in Ref. [179]). The eavesdropper's available strategies become more constrained when the optimal confidence has to be reproduced. In a noncontextual scenario, the constraint on the eavesdropper's strategies grows weaker for both nearly distinguishable and indistinguishable states.

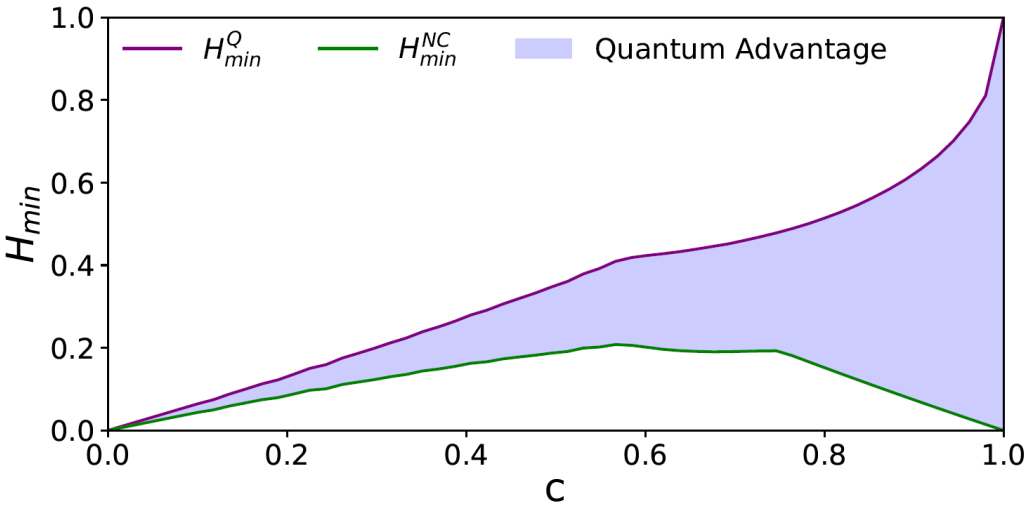



Figure 6.3: Figure extracted from Ref. [17]. Minimum entropy corresponding to the output rates with maximal quantum advantage, for quantum and noncontextual discrimination schemes and a quantum eavesdropper.

## 6.7 Conclusion

In conclusion, we have computed the amount of randomness which can be semi-device-independently certified in maximum-confidence state discrimination setups in both quantum and preparation-noncontextual models. We have derived the maximal randomness within each model, and we find a quantum advantage for MSD-based randomness generation against quantum adversaries. When the adversary in the noncontextual setting is constrained to be noncontextual as well, we find a quantum advantage when the prepared states are difficult to distinguish, but a noncontextual advantage when they are easy to distinguish. In the future, it would be interesting to extend these results to settings with more than two inputs, where more randomness can potentially be generated, and to mixed-state preparations, where correlations between the prepared states and the eavesdropper potentially need to be taken into account.

## 6.8 S1: SDP derivation for quantum randomness certification

In this part of the supplemental material, we show how that the average guessing probability for a quantum eavesdropper can be rendered as a semidefinite

program (SDP). Our derivation closely follows [68].

The objective function (guessing probability) is nonlinear in the variables $q(\omega)$, $\hat{\pi}_b^\omega$, contains a maximisation over $b$, and the number of strategies $\omega$ is a priory unbounded. The latter two issues can be resolved, following [180], by noting that all strategies for which the max occurs for the same $b$ for given $x$ can be lumped together. Hence, only $|b|^{|x|} = 3^2 = 9$ strategies are required. We label each strategies by $(\omega_0, \omega_1)$ where $\omega_x \in \{0, 1, \emptyset\}$ indicates the optimal $b$ given $x$. Thus

$$p_g^Q = \max_{q(\omega_0,\omega_1),\hat{\pi}_b^{\omega_0,\omega_1}} \sum_{x,\omega} p_x q(\omega) \operatorname{Tr}[\rho_x \hat{\pi}_{\omega_x}^{\omega_0,\omega_1}], \tag{6.6}$$

where the distribution over strategies and the POVM elements fulfill

$$\sum_{\omega_0,\omega_1} q(\omega_0, \omega_1) = 1, \tag{6.7}$$

$$q(\omega_0, \omega_1) \geq 0 \ \forall \omega_0, \omega_1, \tag{6.8}$$

$$\hat{\pi}_b^{\omega_0\omega_1} = (\hat{\pi}_b^{\omega_0\omega_1})^\dagger \ \forall \omega_0, \omega_1, \tag{6.9}$$

$$\hat{\pi}_b^{\omega_0\omega_1} \geq 0 \ \forall \omega_0, \omega_1, b, \tag{6.10}$$

$$\sum_b \hat{\pi}_b^{\omega_0\omega_1} = \mathbb{1} \ \forall \omega_0, \omega_1, \tag{6.11}$$

and the observed output rate $\eta_0$ and confidence $C_0^Q$ should be reproduced

$$\sum_{\omega_0,\omega_1} \sum_x p_x q(\omega_0, \omega_1) \operatorname{Tr}\left[\hat{\pi}_0^{\omega_0\omega_1} \rho_x\right] = \eta_0 \tag{6.12}$$

$$\sum_{\omega_0,\omega_1} \frac{p_0}{\eta_0} q(\omega_0, \omega_1) \operatorname{Tr}\left[\hat{\pi}_0^{\omega_0\omega_1} \rho_0\right] = C_0^Q \ . \tag{6.13}$$

Next, $p_g^Q$ and the constraints can be linearised by defining new optimisation variables $\hat{M}_b^{\omega_0\omega_1} = q(\omega_0, \omega_1)\hat{\pi}_b^{\omega_0\omega_1}$. The primal version of the SDP can then

be written:

$$
\begin{aligned}
&\underset{\hat{M}_b^{\omega_0\omega_1}}{\text{maximise}} \qquad p_g^Q = \sum_{x=0}^{1} \sum_{\omega_0,\omega_1} p_x \mathrm{Tr}\left[\hat{M}_{\omega_x}^{\omega_0\omega_1}\rho_x\right] \\
&\text{subject to :} \\
&\hat{M}_b^{\omega_0\omega_1} \geq 0,\ \left(\hat{M}_b^{\omega_0\omega_1}\right)^\dagger = \hat{M}_b^{\omega_0\omega_1},\ \forall \omega_0,\omega_1,b \\
&\sum_b \hat{M}_b^{\omega_0\omega_1} = \frac{1}{2}\mathrm{Tr}\left[\sum_b \hat{M}_b^{\omega_0\omega_1}\right]\mathbb{1},\ \forall \omega_0,\omega_1 \\
&\sum_b \sum_{\omega_0,\omega_1} \sum_x p_x \mathrm{Tr}\left[\hat{M}_b^{\omega_0\omega_1}\rho_x\right] = 1 \\
&\sum_{\omega_0,\omega_1} \sum_x p_x \mathrm{Tr}\left[\hat{M}_0^{\omega_0\omega_1}\rho_x\right] = \eta_0 \\
&\sum_{\omega_0,\omega_1} \frac{p_0}{\eta_0}\mathrm{Tr}\left[\hat{M}_0^{\omega_0\omega_1}\rho_0\right] = C_0^Q\ .
\end{aligned}
\tag{6.14}
$$

The last two constraints can be reduced to

$$
\sum_{\omega_0,\omega_1} p_x \mathrm{Tr}\left[\hat{M}_0^{\omega_0\omega_1}\rho_x\right] = \ \eta_0 C_0^Q \delta_{x,0} + \eta_0\left(1 - C_0^Q\right)\delta_{x,1}\ ,
$$

and normalisation implies

$$
\sum_b \sum_{\omega_0,\omega_1} \sum_x \mathrm{Tr}\left[\hat{M}_b^{\omega_0\omega_1}\rho_x\right] = 2\ . \tag{6.15}
$$

Further on, we formulate the dual version of the problem. From each primal constraint in (6.14), we introduce the dual variables $\hat{G}_b^{\omega_0\omega 1}$, $\hat{H}^{\omega_0\omega 1}$, $\nu_x$

and $\chi$. The corresponding Lagrangian will then be

$$\begin{aligned}\mathcal{L} = &\sum_x \sum_{\omega_0,\omega_1} p_x \mathrm{Tr}\left[\rho_x \hat{M}_{\omega_x}^{\omega_0\omega_1}\right] + \sum_b \sum_{\omega_0,\omega_1} \mathrm{Tr}\left[\hat{G}_b^{\omega_0\omega_1} \hat{M}_b^{\omega_0\omega_1}\right] \\ &+ \sum_{\omega_0,\omega_1} \mathrm{Tr}\left[\hat{H}^{\omega_0\omega_1} \sum_b \left(\hat{M}_b^{\omega_0\omega_1} - \frac{1}{2}\mathrm{Tr}\left[\hat{M}_b^{\omega_0\omega_1}\right]\mathbb{1}\right)\right] \\ &+ \sum_x \nu_x \left(\sum_{\omega_0,\omega_1} p_x \mathrm{Tr}\left[\rho_x \hat{M}_0^{\omega_0\omega_1}\right] - \eta_0\left(\delta_{x,0} C_0^Q + \delta_{x,1}\left(1 - C_0^Q\right)\right)\right) \\ &+ \chi\left(\sum_b \sum_{\omega_0,\omega_1} \sum_x p_x \mathrm{Tr}\left[\rho_x \hat{M}_b^{\omega_0\omega_1}\right] - 1\right) \ .\end{aligned} \tag{6.16}$$

Let us now introduce the supremum of the Lagrangian,

$$\mathcal{S} \equiv \sup_{\hat{M}_b^{\omega_0\omega_1}} \mathcal{L} \ . \tag{6.17}$$

Given any solution $\hat{M}_b^{\omega_0\omega_1}$ of the primal, the last three terms in (6.16) vanish. Thus, as $\hat{M}_b^{\omega_0\omega_1}$ are constrained to be positive semi-definite, the first line in (6.16) yields an upper bound on the guessing probability $p_g^Q$ (only if all $\hat{G}_b^{\omega_0\omega_1}$ are positive semi-definite). The dual can then be formulated by minimising the supremum in (6.17). We re-write it as follows:

$$\begin{aligned}\mathcal{S} = \sup_{\hat{M}_b^{\omega_0\omega_1}} &\sum_{\omega_0,\omega_1} \sum_b \mathrm{Tr}\left[\hat{M}_b^{\omega_0\omega_1} \hat{K}_b^{\omega_0\omega_1}\right] \\ &- \sum_x \nu_x \eta_0 \left(\delta_{x,0} C_0^Q + \delta_{x,1}\left(1 - C_0^Q\right)\right) - \chi,\end{aligned} \tag{6.18}$$

where,

$$\begin{aligned}\hat{K}_b^{\omega_0\omega_1} = &\sum_x p_x \rho_x \left(\delta_{b,\omega_x} + \nu_x \delta_{b,0} + \chi\right) \\ &+ \hat{G}_b^{\omega_0\omega_1} + \hat{H}^{\omega_0\omega_1} - \frac{1}{2}\mathrm{Tr}\left[\hat{H}^{\omega_0\omega_1}\right]\mathbb{1} \ .\end{aligned} \tag{6.19}$$

Now the supremum in (6.18) will diverge, unless $\hat{K}_b^{\omega_0\omega_1} = 0$. We will drop $\hat{G}_b^{\omega_0\omega_1}$, imposing that the remaining expression is negative. This way, the guessing probability can be upper bounded by:

$$p_g \leq p_g^Q = -\sum_{x=0}^{1} \nu_x \eta_0 \left(\delta_{x,0} C_0^Q + \delta_{x,1}\left(1 - C_0^Q\right)\right) - \chi \tag{6.20}$$

for a given value of confidence $C_0$ in discriminating $\rho_1$ and any numbers $\nu_x$ and $\chi$ which fulfil that there exists nine $2 \times 2$ hermitian matrices $\hat{H}^{\omega_0\omega_1}$, with indices $\omega_0, \omega_1 = 0, 1, \emptyset$, such that:

$$\sum_{x=0}^{1} p_x \rho_x \left(\delta_{b,\omega_x} + \nu_x \delta_{b,0} + \chi\right) + \hat{H}^{\omega_0\omega_1} - \frac{1}{2}\mathrm{Tr}\left[\hat{H}^{\omega_0\omega_1}\right] \mathbb{1} \leq 0 \ . \tag{6.21}$$

## 6.9 S2: Matrix notation for noncontextual theory

In this part of the supplemental material, we provide a formalisation of noncontextual state discrimination, paving the way for the comparison with the quantum model.

### 6.9.1 Ontic space division and noncontextuality

The observed data in state discrimination problems are the conditional input-output probabilities. In a noncontextual framework, these take the a form analogous to the Born rule in quantum mechanics. In order to simplify the noncontextual optimisation problem, our goal now is to split up the integral over four different regions, as sketched in Fig. 6.4.

For each epistemic state ($\mu_x(\lambda)$) we define its complementary epistemic state($\mu_x^{\perp}(\lambda)$) which fulfil the orthogonality relation, $\mu_x(\lambda) \cdot \mu_x^{\perp}(\lambda) = 0$. Also, the preparation noncontextuality assumption implies that each pair of complementary epistemic states sum to the *maximally mixed state* ($\mu_{\frac{\mathbb{1}}{2}}(\lambda)$):

$$\frac{1}{2}\mu_0(\lambda) + \frac{1}{2}\mu_0^{\perp}(\lambda) = \frac{1}{2}\mu_1(\lambda) + \frac{1}{2}\mu_1^{\perp}(\lambda) = \mu_{\frac{\mathbb{1}}{2}}(\lambda) \ . \tag{6.22}$$

The maximally mixed state is introduced within noncontextual models analogously to the quantum maximally mixed state [27].

Let us divide the ontic space in four regions on the ontic space (Fig. 6.4). In each region at least two epistemic states will overlap. For example, $\mu_0(\lambda)$ and $\mu_1(\lambda)$ will overlap if $\lambda \in \Lambda_{01}$; or $\mu_0(\lambda)$ and $\mu_1^{\perp}(\lambda)$ overlap if $\lambda \in \Lambda_0$. On the region where two epistemic states overlap they are equal, due to the noncontextuality assumption. Thus, since $\mu_x(\lambda)$ and $\mu_x^{\perp}(\lambda)$ have disjoint supports:

$$\mu_0(\lambda) = \mu_0^{\perp}(\lambda) = \mu_1(\lambda) = \mu_1^{\perp}(\lambda) = 2\mu_{\frac{\mathbb{1}}{2}}(\lambda) \ . \tag{6.23}$$

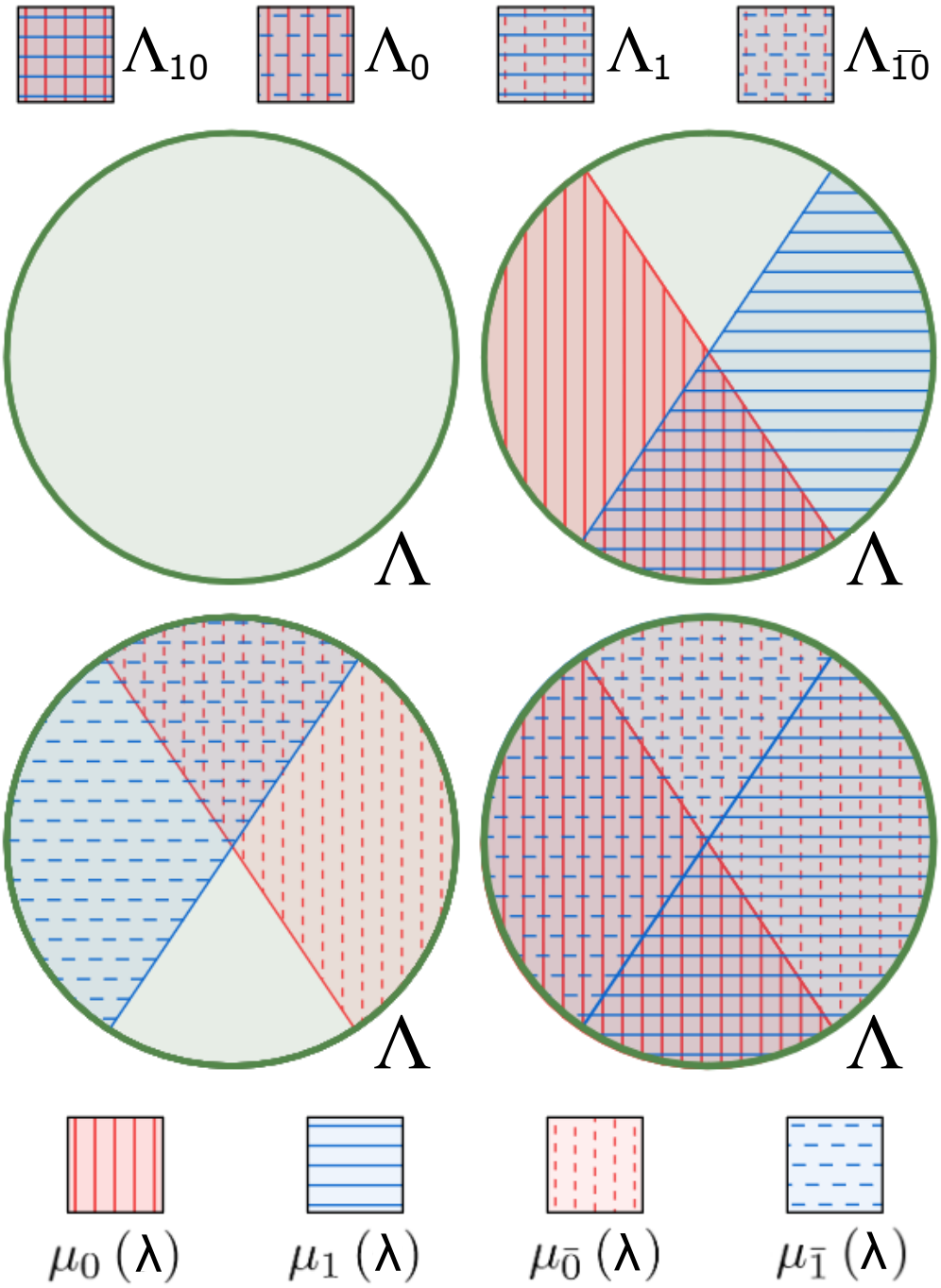



Figure 6.4: Figure extracted from Ref. [17]. Regions on the ontic space ($\Lambda$) according to the overlap of a couple of epistemic states. On the second circle starting form the left, the supports of $\mu_0(\lambda)$ and $\mu_1(\lambda)$ are drawn. On the third, the supports of $\mu_{\bar{0}}^{\perp}(\lambda)$ and $\mu_{\bar{1}}^{\perp}(\lambda)$ are sketched. On the last circle, the supports of all states are drawn.

### 6.9.2 Noncontextual matrix notation

When optimising the noncontextual guessing probability, it is not be convenient to work directly with the response functions $\xi_b(\lambda)$ and epistemic states $\mu_x(\lambda)$. We can reduce the problem to depend on a finite number of real optimisation variables. Let us introduce the following quantities based on integrating the response functions over the regions of the ontic space previously defined:

$$\begin{aligned}
\alpha_{0b} &= \frac{1}{1-c}\int_{\Lambda_0} d\lambda \xi_b(\lambda)\mu_0(\lambda) \underset{\text{n.c.}}{=} \frac{1}{1-c}\int_{\Lambda_0} d\lambda \xi_b(\lambda)\mu_1^{\perp}(\lambda) \\
\alpha_{1b} &= \frac{1}{1-c}\int_{\Lambda_1} d\lambda \xi_b(\lambda)\mu_1(\lambda) \underset{\text{n.c.}}{=} \frac{1}{1-c}\int_{\Lambda_1} d\lambda \xi_b(\lambda)\mu_0^{\perp}(\lambda) \\
\beta_b &= \frac{1}{c}\int_{\Lambda_{10}} d\lambda \xi_b(\lambda)\mu_0(\lambda) \underset{\text{n.c.}}{=} \frac{1}{c}\int_{\Lambda_{10}} d\lambda \xi_b(\lambda)\mu_1(\lambda) \\
\bar{\beta}_b &= \frac{1}{c}\int_{\Lambda_{\bar{1}\bar{0}}} d\lambda \xi_b(\lambda)\mu_0^{\perp}(\lambda) \underset{\text{n.c.}}{=} \frac{1}{c}\int_{\Lambda_{\bar{1}\bar{0}}} d\lambda \xi_b(\lambda)\mu_1^{\perp}(\lambda) \ .
\end{aligned} \tag{6.24}$$

The second equality in each row of (6.24) is fulfilled when preparation noncontextuality is fulfilled, i.e. (6.23). In fact, we can express these terms in a more compact form

$$\begin{aligned}
\alpha_{xb} &= \frac{2}{1-c}\int_{\Lambda_x} d\lambda \xi_b(\lambda)\mu_{\frac{1}{2}}(\lambda), \\
\beta_b &= \frac{2}{c}\int_{\Lambda_{10}} d\lambda \xi_b(\lambda)\mu_{\frac{1}{2}}(\lambda), \\
\bar{\beta}_b &= \frac{2}{c}\int_{\Lambda_{\bar{1}\bar{0}}} d\lambda \xi_b(\lambda)\mu_{\frac{1}{2}}(\lambda) \ .
\end{aligned} \tag{6.25}$$

Probabilities can then be written in terms of these quantities as

$$p(b|x) = \alpha_{xb}(1-c) + \beta_b c \ . \tag{6.26}$$

It is then sufficient to consider the value of the integration of the response functions times the maximally mixed state, over the regions we introduced, to solve the noncontextual state discrimination problem.

Pushing this notation further, we propose a matrix structure which collects the notion of the divisions of the ontic space in Fig. 6.4. Each epistemic state

will be represented by a $2 \times 2$ matrix, $\hat{\mu}_x$, and each term will represent the definite integral over the different regions on the ontic space, as

$$\begin{aligned} \hat{\mu}_x^T &\equiv \begin{pmatrix} \int_{\Lambda_0} d\lambda \mu_x(\lambda) & \int_{\Lambda_{10}} d\lambda \mu_x(\lambda) \\ \int_{\Lambda_{\bar{1}\bar{0}}} d\lambda \mu_x(\lambda) & \int_{\Lambda_1} d\lambda \mu_x(\lambda) \end{pmatrix} \\ &= \begin{pmatrix} \left(\delta_{x,0} + \delta_{x,\bar{1}}\right)(1-c) & \left(\delta_{x,0} + \delta_{x,1}\right) c \\ \left(\delta_{x,\bar{0}} + \delta_{x,\bar{1}}\right) c & \left(\delta_{x,\bar{0}} + \delta_{x,1}\right)(1-c) \end{pmatrix} , \end{aligned} \tag{6.27}$$

for $\delta_{i,j}$ being the Kronecker delta. It is convenient to define the transpose of the matrix form of the epistemic state to ease the notation later on. The orthogonality relation between the complementary and the prepared epistemic states becomes $\hat{\mu}_x \circ \hat{\mu}_x^{\perp} = 0$, where $\circ$ is the element-wise matrix product, commonly known as *Hadamard product*, and the right-hand side is the zero matrix.

We can use the quantities introduced in (6.25) to write down the matrix form of the response functions, as

$$\hat{\xi}_b \equiv \begin{pmatrix} \alpha_{0b} & \beta_b \\ \bar{\beta}_b & \alpha_{1b} \end{pmatrix} . \tag{6.28}$$

The input-output conditional probabilities can thus be written with the form

$$p(b|M,x) = \sum_{ij}^{N} \left[\hat{\xi}_b \circ \hat{\mu}_x^T\right]_{ij} = \mathrm{Tr}\left[\hat{\xi}_b \hat{\mu}_x\right] . \tag{6.29}$$

The first equality can be derived straight from (6.26), by summing up all the terms from the Hadamard product between the response function and epistemic state. The second equality holds for any pair of $N \times N$ matrices, relating the Hadamard product with the usual matrix product. The result from (6.29) allows us to write the input-output probabilities on a form similar to the Born rule in quantum mechanics.

As mentioned in [27], the noncontextual model we are building must reproduce some results from the quantum theory. In the present case, we are interested in two state discrimination. Thus, all statistics from discriminating any two states $|\psi_0\rangle$ and $|\psi_1\rangle$ (together with their orthogonal counterparts)

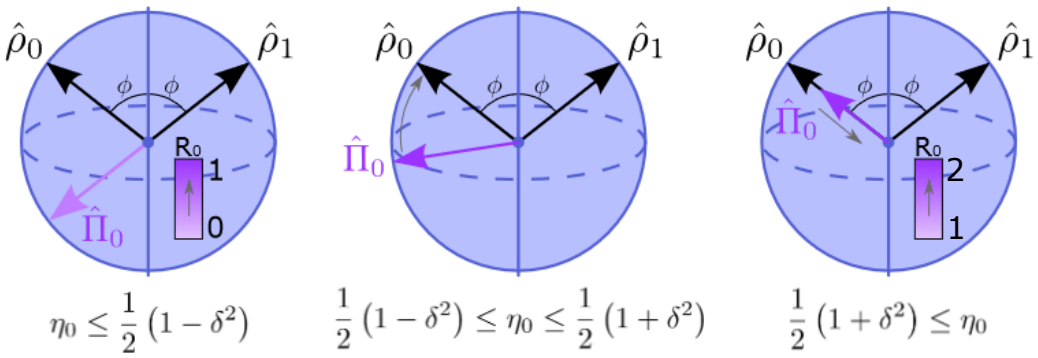



Figure 6.5: Figure extracted from Ref. [17]. Representation of the POVM element $\hat{\pi}_0$ and the quantum states $\rho_0$ and $\rho_1$ on the Bloch sphere for different values of the output rate $\eta_0$.

must be reproducible by the response functions. This condition puts a constraint on the components of $\hat{\xi}_b$. Concretely, $\alpha_{0b} + \alpha_{1b} = \beta_b + \bar{\beta}_b$ shall be fulfilled.

## 6.10 S3: Rates at which states can be unambiguously identified

In this part of the supplemental material, we derive analytical expressions for the output rates at which optimal MCSD unambiguously identifies one of the inputs. We will look separately at the quantum and noncontextual scenarios.

### 6.10.1 Quantum case

We look at the discrimination of two pure states $\rho_0 = |\psi_0\rangle\langle\psi_0|$ and $\rho_1 = |\psi_1\rangle\langle\psi_1|$ randomly prepared with equal probability. The POVM element corresponding to outcome $b$ can be represented on the Bloch sphere as

$$\hat{\pi}_b = \frac{R_b}{2}\left[\mathbb{1} + r_b \sin\theta_b X + r_b \cos\theta_b Z\right] \ , \tag{6.30}$$

where $X$ and $Z$ are the Pauli matrices in (2.6) and $(r_b \sin\theta_b, 0, r_b \cos\theta_b)$ is the Bloch vector (on the X-Z plane), with $|r_b| \leq 1$, $R_b \geq 0$ [181]. Positivity and normalisation imply that

$$\sum_b R_b = 2, \ \sum_b R_b r_b \sin\theta_b = \sum_b R_b r_b \cos\theta_b = 0 \ . \tag{6.31}$$

In Fig. 6.5 we show the POVM element $\hat{\pi}_0$ in the Bloch sphere, together with the quantum states

$$\begin{aligned} \rho_0 &= \frac{1}{2}\left[\mathbb{1} - \sin\phi\, X + \cos\phi\, Z\right], \\ \rho_1 &= \frac{1}{2}\left[\mathbb{1} + \sin\phi\, X + \cos\phi\, Z\right] \ . \end{aligned} \tag{6.32}$$

The overlap is given by $\cos\phi = \delta$. Assuming equal prior probabilities ($p_0 = p_1 = 1/2$), the confidence $C_0$ expressed in terms of $\hat{\pi}_0$ is

$$C_0 = \frac{\mathrm{Tr}\left[\hat{\pi}_0\rho_0\right]}{\mathrm{Tr}\left[\hat{\pi}_0\rho_0\right] + \mathrm{Tr}\left[\hat{\pi}_0\rho_1\right]}. \tag{6.33}$$

The expression in (6.33) is the figure of merit in MCSD. Without loss of generality, we can focus on the POVM element $\hat{\pi}_0$, and consider $\hat{\pi}_1 = \hat{\pi}_\emptyset$.

The maximum value of the confidence ($C_0 = 1$) can be obtained if the measurement device is able to unambiguously discriminate the state $\rho_0$, i.e. $\mathrm{Tr}\left[\hat{\pi}_0\rho_1\right] = 0$. This implies that $\theta_0 = \phi + \pi$ and $r_0 = 1$. The POVM element $\hat{\pi}_0$ has rank 1, and we are left with $0 \leq R_0 \leq 1$. The only possible output rates are

$$0 \leq \eta_0 \leq \frac{1}{2}\left(1 - \delta^2\right) \ . \tag{6.34}$$

For higher rates, we need to allow $\hat{\pi}_0$ to rotate. To keep $C_0$ as large as possible, we need to make sure that the numerator is also at its maximum. Thus, our goal now is to find the maximum value of $p(0|0)$. That is achieved by rotating $\hat{\pi}_0$ towards $\rho_0$. The rotation angle $\theta_0$ can be parametrized in terms of the output rate as, $\cos\theta_0 = (2\eta_0 - 1)/\delta$. This will run within the interval $\phi + \pi \leq \theta \leq 2\pi - \phi$. For the output rate, this means

$$\frac{1}{2}\left(1 - \delta^2\right) \leq \eta_0 \leq \frac{1}{2}\left(1 + \delta^2\right) \ . \tag{6.35}$$

Beyond that point, the output rate saturates when the POVM element $\hat{\pi}_0$ is no longer projective. Thus, $r_0$ will be reduced to zero and $R_0$ will increase from 1 to 2. For the output rate this means

$$\frac{1}{2}\left(1 + \delta^2\right) \leq \eta_0 \leq 1 \ . \tag{6.36}$$

Here state $\rho_1$ can be unambiguously discriminated, i.e. $\mathrm{Tr}\left[\hat{\pi}_1\rho_0\right] = 0$, as $\theta_1 = \theta_\emptyset = \pi - \phi$ because of (6.31).

### 6.10.2 Noncontextual case

In the noncontextual framework, we use (6.25), and the probabilities (6.26). Then, the confidence can be written as

$$C_0 = \frac{(\alpha_{00}\,(1-c) + \beta_0 c)}{(\alpha_{00} + \alpha_{10})\,(1-c) + 2\beta_0 c}\ . \tag{6.37}$$

The maximal value on the confidence ($C_0 = 1$) is achieved when $\alpha_{10} = \beta_0 = 0$. Since $0 \leq \alpha_{00} \leq 1$, this occurs for rates

$$0 \leq \eta_0 \leq \frac{1}{2}\,(1-c)\ . \tag{6.38}$$

For larger rates, we need $\beta_0$ to grow. We can keep $\alpha_{10} = 0$ since it only appears in the denominator. Again, since $0 \leq \beta_0 \leq 1$, the rates at which this is possible are

$$\frac{1}{2}\,(1-c) \leq \eta_0 \leq \frac{1}{2}\,(1+c)\ . \tag{6.39}$$

Finally, for even larger $\eta_0$ we need $\alpha_{10}$ to grow. As $0 \leq \alpha_{10} \leq 1$, we are left with

$$\frac{1}{2}\,(1+c) \leq \eta_0 \leq 1\ . \tag{6.40}$$

## 6.11 S4: SDP for noncontextual randomness certification

In this appendix, we show that the average guessing probability for a noncontextual eavesdropper can be cast as an SDP, similarly to the quantum case. Similarly as in the first section of the additional information, the number of relevant strategies is again 9, labeled by $\omega_x \in \{0, 1, \emptyset\}$ for $x \in \{0, 1\}$. These distributions over strategies and the response functions fulfill

$$\sum_{\omega_0,\omega_1} q(\omega_0, \omega_1) = 1, \tag{6.41}$$

$$q(\omega_0, \omega_1) \geq 0\ \forall \omega_0, \omega_1, \tag{6.42}$$

$$\sum_b \xi_b^{\omega_0\omega_1}(\lambda) = 1\ \forall \omega_0, \omega_1, \lambda, \tag{6.43}$$

$$\xi_b^{\omega_0\omega_1}(\lambda) \geq 0\ \forall \omega_0, \omega_1, b, \lambda, \tag{6.44}$$

and the observed output rate $\eta_0$ and confidence $C_0^{NC}$ should be reproduced,

$$\sum_{\omega_0,\omega_1}\sum_x p_x q(\omega_0,\omega_1)\int d\lambda \xi_0^{\omega_0\omega_1}(\lambda)\mu_x(\lambda) = \eta_0 \tag{6.45}$$

$$\sum_{\omega_0,\omega_1}\frac{p_0}{\eta_0} q(\omega_0,\omega_1)\int d\lambda \xi_0^{\omega_0\omega_1}(\lambda)\mu_0(\lambda) = C_0^Q\ . \tag{6.46}$$

Finally, $p_g^{NC}$ can be linearised by defining $M_b^{\omega_0\omega_1}(\lambda) = q(\omega_0,\omega_1)\xi_b^{\omega_0\omega_1}(\lambda)$. The primal version of the SDP can then be written as follows:

$$\begin{aligned}
&\underset{M_b^{\omega_0\omega_1}(\lambda)}{\text{maximise}} \qquad p_g^{NC} = \sum_{x=0}^{1}\sum_{\omega_0,\omega_1} p_x \int d\lambda M_{\omega_x}^{\omega_0\omega_1}(\lambda)\mu_x(\lambda)\\
&\text{subject to :}\\
&M_b^{\omega_0\omega_1}(\lambda) \geq 0,\ \forall \omega_0,\omega_1,b,\lambda\\
&\sum_b M_b^{\omega_0\omega_1}(\lambda) = \frac{1}{|\Lambda|}\int d\lambda \sum_b M_b^{\omega_0\omega_1}(\lambda),\ \forall \omega_0,\omega_1\\
&\sum_b\sum_{\omega_0,\omega_1}\sum_x p_x \int d\lambda M_b^{\omega_0\omega_1}(\lambda)\mu_x(\lambda) = 1\\
&\sum_{\omega_0,\omega_1}\sum_x p_x \int d\lambda M_0^{\omega_0\omega_1}(\lambda)\mu_x(\lambda) = \eta_0\\
&\sum_{\omega_0,\omega_1}\frac{p_0}{\eta_0}\int d\lambda M_0^{\omega_0\omega_1}(\lambda)\mu_0(\lambda) = C_0^{NC}\ .
\end{aligned} \tag{6.47}$$

The last two constraints can be reduced to:

$$\sum_{\omega_0,\omega_1} p_x \int d\lambda M_0^{\omega_0\omega_1}(\lambda)\mu_x(\lambda) = \eta_0 C_0^{NC}\delta_{x,0} + \eta_0\left(1 - C_0^{NC}\right)\delta_{x,1}\ . \tag{6.48}$$

The explicit use of functions over the ontic space as optimisation variables and the presence of integrals makes the SDP impractical to solve. To avoid

these issues, we introduce the quantities from (6.25) and define:

$$\begin{aligned} A_{xb}^{\omega_0\omega_1} &= q(\omega_0,\omega_1)\alpha_{xb} = \frac{2}{1-c}\int_{\Lambda_x} d\lambda M_b^{\omega_0\omega_1}(\lambda)\,\mu_{\frac{\mathbb{1}}{2}}(\lambda)\,, \\ B_b^{\omega_0\omega_1} &= q(\omega_0,\omega_1)\beta_b = \frac{2}{c}\int_{\Lambda_{10}} d\lambda M_b^{\omega_0\omega_1}(\lambda)\,\mu_{\frac{\mathbb{1}}{2}}(\lambda)\,, \\ \bar{B}_b^{\omega_0\omega_1} &= q(\omega_0,\omega_1)\bar{\beta}_b = \frac{2}{c}\int_{\Lambda_{\bar{1}\bar{0}}} d\lambda M_b^{\omega_0\omega_1}(\lambda)\,\mu_{\frac{\mathbb{1}}{2}}(\lambda)\ . \end{aligned} \tag{6.49}$$

We can now re-write the primal problem in (6.47) with the quantities in (6.49). Since this process in not trivial, we go through it step by step.

- Guessing probability:

$$\begin{aligned} p_g^{NC} &= \sum_x \sum_{\omega_0,\omega_1} p_x \left[\int_{\Lambda_0} d\lambda M_{\omega_x}^{\omega_0\omega_1}(\lambda)\,\mu_x(\lambda) + \int_{\Lambda_{10}} d\lambda M_{\omega_x}^{\omega_0\omega_1}(\lambda)\,\mu_x(\lambda)\right. \\ &\left. + \int_{\Lambda_1} d\lambda M_{\omega_x}^{\omega_0\omega_1}(\lambda)\,\mu_x(\lambda) + \int_{\Lambda_{\bar{1}\bar{0}}} d\lambda M_{\omega_x}^{\omega_0\omega_1}(\lambda)\,\mu_x(\lambda)\right] \\ &= \sum_x \sum_{\omega_0,\omega_1} p_x \left[(1-c)\,A_{x\omega_x}^{\omega_0\omega_1} + cB_{\omega_x}^{\omega_0\omega_1}\right] \end{aligned} \tag{6.50}$$

- Non-negativity constraint:

$$M_b^{\omega_0\omega_1}(\lambda) \geq 0 \Leftrightarrow \begin{cases} A_{xb}^{\omega_0\omega_1} \geq 0 \\ B_b^{\omega_0\omega_1} \geq 0 \quad \forall\omega_0,\omega_1,x,b, \\ \bar{B}_b^{\omega_0\omega_1} \geq 0 \end{cases} \tag{6.51}$$

- Ontic state independence from $q(\omega_0,\omega_1)$:

$$\begin{aligned} \sum_b A_{xb}^{\omega_0\omega_1} &= \frac{1}{1-c}\int_{\Lambda_x} d\lambda \sum_b M_b^{\omega_0\omega_1}(\lambda)\,\mu_x(\lambda) \\ &= \sum_b M_b^{\omega_0\omega_1}(\lambda')\,\frac{1}{1-c}\int_{\Lambda_x} d\lambda\mu_x(\lambda) = \sum_b M_b^{\omega_0\omega_1}(\lambda')\ . \end{aligned} \tag{6.52}$$

Also:

$$\sum_b B_b^{\omega_0\omega_1} = \sum_b \bar{B}_b^{\omega_0\omega_1} = \sum_b M_b^{\omega_0\omega_1}(\lambda')\ . \tag{6.53}$$

On the other hand:

$$\begin{aligned}
&\sum_b \left[(1-c)\left(A_{0b}^{\omega_0\omega_1}+A_{1b}^{\omega_0\omega_1}\right)c\left(B_b^{\omega_0\omega_1}+\bar{B}_b^{\omega_0\omega_1}\right)\right] \\
=&\sum_b\left[\int_{\Lambda_0} d\lambda M_b^{\omega_0\omega_1}(\lambda)\,\mu_{\frac{1}{2}}(\lambda)+\int_{\Lambda_1} d\lambda M_b^{\omega_0\omega_1}(\lambda)\,\mu_{\frac{1}{2}}(\lambda)\right. \\
&\left.+\int_{\Lambda_{10}} d\lambda M_b^{\omega_0\omega_1}(\lambda)\,\mu_{\frac{1}{2}}(\lambda)+\int_{\Lambda_{\bar{1}\bar{0}}} d\lambda M_b^{\omega_0\omega_1}(\lambda)\,\mu_{\frac{1}{2}}(\lambda)\right] \\
=&\sum_b M_b^{\omega_0\omega_1}(\lambda')\int_\Lambda d\lambda 2\mu_{\frac{1}{2}}(\lambda)=2\sum_b M_b^{\omega_0\omega_1}(\lambda')
\end{aligned} \tag{6.54}$$

Thus, combining (6.52), (6.53) and (6.54) one ends up with:

$$\begin{aligned}
&\sum_b A_{xb}^{\omega_0\omega_1}=\sum_b B_b^{\omega_0\omega_1}=\sum_b \bar{B}_b^{\omega_0\omega_1} \\
&=\sum_b\left[\frac{1-c}{2}\left(A_{0b}^{\omega_0\omega_1}+A_{1b}^{\omega_0\omega_1}\right)+\frac{c}{2}\left(B_b^{\omega_0\omega_1}+\bar{B}_b^{\omega_0\omega_1}\right)\right]
\end{aligned} \tag{6.55}$$

- Reproduce the output rates:

$$\eta_b=\sum_{\omega_0\omega_1}\sum_x p_x \tag{6.56}$$

$$\begin{aligned}
&\left[\int_{\Lambda_0} d\lambda M_b^{\omega_0\omega_1}(\lambda)\,\mu_x(\lambda)+\int_{\Lambda_{10}} d\lambda M_b^{\omega_0\omega_1}(\lambda)\,\mu_x(\lambda)\right. \\
&\left.+\int_{\Lambda_1} d\lambda M_b^{\omega_0\omega_1}(\lambda)\,\mu_x(\lambda)+\int_{\Lambda_{\bar{1}\bar{0}}} d\lambda M_b^{\omega_0\omega_1}(\lambda)\,\mu_x(\lambda)\right] \\
&=\sum_{\omega_0\omega_1}\sum_x p_x\left[(1-c)\,A_{xb}^{\omega_0\omega_1}+cB_b^{\omega_0\omega_1}\right]\ .
\end{aligned} \tag{6.57}$$

- Normalisation of the output rates:

$$\begin{aligned}
&\sum_b\sum_{\omega_0\omega_1}\sum_x p_x\left[\int_{\Lambda_0} d\lambda M_b^{\omega_0\omega_1}(\lambda)\,\mu_x(\lambda)+\int_{\Lambda_{10}} d\lambda M_b^{\omega_0\omega_1}(\lambda)\,\mu_x(\lambda)\right. \\
&\qquad\left.+\int_{\Lambda_1} d\lambda M_b^{\omega_0\omega_1}(\lambda)\,\mu_x(\lambda)+\int_{\Lambda_{\bar{1}\bar{0}}} d\lambda M_b^{\omega_0\omega_1}(\lambda)\,\mu_x(\lambda)\right] \\
=&\sum_b\sum_{\omega_0\omega_1}\sum_x p_x\left[(1-c)\,A_{xb}^{\omega_0\omega_1}+cB_b^{\omega_0\omega_1}\right]=1\ .
\end{aligned} \tag{6.58}$$

- Fix the confidence $C_0$ of the measurement device:

$$\begin{aligned}&\sum_{\omega_0\omega_1}\frac{p_0}{\eta_0}\int_\Lambda d\lambda M_0^{\omega_0\omega_1}(\lambda)\,\mu_0(\lambda)\\&=\sum_{\omega_0\omega_1}\frac{p_0}{\eta_0}\left[(1-c)\,A_{0b}^{\omega_0\omega_1}+cB_b^{\omega_0\omega_1}\right]=C_0^{NC}\ .\end{aligned}\tag{6.59}$$

At the end of the day, we can write the primal problem as follows:

$$\begin{aligned}&\underset{\left\{A_{xb}^{\omega_0\omega_1},B_b^{\omega_0\omega_1},\bar{B}_b^{\omega_0\omega_1}\right\}}{\text{maximise}}\quad p_g^{NC}=\sum_x\sum_{\omega_0,\omega_1}p_x\left[(1-c)\,A_{x\omega_x}^{\omega_0\omega_1}+cB_{\omega_x}^{\omega_0\omega_1}\right]\\&\text{subject to :}\qquad A_{xb}^{\omega_0\omega_1}\geq 0,B_b^{\omega_0\omega_1}\geq 0,\bar{B}_b^{\omega_0\omega_1}\geq 0\ \forall\omega_0,\omega_1,b\\&\sum_b A_{0b}^{\omega_0\omega_1}=\sum_b A_{1b}^{\omega_0\omega_1}=\sum_b B_b^{\omega_0\omega_1}=\sum_b\bar{B}_b^{\omega_0\omega_1}=\\&=\sum_b\left[\frac{1-c}{2}\left(A_{0b}^{\omega_0\omega_1}+A_{1b}^{\omega_0\omega_1}\right)+\frac{c}{2}\left(B_b^{\omega_0\omega_1}+\bar{B}_b^{\omega_0\omega_1}\right)\right]\\&\sum_b\sum_{\omega_0\omega_1}\sum_x p_x\left[(1-c)\,A_{xb}^{\omega_0\omega_1}+cB_b^{\omega_0\omega_1}\right]=1\\&\sum_{\omega_0\omega_1}\sum_x p_x\left[(1-c)\,A_{x0}^{\omega_0\omega_1}+cB_0^{\omega_0\omega_1}\right]=\eta_0\\&\sum_{\omega_0\omega_1}\frac{p_0}{\eta_0}\left[(1-c)\,A_{0b}^{\omega_0\omega_1}+cB_b^{\omega_0\omega_1}\right]=C_0^{NC}\ .\end{aligned}\tag{6.60}$$

Using the matrix form of the response functions introduced in (6.28), we define:

$$\hat{M}_b^{\omega_0\omega_1}\equiv q(\omega_0,\omega_1)\hat{\xi}_b^{\omega_0\omega_1}=\begin{pmatrix}A_{1b}^{\omega_0\omega_1} & B_b^{\omega_0\omega_1}\\ \bar{B}_b^{\omega_0\omega_1} & A_{2b}^{\omega_0\omega_1}\end{pmatrix}\ .\tag{6.61}$$

Implementing this matrix notation, together with the matrix form of the

epistemic states in (6.27), we re-write the primal problem in (6.60) as follows:

$$
\begin{aligned}
&\underset{\hat{M}_b^{\omega_0\omega_1}}{\text{maximise}} \qquad p_g^{NC} = \sum_x \sum_{\omega_0,\omega_1} p_x \text{Tr}\left[\hat{M}_{\omega_x}^{\omega_0\omega_1}\hat{\mu}_x\right] \\
&\text{subject to}: \quad \hat{M}_b^{\omega_0\omega_1} \underset{\text{e.w.}}{\geq} 0 \;\forall \omega_0,\omega_1,b \\
&\sum_b \hat{M}_b^{\omega_0\omega_1} = \text{Tr}\left[\sum_b \hat{M}_b^{\omega_0\omega_1}\hat{\mu}_{\frac{\mathbb{1}}{2}}\right]\hat{J}_2 \\
&\sum_b \sum_{\omega_0,\omega_1} \sum_x p_x \text{Tr}\left[\hat{M}_b^{\omega_0\omega_1}\hat{\mu}_x\right] = 1 \\
&\sum_{\omega_0,\omega_1} \sum_x p_x \text{Tr}\left[\hat{M}_0^{\omega_0\omega_1}\hat{\mu}_x\right] = \eta_0 \\
&\sum_{\omega_0,\omega_1} \frac{p_0}{\eta_0} \text{Tr}\left[\hat{M}_0^{\omega_0\omega_1}\hat{\mu}_0\right] = C_0^Q \;.
\end{aligned}
\tag{6.62}
$$

Here, $\hat{J}_2$ denotes a $2 \times 2$ matrix with all entries equal to 1. Also, $\underset{\text{e.w.}}{\geq}$ denotes element-wise inequalities between matrices, and the maximally mixed state in noncontextual theory has been introduced in the matrix notation. It is given by

$$
\hat{\mu}_{\frac{\mathbb{1}}{2}} \equiv \frac{1}{2}\begin{pmatrix} 1-c & c \\ c & 1-c \end{pmatrix} \;. \tag{6.63}
$$

Finally, note that both last constraints in (6.62) can be re-written as:

$$
\sum_{\omega_0,\omega_1} p_x \text{Tr}\left[\hat{M}_0^{\omega_0\omega_1}\hat{\mu}_x^T\right] = C_0^{NC}\delta_{x,0} + \left(1 - C_0^{NC}\right)\delta_{x,1} \;. \tag{6.64}
$$

Also, due to normalization:

$$
\sum_b \sum_{\omega_0,\omega_1} \sum_x \text{Tr}\left[\hat{M}_b^{\omega_0\omega_1}\hat{\mu}_x\right] = 2 \;. \tag{6.65}
$$

We proceed obtaining the dual problem in the noncontextual framework. From each constraint in (6.62), we introduce the dual variables: $\hat{G}_b^{\omega_0\omega 1}$, $\hat{H}^{\omega_0\omega 1}$,

$\nu_x$ and $\chi$. The corresponding Lagrangian will then be:

$$\begin{aligned}\mathcal{L} =& \sum_x \sum_{\omega_0,\omega_1} p_x \mathrm{Tr}\left[\hat{\mu}_x \hat{M}_{\omega_x}^{\omega_0\omega_1}\right] + \sum_b \sum_{\omega_1,\omega_2} \mathrm{Tr}\left[\hat{G}_b^{\omega_0\omega_1} \hat{M}_b^{\omega_0\omega_1}\right] \\ &+ \sum_{\omega_0,\omega_1} \mathrm{Tr}\left[\hat{H}^{\omega_0\omega_1} \sum_b \left(\hat{M}_b^{\omega_0\omega_1} - \mathrm{Tr}\left[\hat{M}_b^{\omega_0\omega_1} \hat{\mu}_{\frac{1}{2}}\right] \hat{J}_2\right)\right] \\ &+ \sum_x \nu_x \left(\sum_{\omega_0,\omega_1} p_x \mathrm{Tr}\left[\hat{\mu}_x \hat{M}_0^{\omega_0\omega_1}\right] - \eta_0 \left(\delta_{x,0} C_0^{NC} + \delta_{x,1}\left(1 - C_0^{NC}\right)\right)\right) \\ &+ \chi \left(\sum_b \sum_{\omega_0,\omega_1} \sum_x p_x \mathrm{Tr}\left[\hat{\mu}_x \hat{M}_b^{\omega_1\omega_2}\right] - 1\right) \ . \end{aligned} \tag{6.66}$$

We write the supremum of the Lagrangian as:

$$\mathcal{S} \equiv \sup_{\hat{M}_b^{\omega_0\omega_1}} \mathcal{L} \ . \tag{6.67}$$

Given any solution $\hat{M}_b^{\omega_0\omega_1}$ of the primal, the last three terms in (6.66) vanish. Thus, as $\hat{M}_b^{\omega_0\omega_1}$ are constrained to be positive semi-definite, the first line in (6.16) yields an upper bound on the guessing probability $p_g^{NC}$ (only if all $\hat{G}_b^{\omega_0\omega_1}$ are positive semi-definite). The dual can then be formulated by minimising the supremum in (6.67). We re-write it as follows:

$$\begin{aligned}\mathcal{S} = \sup_{\hat{M}_b^{\omega_0\omega_1}} & \sum_b \sum_{\omega_0,\omega_1} \mathrm{Tr}\left[\hat{M}_b^{\omega_0\omega_1} \hat{K}_b^{\omega_0\omega_1}\right] \\ & - \sum_x \nu_x \eta_0 \left(\delta_{x,0} C_0^{NC} + \delta_{x,1}\left(1 - C_0^{NC}\right)\right) - \chi, \end{aligned} \tag{6.68}$$

where,

$$\begin{aligned}\hat{K}_b^{\omega_0\omega_1} =& \sum_x p_x \hat{\mu}_x \left(\delta_{b,\omega_x} + \nu_x \delta_{b,0} + \chi\right) \\ &+ \hat{G}_b^{\omega_0\omega_1} + \hat{H}^{\omega_0\omega_1} - \mathrm{Tr}\left[\hat{H}^{\omega_0\omega_1} \hat{J}_2\right] \hat{\mu}_{\frac{1}{2}} \ . \end{aligned} \tag{6.69}$$

The supremum in (6.68) will diverge, unless $\hat{K}_b^{\omega_0\omega_1} = 0$. We will drop $\hat{G}_b^{\omega_0\omega_1}$, imposing that the remaining expression is negative. This way, the guessing probability can be upper bounded by:

$$p_g \leq p_g^{NC} = -\sum_{x=0}^{1} \nu_x \eta_0 \left(\delta_{x,0} C_0^{NC} + \delta_{x,1}\left(1 - C_0^{NC}\right)\right) - \chi \tag{6.70}$$

for a given value of confidence $C_0$ in discriminating $\rho_1$ and any numbers $\nu_x$ and $\chi$ which fulfil that there exists nine $2 \times 2$ matrices $\hat{H}^{\omega_0\omega_1}$, with indices $\omega_0, \omega_1 = 0, 1, \emptyset$, such that:

$$\sum_{x=0}^{1} p_x \hat{\mu}_x \left(\delta_{b,\omega_x} + \nu_x \delta_{b,0} + \chi\right) + \hat{H}^{\omega_0\omega_1} - \mathrm{Tr}\left[\hat{H}^{\omega_0\omega_1} \hat{J}_2\right] \hat{\mu}_{\frac{\mathbb{1}}{2}} \leq 0 \ . \tag{6.71}$$

As a final remark, note that one can straightforwardly switch between quantum and noncontextual models by switching: the bound on the confidence ($C_0^Q \leftrightarrow C_0^{NC}$); the physical state representations ($\rho_x \leftrightarrow \hat{\mu}_x$); the measurement outcome representation ($\hat{\pi}_b^\omega \leftrightarrow \hat{\xi}_b^\omega$); the identity element ($\mathbb{1} \leftrightarrow \hat{J}_2$); the maximally mixed state ($\frac{1}{2}\mathbb{1} \leftrightarrow \hat{\mu}_{\frac{\mathbb{1}}{2}}$); and the positive (negative) semi-definite matrix constraints with the non-negativity (negativity) element-wise restriction ($\geq (\leq) \leftrightarrow \underset{\text{e.w.}}{\geq} \ (\underset{\text{e.w.}}{\leq})$).

## 6.12 S5: Min-entropies for pure and noisy states

We denote the prepared states as an ensemble of pure states ($\rho_x = |\psi_x\rangle \langle\psi_x|$ for $x \in \{0, 1\}$) with distinguishability bounded by $c = |\langle\psi_0|\psi_1\rangle|^2$. These pure states are then mixed with white noise with probability $r$, and result into the following noisy states

$$\rho'_x = r\rho_x + (1 - r)\frac{1}{2}\mathbb{1} \ . \tag{6.72}$$

The confidence in discriminating $\rho'_0$ is then bounded by $C_0'^Q$. It follows that, compared to the pure state bound $C_0^Q$, the noisy states bound on the confidence is strictly $C_0'^Q \leq C_0^Q$, being equal only if $r = 1$.

One can consider the model in the following manner. In some rounds the distinguishability of the prepared states is bounded with respect to Eve and in some other rounds white noise is prepared. She knows whether the pure state $\rho_x$ or white noise $\mathbb{1}/2$ is sent to the measurement device. In this manner, she is as able to discriminate $\hat{\rho_x}'$ as if only the pure states $\rho_x$ where prepared, with probability $r$. Only the observed confidence, therefore, will differ when comparing pure and noisy states in the randomness certification process.

Similarly, in the noncontextual model the following noisy epistemic states are prepared

$$\mu'_x(\lambda) = r\mu_x(\lambda) + (1 - r)\mu_{\frac{\mathbb{1}}{2}}(\lambda) \ . \tag{6.73}$$

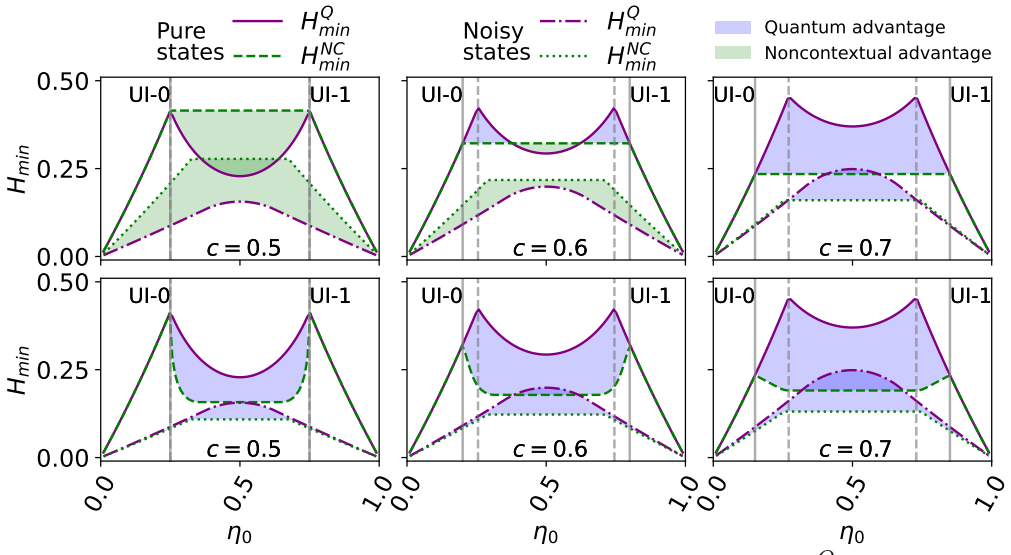



Figure 6.6: Figure extracted from Ref. [17]. Quantum $H^Q_{min}$ and noncontextual $H^{NC}_{min}$ certifiable min-entropies vs. output rate $\eta_0$, for three different confusabilities $c$, optimal confidence $C_0$, equal prior probabilities $p_0 = p_1 = \frac{1}{2}$, pure ($r = 1$) and noisy ($r = 0.7$) states. Solid vertical lines delimit parameter regions in which input $x$ is unambiguously identified, labelled UI-$x$. Dashed vertical lines indicate rates at which $H^Q_{min}$ is maximal. The confidences are maximal in all plots. Top row: eavesdroppers in quantum and noncontextual models are respectively quantum and noncontextual. Bottom row: a quantum eavesdropper is considered in both cases.

The confusability $c$ between epistemic states $\mu_0(\lambda)$ and $\mu_1(\lambda)$ is fixed. The confidence in discriminating $\mu'_0(\lambda)$ is then bounded by $C_0'^{NC}$, which is strictly lower than or equal to the bound on the confidence in discriminating $\mu_0(\lambda)$, namely $C_0^{NC}$. As in the quantum case, the eavesdropper knows whether the epistemic states ($\mu_x(\lambda)$) with bounded confusability or the maximally mixed state ($\mu_{\frac{1}{2}}(\lambda)$) are prepared.

Results on the min-entropies are shown in Fig. 6.6. We look at the cases where pure states ($r = 1$) and noisy states with $r = 0.7$ are prepared. In both cases, the min-entropies on the measurement outcomes of noisy states are lower than those of pure states. Remarkably, even with noisy states a quantum advantage is still found whenever the eavesdropper is quantum in both quantum and noncontextual state discrimination schemes.

# Chapter 7

# More than one bit of semi-device independent randomness from a single qubit

In this chapter we present the results in "More than one bit of semi-device independent randomness from a single qubit" [16], authored by Carles Roch i Carceller, Lucas Nunes Faria, Zheng-Hao Liu, Ulrik Lund Andersen, Jonas Schou Neegaard-Nielsen and Jonatan Bohr Brask. A final version of this work is still in preparation.

## 7.1 Abstract

Certified randomness guaranteed to be unpredictable by adversaries is central to information security. The fundamental randomness inherent in quantum physics makes certification possible from devices that are only weakly characterised hence requiring little trust in their implementation. Here, we demonstrate semi-device independent randomness certification from untrusted measurements on the simplest possible system – a single quantum bit – producing more than one bit of certified randomness per round.

## 7.2 Introduction

Randomness is a fundamental aspect in many physical and mathematical systems, and it is often used to model uncertainty or to generate unique unpredictable values in a wide variety of applications. The role of randomness is central in many applications of the modern world. Cryptographic protocols heavily rely on the unpredictability of a secret key between two parties, a

task for which random number generators are essential [1, 3, 182]. Many algorithms, such as in cryptography or prime factorization, also make use of random sequences of numbers to run optimally [2, 4, 183–187]. In all cases, random numbers with high quality are required, which can be provided through quantum random number generators (QRNG).

The inherent randomness of quantum mechanics can be harnessed to generate "true" random numbers [14, 188–190]. Those are essentially different from pseudo-random numbers which rely on the complexity of a deterministic algorithm to be generated [11, 12]. True random numbers can be generated in a simple prepare-and-measure scenario from the outcomes of a fully characterized measurement of a quantum state. These schemes are often known as device-dependent protocols. Albeit unpredictable by nature, one cannot be sure whether there is not some subtle regularity that would let someone else predict them. This problem is more philosophical and goes back to the beginning of quantum mechanics. It was proven by John Bell [15, 18] and later by Kochen and Specker [19] that quantum mechanics is nondeterministic and does not admit a local hidden-variable model [53]. This statement is the principal aspect which allows for protocols where the devices are allowed to be only partially characterized. Measurement-device independent protocols, for instance, are those where the measurement is completely unknown, but is required to reproduce a set of observable statistics. On the other hand, if the state preparations are the completely unknown part, with a fully revealed measurement, the protocol is said to be source-device independent. Fully device independent (DI) protocols rely on even fewer assumptions, with both the state preparation and measurement parts unknown. Randomness through DI protocols, however, can only be certified in multipartite scenarios, where nonlocality enters into play as an essential ingredient [58, 116, 117]. Such protocols are therefore very demanding, out of reach for current technology and only possible in highly sophisticated environments [163–167]. In single-party prepare-and-measure scenarios, one can relax the DI assumption still retaining the randomness certification capabilities to design semi-DI QRNG protocols with less strict experimental requirements [66, 191–193]. One can partially uncover the preparation by bounding a concrete particularity of the prepared states, such as the overlap or the transmitted energy [61, 65, 194].

In this paper we propose a simple QRNG protocol involving the preparation of three qubit states and a single measurement. What sets our work apart from other semi-DI QRNG proposals is that we are able to observe more than one bit of randomness per round where a single qubit state is measured with a surprisingly simple setup. The protocol is semi-DI in the sense that only the

pair-wise overlap of the prepared states is bounded, leaving the measurement completely un-characterised. This framework allows us to target high values of randomness during the rounds when only one of the states is measured. This is possible thanks to the utilization of the rounds where the other two states are prepared to self-test the measurement device. Also, it is central in our protocol that the measurement outcome is at least threefold. The idea is simple: we target preparations and measurements such that observed correlations are only reproducible with extremal and unique strategies. A measurement is said to be extremal if its positive operator-valued measure (POVM) representation can only be trivially expressed through a convex combination of other valid POVMs. These can be found, for instance, in quantum state-discrimination protocols such as minimum error state discrimination (MESD) [97, 108, 155, 195].

The goal of MESD is to minimise the error in detecting a particular state from an ensemble with a single measurement. The minimum attainable error in MESD is given by the Helstrom bound, which is reached through a projective (hence extremal) and unique measurement [77, 98, 99, 176]. Suppose now that one aims to use MESD to certify the randomness on the measurement outcome. In order to evaluate more than one bit of randomness per round, we would need a three state discrimination scenario with a Helstrom measurement with three different outcomes (one for each state). That measurement is described through a POVM in a qutrit space [196]. In our case we aim to evaluate the randomnness in a qubit discrimination setup, thus making MESD not a good candidate for us. We need to choose a measurement strategy with at least three outcomes. A good option is to use unambiguous state discrimination (USD) during the self-test rounds [79–81]. In USD, the error rate is nullified, at the cost of adopting an additional measurement event which gives no information on which state was prepared (formally called inconclusive event). That way, an optimal three outcome measurement can be implemented in a two-qubit state discrimination framework, making it possible to reach more than one bit of randomness per round. In this work, we implement a protocol which uses USD during the self-test rounds as in [68], allowing us to make use of the additional inconclusive event in two-state discrimination to reach randomness values greater than one bit per round. Then, we choose a third state that yields equiprobable outcomes when involved in the state discrimination scenario.

The paper is organised as follows. In the "Results" section we begin introducing the qubit state discrimination scenario and specifying the measurement strategy. Later we explain how we evaluate the randomness and continue

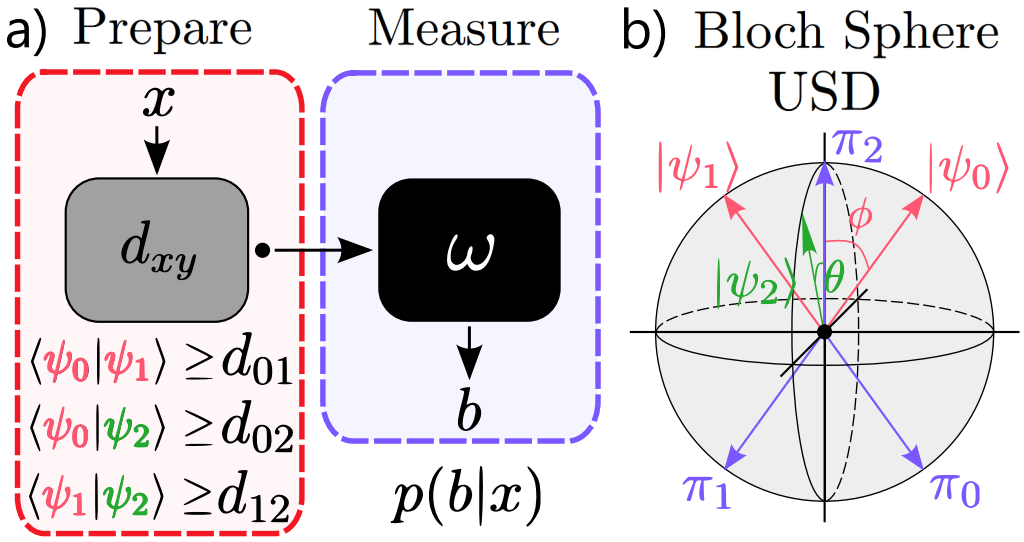



Figure 7.1: Figure extracted from Ref. [16]. **a)** Sketch of the semi-device independent prepare-and-measure scenario with the three preparations and one untrusted measurement (black box). **b)** The implemented USD protocol is illustrated on the Bloch sphere, with the Bloch vectors corresponding to the prepared states $|\psi_x\rangle$ and POVM elements $\hat{\pi}_b$.

with the main semi-device independent assumptions we consider. We end the section by experimentally implementing the protocol with coherent states and presenting the results we observe in the experiment and in simulations. In the following "Discussion" section we explain how we deal with finite size effects assuming independent and identically distributed (i.i.d.) and non-i.i.d. rounds. We end the paper detailing some important particularities involved in the experimental implementation in the "Methods" section.

## 7.3 Results

### 7.3.1 Prepare-and-measure

Consider a device that can receive an input $x \in \{0, 1, 2\}$ with prior probabilities $p_x$ and prepares a pure quantum state $\rho_x = |\psi_x\rangle\langle\psi_x|$. Consider the

following preparations with Bloch representation

$$|\psi_0\rangle = \cos\frac{\phi}{2}|0\rangle + \sin\frac{\phi}{2}|1\rangle \qquad (7.1)$$
$$|\psi_1\rangle = \cos\frac{\phi}{2}|0\rangle - \sin\frac{\phi}{2}|1\rangle \ ,$$

and a third state $|\psi_2\rangle$ with a general qubit structure which shall remain unspecified for now. These states are sent to a second device which will perform a measurement described by the POVM $\{\hat{\pi}_b\}$ for $b \in \{0, 1, 2\}$, as we show in Fig. 7.1. Over many rounds of the experiment, one can gather enough data to build consistent statistics according to the underlying probability distribution $p(b|x) = \mathrm{Tr}\,[\rho_x \hat{\pi}_b]$, according to the Born rule.

Our first task is to find the optimal measurement, according to a particular state discrimination protocol, to identify whether the preparation was $x = 0$ or $x = 1$, ignoring the third possible input $x = 2$. This will yield correlations reproducible only with extremal and unique POVMs. Then, the idea is to properly design a state $|\psi_2\rangle$ such that, with that optimal measurement, all outcomes are equiprobable whenever that state is prepared (i.e. $p(b|x = 2)$ are equal $\forall b$). In order to do that, we take USD as our target strategy.

### 7.3.2 Unambiguous state discrimination

The task in USD is to identify which state was prepared without making any errors, i.e. $p_{\mathrm{err}} = p_0 p(1|0) + p_1 p(0|1) = 0$ [79–81]. That can be done if one pays the price of having some rounds in which the measurement result turns inconclusive. In the present case, USD targets preparations $x = 0$ and $x = 1$. The events that turn inconclusive will be labeled with $b = 2$. The goal of USD is to minimize the rate of inconclusive events, which here we call $p_{\text{ø}} := p_0 p(2|0) + p_1 p(2|1)$. The minimum rate of inconclusive events in two-state discrimination is proportional to the overlap of the prepared states, which is denoted by $|\langle\psi_0|\psi_1\rangle| = \cos\phi$, according to (7.1). In this case, the minimum rate of inconclusive events is lower bounded by

$$p_{\text{ø}} \geq 2\sqrt{p_0 p_1}\cos\phi \ . \qquad (7.2)$$

The POVM that represents an optimal USD measurement must be given by rank-1 POVM elements, proportional to the projectors onto the orthogonal

states (7.1). Concretely, for equiprobable preparations $p_0 = p_1 = 1/2$,

$$\begin{aligned}\hat{\pi}_0 &= \frac{1}{1+\cos\phi}\,|\psi_1^{\perp}\rangle\langle\psi_1^{\perp}| \\ \hat{\pi}_1 &= \frac{1}{1+\cos\phi}\,|\psi_0^{\perp}\rangle\langle\psi_0^{\perp}| \\ \hat{\pi}_2 &= \mathbb{1} - \hat{\pi}_0 - \hat{\pi}_1 \ ,\end{aligned} \tag{7.3}$$

where $\langle\psi_x|\psi_x^{\perp}\rangle = 0$. Consider now we put a third preparation $x = 2$ into play. We aim to find a state $|\psi_2\rangle$ which triggers all three outcomes $b = 0, 1, 2$ of the measurement in (7.3) with the same probability. That means, it must satisfy $\langle\psi_2|\,\hat{\pi}_b\,|\psi_2\rangle = 1/3$. A state that potentially fulfills that condition is given by a state with the form

$$|\psi_2\rangle = \cos\frac{\theta}{2}\,|0\rangle + i\sin\frac{\theta}{2}\,|1\rangle \ . \tag{7.4}$$

Now, all three outcomes will be equiprobably triggered when

$$\cos\theta = \frac{1-2\cos\phi}{3\cos\phi} \ . \tag{7.5}$$

This condition is only valid when $-1 \leq \cos\theta \leq 1$. This means that equiprobable outcomes in this setting are only achievable for $\cos\phi \geq 1/5$. This is only true if the states are constrained to be two-dimensional. In semi-DI randomness certification, one wants to keep the number of assumptions at minimum. We will later show how we can get rid of the assumption of a fixed dimension.

### 7.3.3 Randomness certification

We proceed to explain how we certify the randomness of the measurement outcomes. To do that, we introduce the figure of a third malicious party (Eve) as an adversary that aims to guess the outcome $b$. Eve is given control of the preparation ($x$) and measurement device, and even can hold a state which shares quantum correlations within our setup. We also give her the freedom to change her strategy $\omega$ each round, according to the distribution $q(\omega)$. We certify the amount of randomness in the measurement outcome only when state $|\psi_2\rangle$ is prepared. In this sense, the rounds where states $|\psi_0\rangle$ and $|\psi_1\rangle$ are prepared can be thought as self-test rounds used to build up the observed statistics. During the rounds that the state $|\psi_2\rangle$ is prepared, her best guessing probability averaged through each round can be written as

$$p_g = \sum_{\omega} q(\omega)\max_b \{p(b|x=2,\omega)\} \ , \tag{7.6}$$

for $p(b|x,\omega) = \mathrm{Tr}\left[\rho_x \hat{\pi}_b^\omega\right]$ ($\rho_x = |\psi_x\rangle\langle\psi_x|$) being the probability that the measurement outcome is $b$ given state preparation $x$ and the eavesdropper's strategy $\omega$. Each round of the experiment she can change the measurement strategy $\omega$ that suits her better in order to get the best possible guess of the outcome $b$. That is, as long as she reproduces the observed statistics on average, i.e. $p(b|x) = \sum_\omega q(\omega)\,\mathrm{Tr}\left[\rho_x \hat{\pi}_b^\omega\right]$.

We aim to find an upper bound on $p_g$ in (7.6) for all possible strategies $\omega$, distributions $q(\omega)$ and measurements $\hat{\pi}_b^\omega$. Such optimisation problem can be rendered as a semidefinite program (SDP) [31], as we detail in Sec. 7.8. We further define the dual problem as a semidefinite program which, in turn, allows us to re-write Eve's guessing probability as

$$p_g = -\sum_{b,x} \nu_{bx} p(b|x)\ . \tag{7.7}$$

An upper bound $p_g^* \geq p_g$ can be found by minimizing (7.7) through all possible parameters $\nu_{bx}$ and $2\times 2$ matrices $\hat{H}^\omega$ that fulfil the constraints

$$\rho_2 \delta_{b,\omega} + \sum_x \rho_x \nu_{b,x} + \hat{H}^\omega - \frac{1}{D}\,\mathrm{Tr}\left[\hat{H}^\omega\right]\mathbb{1} \leq 0 \tag{7.8}$$

$$\hat{H}^\omega = \left(\hat{H}^\omega\right)^\dagger\ , \tag{7.9}$$

where $\delta_{x,y}$ is the Kronecker delta and $D$ is the dimension of the eavesdropper. The randomness of the measurement outcomes is quantified through the min-entropy $H_{min} = -\log_2(p_g)$, which gives the number of (almost) uniformly random bits which can be extracted per round of the protocol [177].

### 7.3.4 Semi-device independent certification

To run the semidefinite program one needs to insert specific quantum states. In a semi-device independent approach, however, those should not be completely specified, but only a particularity of the states should be bounded or fixed. For two-state discrimination problems, usually one imposes some dimensionality bounds (which can be translated to energy bounds [61]) and the distinguishability through their overlap. Then, one can specify a couple of states without losing any sense of generality, as unitary rotations on the Bloch sphere will not affect the results. Also, in those cases it is clear that Eve will not gain extra information about the outcome by working with a three-dimensional space. However, that statement does not hold anymore

if we add a third state into play. We want to avoid constraining the dimensionality accessible by the eavesdropper. To do that, suppose that the three prepared states span a three-dimensional space. That is, the third state is

$$|\tilde{\psi}_2\rangle = \sqrt{a}\,|\psi_2\rangle + \sqrt{1-a}\,|2\rangle \ , \tag{7.10}$$

for $|\psi_2\rangle$ in (7.4) having support on the bi-dimensional space spanned by $|\psi_0\rangle$ and $|\psi_1\rangle$ in (7.1), and $|2\rangle$ has only support on an additional orthogonal dimension, such that $\langle\psi_x|2\rangle = 0$, $\forall x$. One can see that, if both overlaps $|\langle\tilde{\psi}_2|\psi_0\rangle|^2$ and $|\langle\tilde{\psi}_2|\psi_1\rangle|^2$ are simultaneously fixed to be

$$|\langle\tilde{\psi}_2|\psi_0\rangle|^2 = |\langle\tilde{\psi}_2|\psi_1\rangle|^2 = \frac{1}{2}\left(1+\cos\phi\cos\theta\right) \ , \tag{7.11}$$

then normalisation in (7.10) imposes $a \Rightarrow 1$, and all three preparations must solely have support on a qubit space. However, strict equalities can be hard to satisfy in the lab. In Sec. 7.6 we show how one can relax this assumption by only bounding the overlaps with some fixed quantities as $|\langle\psi_x|\psi_y\rangle| \geq d_{xy}$. This, in turn, bounds the accessibility to an additional third dimension by the eavesdropper as $a$ turns to be lower-bounded by

$$a \geq \frac{d_{02}^2 + d_{12}^2 - 2d_{01}d_{02}d_{12}}{1-d_{01}^2} \ . \tag{7.12}$$

Also, since $a \leq 1$, one also finds the following relation between the bounds on the overlaps:

$$1 \geq d_{02}^2 + d_{12}^2 + d_{01}^2 - 2d_{01}d_{02}d_{12} \ , \tag{7.13}$$

which must hold true for any dimension. Equation (7.12) defines the surface of a tetrahedron with curved faces (see Sec. 7.6). The amplitude $a$ decreases towards the center of the tetrahedron. Thus, in another perspective, bounding $|\langle\psi_x|\psi_y\rangle| \geq d_{xy}$ implies that we forbid Eve to access the core of the tetrahedron.

Our protocol is semi-device independent in the sense that we do not make any assumption on the devices involved on the experiment, other than only bounding the overlap of the prepared states.

### 7.3.5 From qubit to coherent states

Let $\mathcal{C}_D$ denote the convex set of correlations reproducible by $D$-dimensional states $\rho_x$ and POVM elements $\hat{\pi}_b$. Then, the observed probabilities on the

experiment $p(b|x) = \mathrm{Tr}\left[\hat{\pi}_b \rho_x\right]$ belong to $\mathcal{C}_2$. However, the availability of an additional third dimension to an eavesdropper allows her to access probabilities $p(b|x,\omega) \in \mathcal{C}_3 \supset \mathcal{C}_2$, as long as on average $p(b|x)$ are reproduced. The fact that from our setup we can only constrain correlations in $\mathcal{C}_2$ is what allows us to certify the randomness when qubit states are measured, with unbounded dimensionality to the eavesdropper.

In our model, states are interpreted as qubit preparations according to bounded overlaps. However, the implementation we use in this work makes use of coherent state preparations. To map the theoretical framework to the actual implementation, we collect the overlaps of the prepared coherent states which can be expressed with their corresponding coherent amplitudes. Then, we choose three qubit states $|\psi_x\rangle$ which fulfill the registered overlaps, and a qubit POVM to calculate the observed probabilities $p(b|x)$, which belong to the sub-set of correlations $\mathcal{C}_2$. However, as previously manifested, the relevant Hilbert space in any state discrimination problem is that with dimension equal to the number of involved states. This means that any eavesdropper may benefit from using qutrits and correlations in $\mathcal{C}_3$ to increase the guessing probability. This is accounted for by considering an eavesdropper with access to qutrit states $|\tilde{\psi}_x\rangle$, of the form in (7.10), which satisfy the registered overlaps in the experiment, and qutrit POVMs reproducing, on average, the observed correlations.

### 7.3.6 Implementation

We implement the protocol on a real prepare-and-measure experiment through a very simple optical setup. Our time-bin-encoding method is inspired by previous works with a similar setting [64, 68]. The setup is illustrated in Fig. 7.2; here we briefly sketch its working principle. In the Methods section, we elaborate the detailed experimental setup.

The photon source in our experiment is a 1550 nm continuous-wave laser. We have used an electro-optic modulator (EOM) and an acousto-optic modulator (AOM) to carve the output of the laser into 10 ns-width pulsed coherent states with appropriate amplitude. In each round of the experiment, the coherent states can emerge at an early time-bin and a late time-bin.

The two self-testing states for $x = 0$ and $x = 1$ are prepared with a coherent state with amplitude $\alpha$ encoded in an early and late time-bin respectively, that is, $|\psi_0\rangle = |\alpha 0\rangle$ and $|\psi_1\rangle = |0\alpha\rangle$, where the two entries in the ket denote the amplitude of the coherent state at the two time-bins. On the other hand,

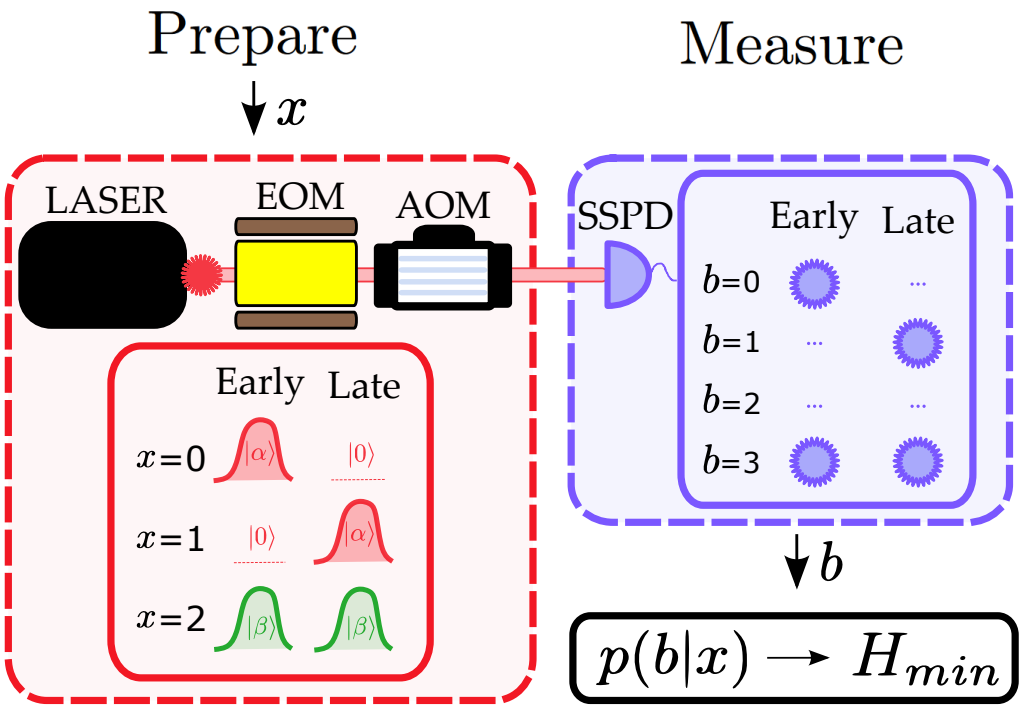



Figure 7.2: Figure extracted from Ref. [16]. Experimental implementation of the protocol. The input states are generated by a sequence of EOM and AOMs. The measurement simply consists of a single-photon detector.

the third state, $x = 2$, used for randomness extraction is prepared with a coherent amplitude $\beta$ in both early and late time-bins: $|\psi_2\rangle = |\beta\beta\rangle$.

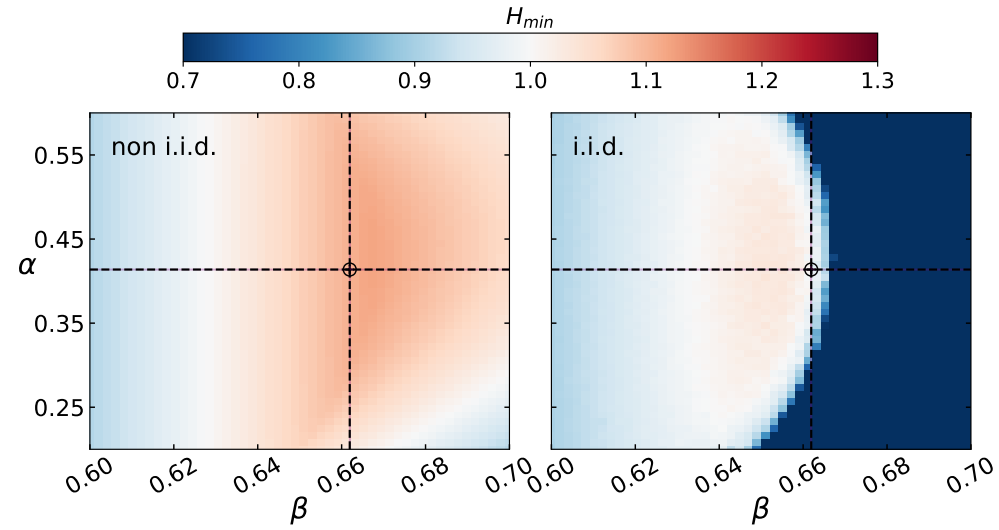
The measurement simply consists of a superconducting single-photon detector (ID Quantique ID281) with a quantum efficiency of 94% which detects photons from the carved, attenuated laser beam. We label the outcome $b$ according to whether the detector clicks on the early or late time bins: we relate the events triggered only on the early time-bin with $b = 0$, those triggered only on the late time-bin with $b = 1$. We label the events where the detector does not click at any time-bin with $b = 2$. Lastly, due to the third state preparation and dark counts we have the possibility that the detector clicks at both early and late time-bins in a single round. We incorporate those events with $b = 3$. In Sec. 7.7 we specify in more detail the different events and their corresponding probabilities with the time-bin coherent state preparations.

### 7.3.7 Simulation and observation

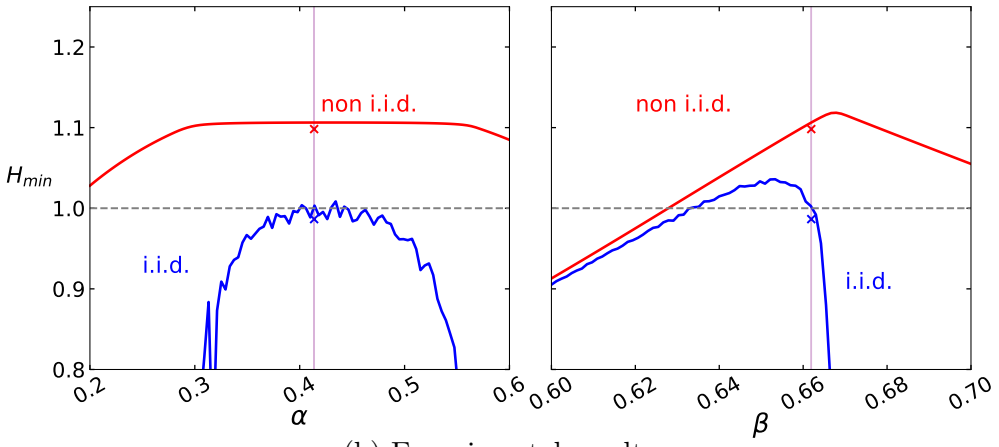
In order to find optimal experimental settings, we first simulate the system by running the SDP in a closed range of coherent amplitudes. In Fig. 7.3 we show the minimum entropy from the SDP under realistic experimental conditions. The results show optimal minimum entropy around the amplitudes $\alpha_T = 0.41$ and $\beta_T = 0.66$ which will be targeted in the experiment. This corresponds to states $|\psi_x\rangle$ with overlaps $|\langle\psi_0|\psi_1\rangle| \simeq 0.84$ and $|\langle\psi_0|\tilde{\psi}_2\rangle| = |\langle\psi_1|\tilde{\psi}_2\rangle| \simeq 0.78$. This value depends on the relative phase between $\alpha$ and $\beta$, which is not controlled in the experiemt. However, it does not matter neither due to the phase-insensitivity of the detector and because the relative phase does not play any role in our protocol. So, we take those as the bounds $d_{xy}$ we consider to certify the randomness. According to the criterion deduced from (7.5), $|\langle\psi_0|\psi_1\rangle| \geq 1/5$, which means that these preparations should allow us to find a set of correlations which yields more than one bit of randomness.

Let us first analyze the effects on the dimensional unboundedness due to the semi-device independent treatment. This is denoted by the minimal amplitude $a$ introduced in (7.10). In principle, the observed probabilities when the third state $|\psi_2\rangle = |\beta\beta\rangle$ is prepared should not depend on $\alpha$, and this should be reflected on the $H_{min}$ as well. However, we see a subtle dependency on $\alpha$ in Fig. 7.3, which becomes more evident for higher values of $\beta$.

The fact that $a < 1$ in the set of sampled amplitudes is not only the main responsible of that dependency. It also makes the observed correlations no




(a) Simulation results.




(b) Experimental results.

Figure 7.3: Figure extracted from Ref. [16]. Bound on the min-entropy after finite-size effects are accounted for assuming independent and identicallu distributed (i.i.d.) rounds (Chernoff-Hoeffding tail inequality) and non i.i.d. rounds (entropy accumulation theorem). **(a)** Result of the SDP for a limited range of coherent amplitudes. The targeted $\alpha = 0.414 \pm 4.38 \cdot 10^{-5}$ and $\beta = 0.662 \pm 3.63 \cdot 10^{-5}$ in the lab are denoted with black-dashed lines. **(b)** Slice of the color plots at the chosen amplitudes. The results of the experiment are showed in a cross with error bars too small to be visible (see main text). The chosen parameters are $\varepsilon = 10^{-9}$ with photon loss $1 - \eta = 0.06$, dark count probability of $p_{\text{dc}} = (4.49 \pm 0.14) \cdot 10^{-6}$ and total number of samples $n = 10^8$.

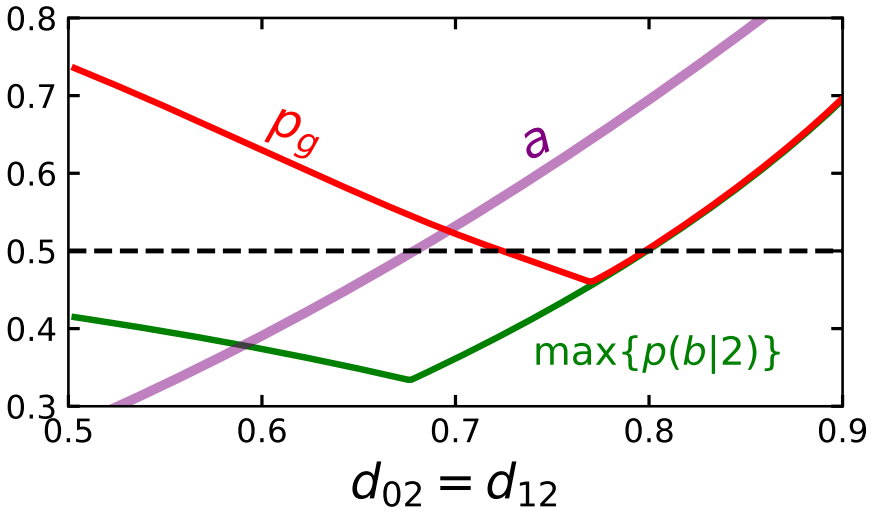



Figure 7.4: Figure extracted from Ref. [16]. Observed correlations $p(b|x)$ and guessing probability $p_g$. Here $\alpha = \alpha_T$. If the dimension were to be constrained, the probabilities would be only reproducible through a unique and extremal POVM. However, $p_g$ deviates from $\max_b \{p(b|x)\}$ as the minimal amplitude $a$ decreases. This evidences the access to an unconstrained third dimension by Eve.

longer uniquely reproducible by unique and extremal POVMs. In Fig. 7.4 we show how the guessing probability deviates from the maximal conditional probability $p(b|2)$, which means that the measurement is reproducible through multiple strategies accessible by Eve. This compromises the secrecy of the outcome, which means a decrease on the expected randomness. Nonetheless, we are able to find a set of amplitudes where, although Eve has unbounded dimensionality, her guessing probability is lower than 1/2.

We run the experiment through the proposed implementation and evaluate the randomness for a single configuration of targeted amplitudes. We show the obtained results in Fig. 7.3 (side and top plots). Our observations agree well with the predictions from the simulations, and show a randomness extraction rate of $1.09817 \pm 0.00012$ and $0.986 \pm 0.008$ bits per round after $10^8$ runs of the experiment, when analyzed using the entropy accumulation theorem (EAT) [197] and considering independent and identically distributed (i.i.d.) rounds, respectively. Here, the error bars (which are too small to be visible in Fig. 7.3) are the $1\sigma$ bootstrap standard deviation obtained via resampling the data.

Although the chosen coherent amplitudes avoid the big drop in randomness from the i.i.d. analysis (blue curve in Fig. 7.3), we are not able to observe more than one bit of randomness per round. This drop is caused by the deviation on the observed correlations due to the Chernoff-Hoeffding tail inequality (see Sec. 7.9). At some point, the dual coefficients $\nu_{bx}$ become so big that small deviations turn into a huge drop in the certifiable randomness. Surprisingly, dealing with finite-size effects through the EAT, without assuming i.i.d. rounds, avoids such drops in the randomness (red curve in Fig. 7.3). It is in this case that we are able to observe more than one bit of randomness per round.

## 7.4 Discussion

We have realized a semi-DI randomness generation protocol that is capable of generating more than one bit of randomness in every round of experiments without assuming fair sampling. Although high-rate randomness generation via multi-outcome POVM could be intuitive, certification of the generated randomness against side information to the adversaries remains challenging. Our contribution in this regard is twofold. First, we used semidefinite programming tools to bound the probability that any potential adversary could guess the measurement outcome in a prepare-and-measure experiment. Second, we have applied the EAT to certify the amount of entropy generated during the experiment from finite-size statistics and without assuming i.i.d. which is often unphysical in a realistic scenario. Our methodology thus provides a pathway for effectively certifying randomness under very simple experimental settings with a minimum number of assumptions.

From a pragmatic perspective, our randomness generation protocol is highly economic and resource-efficient: the preparation of qubit states can be equivalently attained with coherent light (see the "from qubit to coherent states" part in Resutls section), and the entire setup requires no phase stabilization which simplifies the control loop and offered desirable stability in realistic applications. The main technical requirement in our experiment seems to lie in the high-efficiency photodetection part. Commercial superconducting single-photon detectors are able to achieve the efficiency required for the current protocol, but a more intriguing direction would be to further improve the protocol so the requirement of high-efficiency photodetection can be eased. We anticipate the future protocols developed from our results will have a nice prospect for broader practical applications.

Our experimental setup yields a 4 outcome measurement. However, in the

present semi-device independent setting, the certifiable randomness cannot be larger than $\log_2(3)$ [179]. This is because we effectively treat the prepared coherent states as qubits, even through their dimension is unbounded for the eavesdropper. Since a $4^{\text{th}}$ outcome turns useless, we reduce the size of the set of outcomes. We do that by absorbing the outcome $b = 3$ with the events $b = 0$ and $b = 1$ with equal probability so that, at the end of the day, our protocol only yields a ternary outcome. Our measurements consist entirely of a single photo-detector which is subject to photon losses and dark counts. These effects are accommodated as undesired deviations in the observed statistics.

In real-life experiments, the observed probabilities are built from the frequencies of data-points. Due to the finite number of data points, the observed frequency of events $\text{freq.}(b|x) = n_{bx}/n_x$, for $n_{bx}$ denoting the total number of events $b$ given a state preparation $x$ does not exactly represent the true conditional probability $p(b|x)$. We deal with finite-size effects making use of the Chernoff-Hoeffding tail inequality [198], as we detail in the Sec. 7.9. This allows us to quantify the deviation of the sum of data-points from its expected value, assuming these are obtained in independent and identically distributed (i.i.d.) rounds. In Fig. 7.3, we show a simulation of the minimum min-entropy when taking into account finite size effects from the collected data. Due to the pessimistic approach of the Chernoff-Hoeffding tail inequality, the randomness we are able to extract does not go beyond one bit of randomness per round.

As mentioned earlier, in our analysis we go beyond the i.i.d. assumption and perform a second characterization of finite size effects. In this case, we use the EAT [197], which allows us to quantify the amount of entropy accumulated per round when those are not i.i.d. Furthermore, one could harness the asymptotic equipartition property by considering multiple rounds of the experiment to use the Von-Neumann entropy, instead of the min-entropy, plus a correction term to evaluate the randomness [199] (in fact, the EAT is originally formulated with that idea). This approach would yield higher values of randomness. However, we do not find it necessary in this work, and the reason is twofold. First, because we find our methodology more comprehensive by working directly with a guessing probability. Secondly, because even using the min-entropy which is the most pessimistic (conservative) choice, we are able to reach more than one bit of randomness per round as we see in Fig. 7.3, which empowers even more the protocol we propose.

## 7.5 Methods

As sketched in the main text, the input states used in the experiment are created by using an EOM capable of inducing intensity modulation, and a sequence of two AOMs to carve the output of a 1550.32 nm continuous-wave laser into time-bin-encoded series of pulsed weak coherent states. Concretely, the AOMs attenuate the output of the laser by 77 dB when sending a pulse, so the intensities of the output states are at a single-photon level, and it completely shuts off the laser output when no pulse is being sent. In doing so, we achieve a dark count rate of only $\approx 10^{-6}$ per pulse. The EOM, on the other hand, generates the correct amplitudes for the input states during a 10 ns time window, which is chosen as the duration of the time-bin. We have introduced a feedback control loop to stabilize the output power of the EOM; consequently, we are able to keep the standard deviation of the counting rate against time at $(0.29\pm0.22)\%$ during each run of the experiment.

The values for $\alpha$ and $\beta$ are calibrated using a single-photon detector with a specified efficiency of $\eta = 94\%$. Based on the amplitude we desire to achieve, we are able to calculate the click rate related to this coherent state. For the values $\alpha = 0.41$ and $\beta = 0.66$, which optimize the randomness we obtain, we are expected to measure clicks on 14.5% and 33.2% of the coherent state pulses, respectively.

In order to account for finite size effects, we take around $10^8$ rounds of measurements. Since our oscilloscope can only save data from $10^7$ rounds of experiments before the memory is filled, we obtain the amount of data which is possible and then perform a calibration measurement before taking data again. This ensures that we are keeping track of possible fluctuations of parameters on the experiment and we are able to correct for it. Indeed, the standard deviation of the counting rate across all experiment runs is only 0.37%, certifying the consistency between the states prepared at different times.

## 7.6 S1: Unconstrained dimensionality and semi-device independence

In this section we show how the eavesdropper can use the unbounded dimensionality assumption in a semi-device independent setting. Also, we show how bounding the overlaps between the prepared states can limit this the use of any additional dimension by the eavesdropper.

Consider the state discrimination scenario with three preparations. Two of the prepared states $|\tilde{\psi}_0\rangle$ and $|\tilde{\psi}_1\rangle$ have support only on a two-dimensional Hilbert space, but the third state $|\tilde{\psi}_2\rangle$ may have support also on a third dimension. Then,

$$|\tilde{\psi}_0\rangle = \cos\frac{\phi}{2}|0\rangle + \sin\frac{\phi}{2}|1\rangle + 0\,|2\rangle \tag{7.14}$$

$$|\tilde{\psi}_1\rangle = \cos\frac{\phi}{2}|0\rangle - \sin\frac{\phi}{2}|1\rangle + 0\,|2\rangle \tag{7.15}$$

$$|\tilde{\psi}_2\rangle = \sqrt{a}\left(\cos\frac{\theta}{2}|0\rangle + e^{i\varphi}\sin\frac{\theta}{2}|1\rangle\right) + \sqrt{1-a}\,|2\rangle \ . \tag{7.16}$$

Suppose now that we only trust a bound on the overlap of the prepared states. Let us define $|\langle\tilde{\psi}_0|\tilde{\psi}_1\rangle| \geq |d_{01}|$, $|\langle\tilde{\psi}_0|\tilde{\psi}_2\rangle| \geq |d_{02}|$ and $|\langle\tilde{\psi}_1|\tilde{\psi}_2\rangle| \geq |d_{12}|$. This means that

$$a \geq \frac{d_{02}^2 + d_{12}^2}{1 + d_{01}\cos\theta} \ . \tag{7.17}$$

If $a = 1$, the three states will have support on the same bi-dimensional Hilbert space. Whilst the measurement device is treated as a black box, the preparation device is partially characterized through the bounds we place on the overlaps of the prepared states. Then, the eavesdropper has the freedom in choosing the states $|\psi_x\rangle$ in terms of the angles $\phi$, $\theta$ and $\varphi$ that satisfy those bounds. Her probability of guessing the measurement outcome when state $|\tilde{\psi}_2\rangle$ is prepared is

$$\begin{aligned} p_g &= \sum_\omega q(\omega)\max_b \left\{\langle\tilde{\psi}_2|\,\hat{\pi}_b^\omega\,|\tilde{\psi}_2\rangle\right\} \\ &= \sum_\omega q(\omega)\max_b \{a\,\langle\psi_2|\,\hat{\pi}_b^\omega\,|\psi_2\rangle + (1-a)\,\langle 2|\,\hat{\pi}_b^\omega\,|2\rangle + \\ &\qquad \sqrt{a(1-a)}\,(\langle\psi_2|\,\hat{\pi}_b^\omega\,|2\rangle + \langle 2|\,\hat{\pi}_b^\omega\,|\psi_2\rangle)\Big\} \ . \end{aligned} \tag{7.18}$$

The support of the POVM onto the qubit space spanned by the test states $|\tilde{\psi}_0\rangle$ and $|\tilde{\psi}_1\rangle$ is constrained by the reproducibility of the observed statistics $p(b|x)$. However, the support onto the sub-space spanned by $|2\rangle$ does not have any constraints applied. This implies that $p_g$ is maximum whenever the measurement described by the POVM $\{\hat{\pi}_b^\omega\}$ has minimal support on the constrained subspace. Thus, the upper bound on $p_g$ is given whenever $a$ is minimal, which we know is lower bounded indirectly by (7.17) whenever the

overlaps of the prepared states are also bounded.

On the SDP this is reflected by considering the discrimination of the qutrit states $|\tilde{\psi}_x\rangle$. Assume that Eve is even allowed to change the angles $\phi$ and $\theta$, so that the bounds on the overlaps are still satisfied. Eve can pick them to be the ones she wants in order to make the support onto the third dimension as large as she can. Let's see what is the best she can do. We first relate both angles in a single expression by first writing $a = (d_{02}^2 - d_{12}^2) \,/\, \left(\sqrt{1-d_{01}^2}\sin\theta\cos\varphi\right)$ and equating with the right-hand side in (7.17). One gets

$$\cos\theta = \frac{d_{02}^2 - d_{12}^2}{d_{02}^2 + d_{12}^2} \frac{1 + d_{01}\cos\theta}{\sqrt{1-d_{01}^2}\sin\theta} \ . \tag{7.19}$$

If one plots $\cos\theta$ vs. $\cos\varphi$, one will see that: if $d_{02} > d_{12}$, $\cos\theta$ is maximal if $\cos\varphi = 1$; if $d_{02} < d_{12}$, $\cos\phi$ is maximal if $\cos\varphi = -1$; and if $d_{02} = d_{12}$, $\cos\theta$ is maximal if $\cos\varphi = 0$. The maximal value of $\cos\theta$ is the same in the three cases, being

$$(\cos\theta)_{\max} = \frac{2d_{02}d_{12} - d_{01}\left(d_{02}^2 + d_{12}^2\right)}{d_{02}^2 - 2d_{01}d_{02}d_{12} + d_{12}^2} \ , \tag{7.20}$$

which means

$$a \geq \frac{d_{02}^2 + d_{12}^2 - 2d_{01}d_{02}d_{12}}{1 - d_{01}^2} \ . \tag{7.21}$$

Equation (7.21) defines the surface of a tetrahedron with curved faces. The amplitude $a$ decreases towards the center of the tetrahedron non-symmetrically, as we show in Fig. 7.5.

On the SDP, the prepared states shall only be expressed in terms of the bounds of the overlaps $d_{xy}$, which can be done by replacing the angles $\phi$, $\theta$ and $\varphi$ with $d_{01}$, $d_{02}$ and $d_{12}$ accordingly. This yields

$$\begin{aligned}
|\tilde{\psi}_0\rangle =& \sqrt{\frac{1+d_{01}}{2}}\,|0\rangle + \sqrt{\frac{1-d_{01}}{2}}\,|1\rangle + 0\,|2\rangle \\
|\tilde{\psi}_1\rangle =& \sqrt{\frac{1+d_{01}}{2}}\,|0\rangle - \sqrt{\frac{1-d_{01}}{2}}\,|1\rangle + 0\,|2\rangle \\
|\tilde{\psi}_2\rangle =& \frac{1}{\sqrt{2}}\frac{d_{02}+d_{12}}{\sqrt{1+d_{01}}}\,|0\rangle + \frac{1}{\sqrt{2}}\frac{d_{02}-d_{12}}{\sqrt{1-d_{01}}}\,|1\rangle \\
&+ \sqrt{1 - \frac{|d_{02}+d_{12}|^2}{2(1+d_{01})} - \frac{|d_{02}-d_{12}|^2}{2(1-d_{01})}}\,|2\rangle \ .
\end{aligned} \tag{7.22}$$

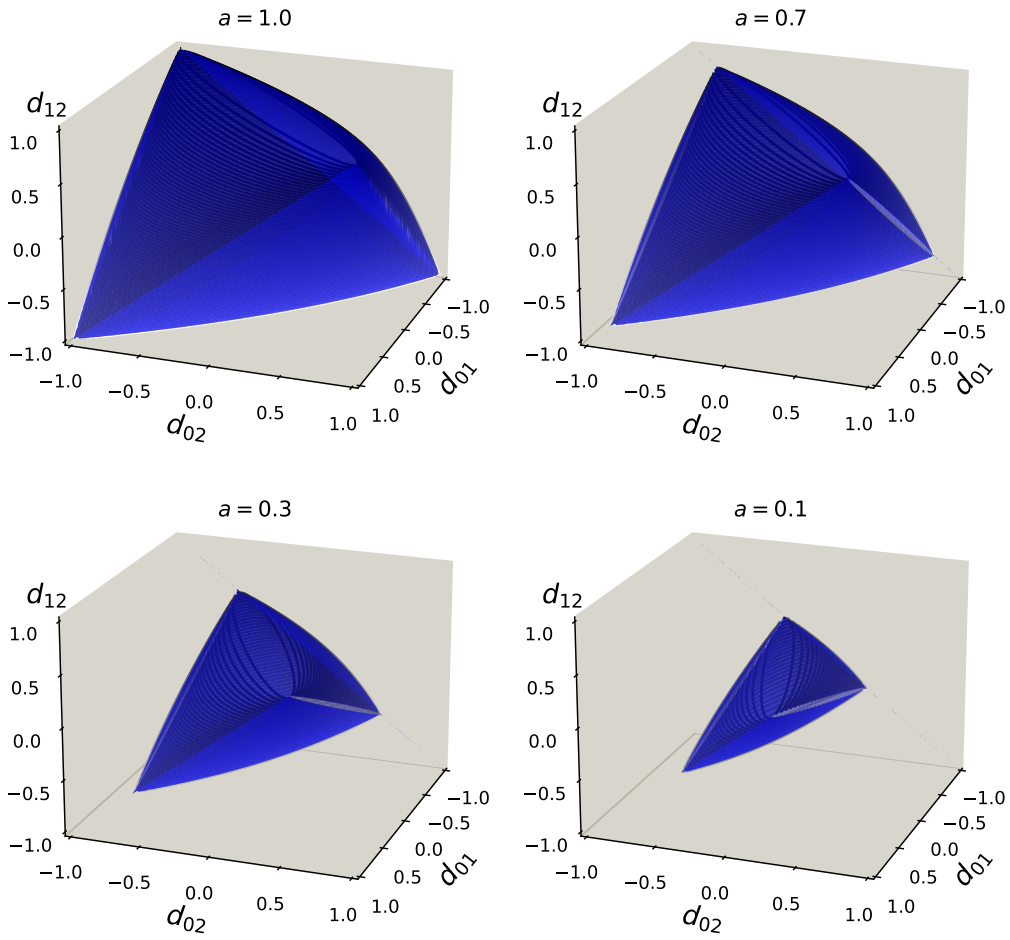



Figure 7.5: Figure extracted from Ref. [16]. Tetrahedron formed by the available overlap configurations in the preparation of three non-orthogonal quantum states. The parameter $a$ indicates the minimal support of the prepared states onto a two-dimensional subspace if only their overlaps are bounded.

A particular choice of bounds on the overlaps and their phases limits the accessibility to a the third dimension by Eve. Concretely, one can tune the states to be symmetric in the sense that $d_{02} = d_{12}^* = \tilde{d}e^{i\gamma}$. Thus, fulfilling the relation $\tilde{d}^2 = (1 - d_{01}^2)/(2(1 - d_{01}\cos 2\gamma))$ makes the third component of $|\tilde{\psi}_2\rangle$ null. This means that we can be sure that any potential eavesdropper will gain no information of the outcome by reaching into an additional third dimension.

## 7.7 S2: Implementations: specific details

In this section of the appendix we detail the specific parameters to be adjusted to obtain the desired statistics in the measurement outcomes on the proposed implementation.

The proposed implementation for the USD setup consists in preparing the equi-probable two-mode coherent states $|\psi_0\rangle = |\alpha\rangle \otimes |0\rangle$ and $|\psi_1\rangle = |0\rangle \otimes |\alpha\rangle$. These can be unambiguously discriminated by means of using only photo detectors in each mode. If only the photo detector in the first mode clicks, that would mean that state $|\alpha 0\rangle$ had been prepared, and thus, we associate the outcome $b = 0$. Otherwise, if only the second photo-detector clicks, means that $|0\alpha\rangle$ was prepared and we associate $b = 1$ as a measurement outcome. Note that if these two states are prepared, there is a possibility that none of the detectors click. If that happens, the measurement is uncertain of which state was prepared and we associate this events with the inconclusive measurement outcome $b = 2$. Assume now that we include the preparation of a third two-mode coherent state $|\psi_2\rangle = |\beta_0\beta_1\rangle$ into play. Whenever this state is prepared, either only one detector can click, the other, none or even both at the same time. Let us go through all possible measurement events

and their probabilities to happen whenever a generic state $|\psi_x\rangle$ is prepared.

$$
\begin{aligned}
1 &= \langle\psi_x| \left(\mathbb{1}\otimes\mathbb{1}\right) |\psi_x\rangle = \langle\psi_x| \left(\sum_{n=0}^{\infty} |n\rangle\langle n|\right) \otimes \left(\sum_{m=0}^{\infty} |m\rangle\langle m|\right) |\psi_x\rangle \\
&= \langle\psi_x| \left(|0\rangle\langle 0| + \sum_{n=1}^{\infty} |n\rangle\langle n|\right) \otimes \left(|0\rangle\langle 0| + \sum_{m=1}^{\infty} |m\rangle\langle m|\right) |\psi_x\rangle \\
&= \langle\psi_x| \Bigg( \underbrace{|0\rangle\langle 0|\otimes|0\rangle\langle 0|}_{\text{Detector does't click}} + \underbrace{|0\rangle\langle 0| \otimes \sum_{m=1}^{\infty} |m\rangle\langle m|}_{\text{Click on late bin}} \\
&+ \underbrace{\sum_{n=1}^{\infty} |n\rangle\langle n|\otimes|0\rangle\langle 0|}_{\text{Click on early bin}} + \underbrace{\sum_{n=1}^{\infty} |n\rangle\langle n| \otimes \sum_{m=1}^{\infty} |m\rangle\langle m|}_{\text{Click on both early and late bins}} \Bigg) |\psi_x\rangle \\
&= |\langle 00|\psi_x\rangle|^2 + \sum_{m=1}^{\infty} |\langle 0m|\psi_x\rangle|^2 + \sum_{n=1}^{\infty} |\langle n0|\psi_x\rangle|^2 + \sum_{n=1}^{\infty}\sum_{m=1}^{\infty} |\langle nm|\psi_x\rangle|^2 \ .
\end{aligned}
\tag{7.23}
$$

The event consisting in a simultaneous click at both time-bins does not come into play when unambiguously discriminating the two-mode coherent states $|\alpha 0\rangle$ and $|0\alpha\rangle$. In fact, that event would correspond in a unambiguous identification of the third state $|\beta_0\beta_1\rangle$. Since our aim is to only consider measurement strategies able to unambiguously discriminate solely states $|\alpha 0\rangle$ and $|0\alpha\rangle$, whenever that event occurs we will consider that $b = 0$ with probability $g_0$, $b = 1$ with the same probability $g_1$ and the rest of the times $b_2$ with probability $b = 2$. No data will be discarded so that the certified randomness is not affected. The considered events and their corresponding probabilities depending on which state was prepared after the post-processing are summarized in table 7.1.

For simplicity and without loss of generality, we will consider non-imaginary coherent amplitudes only. The overlaps of the prepared states are characterized by the amplitudes of the coherent states as follows

$$
d_{01} = e^{-\alpha^2} \qquad d_{02} = e^{-\frac{(\alpha-\beta_0)^2}{2}} e^{-\frac{\beta_1^2}{2}} \qquad d_{12} = e^{-\frac{\beta_0^2}{2}} e^{-\frac{(\alpha-\beta_1)^2}{2}} \ . \tag{7.24}
$$

Over a set of runs, we observed that the best and simplest choice is to pick $g_0 = g_1 = 1/2$ and so $g_2 = 0$.

| Meas. Event / Prepared state | $b = 0$ |
|---|---|
| $\|\psi_0\rangle = \|\alpha\rangle \otimes \|0\rangle$ | $1 - e^{-\|\alpha\|^2}$ |
| $\|\psi_1\rangle = \|0\rangle \otimes \|\alpha\rangle$ | $0$ |
| $\|\psi_2\rangle = \|\beta_0\rangle \otimes \|\beta_1\rangle$ | $\left(1 - e^{-\|\beta_0\|^2}\right) e^{-\|\beta_1\|^2}$ $+g_0 \left(1 - e^{-\|\beta_0\|^2}\right)\left(1 - e^{-\|\beta_1\|^2}\right)$ |
| Meas. Event / Prepared state | $b = 1$ |
| $\|\psi_0\rangle = \|\alpha\rangle \otimes \|0\rangle$ | $0$ |
| $\|\psi_1\rangle = \|0\rangle \otimes \|\alpha\rangle$ | $1 - e^{-\|\alpha\|^2}$ |
| $\|\psi_2\rangle = \|\beta_0\rangle \otimes \|\beta_1\rangle$ | $e^{-\|\beta_0\|^2}\left(1 - e^{-\|\beta_1\|^2}\right)$ $+g_1 \left(1 - e^{-\|\beta_0\|^2}\right)\left(1 - e^{-\|\beta_1\|^2}\right)$ |
| Meas. Event / Prepared state | $b = 2$ |
| $\|\psi_0\rangle = \|\alpha\rangle \otimes \|0\rangle$ | $e^{-\|\alpha\|^2}$ |
| $\|\psi_1\rangle = \|0\rangle \otimes \|\alpha\rangle$ | $e^{-\|\alpha\|^2}$ |
| $\|\psi_2\rangle = \|\beta_0\rangle \otimes \|\beta_1\rangle$ | $e^{-\|\beta_0\|^2} e^{-\|\beta_1\|^2}$ $+g_2 \left(1 - e^{-\|\beta_0\|^2}\right)\left(1 - e^{-\|\beta_1\|^2}\right)$ |

Table 7.1: Summary of the considered measurement events and their corresponding re-normalized probabilities for the USD setup.

## 7.8 S3: Semi-definite program: primal and dual

In this section we formally introduce the semidefinite program we use to bound the certifiable randomness.

### 7.8.1 Primal SDP

We start presenting the primal form of the problem. Our goal is to maximise the guessing probability of the eavesdropper which we can write as

$$p_g = \sum_{\omega} q(\omega) \max_b \left\{ \mathrm{Tr}\left[\rho_2 \hat{\pi}_b^{\omega}\right] \right\} \ . \tag{7.25}$$

The maximisation is done through all possible measurement strategies $\omega$, distributions $q(\omega)$ and POVM elements $\hat{\pi}_b^{\omega}$. These are constrained to be valid distributions and POVMs, which implies

$$q(\omega) \geq 0 \qquad \sum_{\omega} q(\omega) = 1 \qquad q(\omega) \in \mathbb{R} \tag{7.26}$$

$$\hat{\pi}_b^{\omega} \geq 0 \qquad \sum_{b} \hat{\pi}_b^{\omega} \geq 0 \qquad \hat{\pi}_b^{\omega} = (\hat{\pi}_b^{\omega})^{\dagger} \ . \tag{7.27}$$

There is the additional constraint that the observed probabilities must be reproduced on the real experiment. This is reflected in

$$p(b|x) = \sum_{\omega} q(\omega) \, \mathrm{Tr}\left[\rho_x \hat{\pi}_b^{\omega}\right] \ . \tag{7.28}$$

Since states $\rho_x$ are not fully specified, but instead only their overlaps are bounded, we will insert the states in (7.22), so $\rho_x = |\tilde{\psi}_x\rangle \langle \tilde{\psi}_x|$.

This optimisation problem can be rendered as a linear semidefinite program following a couple of steps. First, we will consider only the most relevant steategies, which in our case are those which yield the maximal value $\max_b \{\mathrm{Tr}\left[\rho_x \hat{\pi}_b^{\omega}\right]\}$. This can be done by simply labeling $\omega = b$ the maximal strategy for outcome $b$, i.e. $\max_b \{p(b|x=2,\omega)\} = p(\omega|x=2,\omega)$. This leaves us with only $n_{\mathrm{B}}$ relevant strategies, being $n_{\mathrm{B}}$ the number of different outcomes from the measurement. Secondly, we will absorb the distribution $q(\omega)$ in the POVM element $\hat{\pi}_b^{\omega}$ and define a new quantity $\hat{M}_b^{\omega} = q(\omega)\hat{\pi}_b^{\omega}$. The definition of this new operator changes the above constraints to the following:

$$\hat{M}_b^{\omega} \geq 0, \quad \hat{M}_b^{\omega} = \left(\hat{M}_b^{\omega}\right)^{\dagger} \ \forall b, \omega, \quad \sum_b \hat{M}_b^{\omega} = \frac{1}{D} \mathrm{Tr}\left[\sum_b \hat{M}_b^{\omega}\right] \ \forall \omega \ , \tag{7.29}$$

where $D$ is the dimensionality of the eavesdropper. The useful space accessible by the eavesdropper is that spanned by the states involved in the experiment. Since we are considering a three-state discrimination setting, the dimension can be at maximum the number of states, i.e. $D = 3$. The reproducibility constraint is also changed to simply

$$p(b|x) = \sum_{\omega} \mathrm{Tr}\left[\rho_x \hat{M}_b^{\omega}\right] \ . \tag{7.30}$$

Finally, we can re-write the guessing probability in the following way

$$p_g = \sum_{\omega} \mathrm{Tr}\left[\rho_2 \hat{M}_{\omega}^{\omega}\right] \ . \tag{7.31}$$

An upper bound $p_g^* \geq p_g$ can be found by maximising it through all possible $2 \times 2$ matrices $\hat{M}_b^{\omega}$ that fulfil the constraints above.

### 7.8.2 Dual SDP

We continue by presenting the dual form of the SDP. We begin by writing the Lagrangian corresponding to the present problem:

$$\begin{aligned} \mathcal{L} = & \sum_{\omega} \mathrm{Tr}\left[\rho_2 \hat{M}_{\omega}^{\omega}\right] + \sum_{b,\omega} \mathrm{Tr}\left[G_b^{\omega} \hat{M}_b^{\omega}\right] \\ & + \sum_{\omega} \mathrm{Tr}\left[\hat{H}^{\omega} \sum_{b} \left(\hat{M}_b^{\omega} - \frac{1}{D}\mathrm{Tr}\left[\hat{M}_b^{\omega}\right] \mathbb{1}\right)\right] \\ & + \sum_{b,x} \nu_{b,x} \left(\sum_{\omega} \mathrm{Tr}\left[\rho_x \hat{M}_b^{\omega}\right] - p(b|x)\right) \ . \end{aligned} \tag{7.32}$$

The supremum $\mathcal{S}$ of the Lagrangian, over the primal SDP variables, reads

$$\mathcal{S} = \sup_{\hat{M}_b^{\omega}} \sum_{b,\omega} \mathrm{Tr}\left[\hat{M}_b^{\omega} K_b^{\omega}\right] - \sum_{b,x} \nu_{b,x} p(b|x) \ , \tag{7.33}$$

where we defined

$$K_b^{\omega} = \rho_2 \delta_{b,\omega} + \sum_x \rho_x \nu_{b,x} + G_b^{\omega} + \hat{H}^{\omega} - \frac{1}{D}\mathrm{Tr}\left[\hat{H}^{\omega}\right] \mathbb{1} \ . \tag{7.34}$$

The supremum will diverge unless $K_b^{\omega} = 0$. This is also commonly known as Lagrange stability in convex optimisation. Also, since $G_b^{\omega} \geq 0 \ \forall b, \omega$, we

can drop them and imply $K_b^\omega \neq 0$. With these remarks, the dual form of the SDP reads

$$\begin{aligned} p_g^* = \underset{\hat{H}^\omega, \nu_{b,x}}{\text{minimize}} \quad & -\sum_{b,x} \nu_{b,x} p(b|x) \\ \text{subject to} \quad & \rho_2 \delta_{b,\omega} + \sum_x \rho_x \nu_{b,x} + \hat{H}^\omega - \frac{1}{D} \operatorname{Tr}\left[\hat{H}^\omega\right] \mathbb{1} \leq 0 \\ & \hat{H}^\omega = \left(\hat{H}^\omega\right)^\dagger \ , \end{aligned} \tag{7.35}$$

for $\delta_{i,j}$ being the Kronecker delta and $D$ the dimensionality of the eavesdropper. Observe that the observed statistics $p(b|x)$ only appear in the dual object function and not on the constraints. This implies that, given a dual solution (so a set of $\hat{H}^\omega$ and $\nu_{bx}$ that fulfill the constraints above), valid bounds on $p_g$ can still be computed for any $p(b|x)$ by just evaluating the object function. This form allows us to treat finite size effects as we explain in the following section.

## 7.9 S4: Finite size effects and entropy accumulation

In this section we explain how we treat finite size effects from the data extracted in the experiment. Also, we explain how we can abandon the general assumption of independent and identically distributed rounds (i.i.d) though the entropy accumulation theorem (EAT) as is explained in [197].

### 7.9.1 Finite-size effects under the i.i.d. assumption

In the real life implementation of the protocol, the observed statistics are built from finite sets of collected data. Thus, the entropy is computed based on a finite number of samples. In order to incorporate such finite-size effects into our analysis, we make use of the Chernoff-Hoewffding tail inequality [198]. This allows us to quantify the probability that the sum of data-points deviates from its expected value, assuming these are obtained in independent and identically distributed rounds. From the experiment we collect pairs of data-points for each question (or state preparation $x$) and answer (measurement outcome $b$). We label by $n_{b,x}$ the number of pairs with question $x$ and answer $b$. Then, from the set of questions we label $n_x$ the number of questions $x$, and form the set of answers we label $n_b$ the number of answers $b$. The total number of extracted data-points is $N = \sum_b n_b = \sum_x n_x = \sum_{b,x} n_{b,x}$. From

these, we can obtain the observed frequencies as $\text{freq.}(b|x) = n_{b,x}/\sum_b n_{b,x}$. Then, neglecting finite-size effects, we can compute the guessing probability according to the dual SDP with

$$p_g^* \leq p_g = -\sum_{b,x} \nu_{b,x} \text{freq.}(b|x) \ . \tag{7.36}$$

With the solution of the dual SDP (i.e., with $\hat{H}^\omega$ and $\nu_{b,x}$ that satisfy the dual constraints), we proceed by evaluating a bound on $p_g$, now taking into account the finite-size effects. The Chernoff-Hoeffding inequality allows us to write a bound on the true observed probabilities with respect to the obtained frequencies. The inequality reads

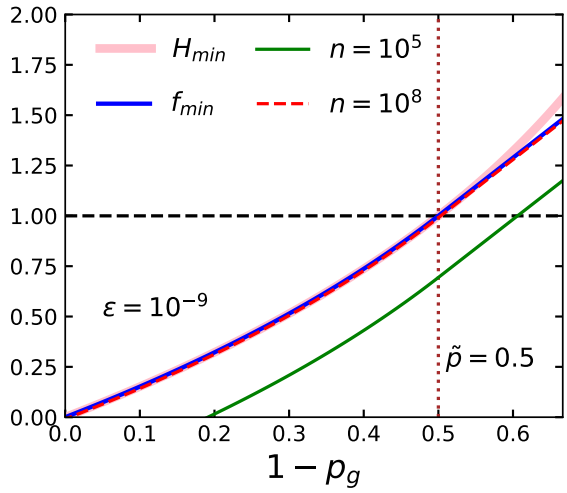
$$\text{freq.}(b|x) - \sqrt{\frac{\log(1/\varepsilon)}{2N}} \leq p_{\text{f.s.}}(b|x) \leq \text{freq.}(b|x) + \sqrt{\frac{\log(1/\varepsilon)}{2N}} \ , \tag{7.37}$$

for here $\varepsilon$ being the probability of this bound not being satisfied (in our results we choose $10^{-9}$). Since coefficients $\nu_{b,x}$ can be negative or positive, we might choose either the lower or upper bound on the Chernoff-Hoeffding inequality. The most conservative choice is to choose the bounds that yield the highest bound in $p_g$. That is, if $\nu_{b,x} \geq 0$ we choose the lower bound and otherwise if $\nu_{b,x} \leq 0$. This yields the new bound on the guessing probability

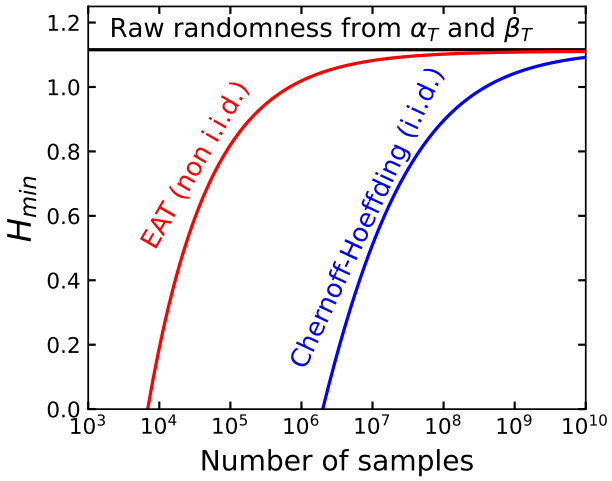
$$p_g^* \leq p_g = -\sum_{b,x} \nu_{b,x} p_{\text{f.s.}}(b|x) \ . \tag{7.38}$$

### 7.9.2 Entropy accumulation theorem: dropping the i.i.d. assumption

The i.i.d. assumption is not very attractive in (semi-)Device Independent protocols. Indeed, this assumes that the eavesdropper can not learn form past rounds to have a better guess in future rounds (sort of if the eavesdropper looses its memory in each round). To get rid of this strong assumption, we refer to the entropy accumulation theorem and its application in [197] for a Device-Independent setting. Here we adapt it to our semi-Device Independent scenario, where nonlocality does not play any role. The EAT places a bound on the *smooth*-min-entropy, that is the min-entropy of a distribution $\varepsilon$-close to the target distribution, per round in a prepare-and-measure experiment with $N$ rounds. The EAT implies that the operationally total relevant uncertainty about the total set of outcomes over $N$ rounds $B_1^N$ corresponds to the sum of the entropies of the individual rounds to first order in $N$ under the i.i.d. assumption, plus a contribution from not assuming the i.i.d. case. This




(a) Trade-off function.




(b) Scaling with number of samples.

Figure 7.6: Figure extracted from Ref. [16]. **(a)** Illustration of the trade-off function $f_{min}$ we choose to characterize the EAT bound. For a large amount of data, the EAT bound becomes practically equivalent to the trade-off function. **(b)** Scaling of the obtained data-points with the number of collected samples. We show the raw randomness (i.e. with finite size effects not accounted for), the i.i.d. case treated with the Chernoff-Hoeffding inequality and the non i.i.d. case treated with the EAT.

contribution is provided given that one quantifies the uncertainty of each individual round with the Von-Neumann entropy of a suitable chosen state. Formally, the EAT is

$$\frac{1}{N}H_{min}^{\varepsilon}(B_1^N|S_1^N E) > t - \frac{\nu}{\sqrt{N}} \ , \tag{7.39}$$

where

$$\begin{aligned} &\nu = 2\left(\log_2\left(1+2d_{B_i}\right) + \lceil ||\nabla f_{min}||_\infty \rceil\right)\sqrt{2\log_2\left(\varepsilon \cdot p_\Omega\right)} \\ &\text{and} \qquad t \leq f_{min}\left(\text{freq}\left(b|x\right)\right) \ . \end{aligned} \tag{7.40}$$

Here, $d_{B_i}$ is the dimension of $B_i$, which in our case is the number of different outcomes: 3. Also, $\varepsilon$ the smoothing of the min-entropy. Finally, $p_\Omega$ is the probability of the event $\Omega$ of winning a game, or in our case the probability that the eavesdropper does not guess toe outcome (i.e. $p_\Omega = 1 - p_g$).

A trade-off function for EAT channels $f_{min}$ is formally defined as $f_{min}(p) \leq \inf_{\sigma\in\Sigma_i(p)} H(B_i|S_iR')_\sigma$. The infimum of the Von-Neumann entropy is performed over all post-measurement states $\sigma$ after each EAT channel (that is, every round in our experiment). In our case we can, instead of the Von-Neumann entropy, consider the min-entropy which we know is the lower bound on the Von-Neumann entropy. We do that for two main reasons. First, we do not need to extend our results through the Asymptotic Equipartition Property (AEP) from characterising the randomness of the outcome with the min-entropy to the Von-Neumann entropy. Our results are robust and good enough for the purpose of this work: certify more than one bit of randomness. The second, and more important reason, is that we already know which is the minimum min-entropy per round with finite-size effects already accounted for. Thus, we consider

$$\begin{aligned} f_{min} \leq \inf_{\sigma\in\Sigma_i(p)} H_{min}(B|E)_\sigma &= -\log_2\left(-\sum_{b,x}\nu_{b,x}p_{\text{f.s.}}(b|x)\right) \\ &\leq \inf_{\sigma\in\Sigma_i(p)} H(B_i|S_iR')_\sigma \ . \end{aligned} \tag{7.41}$$

The next step is to find a good candidate for $f_{min}$ which satisfies that condition. We see that the maximal value of its derivative $\lceil ||\nabla f_{min}||_\infty \rceil$ appears as a negative term in the lower bound. This derivative is performed over the probability of event $\Omega$ ($p_\Omega$) which in our case is the event that the eavesdropper does not guess the measurement outcome (i.e. $p_\Omega = 1 - p_g$). A good candidate is then than that is lower than the min-entropy and has

a derivative with a not-so-big maximal value. We plot in Fig. 7.6 the min-entropy vs. $1 - p_g$. We see that its derivative increases as the guessing probaiblity decreases. Thus, since our goal is to find more than one bit of ranodmness, we pick $f_{min} = H_{min}$ for $1 - p_g \leq \tilde{p} = 1/2$, and an increasing straight line otherwise. This is pictured in Fig. 7.6. Specifically,

$$f_{min} = \begin{cases} H_{min} & \text{if } 1 - p_g \leq \tilde{p}, \\ \dfrac{1 - p_g - \tilde{p}}{(1 - \tilde{p}) \log 2} - \log_2(1 - \tilde{p}) & \text{if } 1 - p_g \geq \tilde{p} \ . \end{cases} \tag{7.42}$$

With all this, we can find a certifiable lower bound on the smooth-min-entropy, exempt of the i.i.d. assumption.

# Chapter 8

# Conclusion and outlook

In this thesis, I investigated two main branches in quantum information science: quantum randomness certification and quantum contextuality, form the perspective of quantum state discrimination. Here, I take an overview of all the works presented in this thesis to reflect on their potential, extract some conclusions and consider future related routes of investigation.

### Maximum confidence state discrimination

The problem of discriminating quantum states has been one of the central topics in quantum information science since the arrival of the quantum theory. This motivated the development of many state discrimination protocols, the most remarkable being minimum error state discrimination (MESD) and unambiguous state discrimination (USD). Both protocols have been presented in this thesis, and there has been a lot of progress during the recent years. Here, I centered on studying maximum confidence state discrimination (MCSD) [183], and the reason in twofold. First, because MCSD generalises the notion of MESD and USD, predicting the best possible measurement for a flexible rate of inconclusive outcomes or undetected events. It follows that, this flexibility, might facilitate potential implementations, which makes all works presented in this thesis good candidates to be implemented in the laboratory. The second reason is that MCSD proves to be the best measurement strategy in terms of randomness certification and finding contextual advantages, at least in two state discrimination scenarios. We base this statement from the observation that statistics in MCSD cover the limits in the correlation space parameterized with error and success probabilities [26]. There, measurement statistics are only reproducible by extremal and unique POVMs, making them the best candidates for randomness certification.

In the works presented in this thesis I centered in studying the simplest case of qubit state discrimination. A natural continuation would therefore be to extend the scenario to $D > 2$ dimensional Hilbert spaces. Maximum confidence measurements have been already been studied in a multidimensional system [157], but only in the symmetric noiseless case. The main obstacle in examining a more general scenario is that the intuitive Bloch representation of does not hold anymore. Even for the nearest extension of qutrit states, an analogous Bloch representations requires eight dimensions. In this case, a study of the geometry of a noisy and general $D$-dimensional maximum confidence measurement remains still a difficult challenge.

Following up, the search for extremal and unique measurements in qubit spaces through the investigation of quantum correlations (as in Refs. [26, 200]) could be extended to $N > 2$ states. The challenging part would be to perform this search at the level of probabilities, and properly find a good parametrization of the correlation space. For two-state discrimination scenarios, by simply defining success and error probabilities for both states, one can draw a comprehensive two-dimensional correlation space. The choice of error probabilities, in a general $N$ state discrimination case, is not trivial.

**Contextual advantages in state discrimination**

The context-dependent nature of quantum observables has been one of the main gaps between classical and quantum models. As so in quantum information science, Bell-Kochen-Specker contextuality [18, 19] served to propel the search for advantages in many aspects versus classical models. Generalised contextuality [20] has been found to provide certain advantages, for instance, in random-access codes [142], parity-oblivious multiplexing [140, 141], and certain advantages in quantum communication [143, 145] and state discrimination [27]. In this thesis we elaborated further in finding contextual advantages in state discrimination. Concretely in Ref. [25] we extended the work in Ref. [27] finding contextual advantages for USD and MCSD. Moreover, in Ref. [17] we studied the contextual advantages in using MCSD for a particular state in terms of randomness certification. A simple an interesting continuation of this work would consist in testing the contextual advantages in randomness certification when implementing MCSD for the whole ensemble. In [26] we found that this strategy accesses the limits of the space of correlations in two-state discrimination, which is also very convenient for randomness generation.

An alternative interesting route of research would be to find other practical

uses for contextual advantages in state discrimination. For instance, contextuality has been found to provide advantages in quantum key distribution (QKD) in the notion introduced by Kochen and Specker [201] and in the Klyachko-Can-Binicioglu-Shumovsky [138] contextuality scenario [202]. Generalised contextuality advantages in state discrimination could also potentially be used in communication tasks with QKD protocols.

**Quantum randomness certification**

Quantum randomness has also been a rapidly developing field of quantum information science, especially showing significant progress, both fundamentally and experimentally, over the recent years. There exist, however, many challenges in finding feasible implementations of randomness certification protocols, from which I will highlight two. First, randomness protocols are theoretically designed at the level of operators, POVMs and state preparations. It is not usually trivial to find a direct implementation of a particular measurement from a POVM description. All that, without mentioning the experimental challenges entailed in concrete platforms, to generate and maintain quantum states with tolerable, albeit inevitable, noise values. This introduces the noise robustness as another main challenge. We know that high values for randomness can be achieved by extremal and unique measurements, reaching the limits of the correlation space. Any kind of noise affecting our setup disturbs the targeted statistics. A small displacement on the correlation space away from the border of the convex set involves a prominent drop on the certifiable randomness. The protocol in Ref. [16] is implemented in an optical platform, using coherent states of light. In this case, the main sources of noise are photon losses and dark counts. There, we managed to calibrate our experimental setup to tolerate affordable losses and dark counts, and still obtain extremely good values of randomness per round.

Moreover, in this thesis we considered only semi-device independent settings in prepare-and-measure scenarios. In Ref. [17] for instance, only the overlap of the two possible preparations is bounded. Since two states can only span at maximum a two-dimensional Hilbert space, the dimension is automatically bounded to a qubit-space. This is not the case in Ref. [16], where we aim to certify the randomness for three qubit-state preparations. There, we used a trick to bound the dimensionality of the eavesdropper by only using a bound on the overlap of the three prepared states. Nonetheless, there is a potential alternative method, which involves a hierarchy of semidefinite programs. The Navascués-Pironio-Acín (NPA) hierarchy [203, 204] is a method to approximate the quantum set of correlations in bi-partite

scenarios which does not require to specify either the states or measurements involved in the experiment, only the observable probabilities. This means that the dimension of the Hilbert space is completely uncharacterised. Recently, many works implemented the NPA hierarchy for device-independent randomness certification [205–208]. A similar implementation of the NPA, but in a semi-device independent single-party prepare-and-measure scenario such as in Ref. [16], would also solve the issue with the unbounded Hilbert space dimension. An SDP hierarchy in this case would approximate the quantum set of correlations reproducible in a prepare-and-measure scenario from the exterior. Thus, it would provide an upper-bound on the guessing probability, which gets tighter as the hierarchy level increases. This method would also facilitate the randomness certification semi-device independently if the scenario is extended for more than three state preparations.

# References


[1] Charles H Bennett and Gilles Brassard. "Quantum cryptography and its application to provably secure key expansion, public-key distribution, and coin-tossing". *Proceedings of IEEE International Symposium on Information Theory, St Jovite, Canada*. 1983.

[2] Nicholas Metropolis and S. Ulam. "The Monte Carlo Method". *Journal of the American Statistical Association* 44.247 (1949). PMID: 18139350, pp. 335–341.

[3] C. E. Shannon. "Communication theory of secrecy systems". *The Bell System Technical Journal* 28.4 (1949), pp. 656–715.

[4] R. Gennaro. "Randomness in cryptography". *IEEE Security and Privacy* 4.2 (2006), pp. 64–67.

[5] Sen Tian et al. "Random Sampling-Arithmetic Mean: A Simple Method of Meteorological Data Quality Control Based on Random Observation Thought". *IEEE Access* 8 (2020), pp. 226999–227013.

[6] Lyn Carson and Brian Martin. *Random Selection in Politics*. Greenwood Publishing Group, 1999.

[7] Pei Zhang et al. "Quantum gambling based on Nash-equilibrium". *npj Quantum Information* 3.1 (2017), p. 24.

[8] Calvin T. Long. *Elementary Introduction to Number Theory (2nd ed.)* D. C: Heath, 1972.

[9] Anthony J. Pettofrezzo. *Elements of Number Theory*. Englewood Cliffs, N.J. :Prentice-Hall, 1970.

[10] David Burton. *The History of Mathematics: An Introduction*. McGraw Hill, 2010.

[11] F. James. "A review of pseudorandom number generators". *Computer Physics Communications* 60.3 (1990), pp. 329–344.

[12] Luc Devroye. "Non-Uniform Random Variate Generation". 1986.

[13] A. Einstein, B. Podolsky, and N. Rosen. "Can Quantum-Mechanical Description of Physical Reality Be Considered Complete?" *Phys. Rev.* 47.10 (1935), pp. 777–780.

[14] Manabendra Nath Bera et al. "Randomness in quantum mechanics: philosophy, physics and technology". *Reports on Progress in Physics* 80.12 (2017), p. 124001.

[15] J. S. Bell. "On the Einstein Podolsky Rosen paradox". *Physics Physique Fizika* 1.3 (1964), pp. 195–200.

[16] Carles Roch i Carceller et al. "More than one bit of semi-device-independent randomness from a single qubit" (In preparation).

[17] Carles Roch i Carceller et al. "Quantum vs Noncontextual Semi-Device-Independent Randomness Certification". *Phys. Rev. Lett.* 129.5 (2022), p. 050501.

[18] John S. Bell. "On the Problem of Hidden Variables in Quantum Mechanics". *Rev. Mod. Phys.* 38.3 (1966), pp. 447–452.

[19] S. Kochen and E. Specker. "The Problem of Hidden Variables in Quantum Mechanics". *Indiana Univ. Math. J.* 17 (1968), pp. 59–87.

[20] R. W. Spekkens. "Contextuality for preparations, transformations, and unsharp measurements". *Phys. Rev. A* 71.5 (2005), p. 052108.

[21] David A. Meyer. "Finite Precision Measurement Nullifies the Kochen-Specker Theorem". *Phys. Rev. Lett.* 83.19 (1999), pp. 3751–3754.

[22] Adrian Kent. "Noncontextual Hidden Variables and Physical Measurements". *Phys. Rev. Lett.* 83.19 (1999), pp. 3755–3757.

[23] Clifton Rob and Kent Adrian. "Simulating quantum mechanics by non-contextual hidden variables". *Proc. R. Soc. Lond. A.* 456 (2000), pp. 2101–2114.

[24] Jonathan Barrett and Adrian Kent. "Non-contextuality, finite precision measurement and the Kochen–Specker theorem". *Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics* 35.2 (2004), pp. 151–176.

[25] Kieran Flatt et al. "Contextual Advantages and Certification for Maximum-Confidence Discrimination". *PRX Quantum* 3.3 (2022), p. 030337.

[26] Carles Roch i Carceller and Jonatan Bohr Brask. "A contextuality witness inspired by optimal state discrimination" (In preparation).

[27] David Schmid and Robert W. Spekkens. "Contextual Advantage for State Discrimination". *Phys. Rev. X* 8.1 (2018), p. 011015.

[28] Michael A. Nielsen and Isaac L. Chuang. *Quantum Computation and Quantum Information: 10th Anniversary Edition*. Cambridge University Press, 2010.

[29] Stephen M. Barnett and Sarah Croke. "Quantum state discrimination". *Adv. Opt. Photon.* 1.2 (2009), pp. 238–278.

[30] "Semidefinite programming". 38.1 (1996), pp. 49–.

[31] Paul Skrzypczyk and Daniel Cavalcanti. *Semidefinite Programming in Quantum Information Science*. 2053-2563. IOP Publishing, 2023.

[32] Masahito Hayashi. *Quantum Information Theory: Mathematical Foundation*. Springler Nature, 2017.

[33] Leslie E. Ballentine. *Quantum Mechanics: A Modern Development*. Simon Fraser University, Canada, 2014.

[34] Rodrigo A. Thomas et al. "Entanglement between distant macroscopic mechanical and spin systems". *Nature Physics* 17.2 (2021), pp. 228–233.

[35] Max Born. "Quantenmechanik der Stossvorghange". *Z. Phys.* 38.11-12 (1926), pp. 803–827.

[36] David Griffiths. *Introduction to Elementary Particles (2nd ed.)* John Wiley and Sons, 2008.

[37] Marie Ioannou et al. "Steering-based randomness certification with squeezed states and homodyne measurements". *Phys. Rev. A* 106.4 (2022), p. 042414.

[38] Zheng-Da Li et al. "Measurement-Device-Independent Entanglement Witness of Tripartite Entangled States and Its Applications". *Phys. Rev. Lett.* 124.16 (2020), p. 160503.

[39] Weilong Wang, Kiyoshi Tamaki, and Marcos Curty. "Measurement-device-independent quantum key distribution with leaky sources". *Scientific Reports* 11.1 (2021), p. 1678.

[40] Wei Li and Shengmei Zhao. "Security research on practical measurement-device-independent quantum key distribution". *Phys. Rev. A* 106.4 (2022), p. 042445.

[41] Hoi-Kwong Lo, Marcos Curty, and Bing Qi. "Measurement-Device-Independent Quantum Key Distribution". *Phys. Rev. Lett.* 108.13 (2012), p. 130503.

[42] Hua-Lei Yin and Yao Fu. "Measurement-Device-Independent Twin-Field Quantum Key Distribution". *Scientific Reports* 9.1 (2019), p. 3045.

[43] Marco Avesani et al. "Source-device-independent heterodyne-based quantum random number generator at 17 Gbps". *Nature Communications* 9.1 (2018), p. 5365.

[44] D. G. Marangon, G. Vallone, and P. Villoresi. "Source-Device-Independent Ultrafast Quantum Random Number Generation". *Phys. Rev. Lett.* 118.6 (2017), p. 060503.

[45] P. R. Smith et al. "Simple source device-independent continuous-variable quantum random number generator". *Phys. Rev. A* 99.6 (2019), p. 062326.

[46] Jialin Cheng et al. "Mutually testing source-device-independent quantum random number generator". *Photon. Res.* 10.3 (2022), pp. 646–652.

[47] Jakub J. Borkala et al. "Device-Independent Certification of Maximal Randomness from Pure Entangled Two-Qutrit States Using Non-Projective Measurements". *Entropy* 24.3 (2022).

[48] Iris Agresti et al. "Experimental device-independent certified randomness generation with an instrumental causal structure". *Communications Physics* 3.1 (2020), p. 110.

[49] Antonio Acín et al. "Device-Independent Security of Quantum Cryptography against Collective Attacks". *Phys. Rev. Lett.* 98.23 (2007), p. 230501.

[50] Stefano Pironio et al. "Device-independent quantum key distribution secure against collective attacks". *New Journal of Physics* 11.4 (2009), p. 045021.

[51] Victor Zapatero et al. "Advances in device-independent quantum key distribution". *npj Quantum Information* 9.1 (2023), p. 10.

[52] Lluís Masanes, Stefano Pironio, and Antonio Acín. "Secure device-independent quantum key distribution with causally independent measurement devices". *Nature Communications* 2.1 (2011), p. 238.

[53] Nicolas Brunner et al. "Bell nonlocality". *Rev. Mod. Phys.* 86.2 (2014), pp. 419–478.

[54] Simon Storz et al. "Loophole-free Bell inequality violation with superconducting circuits". *Nature* 617.7960 (2023), pp. 265–270.

[55] B. Hensen et al. "Loophole-free Bell inequality violation using electron spins separated by 1.3 kilometres". *Nature* 526.7575 (2015), pp. 682–686.

[56] Marissa Giustina et al. "Significant-Loophole-Free Test of Bell's Theorem with Entangled Photons". *Phys. Rev. Lett.* 115.25 (2015), p. 250401.

[57] Gonzalo Carvacho et al. "Postselection-Loophole-Free Bell Test Over an Installed Optical Fiber Network". *Phys. Rev. Lett.* 115.3 (2015), p. 030503.

[58] R. Colbeck. *Quantum And Relativistic Protocols For Secure Multi-Party Computation*. Ph.D. Thesis, University of Cambridge. 2009.

[59] Dominic Mayers and Andrew Yao. "Quantum Cryptography with Imperfect Apparatus". *arXiv e-prints* (1998), quant–ph/9809039.

[60] Marco Avesani et al. "Semi-Device-Independent Heterodyne-Based Quantum Random-Number Generator". *Phys. Rev. Appl.* 15.3 (2021), p. 034034.

[61] Davide Rusca et al. "Self-testing quantum random-number generator based on an energy bound". *Phys. Rev. A* 100.6 (2019), p. 062338.

[62] Thomas Van Himbeeck and Stefano Pironio. *Correlations and randomness generation based on energy constraints*. 2019. arXiv: `1905.09117 [quant-ph]`.

[63] Ming-Han Li et al. "Experimental Realization of Device-Independent Quantum Randomness Expansion". *Phys. Rev. Lett.* 126.5 (2021), p. 050503.

[64] Hamid Tebyanian et al. "Semi-device independent randomness generation based on quantum state's indistinguishability". *Quantum Science and Technology* 6.4 (2021), p. 045026.

[65] Thomas Van Himbeeck et al. "Semi-device-independent framework based on natural physical assumptions". *Quantum* 1 (2017), p. 33.

[66] T. Lunghi et al. "Self-Testing Quantum Random Number Generator". *Phys. Rev. Lett.* 114.15 (2015), p. 150501.

[67] Piotr Mironowicz et al. "Quantum randomness protected against detection loophole attacks". *Quantum Information Processing* 20 (2021), p. 39.

[68] Jonatan Bohr Brask et al. "Megahertz-Rate Semi-Device-Independent Quantum Random Number Generators Based on Unambiguous State Discrimination". *Phys. Rev. Applied* 7.5 (2017), p. 054018.

[69] Stephen Boyd and Lieven Vandenberghe. *Convex Optimisation*. Cambridge University Press, 2004.

[70] Miguel Navascués, Sukhbinder Singh, and Antonio Acín. "Connector Tensor Networks: A Renormalization-Type Approach to Quantum Certification". *Phys. Rev. X* 10.2 (2020), p. 021064.

[71] MOSEK APS. "MOSEK Optimizer API for Python (Release 10.0.45)" (2023).

[72] Brendan O'Donoghue et al. "Conic Optimization via Operator Splitting and Homogeneous Self-Dual Embedding". *Journal of Optimization Theory and Applications* 169.3 (2016), pp. 1042–1068.

[73] Guido Van Rossum and Fred L Drake Jr. *Python reference manual*. Centrum voor Wiskunde en Informatica Amsterdam, 1995.

[74] Steven Diamond and Stephen Boyd. "CVXPY: A Python-embedded modeling language for convex optimization". *Journal of Machine Learning Research* 17.83 (2016), pp. 1–5.

[75] Akshay Agrawal et al. "A rewriting system for convex optimization problems". *Journal of Control and Decision* 5.1 (2018), pp. 42–60.

[76] Yurii Nesterov and Arkadii Nemirovskii. *Interior-Point Ploynomial Algorithms in Convex Programming*. Society for Industrial and Applied Mathematics, 1993.

[77] Carl W. Helstrom. "Quantum detection and estimation theory". *Journal of Statistical Physics* 1.2 (1969), pp. 231–252.

[78] Stephen M. Barnett and Erling Riis. "Experimental demonstration of polarization discrimination at the Helstrom bound". *Journal of Modern Optics* 44.6 (1997), pp. 1061–1064.

[79] Igor D Ivanovic. "How to differentiate between non-orthogonal states". *Physics Letters A* 123.6 (1987), pp. 257–259.

[80] Dennis Dieks. "Overlap and distinguishability of quantum states". *Physics Letters A* 126.5 (1988), pp. 303–306.

[81] Asher Peres. "How to differentiate between non-orthogonal states". *Physics Letters A* 128.1 (1988), p. 19.

[82] Roger B. M. Clarke et al. "Experimental demonstration of optimal unambiguous state discrimination". *Phys. Rev. A* 63.4 (2001), p. 040305.

[83] Hanwool Lee et al. "Maximum-confidence measurement for qubit states". *Phys. Rev. A* 106.3 (2022), p. 032422.

[84] W. K. Wootters and W. H. Zurek. "A Single Quantum Cannot Be Cloned". *Nature* 299.5886 (1982), p. 802.

[85] N Gisin. "Quantum cloning without signaling". *Physics Letters A* 242.1 (1998), pp. 1–3.

[86] Charles H. Bennett and Gilles Brassard. "Quantum cryptography: Public key distribution and coin tossing". *Proceedings of IEEE International Conference on Computers, Systems, and Signal Processing, Bangalore, India (IEEE, New York, 1984), pp. 175–179.* (1984).

[87] Charles H. Bennett, Gilles Brassard, and N. David Mermin. "Quantum Cryptography without Bell's Theorem". *Phys. Rev. Lett.* 68.5 (1992), pp. 557–559.

[88] H. Barnum and E. Knill. "Reversing quantum dynamics with near-optimal quantum and classical fidelity". *Journal of Mathematical Physics* 43.5 (2002), pp. 2097–2106. eprint: `https://aip.scitation.org/doi/pdf/10.1063/1.1459754`.

[89] Charles H. Bennett et al. "Quantum nonlocality without entanglement". *Phys. Rev. A* 59 (2 1999), pp. 1070–1091.

[90] William Matthews, Stephanie Wehner, and Andreas Winter. "Distinguishability of Quantum States Under Restricted Families of Measurements with an Application to Quantum Data Hiding". *Communications in Mathematical Physics* 291.3 (2009), pp. 813–843.

[91] D Spehner and M Orszag. "Geometric quantum discord with Bures distance". *New Journal of Physics* 15.10 (2013), p. 103001.

[92] Dominique Spehner. "Quantum Correlations and Distinguishability of Quantum States". *Journal of Mathematical Physics* 55.7 (2014), p. 075211.

[93] Anthony Chefles. "Quantum State Discrimination". *Contemp. Phys.* 41.6 (2000), pp. 401–424.

[94] Janos A. Bergou, Ulrike Herzog, and Mark Hillery. "11 Discrimination of Quantum States". *Quantum State Estimation*. Ed. by Matteo Paris and Jaroslav Rehacek. Lecture Notes in Physics. Berlin, Heidelberg: Springer Berlin Heidelberg, 2004, pp. 417–465.

[95] Janos A Bergou. "Quantum state discrimination and selected applications". *Journal of Physics: Conference Series* 84 (2007), p. 012001.

[96] Janos A. Bergou. "Discrimination of Quantum States". *J. Mod. Opt.* 57.3 (2010), pp. 160–180.

[97] Joonwoo Bae and Leong-Chuan Kwek. "Quantum state discrimination and its applications". *Journal of Physics A: Mathematical and Theoretical* 48.8 (2015), p. 083001.

[98] Carl W. Helstrom. "Detection theory and quantum mechanics". *Information and Control* 10.3 (1967), pp. 254–291.

[99] Carl W. Helstrom. "Detection theory and quantum mechanics (II)". *Information and Control* 13.2 (1968), pp. 156–171.

[100] Stephen M. Barnett, John Jeffers, and David T. Pegg. "Quantum Retrodiction: Foundations and Controversies". *Symmetry* 13.4 (2021).

[101] Sarah Croke et al. "Maximum Confidence Quantum Measurements". *Phys. Rev. Lett.* 96.7 (2006), p. 070401.

[102] Ulrike Herzog. "Discrimination of two mixed quantum states with maximum confidence and minimum probability of inconclusive results". *Phys. Rev. A* 79.3 (2009), p. 032323.

[103] Ulrike Herzog. "Optimized maximum-confidence discrimination of $N$ mixed quantum states and application to symmetric states". *Phys. Rev. A* 85.3 (2012), p. 032312.

[104] Ulrike Herzog. "Optimal measurements for the discrimination of quantum states with a fixed rate of inconclusive results". *Phys. Rev. A* 91.4 (2015), p. 042338.

[105] N. R. Kenbaev and D. A. Kronberg. "Quantum postselective measurements: Sufficient condition for overcoming the Holevo bound and the role of max-relative entropy". *Phys. Rev. A* 105.1 (2022), p. 012609.

[106] E. Bagan et al. "Optimal discrimination of quantum states with a fixed rate of inconclusive outcomes". *Phys. Rev. A* 86.4 (2012), p. 040303.

[107] Matthieu E. Deconinck and Barbara M. Terhal. "Qubit state discrimination". *Phys. Rev. A* 81.6 (2010), p. 062304.

[108] Joonwoo Bae. "Structure of minimum-error quantum state discrimination". *New Journal of Physics* 15.7 (2013), p. 073037.

[109] Donghoon Ha and Younghun Kwon. "Complete analysis for three-qubit mixed-state discrimination". *Phys. Rev. A* 87.6 (2013), p. 062302.

[110] ALAN JEFFREY and YUSUKE KATO. "Liapunov's Direct Method in Stability Problems for Semilinear and Quasilinear Hyperbolic Systems". *Journal of Mathematics and Mechanics* 18.7 (1969), pp. 659–682.

[111] H.W. Kuhn and A.W. Tucker. "Nonlinear Programming". *Proceedings of the 2nd Berkeley Symposium on Mathematics* (1969), pp. 481–492.

[112] Richard W. Cottle. "Linear complementarity problemLinear Complementarity Problem". *Encyclopedia of Optimization*. Ed. by Christodoulos A. Floudas and Panos M. Pardalos. Boston, MA: Springer US, 2009, pp. 1873–1878.

[113] Y.C. Eldar, A. Megretski, and G.C. Verghese. "Optimal detection of symmetric mixed quantum states". *IEEE Transactions on Information Theory* 50.6 (2004), pp. 1198–1207.

[114] Joseph M. Renes et al. "Symmetric informationally complete quantum measurements". *Journal of Mathematical Physics* 45.6 (2004), pp. 2171–2180. eprint: https://doi.org/10.1063/1.1737053.

[115] Artur K. Ekert. "Quantum Cryptography Based on Bell's Theorem". *Phys. Rev. Lett.* 67.6 (1991), pp. 661–663.

[116] S. Pironio et al. "Random numbers certified by Bell's theorem". *Nature* 464 (2010), p. 1021.

[117] Antonio Acín and Lluis Masanes. "Certified randomness in quantum physics". *Nature* 540.7632 (2016), pp. 213–219.

[118] Joseph Bowles et al. "Device-Independent Entanglement Certification of All Entangled States". *Phys. Rev. Lett.* 121.18 (2018), p. 180503.

[119] Joonwoo Bae and Antonio Acín. "Asymptotic Quantum Cloning Is State Estimation". *Phys. Rev. Lett.* 97.3 (2006), p. 030402.

[120] Won-Young Hwang. "Helstrom theorem from the no-signaling condition". *Phys. Rev. A* 71.6 (2005), p. 062315.

[121] Joonwoo Bae, Won-Young Hwang, and Yeong-Deok Han. "No-Signaling Principle Can Determine Optimal Quantum State Discrimination". *Phys. Rev. Lett.* 107.17 (2011), p. 170403.

[122] Armin Tavakoli. "Semi-Device-Independent Framework Based on Restricted Distrust in Prepare-and-Measure Experiments". *Phys. Rev. Lett.* 126.21 (2021), p. 210503.

[123] Joonwoo Bae and Won-Young Hwang. "Minimum-error discrimination of qubit states: Methods, solutions, and properties". *Phys. Rev. A* 87.1 (2013), p. 012334.

[124] Graeme Weir et al. "Optimal measurement strategies for the trine states with arbitrary prior probabilities". *Quantum Science and Technology* 3.3 (2018), p. 035003.

[125] "Quantum Theory: Concepts and Methods". *Fundemental Theories of Physics (Springler, Dordrecht, Netherlands)* (1995).

[126] Matteo Lostaglio and Gabriel Senno. "Contextual advantage for state-dependent cloning". *Quantum* 4 (2020), p. 258.

[127] Farid Shahandeh. *Quantum computational advantage implies contextuality*. 2021. arXiv: 2112.00024 [quant-ph].

[128] David Schmid et al. *A structure theorem for generalized-noncontextual ontological models*. 2020. arXiv: 2005.07161 [quant-ph].

[129] M. S. Leifer and O. J. E. Maroney. "Maximally Epistemic Interpretations of the Quantum State and Contextuality". *Phys. Rev. Lett.* 110.12 (2013), p. 120401.

[130] H. M. Wiseman, S. J. Jones, and A. C. Doherty. "Steering, Entanglement, Nonlocality, and the Einstein-Podolsky-Rosen Paradox". *Phys. Rev. Lett.* 98.14 (2007), p. 140402.

[131] Matteo Lostaglio and Gabriel Senno. "Contextual advantage for state-dependent cloning". *Quantum* 4 (), p. 258.

[132] Seth Lloyd et al. "Quantum embeddings for machine learning" (2020). arXiv: 2001.03622 [quant-ph].

[133] Mark Howard et al. "Contextuality supplies the 'magic'for quantum computation". *Nature* 510.7505 (2014), pp. 351–355.

[134] Helle Bechmann-Pasquinucci and Asher Peres. "Quantum Cryptography with 3-State Systems". *Phys. Rev. Lett.* 85.15 (2000), pp. 3313–3316.

[135] Karol Horodecki et al. *Contextuality offers device-independent security*. 2010. arXiv: 1006.0468 [quant-ph].

[136] Alastair A. Abbott et al. "Strong Kochen-Specker theorem and incomputability of quantum randomness". *Phys. Rev. A* 86.6 (2012), p. 062109.

[137] André Chailloux et al. "Optimal bounds for parity-oblivious random access codes". *New Journal of Physics* 18.4 (2016), p. 045003.

[138] Alexander A. Klyachko et al. "Simple Test for Hidden Variables in Spin-1 Systems". *Phys. Rev. Lett.* 101.2 (2008), p. 020403.

[139] Adán Cabello et al. "Simple Hardy-Like Proof of Quantum Contextuality". *Phys. Rev. Lett.* 111.18 (2013), p. 180404.

[140] Robert W. Spekkens et al. "Preparation Contextuality Powers Parity-Oblivious Multiplexing". *Phys. Rev. Lett.* 102.1 (2009), p. 010401.

[141] Shouvik Ghorai and A. K. Pan. "Optimal quantum preparation contextuality in an $n$-bit parity-oblivious multiplexing task". *Phys. Rev. A* 98.3 (2018), p. 032110.

[142] Andris Ambainis et al. "Parity oblivious d-level random access codes and class of noncontextuality inequalities". *Quantum Information Processing* 18.4 (2019), p. 111.

[143] Debashis Saha, Paweł Horodecki, and Marcin Pawłowski. "State independent contextuality advances one-way communication". *New Journal of Physics* 21.9 (2019), p. 093057.

[144] Alley Hameedi et al. "Communication Games Reveal Preparation Contextuality". *Phys. Rev. Lett.* 119.22 (2017), p. 220402.

[145] Debashis Saha and Anubhav Chaturvedi. "Preparation contextuality as an essential feature underlying quantum communication advantage". *Phys. Rev. A* 100.2 (2019), p. 022108.

[146] Manik Banik et al. "Limited preparation contextuality in quantum theory and its relation to the Cirel'son bound". *Phys. Rev. A* 92.3 (2015), p. 030103.

[147] Anirudh Krishna, Robert W Spekkens, and Elie Wolfe. "Deriving robust noncontextuality inequalities from algebraic proofs of the Kochen–Specker theorem: the Peres–Mermin square". *New Journal of Physics* 19.12 (2017), p. 123031.

[148] Ravi Kunjwal and Robert W. Spekkens. "From the Kochen-Specker Theorem to Noncontextuality Inequalities without Assuming Determinism". *Phys. Rev. Lett.* 115.11 (2015), p. 110403.

[149] Ravi Kunjwal and Robert W. Spekkens. "From statistical proofs of the Kochen-Specker theorem to noise-robust noncontextuality inequalities". *Phys. Rev. A* 97.5 (2018), p. 052110.

[150] Michael D. Mazurek et al. "An experimental test of noncontextuality without unphysical idealizations". *Nature Communications* 7.1 (2016), ncomms11780.

[151] Matthew F. Pusey. "Robust preparation noncontextuality inequalities in the simplest scenario". *Phys. Rev. A* 98.2 (2018), p. 022112.

[152] David Schmid, Robert W. Spekkens, and Elie Wolfe. "All the noncontextuality inequalities for arbitrary prepare-and-measure experiments with respect to any fixed set of operational equivalences". *Phys. Rev. A* 97.6 (2018), p. 062103.

[153] Joonwoo Bae, Dai-Gyoung Kim, and Leong-Chuan Kwek. "Structure of Optimal State Discrimination in Generalized Probabilistic Theories". *Entropy* 18.2 (2016).

[154] Gen Kimura, Takayuki Miyadera, and Hideki Imai. "Optimal state discrimination in general probabilistic theories". *Phys. Rev. A* 79.6 (2009), p. 062306.

[155] Elena R. Loubenets. "General lower and upper bounds under minimum-error quantum state discrimination". *Phys. Rev. A* 105.3 (2022), p. 032410.

[156] Ulrike Herzog. "Minimum-error discrimination between a pure and a mixed two-qubit state". *Journal of Optics B: Quantum and Semiclassical Optics* 6.3 (2004), S24–S28.

[157] O. Jiménez et al. "Maximum-confidence discrimination among symmetric qudit states". *Phys. Rev. A* 84.6 (2011), p. 062315.

[158] Daowen Qiu. "Minimum-error discrimination between mixed quantum states". *Phys. Rev. A* 77.1 (2008), p. 012328.

[159] Matthias Kleinmann, Hermann Kampermann, and Dagmar Bruß. "Unambiguous discrimination of mixed quantum states: Optimal solution and case study". *Phys. Rev. A* 81.2 (2010), p. 020304.

[160] R. Salazar and A. Delgado. "Quantum tomography via unambiguous state discrimination". *Phys. Rev. A* 86.1 (2012), p. 012118.

[161] Robert W. Spekkens. "Negativity and Contextuality are Equivalent Notions of Nonclassicality". *Phys. Rev. Lett.* 101 (2008), p. 020401.

[162] Brian Hayes. "Randomness as a resource". *American Scientist* 89.4 (2001), pp. 300–304.

[163] B. G. Christensen et al. "Detection-Loophole-Free Test of Quantum Nonlocality, and Applications". *Phys. Rev. Lett.* 111.13 (2013), p. 130406.

[164] Yang Liu et al. "Device-independent quantum random-number generation". *Nature* 562.7728 (2018), pp. 548–551.

[165] P. Bierhorst et al. "Experimentally Generated Randomness Certified by the Impossibility of Superluminal Signals". *Nature* 556 (2018), pp. 223–226.

[166] Lynden K. Shalm et al. "Device-independent randomness expansion with entangled photons". *Nature Physics* 17.4 (2021), pp. 452–456.

[167] Wen-Zhao Liu et al. “Device-independent randomness expansion against quantum side information”. *Nature Physics* 17.4 (2021), pp. 448–451.

[168] H.-W. Li et al. “Semi-device-independent random-number expansion without entanglement”. *Phys. Rev. A* 84.3 (2011), p. 034301.

[169] T. Michel et al. “Real-Time Source Independent Quantum Random Number Generator with Squeezed States”. *arXiv:1903.01071* (2019).

[170] David Drahi et al. “Certified Quantum Random Numbers from Untrusted Light”. *Phys. Rev. X* 10 (2020), p. 041048.

[171] G. Vallone et al. “Quantum randomness certified by the uncertainty principle”. *Phys. Rev. A* 90.5 (2014), p. 052327.

[172] Z. Cao, H. Zhou, and X. Ma. “Loss-tolerant measurement-device-independent quantum random number generation”. *New J. Phys.* 17.12 (2015), p. 125011.

[173] Z. Cao et al. “Source-Independent Quantum Random Number Generation”. *Phys. Rev. X* 6.1 (2016), p. 011020.

[174] F Xu, J. H. Shapiro, and F. N. C. Wong. “Experimental fast quantum random number generation using high-dimensional entanglement with entropy monitoring”. *Optica* 3.11 (2016), pp. 1266–1269.

[175] C. Budroni et al. *Quantum Contextuality*. 2021. arXiv: `2102.13036 [quant-ph]`.

[176] C. W. Helstrom. *Quantum Detection and Estimation Theory*. Academic Press, 1976.

[177] Robert Konig, Renato Renner, and Christian Schaffner. “The Operational Meaning of Min- and Max-Entropy”. *IEEE Transactions on Information Theory* 55 (2009), pp. 4337–4347.

[178] Carles Roch i Carceller et al. “Supplemental Material for: Quantum vs. noncontextual semi-device-independent randomness certification, which includes [180, 181]” (2022).

[179] Marie Ioannou, Jonatan Bohr Brask, and Nicolas Brunner. “Upper bound on certifiable randomness from a quantum black-box device”. *Phys. Rev. A* 99 (2019), p. 052338.

[180] Jean-Daniel Bancal, Lana Sheridan, and Valerio Scarani. “More randomness from the same data”. *New Journal of Physics* 16.3 (2014), p. 033011.

[181] Armin Tavakoli et al. “Self-testing nonprojective quantum measurements in prepare-and-measure experiments”. *Science Advances* 16 (2020), eaaw6664.

[182] Bruce Schneier Niels Ferguson and Tadayoshi Kohno. *Cryptography engineering: design principles and practical applications*. John Wiley and Sons, 2011.

[183] Kyungroul Lee and Manhee Lee. "True Random Number Generator (TRNG) Utilizing FM Radio Signals for Mobile and Embedded Devices in Multi-Access Edge Computing". *Sensors* 19.19 (2019).

[184] Koichi Miyamoto and Kenji Shiohara. "Reduction of qubits in a quantum algorithm for Monte Carlo simulation by a pseudo-random-number generator". *Phys. Rev. A* 102.2 (2020), p. 022424.

[185] Montanaro Ashley. "Quantum speedup of Monte Carlo methods". *Proc. R. Soc. A.* 471.2181 (2015).

[186] Radomir Stevanović et al. "Quantum Random Bit Generator Service for Monte Carlo and Other Stochastic Simulations". *Large-Scale Scientific Computing*. Ed. by Ivan Lirkov, Svetozar Margenov, and Jerzy Waśniewski. Berlin, Heidelberg: Springer Berlin Heidelberg, 2008, pp. 508–515.

[187] Dario Ghersi, Abhishek Parakh, and Mihaly Mezei. "Comparison of a quantum random number generator with pseudorandom number generators for their use in molecular Monte Carlo simulations". *Journal of Computational Chemistry* 38.31 (2017), pp. 2713–2720. eprint: `https://onlinelibrary.wiley.com/doi/pdf/10.1002/jcc.25065`.

[188] Grangier P. and Auffèves A. Phil. Trans. R. Soc. A., 2018.

[189] Xiongfeng Ma et al. "Quantum random number generation". *npj Quantum Information* 2.1 (2016), p. 16021.

[190] Miguel Herrero-Collantes and Juan Carlos Garcia-Escartin. "Quantum random number generators". *Rev. Mod. Phys.* 89.1 (2017), p. 015004.

[191] Felix Bischof, Hermann Kampermann, and Dagmar Bruß. "Measurement-device-independent randomness generation with arbitrary quantum states". *Phys. Rev. A* 95.6 (2017), p. 062305.

[192] Elsa Passaro et al. "Optimal randomness certification in the quantum steering and prepare-and-measure scenarios". *New Journal of Physics* 17.11 (2015), p. 113010.

[193] Matej Pivoluska et al. "Semi-device-independent random number generation with flexible assumptions". *npj Quantum Information* 7.1 (2021), p. 50.

[194] Thomas Van Himbeeck and Stefano Pironio. *Correlations and randomness generation based on energy constraints*. 2019. arXiv: 1905.09117 [quant-ph].

[195] Won-Young Hwang and Joonwoo Bae. "Minimum-error state discrimination constrained by the no-signaling principle". *Journal of Mathematical Physics* 51.2 (2010), p. 022202. eprint: https://doi.org/10.1063/1.3298647.

[196] Erika Andersson et al. "Minimum-error discrimination between three mirror-symmetric states". *Phys. Rev. A* 65.5 (2002), p. 052308.

[197] Rotem Arnon-Friedman et al. "Practical device-independent quantum cryptography via entropy accumulation". *Nature Communications* 9.1 (2018), p. 459.

[198] Wassily Hoeffding. "Probability Inequalities for Sums of Bounded Random Variables". *Journal of the American Statistical Association* 58.301 (1963), pp. 13–30.

[199] Marco Tomamichel, Roger Colbeck, and Renato Renner. "A Fully Quantum Asymptotic Equipartition Property". *IEEE Transactions on Information Theory* 55.12 (2009), pp. 5840–5847.

[200] Thomas Van Himbeeck et al. "Semi-device-independent framework based on natural physical assumptions". *Quantum* 1 (2017), p. 33.

[201] Shashank Gupta et al. "Quantum Contextuality Provides Communication Complexity Advantage". *Phys. Rev. Lett.* 130.8 (2023), p. 080802.

[202] Jaskaran Singh, Kishor Bharti, and Arvind. "Quantum key distribution protocol based on contextuality monogamy". *Phys. Rev. A* 95.6 (2017), p. 062333.

[203] Miguel Navascués, Stefano Pironio, and Antonio Acín. "Bounding the Set of Quantum Correlations". *Phys. Rev. Lett.* 98.1 (2007), p. 010401.

[204] 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 (2008), p. 073013.

[205] Peter J. Brown, Sammy Ragy, and Roger Colbeck. "A Framework for Quantum-Secure Device-Independent Randomness Expansion". *IEEE Transactions on Information Theory* 66.5 (2020), pp. 2964–2987.

[206] Joseph Bowles, Flavio Baccari, and Alexia Salavrakos. "Bounding sets of sequential quantum correlations and device-independent randomness certification". *Quantum* 4 (2020), p. 344.

[207] Yun Zhi Law et al. “Quantum randomness extraction for various levels of characterization of the devices”. *Journal of Physics A: Mathematical and Theoretical* 47.42 (2014), p. 424028.

[208] Youwang Xiao et al. “Device-independent randomness based on a tight upper bound of the maximal quantum value of chained inequality”. *Phys. Rev. A* 107.5 (2023), p. 052415.