De l'utilité des petits états quantiques : applications à l'optique linéaire et à la vérification de la position

Andrea Olivo

université PARIS-SACLAY

Leveraging small quantum states:
applications to linear optics and position verification

De l'utilité des petits états quantiques :
applications à l'optique linéaire et à la vérification de la position

Thèse de doctorat de l'université Paris-Saclay

École doctorale n°572 : ondes et matière (EDOM)
Spécialité de doctorat : Physique
Graduate School : Physique, Référent : Faculté des sciences d'Orsay

Thèse préparée dans les unités de recherche : Laboratoire de physique des gaz et des plasmas (Université Paris-Saclay, CNRS) sous la direction de Jacques ROBERT, professeur, le co-encadrement de Frédéric GROSSHANS, chargé de recherche et André CHAILLOUX, chargé de recherche

Thèse soutenue à Paris-Saclay, le 15 juin 2022, par

Andrea OLIVO

Composition du jury

Rosa TUALLE-BROURI Professeure, Université Paris-Saclay, France	Présidente
Matthias CHRISTANDL Professeur, Københavns Universitet, Denmark	Rapporteur et examinateur
Peter VAN LOOCK Professeur, Johannes Gutenberg-Universität Mainz, Allemagne	Rapporteur et examinateur
Anne BROADBENT Professeure agrégée, University of Ottawa, Canada	Examinatrice
Mercedes GIMENO-SEGOVIA PhD, PsiQuantum, California	Examinatrice
Jacques ROBERT Professeur, Université Paris-Saclay, France	Directeur de thèse

THESE DE DOCTORAT
NNT : 2022UPASP055

université PARIS-SACLAY

ÉCOLE DOCTORALE
Ondes et matière (EDOM)

Titre : De l'utilité des petits états quantiques : applications à l'optique linéaire et à la vérification de la position

Mots clés : optique linéaire, cryptographie quantique, optimisation

Résumé : Cette thèse vise à étudier, en employant des méthodes tant analytiques que numériques, l'efficacité de petits systèmes quantiques pour deux applications dans le domaine de l'information quantique. La première en tant qu'états auxiliaires pour une mesure de Bell en optique linéaire, qui est une primitive d'une importance cruciale pour le calcul quantique avec des photons et pour la réalisation de l'internet quantique à venir. La deuxième en tant que ressources intriquées que des attaquants peuvent utiliser pour casser une primitive cryptographique appelé «vérification de la position», qui fait appel simultanément à des contraintes quantiques et relativistes pour permettre un nouveau type d'authentification.

Title : Leveraging small quantum states : applications to linear optics and position verification

Keywords : linear optics, quantum cryptography, optimization

Abstract : This thesis work explores with both analytical and numerical methods the power of small quantum systems in two applications in the field of quantum information. The first is as auxiliary states in linear optical Bell measurement, which is a primitive of paramount importance for quantum computing with photons and the coming quantum internet. The second is as entangled resources which attackers can use to break a cryptographic primitive known as quantum position verification, which simultaneously leverages quantum and relativistic constraints to enable a new kind of authentication token.

The present work is the account of my research as a graduate student, in collaboration with my supervisors. The contribution of other researchers has been adequately acknowledged. The official duration of my studies ranged from November 2017 to June 2022. However, this span includes a six-months hiatus from August 2019 to January 2020, which I spent as an intern at the California-based quantum computing company PsiQuantum, as well as a two-months extension granted due to the Covid-19 pandemic. Other than at INRIA Paris and at the Laboratoire Aimé Cotton in Paris-Saclay University, I spent a large portion of my research days as a visitor of the LIP6 research group in Sorbonne Université, which I warmly thank for the hospitality. Finally, I acknowledge financial support from the ANR project ANR-16-CE39-0001 DEREC.

This thesis has been composed by myself and the work will not be submitted for any other degree or professional qualification. Appropriate credit has been given within the manuscript where reference has been made to the work of others. Chapter 2 and Chapter 3 contain the two main research topics as well an introduction of the literature on them. In particular we declare that, as far as we know, the following Sections contain original work:

- Sections 2.4 to 2.7, which partially correspond to work also published in our paper [OG18] in collaboration with my co-supervisors.
- Sections 3.3 to 3.6, which is partially covered by our preprint [Oli+20] in collaboration with my co-supervisors and Ulysse Chabaud.

To my ever-supportive family

Résumé du manuscrit

Les récentes avancées théoriques et technologiques permettant la manipulation de l'information quantique peuvent donner accès à une vaste gamme d'applications. Ils promettent d'améliorer les mesures de précision (*quantum metrology*), la sécurité des communications numériques (*quantum internet*) et la vitesse de certains algorithmes (*quantum computing*) – transformant ainsi plusieurs domaines de recherche. Cependant, on ne sait toujours pas avec certitude s'il sera possible d'obtenir un avantage quantique pour des problèmes pratiques avec les technologies disponibles aujourd'hui ou dans un futur proche.

Dans cette thèse de doctorat, nous étudions deux de ces applications : la mesure de Bell en optique linéaire, une tâche essentielle à la base de la plupart des protocoles quantiques, et la vérification de la position dans le cadre quantique (QPV), une primitive cryptographique – inefficace dans le cadre classique – qui permettrait d'utiliser sa position dans l'espace comme identifiant. Pour les deux tâches nous analysons dans quelle mesure l'utilisation de petits états quantiques auxiliaires, relativement simples à produire dans le laboratoire, peut influencer leurs performances. Dans le premier Chapitre, nous présentons une brève introduction aux concepts généraux de la mécanique quantique. Les Chapitres 2 et 3 commencent avec une introduction plus spécifique des sujets respectifs et une revue de l'état de l'art dans la littérature, suivie par l'exposition du travail original et une conclusion.

En optique classique, les outils et matériaux possédant une réponse linéaire par rapport à l'intensité de la lumière incidente sont de loin le plus étudiés et utilisés dans le laboratoire. Quand la lumière encode de l'information quantique, la restriction à ces matériaux (avec l'ajout des sources et détecteurs de photons) impose une structure dans l'espace des états réalisables tout aussi simple sur le plan théorique. Même si l'optique linéaire est en principe suffisante pour arriver au calcul quantique universel, sa nature probabiliste implique un coût très élevé. Il est donc intéressant de

se focaliser sur des tâches plus spécifiques pour réduire ce coût, comme la mesure de Bell analysée dans ce manuscrit.

Une *mesure de Bell non ambigüe* est la projection d'un état bipartite de deux qubits sur une base d'états maximalement intriqués. Dans le cas où seuls les photons à mesurer en entrée peuvent interférer dans le réseau optique, une borne supérieure à sa probabilité de succès de 50% est connue. Cependant, il est possible de profiter d'états auxiliaires préparés à l'avance pour améliorer la probabilité de réussite : nous exposons une revue de deux stratégies connues dans la littérature en les mettant dans un cadre commun. En ajoutant une restriction sur le type de réseau optique admis, nous prouvons une borne supérieure à la probabilité de succès qui dépend de la forme de l'état auxiliaire employé. Comme cette restriction est difficile à justifier d'un point de vue expérimental, nous employons une stratégie computationnelle pour explorer l'espace des réseau optiques génériques. On expose les difficultés rencontrées pour réduire la taille du problème de recherche d'une solution à un niveau raisonnable pour le nœud de calcul dont nous disposons. Enfin, nous présentons nos résultats, qui incluent la (re)découverte d'une stratégie "intermédiaire" à efficacité 62.5% et l'analyse de plusieurs types d'états auxiliaires.

La deuxième partie de notre travail concerne la possibilité de vérifier de manière cryptographique sécurisée la position d'un dispositif (ou d'une personne) dans l'espace. Dans un tel protocole, des vérificateurs $V_1, V_2, \dots$ envoient des messages $x_1, x_2, \dots$ vers la position P revendiquée par le prouveur, et ils reçoivent des réponses $f(x_1, x_2, \dots) = (y_1, y_2, \dots)$ de sa part. Les messages sont synchronisés de sorte qu'ils arrivent à P tous simultanément, et les temps d'arrivée des réponses sont enregistrés. Le protocole exploite la relativité restreinte (sous la forme du principe de *no-signaling*) pour borner la région autour de P d'où les réponses auraient pu être envoyées. Dans le cas classique, l'impossibilité pour ces protocoles d'atteindre un niveau de sécurité souhaitable est connue : il est toujours possible pour une coalition de agents malveillant, dont aucun entre eux se trouve à P, de se coordonner pour donner les bonnes réponses au bon moment. Cependant, si la fonction f ainsi que les messages peuvent être quantiques, la meilleure attaque connue demande aux attaquants une quantité de ressources exponentielle –

plus précisément, des états intriqués – par rapport à la taille des messages des vérificateurs. Nous formalisons un protocole simple de QPV paramétré par un angle θ, et nous développons un langage graphique pour caractériser ses attaques exactes. Avec une méthode numérique, nous trouvons de nouvelles attaques pour plusieurs angles, qui dépassent largement en efficacité ceux présents dans la littérature. En assouplissant la condition d'attaque exacte, nous trouvons que deux *ebit* partagés par qubit envoyé sont suffisant pour permettre aux adversaires de casser tous les angles θ avec une probabilité $> 99.5\%$, et nous discutons les implication expérimentales à court terme de ces attaques.

Contents

Résumé du manuscrit vi

1 Introduction **1**

1.1 A brief toolbox 3

1.1.1 The qubit 4

1.1.2 Occupation number representation 4

1.1.3 Second quantization 7

1.1.4 Bogoliubov transformations 8

1.1.5 State discrimination 9

2 Linear optical Bell measurement **13**

2.1 Brief history of optics 13

2.1.1 Why *linear* optics? 15

2.2 Quantum linear optics, formally 17

2.2.1 Encodings 17

2.2.2 Algebraic structure 19

2.2.3 Polynomial representation 22

2.2.4 Linear optical elements 22

2.2.5 Auxiliary states 28

2.2.6 Hong–Ou–Mandel effect 29

2.2.7 Linear optical network 31

2.2.8 Quantum computation with light 32

2.3 Bell measurement 36

2.3.1 Bell measurement in linear optics 38

2.3.2 A simple Bell analyzer 40

2.3.3 No-go theorem for perfect Bell analyzers 41
2.3.4 Calsamiglia–Lütkenhaus 1/2 upper bound 41
2.3.5 A $p_{\text{succ}} = 1/2$ analyzer: the Innsbruck scheme 45
2.3.6 Beating the 1/2 limit 46
2.3.7 Grice's approach 47
2.3.8 Ewert and van Loock's approach 50
2.4 A new polarization-preserving bound 55
2.4.1 Bound based on photon number 60
2.5 Computer to the rescue: an optimization approach 63
2.5.1 Overview 63
2.5.2 Symbolic computation 66
2.5.3 Function compilation 68
2.5.4 Optimizations and symmetries 69
2.5.5 Numerical optimization 73
2.6 Optimization results 84
2.6.1 Vacuum and eigenstates of n_h 84
2.6.2 Extra Bell pairs 86
2.6.3 Extra single photons 88
2.6.4 GHZ and W states 90
2.7 Summary and conclusion 94

3 Quantum position verification **99**
3.1 Introduction 100
3.1.1 General features 102
3.2 The PV protocol zoo 104
3.2.1 Proto-PV protocols: distance-bounding 105
3.2.2 PV from distance bounding: a no-go theorem 106
3.2.3 Malaney and Chandran *et al.*: a new hope 108
3.2.4 Kent, Munro, Spiller: first insecurity proof 109
3.2.5 Lau and Lo: going $D > 1$ and qutrit security proof . . 111
3.2.6 Buhrman *et al.*: good news and bad news 113
3.2.7 Doubly-exponential INQC attack 115
3.2.8 Beigi and König: a "just exponential" attack 119
3.2.9 Chakraborty–Leverrier: INQC in CH 121

3.2.10 Speelman: a polynomial attack 124
3.2.11 Gonzales-Chitambar: attacking a two-qubit protocol 126
3.2.12 Linear lower bounds 128
3.2.13 Recent results 132
3.3 Defining QPV_θ and its attack model 134
3.3.1 The protocol 134
3.3.2 Attack model 136
3.3.3 Circuit picture 138
3.4 Exact attacks against QPV_θ 141
3.4.1 Graphical language 144
3.4.2 Numerical optimization: a comeback 150
3.4.3 Circuit solutions 158
3.4.4 QPV_θ in the INQC picture 162
3.5 Approximate attacks 165
3.5.1 Figure of merit 165
3.5.2 Methods 166
3.5.3 Results 169
3.5.4 $QPV_{(n)}$: a better protocol with little effort? 172
3.6 Summary and conclusion 175

Bibliography **179**

Chapter 1

Introduction

For a physicist working at the end of the 1970s, the successes of quantum mechanics and its role in fundamental physics were indisputable—in fact, they were updated by the day. Just a decade before, Bell had "proven Einstein wrong" (quotes intentional) by showing its eponymous inequality, which at the time was in the process of being experimentally validated in favor of quantum mechanics. Meanwhile, the Standard Model was in active development, showing that quantum field theories were a promising road to the unification of fundamental particles and interactions. Quantum mechanics' role was becoming prominent outside of just fundamental research, too: technology had fully embraced its counterintuitive properties, enabling revolutionary applications the likes of lasers, magnetic resonance imaging and ever smaller integrated circuits.

By contrast, a researcher in computer science of the time could have been largely unaware of these ongoing successes, shielded by the thought that their field of research resided on a different abstraction level. They would have had good reasons for this: building on the work of, among others, Turing, Shannon and von Neumann, it had been made abundantly clear that it was possible to treat computation and information as purely mathematical theories, parting ways with their physical implementation. Over the years analog computers, digital ones, mechanical, electronic, and many other platforms were all shown to be Turing machines, all fundamentally

equally powerful. Besides, if they happened to work on cryptography, there was no shortage of big news; public-key cryptography had recently been discovered, a new way—counterintuitive in its own terms—of exchanging messages over public channels, and the advent of computers was giving a way of proving the usefulness of this and many other results (e.g. in complexity theory) in real-world applications. However, quantum mechanics was about to, if not revolutionize, at least shake this assumption from its foundations.

In a pioneering 1982 paper, Feynman noticed the apparent exponential complexity of simulating interacting quantum systems even approximately, thus challenging the notion that Turing machines could *efficiently* simulate any physical process [Fey82]. For dealing with a "probabilistic nature", he proposed, we need a "probabilistic computer"; but since classical probabilities would not do, it should be quantum mechanical. A model for a quantum Turing machine was developed [Ben80; Deu85], and in the following years it was shown to support (relative to an oracle) efficient algorithms for problems outside P (Deutsch–Josza [DJ92]), and BPP (Bernstein–Vazirani [BV97] and Simon [Sim97]). If any of these sounded rather contrived to the computer scientist of the time, Shor helped clear any doubt with his work on factorization [Sho94] and the possibility of fault-tolerant quantum computation [Sho96]. Meanwhile, it was being discovered that quantum phenomena could support a holy grail of cryptography: quantum key distribution (QKD [BB84]), an *information-theoretic* secure protocol to exchange keys over a public channel.

Forty years after Feynman's paper, it is clear that quantum information is going to stay. Its development promises to bring considerable advances in many areas of research, in academia as well as the industry. QKD has seeded the whole field of quantum cryptography and quantum communication. Applications of quantum computing will include the efficient simulation of quantum systems, a compelling tool for, among others, the development of new materials and a better understanding of many-body chemical processes. However, the journey of bringing universal, error-corrected quantum computers to light (which, as far as we know, most of the above requires) is far from over. Many quantum computing platforms have been

proposed, differing in architecture and in the physical implementations of the basic units of quantum information processing—qubits and quantum gates. Broadly speaking, the main challenge all platforms have to overcome [DiV00] is the design of physical qubits that can be well isolated from the environment, to preserve the fragile coherence of quantum information, whilst also providing a mechanism for qubit-qubit interactions, necessary for the implementation of logic gates.

We are not there yet. Today, the available platforms consist of less than a hundred noisy qubits, a situation that has been dubbed the "NISQ era" (acronym of *noisy intermediate-scale quantum*, [Pre18]). It sees theoreticians from one side, lowering the amount of resources needed for running useful algorithms, and experimentalists from the other, raising the capabilities of quantum hardware. This near-term quantum hardware could enable advanced sensors, breaking classical limits [DRC17; GLM11] and support the huge undertaking of building a *quantum internet* [WEH18]. As private companies join the race alongside academia, NISQ hardware is now available to researchers in the cloud (among others, [Xan; IBM]).

In this context, this thesis explores the capabilities of small, near-term quantum systems in two specific tasks. After a brief introduction in Section 1.1 below, we tackle the problem of enhancing the success probability of linear optical Bell measurement via entangled auxiliary states in Chapter 2. Then, in Chapter 3 we show how small entangled states can be used to attack a class of protocols implementing a cryptographic primitive known as *quantum position verification*. Both problems are presented and motivated in the respective introductions to the two Chapters.

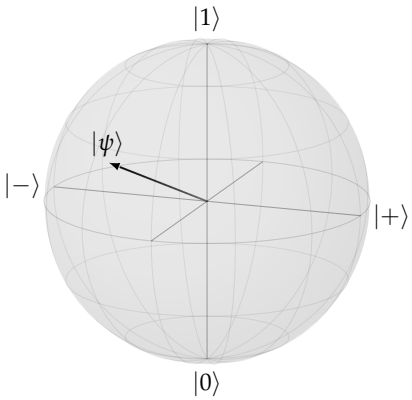
1.1 A brief toolbox

We assume the reader is familiar with the usual objects of quantum mechanics (QM) and their standard notation, such as pure ($|\psi\rangle$) and mixed (ρ) states, Hilbert spaces ($\mathcal{H}$), measurements of quantum systems and their effect on the wavefunction, tensor products ($\otimes$) as a mean to describe joint quantum systems and other basic notions which can be found in an introductory course. A (by no means complete) selection of textbooks on the topic is

Griffiths' *Introduction to quantum mechanics* [GS18], which provides a very general viewpoint on QM, Nielsen and Chuang's *quantum computation and quantum information* [NC10] and Wilde's *from classical to quantum Shannon theory* [Wil13], which focus more on quantum information (QI).

1.1.1 The qubit

The simplest quantum system used in quantum information is the *qubit*, which is a two-level state analogous to the classical bit. A useful and complete representation of the qubit's state space is the Bloch sphere:



Here, $|0\rangle$ and $|1\rangle$ form the *computational basis* and $|\pm\rangle = \frac{|0\rangle \pm |1\rangle}{\sqrt{2}}$ is the *Hadamard basis*, an example of a *superposition* of states in the computational basis. Pure states are found on the surface of the sphere, while the internal Bloch ball accommodates mixed states; the center is the maximally mixed state $\rho = \frac{1}{2} I$.

1.1.2 Occupation number representation

One of the nonclassical features of quantum systems that a quantum mechanical description of nature has to accommodate is the notion of *indistinguishability*. Two hydrogen atoms in the ground state are identical, as in: they are completely described by the very same set of quantum num-

bers.[1] Accounting for indistinguishability puts restrictions on the allowed quantum states for a collection of identical particles. As a consequence, writing the state in the computational basis is much less natural, and a better representation can be used.

Mathematically, the situation can be illustrated as follows. Given quantum systems A and B, the joint state of system A in state $|\psi\rangle$ and system B in state $|\phi\rangle$ is described by their tensor product:

$$|\Psi^{\text{dist}}\rangle_{\text{AB}} = |\psi\rangle_{\text{A}} \otimes |\phi\rangle_{\text{B}} \,. \tag{1.1}$$

If A and B are distinguishable, we should expect their joint state to be different if we swap the two systems. That is indeed the case: the tensor product is not commutative, so $|\phi\rangle_{\text{A}} \otimes |\psi\rangle_{\text{B}} \neq |\psi\rangle_{\text{A}} \otimes |\phi\rangle_{\text{B}}$. For identical systems, on the other hand, the joint state must be invariant (up to an unobservable global phase) under this exchange, as otherwise we could tell the systems apart. In three space dimensions, there are two[3] ways to do this:

$$|\Psi^{\text{id}}\rangle_{\text{AB}} = |\psi\rangle_{\text{A}} \otimes |\phi\rangle_{\text{B}} \pm |\phi\rangle_{\text{A}} \otimes |\psi\rangle_{\text{B}} \,, \tag{1.2}$$

depending on which sign $|\Psi^{\text{id}}\rangle$ acquires after the exchange. For identical systems, the operator exchanging the two systems commutes with the Hamiltonian (which is an observable itself), meaning that the sign choice does not change during the state's evolution. Two classes of quantum systems then arise:

- **Bosons** when choosing $+$, leading to a symmetric eq. (1.2);
- **Fermions**, choosing $-$, are instead antisymmetric under exchange.

[1] Borrowing a term from an unrelated[2] area of physics, they have "no hair". Even tracking the two atoms' worldlines backwards up to their creation can't always help us: due to the uncertainty principle, the lines have an intrinsic "fuzziness" which may lead to them switching places without us noticing [remark from GS18].

[2] Hopefully not for much longer

[3] In principle, the only requirement is that observables quantities are unchanged, which allows $|\Psi^{\text{id}}\rangle$ to acquire any global phase factor $e^{i\theta}$ after particle exchange. In three dimensions, however, the *spin-statistics theorem* from relativistic quantum mechanics [Pau40] effectively restricts the acquired phase after a second exchange to be 1, leaving $\theta = \pm\pi$. In 2D this result does not apply and θ is an additional continuous degree of freedom, which gives rise to exotic solutions called *anyons* [Wil82].

The most common example of bosons are photons, the protagonists of the first Section of this thesis (Chapter 2).

When working with k identical particles, the notation in eq. (1.2) quickly becomes cumbersome as k gets large. It is convenient in this case to switch to the *occupation number representation*. Given a basis $\{|e_i\rangle\}_{i=1}^m$ of the single particle Hilbert space $\mathcal{H}$,[4] due to indistinguishability we can only keep track of *how many* particles are in each state, not *which* particle. We can capture this property by using *Fock states*:

$$|n_1\, n_2 \ldots n_m\rangle \quad \text{with} \quad \sum_{i=1}^{m} n_i = k, \tag{1.3}$$

where n_i is the number of particles, or *occupation number*, in state $|e_i\rangle$. The single particle basis states $|e_i\rangle$ are called *modes* and they correspond to other degrees of freedom differentiating the particles (often location in space or time, or their spin). For fermions, all n_i are restricted to $\{0,1\}$: as a consequence of antisymmetrization, no two particles can occupy the same state as eq. (1.2) would just be identically zero. Fock states form a basis of the correct "symmetrized" $\mathcal{S}(\mathcal{H}^{\otimes k})$ or "antisymmetrized" $\mathcal{A}(\mathcal{H}^{\otimes k})$ Hilbert space. Their dimension is as follows:

- For fermions, it is the number of ways k particles can occupy k different modes out of m, giving:

$$\dim \mathcal{A}(\mathcal{H}^{\otimes k}) = \binom{m}{k} \tag{1.4}$$

- For bosons, several of the k particles can occupy the same mode. A standard combinatorial argument (*stars and bars*) gives:

$$\dim \mathcal{S}(\mathcal{H}^{\otimes k}) = \binom{m+k-1}{k}. \tag{1.5}$$

Both dimensions are reduced compared to the case of distinguishable particles, giving $\dim \mathcal{H}^{\otimes k} = m^k$. From now on, we focus on bosonic states.

[4] We assume the dimension of the single particle Hilbert space to be finite, as this is sufficient for our purpose. With due care, all the definitions generalize to infinite-dimensional systems.

1.1.3 Second quantization

The k-particle space in eq. (1.5) is convenient when dealing with evolutions which do not change the number of bosons, for instance the ones allowed by linear optics (which will be useful later). In full generality however, bosonic operators live in the *symmetric Fock space*

$$\mathcal{F}(\mathcal{H}) = \bigoplus_{k=0}^{\infty} \mathcal{S}(\mathcal{H}^{\otimes k}) = \mathbb{C} \oplus \mathcal{H} \oplus \mathcal{S}(\mathcal{H} \otimes \mathcal{H}) \oplus \dots \tag{1.6}$$

The most important bosonic operators acting in Fock space are the *creation* and *annihilation* operators, which add or remove a boson in mode i:

$$\begin{aligned} a_i^\dagger \left|n_1 \dots n_m\right\rangle &= \sqrt{n_i + 1} \left|n_1 \dots (n_i + 1) \dots n_m\right\rangle, \\ a_i \left|n_1 \dots n_m\right\rangle &= \sqrt{n_i} \left|n_1 \dots (n_i - 1) \dots n_m\right\rangle. \end{aligned} \tag{1.7}$$

Together, they provide a way to count the number of systems in mode i, as well as the total number of bosons:

$$\langle a_i^\dagger a_i \rangle = n_i, \qquad \left\langle \sum_i^m a_i^\dagger a_i \right\rangle = k. \tag{1.8}$$

From eq. (1.7), the *canonical commutation relations* can be derived:

$$\forall\, i, j \qquad \begin{aligned} \left[a_i, a_j^\dagger\right] &= \delta_{ij}, \\ \left[a_i, a_j\right] &= \left[a_i^\dagger, a_j^\dagger\right] = 0. \end{aligned} \tag{1.9}$$

Notice that eq. (1.7) also means that all Fock states can be built up from the vacuum by applying powers of the $a_i^\dagger$:

$$\left|n_1 \dots n_m\right\rangle = \prod_{i=1}^{m} \frac{(a_i^\dagger)^{n_i}}{\sqrt{n_i!}} \left|0\right\rangle, \tag{1.10}$$

while arbitrary superpositions (all bosonic quantum states) can be represented by a *polynomial* in the creation operators:

$$\left|\Psi\right\rangle = \sum_{n_1,\dots,n_m}^{\infty} \alpha_{n_1 \dots n_m} \left|n_1 \dots n_m\right\rangle = P_\psi(a_1^\dagger, \dots, a_m^\dagger) \left|0\right\rangle. \tag{1.11}$$

This representation, further developed in Section 2.2.3, is at the core of the notation used in the first part of this work.

When dealing with conceptually distinct groups of modes (e.g. the input and output modes of an m-port interferometer), in order to keep a clean notation we will make use of other letters than a for the bosonic operators. For instance, as it is common in the literature, we use $h_i^\dagger$ and $v_i^\dagger$ for the creation operators respectively of a horizontally and of a vertically polarized photon in the spatial mode indexed by i.

1.1.4 Bogoliubov transformations

In Hilbert space, quantum states evolve through the Hamiltonian, describing the dynamic of the system:

$$|\psi(t)\rangle = e^{-iHt}\,|\psi(0)\rangle = U(t)\,|\psi(0)\rangle\,. \tag{1.12}$$

In order to get meaningful information about a system, we often choose a convenient basis which diagonalizes the Hamiltonian. In the above construction of the creation and annihilation operators $a_i, a_i^\dagger$, we similarly had to fix, somewhat arbitrarily, a set of modes $|e_i\rangle$. How does the change of basis looks like from the point of view of the bosonic operators? Let us first introduce a widely-used vectorial notation:

$$\mathbf{a} = (a_1, \ldots, a_m)^\mathsf{T} \qquad \mathbf{a}^\dagger = (a_1^\dagger, \ldots, a_m^\dagger)^\mathsf{T} \tag{1.13}$$

A linear transformation of the modes can be written:

$$\begin{pmatrix} \mathbf{c} \\ \tilde{\mathbf{c}} \end{pmatrix} = T \begin{pmatrix} \mathbf{a} \\ \mathbf{a}^\dagger \end{pmatrix} = \begin{pmatrix} T_{00} & T_{01} \\ T_{10} & T_{11} \end{pmatrix} \begin{pmatrix} \mathbf{a} \\ \mathbf{a}^\dagger \end{pmatrix}, \tag{1.14}$$

where $\mathbf{c}$ and $\tilde{\mathbf{c}}$ are the creation and annihilation operators on a new set of modes.[5] Not all linear maps are physical: the commutation relations in eq. (1.9) have to be preserved.[6] A full derivation [Bog58] shows that T has

[5] The new mode operators c_j and $\tilde{c}_j$ are not necessarily mutually adjoint anymore, hence the difference in notation.

[6] This is reminiscent of the canonical transformation of coupled modes in classical mechanics [LL82], which preserve the *Poisson brackets*.

to obey:

$$T^{\mathsf{T}}\,\Omega\,T=\Omega,\qquad \Omega=\begin{pmatrix}0 & I\\ -I & 0\end{pmatrix}; \tag{1.15}$$

namely $T\in \mathrm{Sp}(2m,\mathbb{C})$, the symplectic group. Equation (1.14) is the general form of a *Bogoliubov-Valatin transformation* [Bog58; Val58]. In the special case of $\tilde{\mathbf{c}}=\mathbf{c}^\dagger$, eq. (1.15) implies:

$$\begin{aligned} &T_{00}=T_{11}^* \qquad T_{00}\,T_{01}^{\mathsf{T}} \quad \text{symmetric}\\ &T_{01}=T_{10}^* \qquad T_{00}\,T_{00}^\dagger - T_{01}\,T_{01}^\dagger = I, \end{aligned} \tag{1.16}$$

which, for example, include all *gaussian operations* from continuous variables quantum optics [Wee+12]. When $T_{01}=T_{10}=0$, we get operations which only map creation operators to creation operators and annihilation operators to annihilation operators, i.e. they conserve the number of bosons. In this case, the last of eq. (1.16) implies $T_{00}\in \mathrm{U}(m)$, the unitary group.

1.1.5 State discrimination

Suppose we are given a quantum system whose state is taken from a set $\{\rho_i\}$, and we are asked to determine the index i. Contrary to classical mechanics, it is a well known fact that this *quantum state discrimination* task can be performed perfectly if and only if the states are all orthogonal. The existence of non-orthogonal states is central to both information processing, where it enables advantages over many classical protocols, and foundations of quantum theory, e.g. playing a central role in ruling out "epistemic-only" interpretations of the wavefunction [PBR12]. A great deal of theoretical effort has been spent deriving upper and lower bounds on the optimal discrimination strategy in the general case; unfortunately, with the exception of simple systems (e.g. two-state discrimination, qubit states), exact analytical results are hard to come by. For a recent review, see [BK15]. Another interesting research direction looks at restrictions of QM with reduced power, where even some sets of orthogonal states are not perfectly distinguishable; this is the case for linear optics, which we discuss in Chapter 2. When dealing with non-perfect discrimination, the relevant parameter to optimize depends on

the application. For example, the discrimination strategy would be different if multiple copies of the state are available at once, as opposed to sequential *single-shot* measurements; and in the latter case, a further distinction could be made between the adaptive and nonadaptive setting. For our purposes, we are interested in measurements in which errors in the discrimination are not allowed: instead, we cope with non-perfect discrimination by adding a "failure" channel, yielding an indeterminate outcome. This setting is called *unambiguous discrimination*, and its main interest is in applications in which information loss is more prominent than errors. In general, this restriction leads to a lower overall success probability than in the ambiguous case (called *minimum-error discrimination*) [BC09].

Notation and definitions

In a basic setting, a source emits one of the states $\{\rho_i\}_{i=1}^n$ at random, each with probability q_i such that $\sum_i^n q_i = 1$. Often, we choose the uniform distribution, where $q_i = \frac{1}{n}\,\forall i$. The state of the source is the ensemble $\{q_i, \rho_i\}_i$ which corresponds to the mixed state:

$$\rho = \sum_i^n q_i \rho_i, \tag{1.17}$$

and the discrimination procedure can be described by a POVM, i.e. a set of positive operators $\{\Pi_j\}_{j=1}^m$ such that $\sum_j^m \Pi_j = I$ and for which

$$p(j|i) = \mathrm{Tr}\left[\Pi_j \rho_i\right] \tag{1.18}$$

is the probability of getting outcome j when the source outputs ρ_i. For minimum-error discrimination, we associate each outcome to one input, i.e. $m = n$. The overall success probability is:

$$p_{\mathrm{succ}}^{min\,err} = \sum_i q_i\, p(i|i) = \sum_i q_i\, \mathrm{Tr}\left[\Pi_i \rho_i\right]. \tag{1.19}$$

For unambiguous discrimination, we designate one of the outcomes to represent failure, such that $\Pi_{\mathrm{fail}} + \sum_i^n \Pi_i = I$. Defining the probability of success in this case requires a bit of care. In order to guarantee that the

outcome i can be unmistakably associated to state ρ_i, we need to impose:

$$p(i|j) = \mathrm{Tr}\left[\Pi_i \rho_j\right] = 0 \quad \forall j \notin \{i, \mathrm{fail}\}. \tag{1.20}$$

This cannot (in general) be satisfied for all i; however, a sufficient condition for pure states is their linear independence. The success probability is then defined the same way as eq. (1.19), summing only over the indices for which eq. (1.20) is fulfilled, or alternatively, through the probability of the indeterminate outcome:

$$p_{\mathrm{succ}}^{unamb} = 1 - p_{\mathrm{fail}} = 1 - \mathrm{Tr}\left[\Pi_{\mathrm{fail}} \rho\right]. \tag{1.21}$$

Optimal two-state discrimination

A special case which will be useful in the following is the discrimination of two pure states $|\psi_1\rangle$ and $|\psi_2\rangle$ with prior probability $q_1 = q_2 = \frac{1}{2}$. In this case, the optimal success probabilities, as well as a POVM attaining them, are analytically known.

- **Min-error discrimination**
  The probability is given by the *Helstrom bound* for pure states [Hel69]:

$$p_{\mathrm{succ}}^{min\,err} \leq \frac{1}{2}\left(1 + \sqrt{1 - |\langle\psi_1|\psi_2\rangle|^2}\right). \tag{1.22}$$

- **Unambiguous discrimination**
  In this case we have the *Ivanovic–Dieks–Peres limit* [Iva87; Die88; Per88]:

$$p_{\mathrm{succ}}^{unamb} \leq 1 - |\langle\psi_1|\psi_2\rangle|. \tag{1.23}$$

Chapter 2

Linear optical Bell measurement

2.1 Brief history of optics

> *Strepsiades:*
> Have you ever seen this stone in the chemist's shops, the beautiful and transparent one, from which they kindle fire?
>
> *Socrates:*
> Do you mean the burning-glass?
>
> Aristophanes, *The Clouds*, 424 BC

Optics, the field of knowledge that deals with the manipulation of light and its interactions with the rest of matter, has been on the shelves of philosophers and scholars since the infancy of natural sciences. As far as optical artifacts are concerned, we know mirrors and other rudimentary devices predate the first millennium BC; one of the first written accounts of glass lenses being used as fire starters (ὕαλον, "burning-glass") appears as soon as the 4^{th} century BC in Aristophanes' play *The Clouds*. Around the same time a few thinkers started changing perspective on the everyday experiences of reflections, rainbows and the very process of human vision from supernatural explanations to natural ones, beginning what we could call a theoretical program to explain optical phenomena.

The study of optics as an axiomatic geometrical theory arose in the Greek world from the likes of Euclid (Ὀπτικά, *"Optics"*, ca 300 BC) and was adapted into a full theory of vision by Ptolemy (*Optics*, ca 2nd century AC).[1] Their works resurfaced during the Islamic Golden Age at the turn of the millennium, when they were studied by Islamic scholars like Ibn al-Hayṯam (كتاب المناظر, *"Book of Optics"*, 1021), referred to as "the father of modern optics". In the centuries to follow lens-making technology advanced at a steady pace, which resulted in the invention of vision-aid systems like reading stones and spectacles (13th century), microscopes (1595) and refraction-based telescopes (1608). On the theoretical side, we owe the foundation of modern classical optics to renaissance mathematicians and astronomers. With Kepler, dedicating a great deal of work on describing optical phenomena involved in astronomical observations (*Astronomie Pars Optica*, 1604) and Snell, giving a proper mathematical description of refraction almost 1500 years after its first qualitative description by Ptolemy, the way was paved for Huygens and Newton, whose debate on the nature of light itself had enormous influence on subsequent theories.

Superseding a purely geometrical view of optics, Newton was concerned with explaining phenomena like diffraction and dispersion through a theory of *physical optics*, where the geometric approximation is not valid. In pure atomistic spirit, he embraces the idea that light is made of indivisible corpuscles (*Opticks*, 1704), failing however to completely remove a wave description of light from all his observations. Indeed, in the same years Huygens had managed to give a satisfactory explanation of many of the same phenomena by modeling light as a wave, deriving important mathematical results that are still used today (*Traité de la lumière*, 1690). Newton's view held its ground through the 17th and 18th century, despite the discovery of more optical phenomena which were difficult to account for—Young's 1801 *double-slit experiment* is just one example. Eventually, the matter could be settled by experiment: in 1850, Fizeau and Foucault [Fou53] measured the difference between the speed of light in air and water, favoring Huygens' wave theory by observing a lower velocity through water (while corpuscular theory predicted the opposite). Only half a century later, the

[1] An introduction to the history of optics can be found in [Dar12].

reconciliation of these two views and their extension to the nature of all matter under the concept of *wave-particle duality* was one of the main forces driving forward the quantum revolution.

Today, applications of optics are everywhere: a prime example are the many uses of lasers for metrology, communication, imaging of the farthest objects (astronomy) and of the smallest (microscopy), and even manufacturing. The modern world of computation seem to have little to share with light, given the ubiquity of electrical transistor-based devices. However, as advances in the field of optoelectronics enabled high-speed optical fiber communication, they are making the case for the further development of all optical logic which, aside from the promise of faster processing, would eliminate the inefficiencies associated to the conversion between electrical and optical signals at the interfaces of information processing and its long-distance distribution. More closely related to the argument of this Chapter are the promises for optical devices to be a main actor for *quantum* processing of information, which range from satellite- and fiber optics-based entanglement distribution to the development of integrated optics for computation.

2.1.1 Why *linear* optics?

In the second half of the 19^{th} century, light was the protagonist of Maxwell's unification of electric and magnetic phenomena, forming the foundation of classical electromagnetism. Now modeled as an electromagnetic wave, light's interaction with a known distribution of charges ρ and currents $\vec{J}$ can be derived from Maxwell's equations:

$$\begin{aligned} \nabla \cdot \vec{E} &= \frac{\rho}{\epsilon_0}, & \nabla \cdot \vec{B} &= 0, \\ \nabla \times \vec{E} &= -\frac{\partial \vec{B}}{\partial t}, & \nabla \times \vec{B} &= \mu_0 \left(\vec{J} + \epsilon_0 \frac{\partial \vec{B}}{\partial t} \right). \end{aligned} \tag{2.1}$$

The above equations are *linear* and their solutions $\vec{E}(t)$ and $\vec{B}(t)$ enjoy a simple mathematical structure. At first glance however they seem to be of little use to describe how light propagates through a medium, where a

precise description of ρ and $\vec{J}$ may be difficult to obtain and may depend on the electric and magnetic field themselves, i.e. $\rho(\vec{E}, \vec{B})$ and $\vec{J}(\vec{E}, \vec{B})$. The contribution to ρ and $\vec{J}$ due to the material can be integrated into $\vec{E}$ and $\vec{B}$ through the auxiliary fields $\vec{D}$ and $\vec{H}$:

$$\begin{aligned} \nabla \cdot \vec{D} &= \frac{\rho}{\epsilon_0}, & \nabla \cdot \vec{B} &= 0, \\ \nabla \times \vec{E} &= -\frac{\partial \vec{B}}{\partial t}, & \nabla \times \vec{H} &= \mu_0 \left(\vec{J} + \epsilon_0 \frac{\partial \vec{D}}{\partial t} \right), \end{aligned} \tag{2.2}$$

where now ρ and $\vec{J}$ represent only external charges and currents. Now, the relation between $\vec{D}, \vec{H}$ in the material and the true $\vec{E}, \vec{B}$ has to be specified in order to solve eq. (2.2). Depending on the material, this dependency can be arbitrarily complex, breaking linearity. However, it turns out that *most* materials approximately satisfy:

$$\vec{D} = \epsilon \vec{E} \qquad\qquad \vec{H} = \frac{1}{\mu} \vec{B} \tag{2.3}$$

that is, their response is proportional to the fields applied. This restores the linear form of eq. (2.1), with the new constants ϵ and μ describing the first-order response of the material. In fact, this approximation was extremely well justified for a long time: experimental access to nonlinear optical behavior only became commonplace with the advent of the laser in 1960. A comprehensive overview of linear optical phenomena can be found in [Dor07].

Superposition

One of the main reason focusing on linear phenomena is particularly interesting is the *superposition principle*. It states that if E_1, E_2 are two solutions of our linear equations, linear combinations of them are solutions too. This property is immensely useful in many areas, not only in electromagnetism, and it allows to characterize all solutions in terms of sums of simpler ones. In the introduction to quantum linear optics that follows, we implicitly make use of it many times: it allows us for example to completely specify linear optical elements by only describing their action on a reduced, simple

set of inputs (Section 2.2.4). It also gives rise to a nice algebraic structure of linear optical operations which we introduce in Section 2.2.2.

2.2 Quantum linear optics, formally

The following Sections are meant to provide the tools needed for understanding our results, and assumes familiarity with the notation and concepts we briefly reviewed in the Introduction Chapter 1. This is by no means a complete treatment, which the interested reader can find in introductory textbooks such as [KL10].

2.2.1 Encodings

Discrete vs. continuous variable encoding

The quantized electromagnetic field supports multiple ways to encode quantum information, depending on the specific degree of freedom used. We can identify two main categories dividing the space of encodings: *continuous variables* (CV) and *discrete variables* (DV). This subdivision is reminiscent of the analog vs. digital encoding of *classical* information and, just like in that case, it leads to radically different experimental requirements and theoretical formalisms. In DV systems the fundamental unit of quantum information is the qubit, and its discrete nature makes algebra the tool of choice; CV systems [ARL14], on the other hand, deal with continuous functions and draw fully from the machinery of analysis and calculus for their manipulation.[2] The theoretical simplicity of DV, which only deals with finite-dimensional Hilbert spaces, made it suited for the design and early implementation of quantum computing tasks (Section 2.2.8). However, optical CV systems have been investigated for the practical convenience they offer for communication and cryptography purposes (a prime example is *continuous variables quantum key distribution*, or CV-QKD [LGG10]). Recently, there has been a great deal of effort [Cha21] to bring the experimental advantages of CV to the field of computation too in the hope that, in the

[2] This is not to say that the division in mathematical formalism is perfectly net: there are naturally cases in which the two blend together.

quantum world, analog systems will play a more important role than in the classical one; though at the time of writing, the computing landscape is still skewed towards DV. For the purpose of this thesis we only deal with the discrete variables encoding.

Single vs. dual rail encoding

Choosing the photon as carrier of quantum information in DV still leaves us with different choices for the physical states representing the logical $|0\rangle_L$ and $|1\rangle_L$ states of the computational basis of a qubit. In *single-rail encoding*, the logical states closely match their physical implementation:

$$|0\rangle_L = |0\rangle_p \qquad |1\rangle_L = |1\rangle_p \tag{2.4}$$

meaning the orthogonal states of the two-level system are encoded by the presence or absence of the photon in a single mode of the electromagnetic field—usually a spatially well-defined location, like a waveguide or a path in free space. While single-rail qubits seem the most natural representation as far as state preparation and measurements are concerned, they are clearly not well-suited for operations which conserve the number of photons like the linear optical ones. *Dual-rail encoding*, instead, overcomes this limitation by exploiting "which-mode" information about the photon. This time a photon is prepared in a superposition of a pair of distinguishable and not necessarily spatially separated physical degrees of freedom, such that all single-qubit states contain exactly one photon. As far as concrete implementations are concerned, various choices of optical modes can be made. Among the most widely used we recall:

- **Path encoding**, where the modes are separated in the spatial degree of freedom, e.g. in pairs of adjacent waveguides:

  $$|0\rangle_L = |1\rangle_p |0\rangle_p , \qquad |1\rangle_L = |0\rangle_p |1\rangle_p . \tag{2.5}$$

  This encoding is often referred to as just "dual-rail" in the literature, as it's the one that originally gave the name to this qubit representation.

- **Polarization encoding**, where the qubit basis states correspond to two

orthogonal polarizations. These are often taken to be the horizontal and vertical orientations orthogonal to the wave vector in a chosen frame of reference:

$$|0\rangle_{\mathrm{L}} = |H\rangle_{\mathrm{p}}\,, \qquad |1\rangle_{\mathrm{L}} = |V\rangle_{\mathrm{p}}\,, \tag{2.6}$$

but other kinds of polarization states are used (e.g. circular polarization). Multiple qubits are still usually spatially separated, but polarization encoding halves the spatial modes count compared to path encoding.

- **Time-bin encoding**, in which the time of arrival of the photons is partitioned in discrete time steps (*bins*), all orthogonal to each other. This encoding is particularly well-suited for d-level systems (qudits), where the state space of a single photon is made up of a large number of orthogonal states instead of just two; however, logical one-qubit operations are not as readily available as with the other two encodings [Kok+07].

All of these choices admit a set of optical elements which generate the whole space of possible transformations allowed by linear optics. The schemes in this thesis will focus on polarization and path encoding.

2.2.2 Algebraic structure

While linear optics imposes restrictions on the allowed evolution of the photonic states with respect to full-fledged quantum mechanics, it turns out these restrictions are captured by a nice algebraic structure. In the following, we use the notation introduced in Section 1.1.2, where we presented the formalism to work with identical bosons. In Section 1.1.4 we found these evolutions to be the special case of a Bogoliubov transformation which does not mix the creation/annihilation operators, conserving the number of photons. In particular, the transformation in eq. (1.14) simplifies to:

$$\begin{pmatrix} \mathbf{c} \\ \mathbf{c}^\dagger \end{pmatrix} = \begin{pmatrix} U & 0 \\ 0 & U^* \end{pmatrix} \begin{pmatrix} \mathbf{a} \\ \mathbf{a}^\dagger \end{pmatrix}, \tag{2.7}$$

where U is a unitary matrix. Letters in boldface stand for formal vectors of bosonic operators, i.e. $\mathbf{a} = (a_1, \ldots, a_m)$. Explicitly, we can thus write:

$$\forall i, \qquad c_i = \sum_{j=1}^{m} u_{ij}\, a_j. \tag{2.8}$$

In Section 2.2.4 we show the most common linear optical components, alongside the unitary matrix they implement. Chaining single and two-mode elements together forms an **interferometer**.[3] In a generic multimode interferometer, the action of these components on the pair of input modes (j_1, j_2) to output modes (i_1, i_2) can be represented by an $m \times m$ extended unitary of this form:

$$\begin{array}{c} \\ i_1 \\ \\ i_2 \\ \\ \end{array} \begin{array}{c} \begin{array}{ccccccc} & & j_1 & & j_2 & & \end{array} \\ \begin{pmatrix} 1 & \cdots & 0 & \cdots & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & & \vdots & & \vdots \\ 0 & \cdots & u_{11} & \cdots & u_{12} & \cdots & 0 \\ \vdots & & \vdots & \ddots & \vdots & & \vdots \\ 0 & \cdots & u_{21} & \cdots & u_{22} & \cdots & 0 \\ \vdots & & \vdots & & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & \cdots & 0 & \cdots & 1 \end{pmatrix} \end{array}, \tag{2.9}$$

i.e. which acts nontrivially only on a 2×2 subspace. Sequential execution of components is equivalent to multiplying the extended matrices together, which preserves unitarity, while parallel execution on separate sets of modes corresponds to their direct sum (contrasting with the tensor product for general quantum mechanical evolutions). The converse is also true: any $\mathrm{U}(m)$ unitary decomposes into a series of two-mode optical components of the form in eq. (2.19), preceded by an array of single-mode components (phase shifters) on the input modes. Furthermore, the decomposition is

[3] The name *interferometer* comes from classical optics, when already around 1850 physicists like Fizeau and later, Michelson used the interference properties of waves to accurately measure distances, shapes and velocities.

efficient: in the worst case, it uses

$$\binom{m}{2} = \frac{m(m-1)}{2} \sim O(m^2) \tag{2.10}$$

generalized beamsplitters and m additional phase shifters. This important result was first shown by Reck *et. al* [Rec+94], establishing an operational equivalence which lets us abstract away from the optical components and only look at the structure of the unitary group. Recently, the amount of resources needed to implement the scheme has been improved [Cle+16].

A tale of two unitaries

At this point, an important clarification has to be made, in order to avoid confusion. Is not the result in eq. (2.7) trivial, at first glance? After all, we already know that quantum mechanics allows for unitary evolutions. Moreover, this seems to contrast the claim made throughout this Section, that linear optics has limited computational power compared to full-fledged quantum computers. In fact, this confusion vanish quickly as we realize that those two unitaries act on very different spaces. In QM, an evolution involving k qubits in m modes lives in the (infinite-dimensional) Fock space. Even focusing on photon-number preserving transformations, we get an Hilbert space of dimension $\binom{m+k-1}{k}$, as in eq. (1.5). Whereas in linear optics, only the m-dimensional space spanned by the single-photon states $a_i^\dagger |0\rangle$ is involved. In the literature on interferometers, the latter is sometimes called *transfer matrix* and written as S or T to distinguish it from the former.[4] The ambiguity cleared out, we will instead default to the use of "unitary" and the symbol U, when it is clear form the context that we're talking about linear optical evolutions.

One might furthermore notice that the action of a linear interferometer on single photon states on m modes can be modeled by a QM unitary acting on m-dimensional qudits. This observation leads to an alternative theoretical framework to study linear optical evolutions, which makes use of the *Schur-Weyl duality* to map them to the quantum circuit formalism [MT18].

[4] The term "transfer matrix" is more general, and can include non-unitary transformations which may involve photon loss and other non-linear processes.

2.2.3 Polynomial representation

As seen in Section 1.1.3, it is useful to describe photonic states on m modes through polynomials in the creation operators:

$$|\Psi\rangle = \sum_{n_1,\dots,n_m}^{\infty} |n_1 \dots n_m\rangle = P_\psi(a_1^\dagger, \dots, a_m^\dagger)\,|0\rangle\,. \tag{1.11}$$

We will make heavy use of this representation throughout the thesis. For our purposes, virtually all of the states that we encounter contain a definite number of photons k, which means they are represented by homogeneous polynomials of degree k in m variables. The amplitude α for a Fock state outcome can be computed from the coefficient of the corresponding monomial, using eq. (1.10):

$$\begin{aligned} \alpha_{n_1\dots n_m}\,|n_1 \dots n_m\rangle &= \alpha_{n_1\dots n_m} \prod_{i=1}^{m} \frac{(a_i^\dagger)^{n_i}}{\sqrt{n_i!}}\,|0\rangle \\ &= \left(\frac{\alpha_{n_1\dots n_m}}{\sqrt{\prod_{i=1}^{m} n_i!}}\right) \prod_{i=1}^{m} (a_i^\dagger)^{n_i}\,|0\rangle\,. \end{aligned} \tag{2.11}$$

When doing so extra care has to be taken in order to properly account for the *bosonic normalization factor* $1/\sqrt{\prod_i n_i!}$, which arises from the indistinguishability of multiple photons in the same mode.

2.2.4 Linear optical elements

In Section 2.2.2 we showed how all linear optical transformations can be built by composing one- and two-mode basic elements. Depending on the platform, some of them have a direct counterpart on an optical table or in an integrated optical circuit. In the following, we list the conventions which we use in the rest of the thesis when drawing optical diagrams and making calculations. It should be noted that these choices might be different to the ones used in other works (e.g. [Kok+07], from which we adapted some of the diagrams); we list a few options when applicable.

Except when noted otherwise we label the input modes with integers $(1, 2, \dots)$ and the output modes with primed integers $(1', 2', \dots)$. Spatial

modes are represented by lines suggesting the optical path of the photons. We draw them blue when the modes support path-encoded qubits or when encoding is not important; we instead use red to highlight polarization modes, which may share the same spatial mode.

Beamsplitter

The beamsplitter (BS) is designed to partially reflect and partially transmit incident light. It is commonly made from two triangular glass prisms, or by a thin metal coated optical substrate (a *half-silvered* mirror). In integrated optics, the beamsplitter can be realized by bringing together two waveguides. Its diagrammatic representation is:

with $\{1,2\}$ the incident spatial modes, $\{1',2'\}$ the output modes and θ parametrizing the reflectivity $R = (\sin\theta)^2$. The unitary associated with the action of the beamsplitter on the input modes follows somewhat different conventions throughout the literature, depending whether a dephasing ϕ is also introduced between the reflected and transmitted modes:

$$U_{\mathrm{BS}} = \begin{pmatrix} \cos\theta & -\mathrm{e}^{-i\phi}\sin\theta \\ \mathrm{e}^{i\phi}\sin\theta & \cos\theta \end{pmatrix}. \tag{2.12}$$

Common choices for the dephasing include $\phi = \frac{\pi}{2}$, which leads to the *symmetric* beamsplitter, and $\phi = 0$ (our choice for this work), resulting in a real transfer matrix:

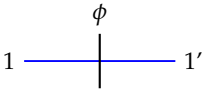
$$\phi = \frac{\pi}{2} \longrightarrow \begin{pmatrix} \cos\theta & i\sin\theta \\ i\sin\theta & \cos\theta \end{pmatrix}, \tag{2.13}$$

$$\phi = 0 \longrightarrow \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}. \tag{2.14}$$

The beamsplitters we use are meant to be polarization-independent, i.e. they only mix spatial modes and act identically on horizontal and vertical modes if fed with polarized light. When $\theta = \frac{\pi}{4}$ we have $\cos\theta = \sin\theta = \frac{1}{\sqrt{2}}$, and the resulting beamsplitter is called **balanced** or **50:50**. In this case, we omit the angle θ from the diagram and sometimes use the Hadamard matrix $H = \frac{1}{\sqrt{2}}\left(\begin{smallmatrix} 1 & 1 \\ 1 & -1 \end{smallmatrix}\right)$ as the beamsplitter unitary.

Phase shifter

This optical element introduces a dephasing of an angle ϕ on the mode on which it acts. It usually consists of a transparent material with a different index of refraction than free space, where the length of the material determines the amount of accumulated phase. It is represented as:

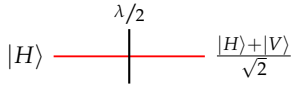


When acting on a single mode, it is described by an element of $\mathrm{U}(1)$, i.e. a unit scalar $\mathrm{e}^{i\phi}$:

$$a_{1'}^\dagger = \mathrm{e}^{i\phi} a_1^\dagger \,. \tag{2.15}$$

If applied on one path of a path-encoded dual-rail qubit, it can be used to rotate the state around the Z direction in the Bloch sphere.

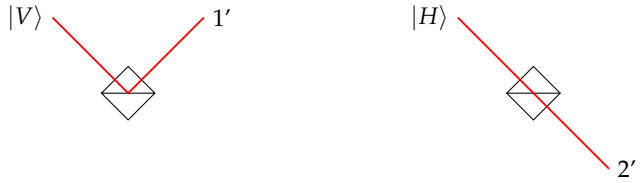
On polarization-encoded qubits on the other hand, we need to selectively apply a phase to one of the two polarization modes. A solution is to use a **waveplate**, which typically consists of a slab of birefringent material with different indices of refraction along two orthogonal directions, called *fast* and *slow axis*. When the applied phase is exactly half the wavelength of the incoming photon ($\phi = \pi$), it implements the same transformation as a beamsplitter of angle 2θ, where θ is the angle between the fast axis and the horizontal polarization mode. When $2\theta = \frac{\pi}{4}$ (corresponding to a balanced beamsplitter) we will show it as:



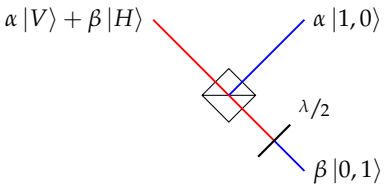
where $\frac{\lambda}{2}$ refers to its name, the *half-wave* plate.

Polarizing beamsplitter

The polarizing beamsplitter (PBS) is the bridge connecting spatial and polarization modes. It acts differently depending on the incoming polarization:



namely, it reflects vertically polarized photons and transmits horizontally polarized ones (the action is analogous for second spatial input mode, not shown above). The PBS is not enough to obtain a path-encoded qubit from a polarization-encoded one (and viceversa); a half-wave plate is needed to align the polarization of the output arms:

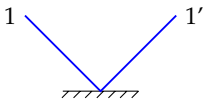


(2.16)

This step is important, as otherwise the two modes would be distinguishable and would not interfere anymore on a subsequent beamsplitter.

Mirror

The name says it all:



which implements the identity operation $U = I$.

Photon sources and detectors

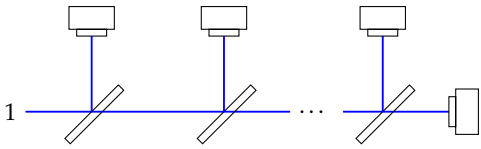
Strictly speaking, these are not linear devices, but are essential to any quantum linear optics experiment. Their physical implementations are very diverse, depending on the wavelength of operation and required purity or noise level: a review can be found in [Eis+11]. In our optical diagrams the sources are omitted, showing instead the initial state of the mode beside the input paths when useful. As far as detectors are concerned, their main job is to convert the incoming photons into an electrical signal that can be read classically. They can have different features, depending on how much information they can extract from the input. We use:

- **Non-photon number resolving detectors**, which fire if the mode is occupied by at least one photon:

  They implement the Fock space projection $\{|0\rangle\langle 0|, I - |0\rangle\langle 0|\}$.

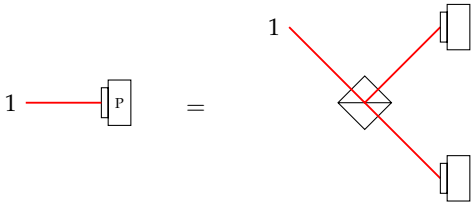
- **Photon number resolving detectors** (PNRD), capable of counting the number of photons:

  Ideal PNRD measure the observable $N = a^\dagger a$. In practice, their resolution is usually limited to a maximum number of photons. A common way to implement a (non-ideal) PNRD is to use a tree of beamsplitters and regular non-number-resolving detectors:



- **Polarization-resolving detectors**, which additionally extract polarization information:

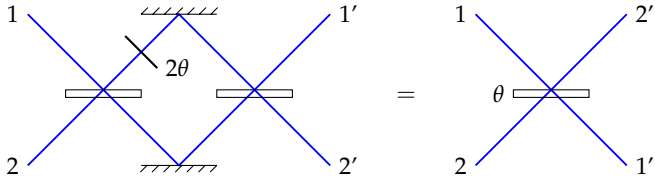
Physically, they can be realized via a PBS and two non-polarization-resolving detectors:



(2.17)

Single-qubit operations

Often, a variable-reflectivity beamsplitter is not available; in practice, the following setup (the *Mach–Zehnder interferometer* [Zeh91; Mac92; Kok+07]) can be used to implement it from two balanced beamsplitters and a variable phase shifter:



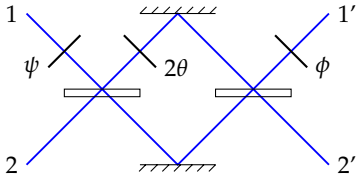
(2.18)

With it, a general rotation in the Bloch sphere:

$$R(\theta,\phi,\psi) = \begin{pmatrix} 1 & 0 \\ 0 & \mathrm{e}^{i\phi} \end{pmatrix} \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & \mathrm{e}^{i\psi} \end{pmatrix}, \tag{2.19}$$

can be implemented by adding with phase shifters of phase ψ and ϕ on one of the input and one of the output modes. Ignoring the global phase, the

two-mode transformation looks like this:



(2.20)

Two-qubit operations

Besides single-qubit gates, any full implementation of a quantum computer should be able to execute entangling two-qubit operations. The linear optical platform is at a disadvantage here, due to the lack of direct interaction between photons, which interact only indirectly through interference. An entangling gate can nonetheless be achieved by complementing interference with single-photon measurements, which act as a source of nonlinearity. The resulting gates are inherently *probabilistic* [KLM01; Kok+07]; lowering their failure probability to a manageable level requires dynamic control and complex auxiliary states. In Section 2.2.8 we briefly review the challenges and improvements which, in recent years, led to more and more efficient implementations of these gates. However, another perspective on the matter is that often we do not need a general purpose quantum computer: in many useful applications, e.g. quantum metrology [GLM11] and quantum simulation [GAN14], it makes sense to expect large gains in efficiency if we focus instead on non-universal schemes dedicated to the specific problem at hand. Indeed, this turns out to be the case for *Bell measurement*, the primitive we explore in Section 2.3.

2.2.5 Auxiliary states

The capabilities of linear optical networks can be augmented by injecting so-called *auxiliary states*[5] which can be prepared beforehand and are in general independent of the rest of the input. Auxiliary states can be arbitrarily

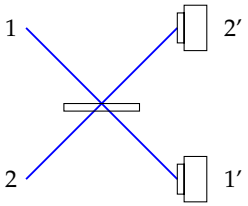
[5]These are usually called *ancillary states* or *ancillæ* in a large fraction of the literature, including our paper [OG18]. Given its problematic etymological origin [Wie17], we chose a more neutral term here, which is nonetheless quite commonly used.

complex and may require nonlinear processes to be produced; however, they can be prepared *offline* (i.e. before the actual input is available), leveraging for example probabilistic processes. A similar phenomenon in the theory of quantum computation provides an analogy: there, the celebrated *Gottesman-Knill theorem* [Got98] shows how a circuit composed by gates from a restricted set (*Clifford gates*) provides no quantum computational advantage, but injecting so-called *magic states* restores universality [Kni04; BK05].

In linear optics, the experimentally-friendly characteristics mentioned above encourage theorists to asses exactly how well auxiliary states can replace nonlinear operations on the input. This Chapter is about the power of linear optics + auxiliary states for performing Bell measurement.

2.2.6 Hong–Ou–Mandel effect

One of the simplest example showcasing a departure of quantum linear optics from its classical counterpart is the Hong–Ou–Mandel (HOM) effect [HOM87], which involves a balanced beamsplitter, two single photon sources and two photon detectors:



(2.21)

The beamsplitter acts on the input modes in this way:

$$\begin{pmatrix} a^\dagger_{1'} \\ a^\dagger_{2'} \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} a^\dagger_1 \\ a^\dagger_2 \end{pmatrix} \tag{2.22}$$

When the two photons which interfere on the beamsplitter are indistinguishable (i.e. they have the same frequency, phase and are correctly aligned), a phenomenon called *bunching* occurs: they are always both detected in the

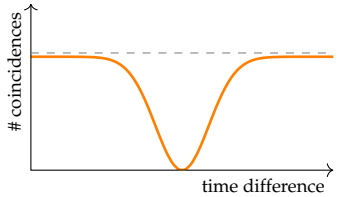
same output mode. In the polynomial representation:

$$a_1^\dagger a_2^\dagger \xrightarrow{\text{BS}_{50:50}} \frac{1}{2}(a_{1'}^\dagger + a_{2'}^\dagger)(a_{1'}^\dagger - a_{2'}^\dagger) = \frac{1}{\sqrt{2}}\left[\frac{(a_{1'}^\dagger)^2}{\sqrt{2}} - \frac{(a_{2'}^\dagger)^2}{\sqrt{2}}\right], \tag{2.23}$$

which shows that the amplitudes of the *coincidence* events (both detectors click) cancel out. Using Fock states, eq. (2.23) is commonly written as:

$$|11\rangle \xrightarrow{\text{BS}_{50:50}} \frac{1}{\sqrt{2}}(|20\rangle - |02\rangle). \tag{2.24}$$

The more the photons are indistinguishable, the more pronounced is the effect, which produces the following curve, known as *HOM dip*, when one of the photon's parameters (for example, the time of arrival on the beamsplitter) is varied:



BosonSampling [AA13], which we briefly introduce in Section 2.2.8, can be viewed as a generalization of the HOM effect. Indeed, its quantum advantage revolves around the indistinguishability of the photons: above a distinguishability threshold, BosonSampling can be classically simulated [Ren+18].

Entangled modes vs. entangled qubits

It is helpful to point out an additional subtlety here which will be relevant in the following for some of the auxiliary states used in Bell measurement (Section 2.3.8). Consider the action of a balanced beamsplitter which mixes the two modes of a dual-rail, path-encoded (eq. 2.5) state of one logical qubit:

$$|0\rangle_\text{L} = |1\rangle_\text{p}|0\rangle_\text{p} \xrightarrow{\text{BS}_{50:50}} \frac{1}{\sqrt{2}}\Big(|1\rangle_\text{p}|0\rangle_\text{p} + |0\rangle_\text{p}|1\rangle_\text{p}\Big) = |+\rangle_\text{L}. \tag{2.25}$$

In this encoding, this looks like a regular, deterministic single-qubit operation. The same physical modes and photons can be used with single-rail encoding, to represent *two* logical qubits:

$$|10\rangle_{\mathrm{L}} \xrightarrow{\mathrm{BS}_{50:50}} \frac{1}{\sqrt{2}}\Big(|10\rangle_{\mathrm{L}} + |01\rangle_{\mathrm{L}}\Big). \tag{2.26}$$

Here, our beamsplitter took an unentangled input and produced a maximally entangled state! The two situations in eq. (2.25) and eq. (2.26) help stress an important point, which is that *with an encoding, comes an implicit grouping* of the modes. When talking about entangled states only entanglement between separate groups is relevant, while the modes of a group might be "locally" entangled.

In fact, linear operations on the modes might just not keep you inside the qubit subspace. If we stick to these two encodings, the HOM state in eq. (2.24) is of difficult interpretation, having two photons in one mode—which do not correspond to a logical qubit state. Can't we just use single-rail encoding for everything, as it lets us create maximally entangled states? The catch is that we just traded the hardness of creating entanglement with the possibility of performing simple qubit rotations. As already observed, in single-rail encoding single qubit gates do not conserve the number of photons, and can only be implemented probabilistically within linear optics [Par00].

2.2.7 Linear optical network

Definition 1. We define a *linear optical network* (Fig. 2.1) as a device composed of the following parts:

- A series of (nonlinear) single photon sources, capable of injecting photons in a subset of the input modes;
- Any arrangement of linear optical elements (Section 2.2.4), which form an interferometer implementing the unitary U_i.
- Photon-number-resolving detectors (PNRD), which destructively measure the number of photons in each mode;

- The three components above combined to form *stages*, each possibly depending on the result of the measurements performed in the previous stage.[6]

In the following we mainly use single-stage networks, described by a single U.

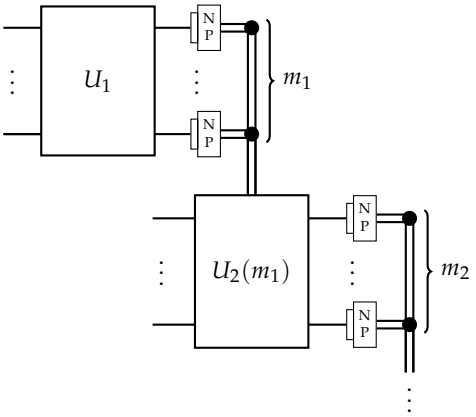


Figure 2.1: A multi-stage linear-optical network. The single-photon sources are not shown.

2.2.8 Quantum computation with light

At this point, it should not be difficult to argue that the linear optical platform has several features which renders it a good candidate for quantum computing purposes. They have a pretty good historical record too, as the first proposal for a quantum gate implementation involved photonic qubits [Mil89]. Nonetheless, quantum computing is hard: a physical platform has to simultaneously possess many desirable properties in order to attain universality [DiV00]. How well light-based qubits fare against other platforms? Photons interact very little with the environment even at room temperature,

[6] In practice, this adaptivity (also known as *feedforward*) can be implemented by optical switches, which reroute photons to different static interferometers, or by building a programmable interferometer [Har+16].

and their quantum state tends to be almost free from decoherence. They can be transmitted over long distances, produced in a given fiducial state with high accuracy, processed at high repetition rates and measured with a variety of techniques [Eis+11]. Single-qubit operations are readily available using the linear optical components we just saw, which come straight from classical optics. Their biggest crux is the realization of the required two-qubit gates. We already mentioned how these can be implemented probabilistically, using linear optical evolutions and measurements; other solutions exists, which include the use of nonlinear elements based on Kerr nonlinearity [Ker75] or optical-matter interactions [Pey+12]. Here, the catch is that the strength of these effects in available materials is often really weak, which imposes other kinds of design constraints. Despite these difficulties, a great deal of theoretical and experimental efforts have been put in recent years to try to overcome them. We briefly introduce two examples in the following.

KLM-like schemes

Do single-photon generation and (destructive) photon detection introduce powerful enough nonlinearities to actually achieve *efficient* universal quantum computation? If all we can do are two-qubit gates with a constant probability of failure, it seems there's no hope: if we want N gates in a row to succeed with high probability, we need to replicate the circuit $\sim O(2^N)$ times, effectively killing any advantage that quantum computation can bring to the table. To show that efficient linear optical quantum computation (LOQC) is indeed possible, Knill, Laflamme and Milburn developed the *KLM scheme* in 2001 [KLM01]. Their quantum computer is a full-fledged linear optical network (Definition 1): they cleverly make use of auxiliary states prepared offline (Section 2.2.5) along with dynamically applied operations which depend on previous measurements (feedforward) and techniques like gate teleportation, redundant encoding and error correction. This way, they can approach near-deterministic gates with a polynomial overhead in terms of optical components. The original KLM scheme however is all but experimentally feasible, requiring tens of thousands of components for

lowering the failure probability of a gate under 5% [KLM01]. Their work has been nonetheless of huge theoretical importance in the following years, when various improvements made the requirements for LOQC increasingly smaller. A thorough overview of this "race" to lower the resource count, which is ongoing to this day, can be found in Gimeno-Segovia's PhD thesis [Gim15, Chapter 2 and Appendix B].

Boson sampling

From the point of view of theoretical computer science, the linear optical platform has recently proven to be of foundational interest. In 2010, Aaronson and Arkhipov [AA13] showed that linear optics supports a restricted "ballistic" model of computation, whose experimental requirements are much lower than in the KLM scheme but which nonetheless appears to efficiently solve BosonSampling, a classically intractable problem. While traditional complexity classes like P and NP are concerned with *decision problems*, this model involves *sampling problems*: their input is a probability distribution, and their output is a procedure (that is, a circuit in the model of computation at hand) to output samples from the distribution. In order to picture this kind of computation, it can be useful to think in terms of a classical analog, the *Galton board* [Gal89, p. 63] (Fig. 2.2), in which a collection of beads moving through a network of pegs ends up extracting samples from the binomial distribution.

In BosonSampling, k photons enter the first k modes of an optical network with $m \gg k$ modes. After scattering through the interferometer—where one might think of beamsplitters as analogous to the pegs in a Galton board—the photons hit some subset of the single photon detectors at the output of each mode, generating a *sample*. The distribution of clicks in a BosonSampling experiment depends on the particular network of optical elements implemented, the same way the Galton board's final distribution is determined by the arrangement of the pegs. We will see in Section 2.2.2 that each linear optical arrangement can be described by a unitary matrix U, which fully encodes the expected output probability distribution. In particular, this distribution is linked to the *permanent* of $k \times k$ submatrices

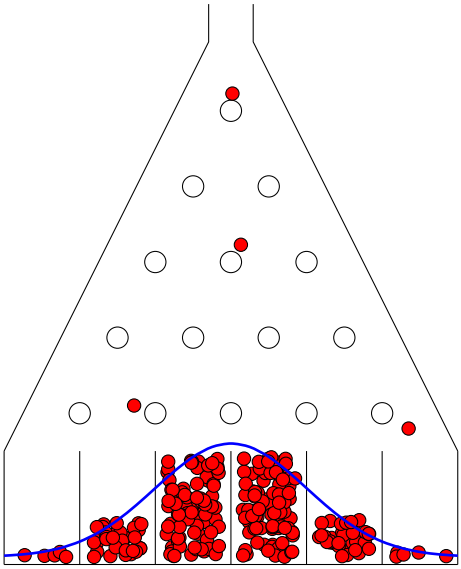


Figure 2.2: The Galton board.

A_k of U, which is a polynomial of the entries a_{ij} defined as:

$$\text{perm}(A_k) = \sum_{\sigma \in S_k} \prod_{i=1}^{k} a_{i,\sigma(i)} \tag{2.27}$$

where $\sigma \in S_k$ are all the permutations of $(1, \ldots, k)$. Valiant had showed in 1979 that computing permanents is a hard problem, in the complexity class #P [Val79]. However, the existence of an efficient classical algorithm for BosonSampling lets us approximate such permanents in the complexity class BPP^{NP}, which is widely believed to be much less powerful than #P. Instead, the existence of an efficient quantum algorithm has no such consequence: then, BosonSampling represent a promising separation between classical and quantum computation.[7,8] It is unclear if this restricted model

[7] Despite its astounding implications on cryptography, we still do not know which consequences a classical polynomial algorithm for factoring would have on the known complexity classes. In this regard, BosonSampling is better evidence, if probably of low practical value.

[8] The result in [AA13] actually concerns *approximate* BosonSampling, which better represent a

of computation (which can be implemented on current NISQ hardware [Zho+21]) can compute problems which are of practical interest outside the theoretical applications on establishing quantum computational advantage. That calculating the output probabilities of an interferometer is hard will be important in the following, when it will be reflected in the scalability of our Bell measurement optimization algorithm (Section 2.5.5).

2.3 Bell measurement

Among the features of quantum mechanics which exhibit nonclassical behavior, it is hard to find a more compelling and historically significant example than the existence of the set of maximally entangled two-qubit states known as the *Bell basis*.

Definition 2 (Bell basis)**.** The following pure entangled states:

$$|\phi^+\rangle = \frac{1}{\sqrt{2}}\Big(|0\rangle\,|0\rangle + |1\rangle\,|1\rangle\Big), \tag{2.28a}$$

$$|\phi^-\rangle = \frac{1}{\sqrt{2}}\Big(|0\rangle\,|0\rangle - |1\rangle\,|1\rangle\Big), \tag{2.28b}$$

$$|\psi^+\rangle = \frac{1}{\sqrt{2}}\Big(|0\rangle\,|1\rangle + |1\rangle\,|0\rangle\Big), \tag{2.28c}$$

$$|\psi^-\rangle = \frac{1}{\sqrt{2}}\Big(|0\rangle\,|1\rangle - |1\rangle\,|0\rangle\Big), \tag{2.28d}$$

form an orthonormal basis of the Hilbert space of two (logical) qubits. We collectively refer to them as *Bell states* and we will use the shorthand $|\beta_i\rangle$, where $i = 1 \dots 4$, to label them in the order of eq. (2.28).

A local measurement of any one of the two qubit in a Bell state gives the outcome 0 or 1 with $1/2$ probability, collapsing the other to the same (for $|\phi^\pm\rangle$) or opposite (for $|\psi^\pm\rangle$) state; moreover, similar correlations are present when measured in a different basis than the computational basis, e.g. the Hadamard basis ($|+\rangle$, $|-\rangle$). The states are also called *EPR pairs*, from the 1935 paper by Einstein, Podolsky and Rosen [EPR35] which first

real-world non-ideal experiment, and relies on a few additional conjectures on the distribution of permanents of random matrices.

highlighted their relevance concerning foundational questions in quantum mechanics. In 1964, Bell [Bel64] proved that the measurement correlations produced by the Bell states reject any explanation in terms of a deeper classical theory involving local, hidden variables (i.e. obeying "local realism"), suggesting for the first time the possibility that quantum behavior could be experimentally discriminated from classical, by looking for the violation of what is now known as a *Bell inequality*. Later advancements in theory [Cla+69] made Bell inequalities accessible to experiments, which settled the matter in the eighties in favor of quantum mechanics [ADR82] by generating entangled photons in a Bell state and measuring them in different bases at spacelike-separated locations. The (possibly) last word, closing all reasonable loopholes, is from a recent experiment in Delft [Hen+15].

Bell states are used throughout quantum information protocols, including communication tasks like teleportation [Ben+93] and quantum repeaters [San+11], and cryptographic primitives like quantum key distribution [Eke91] and self-testing of quantum states [ŠB20]. A Bell measurement also provides a primitive for measurement-based quantum computation to grow *cluster states* from smaller entangled states [RHG07; Bar+21]. They are both an essential resource for these tasks, as they cannot be prepared remotely by local operations and classical communication (LOCC) [Chi+14], as well as a theoretical tool in proving the protocols' correctness: while today most QKD experimental demonstrations are based on coherent states, their security is based on an equivalence with their Bell states-based counterpart. An EPR pair shared among two parties is also known as an *ebit*; there exists protocols [Ben+96] which can "distill" ebits from other entangled resources by LOCC. The corresponding projective measurement, described by the four orthogonal operators $\{\Pi_i = |\beta_i\rangle \langle\beta_i|\}$ is called *Bell measurement*. If performed nondestructively, it leaves the system in one of the four Bell states; in this case, the measurement itself is an entangling operation. In the circuit model of quantum computation, Bell states can be generated and measured by the simple circuits shown in Fig. 2.3.

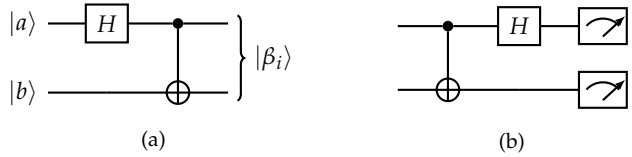


Figure 2.3: Quantum circuits for (a) generating the Bell state $|\beta_i\rangle$ depending on the computational basis states at the input; (b) projective measurement onto the Bell basis.

2.3.1 Bell measurement in linear optics

To work with Bell states in the linear optical setting, we need to write eq. (2.28) using mode operators. Depending on the chosen encoding:

Path encoding (2.29)	**Polarization encoding** (2.30)
$\left\|\phi^{\pm}\right\rangle = \frac{1}{\sqrt{2}}(a_1^\dagger a_3^\dagger \pm a_2^\dagger a_4^\dagger)\left\|0\right\rangle$	$\left\|\phi^{\pm}\right\rangle = \frac{1}{\sqrt{2}}(h_1^\dagger h_2^\dagger \pm v_1^\dagger v_2^\dagger)\left\|0\right\rangle$
$\left\|\psi^{\pm}\right\rangle = \frac{1}{\sqrt{2}}(a_1^\dagger a_4^\dagger \pm a_2^\dagger a_3^\dagger)\left\|0\right\rangle$	$\left\|\psi^{\pm}\right\rangle = \frac{1}{\sqrt{2}}(h_1^\dagger v_2^\dagger \pm v_1^\dagger h_2^\dagger)\left\|0\right\rangle$

where $(a_1^\dagger, a_2^\dagger)$, $(h_1^\dagger, v_1^\dagger)$ and $(a_3^\dagger, a_4^\dagger)$, $(h_2^\dagger, v_2^\dagger)$ are the creation operators on the modes that support, respectively, the first and the second logical qubit. Remember that the two encodings can be always converted into each other by the interferometer in Diagram (2.16); depending on the situation, we will find more convenient to use one or the other. Typically, a nonlinear process like *spontaneous parametric down-conversion* (SPDC) [BP08] can be used as a source of pairs of polarization-entangled photons, together with postselection. Deterministic creation of the above states cannot be done in linear optics; this immediately rules out the possibility of a deterministic, non-destructive Bell measurement. Nevertheless for most applications, including e.g. quantum teleportation, *destructive* Bell measurement is enough, which does not leave the photons in an entangled state.

The destructive, non-ideal version of the measurement is described in general by a POVM. We characterize the success of a measurement strategy by how well it discriminates the Bell states, when acting on an equal mixture of them. This motivates the definition of the following linear optical device.

Definition 3 (Bell analyzer)**.** A (potentially multistage) linear optical network of $m \geq 4$ modes, fed with one of the input states:

$$\forall \beta_i, \quad |\Psi^{\text{in}}_{\beta_i}\rangle = |\beta_i\rangle\,|\Gamma\rangle = P_{\beta_i}(a_1^\dagger, \ldots, a_4^\dagger)\, Q(a_5^\dagger, \ldots, a_m^\dagger)\,|0\rangle\,, \tag{2.31}$$

each with probability p_i, is called *Bell analyzer* (BA). P_{β_i} is the polynomial representation of the Bell state $|\beta_i\rangle$, while $|\Gamma\rangle = Q\,|0\rangle$ describes an auxiliary state on the rest of the modes. The aim of a BA is to guess which Bell state $|\beta_i\rangle$ entered the network. When evaluating the performance (or *efficiency*) of a BA, we will always assume each $|\beta_i\rangle$ is chosen with probability $p_i = 1/4$.

After passing through the interferometers the photons hit the PNRDs, which measure the occupation number of each output mode. A configuration of clicks $e := n_1 \ldots n_m$ is a *detection event*, and constitute the output of a Bell analyzer from which we need to deduce the index β_i. The quantities we are interested in are the probabilities $p(e|\beta_i)$ of the event e occurring conditioned on having Bell state β_i at the input. Notice that within each stage of a Bell analyzers (i.e. considering one interferometer at a time) detection events are in one-to-one correspondence with Fock states of the output modes. The probability of each event can be computed by evolving $|\Psi^{\text{in}}_{\beta_i}\rangle$ through the interferometer, obtaining the output state before the PNRD array:

$$|\Psi^{\text{in}}_{\beta_i}\rangle \xrightarrow{U} |\Psi^{\text{out}}_{\beta_i}\rangle = T_{\beta_i}(c_1^\dagger, \ldots, c_m^\dagger)\,|0\rangle\,, \tag{2.32}$$

where the polynomial T_{β_i} is the result of applying the substitution in eq. (2.8) to the input creation operators' monomials in $P_{\beta_i} Q$. The probabilities $p(e|\beta_i)$ are then computed from the coefficients of T, taking care of the bosonic normalization factor as shown in Section 2.2.3. From them, we can assess the *efficiency* of a Bell analyzer $\mathcal{B}$, namely the probability $p(\mathcal{B})$[9] of a successful unambiguous discrimination (Section 1.1.5).

Remark. As Calsamiglia and Lütkenhaus [LCS99]) point out, a perfect Bell analyzer (i.e. with unit efficiency) should not be *a priori* ruled out on a linear optical basis. It could be possible that a sufficiently large network,

[9] We will mostly deal with single-stage analyzers described by U, in which case we will indicate the efficiency as $p(U)$.

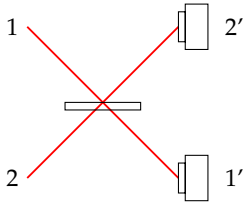


Figure 2.4: The simplest BA is just a beamsplitter, achieving $p_{\text{succ}} = 1/4$. It works by discriminating $|\psi^-\rangle$, which is the only state for which both detectors click, from the other Bell states.

which interferes the states with carefully chosen (and potentially entangled) auxiliary states, can be used to separate the events in the output space far enough to allow us to always correctly identify the input β_i. Nevertheless, they show that such no-error discrimination is impossible (Section 2.3.3).

2.3.2 A simple Bell analyzer

Figure 2.4 shows the (arguably) simplest Bell analyzer which achieves non-zero discrimination probability. This scheme is widely used in experiments [Bou+97; Pir+15], where its simplicity outweighs the low 25% efficiency. Moreover, it can be easily upgraded to a 50% scheme (which, as we will see, is optimal in a specific context) as shown in Section 2.3.5. Let's look at out how it works. The setup looks the same as the Hong–Ou–Mandel experiment, consisting of a beamsplitter and two photon detectors.

The Bell states enter in polarization encoding in spatial modes 1 and 2, and are measured in modes 1′ and 2′. The beamsplitter acts identically on both sets $\{h_1^\dagger, h_2^\dagger\}$ and $\{v_1^\dagger, v_2^\dagger\}$ of creation operators, as in eq. (2.22). The only discriminating event turns out to be a coincidence detection, i.e. one photon (of any polarization) in each of the two detectors. Introducing variables $b, d \in \{h, v\}$, the monomials corresponding to the coincidence event can be written as $(b_{1'}^\dagger d_{2'}^\dagger)$. Working backwards, we can get the possible contributions from the input state:

$$b_{1'}^\dagger d_{2'}^\dagger \to \frac{1}{2}(b_1^\dagger + b_2^\dagger)(d_1^\dagger - d_2^\dagger) = \frac{1}{2}(b_1^\dagger d_1^\dagger + b_2^\dagger d_1^\dagger - b_1^\dagger d_2^\dagger - b_2^\dagger d_2^\dagger). \quad (2.33)$$

The polarization-encoded Bell states have one photon in each spatial mode, so we can discard $b_1^\dagger d_1^\dagger$ and $b_2^\dagger d_2^\dagger$. The remaining terms are nonzero iff $b \neq d$, i.e. when the photons have different polarization. In that case the state is proportional to $|\psi^-\rangle$, which can then be unambiguously discriminated by witnessing this detection event. Later, we will work forward from the Bell states and obtain the amplitude for all output events (Section 2.3.5). From a different viewpoint, $|\psi^-\rangle$ is antisymmetric under mode exchange; this eludes the bunching feature of the HOM effect.

2.3.3 No-go theorem for perfect Bell analyzers

The first general result on the efficiency of a Bell analyzer was given in 1999 by Lütkenhaus, Calsamiglia and Suominen [LCS99]. A similar conclusion had been reached before by Vaidman and Yoran [VY99] in the much more restrictive setting of no auxiliary modes ($Q = 1$) and no feedforward. A few years later, van Loock and Lütkenhaus derive general criteria to decide if a specific set of state is (perfectly) distinguishable via linear optics [LL04], extending the following no-go theorem.

Theorem 1 (No perfect BA [LCS99])**.** *Let $\mathcal{B}$ be a Bell analyzer according to Definition 3. Then within each stage there is at least one detection event e such that $p(e|\beta_i) > 0$ for at least two input states β_a, β_b.*

Whereas Theorem 1 is an important theoretical result, by itself it does not rule out the possibility of a simple, experimentally viable Bell analyzer with efficiency arbitrarily close to 100%. However, their follow-up result discussed in the next Section puts a further nail in that coffin.

2.3.4 Calsamiglia–Lütkenhaus 1/2 upper bound

An experimentally interesting subclass of Bell analyzers is the one obtained by considering networks with vacuum state in the auxiliary modes, i.e. setting $Q(a_5^\dagger, \ldots, a_m^\dagger) = 1$ in eq. (2.31). In this case, the input states enjoy a particularly symmetric description, and a stronger upper bound can be proven. As this result is important for our work, we report a summarized version of the proof, adapting it to our notation.

Theorem 2 ([CL01])**.** *The maximum efficiency of a Bell analyzer $\mathcal{B}$ with $Q = 1$ (no extra photons in the auxiliary modes) is $p(\mathcal{B}) = 1/2$.*

Proof. The goal is to obtain an expression for $|\Psi^{\text{in}}_{\beta_i}\rangle$ which is easy to evolve through the interferometer's unitary U. Notice that in general, an optical state containing two photons is described by an homogeneous degree-2 polynomial (Section 2.2.3), which can be written as a symmetric bilinear form of the "formal" vector[10] of degree-1 monomials:

$$P(a_1^\dagger, \ldots, a_m^\dagger) = \sum_{i,j=1}^{m} N_{ij}\, a_i^\dagger a_j^\dagger = \mathbf{a}^{\mathsf{T}} \mathbf{N}\, \mathbf{a}. \tag{2.34}$$

In the case at hand, only the modes 1 through 4 are occupied, the rest being empty. Thus, only a 4×4 corner of the matrix $\mathbf{N}$ is relevant:

$$\mathbf{N} = \frac{1}{2} \left[\begin{array}{c|c} \mathbf{W} & 0 \cdots\cdots 0 \\ \hline \begin{matrix} 0 \\ \vdots \\ \vdots \\ \vdots \\ 0 \end{matrix} & \begin{matrix} \\ \\ 0 \\ \\ \\ \end{matrix} \end{array}\right] \tag{2.35}$$

where the $1/2$ factor prevents double counting. For Bell state β_i, the block $\mathbf{W}^{\beta_i}$ is can be written in a compact form:

$$\mathbf{W}^{\beta_i} = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 & & \delta_{i1}+\delta_{i2} & \delta_{i3}+\delta_{i4} \\ & & \delta_{i3}-\delta_{i4} & \delta_{i1}-\delta_{i2} \\ \delta_{i1}+\delta_{i2} & \delta_{i3}-\delta_{i4} & 0 & \\ \delta_{i3}+\delta_{i4} & \delta_{i1}-\delta_{i2} & & \end{pmatrix}, \tag{2.36}$$

where $\delta_{ij} = 1$ for $i = j$ and 0 otherwise. Notice that, when correcting for the bosonic normalization factor, $\sqrt{2}\,\mathbf{W}^{\beta_i}$ is *also* unitary.

Let's work out the polynomial $T(c_1^\dagger, \ldots, c_m^\dagger)$ describing the state at the output of the first stage of transfer matrix $U^\dagger$. We have, by eq. (2.8):

$$a_i^\dagger = \sum_{j=1}^{m} u_{ij}\, c_j^\dagger \quad \longrightarrow \quad \mathbf{a} = U\mathbf{c}, \tag{2.37}$$

[10] In order to ease the notation we write $\mathbf{a} = (a_1^\dagger, \ldots, a_m^\dagger)^{\cdot}$ instead of $\mathbf{a}^\dagger$. We will never have to deal with annihilation operators in the following, so no ambiguity will arise.

then, substituting in eq. (2.34):

$$T(\mathbf{c}) = (U\mathbf{c})^{\mathsf{T}}\mathbf{N}\,(U\mathbf{c}) = \mathbf{c}^{\mathsf{T}}U^{\mathsf{T}}\mathbf{N}^{\beta_i}U\mathbf{c} = \mathbf{c}^{\mathsf{T}}\mathbf{M}^{\beta_i}\mathbf{c}. \tag{2.38}$$

Due to eq. (2.35), $\mathbf{M}^{\beta_i}$ simplifies to

$$\mathbf{M}^{\beta_i} = \frac{1}{2}\,\overline{U}^{\mathsf{T}}\mathbf{W}^{\beta_i}\,\overline{U}, \tag{2.39}$$

with $\overline{U}$ the $4 \times n$ matrix obtained by truncating U at the fourth row.

Now we can compute the output amplitudes from the coefficients of $T(\mathbf{c})$. In particular, for the case of both photons bunching in the same mode k we need the coefficient of $c_k^{\dagger 2}/\sqrt{2!}$:

$$\left(M_{kk}^{\beta_i}\right)c_k^{\dagger 2} = \left(\sqrt{2}\,M_{kk}^{\beta_i}\right)\frac{c_k^{\dagger 2}}{\sqrt{2}} = \left(\frac{1}{\sqrt{2}}\,\overline{\mathbf{u}}_k^{\mathsf{T}}\mathbf{W}^{\beta_i}\,\overline{\mathbf{u}}_k\right)\frac{c_k^{\dagger 2}}{\sqrt{2}} \tag{2.40}$$

where $\overline{\mathbf{u}}_k$ is the k^{th} column of $\overline{U}$, which appears when we use eq. (2.39) to expand the diagonal element of $\mathbf{M}^{\beta_i}$. Labeling this kind of detection event $\alpha = 2_k$, we have:

$$p(2_k|\beta_i) = \frac{1}{2}\left|\,\overline{\mathbf{u}}_k^{\mathsf{T}}\mathbf{W}^{\beta_i}\,\overline{\mathbf{u}}_k\right|^2. \tag{2.41}$$

In order to contribute to the overall success probability these events have to be discriminating, but it is easy to check that this is not the case. Imposing eq. (2.41) to be zero for three out of four Bell states is very restrictive on the elements of U:

$$\overline{\mathbf{u}}_k = (u_{1k}, u_{2k}, 0, 0)^{\mathsf{T}} \quad \text{or} \quad \overline{\mathbf{u}}_k = (0, 0, u_{3k}, u_{4k})^{\mathsf{T}}, \tag{2.42}$$

which inevitably give a zero probability for the remaining Bell state too.

Let us move to the other case, namely events where two different detectors click. Given a first detection in mode k (event $e = 1_k$), we can get the state on the remaining modes before the second detection by summing all the elements on the k-th row and the k-th column of $\mathbf{M}$, except for the

two-photon term M_{kk} on the diagonal. As $\mathbf{M}$ is symmetric, we have:

$$\begin{aligned} |\Phi_k^{\beta_i}\rangle &= 2\left[\textstyle\sum_j M_{jk}^{\beta_i} c_j^\dagger - M_{kk}^{\beta_i} c_k^\dagger\right] |0\rangle \\ &= \left[(\overline{U}^\mathsf{T}\mathbf{W}^{\beta_i}\overline{\mathbf{u}}_k)\cdot\mathbf{c} - (\overline{\mathbf{u}}_k^\mathsf{T}\mathbf{W}^{\beta_i}\overline{\mathbf{u}}_k)\, c_k^\dagger\right] |0\rangle \\ &= \left[(\overline{U}^\mathsf{T}\mathbf{s}_k^{\beta_i})\cdot\mathbf{c} - (\overline{\mathbf{u}}_k\cdot\mathbf{s}_k^{\beta_i})\, c_k^\dagger\right] |0\rangle \end{aligned} \tag{2.43}$$

where $\mathbf{s}_k^{\beta_i} := \mathbf{W}^{\beta_i}\overline{\mathbf{u}}_k$. They occur with probability:

$$p(1_k|\beta_i) = \langle\Phi_k^{\beta_i}|\Phi_k^{\beta_i}\rangle = |\mathbf{s}_k^{\beta_i}|^2 - |\overline{\mathbf{u}}_k\cdot\mathbf{s}_k^{\beta_i}|^2. \tag{2.44}$$

We can bound the distinguishability of the four $|\Phi_k^{\beta_i}\rangle$ by showing that they are not all linearly independent. It can be easily checked that the determinant of the matrix formed by juxtaposing the four $\mathbf{s}_k^{\beta_i}$ is zero. Moreover, they all have the same nonzero norm $|\mathbf{s}_k^{\beta_i}|^2 = \frac{1}{2}|\overline{\mathbf{u}}_k|^2$. Then, the post-detection states satisfy:

$$\sum_{i=1}^{4} \lambda_{\beta_i} |\Phi_k^{\beta_i}\rangle = 0, \tag{2.45}$$

with at least two nonzero λ, i.e. at least two out of four states are linearly dependent. This implies that for each k, at best, only two states can be unambiguously discriminated from the others. Assuming we can do this perfectly for β_{a_k} and β_{b_k}, the success probability after detection in mode k averaged over β_i is:

$$p_{\text{succ}}(1_k) \leq \frac{1}{4}\left[p(1_k|\beta_{a_k}) + p(1_k|\beta_{b_k}) + 0 + 0\right] \leq \frac{1}{4}|\overline{\mathbf{u}}_k|^2 \tag{2.46}$$

which, summing over all the modes:

$$p_{\text{succ}} = \frac{1}{2}\sum_{k=1}^{m} p(1_k) \leq \frac{1}{8}\sum_{k=1}^{m} |\overline{\mathbf{u}}_k|^2, \tag{2.47}$$

where the additional $\frac{1}{2}$ factor avoids double counting the modes, noticing that a successful discrimination always involve two modes. The last sum is readily evaluated as the sum of the norm of the first four rows of U, which

is unitary:

$$\sum_{k=1}^{m} |\overline{\mathbf{u}}_k|^2 = \sum_{k=1}^{m} \Big(\sum_{j=1}^{4} |u_{jk}|^2 \Big) = \sum_{j=1}^{4} 1 = 4. \tag{2.48}$$

So, finally, $p_{\text{succ}} \leq 1/2$ and the bound is proven. □

2.3.5 A $p_{\text{succ}} = 1/2$ analyzer: the Innsbruck scheme

The bound imposed by Theorem 2 is tight. Indeed, there is a $p_{\text{succ}} = 1/2$ Bell analyzer which saturates it, which improves upon the scheme presented in Section 2.3.2. As far as we can tell, this strategy seems to have first appeared in an article by Weinfurter [Wei94] in 1994, and was seemingly independently rediscovered the following year by Braunstein and Mann [BM95]. As such, it is referred to as *Innsbruck* or *Braunstein-Mann* scheme in the literature on Bell measurement. The corresponding interferometer is shown in Fig. 2.5, in polarization encoding. While the bound is valid for much more general analyzers, the Innsbruck scheme already saturates it, despite consisting of just a single stage and not needing any auxiliary mode. The setup is the same as the $p_{\text{succ}} = 1/4$ protocol of Section 2.3.2, but the detectors have been upgraded to be capable of resolving polarization of the incoming photons—for example, using the equivalence in Diagram (2.17).

Instead of working backwards from the detection events, this time we look at how the Bell states evolve through the beamsplitter unitary (eq. 2.22). We already know about $|\psi^-\rangle$:

$$\begin{aligned} \frac{1}{\sqrt{2}}(h_1^\dagger v_2^\dagger - v_1^\dagger h_2^\dagger) \longrightarrow & \frac{1}{2\sqrt{2}}\Big[(h_{1'}^\dagger + h_{2'}^\dagger)(v_{1'}^\dagger - v_{2'}^\dagger) - (v_{1'}^\dagger + v_{2'}^\dagger)(h_{1'}^\dagger - h_{2'}^\dagger)\Big] \\ = & -\frac{1}{\sqrt{2}}(h_{1'}^\dagger v_{2'}^\dagger - v_{1'}^\dagger h_{2'}^\dagger), \end{aligned} \tag{2.49}$$

namely it is unaffected by the transformation (up to a global phase). Conversely, the symmetric states $|\psi^+\rangle, |\phi^\pm\rangle$ photons bunch in the same mode:

$$\begin{aligned} \frac{1}{\sqrt{2}}(h_1^\dagger v_2^\dagger + v_1^\dagger h_2^\dagger) \longrightarrow & \frac{1}{2\sqrt{2}}\Big[(h_{1'}^\dagger + h_{2'}^\dagger)(v_{1'}^\dagger - v_{2'}^\dagger) + (v_{1'}^\dagger + v_{2'}^\dagger)(h_{1'}^\dagger - h_{2'}^\dagger)\Big] \\ = & \frac{1}{\sqrt{2}}(h_{1'}^\dagger v_{1'}^\dagger - h_{2'}^\dagger v_{2'}^\dagger), \end{aligned} \tag{2.50}$$

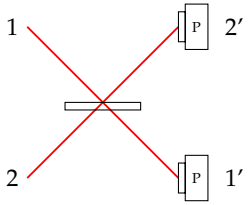


Figure 2.5: The Innsbruck scheme.

for $|\psi^+\rangle$, and

$$\frac{1}{\sqrt{2}}(h_1^\dagger h_2^\dagger \pm v_1^\dagger v_2^\dagger) \longrightarrow \frac{1}{2\sqrt{2}}\left[(h_{1'}^\dagger + h_{2'}^\dagger)(h_{1'}^\dagger - h_{2'}^\dagger) \pm (v_{1'}^\dagger + v_{2'}^\dagger)(v_{1'}^\dagger - v_{2'}^\dagger)\right]$$
$$= \frac{1}{2\sqrt{2}}(h_{1'}^{\dagger\,2} - h_{2'}^{\dagger\,2} \pm v_{1'}^{\dagger\,2} \mp v_{2'}^{\dagger\,2}), \tag{2.51}$$

for $|\phi^\pm\rangle$. It seems the situation has not improved, as all symmetric states still register two photons in the same mode. This time, though, we have access to the polarization of the output photons. Fortunately, we can see it is unchanged with respect to the input states. This is not a coincidence: the beamsplitter is an example of a *polarization-preserving* transformation. It is then immediate to distinguish the pair $|\phi^\pm\rangle$ from $|\psi^\pm\rangle$, by just checking whether the photons are detected with matching or opposite polarization. By joining this with photon bunching, it is then possible to single out the events coming from $|\psi^+\rangle$. This scheme thus allows for perfect unambiguous identification of 2 out of 4 Bell states, matching the upper bound.

2.3.6 Beating the 1/2 limit

The limits imposed by Calsamiglia and Lütkenhaus' result (Theorem 2) were without a doubt influential, possibly downsizing, at first, the promises of the linear optical platform. As with all theorems, the ways around it focus on relaxing one of the hypotheses. We already discussed how allowing for more complex components such as squeezing and nonlinear interactions can help; however, these effects are tiny and present challenges of their own. For example, Zaidi and van Loock show [ZL13] that reaching the

probability of 62.5% needs 8.69 dB of squeezing, which is experimentally quite demanding. One of the most compelling way to circumvent the bound is the use of auxiliary states. The authors had indeed already noticed [CL01] that full-fledged LOQC schemes like KLM and its improvements allow for a Bell analyzer arbitrarily close to perfect. Still, the enormous overhead in which these techniques incur means no scheme achieving more than 50% probability could be deemed practical. However, LOQC is general purpose: the hope was that a scheme focused on Bell measurement could achieve better efficiencies.

It might seem of very little value to add expensive resources, which could increase losses and do significantly increase complexity, just to achieve small gains in efficiency which, realistically, would always be far from 100%. Many experiments indeed only use the 25% for simplicity, at the cost of efficiency. After all, an argument could be made that any non-unit success probability gets quickly reduced to zero in practical applications, where one needs hundreds, if not thousands of Bell measurements. Yet in many of those cases (e.g. quantum repeaters [GEW21; HBE21]) any small improvement on the quality of the physical primitive leads to an exponential reduction of the protocol's total resources. Even better reasons for improving Bell measurement efficiency are fault tolerance and percolation schemes [Rud17], where overtaking a target threshold success probability is required for a phase transition to happen.

2.3.7 Grice's approach

Indeed, in 2011 a paper by Grice constructs a hierarchy of measurement strategies whose discrimination probability approaches unity [Gri11]. At each level of the hierarchy, increasingly complex entangled auxiliary states are needed, which are not necessarily easy to prepare. Nonetheless, the importance of Grice's approach is twofold: from the theoretical point of view, it is the first scheme breaking through the $1/2$ barrier without the KLM-like overhead, while most of the experimental interest resides in its relatively practical $p_{\text{succ}} = 3/4$ strategy ("level two" in the following).

Level one

The starting point of the hierarchy is the Innsbruck scheme, which consists of a single beamsplitter (Section 2.3.5). As usual, we label the output modes with a prime, i.e. $1', 2'$. In order to establish a convenient notation when extending the scheme to higher levels, Grice reclassifies the discriminating events in terms of n_h, n_v and $n_{[1']}$, which represent respectively the number of photons detected with horizontal and vertical polarization and the total number of photons (of both polarizations) counted by the detector on mode $1'$. We can immediately see that:

- n_h and n_v are odd for $|\psi^{\pm}\rangle$ and even for $|\phi^{\pm}\rangle$;
- $n_{[1']}$ is even for $|\psi^{+}\rangle$ and odd for $|\psi^{-}\rangle$.

Level two: auxiliary Bell pair

Grice realized that part of the degeneracy in the output states corresponding to $|\phi^{\pm}\rangle$ could be lifted by first interfering the unknown input state with a state from that same set. The rationale here is that he wants to preserve the "discrimination-by-polarization" feature of the Innsbruck scheme. This way, by keeping the interferometer polarization-preserving and adding a pair of auxiliary photons of matching polarization, the parity of n_h and n_v

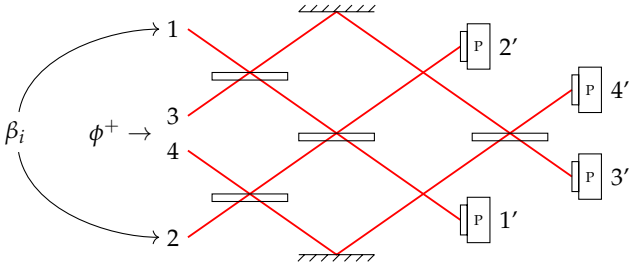


Figure 2.6: The "level two" Grice scheme, achieving $p_{\text{succ}} = 75\%$. The unknown state is input in modes 1 and 2, while modes 3 and 4 host the auxiliary state $|\phi^{+}\rangle$. Figure adapted from [Gri11], with a different layout better matching our presentation in the text.

is unaffected and will always let us separate $|\psi^{\pm}\rangle$ from $|\phi^{\pm}\rangle$. Referring to Definition 3 of a Bell analyzer, let us choose the auxiliary polynomial $Q = P_{\beta_1} = \frac{1}{\sqrt{2}}(h_3^\dagger h_4^\dagger + v_3^\dagger v_4^\dagger)$, so that the representation of the input state is $P_{\beta_i} P_{\beta_1}\ \forall i$. The idea is to first pairwise mix the photons in the unknown state and the ones in the auxiliary state, i.e. interfere mode 1 and mode 3 at a 50:50 beamsplitter and likewise mode 2 and 4, and then direct each output to a "level one" (i.e. Innsbruck) stage. The setup is illustrated in Fig. 2.6, and the corresponding unitary is[11]:

$$\begin{pmatrix} b_{1'}^\dagger \\ b_{2'}^\dagger \\ b_{3'}^\dagger \\ b_{4'}^\dagger \end{pmatrix} = \frac{1}{2} \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 \end{pmatrix} \begin{pmatrix} b_1^\dagger \\ b_2^\dagger \\ b_3^\dagger \\ b_4^\dagger \end{pmatrix}, \tag{2.52}$$

acting the same way on $b_i^\dagger = h_i^\dagger$ and $b_i^\dagger = v_i^\dagger$ for $i = 1, \ldots, 4$. The resulting detection patterns show that disambiguation between $|\psi^{\pm}\rangle$ is also preserved, this time by checking the parity of $n_{[1',3']}$—the total number of photons in modes $1' + 3'$. Now, the advantage over the first level is found by the outputs of $|\phi^{\pm}\rangle$: at variance with the previous case, some monomials are unique to $|\phi^{+}\rangle$ or $|\phi^{-}\rangle$. In particular, they are the terms corresponding to the detection of two horizontal and two vertical photons, happening with 50% probability. Then, the parity of $n_{[1',2']}$ can be used as a further source of discrimination. With this construction, Grice thus showed that an auxiliary Bell pair is sufficient to cut in half the failure probability, achieving on average $p_{\text{succ}} = 3/4 = 0.75$.

Levels $N \geq 3$

By adding more and more complex auxiliary states, Grice showed that it is possible to progressively halve the error probability of discriminating $|\phi^{+}\rangle$ from $|\phi^{-}\rangle$, while preserving the separability of $|\psi^{\pm}\rangle$. The interferometer

[11] For consistency with the rest of the Chapter, we use a real beamsplitter unitary here, as opposed to the complex one (eq. 2.13) which is used in Grice's paper, but the conclusions are otherwise the same.

unitary for the N-th level is defined by the following recursive relation:

$$U^{(N)} = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \otimes U^{(N-1)} = \frac{1}{\sqrt{2}} \begin{pmatrix} U^{(N-1)} & U^{(N-1)} \\ U^{(N-1)} & -U^{(N-1)} \end{pmatrix}, \quad (2.53)$$

which leads to a network on $m = 2^N$ spatial modes. The auxiliary state Γ_N is the product of $N-1$ polynomials Υ_j of the form:

$$\Gamma_N = \prod_{j=1}^{N-1} \Upsilon_j = \prod_{j=1}^{N-1} \frac{1}{\sqrt{2}} \left[h^\dagger_{(2^j)+1} \dots h^\dagger_{2^{j+1}} + v^\dagger_{(2^j)+1} \dots v^\dagger_{2^{j+1}} \right], \quad (2.54)$$

where each Υ_j is a GHZ-like state of 2^j photons:

$$\Upsilon_j |0\rangle = \frac{1}{\sqrt{2}} \Big(|00\dots0\rangle + |11\dots1\rangle \Big). \quad (2.55)$$

The only non-discriminating events—that is, the events which have a non-zero probability of occurring for at least two Bell states—happen when all photons impinge on the detectors with the same polarization. Grice's construction is able to suppress the probability of this kind of event happening, which decreases as $1/2^N$ at level N. This strategy achieves $p_{\text{succ}} = 1 - \frac{1}{2^N}$, using additional $\sum_{j=1}^{N-1} 2^j = 2^N - 2$ auxiliary entangled photons.

2.3.8 Ewert and van Loock's approach

Can we get rid of entanglement in the auxiliary state in order to break the $1/2$ barrier? In their 2014 paper Ewert and van Loock try to improve on Grice's design, providing a positive answer to the question [EL14]. Inspired by Grice's approach, they too build an hierarchy of Bell analyzers which use increasingly complex auxiliary states. Surprisingly, their $p_{\text{succ}} = 3/4$ scheme has a nice extra feature: its auxiliary state can be straightforwardly generated from 4 single, unentangled photons through a simple linear optical preprocessing. Higher-order schemes still seem to require entanglement. Interestingly, they show an unentangled scheme beating 75% (albeit only slightly), leaving the question open to what is the true limit of Bell analyzers without feedforward and unentangled auxiliary states.

At variance with the polarization encoding used in Grice's paper, the schemes in [EL14] are presented using path encoding (eq. 2.29), similar to the original paper by Weinfurter [Wei94]. Nonetheless, we chose to stick to polarization encoding here, in order to ease the comparison between the two. This choice also leads to more compact optical diagrams, halving the number of paths and beamsplitters which need to be drawn.

Level one

When converted to polarization encoding, Ewert and van Loock's $p_{\text{succ}} = 1/2$ protocol is the same as Grice's (the Innsbruck scheme), only consisting of a single beamsplitter and polarization-resolving detectors.

Level two: four auxiliary photons

Instead of building directly on the Innsbruck scheme, we start by going one step down, remembering what happens in the 25% scheme (Section 2.3.2). There, we did not look at the photons' polarization, resulting in only $|\psi^-\rangle$ being distinguishable from the other Bell states, as it sends one photon in each of the two output modes. The other three input states instead result in one of the following 2-photon states in either one of the two modes after the beamsplitter:

$$\psi^- \longrightarrow h^\dagger v^\dagger =: \alpha, \qquad \phi^\pm \longrightarrow \frac{1}{2}\Big[(h^\dagger)^2 \pm (v^\dagger)^2\Big] =: \gamma^\pm. \tag{2.56}$$

The idea here is to figure out a strategy to discriminate those, and then simply replicate it on both spatial modes—as we do not know *a priori* in which arm the photons will be found after the beamsplitter. The situation is depicted in Fig. 2.7. If we just choose to measure the photons' polarization, we recover the Innsbruck scheme. Instead, Ewert and van Loock chose a different route. They interfere at a beamsplitter the states in eq. (2.56) with the auxiliary state $\mathrm{Y}_1 = \frac{1}{2}[(h_2^\dagger)^2 + (v_2^\dagger)^2] = \gamma^+$. A lengthy but straightforward calculation shows that this lets them resolve γ^+ from γ^- half of the time, while α can be discriminated, as usual, by the polarization distribution. Therefore, two copies of Y_1 (one per arm) enable 75% overall success

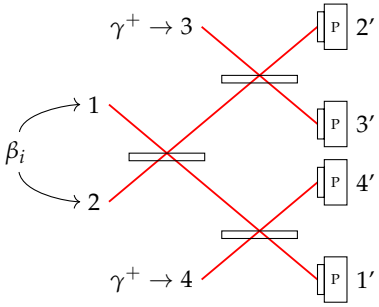
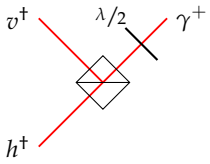


Figure 2.7: The Ewert–van Loock $p_{\text{succ}} = 75\%$ scheme, without the auxiliary state preprocessing. Figure adapted to polarization encoding from [EL14], where it is shown in path encoding.

probability.

So far, their approach looks like a "delayed" Grice, with a different set of states to be discriminated. If anything, the four-photon auxiliary state they need looks more complex than Grice's, which used a two-photon Bell pair. At second glance however (remembering the remarks about entangled modes vs entangled qubits in Section 2.2.6), Y_1 turns out to be much more linear optical-friendly than a Bell pair. Indeed, it can be deterministically produced from two single, unentangled photons:



(2.57)

Remembering that a half-wave plate acts like a beamsplitter on polarization modes, this is how the HOM effect looks like in polarization encoding. Through the transformation in Diagram (2.57), they can attain the same 75% success probability, *without* using entangled auxiliary states. The authors further note that this approach is reasonably robust to errors when non-ideal single-photon sources are employed.

Level $N \geq 3$: additional photons

The main task is to discriminate the states in eq. (2.56) which arise after the first beamsplitter. Then, similar to what happened in Level 2, the following only applies to the upper arm and the resource count will have to be doubled on the lower arm as well. The generalization to near-unit efficiency is again based on recursively embedding the previous interferometer, each time adding a more complex auxiliary state. In particular, the N^{th} level is built from two copies of the apparatus for the $(N-1)^{\text{th}}$ level (without the detectors):

1. The first copy acts on the $(N-1)^{\text{th}}$ input state, which is the product of one of the unknown states $\{\alpha, \gamma^{\pm}\}$ on the first input mode and the auxiliary state $\mathrm{Y}_1 \cdots \mathrm{Y}_{N-2}$, where

$$\mathrm{Y}_j = \frac{1}{\sqrt{2}} \left[\prod_{k=2^{j-1}+1}^{2^j} \frac{(h_k^\dagger)^2}{\sqrt{2}} + \prod_{k=2^{j-1}+1}^{2^j} \frac{(v_k^\dagger)^2}{\sqrt{2}} \right] \tag{2.58}$$

   on modes 2 to 2^{N-2}.

2. The second copy is just fed the next auxiliary state Y_{N-1} up to mode 2^{N-1}.

3. The output modes of the two copies are then pairwise mixed at beamsplitters.

When completed by duplicating on the lower arm, this approach uses the same number of spatial modes as Grice's but twice as many auxiliary photons, respectively $m = 2^N$ and $k = 2(2^N - 2)$, in order to attain the same success probability $p_{\text{succ}} = 1 - \frac{1}{2^N}$. Unfortunately, the states Y_j for $j \geq 2$ do not appear to enjoy the same linear-optical manufacturability of Y_1. A lengthy calculation shows that a probability of $25/32 \backsimeq 0.78$ can be achieved by substituting $|\mathrm{Y}_2\rangle$ with $|\mathrm{Y}_1\rangle\,|\mathrm{Y}_1\rangle$ in the $N = 3$ level. Even if this minuscule improvement over $3/4$ requires $2(2+4) = 12$ single photons, it shows that the former is not a probability barrier for unentangled auxiliary states, and gives hope for better Bell analyzers using only single photons.

Optimality

At variance with the limits imposed by Theorem 2, no probability bound is known when auxiliary states are introduced. Consequently, the schemes in Sections 2.3.7 and 2.3.8 are not known to be optimal in terms of photon consumption, or even assuming the specific auxiliary state used. In fact, both protocols appear to have arisen from symmetry considerations, while the space of *possible* protocols remain largely unexplored.

We now present our research, originally motivated by the desire of collecting evidence, through analytical and numerical means, about the optimality of known schemes, while potentially looking for better ones.

2.4 A new polarization-preserving bound

As we already noted, an interesting feature of all the schemes considered up to now is that the interferometers which implement them preserve the polarization of the input photons. In particular, they are strictly *polarization-independent*, meaning they act identically on photons of any polarization.

Remark. For schemes defined in path encoding, this property does not look very natural. Formally, if the modes can be divided equally in two groups A and B which never mix—in other words, if the interferometer's unitary U can be decomposed in U_A and U_B acting on the respective mode group—then there is an equivalent polarization-preserving interferometer for polarization-encoded qubits through the conversion in Diagram (2.16). This is indeed the case in Ewert and van Loock's original protocols (but not in the polarization-encoded version we give in Section 2.3.8), where the interferometer acts the same way on odd- and even-numbered modes.

While clearly not representative of all linear optical transformations, polarization-preserving interferometers proved to be useful for a Bell analyzer: they ensure we can separate $|\phi^{\pm}\rangle$ and $|\psi^{\pm}\rangle$ by looking at the polarization distribution of the detected photons. It turns out that this subset of interferometers has nice analytical properties: the condition is restrictive enough that it is possible to derive a non-trivial upper bound to the discrimination probability.

However, from the experimental point of view it might be difficult to justify this restriction. Indeed in a typical setup on an optical table no significant simplification is achieved if, for example, half- and quarter-wave plates are not used. Nonetheless, on the integrated optics platform there might be a tangible advantage in only having to fabricate beamsplitters (which are often just properly positioned waveguides), especially for a fixed-function device like a Bell analyzer[12]

The proof we give is not constructive, and we do not expect the bound to be saturated by explicit linear optical schemes for all auxiliary states. The known schemes however match the bound's value.

[12]When using path encoded qubits the equivalent requirement of not mixing even and odd modes looks even more difficult to justify as advantageous.

Theorem 3 (Polarisation-preserving upper bound)**.** *Let $\mathcal{B}$ be a Bell analyzer which preserves the polarization of the input photons, equipped with the k-photon auxiliary state $|\Gamma\rangle$. Then the probability of unambiguous Bell-state discrimination is upper bounded by:*

$$p_{\text{succ}} \leq \frac{1}{2} + \frac{1}{2} \sum_{\lambda=0}^{k-2} \min\{|\gamma_\lambda|^2, |\gamma_{\lambda+2}|^2\}, \tag{2.59}$$

where the γ_λ are the coefficients of the expansion $|\Gamma\rangle = \sum_{\lambda=0}^{k} \gamma_\lambda |\mathrm{Y}_\lambda\rangle$ over orthogonal states $|\mathrm{Y}_\lambda\rangle$ of λ horizontally and $k-\lambda$ vertically polarized photons.

Theorem 3 is not straightforward to use *as is*, so we will derive later two corollaries which establish looser bounds based on more easily obtainable properties of the auxiliary state (e.g. its photon content).

Proof. Schematically, the proof works by disentangling the input states using a particular projection and then bounding their quantum mechanical discrimination probability (Fig. 2.8).

First, suppose we run the first stage of $\mathcal{B}$ until just before the measurement, and we performed at that point a quantum non-demolition (QND) measurement of n_h, the number of horizontally polarized photons.[13] Since the photon detectors at the end are polarization-resolving by construction, this information could have been obtained later, meaning that the addition of the QND projection does not change the output probabilities. Now we can make use of the polarization-preserving hypothesis, which ensures the polarization measurement commutes with the action of the interferometer (Fig. 2.8). As such, we could have performed it *immediately* on the input state. We can expand the auxiliary state (consisting of k total photons[14]) in terms of states $|\mathrm{Y}_\lambda\rangle$, eigenstates of n_h of definite number λ of horizontally

[13] While the reader might have been under the impression that measuring the state of a photon always destroys it, this is not necessarily the case: indeed, quantum mechanics allows such non-destructive measurement. Realizing this kind of measurement in the laboratory is really challenging, due to how fragile photon states are. In 2012, a Nobel prize was awarded to Serge Haroche for pioneering such experimental techniques [RBH01]. Here, QND measurement is only used as a proof device.

[14] The assumption that the auxiliary state has a well defined and known photon number is not a restriction, as can be easily deduced from the fact that the total photon number operator, which we measure, commutes with linear optical evolutions, and the fact that the Bell states have exactly two photons.

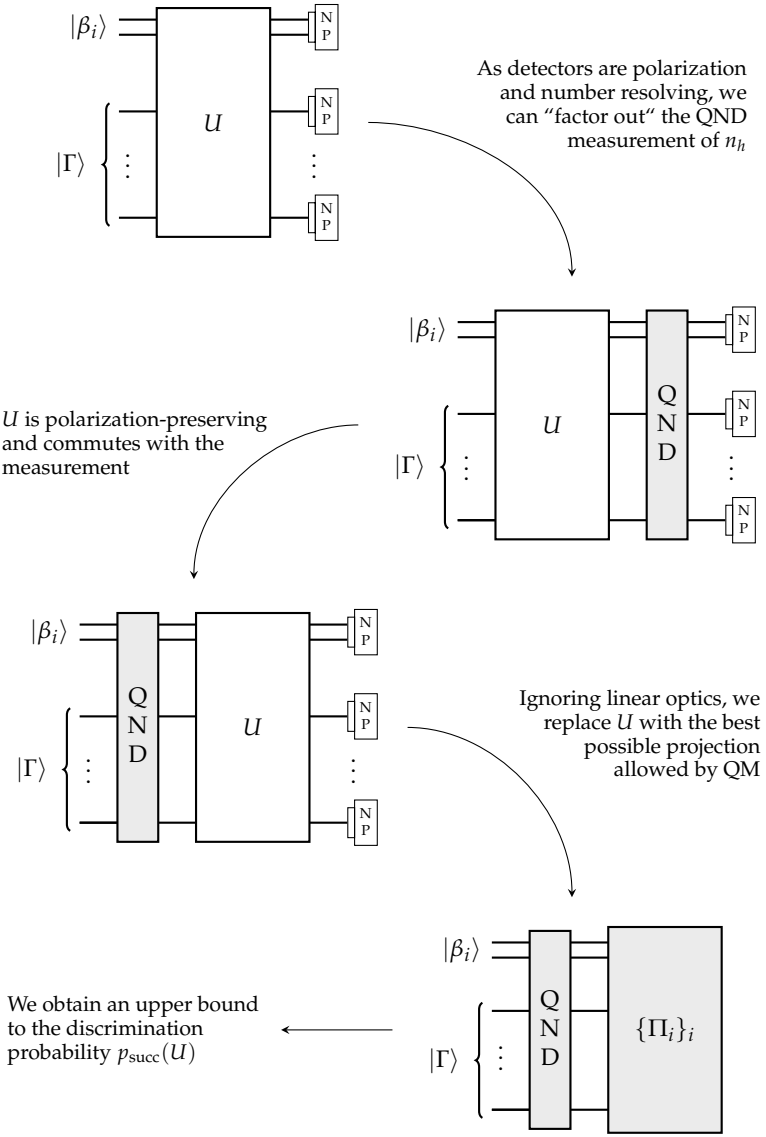


Figure 2.8: Overview of the proof. Nonlinear operations are shaded.

polarized photons:

$$|\Gamma\rangle = \sum_{\lambda=0}^{k} \gamma_\lambda \, |\mathrm{Y}_\lambda\rangle \, . \tag{2.60}$$

For convenience, we define $\gamma_\lambda = 0$ for $\lambda < 0$ and $\lambda > k$ with a slight abuse of notation. The four complete inputs to the Bell analyzer are then $(k+2)$-photon states, which we can also write as a sum of eigenstates of n_h:

$$\begin{aligned} |\phi^\pm\rangle \, |\Gamma\rangle &= \left[\frac{1}{\sqrt{2}}(h_1^\dagger h_2^\dagger \pm v_1^\dagger v_2^\dagger)\,|0\rangle\right] \sum_{\lambda=0}^{k} \gamma_\lambda \, |\mathrm{Y}_\lambda\rangle \\ &= \sum_{\bar{\lambda}=0}^{k+2} \left(\frac{\gamma_{\bar{\lambda}-2}}{\sqrt{2}} \, |HH\rangle \, |\mathrm{Y}_{\bar{\lambda}-2}\rangle \pm \frac{\gamma_{\bar{\lambda}}}{\sqrt{2}} \, |VV\rangle \, |\mathrm{Y}_{\bar{\lambda}}\rangle \right), \end{aligned} \tag{2.61}$$

$$\begin{aligned} |\psi^\pm\rangle \, |\Gamma\rangle &= \left[\frac{1}{\sqrt{2}}(h_1^\dagger v_2^\dagger \pm v_1^\dagger h_2^\dagger)\,|0\rangle\right] \sum_{\lambda=0}^{k} \gamma_\lambda \, |\mathrm{Y}_\lambda\rangle \\ &= \sum_{\bar{\lambda}=1}^{k+1} \frac{\gamma_{\bar{\lambda}-1}}{\sqrt{2}} \Big[|HV\rangle \pm |VH\rangle \Big] \, |\mathrm{Y}_{\bar{\lambda}-1}\rangle \, , \end{aligned} \tag{2.62}$$

where each term in the sum has $\bar{\lambda}$ horizontally polarized photons and the mode numbering is implied.

Notice that after the QND measurement the inputs corresponding to $|\psi^+\rangle$, $|\psi^-\rangle$ and $|\phi^\pm\rangle$ collapse onto three orthogonal subspaces. Clearly, the Bell analyzer cannot distinguish them better than what is allowed by quantum mechanics; this way we can get a bound on the discrimination probability, relaxing the restrictions imposed by linear optics. The orthogonal subspaces can be perfectly discriminated, and the only remaining ambiguity is among the (unnormalized) states $|\Lambda^\pm\rangle$ arising from input $|\phi^\pm\rangle \, |\Gamma\rangle$ after the projection:

$$|\Lambda^\pm\rangle = \frac{\gamma_{\bar{\lambda}-2}}{\sqrt{2}} \, |HH\rangle \, |\mathrm{Y}_{\bar{\lambda}-2}\rangle \pm \frac{\gamma_{\bar{\lambda}}}{\sqrt{2}} \, |VV\rangle \, |\mathrm{Y}_{\bar{\lambda}}\rangle \, , \tag{2.63}$$

which are not orthogonal. Their squared norm and overlap are:

$$\|\Lambda\|^2 = \langle\Lambda^+|\Lambda^+\rangle = \langle\Lambda^-|\Lambda^-\rangle = \frac{1}{2}\Big(|\gamma_{\bar{\lambda}-2}|^2 + |\gamma_{\bar{\lambda}}|^2\Big), \tag{2.64}$$

$$|\langle\Lambda^+|\Lambda^-\rangle| = \frac{1}{2}\Big||\gamma_{\bar{\lambda}-2}|^2 - |\gamma_{\bar{\lambda}}|^2\Big|. \tag{2.65}$$

As shown in Section 1.1.5, they can be unambiguously distinguished with probability at most:

$$p_{\text{disc}}(\bar{\lambda}) \leq \|\Lambda\|^2 - |\langle\Lambda^+|\Lambda^-\rangle| = \min\{|\gamma_{\bar{\lambda}-2}|^2, |\gamma_{\bar{\lambda}}|^2\}, \tag{2.66}$$

where we adapted eq. (1.23) to unnormalized states. Assuming perfect discrimination of $|\psi^+\rangle$ and $|\psi^-\rangle$, and summing over $\bar{\lambda}$, the total success probability of the Bell analyzer is upper bounded by:

$$\begin{aligned} p_{\text{succ}} &\leq \frac{1}{4}\Big(1+1+2\sum_{\bar{\lambda}=0}^{k+2}\min\{|\gamma_{\bar{\lambda}-2}|^2, |\gamma_{\bar{\lambda}}|^2\}\Big) \\ &= \frac{1}{2} + \frac{1}{2}\sum_{\bar{\lambda}=2}^{k}\min\{|\gamma_{\bar{\lambda}-2}|^2, |\gamma_{\bar{\lambda}}|^2\}, \end{aligned} \tag{2.67}$$

where we excluded from the sum the terms which are guaranteed to be zero. Then, eq. (2.59) follows by the substitution $\lambda = \bar{\lambda} - 2$. □

We can derive a more insightful version of the bound by noticing that, since the minimum is taken every two values of λ, it is useful to split the sum over even and odd indices:

$$\begin{aligned} p_{\text{succ}} \leq \frac{1}{2} &+ \frac{1}{2}\sum_{\lambda\ \text{even}}\min\{|\gamma_\lambda|^2, |\gamma_{\lambda+2}|^2\} \\ &+ \frac{1}{2}\sum_{\lambda\ \text{odd}}\min\{|\gamma_\lambda|^2, |\gamma_{\lambda+2}|^2\}. \end{aligned} \tag{2.68}$$

Among each group, it is not difficult to convince oneself that every γ_λ appear once in the sum *except* for local maxima, which are excluded, and local minima, which are counted twice.[15] Remembering that $\sum_\lambda |\gamma_\lambda|^2 = 1$, we have:

$$p_{\text{succ}} \leq 1 - \frac{1}{2}\Big(\sum_{\substack{\lambda\ \text{even}\\ |\gamma_\lambda|^2\ \text{loc max}}}|\gamma_\lambda|^2 - \sum_{\substack{\lambda\ \text{even}\\ |\gamma_\lambda|^2\ \text{loc min}}}|\gamma_\lambda|^2 + \sum_{\substack{\lambda\ \text{odd}\\ |\gamma_\lambda|^2\ \text{loc max}}}|\gamma_\lambda|^2 - \sum_{\substack{\lambda\ \text{odd}\\ |\gamma_\lambda|^2\ \text{loc min}}}|\gamma_\lambda|^2\Big), \tag{2.69}$$

[15] In order to deal with the boundaries (i.e. if they should be counted as minima or maxima), remember that we defined $\gamma_\lambda = 0$ for $\lambda < 0$ and $\lambda > k$. Consecutive optima with the same value are only included (or excluded) once.

which leads to the following corollary, establishing a (looser) bound on p_{succ}.

Corollary 3.1. *Under the same notation of Theorem 3, the unambiguous discrimination probability of a polarization-preserving Bell analyzer with auxiliary state* $|\Gamma\rangle = \sum_\lambda \gamma_\lambda |\mathrm{Y}_\lambda\rangle$ *is at most:*

$$p_{\text{succ}} \leq 1 - \frac{1}{2}\Big(\max_{\lambda \text{ even}} |\gamma_\lambda|^2 + \max_{\lambda \text{ odd}} |\gamma_\lambda|^2\Big). \tag{2.70}$$

Proof. As $|\gamma_\lambda|^2 = 0$ for $\lambda \notin [0,k]$ and $|\gamma_\lambda|^2 \geq 0$ otherwise, the number of local optima in each index group (λ even and λ odd) is always odd. We can always pair each local minimum λ_{min} in eq. (2.69) with a neighboring local maximum λ_{max} such that $|\gamma_{\lambda_{\text{max}}}|^2 - |\gamma_{\lambda_{\text{min}}}|^2 > 0$. Therefore, the extra optimum has to be a local maximum. The corollary is then proved by only keeping the global maximum for each index group. □

The bound reduces to eq. (2.59) whenever there is a single local maximum among each index group, which is the case for many of the auxiliary states considered up to now.

2.4.1 Bound based on photon number

In order to compare various schemes which use different auxiliary states, it is useful to obtain a bound which is independent on the specific form of $|\Gamma\rangle$, being instead only function of its photon count k.

- If k is odd, we have at most $\frac{k+1}{2}$ even and $\frac{k+1}{2}$ odd values of λ for which $|\gamma_\lambda|^2 \neq 0$. Defining $\sum_{\lambda \text{ odd}} |\gamma_\lambda|^2 = S_{\text{odd}}$, then $(\max_{\lambda \text{ odd}} |\gamma_\lambda|^2)$ has to be at least $\frac{2\,S_{\text{odd}}}{k+1}$, and similarly for S_{even}. Given that $S_{\text{odd}} + S_{\text{even}} = 1$, the sum of the maxima in eq. (2.70) is lower bounded by $\frac{2}{k+1}$, and we get:
$$p_{\text{succ},\, k \text{ odd}} \leq 1 - \frac{1}{k+1}, \tag{2.71}$$
which can for example be saturated by auxiliary states where all eigenstates $|\mathrm{Y}_\lambda\rangle$ are equiprobable, i.e. $|\gamma_\lambda|^2 = \frac{1}{k+1}\ \forall\lambda$.

- Similarly, if k is even, we have at most $\frac{k}{2}$ odd and $\frac{k}{2}+1$ even λs with $|\gamma_\lambda|^2 \neq 0$. This time we have:

$$p_{\text{succ},\, k \text{ even}} \leq 1 - \frac{1}{k+2}, \tag{2.72}$$

  which is best saturated when all odd components are zero and the rest are equiprobable:

$$|\gamma_\lambda|^2 = \begin{cases} \frac{2}{k+2} & \lambda \text{ even}, \\ 0 & \lambda \text{ odd}. \end{cases} \tag{2.73}$$

Merging the two cases, we have our simplest (but "loosest") bound:

Corollary 3.2 (Photon-number based bound)**.** *The unambiguous discrimination probability of a polarization-preserving Bell analyzer with an auxiliary k-photon state is at most:*

$$p_{\text{succ}} \leq 1 - \frac{1}{\lceil k+1 \rceil_{\text{even}}}, \tag{2.74}$$

where $\lceil \cdot \rceil_{\text{even}}$ *is the smallest even integer greater or equal to its argument.*

In the trivial case of no auxiliary photons at all, i.e. $k = 0$, this result matches the Calsamiglia-Lütkenhaus $1/2$ bound as expected. It is not difficult to verify that in this case the proof reduces to just a couple of lines (Section 2.6.1), which when compared to the involved Theorem 2 constitutes further evidence of the strength (for better or for worse) of the polarization-preserving restriction. Another interesting observation is that, at least for this kind of Bell analyzers, a single extra photon does not help: the smallest state beating $1/2$ has to contain at least two photons. Despite all the simplifications made along the way, though, the bound still proves to be useful! In fact, we know of at least one series of strategies which saturate it for all k which are powers of 2, namely the Grice schemes in Section 2.3.7. For the states in Ewert and van Loocks schemes, instead, the bound in eq. (2.59) is tight, while eq. (2.74) is not. We will come back to this in Section 2.6.3.

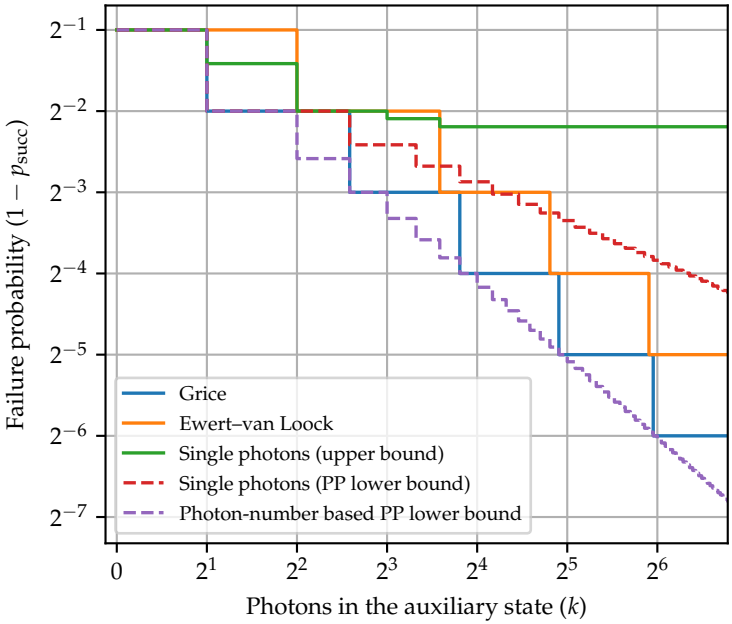


Figure 2.9: The failure probability of known explicit schemes (solid lines) and our polarization-preserving (PP) bound (dashed lines). The green and red curves (the single photons schemes) are based on a more careful analysis of Ewert and van Loock's single photons results, and how our bound can be specialized to them (Section 2.6.3).

2.5 Computer to the rescue: an optimization approach

Notwithstanding the importance of analytical results like our polarization-preserving bound, the restriction imposed by the proof technique is not particularly desirable in a practical setting. While obtaining a non-trivial upper bound for generic BA would be ideal, the only known proof in this setting (Theorem 2) does not easily generalize to networks with more than two photons. Moreover, it is a non-constructive proof: it just so happens that the bound can be saturated by a simple known scheme (the Innsbruck scheme). As a matter of fact, to our knowledge the space of possible BA had not been systematically explored before our work; each new scheme found would represent a lower bound on the achievable Bell measurement efficiency for a given auxiliary resource.

With this in mind we built `solon` (Simulation Of Linear Optical Networks), a custom software tailored to the optimization of Bell analyzers. With it, we set ourselves three main goals:

- Collecting numerical evidence towards the optimality of Grice's and Ewert and van-Loock's strategies;
- Finding new interesting BA, using different kinds of auxiliary resources;
- Implement an easy-to-use tool to work with linear optical networks, opening the doors to applications other than Bell measurement.[16]

In the following, we present the inner workings of the program and discuss some of the challenges we had to overcome in order to decrease its complexity to a manageable level. We then analyze in detail the results of the numerical investigation in Section 2.6, comparing the data with the analytical bound; we organize our findings in Table 2.3.

2.5.1 Overview

Our approach is composed of multiple steps (Fig. 2.10).

[16]The extension to generic inputs and general packaging and code cleanup was delivered by Kim Vallée as part of his *Licence 3* internship.

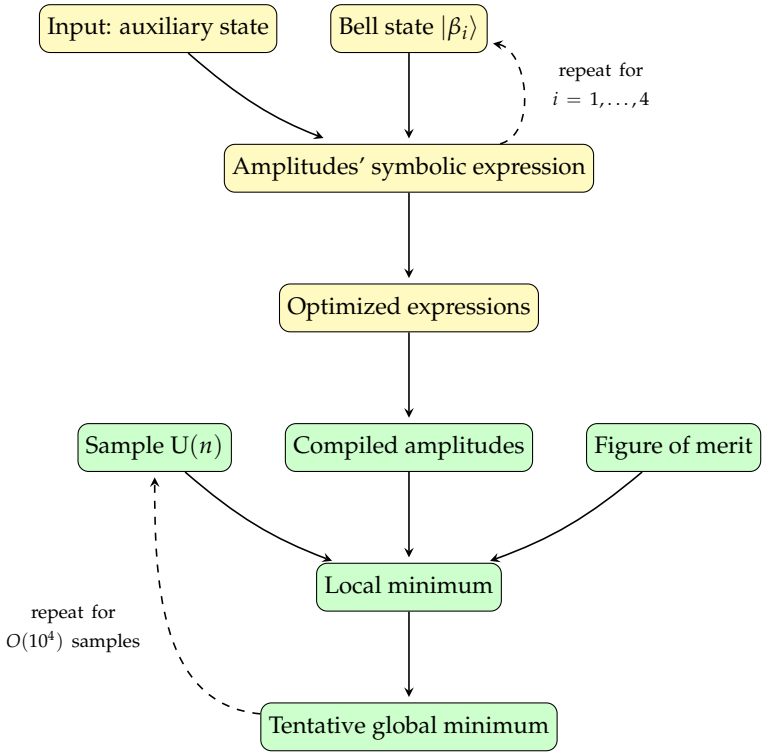


Figure 2.10: The pipeline used by `solon`. The yellow boxes are symbolic calculations done in `SymPy`, while the green ones correspond to numerical calculations in a combination of `numpy` and `C/theano`.

1. For each auxiliary state we analyze, and for each input Bell state, we generate a symbolic expression for the amplitudes of all output events in terms of the entries of the unitary U associated to the BA.[17]

2. Those symbolic expressions, along with their gradient with respect to the entries of U, are optimized in order to reduce the number of elementary operations.

3. The optimized functions are then automatically translated into a low-level language and compiled.

Due to the heavily non-smooth character of the success probability for unambiguous measurements, we construct a meaningful figure of merit, function of the previously obtained probability amplitudes.

4. A constrained numerical optimization using a nonlinear method is then performed multiple times from randomly sampled starting points.

All the above steps are automated; the only "manual" input is the auxiliary polynomial Q. While this can seem at first glance an overkill *brute-force* approach, the problem present important symmetries that we exploit, gaining up to two orders of magnitude in computation time. Once an optimum has been reached, the output looks like the one in Fig. 2.11, which is a detailed description of the BA's performance.

Choice of programming language

In accordance with a rising trend in the scientific domain, we chose to write `solon` in Python. As a high-level language, it seems unsuitable for intense computational tasks; however, it comes with a good selection of open-source scientific libraries written in C which implement a large portion of the heavy number crunching we need, and offers tools to automatically produce efficient compiled code when needed. Additionally, Python enforces good programming practices and the resulting code is clear and easy to understand, in a way that eases the process of checking its correctness, which is an

[17] In the following, we only consider single-stage Bell analyzers, i.e. we ignore the possibility of feedforward operations.

```
~$ python 4_modes_no_aux.py

(0, 0, 1, 1) [0._____ 0._____ 0.5____ 0._____]
(0, 1, 0, 1) [0._____ 0._____ 0._____ 0._____]
(0, 1, 1, 0) [0._____ 0._____ 0._____ 0.5____]
(1, 0, 0, 1) [0._____ 0._____ 0._____ 0.5____]
(1, 0, 1, 0) [0._____ 0._____ 0._____ 0._____]
(1, 1, 0, 0) [0._____ 0._____ 0.5____ 0._____]
(0, 0, 0, 2) [0.25___ 0.25___ 0._____ 0._____]
(0, 0, 2, 0) [0.25___ 0.25___ 0._____ 0._____]
(0, 2, 0, 0) [0.25___ 0.25___ 0._____ 0._____]
(2, 0, 0, 0) [0.25___ 0.25___ 0._____ 0._____]
Discrimination probability (zero=1e-08):
             [0._____ 0._____ 1._____ 1._____], p = 0.5
```

Figure 2.11: The output of the simulator when it finds the Innsbruck scheme, if initialized with no auxiliary modes. The underscore "_" replaces the digit "0" in the decimal expansion for improved legibility. From left to right: the output events and their probability on the four different inputs, in the order $(\phi^+, \phi^-, \psi^+, \psi^-)$.

important part of the scientific review process. Following a common design pattern, we started by building a naïvely written version which (slowly) worked for a particular use case, which underwent subsequent refining and generalizations in order to improve speed.[18] The identification of the slow parts of the program was possible thanks to the use of *code profiling* tools, which monitor the execution time of the various portions of the code.

2.5.2 Symbolic computation

The purpose of the method presented here is to provide the optimum-finding algorithm described in the next subsection with a fast, optimized function returning all the detection event probabilities, along with their gradients with respect to the entries of U, from which a figure of merit $f(U)$ will be constructed. We already know a closed formula for the amplitudes, as they are related to the permanents of submatrices of U. However,

[18] From a Donald Knuth's catchphrase, "premature optimization is the root of all evil" [Knu74]

working on optimizing the symbolic expressions separately enabled us to carefully analyze the specific problem and implement some analytical shortcuts with which we could speed up the search for optima. Symbolic operations are significantly slower than their numerical counterpart, but they have to be performed just once per auxiliary state.

We coded this part of the program in `SymPy` [Meu+17], an open-source *Computer algebra system* (CAS) for Python, which provides similar functionalities of proprietary tools like Mathematica or Matlab. `SymPy` is written in pure python, which makes extending its features easy for us; however, it is not ideal for large calculations *as is*, due to the overhead of Python code. Fortunately, the project `PyPy` [Ped] provides a Just-in-Time compiler for Python which helps achieving similar performance to the above mentioned proprietary software, albeit at the cost of some memory overhead.[19] In the following, we use the same notation conventions for the input and output polynomials introduced in Definition 3. In particular, we work with the path-encoded Bell states of eq. (2.29); this choice better reflects the loss of the polarization-preserving structure for the generic interferometers we want to explore. Here, m is the number of modes while k refers to the number of photons in the auxiliary state, making $k+2$ the total number of photons in the BA. Briefly, this section of the code start by taking the symbolic expression of the four input polynomials $P_{\beta_i}(a_1^\dagger, \ldots, a_4^\dagger)Q(a_5^\dagger, \ldots, a_m^\dagger)$ and performing the substitution $\mathbf{a} = U\mathbf{c}$ of eq. (2.8), where each entry u_{ij} of U is itself defined as a symbolic variable. It automatically expands the expression in the new indeterminates and collects the coefficients of the output polynomial $T(c_1^\dagger, \ldots, c_m^\dagger)$, of which there are:

$$N = \binom{m+k+1}{k+2} \tag{2.75}$$

for each Bell state. The symbolic expressions (in u_{ij}) for the amplitudes $\alpha_e(U)$ of each detection event e are obtained after correctly accounting for the bosonic normalization factor for the monomials containing more than

[19] As an example, finding the second level of the Ewert–van Loock strategy takes 7 minutes and 30 seconds and about 150 MB of RAM on our laptop (the specifications of which are reported in Table 2.2), using the standard cPython interpreter. Using `PyPy` the time is cut down to 45 seconds, with a memory consumption of 250 MB.

one photon in the same mode, as in eq. (2.11). We can directly compile an expression for the probabilities $p(e|\beta_i)$ by taking their square moduli, and this is indeed the strategy we followed initially. However, it turns out that dealing directly with the amplitudes at this stage is much more efficient when the gradients are taken into account.

A simple setup

Let us consider the simplest example, to better understand the principle. Consider a network with $m = 4$ modes, with $|\phi^+\rangle$ as input and no auxiliary state, i.e. $P_{\phi^+} = \frac{1}{\sqrt{2}}(a_1^\dagger a_3^\dagger + a_2^\dagger a_4^\dagger)$, $Q = 1$ and $k = 0$. At the substitution step, `solon` computes the expression:

$$\begin{aligned} T(c_1^\dagger, c_2^\dagger, c_3^\dagger, c_4^\dagger) = &\frac{1}{\sqrt{2}} \Big(\sum_{j_1} u_{1j_1}\, c_{j_1}^\dagger\Big)\Big(\sum_{j_2} u_{3j_2}\, c_{j_2}^\dagger\Big) \\ &+ \frac{1}{\sqrt{2}} \Big(\sum_{j_3} u_{2j_3}\, c_{j_3}^\dagger\Big)\Big(\sum_{j_4} u_{4j_4}\, c_{j_4}^\dagger\Big). \end{aligned} \tag{2.76}$$

After expanding all the products, we obtain a polynomial of degree $k+2 = 2$ in four variables, with $N = 10$ terms. The coefficient of the monomial $c_1^\dagger c_3^\dagger$ is, for example, the amplitude of the detection event 1010. After expanding eq. (2.76), all the events' amplitudes looks something like the ones in Table 2.1. Analogous expressions are produced for ϕ^-, ψ^+ and ψ^-. While the above expansion could have been done by hand, the complexity quickly increases: already the smallest interesting setup with $Q = a_5^\dagger$ (one auxiliary photon) results in four degree-3 polynomial in 5 variables, with $N = 35$ terms.

2.5.3 Function compilation

In order to get a compiled, serializable and importable function for the amplitudes, we wrote a custom routine based on `f2py` and `SymPy` internal code-generation tools, which we later replaced by the third party library `theano` [Tea16] once SymPy integration matured enough. The latter implements additional optimizations based on the construction of a graph-based

$$
\begin{aligned}
2000 &\longrightarrow u_{11}u_{31} + u_{21}u_{41} \\
0200 &\longrightarrow u_{12}u_{32} + u_{22}u_{42} \\
0020 &\longrightarrow u_{13}u_{33} + u_{23}u_{43} \\
0002 &\longrightarrow u_{14}u_{34} + u_{24}u_{44} \\
1100 &\longrightarrow (u_{11}u_{32} + u_{12}u_{31} + u_{21}u_{42} + u_{22}u_{41})/\sqrt{2} \\
1010 &\longrightarrow (u_{11}u_{33} + u_{13}u_{31} + u_{21}u_{43} + u_{23}u_{41})/\sqrt{2} \\
1001 &\longrightarrow (u_{11}u_{34} + u_{14}u_{31} + u_{21}u_{44} + u_{24}u_{41})/\sqrt{2} \\
0110 &\longrightarrow (u_{12}u_{33} + u_{13}u_{32} + u_{22}u_{43} + u_{23}u_{42})/\sqrt{2} \\
0101 &\longrightarrow (u_{12}u_{34} + u_{14}u_{32} + u_{22}u_{44} + u_{24}u_{42})/\sqrt{2} \\
0011 &\longrightarrow (u_{13}u_{34} + u_{14}u_{33} + u_{23}u_{44} + u_{24}u_{43})/\sqrt{2}
\end{aligned}
$$

Table 2.1: Symbolic expressions for the amplitudes of a low-dimensional example calculation. The polynomials are hard-compiled into a fast C expression which will be used by the gradient descent evaluation routine.

function representation. The two approaches turned out to have similar performances: while our implementation is more memory-efficient as no intermediate graph representation has to be built, we chose `theano` for the minor speed benefits and the need to maintain less code. To give a sense of the scale of the polynomials involved, the amplitudes for the largest computation we managed to complete—involving $m = 16$ modes and $2 + k = 8$ total photons[20]—contains about $1.8 \cdot 10^6$ symbolic operations, while their Jacobian totals $\sim 17 \cdot 10^6$ operations.

2.5.4 Optimizations and symmetries

Strictly speaking, all the numerical part of the program will need for the maximizing our figure of merit is an array of functions like those in Table 2.1. However, there are some insights we can exploit to save computation time and improve convergence.

[20] This is for the computation of the $N = 3$ level of Grice's schemes

Gradients

The most expensive part of the gradient descent method we employ is the computation of the gradient of the objective function $f(U)$ at each iteration. If it is not provided with an analytical expression, the algorithm can estimate the partial derivatives by evaluating the figure of merit at nearby points. Using the *finite difference* method, this amounts to two additional evaluations of $f(U)$ per variable, which is fine for most problems for which the objective is fast to compute, but very expensive in our case. Furthermore, the Jacobian would only be calculated to some fixed accuracy which depends on the size of the step chosen for the finite difference, which adds noise to the optimization algorithm and worsens its convergence. We can exploit the fact that we have the expensive parts of the figure of merit at our disposal in symbolic form in order to also obtain a fast, analytical compiled expression for its gradient. Ultimately, $f(U)$ depends on the square moduli of the amplitudes $|\alpha_e(U)|^2$, which are not differentiable with respect to the entries u_{ij} of U in the complex sense. In general this is not an issue; when dealing with functions of complex input, a common practice in optimization is to split the independent variable into a real and an imaginary part, with respect to which the derivatives of $|\alpha_e(U)|^2$ are well defined. While in general the derivatives involving the square modulus can be cumbersome, the complex-valued amplitudes $\alpha_e(U)$ themselves are always holomorphic functions of u_{ij}—specifically, complex polynomials. We can use this property to obtain a compact expression for the gradient $\nabla f(U)$.

In general, for a holomorphic function $\alpha(u_{00}, u_{01}, \dots)$:

$$\frac{\partial \alpha}{\partial u_{ij}} = \frac{\partial \alpha}{\partial \Re\{u_{ij}\}} = -i \frac{\partial \alpha}{\partial \Im\{u_{ij}\}} \tag{2.77}$$

and, given $|\alpha|^2 = \alpha\alpha^*$:

$$\frac{\partial |\alpha|^2}{\partial \Re\{u_{ij}\}} = \frac{\partial (\alpha\alpha^*)}{\partial \Re\{u_{ij}\}} = \alpha \frac{\partial \alpha^*}{\partial u_{ij}} + \frac{\partial \alpha}{\partial u_{ij}} \alpha^* = 2\,\Re\left\{ \alpha \frac{\partial \alpha^*}{\partial u_{ij}} \right\}, \tag{2.78}$$

$$\frac{\partial |\alpha|^2}{\partial \Im\{u_{ij}\}} = \frac{\partial (\alpha\alpha^*)}{\partial \Im\{u_{ij}\}} = i \left(-\alpha \frac{\partial \alpha^*}{\partial u_{ij}} + \frac{\partial \alpha}{\partial u_{ij}} \alpha^* \right) = 2\,\Im\left\{ \alpha \frac{\partial \alpha^*}{\partial u_{ij}} \right\}. \tag{2.79}$$

We can thus obtain all the information we need from an expression for $\alpha_e(U)$ and all its partial derivatives $\frac{\partial \alpha_e(U)}{\partial u_{ij}}$, for each event e and for each Bell state. A significant fraction of them are identically zero: as a matter of fact, $\alpha_e(U)$ only involves the columns of U which have one or more photons in the corresponding output mode. Moreover, obtaining the partial derivative of polynomials automatically can be done very cheaply.

Equivalence under mode permutation

Looking closely at the expression for the amplitudes in Table 2.1, it can be noticed that they neatly divide in two groups, which are related by a permutation of the variables u_{ij}. For example, the amplitude for the event 1100 is:

$$\alpha_{1100}(U) = \frac{1}{\sqrt{2}}\big(u_{11}\,u_{32} + u_{12}\,u_{31} + u_{21}\,u_{42} + u_{22}\,u_{41}\big), \tag{2.80}$$

and the one for 1001:

$$\alpha_{1001}(U) = \frac{1}{\sqrt{2}}\big(u_{11}\,u_{34} + u_{14}\,u_{31} + u_{21}\,u_{44} + u_{24}\,u_{41}\big). \tag{2.81}$$

If we have a function which computes eq. (2.80) for a given U, we do not need to build another function for eq. (2.81). We can just evaluate the former on a different unitary, where we swapped the second column with the fourth:

$$\begin{pmatrix} u_{11} & u_{12} & u_{13} & u_{14} \\ u_{21} & u_{22} & u_{23} & u_{24} \\ u_{31} & u_{32} & u_{33} & u_{34} \\ u_{41} & u_{42} & u_{43} & u_{44} \end{pmatrix} \longrightarrow \begin{pmatrix} u_{11} & u_{14} & u_{13} & u_{12} \\ u_{21} & u_{24} & u_{23} & u_{22} \\ u_{31} & u_{34} & u_{33} & u_{32} \\ u_{41} & u_{44} & u_{43} & u_{42} \end{pmatrix}.$$

This is not a coincidence: we expect the amplitudes for two detection events to be related if one can be obtained from the other by a permutation of the output modes. This separates the events into equivalence classes, where only one representative of each class has to be symbolically computed. In the example of Table 2.1, we only need $\alpha_{2000}(U)$ and $\alpha_{1100}(U)$. It is possible to get all the others by listing all the unique permutations of each event

string—formally, they are called *multiset permutations*.

How much does this helps us? In the example above, we reduced the functions to compile for each Bell state from 10 to 2. In order to get an expression for the general case, we need to define the *partitions* of an integer:

Definition 4. A *partition* of $n \in \mathbb{N}$ is any set of positive integers which sum to n.

The number of equivalence classes is exactly the number of unique partitions of the total number of photons $(k+2)$. It does not depend on the number of modes m, and it grows much more slowly than the number of detection events N (albeit still exponentially). Specifically, its growth is approximately [HR18]:

$$\#\text{part}(k+2) \underset{k\to\infty}{\sim} \frac{1}{4k\sqrt{3}} \exp\left\{\pi\sqrt{\frac{2k}{3}}\right\}, \tag{2.82}$$

while the binomial coefficient is lower bounded by

$$\binom{m+k+1}{k+2} > \left(1+\frac{m-1}{k+2}\right)^{k+2}, \tag{2.83}$$

which, for the BA we study (m always at least $4+k$), is bounded from below by (2^{k+2})—meaning we gain an exponential factor over eq. (2.82) nonetheless. This optimization enabled us to access Bell analyzers with many more modes and bigger auxiliary states. For $m = 8$ and $k = 2$ (e.g. the smallest auxiliary state for Grice's schemes) the number of events per state is 330, while the number of partitions of $k+2$ is just 5, leading to a minimal set of events:

$$40000000 \quad 31000000 \quad 22000000 \quad 21100000 \quad 11110000. \tag{2.84}$$

The number of elements of the gradient to compile can be also reduced through output mode permutation symmetry, from n^2 to (at most) $n(k+2)$. Clearly, the huge time savings on the symbolic part are met with a slight increase of the evaluation time of the objective function in the optimization step, as it also has to permute at runtime the columns of U at each

evaluation. While for small BA this is not negligible, having to load fewer compiled functions is much more cache-friendly and leads to performance improvements in our tests for all but the smallest cases.

Bell state symmetry

A further factor of four can be gained by noticing that, for a given auxiliary state, we actually only need the symbolic expression of the amplitudes for one of the four inputs. This is due to a symmetry in the Bell states, and does not generalize when `solon` is used for the discrimination of a different set of states. Looking at the Bell states in path encoding, eq. (2.29), we can see that given:

$$|\phi^+\rangle = \frac{1}{\sqrt{2}} \left(a_1^\dagger a_3^\dagger + a_2^\dagger a_4^\dagger\right) |0\rangle \,, \tag{2.85}$$

the others can be obtained by appropriate substitutions:

$$|\phi^-\rangle : a_2^\dagger \longrightarrow -a_2^\dagger \quad |\psi^+\rangle : a_1^\dagger \longrightarrow a_2^\dagger \quad |\psi^-\rangle : a_1^\dagger \longrightarrow -a_2^\dagger. \tag{2.86}$$

Similarly to how the output modes are connected to the columns of U, the input modes are related to the rows. If we only have a function for the amplitudes from $|\phi^+\rangle |\Gamma\rangle$, we can use it for the other three by swapping and/or changing the sign of the first two rows of U, e.g. for $|\psi^-\rangle |\Gamma\rangle$:

$$\begin{pmatrix} u_{11} & u_{12} & \dots & u_{1m} \\ u_{21} & u_{22} & \dots & u_{2m} \\ \vdots & \vdots & & \vdots \end{pmatrix} \longrightarrow \begin{pmatrix} -u_{21} & -u_{22} & \dots & -u_{2m} \\ u_{11} & u_{12} & \dots & u_{1m} \\ \vdots & \vdots & & \vdots \end{pmatrix}.$$

A similar reasoning applies to the gradient, meaning we can further skip 3/4 of the symbolic work, with analogous (minor) drawbacks as the ones explained in the previous paragraph.

2.5.5 Numerical optimization

The main task of `solon` is to search the space of unitaries, finding good candidates for Bell analyzers. A possible strategy could be a simple bruteforce search in a (discretized) space which covers the unitary group $\mathrm{U}(m)$, taking

advantage of the fact that unitary matrices have bounded entries. However we would face several problems:

- While the dimension of the space of $m \times m$ unitaries only grows quadratically, the complexity of bruteforcing is exponential in the number of variables, i.e. $O(e^{m^2})$. Raising m would quickly lead to infeasible running times. Moreover, we have no way of knowing beforehand which discretization size is reasonable for the problem.

- It is difficult to obtain a good minimal parametrization of unitary matrices which also preserves the structure of the discretization step well enough.

- Most matrices, if randomly sampled from a uniform distribution, will result in zero discrimination probability, due to the unambiguity requirement. In other words, p_{succ}^{unamb} of eq. (1.21), which has to satisfy eq. (1.20), is not well suited for optimization. We have to choose something else as our figure of merit.

These issues are pretty standard in the nonlinear optimization domain, and there are multiple ways to get around them, each with its advantages and shortcomings. In general, scalable non-convex numerical methods can only reach local minima [PS88]. A common strategy (which we follow) is to repeat the optimization process with thousands adequately sampled starting points, only keeping the best optimum found. This kind of global optimization process is sometimes called *multistart optimization* in the literature.

Parametrizing $\mathbf{U}(m)$

The gradient descent needs a way to navigate the space of unitary matrices. In general, we want to choose a set of real-valued parameters (as they need to be ordered), from which U can be uniquely reconstructed. The space $\mathrm{U}(m)$[21] can be thought as a subspace of $\mathbb{C}^m \cong \mathbb{R}^{2m}$, of (real) dimension

$$\dim_{\mathbb{R}} \mathrm{U}(m) = m^2. \tag{2.87}$$

[21] With some abuse of notation, $\mathrm{U}(m)$ will implicitly refer to the representation of the unitary group as unitary $m \times m$ matrices.

We explore below three different ways to choose the real parameters.

- There exist direct parametrizations of the matrix entries of unitaries in terms of m^2 real parameters. An example are the *Givens rotations* [Cyb01]. They can be seen as a generalization of the common parametrization of $\mathrm{U}(2)$ in terms of angles $(\phi, \phi_1, \phi_2, \theta)$:

$$U = e^{i\frac{\phi}{2}} \begin{pmatrix} e^{i\phi_1}\cos\theta & e^{i\phi_2}\sin\theta \\ -e^{-i\phi_2}\sin\theta & e^{-i\phi_1}\cos\theta \end{pmatrix}. \tag{2.88}$$

  However, for $m > 2$ the analytical form of the entries of U gets more and more cumbersome to work with, if we want to use it in the symbolic part of the program to generate the output probabilities. Furthermore, to the best of our knowledge it is not easy to invert the transformation, i.e. in order to get the values of the parameters for a given U with known matrix entries we have to solve a system of coupled equations. This adds unnecessary complexity: for example, it becomes difficult to test the code with the known strategies.

  The symbolic generation and the numerical optimization are distinct parts of the program: strictly speaking, we do not need to invert the parametrization, as we could work with m^2 parameters only at optimization time, always converting them to an explicit U to be input to the compiled functions. Doing so, however, it becomes significantly more expensive to use our symbolically computed gradient, which can only be easily provided with respect to the explicit entries of the matrix, $(\Re\{u_{00}\}, \Im\{u_{00}\}, \dots)$ and not with respect to the new parameters. For these reasons, we pursued a different approach.

- The unitary group $\mathrm{U}(m)$ is a Lie group, generated from the associated Lie algebra $\mathfrak{u}(m)$ of skew-hermitian matrices through the exponential map. The elements of the algebra can be represented as iA, with A hermitian ($A^\dagger = A$).

  Hermitian matrices can be constructed by choosing $(m^2 - m)/2$ complex numbers for the upper triangular part and m real numbers for

the diagonal, totaling

$$2\,\frac{m^2-m}{2}+m=m^2 \tag{2.89}$$

real parameters. The lower triangular part is fixed by the hermitian condition. The matrix exponential can be computed efficiently, with negligible overhead; more importantly, it is easily invertible (by the matrix logarithm). However, it is still very difficult to make use of an analytical gradient. Nonetheless, this approach is easily implementable if we rely on numerical estimation of the gradient. It is one of the two strategies we used, along with the next one.

- Unitary matrices can always be seen as the space of complex invertible $m \times m$ matrices satisfying the unitarity condition:

$$U^\dagger = U^{-1} \quad \Longrightarrow \quad UU^\dagger = I. \tag{2.90}$$

Because the space of $m \times m$ complex matrices is spanned by $2m^2$ real parameters, using these (non-independent) variables instead of m^2 independent ones means *doubling* the dimension of the space that the optimization algorithm has to explore. The main advantage however is that we can use our analytical gradient, because the optimization variables now coincide with the matrix entries. For this, we need a gradient descent algorithm which supports (quadratic) constraints. We chose *Sequential Least-SQuare Programming*, (SLSQP [Kra88]), which can be directly called from SciPy's optimization collection [Vir+20]. The entry-by-entry condition on U in eq. (2.90) produces a system of $2m^2$ equations of $2m^2$ real variables; given that the number of independent variables is m^2, this system is overdetermined, which can cause problems for SLSQP. We can instead get a set of independent constraints for unitarity by assembling U from a set $\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_m\}$ of m orthonormal vectors of length m:

$$|\mathbf{u}_i|^2 = 1 \qquad m \text{ real equations}, \tag{2.91}$$

$$\mathbf{u}_i \cdot \mathbf{u}_j = 0 \qquad \binom{m}{2} \text{ complex equations}, \tag{2.92}$$

giving a total of

$$m + 2\,\frac{m!}{2!(m-2)!} = m + m(m-1) = m^2 \tag{2.93}$$

independent real equations. We get an added benefit, as the gradients of eq. (2.91) and eq. (2.92) are simple bilinear functions, which can themselves be analytically derived without the need of numerical estimation.

Despite needing twice as many parameters, this approach proved comparable in speed and convergence accuracy to the previous one, and was our preferred choice.

Sampling from $\mathbf{U}(m)$

In order for multistart optimization to be effective in exploring the space, we have to make sure the starting points are fairly sampled. Fortunately, uniformly sampling from the Haar measure over the unitary group amounts to just two steps [Ozo09; Mez07]:

1. Generate an $n \times n$ complex matrix A, sampling the real and imaginary part of the entries from the standard normal distribution (mean 0 and variance 1);

2. Through the Gram–Schmidt method applied to the columns of A, produce a set of orthonormal vectors which will form the columns of the random unitary matrix U.

In most scientific programming languages, a quick way to get an orthogonal matrix out of a generic one is the *QR decomposition*:

$$A = QR \qquad Q \text{ unitary} \quad R \text{ upper triangular}, \tag{2.94}$$

but the decomposition is not unique and there is usually no control on which algorithm is applied under the hood. For example, using the QR decomposition routine *as is* in SciPy produces the wrong distribution. The resulting matrix Q can be however "canonicalized", by multiplying each column by

the phase of the corresponding diagonal entry of R. Scipy implements the correct sampling method in the helper function `stats.unitary_group`.

Figure of merit

Concretely, the unambiguity condition eq. (1.20) in the Bell measurement case means that at least one output event has to be discriminating, i.e. has to occur with nonzero probability for *one and only one* of the four inputs. The total probability of successful discrimination we want to maximize is

$$p_{\text{succ}}(U) = \frac{1}{4} \sum_{(e,\beta_j)\in S_{\text{disc}}} p(e|\beta_j), \tag{2.95}$$

where we sum over S_{disc} which keeps track of the discriminating events along with the correspondent discriminated state

$$S_{\text{disc}} = \left\{ (e,\beta_j) \;\middle|\; \forall \beta_i \neq \beta_j,\ p(e|\beta_j) = 0 \right\}. \tag{2.96}$$

It is tempting to just provide $p_{\text{succ}}(U)$ as objective function for the gradient descent. However, for this to work properly, the function should be at least continuous over the unitary domain, and differentiable almost everywhere. From its definition, we can see that not only p_{succ} is not smooth, it is not even continuous. Moreover, its value is identically zero on a large portion of its domain: actually, picking a random unitary leads with high probability to $p(e|\beta_i) > 0$ for all e, β_i (Fig. 2.12). The optimization would stop immediately, as the gradient would be zero in all directions.

Given that we cannot directly use the success probability, we have to devise a $f(U)$ that has (ideally) the same extrema of $p_{\text{succ}}(U)$, but is at least continuous with nonzero gradient almost everywhere. The choice of f is not just a technicality: it might affect the final result, and involves some empirical tinkering. While we want f and its gradient to be as simple as possible, we do not have to worry too much about their computational efficiency, because for big BA the bottleneck will always be the computation time of the $p(e|\beta_i)$. In the following the dependency of f on U will be understood, and we will write it instead as a function of the event probabilities.

```
~$ python 4_modes_random_unitary.py

(0, 0, 1, 1) [0.002043 0.046756 0.014373 0.078176]
(0, 1, 0, 1) [0.046617 0.047700 0.019362 0.269098]
(0, 1, 1, 0) [0.181250 0.018851 0.168217 0.124671]
(1, 0, 0, 1) [0.306116 0.060476 0.074817 0.056033]
(1, 0, 1, 0) [0.022079 0.086799 0.033373 0.228691]
(1, 1, 0, 0) [0.012422 0.036992 0.010052 0.072311]
(0, 0, 0, 2) [0.072612 0.172534 0.195724 0.048346]
(0, 0, 2, 0) [0.147315 0.173797 0.142019 0.034231]
(0, 2, 0, 0) [0.129856 0.198228 0.151184 0.016960]
(2, 0, 0, 0) [0.079692 0.157867 0.190879 0.071482]
Discrimination probability (zero=1e-08):
             [0._______ 0._______ 0._______ 0._______], p = 0
```

Figure 2.12: Event probabilities for a random unitary. None of the events is discriminating.

- The first piece we need is something which is "zero if $p(e|\beta_i)$ is nonzero for at least two values of β_i", which is how $p_{\text{succ}}(U)$ behaves. In order to enlighten the notation, we switch to $P_e^i := p(e|\beta_i)$ and $\mathbf{p}_e = (P_e^1, P_e^2, P_e^3, P_e^4)$. We can start with this sum:

$$P_e^1 P_e^2 + P_e^1 P_e^3 + P_e^1 P_e^4 + P_e^2 P_e^3 + P_e^2 P_e^4 + P_e^3 P_e^4 = \sum_{\alpha \neq \beta}^{4} P_e^\alpha P_e^\beta, \quad (2.97)$$

  noticing that it has a positive value as long as only two of the P are close to zero, but it drops fast to zero when a third one approaches zero. Minimizing eq. (2.97) for an event e is not exactly what we want, because it also selects unitaries in which all the probabilities for that event are zero (in which case the event does not occur and it is not useful). We modify it by including a direct contribution for the sum of the probabilities:

$$f_e(\mathbf{p}_e) = \frac{P_e^1 + P_e^2 + P_e^3 + P_e^4}{\sqrt{\sum_{\alpha \neq \beta} P_e^\alpha P_e^\beta + \epsilon^2}}. \quad (2.98)$$

For one specific event e, maximizing eq. (2.98) ensures we get one positive probability (if possible). The square root regularizes the behavior around 0, such that the numerator and the denominator stay of the same magnitude; the added ϵ ensures good behavior when we have only one nonzero P and the sum in eq. (2.97) is zero. Then, if for example $P_e^1 > 0$ and $P_e^2 = P_e^3 = P_e^4 = 0$, $f_e(\mathbf{p}_e) \sim P_e^1/\epsilon$. We can make the latter large by choosing a small ϵ. However ϵ cannot be made arbitrarily small, as it serves another purpose: it prevents the occurrence of numerical overflows (that is, "infinite" values) during the optimization. We would also like ϵ to have a negligible effect on the value of f_e in the rest of the space (far-from-discriminating events), so we chose it in such a way that ϵ^2 is small compared to the typical size of the sum under the square root.

We do not know in advance which events will become discriminating at the (local) optimum. To get a figure of merit which is symmetric under the choice of discriminating events, we have to sum over all of them:

$$f(\mathbf{P}) = \sum_e f_e(\mathbf{p}_e), \tag{2.99}$$

where $\mathbf{P}$ is the (formal) matrix of all probabilities obtained by vertically stacking the $\mathbf{p}_e$, i.e. $P_{ij} = p(i|\beta_j)$.

This figure of merit proved useful for small BA, and was the first we successfully employed. However when the number N of events rises, the total probability dilutes among the detection events and we need to scale ϵ accordingly. Due to the properties of the square root derivative, which explodes for small values, there is a limit on how small ϵ can be without causing bad convergence of the gradient descent.

- In order to solve the latter issue, for bigger BA we substitute f_e in eq. (2.98) with its square, obtaining the objective function:

$$f(\mathbf{P}) = \sum_e \frac{(P_e^1 + P_e^2 + P_e^3 + P_e^4)^2}{\sum_{\alpha\neq\beta} P_e^\alpha P_e^\beta + \frac{\epsilon}{N}}, \tag{2.100}$$

which is also easier to compute along with its gradient. This solves the convergence issue, with most of the starting point now yielding a "good" optimal unitary at the end of the algorithm.

However when we moved to even higher dimension with non trivial auxiliary states, where we knew there were schemes breaking the $p_{\text{succ}} = 1/2$ limit, we noticed that convergence to those schemes was extremely rare (just a couple in thousands of optimizations). Now the square at the numerator is giving numerical issues; given the same total success probability, $f(\mathbf{P})$ is bigger for unitaries that concentrate that probability in just a few discriminating events rather than distributing it over many. This introduces an optimization bias which should be removed. With this figure of merit `solon` successfully finds the first iteration of both Grice's and Ewert–van Loock's schemes ($p = 0.75$, Section 2.3.6).

- If we drop the requirement of smooth gradient, we can access a much simpler objective function:

$$f(\mathbf{P}) = \sum_e \Big(2 \max_{\beta_j} \{P_e^j\} - \sum_i P_e^i \Big). \tag{2.101}$$

This expression is compelling because it is positive with value P_e^a when P_e^a is the only nonzero probability, making $f_e = (p_{\text{succ}})_e$ when $(p_{\text{succ}})_e \neq 0$. We have to use extra care however, as f_e contains the max function and its gradient is of this form:

$$\frac{\partial f_e}{\partial P_e^a} = \begin{cases} +1 & \text{if } P_e^a = \max_i \{P_e^i\} \\ -1 & \text{otherwise.} \end{cases} \tag{2.102}$$

Even if the expression above is not continuous, this choice of figure of merit improves greatly on the number of iterations needed to reach convergence, probably due to its (piecewise) linearity.

By choosing an optimization algorithm that operates over the full space of $m \times m$ matrices, we cannot rely on having bounded values for the probabilities P_e^i; this is an important issue to address for SLSQP, which is al-

lowed to wander off the unitarity constraints as long as they are satisfied at convergence. Unfortunately, naturally unbounded (from below) objective functions like eq. (2.101) suffer greatly from this kind of issue when maximized, preventing convergence. At first, the issue was addressed by the introduction of a *regulator*,

$$f'(\mathbf{P}) \coloneqq f(\mathbf{P}) \exp\left\{ -k \sum_{i,e} P_e^i \right\}, \tag{2.103}$$

maximizing $f'(\mathbf{P})$ instead. By tuning k accordingly, the regulator is close to 1 when the total probability is bounded by $\sum_i^4 1 = 4$, while it quickly kills $f(\mathbf{P})$ when it starts to be evaluated over non-unitary regions. The regulator has to be incorporated in the gradient too, increasing complexity and adding yet another parameter to be tuned heuristically. A cleaner approach is to add a series of *box constraints*, which SLSQP can strictly enforce, by exploiting how the unitary subspace is contained in the subset B of matrices with modulus-bounded entries,

$$\mathrm{U}(m) \subset B(m), \quad B(m) \coloneqq \left\{ A \in \mathbb{C}^{m \times m} \;\middle|\; |A_{ij}| \leq 1 \quad \forall i, j \right\}. \tag{2.104}$$

Detection events where all photons bunch in the same mode are always non-discriminating, which we can see by slightly generalizing eq. (2.41) in the $p_{\text{succ}} \leq \frac{1}{2}$ proof of Theorem 2. We could then safely remove such events from the sum in the objective function, because we know their contribution is always null. They are however a tiny fraction of all events for big BA, and accordingly we did not find any measurable difference in performances after this modification.

We only provided all figures of merit as function of the event probabilities P_e^i, which are available from the symbolic generation via a wrapper which directly exposes the array $\mathbf{P}$, such that $f(U)$ is actually $f(\mathbf{P}(U))$. However as far as the gradient is concerned, only the $\partial P_e^i \Big/ \partial u_{ij}$ are provided. We have to compute it through the chain rule for partial derivatives,

$$\frac{\partial f(U)}{\partial u_{ij}} = \sum_i \frac{\partial f(\mathbf{P})}{\partial P_e^i} \frac{\partial P_e^i}{\partial u_{ij}}. \tag{2.105}$$

Technically, this is suboptimal as we could save some computation by directly providing a compiled function for $f(U)$. However, this separation of roles is essential in that it lets us experiment with the objective function without having to recalculate the computationally intensive symbolic functions.

We nonetheless have to implement eq. (2.95)—the true efficiency of the BA—as reference for validating the optimization results. Some numerical subtleties have to be taken care of here when checking the discrimination condition eq. (2.96), namely on the definition of "zero" probability. As a matter of fact, convergence to an optimal unitary is only possible up to a finite accuracy.[22] Even if arbitrary precision numerical techniques exist,[23] they incur in a prohibitive overhead for our application. The choice we make for the bigger $\varepsilon > 0$ we still consider as zero probability can demonstrably affect our analysis, for example inducing us to be much too severe in dropping the results of each run if we were to set it to an unreasonably low value. To be sure to have some leeway, we always compute p_{succ} for three different choices of ε, $[10^{-4}, 10^{-6}, 10^{-8}]$.

Using $f(U)$ as makeshift $p_{\text{succ}}(U)$ we have no guarantee that the extrema of the latter coincide with the ones of the former:

$$E_p \coloneqq \{\, U \mid \nabla p_{\text{succ}}(U) = 0 \,\} \;\neq\; E_f \coloneqq \{\, U \mid \nabla f(U) = 0 \,\}. \tag{2.106}$$

However, from the definition of all the alternatives we discussed above (eqs. (2.98) to (2.101)) we can at least expect $E_p \subseteq E_f$. Unfortunately, we did find some explicit instances of "false positives", where the objective functions gave a better score for an optimum which actually had a lower efficiency, i.e. for a pair of optimal unitaries U_1 and U_2:

$$f(U_1) > f(U_2) \quad \text{while} \quad p_{\text{succ}}(U_1) < p_{\text{succ}}(U_2). \tag{2.107}$$

[22] Usually, we are limited by the *machine epsilon* of the `double` datatype we use ($\simeq 10^{-16}$), which is the maximal relative error due to fixed-precision floating point arithmetic. In practice however, the presence of quadratic expressions in most gradient descent routines limits the achievable accuracy on the optimization variables to the square root of it ($\simeq 10^{-8}$).

[23] For example, the *GNU multiple precision arithmetic library* offers arbitrary-precision integers and float types, and `Julia`'s package `Optim.jl` [MR18] provides interfaces to use them in optimization problems.

Consequently, we cannot trust the global optima of $f(U)$ to be of any use, and we will only make use of its local properties. This rules out global optimization methods such as *simulated annealing* [NW06]—which we implemented anyway with little benefit (as expected).

2.6 Optimization results

This Section is devoted to summarizing our numerical and analytical investigation of Bell analyzers. For various choices of auxiliary states, we work out the specific upper bound for polarization-preserving BA we derived in Section 2.4 and we compare it to the outcome of the numerical optimization over generic interferometers. The bulk of our analysis for different auxiliary states is showcased in Table 2.3. For each one of them, we collected the local optima from about ten thousand successful maximizations, from randomly sampled starting points. In the table the maximum value achieved for each input is shown; For the cases already known in the literature, we find the same maximal discrimination probability, sometimes achieved through different schemes. We regard this as (numerical) evidence of their optimality.

When it converges to a solution, `solon` reports the corresponding conditional probabilities of unambiguous detection for each Bell state (e.g. Fig. 2.11). We use the same format in the following to characterize a discrimination scheme, namely a tuple $(p_{\phi^+}, p_{\phi^-}, p_{\psi^+}, p_{\psi^-})$, when the overall discrimination probability $p_{\text{succ}} = \frac{1}{4}\sum_i p_i$ is not specific enough.

2.6.1 Vacuum and eigenstates of n_h

By virtue of Calsamiglia and Lütkenhaus' result (Theorem 2), the analytical upper bound $p_{\text{succ}} \leq 1/2$ is known to hold for unrestricted BA, equipped with an unlimited number of extra empty modes. We can work out an extended version of this bound (only valid in the polarization-preserving case) following the reasoning laid out in Section 2.4, looking at the distinguishability of the Bell states in eq. (2.30) after a projection onto the basis of the "number of horizontally-polarized photons" operator n_h. If the auxiliary

	Processor model	Core count	Freq. (GHz)	RAM (GB)
Laptop	Intel Core i7-4710MQ	4	2.5	16
Cluster	Intel Xeon E5-2670	12	2.3	256

Table 2.2: Specifications of the two computers we refer to in the text; the frequency shown is the nominal frequency of the processor. The cluster is the `gmpcs-206` branch of the computing center MésoLUM of the LUMAT research federation [Més], and the specifications only refer to a single node.

state is the vacuum or any other state with a fixed number $\bar{\lambda}$ of horizontally polarized photons, $\gamma_{\bar{\lambda}}$ is the only nonzero coefficient of the expansion in eq. (2.60). This leads to just two potentially discriminating outcomes of the projection:

$$|HH\rangle\,|\Gamma\rangle \quad \text{for } |\phi^{\pm}\rangle, \tag{2.108}$$

and

$$+\,|VV\rangle\,|\Gamma\rangle \quad \text{for } |\phi^{+}\rangle, \qquad -\,|VV\rangle\,|\Gamma\rangle \quad \text{for } |\phi^{-}\rangle. \tag{2.109}$$

The term in eq. (2.108) is identical for both inputs, while the two in eq. (2.109) differ by a global phase. Therefore they are not distinguishable at all: in this case the auxiliary state cannot help resolving the $|\phi^{\pm}\rangle$ degeneracy, so $p_{\text{succ}} \leq 1/2$.

As a sanity check of our numerical toolchain, the maximum discrimination probability that we find without extra photons is indeed $1/2$, for any number of modes we could reach. We quickly achieve this maximum on our laptop (see Table 2.2), and we collect a thousand successful runs in a matter of minutes for different values of m up to $m = 14$. With just two photons in the BA, even in the $m = 14$ case we spent about an hour of computation time on a laptop, and a few minutes on a cluster from the LUMAT research federation (specifications in Table 2.2). The large amount of RAM in the cluster is needed for some of our biggest optimizations, which is a drawback of using large compiled functions.

2.6.2 Extra Bell pairs

The choice of auxiliary resources in this Section are inspired by Grice's approach (Section 2.3.7), but motivated by possibly easier-to-implement schemes. While Grice shows that adding one extra $|\phi^+\rangle$ cuts the degeneracy of $|\phi^\pm\rangle$ by half, achieving $p_{\text{succ}} = 3/4$, the auxiliary states used to increase its success probability past this new ceiling become more complex at each iteration. From an experimental point of view, it would be interesting to assess the impact of adding multiple auxiliary Bell pairs, which are less of a challenge to produce than 2^n-GHZ states.

We start by working out our analytic bound, for a slightly more general auxiliary state, a product of $(k/2)$ copies of[24]

$$|\Gamma_1\rangle = \frac{1}{\sqrt{2}}(|2H\rangle + |2V\rangle), \tag{2.110}$$

where with $|2H\rangle$ (resp. $|2V\rangle$) we include any state of two horizontally (resp. vertically) polarized photons. Both Grice's and Ewert and van Loock's $3/4$ constructions use special cases of $|\Gamma_1\rangle$. We have:

$$\begin{aligned} |\Gamma_1\rangle^{\otimes(k/2)} &= 2^{-k/4}(|2H\rangle + |2V\rangle)^{\otimes(k/2)} \\ &= 2^{-k/4} \sum_{\lambda \text{ even}}^{k} \sqrt{\binom{k/2}{\lambda/2}} |\Upsilon_\lambda\rangle\,, \end{aligned} \tag{2.111}$$

where $|\Upsilon_\lambda\rangle$ is the (normalized) uniform superposition of the terms with λ horizontally polarized photons, of which there are $\binom{k/2}{\lambda/2}$—one for every possible subsystem ordering resulting from the (noncommutative) tensor product in eq. (2.111). We can now use Theorem 3, in the form of eq. (2.69). Our auxiliary state only has terms with even λ, therefore

$$p_{\text{succ}} \leq 1 - \frac{1}{2}\Big(\sum_{\substack{\lambda \text{ even} \\ |\gamma_\lambda|^2 \text{ loc max}}} |\gamma_\lambda|^2 - \sum_{\substack{\lambda \text{ even} \\ |\gamma_\lambda|^2 \text{ loc min}}} |\gamma_\lambda|^2 \Big), \tag{2.112}$$

which can be further specialized by instantiating $|\Gamma_1\rangle$, e.g. $|\Gamma_1\rangle = |\phi^+\rangle$. In

[24] The choice of using $(k/2)$ instead of k is so that $|\Gamma_1\rangle^{\otimes k/2}$ is a k-photon state, consistently with the rest of the notation used in the Chapter.

this case there are no local minima, and we get:

$$p_{\text{succ } |\phi^+\rangle^{\otimes k/2}} \leq 1 - 2^{-(\frac{k}{2}+1)} \binom{k/2}{\lfloor k/4 \rfloor}. \tag{2.113}$$

Fixing k multiple of 4 for convenience, we can apply a second-order version of Stirling's approximation [Rob55],

$$\sqrt{2\pi n} \left(\frac{n}{\mathrm{e}}\right)^n \leq n! \leq \sqrt{2\pi n} \left(\frac{n}{\mathrm{e}}\right)^n \mathrm{e}^{\frac{1}{12n}}, \tag{2.114}$$

obtaining the asymptotic behavior:

$$\begin{aligned} p_{\text{succ } |\phi^+\rangle^{\otimes k/2}} &\leq 1 - 2^{-(\frac{k}{2}+1)} \left(\frac{1}{\sqrt{\pi k}} 2^{\frac{k}{2}+1} \mathrm{e}^{-2/3k}\right) \\ &= 1 - \frac{1}{\sqrt{\pi k}} \mathrm{e}^{-2/3k}. \end{aligned} \tag{2.115}$$

When k is even but not a multiple of 4, the inequality (2.115) is invalid, but we still have

$$p_{\text{succ } |\phi^+\rangle^{\otimes k/2}} \leq 1 - \frac{1}{\sqrt{\pi k}} \left[1 + O\left(\frac{1}{k}\right)\right]. \tag{2.116}$$

The $1/\sqrt{k}$ scaling allowed by the above bound is looser than the $1/k$ scaling achieved by Grice schemes. Nevertheless, it does not rule out strategies approaching success probabilities arbitrarily close to 1 by using such auxiliary states, much simpler than the ones needed for Grice schemes.

Our numerical search converges in just about a minute to the $p_{\text{succ}} = 3/4$ schemes on our laptop, with a memory consumption of about 300 MB. Using two auxiliary Bell pairs ($k/2 = 2$) we could not find any improvement over $k/2 = 1$ when including non-polarization-preserving BA. This optimization uses significantly more resources: we spent about 4 hours with 20 parallel threads on the cluster, each using 3 GB of RAM, for the collection of a thousand successful runs. For three Bell pairs, our polarization-preserving bound in eq. (2.113) gives $p_{\text{succ}} \leq 13/16 = .8125$, this time allowing for a scheme beyond 3/4. However, a Bell analyzer this big is barely out of reach for our program, even on the cluster. For comparison with a similar-sized case, the second iteration of Grice's strategy takes about 48 hours *for each run* to converge to a local optimum. We could collect just 12 optimizations,

obtaining $p_{\text{succ}} = 9/16$; unfortunately this is well below the already known $7/8$ scheme using this resource.

2.6.3 Extra single photons

The possibility of improving the discrimination probability through the use of extra unentangled single photons is of great experimental interest, especially with the recent development of high-efficiency single photon sources with near ideal indistinguishability [SSW17]. This kind of auxiliary state would indeed be the first choice for a real-world implementation of a better-than-$\frac{1}{2}$ Bell analyzer. As a matter of fact, without some kind of initial manipulation of the auxiliary photons, we already know that polarization-preserving transformations are useless (as they fall under the case of Section 2.6.1).

Ewert and van Loock explore the use of two single photons per auxiliary mode pair [supp. mat of EL14, section D] as alternatives to their auxiliary states (we reviewed their approach in Section 2.3.8). Their trick is to apply a leading polarization-dependent transformation—which in path-encoding means mixing together the mode pair forming the qubit—obtaining the Hong–ou–Mandel state $|\Gamma_1\rangle_{\text{EvL}} = \frac{1}{\sqrt{2}}(|20\rangle + |02\rangle)$. Then, the γ_λ coefficients for the k-photon state $|\Gamma_1\rangle_{\text{EvL}}^{\otimes k/2}$ are analogous to the ones for $k/2$ Bell states $|\phi^+\rangle^{\otimes k/2}$. We can therefore apply the same reasoning laid out in Section 2.6.2, arriving at the same bound of eq. (2.113):

$$p_{\text{succ}\ |1\rangle^{\otimes k}} \leq 1 - 2^{-(\frac{k}{2}+1)} \binom{k/2}{\lfloor k/4 \rfloor}, \tag{2.117}$$

which is how we obtained the red curve in Fig. 2.9.

We can try to exploit in a different way our bound for $k/2$ photon pairs if we subscribe to a similar preprocessing, namely that each photon enters the network polarized at an angle $\theta = \pm\frac{\pi}{4}$. With this restriction in place, starting from $k \geq 4$ we get a tighter upper bound to p_{succ}, compared to e.g. the photon-number based bound in eq. (2.74). For example, with 4 photons the latter gives $p_{\text{succ}} \leq 5/6$, while eq. (2.117) gives $p_{\text{succ}} \leq 1 - (2^{-3}\binom{2}{1}) = 3/4$: the latter is actually saturated by Ewert and van Loock's

4-single-photon variant. It seems interesting to apply the bound to the state of 12 single photons $|1\rangle^{\otimes 12} \to |\Gamma_1\rangle^{\otimes 6}$, which they use to slightly break the $3/4$ barrier. In this case, a direct application of eq. (2.117) leads to $p_{\text{succ}} \leq 1 - (2^{-7}\binom{6}{3}) = 27/32$, which is indeed larger than the $25/32$ efficiency they give an explicit scheme for. We can do better if we enforce the BA to have the same symmetry they use after the first beamsplitter (Fig. 2.7). In this case, we apply the bound to half of the auxiliary state (and assume the same probability of discrimination for the other symmetrical arm). Under this further restriction, $p_{\text{succ}} \leq 1 - (2^{-4}\binom{3}{1}) = 13/16$, which is closer to (but still above) $25/32$.

On the numerical side, we indeed find the $(1, 1, \frac{1}{2}, \frac{1}{2})$ scheme when initialized with a 4 single photon auxiliary state. We do not manage to improve its probability of success; as before, we regard it as evidence of optimality even in the polarization-dependent case. This time we find another scheme achieving the same total success probability with a different discrimination pattern: $(1, \frac{3}{4}, \frac{3}{4}, \frac{1}{2})$. Are there schemes which improve on $1/2$ with just two single photons, instead of 4? As a matter of fact we find two of them, $(1, 1, \frac{1}{4}, \frac{1}{4})$ and $(1, \frac{3}{4}, \frac{1}{2}, \frac{1}{4})$, achieving $p_{\text{succ}} = 5/8 = 0.625$. In retrospect, the first of the two can be easily derived by "halving" the 4-photon Ewert–van Loock scheme, i.e. by just inputting the vacuum in one of the two symmetric arms; it was indeed independently obtained by van Loock in [Loo17]. We lay out an explicit interferometer for it in Fig. 2.13. It is especially relevant experimentally, since it is (as far as we know) the simplest scheme achieving a success rate above $1/2$. It can be noted that this halving technique can be applied to all of their constructions: for the 12-photon scheme, which uses $(2+4)+(2+4)$ single photons and achieves a probability of $25/32$, we can work out a similar intermediate scheme with $(2+4)+(2) = 8$ extra photons, achieving $p_{\text{succ}} = 49/64$. While this scheme would be interesting to probe numerically, the size of this BA ($m = 12, k = 8$) proved to be computationally unfeasible given our resources.

An odd number $k+1$ of single photons in the auxiliary state does not improve the discrimination probability over k, in all our numerical experiments. This is in line with the polarization-preserving bound (we already

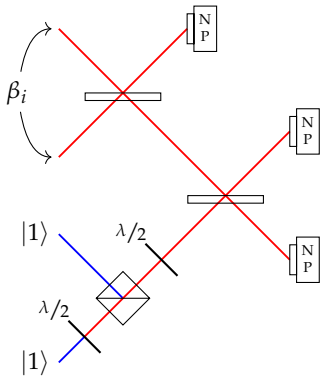


Figure 2.13: The first "half" Ewert and van Loock single-photons Bell analyzer, in polarization encoding. It performs a Bell measurement with $p_{\text{succ}} = 5/8$ using two unentangled extra photons, two beamsplitters, a polarizing beamsplitter, two phase shifters and three polarization-resolving detectors.

noticed it at the end of Section 2.4), and was independently observed by Smith and Kaplan [SK18]. Recently, the latters tackled with a substantially different approach a similar optimization for auxiliary states of single photons, but for ambiguous measurements: to this aim, they work in full Fock space and they maximize the classical mutual information between state preparation and measurement. Remarkably, despite this difference we find corresponding results for auxiliary states up to five photons. They find a slight improvement of their mutual information at the six photons mark (again, for ambiguous measurements [Smi17]). For our part, with six photons we could not find any scheme which goes beyond $p_{\text{succ}} = 3/4$ (Table 2.3). The polarization-preserving bound allows for a scheme with $p_{\text{succ}} \leq 13/16$, which does not exclude an improvement over $3/4$.

2.6.4 GHZ and W states

The last kind of auxiliary resource we investigated are some families of multipartite entangled states. From the start, a 3-GHZ state

$$|\text{GHZ}_3\rangle = \frac{1}{\sqrt{2}}\Big(|000\rangle + |111\rangle\Big) \tag{2.118}$$

does not seem to help with respect to a simple Bell pair, and we still only attain $3/4$ discrimination probability. And neither does a 4-GHZ, at the expense of more computational power; we wrongly expected the latter to give better results, given its use (along with a companion Bell pair) in the $p_{\text{succ}} = 7/8$ iteration of Grice's schemes (Section 2.3.7). The polarization-preserving bound paints an even grimmer picture: it predicts $p_{\text{succ}} \leq 1/2$ for all $|\text{GHZ}_k\rangle$ bigger than a Bell pair ($k > 2$). At least for odd k, pre-rotating the polarization of one of the photons by $\pm\frac{\pi}{4}$ (in the polarization-encoding equivalent scheme) correctly raises the bound to $3/4$. The same value can be achieved by a trivial interferometer applying a $\frac{\pi}{4}$ rotation on $k-2$ auxiliary modes, which leaves the remaining two photons in the $|\phi^{\pm}\rangle$ state.[25] The latter can then be used as described above to achieve $p_{\text{succ}} = 3/4$.

Another interesting state to investigate is the three-photon W state,

$$|\text{W}_3\rangle = \frac{1}{\sqrt{3}}\Big(|100\rangle + |010\rangle + |001\rangle\Big). \tag{2.119}$$

Like $|\text{GHZ}_3\rangle$, it is a genuine 3-party entangled state but, unlike all other states studied above, it is neither a graph state nor a stabilizer state. Its peculiar symmetry is likely the source of the optima we find (end of Table 2.3). Having the same number of horizontally polarized photons in each term, this state is as useless as the vacuum for polarization-preserving interferometers, as showed in Section 2.6.1. Rotating the polarization of two photons by $\frac{\pi}{4}$ the same way as before results in the more interesting bound $p_{\text{succ}} \leq 2/3$, and further manipulation (explained below) raises it to $p_{\text{succ}} \leq 3/4$. The best optimum we find numerically, using the minimum amount of modes ($m = 10$), is $p_{\text{succ}} = 5/9$, significantly lower than $3/4$. This optimum is extremely rare (it occurred only once in more than 20 000 runs). We observe the figures of merit in this case to heavily suffer from the issues described in eq. (2.107), about the relationship between $f(U)$ and $p_{\text{succ}}(U)$.

However, in this case we could find a better scheme by manipulating the state "by hand". By measuring the last two spatial modes we can apply

[25] The phase of the Bell pair is determined by the parity of the measurement of the $k-2$ photons, and its effect is to simply exchange the photon patterns for the detection of $|\phi^+\rangle$ and $|\phi^-\rangle$.

a transformation such that the remaining modes can be, depending on the result of the measurement, either in the state $\frac{1}{2}\left[(a^\dagger_{m-1})^2 - (a^\dagger_m)^2\right]|0\rangle$ or in $|\phi^+\rangle$. Applying to these modes the unitary used in the $p_{\text{succ}} = 3/4$ Grice strategy gives a scheme for $|\mathrm{W}_3\rangle$ with $p_{\text{succ}} = 7/12$. While `solon` correctly identifies this scheme as a local optimum when we put it in as the starting point of a run, an added Gaussian noise of average magnitude well below the requested convergence accuracy is sufficient for the gradient descent to diverge from its narrow valley. This numerical fragility may be the reason why we could not find this optimum from random starting points. Applying the analytical bound to such transformed auxiliary state gives us $p_{\text{succ}} \leq 3/4$. Interestingly, adding at least two vacuum modes ($m \geq 12$) allows `solon` to reach the improved discrimination probability of $0.5785508(2)$, which is slightly below our explicit $7/12 = 0.58\overline{3}$.

Table 2.3: Summary of known analytical and numerical results for different auxiliary states. As usual, m is the number of modes (in path encoding) and k the number of auxiliary photons. $p_{\text{succ}}^{\text{num}}$ is the optimum obtained through `solon`; when a fraction is given, it agrees with the optimum found up to 9 decimals. $p_{\text{succ}}^{\text{ana}}$ is the best known explicit analytical result. $p_{\text{succ}}^{\text{u.b.}}$ is our analytical upper bounds for polarization-preserving Bell analyzers (Section 2.4), and $p_{\text{succ}}^{\text{u.b.}}(k)$ is our bound for arbitrary auxiliary states with the same number of photons. When matching the best known result, the bounds are marked in **bold**.

State	m	k	$p_{\text{succ}}^{\text{num}}$	$p_{\text{succ}}^{\text{ana}}$	$p_{\text{succ}}^{\text{u.b.}}$	$p_{\text{succ}}^{\text{u.b.}}(k)$
Auxiliary vacuum modes						
$\|0\rangle$	4–14	0	1/2	$1/2$ [CL01]	**1/2** [CL01] [a]	**1/2**
***k*/2 extra Bell pairs**						
$\|\phi^+\rangle^{\otimes k/2}$	$2k+4$	even	—	[b]	$\simeq 1-\frac{1}{\sqrt{\pi k}}$ [c]	$\frac{k+1}{k+2}$
$\|\phi^+\rangle = \|\mathrm{Y}_1\rangle_{\mathrm{G}}$	8	2	3/4	$3/4$ [Gri11]	**3/4**	**3/4**
$\|\phi^+\rangle^{\otimes 2}$	12	4	3/4	[d]	**3/4**	5/6
$\|\phi^+\rangle^{\otimes 3}$	16	6	[e]	[d]	13/16	7/8
***k* extra photons**						
$\|1\rangle^{\otimes k}$	$k+4$	even	—	[b]	$\simeq 1-\frac{1}{\sqrt{\pi k}}$ [c]	$\frac{k+1}{k+2}$
$\|1\rangle^{\otimes k}$	$k+4$	odd	—	[b]	same as above, for $k-1$	
$\|1\rangle$	5	1	1/2 [d]	[d]	**1/2**	**1/2**
$\|1\rangle^{\otimes 2}$	6	2	5/8	5/8	3/4 [c] (**5/8**) [f]	3/4
$\|1\rangle^{\otimes 3}$	7	3	5/8 [d]	[d]	3/4 [c] (**5/8**) [f]	3/4
$\|1\rangle^{\otimes 4}$ ($\to \|\mathrm{Y}_1\rangle_{\mathrm{EvL}}^{\otimes 2}$)	8	4	3/4	$3/4$ [EL14]	**3/4** [c]	5/6
$\|1\rangle^{\otimes 6}$	10	6	3/4 [d]	[d]	13/16 [c]	7/8
$\|1\rangle^{\otimes 8}$	12	8	[e]	49/64	13/16 [c] (25/32) [f]	9/10
$\|1\rangle^{\otimes 12}$	16	12	[e]	$25/32$ [EL14]	27/32 [c] (13/16) [f]	13/14
Grice Schemes [Gri11] (including $\|\mathrm{Y}_1\rangle_G$ above)						
$\|\mathrm{Y}_1\rangle_{\mathrm{G}} \cdots \|\mathrm{Y}_{N-1}\rangle_{\mathrm{G}}$	2^{N+1}	2^N-2	—	$\frac{k+1}{k+2}$	$\frac{\boldsymbol{k+1}}{\boldsymbol{k+2}}$	$\frac{\boldsymbol{k+1}}{\boldsymbol{k+2}}$
$\|\mathrm{Y}_1\rangle_{\mathrm{G}} \|\mathrm{Y}_2\rangle_{\mathrm{G}}$	16	6	9/16 [g] [h]	7/8	**7/8**	**7/8**
Ewert–van Loock schemes [EL14] (including $\|\mathrm{Y}_1\rangle_{\mathrm{EvL}}^{\otimes 2}$ above)						
$(\|\mathrm{Y}_1\rangle_{\mathrm{EvL}} \cdots \|\mathrm{Y}_{N-1}\rangle_{\mathrm{EvL}})^{\otimes 2}$	2^{N+1}	$2(2^N-2)$	—	$\frac{k+2}{k+4}$	$\frac{\boldsymbol{k+2}}{\boldsymbol{k+4}}$	$\frac{k+1}{k+2}\left(\frac{\boldsymbol{k+2}}{\boldsymbol{k+4}}\right)$ [f]
GHZ states						
$\|\mathrm{GHZ}_k\rangle$	$2k+4$	k	—	$3/4$ [i]	**3/4** [c]	$1-\frac{1}{\lceil k+1\rceil_{\text{even}}}$
$\|\mathrm{GHZ}_3\rangle$	10	3	3/4	$3/4$ [i]	**3/4** [c]	3/4
$\|\mathrm{GHZ}_4\rangle = \|\mathrm{Y}_2\rangle_{\mathrm{G}}$	12	4	3/4	$3/4$ [i]	**3/4** [c]	5/6
W State						
$\|\mathrm{W}_3\rangle$	10–11 12–14	3	5/9 [h] 0.5785508(2) [h]	7/12	2/3 [c] (3/4) [j]	3/4

[a] Also holds for polarization non-preserving interferometers.
[b] No generic scheme is known.
[c] Polarization-preserving bound, obtained after rotating the polarization of some or all modes by $\frac{\pi}{4}$.
[d] The best known interferometer corresponds to a smaller auxiliary state, together with ignoring extra modes.
[e] Computation out of reach for our program.
[f] For networks which start by interfering the unknown state on a balanced beamsplitter, analyzing each half separately.
[g] Computation at the borderline of our computing capacity: best of a three weeks-long batch of 12 runs.
[h] Numerical result worse than the best known analytical scheme.
[i] Achieved by measuring all auxiliary photons and using the remaining state in a "one extra Bell pair" scheme.
[j] Obtained through a more complex transformation of the input, exposed in the main text.

2.7 Summary and conclusion

We conclude the Chapter with a review of our results on linear optical Bell measurement with auxiliary states, and a brief discussion of their future applications and extensions.

Our findings

After a review of the main theorems and measurement schemes present in the literature, we started by proving in Section 2.4 an analytical upper bound to their success probability, in the restricted case of polarization-preserving Bell analyzers. We have provided various forms of the bound: a tighter one, based on the distribution of the photons' polarization in the auxiliary state, and a looser one, based on just the number of photons. We noticed that the tight bound matches some published schemes, and the looser one shows similar performances if the auxiliary state is preprocessed by applying a very simple rotation in polarization space. While we could not find a proof this preprocessing is optimal, our numerical results support this conclusion for the cases we analyzed.

Since analytical results could only take us this far, in Section 2.5 we have presented the development of `solon`, a numerical package for the optimization of Bell analyzers over the space of all interferometers. `solon` works in the polynomial representation and is capable of symbolically evolving a generic input state through the interferometer, computing an analytical expression for the probabilities of each output detection event. We discussed how to reduce the overall computational cost, by exploiting symmetries of the problem. Then, we conducted a numerical search for the candidate optimal value of $p_{\text{succ}}(U)$ for various auxiliary states of interest (Section 2.6). Our final product is Table 2.3, which also summarizes the analytical bounds.

Through both the analytical study and the numerical optimization we find evidence (but no proofs) for the optimality of known small schemes in the general case. The schemes with two extra photons are the most promising for near-term implementations: we found no evidence of schemes

beating Grice's 75%, which uses an auxiliary Bell pair, but we discovered the simplest better-than-50% scheme known so far, using two unentangled auxiliary photons.[26] Just as interesting, we showed evidence that employing many copies of a Bell pair leads to a worse scaling of p_{succ} than the one achieved with Grice's $|\text{GHZ}_n\rangle$ states, which may give insights into why the complex entangled states used in both Grices's [Gri11] and Ewert and van Loock's [EL14] papers seem to be needed to approach $p_{\text{succ}} \to 1$ as quickly as they do.

We built `solon` to be fast at evaluating Bell analyzers, but the flexibility of Python meant that it could be readily adapted to other scenarios. As an example, we have used it during discussions with Chabaud *et al.* [Cha+18], helping them to gain insights on the effect of Hadamard networks, helping in the design of their linear optical *swap test*.

Computational resources

Some interesting cases lie beyond the computational capabilities at our disposal. There is certainly room for improvement, e.g. by further optimization of the code or by just employing more CPU time. It is unfortunately unlikely we could get rid of the inherent exponential scaling, which is tied to the hardness of calculating permanents (see Section 2.2.8). However the symmetry of the Bell states and the unambiguity constraints, which enforce a structure on the matrix entries—by imposing many null probabilities—may enable significant speedups (even exponential ones), even if the overall scaling stays exponential. Recent work in [Tic11, Appendix B] and [Shc13, Appendix D] suggest optimized algorithms for computing the permanent of matrices with repeated columns/rows; in the event of a rewriting of `solon`'s core, they may help improve our computation time. The very recent (2021) result in [GGM21] suggest that it might be possible to replace the numerical optimization gradient descent with an iterative process which converges more efficiently.

[26] This scheme was also discovered independently by van Loock [Loo17]

Similar projects

Given that optimization over interferometers has a wide range of applications, other projects have implemented similar functionalities since we wrote `solon`. As far as we know, none is specifically tailored to Bell measurement. We are aware of `linopt` [SD18], written in C++ with Python bindings, and the more recent `bolt` [MY21], which uses (among other optimizations) fast gradient descent techniques from machine learning. We did not compare the performances of these approaches, but we can note that leveraging the latter (`bolt`) could be useful in finding optimal schemes for Bell state *generation*, other than measurement.

Chapter 3

Quantum position verification

Paris-la-neuve, year 2250

FGR22 took off his glasses and left his desk, ready for yet another cup of coffee.

It had been a long day. It wasn't just that the police needed his services: after all, a cryptocommunication expert—be it of the quantum or regular type—was an obvious choice for this matter. In fairness, he *was* more used to dealing with machines than your everyday investigator. His machines, though, did not usually decide to go crazy and lock an entire floor of the federal bank, along with twenty-three hostages, on a calm Sunday night…

An android, gone rogue! As a robot himself, he found the concept difficult to grasp. He went through the evidence another time, pondering the events of the past hours. No one, not even the hostages, had actually seen the robot, which was likely hiding in one of the many locked offices. However, it had not been difficult to verify his identity: his name was ACH46, a recent model. With microsecond speed and cold precision, he had answered each and every question FGR22 had asked over the reserved frequency. He wanted money, a lot of it: he was prepared to kill the hostages with his bare hands if the police didn't comply. Clearly, some serious malfunctioning had occurred—usually, robots weren't even able to *lie*, let alone kill. In this situation, the police couldn't risk an assault to the building.

However, FGR22 felt that something was off about the whole story. How did ACH46 got there in the first place? The latest records tracked him in San Francisco, thousands of kilometers away, mere days before he showed up at the bank. No airport nor spaceport had registered him since. And even if the police had decided to pay, how on Earth was he thinking to escape capture? It's not as if he could just vanish... no, there had to be a bad assumption somewhere.

An idea was starting to form in his head. He was going to need a better clock.

***

He hopped on the reserved frequency. ACH46 was still there.
"Did you make up your mind? I'm serious about the hostages" he said.
"We are working on it" FGR22 replied. "Now, I need you to answer these questions. They are simple calculations, and I want you to answer as fast as you can." He plugged himself to the atomic clock he managed to find among his old stuff, and sent the first question. The answer was fast, but not *quite* as fast. He asked the second question, for good measure. He did not need to send a third.

***

FGR22 slammed the door of the precinct, a triumphant grin on his face. "He tricked us all! He cannot be in the building. *He is not there!*"

3.1 Introduction

It is difficult to understate how essential cryptographic protocols are in the modern digital world, where most communication is conducted remotely. One the main problems they solve is establishing trust and protecting against a lack thereof: in the first case, through public-key schemes it is possible to *authenticate* a user by certifying its possession of a special secret, without ever needing to reveal it. Then, private (or symmetric) key schemes ensure communication cannot be understood by unwanted listeners. Most of these features (with some exceptions) are enabled by mathematical assumptions which, albeit widely believed to be true, are still object of debate. This is especially true of public-key schemes, of which the

most popular (and widely used in applications today) are already in need of replacement with the advent of quantum computers [Ala+19], due to Shor's factorization algorithm [Sho94]. It still isn't clear if the replacements—which go by the umbrella-term *post-quantum cryptography*—will hold ground in the years to come [BL17].[1]

However, we can imagine different ways to gain trust in a third party. One way is through *position-based authentication*, which aims at using the physical location of a party as his only token of trust. Position verification, which refers to the possibility of securely convincing a third party about one's position, is the cryptographic primitive at the heart of it.[2] The applications of position-based schemes have some overlap with existing public-key schemes, but they are not a subset nor a complete replacement of them. For example, a connection with servers of a bank which can be certified to be physically located in the bank's building could be trusted in a similar fashion (or even more) than one authenticated through their private key, which might have been stolen—or one of the signing authorities compromised. Many of the services used everyday use self-reported location as an essential component: from ride-share apps to interactive augmented reality games.[3,4]

Finally, as the following review of the literature (Section 3.2) will hopefully show, asking whether position verification is even *possible* in the relativistic (quantum) setting opened a fruitful research direction, with deep connection to fundamental questions in both quantum and computational complexity theory.

[1] The confidence we can gain about the security of these kind of protocols is ultimately based on the time and effort spent trying to break them, as rigorous proofs are often hard to come by; it is the opinion of the author that only the wide availability of error-corrected quantum hardware would the raise the amount of trust we can put into them to today's level of confidence in the security of RSA against classical attacks.

[2] This Chapter only deals with position verification, from which a position-based authentication scheme can be constructed [Cha+09].

[3] Another example is the *pizza-delivery problem* [Sch11], which requires a weaker version of position verification called *distance bounding* and discussed in Section 3.2.1.

[4] In both cases, the possibility of securely verifying the self-reported position would be useful, as it would avoid certain kinds of attacks. A driver might want to report itself in multiple locations at once, in order to get more rides than the competition; or, in the case of augmented reality games, one of the most popular cheating methods to progress unfairly is to modify GPS location data on the fly (a weak form of *GPS spoofing* [Tip+11]).

3.1.1 General features

While the various position verification protocols discussed in the following each have their own peculiarity, we can sketch here some general features of both the protocols and the attack models which should be common to most of them. If needed, we will be explicit about the exceptions. In this Section we also define the notation and conventions used throughout the thesis, which serves the additional purpose of standardizing the overview of the literature we make in Section 3.2. A potential source of confusion, for example, is the naming choice for the actors: depending on the focus of each papers, the prototypical *Alice* and *Bob* have been used either for the protocol's verifiers, for the prover or for the colluding attackers. When dealing with 1D protocols, we choose the latter convention.

Protocol

First, the stage of almost all protocols (with the notable exception of [Unr14]) will be flat Minkowski spacetime of D spatial and 1 time dimension. A (quantum) PV protocol involves two main actors:

- A *prover*, which publicly claims to control a region P in space. At that location, which can potentially move around (i.e. $P(t)$), it is expected to be able to send and receive classical and/or quantum messages from and towards all directions, and perform classical and/or quantum computation. How the prover responds to inputs is part of the protocol, and as such is public knowledge. The prover is assumed *untrusted* by the verifiers, which also means it shares no secret with them which could be used to prove its identity.

- A group of k *verifiers*, which would like to acquire cryptographical evidence about the correctness of the prover's claim. Each verifier controls a station in a small neighborhood around location V_i, which we will always assume to be a point—or, equivalently, much smaller than the distance among them. From there, they send some *challenges* in the form of bitstrings or quantum states, timed in such a way to simultaneously arrive at P (in its frame of reference). Only the regions

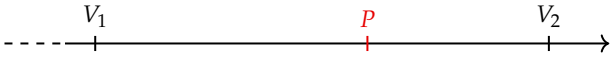


Figure 3.1: The setting for most of the PV protocols in this thesis is the 1D real line.

the verifiers control are assumed to be trusted. They share secure channels among themselves, which they can use to synchronize their clocks, share resources (often in the form of classical randomness) and check the prover's answers. However, the communications with the (untrusted) prover cross through uncontrolled territory, and their messages can be modified, jammed and blocked in any way by a potential attacker impersonating the prover. The verifiers are assumed to be static: in general, we will refer to them by their location V_i, with a slight abuse of notation.

Geometrical consideration show that $k \geq D + 1$ is needed to ensure the region P is small. Usually, k is set to the smallest possible value, $k = D + 1$. In most of this introduction (and in all of our results), $D = 1$: all parties are constrained on a line (Fig. 3.1). In this case, two verifiers V_1 and V_2 suffice. In order to better take advantage of the restrictions imposed by special relativity, all communication is expected to occur at the speed of light and all computation to be performed in negligible time, if compared to the signals' round trip time. For completeness, we note that deviations from this ideal case can still result in functional protocols, usually leading to a looser bound on the region P that can be authenticated.[5] The final verification process is a collaborative effort by the verifiers which is carried out over a private channel *after* interacting with the prover.

Due to the need to accommodate imperfections and catch lucky adversaries, a protocol might use either sequential or parallel repetitions of the same basic challenge-and-response unit, in order to amplify the overall probability of success. In this work we will sometimes commit abuse of "definition" by identifying the protocol with one of its rounds, in order to ease the comparison among them. Nonetheless, it should be kept in mind

[5] It is not straightforward in general to define the verifiable region for a given configuration of verifiers; the issue is tackled formally for example in [LL11; Unr14].

that the security of the single round does not directly translates into the security of its parallel (or serial) composition. Indeed, this feature has to be proven on a per-protocol basis, as it is done for example in [BK11].

Attack model

The space around P could be controlled by malicious actors (also called *adversaries*). The prover itself could lie about its position, or it could be that third parties want to fool the verifiers into believing they are also located at P. We can model their actions by placing them without loss of generality at a number of locations E_i, anywhere in the space outside of P and the V_i. In practice, we will assume each adversarial station E_i to be located somewhere along the geodesic connecting V_i and P. Ideally they are only constrained by the laws of physics: they are allowed to interfere with any communication, share any kind of correlation and compute any (computable) function. It will be clear in the following that allowing this much power turns out to prevent the possibility of secure PV altogether, albeit at a large resource cost for the adversaries. In order to design practical protocols, different kinds of restrictions have been imposed on the adversaries' resources. The most common limitation the literature (and us) has focused on is the amount of quantum correlations—in the form of entangled states—that attackers are allowed to share at the start of a verification round. Except when explicitly stated, classical communication and computational power are commonly considered free resources.

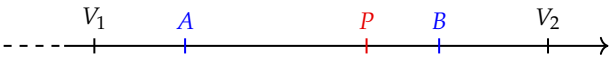


Figure 3.2: The space around V_1 and V_2 is not trusted: here, adversaries A and B set up stations of their own.

3.2 The PV protocol zoo

In this Section we present an overview of the recent results about (quantum) position verification, including the fascinating history of its origins, pieced

together from the early papers. It is not intended to be a full review:[6] its aim is to provide justification for our work, and the focus will be on the papers which had some influence on it—albeit sometimes indirectly.

Remark. While we occasionally mention error-tolerance, most results concerning the other experimental aspect of position verification are absent from this review. This include its loss-tolerance properties [QS15; CL15; Lim+16; All+21] and the discussion around its practical implementations [GLW13; DS21]. This is only justified by the narrow scope of our work, not by lack of interest in the community.

3.2.1 Proto-PV protocols: distance-bounding

In order to get some intuition about the role played by relativistic constraints on PV protocols, we briefly introduce *distance bounding*, a related but weaker task. As the name suggests, a distance bounding protocol should provide a certificate that the prover cannot be located farther than a certain distance r from the verifier. This kind of protocol was first introduced in [BC94], and it was designed to try to address a class of man-in-the-middle attacks.[7] Broadly speaking, distance bounding consists in prepending a signature-based identification scheme (e.g. Feige–Fiat–Shamir [FFS88]) with carefully timed call-and-response rounds of communication between the verifier and the alleged honest prover, which then signs the bits he sent and received. The physical principle on which its security is based is called *no-signalling*. Provided that the laws of nature do not permit faster-than-light communication, the *round trip time* (RTT) of the verifier's challenge can be converted to an upper bound on the physical distance between the prover and the verifier. While distance bounding can be performed by entirely classical means, its implementation in the real world is not trivial, requiring (for the prover) dedicated electronics to deal with the strict constraints ($\sim 1\,\mathrm{ns}$) on the processing delays.[8] For a recent review, see [Avo+18].

[6] Which, by the way, the field is definitely in need of!

[7] In a *man-in-the-middle* (MITM) attack, a malicious party inserts itself into a conversation between A and B, playing the role of B when talking to A and vice versa.

[8] This explains why, in the sci-fi story at the start of this Chapter, FRG22 needed access to an atomic clock to distance-bound ACH46—providing evidence that, barring a violation of special relativity or an error on his part, his claimed position was incorrect.

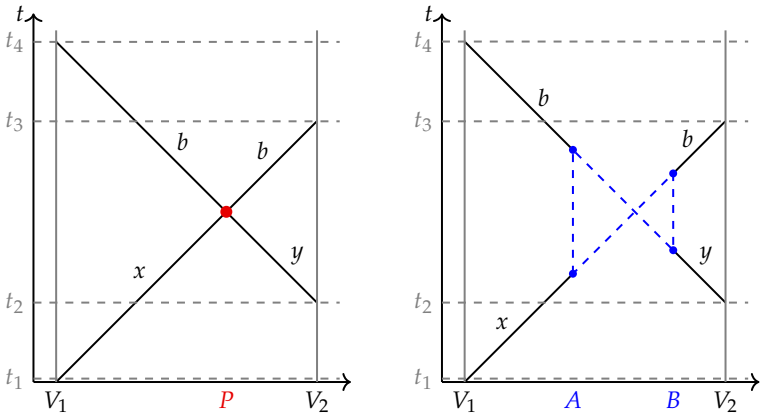
3.2.2 PV from distance bounding: a no-go theorem

We might be tempted to construct a PV protocol by simply composing two or more distance bounding protocols and then, similarly to how timing-based technologies like GPS work, performing a trilateration to obtain the secure region P. In 1D, this might look something like Fig. 3.3a. One could argue its security by noticing that any prover which is not entirely located at P is bound to fail to convince at least one of the two verifiers V_1, V_2. This intuition turns out to be wrong when applied to not one, but two *colluding* adversaries, which can act in concert—each one separately impersonating the prover. Now the verifiers have a new option: they could *also* collaborate, asking the prover to compute a function $f(x, y) = b$, sending the inputs x, y from opposite directions. Unfortunately, this also does not help: it is not difficult to see from the spacetime diagram in Fig. 3.3b that both attackers can receive both inputs in time to simulate the prover successfully. Can we find a secure protocol, or is it the case that adversaries can always break it? Indeed, the following general result can been proven:

Theorem 4 (Classical PV is impossible [Cha+09])**.** *No classical protocol can achieve secure PV, even under strict computational assumptions (unrestricted verifiers,* BPP *attackers/provers).*

This no-go theorem can be circumvented in a couple of ways. One is to find the correct restriction on the adversaries' ability to manipulate information, one powerful enough to enable PV. The latter is found by the same authors [Cha+09] in the *Bounded Retrieval Model* (BRM), which only allows the attackers to retrieve (i.e. read and process) a constant fraction of the information passing through their location. While this assumption could be expected to hold in specific situations, it seems difficult to enforce in most practical cases.

The structure of the known attacks however suggests a different approach. In order for both adversaries to correctly compute $f(x, y)$, they each need a copy of the two inputs. If at least one of them was encoded in a quantum state, the *no-cloning* theorem would prevent its duplication, rendering this specific attack impossible. This kind of observation sparked



(a) honest prover at P (b) colluding attackers at A, B

Figure 3.3: Spacetime diagrams of a 1D PV protocol which asks the prover to compute $f(x, y) = b$, and a general attack from colluding adversaries A and B. Lines at 45° represent lightspeed communication.

the search for secure *quantum* position verification (QPV) protocols. In the next Section, we summarize its history and early development.

Finally, a more subtle analysis of the assumptions reveals that Theorem 4 requires all parties to be fully classical, which includes both communication *and* computation. Allowing the prover quantum computation enables *classically-verifiable* PV protocols, which were recently constructed by Liu *et al.* in [LLQ21].

QPV: an origin story

The history of position verification in the quantum setting has been, to date, relatively short: it mainly involves work published in the last decade. Nonetheless, its first few years have been quite interesting—one might say exemplary—from the point of view of scientific research, as protocols were being proposed as secure just to be broken a short while after with the development of a new technique. We will present a short summary of

how the field evolved since; a useful resource to have on hand is Christian Schaffner's dedicated webpage [Sch11], where a timeline is maintained.

3.2.3 Malaney and Chandran *et al.*: a new hope

The very first works on QPV to appear in the scientific literature (as preprint, in early 2010) are Malaney's articles on "quantum location verification" [Mal10a; Mal10b]. He discusses a protocol based on sharing encrypted entangled states between the verifier and the prover, without giving a rigorous proof of its security. Nonetheless, he claims it secure on the ground of a similar no-cloning argument as the observation made in the previous Section: QM forbids the general attack at the core of the classical no-go result (Theorem 4). Independently and in parallel, a paper by Chandran *et al.*[9] [Cha+10] appeared as preprint giving an analogous quantum protocol[10] but, this time, providing a rigorous proof. Position verification was becoming, in the authors' words:

> ... one of the rare examples besides QKD for which there is a strong separation between classical and quantum cryptography.

But that feeling did not last for long. In an unexpected turn of events, Kent, Munro and Spiller [KMS11] responded in August of the same year with an attack to both protocols, despite their alleged security. How? It turns out that Kent had been thinking about quantum position verification since 2002, when he called it *quantum tagging*. In various discussions, the authors of [KMS11] had discovered the protocols above and their attacks; in 2006 they had filed and were granted a patent for a quantum tagging device [Ken+06]. In the 2010 paper, they observe that both of the previous results had made the hidden assumption that quantum information always needs to follow a definite spacetime path due to its unclonability, or in other words, that it has to be uniquely localized at all times. In fact, they continue, this might not to be the case: a qubit can be placed in a superposition of trajectories, its information content tied to a classical (clonable) variable—as

[9] This work shares most of its authors with the paper containing the classical impossibility result in Theorem 4 ([Cha+09]).

[10] The protocol is essentially Protocol 2, defined in Section 3.2.5.

it happens in the teleportation protocol [Ben+93]. The next Section gives an overview of their attack, which requires entanglement to be shared among the colluding adversaries. Was this the only hidden assumption in [Cha+10]? Indeed, it was. It turns out that assuming unentangled adversaries is enough to restore security in Chandran *et al.*'s result, which they show in a massive update to their paper [Buh+14], presented in Section 3.2.6.

3.2.4 Kent, Munro, Spiller: first insecurity proof

After having introduced the context in which the quantum tagging paper [KMS11] was published, we summarize one of the example protocols they formalize, along with its attack. From this paper (along with Lau and Lo's [LL11], see Section 3.2.5) stems the generalization we made and analyzed in this thesis (Section 3.3). The authors discuss a total of six protocols:

- The first three are "sensible" protocols which appear secure following the (flawed) reasoning of Malaney's and Chandran *et al.*'s. They can all be attacked by a kind of teleportation strategy. Of them, scheme III is inspired by six-states BB84 [BB84] and is the most interesting to us.

- They then present three protocols which cannot be attacked with their simple teleportation strategy. They leave the matter of their security open, but they add some features in the hope it may ease future security proofs.

In the following, we work in one spatial dimension and we will adapt the original paper's notation to the conventions made in Section 3.1.1. All remarks we noted there about signal timings and computational delay apply also here. A peculiarity of this protocol with respect to the generic model is that the verifiers do not need shared randomness, which comes at the cost of a higher rounds count needed to certify honest behavior.

Protocol 1 ($\text{QPV}_{\text{six-states}}$ [KMS11, scheme III])**.** The verifiers control stations V_1 and V_2, respectively to the left and to the right of P. During each round of the protocol:

1. From V_1, a qubit $|\psi\rangle$ randomly chosen from the set

$$S = \{|0\rangle, |1\rangle, |+\rangle, |-\rangle, |i\rangle, |i^*\rangle\}, \tag{3.1}$$

   where $|i\rangle = \frac{1}{\sqrt{2}}(|0\rangle + i|1\rangle)$, is sent towards P.

2. An independently chosen classical message $b \in \{0, 1, 2\}$ is sent from V_2, which instructs the prover to measure $|\psi\rangle$ in one of three bases:

$$B_0 = \{|0\rangle, |1\rangle\}, \quad B_1 = \{|+\rangle, |-\rangle\}, \quad B_2 = \{|i\rangle, |i^*\rangle\}. \tag{3.2}$$

3. V_1 and V_2 check they receive the resulting classical bit x on time.

After multiple rounds, on a secure channel the verifiers exchange the bits they received and check that the measurement statistics agree with the states and basis information they sent.

Colluding adversaries Alice and Bob have to control stations A and B, respectively between V_1 and P and between P and V_2, in order to spoof the timing check. The naïve security argument goes: as the states S are not all orthogonal, they cannot be distinguished immediately at A, nor a copy can be sent to B. Namely, the correct measurement outcome x can only be obtained at a location in which both b and $|\psi\rangle$ are present at the same time, so that either A has to wait for b or B has to wait for $|\psi\rangle$, necessarily failing to answer in time to either V_2 or V_1.

Teleportation attack

However, Alice and Bob are allowed to also locally manipulate some quantum resources, based on the information they receive. In particular, let them share a bipartite entangled state, specifically an EPR pair $|\phi^+\rangle_{\mathrm{AB}}$. Alice can then perform a teleportation measurement (a projection onto the Bell basis) on the joint system formed by the qubit she receives and her half of the entangled pair, which gives her one of four classical outcomes $u = 1 \ldots 4$ and teleports $|\psi\rangle$ on Bob's side with an applied teleportation correction—which depends on u, unknown to Bob—taken from the set $\mathcal{U} = \{I, X, Z, XZ\}$. Now, the usual teleportation should be completed by Alice sending u to

Bob, which will apply the corresponding correction in order to get $|\psi\rangle$; transmitting u from Alice to Bob however takes time, and we would just fall back to the naïve argument. Is Bob really powerless before receiving u? The key insight of KMS is that all the teleportation corrections in $\mathcal{U}$ map the bases B_i to themselves. This enables Bob to make use of the basis information b as soon as he receives it: he immediately measures his half of the EPR pair in the corresponding basis. The outcome $s \in \{0,1\}$ he obtains still contains no information about $|\psi\rangle$ without u (which would otherwise violate no-signaling), because the occurrence of X and XZ as corrections would have flipped the resulting bit. However at this point all the information Alice and Bob need to correctly output x (namely b, u and s) is classical, thus can be copied and shared among them. Now *both of them* satisfy the timing constraint, spoofing the protocol.

A basis issue The protocol $\mathrm{QPV}_{\text{six-states}}$ is flawed by the choice of bases, which just so happen to be invariant under the action of the standard teleportation corrections. KMS propose a fix by letting the verifiers choose $|\psi\rangle$ and the B_i from a set lacking this property. For example, an idealized protocol could sample the state and the basis at random over the entire Bloch sphere; we will call it $\mathrm{QPV}_{\text{Bloch}}$. As a discrete, more realistic version, they propose to take three bases separated by an angle of $\pi/6$, on the basis that there exist no unitary operation leaving all of them invariant (a rigorous proof of this fact was only provided later [LL11], see Section 3.2.5). They leave the security of these protocols against generic adversaries, which could share more complex resources than a maximally entangled pair of qubits, to future work.

3.2.5 Lau and Lo: going $D > 1$ and qutrit security proof

In September 2010 another preprint appeared in response to the flawed early protocols [Mal10a; Cha+10], by Lau and Lo [LL11]. Independently from [KMS11], they properly define generalized versions of the protocols and formalize known entangled attacks. One of the base protocols they use presents some differences with respect to Kent *et al.*; instead, they

more closely follow the one from [Cha+10]. First, in [LL11] (and in all the protocols we analyze from now on) the verifiers preshare classical randomness in order to synchronize their basis choice, which ensures that the expected message from the prover is always deterministic. Second, they use a simpler version of $\text{QPV}_{\text{six-states}}$ which only involves the four BB84 states, eigenstates of the X and Z Pauli operators. As it will come in handy in the future, we explicitly restate the 1D version here as reference.

Protocol 2 (QPV_{BB84} [Cha+10])**.** Under the usual remarks about timings, etc. (Section 3.1.1), each round of the protocol proceeds as follows:

1. V_1 and V_2 make use of shared classical randomness to agree on two bits x and b.

2. V_1 sends qubit $|\psi\rangle = H^b |x\rangle$ towards P, where H is the Hadamard gate. V_2 sends the basis information b.

3. At P, the prover is expected to measure $|\psi\rangle$ in the basis specified by b, i.e. apply $(H^\dagger)^b = H^b$ and measure in the Z basis, and broadcast the resulting bit x' to both verifiers.

4. Each verifier accepts if and only if $x' = x$ and the timing check is satisfied.

Lau and Lo's contribution is at least twofold:

- They address some (but not all, see [Unr14]) of the subtleties which arise for protocols in multiple spatial dimensions, where more than 2 stations come into play. They show the early protocols to be breakable for all D, finding the correct generalization of the $D = 1$ teleportation strategy in a known cryptographic technique called *quantum secret sharing* [HBB99] coupled with cluster state quantum computation [RBB03].

- They show, under a reasonable attack model, that a version of the $\text{QPV}_{\text{Bloch}}$ protocol proposed by KMS—which samples the encoding basis from the Bloch sphere—is indeed safe against exact attacks from adversaries sharing a single entangled qubit or qutrit pair. For

the qubit case, they numerically find the attackers' average success probability under their model to be about 85%.

Having shown that an entangled qutrit pair is of even less use than a qubit pair for this kind of attack, Lau and Lo conjecture their protocol to be secure. We will see in the following (Section 3.2.7) that not long after, a generic technique was discovered to attack it—which in this case produces an approximate attack. Part of our work in Section 3.3 was to study QPV(n), a special case of $\text{QPV}_{\text{Bloch}}$ where the encoding basis is sampled from a discrete set of n angles spanning a Bloch circle.

QPV is not (unconditionally) secure

3.2.6 Buhrman *et al.*: good news and bad news

Now aware of the hidden assumption undermining their claim, the authors of [Cha+10] came up with a new paper in September, teaming up with Schaffner and Buhrman from the university of Amsterdam [Buh+14]. In the updated paper, they properly restate their security result... while also showing a universal attack if entanglement is allowed, shattering the dream of finding a protocol with unconditional security. How much entanglement? A lot: their technique, *instantaneous nonlocal quantum computation* (INQC)—based on a result by Vaidman involving nonlocal measurements [Vai03]—requires the adversaries to share a number of EPR pairs which is doubly exponential in the number of qubits used in the honest case. Additionally, they design position-based authentication and key-exchange protocols building on position verification, which we do not review here, and generalize the attack to higher dimensions. We will give an overview of the first two contributions.

First, they show that QPV_{BB84} (Protocol 2) is secure in the *no preshared entanglement* (no-PE) attack model, where the adversaries are not allowed to share entangled states at the start of each round. The gist of their security result is the following theorem:

Theorem 5 ([Buh+14], informal)**.** *A dishonest prover can successfully spoof one round of* QPV_{BB84} *with a success probability of at most* ~ 0.89.

The proof of Theorem 5 is surprisingly involved and is based on an entropic result which had only been proven very recently, the *strong complementary information tradeoff* [RB09].

The second result is an attack strategy which shows that no reasonable PV protocol is secure against attackers sharing a large amount of entanglement. In order to show the universality of their claim, they first define a general PV scheme which captures a vast class of realistic protocols. We will report here a slightly simplified and specialized 1D version of it, to better highlight the connection to INQC.

Protocol 3 ($\mathrm{QPV}_{\mathrm{full}}$)**.** As usual, the generic remarks in Section 3.1.1 apply. As part of the public specification of the protocol, we choose a family of quantum channels $\{\mathcal{N}_{x,y}\}$ acting on an n-qubit register.

1. Before the start of the protocol, V_1 (resp. V_2) holds subsystem A (resp. B) of an n-qubit bipartite[11] state ρ_{AB}. They additionally share random classical information x, y, which select a transformation $\mathcal{N}_{x,y}$ out of the family.

2. V_1 (resp. V_2) sends their local resources $\{\rho_{\mathrm{A}}, x\}$ (resp. $\{\rho_{\mathrm{B}}, y\}$) towards P, synchronized to arrive at P at the same time.

3. As soon as the quantum and classical information arrive, the prover acts on ρ_{AB} with $\mathcal{N}_{x,y}$, obtaining $\rho_{\mathrm{AB}}^{x,y}$. He immediately sends back subsystem A to V_1 and subsystem B to V_2.

4. By communicating over their private channel, the verifiers accept the round iff the received state is close (in some measure) to the expected state $\mathcal{N}_{x,y}(\rho_{\mathrm{AB}})$.

It is not difficult to see that $\mathrm{QPV}_{\mathrm{full}}$ can be specialized to protocols with one-sided quantum information; for example, one round of $\mathrm{QPV}_{\mathrm{BB84}}$ is obtained by choosing $n = 1$, an empty subsystem B and an empty x with $y \coloneqq b$, the pure state $H^y |0\rangle_{\mathrm{A}}$ as the quantum state ρ_{AB} and the projection $\sum_i H^y |i\rangle\langle i| H^y$ for $\mathcal{N}_y$, after which register A is left in a classical state.

[11] The system may be additionally entangled with a local register E of arbitrary size, which remains with the verifiers during the execution of the protocol.

$\mathrm{QPV}_{\mathrm{full}}$ also includes their parallel repetition, multiplying n by the number of rounds and similarly accommodating the rest of the parameters.

3.2.7 Doubly-exponential INQC attack

Let Alice and Bob share an n-qubit bipartite state $|\psi\rangle_{\mathrm{AB}}$, with $n = n_A + n_B$, which we suppose pure for the sake of simplicity. The goal of instantaneous nonlocal quantum computation is to apply a unitary $U_{x,y}$ to the state, where $x = 1, \ldots, m$ is only known to Alice (resp. $y = 1, \ldots, m$ to Bob), such that at the end they hold $|\phi\rangle_{\mathrm{AB}} = U_{x,y} |\psi\rangle_{\mathrm{AB}}$ with high probability. We can immediately convince ourselves that:

- *Some* communication is needed, otherwise completing such a task would let Alice and Bob violate causality (e.g. U could swap a qubit from Alice to Bob).
- Two one-way communication rounds suffice: Alice can first send (or teleport) her subsystem A and x to Bob, which will locally apply $U_{x,y}$ on the whole state and send A back in the second round.

INQC deals with the intermediate option: one round of simultaneous two-way communication. When the communication is classical, this setting has been recently called *local operations and broadcast communication* (LOBC) in [GC20].[12] The price to pay with respect to two communication rounds is the need for Alice and Bob to share many EPR pairs,[13] which can be used to perform a kind of back-and-forth "teleportation-without-communication", where the measurement outcomes are not communicated right away but all at once at the end of the protocol. For brevity, we use quotes in the following to refer to only the local teleportation *measurement*, without communication of the outcome and correction of the teleported state.

In order to get an intuition for how it works, let us consider the case in which the unitary to compute is independent from x and y, i.e. $U_{x,y} = U$. If

[12] A closely related scenario in the communication complexity world is the *simultaneous message passing* (SMP) model [JK09], where a *referee* takes up the role of Alice and Bob after the communication phase.

[13] That entanglement has to be consumed is to be expected: the unitary U can be an entangling gate itself and entanglement, by definition, cannot increase under local operations and classical communication [Wil13].

we relax the "high probability" constraint, the usual teleportation protocol will sometimes—very rarely—succeed.

Protocol 4 ($\mathrm{INQC}_{\text{low prob}}$)**.**

1. Alice "teleports" her register A to Bob using n_A EPR pairs.

2. Bob applies U to the joint system composed by his half of the EPR pairs and his register B.

3. Bob "teleports"[14] back subsystem A via a second group of n_A EPR pairs. Alice and Bob now hold $|\phi'\rangle = \Sigma_b U \Sigma_a |\psi\rangle$, where Σ_a and Σ_b are tensor products of Pauli operators.

Notice that steps 1 to 3 happen locally and simultaneously, as each party is unaware of the outcomes obtained by the other.

4. They now exchange their teleportation measurement outcomes, which lets Alice correct at least Σ_b. While Bob now knows Σ_a, he still cannot correct it given that, in general, it does not commute with U. In the extremely lucky case in which $\Sigma_a = I \otimes I \otimes \ldots$, which happens with probability $1/4^{n_A}$, the protocol has succeeded; otherwise, they abort.

Notice that if Alice and Bob hold the state $\Sigma |\phi\rangle$ for any Σ before the round of mutual communication, they can correct it and obtain $|\phi\rangle$ after exchanging all measurement results. As shown below, the success probability of $\mathrm{INQC}_{\text{low prob}}$ can be improved following a trick due to Vaidman [Vai03], which deals with the fact that, in general, $\Sigma_a \neq I$. The intuition goes like this: we can setup many *teleportation channels*, namely 4^n groups of n EPR pairs each, labeling each group by one of the 4^n possible Pauli correction. Alice teleports the uncorrected state using the channel indexed by their local correction arising from the previous teleportation, while Bob blindly applies each possible corrections to each of of his local halves—discarding all but the correct one when they finally exchange all outcomes. This gives Alice another try at getting the lucky outcome $I \otimes I \otimes \ldots$, leading to:

[14] At this point, Bob could also directly send the state to Alice via a quantum channel, or use regular teleportation. This way however it is possible to clearly separate the communication round from the rest of the protocol.

Protocol 5 ($\mathrm{INQC}_{\text{double exp}}$)**.**

1, 2. Same as in Protocol 4.

3. Bob teleports the entire state instead of just Alice's subsystem. Now Alice holds $|\phi^1\rangle = \Sigma_{b_1} U \Sigma_{a_1} |\psi\rangle$.

Suppose Σ_{a_1} is not the identity, which is the overwhelmingly likely outcome. Their objective now is to "discard" the run, i.e. revert to something resembling the initial state, and have another chance at getting $\Sigma_a = I$. However, Alice does not know Σ_{b_1} at this point; she needs Bob to correct it.

4. Alice teleports $|\phi^1\rangle$ using the teleportation channel corresponding to her previous outcome a_1, getting a new outcome a_2.

Now the state $\Sigma_{a_2} |\phi^1\rangle = \Sigma_{a_2} \Sigma_{b_1} U \Sigma_{a_1} |\psi\rangle$ has appeared on Bob's channel a_1 (which is unknown to him). If they want to end up with $|\phi\rangle = U |\psi\rangle$, he has to undo the distortion caused by the previous round.

5. On every channel i, Bob does the following: he first applies $\Sigma_i U^\dagger \Sigma_{b_1}$, which brings channel a_1 back to the state

$$\begin{aligned} |\phi^2\rangle &= \Sigma_{a_1} U^\dagger \Sigma_{b_1} \Sigma_{a_2} \Sigma_{b_1} U \Sigma_{a_1} |\psi\rangle \\ &= |\psi\rangle \quad \text{if } \Sigma_{a_2} = I. \end{aligned} \tag{3.3}$$

   He then applies U and sends back the state via channel i of a *different* batch of 4^n channels, recording the outcome $b_2[i]$.

Alice now can find the state $\Sigma_{b_2[a_1]} U |\phi^2\rangle$ on channel a_1 of the last batch sent by Bob. With the same $1/4^n$ probability, her previous outcome Σ_{a_2} was the identity; in this case Alice holds $|\phi\rangle$ up to Bob's teleportation corrections and stops teleporting. Otherwise, she picks one new set of 4^n teleportation channels *for each channel involved the previous step*, indexed in such a way to keep track of both corrections. They restart the protocol from step 4.

N. Bob cannot know when Alice would have finally got a good outcome and stopped teleporting, so he continues up to an agreed number of iterations. At this point Alice teleports subsystem B to Bob and they exchange the teleportation corrections.

Notice that Alice has a constant (small) probability of success with each iteration. In order to succeed with non-negligible probability $(1-\varepsilon)$ overall, Bob has to continue for an exponential number of iterations. Additionally, each round requires a factor of 4^n more teleportation channels. Therefore, Protocol 5 consumes a number of EPR pairs which is roughly $2^{\log(1/\varepsilon)2^{4n}}$, doubly exponential in n.

Remark. $\mathrm{INQC}_{\text{double exp}}$, when used to attack $\mathrm{QPV}_{\text{full}}$, results in general in an approximate (and expensive) attack. This contrasts with the specialized attacks to $\mathrm{QPV}_{\text{six-states}}$ and $\mathrm{QPV}_{\text{BB84}}$, which are exact and much more efficient.

INQC and quantum foundations

The INQC task is a natural evolution of a much older and deeper problem in the foundation of physics, which concerns the subtle interplay between quantum constraints on measurements and relativistic effects—as witnessed, among others, by the famous Bohr–Einstein debate [Boh49] and the EPR paradox [EPR35]. As early as 1931 Landau and Peierls showed the measurement of the electromagnetic field at a specific location to be nonlocal and, therefore, deduced its impossibility [LP31]. However, in 1980 Aharonov and Albert started a line of research investigating how to harness entanglement to perform non-local measurements and operations without violating causality [AA80]. In 1981, they showed how to perform what we now call a Bell measurement between two distant particles, using an entangled pair of qubits [AA81]. These results were generalized to other observables [AA84a; AA84b; AAV86; PV94; GV01], until Vaidman finally showed in 2003 how to approximate any nonlocal measurement using teleportation and causal classical communications [Vai03]. Because of its universality, this turned out to be an effective way to attack QPV protocols and prompted new, quantitative investigations (which the next Sections will attempt to describe) into the amount of resources needed.

Giving a well-defined foundation to nonlocal measurements is not the only direction in which INQC has surprises in store. Since maximally entangled states violate Bell inequalities [Bel64], they give correlations which

are impossible to achieve in a classical theory. For example, while classical correlations only allow a 75% probability of success at the CHSH game[15] [Cla+69], Bell states raise this to $\cos(\pi/8)^2$, or about 85%. However, a theory can exhibit even stronger correlations than QM (called *supra-quantum correlations*) without violating causality. This was shown by Popescu and Rohrlich [PR94] through the definition of an imaginary nonlocal device which can win the CHSH game with unit probability. Such strong correlations would imply the collapse of classical communication complexity, as they enable two distant parties to compute any boolean function by exchanging only one bit [Van13], and are thus regarded as implausible. Recently, Broadbent showed that a similar collapse happens in the context of INQC [Bro16]: if the adversaries have access to Popescu-Rohrlich boxes, the entanglement requirement for a universal attack reduces to linear. For protocols which expect a single qubit (or bit) from the prover, the amount of communication needed is just two bits (or one).

3.2.8 Beigi and König: a "just exponential" attack

Can the enormous overhead of $\mathrm{INQC}_{\text{double exp}}$ be reduced? The source of one of the two exponential factors is the recursive nature of the protocol, which appears to be required to keep track of the Pauli corrections. In January 2011, Beigi and König [BK11] set out to overcome that issue, by employing a different kind of teleportation scheme as the main subroutine: *port-based teleportation*, another technique which was very recently discovered by Ishizaka and Hiroshima [IH09] (for a recent review, see [Chr+21]).

The idea behind port-based teleportation (PBT) is to make Bob's life much easier with respect to standard teleportation, at the cost of more entanglement and a more complex measurement on Alice's side. In a PBT protocol, Alice and Bob share many d-dimensional[16] entangled states called *ports*, labeled $i = 1, \ldots, N$. In order to teleport a state $|\psi\rangle$, Alice performs a particular measurement[17] involving $|\psi\rangle$ and *all* of her local ports; her

[15] In the CHSH game, two non-communicating players are sent respectively a bit x and y and should respond with bits a and b such that a XOR $b = x$ AND b.

[16] Here d refers to the dimension of the Hilbert space, i.e. $d = 2^n$ for n EPR pairs.

[17] One option is the *pretty good measurement* [HW94].

outcome is the index of a specific port k, where the state has been teleported. After receiving the index from Alice, Bob just discards all of his local ports but k. Depending on Alice's measurement, two scenarios are possible:

1. In *probabilistic* (or *heralded*) PBT, the state $|\psi\rangle$ is teleported perfectly to Bob's port k but the process is allowed to fail with probability p, i.e. Alice's measurement includes a failure outcome $\perp$.

2. In *deterministic* (or *approximate*) PBT, the state is always teleported but with non-unit fidelity.

With a finite number of ports N, perfect PBT is impossible. When optimizing over both the measurement and the resource state, the error at fixed d scales as $1/N$ for the heralded protocol and as $1/N^2$ for the approximate one [Chr+21].

For the purpose of INQC, port-based teleportation is very appealing as it lets Bob apply a unitary to a bare $|\psi\rangle$ before waiting for Alice's port index to arrive. The state is teleported in one go, without having to recursively build the exponential tree of Pauli corrections used by $\text{INQC}_{\text{double exp}}$. However, the number of ports needed to achieve a desired error still scales exponentially in n. Beigi and König can nonetheless prove the following improved result about position verification (in)security:

Theorem 6 ([BK11], adapted)**.** *Dishonest adversaries can break Protocol 3 with probability* $(1-\varepsilon)$ *by sharing* $n(1+\frac{2^{8n+5}}{\varepsilon^2})$ *ebits of entanglement.*

The second important contribution in [BK11] is the first QPV protocol with a proof for a lower (nonzero) bound on the amount of ebits required for an attack (Section 3.2.12).

Specialized attacks

The exponential scaling of the known universal attacks by INQC surely seems promising: while to date they have not been proven optimal for any explicit protocol, one can reasonably take that as evidence that PV could be shown to be secure in the quantum settings *for all practical purposes*. On deeper scrutiny, though, even if an exponential lower bound could be

shown one day for some class of unitaries,[18] there is no guarantee that the resulting protocol will be practical *in the honest case*. Indeed, most unitaries have an exponential circuit complexity [Kni95]; it does not seem reasonable to ask the verifiers and the prover to perform a protocol which involves the computation of, e.g. Haar-random unitaries on n qubits.

The analysis of variants of $\mathrm{QPV}_{\mathrm{full}}$ in which the protocol's unitaries are restricted to a particular class serves then two purposes: it ensures we only deal with practical protocols and it might lead to the discovery of more efficient attacks. For example, we already know that *some* protocols (e.g. if U is in the Clifford group [LL11]) can be attacked with a linear amount of entanglement. This is the line of research followed by many subsequent papers [Buh+13; CL15; Spe16; GC20; BCS21] and by us [Oli+20] (Section 3.3).

Remark. Due to scope constraints, we left out the treatment of a major family of QPV strategies, the *qubit routing protocols*, despite their importance: in fact a protocol of this kind was already constructed in Kent's seminal work [KMS11]. In these protocols, all the honest prover is asked to do is reroute an incoming quantum state to one of the verifiers, depending on a function of incoming classical inputs. We would like to stress that our omission does not reflect lack of interest in the community: among other achievements, designing attacks to these kind of protocols motivated the invention of an entirely new communication complexity measure for boolean functions—the *garden-hose complexity* [Buh+13]—which in turn sparked interest in areas unrelated to PV. Moreover, during the writing of this dissertation the first lower bound showing unbounded separation between a prover's quantum resources the adversaries' required entangled system size was shown [Jun+21], precisely for a routing protocol. We decided to discuss this last result briefly in Section 3.2.13.

3.2.9 Chakraborty–Leverrier: INQC in CH

In July 2015, Chakraborty and Leverrier [CL15] started considering the INQC implementation of unitaries in the *Clifford hierarchy*, inspired by

[18] This could happen sooner rather than later: see Section 3.2.13 for some exciting recent developments.

the structure of the original doubly-exponential protocol [Buh+14] in Section 3.2.7. The Clifford hierarchy [GC99] on n qubits (CH from now on) is an infinite sequence of unitary families $\mathcal{C}_1 \subset \mathcal{C}_2 \subset \mathcal{C}_3 \subset \dots$, defined by the recursive relation[19]:

$$\mathcal{C}_k = \{\, U \mid U\Sigma U^\dagger \in \mathcal{C}_{k-1} \quad \forall\, \Sigma \in \mathcal{P}_n \,\}\,, \tag{3.4}$$

where $\mathcal{P}_n$ is the Pauli group on n-qubits, generated by all possible tensor products of single-qubit Pauli operators $\sigma_i \in \{I, X, Y, Z\}$. The definition is completed by setting the Pauli group as the bottom rung of the hierarchy, i.e. $\mathcal{C}_1 \coloneqq \mathcal{P}_n$. Then, $\mathcal{C}_2$ is the usual Clifford group. None of the $\mathcal{C}_k$ with $k \geq 3$ form a group [ZCC08]. Gates from low levels in CH are usually easier to implement for the honest prover, which motivates the study of their attacks.

The conjugation of a Pauli with U in eq. (3.4) might ring a bell: in fact, it is reminiscent of the correction that Bob is required to apply during the recursive step of $\text{INQC}_{\text{double exp}}$ (Protocol 5). Recall that the objective of Alice and Bob is to obtain $|\phi\rangle = U|\psi\rangle$ up to Pauli corrections. In case U is at level k of CH, this leads to an interesting way of modifying the termination condition which does not rely on luck. At step 3, Alice holds the state:

$$|\phi^1\rangle = \Sigma_{b_1} U \Sigma_{a_1} |\psi\rangle = \Sigma_{b_1} U \Sigma_{a_1} U^\dagger |\phi\rangle = \Sigma_{b_1} \tilde{U}_1 |\phi\rangle\,, \tag{3.5}$$

where by definition of the hierarchy, $\tilde{U}_1 \in \mathcal{C}_{k-1}$. Therefore on the first recursive step Bob holds in channel a_1 the state $\Sigma_{a_2}|\phi^1\rangle$, to which he can apply $\tilde{U}_1^\dagger$:

$$\tilde{U}_1^\dagger \Sigma_{a_2} |\phi^1\rangle = \tilde{U}_1^\dagger \Sigma_{a_2} \Sigma_{b_1} \tilde{U}_1 |\phi\rangle = \tilde{U}_2 |\phi\rangle \tag{3.6}$$

and we went down another level, i.e. $\tilde{U}_2 \in \mathcal{C}_{k-2}$. Notice that multiplying $\tilde{U}$ by any Σ always result in another unitary at the same level of the hierarchy. Then, at iteration $j-1$ the state Alice and Bob are teleporting back and forth is of the form $\tilde{U}_j |\phi\rangle$, with $\tilde{U}_j \in \mathcal{C}_{k-j}$. Therefore, they can stop at round $k-2$: at this point, the residual unitary in front of $\tilde{U}_{k-1}|\phi\rangle$ is just another Pauli correction determined by their measurement outcomes, i.e. $\tilde{U}_{k-1} = \Sigma \in \mathcal{C}_1$, and they can correct it perfectly after the classical communication round.

[19] We omit specifying the number of qubits n in the following.

The procedure above also reduces the resources required by universal INQC to exponential (in n and k). While restricted to unitaries in $\mathcal{C}_k$, the main advantage with respect to port-based teleportation is that the resulting attack is not approximate, but exact. By following this argument, Chakraborty and Leverrier prove an upper bound to the number of EPR pairs for an instance of $\mathrm{QPV}_{\mathrm{full}}$ in which one side sends a computational state encoded in $|\psi\rangle = U|x\rangle$, the other sends a description of $U \in \mathcal{C}_k$ and the prover is asked to invert U, measure and broadcast x. In this case, the adversaries can break the protocol perfectly by sharing $4n\,4^{n(k-2)}$ ebits. They show similar reductions for protocols selecting U from another practical family of unitaries, namely those which can be computed with a circuit of a fixed layout.[20]

Their second important contribution is to define the *interleaved product* protocol. In $\mathrm{QPV}_{\mathrm{IP}}$, the verifiers encode the n-qubit secret $|x\rangle$ via the tensor product of n copies of a one-qubit unitary, sending $|\psi\rangle = U^{\otimes n}|x\rangle$. This ensures the protocol is practical for the honest prover. In order to strengthen its security, the verifiers send the description of U in a distributed fashion: V_1 sends unitaries $u_1, \ldots, u_t$ and V_2 sends $v_1, \ldots, v_t$, such that

$$U = \prod_{i=1}^{t} u_i v_i. \tag{3.7}$$

They notice that all attack techniques known at the time applied to $\mathrm{QPV}_{\mathrm{IP}}$ scale polynomially in n, as expected by U's small size, but *exponentially* in the number of terms t in the product (which only represent an amount of classical information in the honest protocol). They thus conjecture it secure. Following a trend which might start to become evident at this point, the claim held for just a few months, before a paper by Speelman [Spe16] showed otherwise.

[20] This could be very common if the prover processes the input "ballistically" with e.g. integrated optics, or in situations in which the selection of U is made by ranging over the values of single and two-qubit gates in a fixed layout.

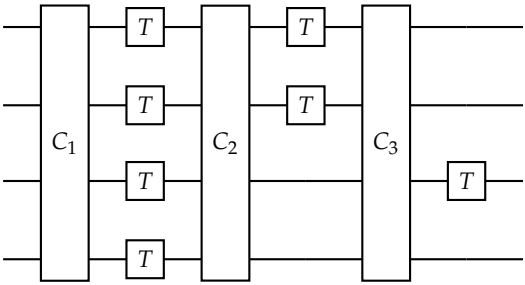


Figure 3.4: A circuit implementing a unitary U with T-count $t = 7$ and T-depth $d = 3$, figure adapted from [Spe16]. The compilation of unitaries to circuits of low T-count/T-depth (and finding lower bounds on these quantities) is an active field of research [Bev+20], as the performance of most fault-tolerant quantum computing architectures is mainly determined by non-Clifford operations.

3.2.10 Speelman: a polynomial attack

Are efficient INQC implementation possible for *any* interesting family of unitaries, besides Clifford? Speelman set out to answer this question for the positive, with a new attack on *low T-depth circuits* [Spe16]. In the circuit model, general unitary operations are "compiled" into a sequence of gates picked from some (usually application-dependent) limited set. Gates in the Clifford group $\mathcal{C}_2$ are usually easier to implement, but do not suffice for universal computation [Got98; AG04]; adding any gate from $\mathcal{C}_3$ however promotes the set to universal, meaning that any unitary can be efficiently (approximately) implemented by composing gates from it. A common choice is the *T gate*, a $\pi/4$ rotation in the Bloch sphere, resulting in the `Clifford+T` gate set. A general computation may then be viewed as alternating blocks of Clifford gates with layers of T gates on one or more qubits (Fig. 3.4). The total number of T gates is called the T-count, while the number of T layers is the T-depth. It should be noted that in some cases the T-depth can be much smaller than the T-count.

An important observation about INQC to keep in mind is its "uncomposability": that is, given a procedure for the nonlocal computation of U_1

and one for U_2, an INQC protocol for the product U_1U_2 is not guaranteed in general, as the two independently require their own round of classical communication. However, we know of at least one exception: the commutation properties of the Clifford group with the Pauli corrections imply composability for $U_1, U_2 \in \mathcal{C}_2$. If a circuit for U was only composed of Clifford operations, Alice could apply all of them in one go to the (still uncorrected) state, and later compute the correct Pauli after the communication phase.

Otherwise, they need a way to account for the T gates sandwiched between the Clifford layers, which do not freely commute with the Paulis. More specifically, $TX = PXT$ up to a global phase, where P is the phase gate; if Alice wants to apply the next Clifford layer before the communication round, they need to remove the extra P, which only appears if X is part of the Pauli corrections—unknown to Alice, known to Bob. In the paper, Speelman gives two ways of dealing with them:

- By building the usual tree of possible corrections, which doubles in size for each T gate applied. This leads to an efficient INQC implementation of unitaries computable by a circuit of low T-count k, which consumes $O(n2^k)$ ebits.
- By a clever application of Buhrman *et al.*'s garden-hose techniques [Buh+13]. This leads to an efficient INQC implementation of unitaries computable by a circuit of low T-depth d, which consumes $O((68n)^d)$ ebits.

It should be noted that the exponential dependence in t and k is crucial: if they grow at least linearly with n, these attacks are not efficient anymore. However, having a slowly (i.e. polylogaritmically) growing T-depth or T-count might be desirable also for the honest prover, if the protocol is to be considered practical.

Armed with these results, Speelman presents an attack to Chakraborty and Leverrier's QPV_{IP} protocol (Section 3.2.9) which uses an amount of EPR pairs polynomial in both n and the amount of classical information needed to specify the interleaved product unitaries. Other than further garden-hose magic, the proof relies on a particularly good (in terms of T-count vs. accuracy) approximate decomposition of any one-qubit unitary.

3.2.11 Gonzales-Chitambar: attacking a two-qubit protocol

Things went quiet for a while, and in 2018 a paper by Gonzales and Chitambar appeared in preprint with new attacks and lower bounds [GC20]. First, they formalize the most common attack model (the ground of INQC) as the task of implementing a unitary through *local operations and broadcast communication* (LOBC), in order to make precise statements about entanglement cost separations with respect to the more widely studied LOCC model (which allows interactive communication).[21]

For the first result involving two-qubit unitaries, they define the family **L**, sporting an *ad hoc* property which makes it easily implementable by INQC:

$$\mathbf{L} = \left\{ \begin{array}{l} U \mid \exists\, R, T_i, V_i \in \mathrm{U}(2) \\ \text{s.t.} \quad U(R\sigma_i R^\dagger \otimes I) = (T_i \otimes V_i)U \quad \forall \sigma_i \in \mathcal{P}. \end{array} \right\} \tag{3.8}$$

In other words, $U \in \mathbf{L}$ if the act of commuting it past a Pauli correction, possibly coming from a teleportation measurement in a rotated basis R, only results in unwanted *local* operations. They show an INQC protocol which implements U exactly, consuming only two ebits. If this property reminds the reader of the behavior of Clifford operators, they are on the right track: it turns out that $U \in \mathbf{L}$ if and only if U is *locally equivalent* to a Clifford operator, namely if:

$$(R_1 \otimes S_1)U(R_2 \otimes S_2) \in \mathcal{C}_2 \quad \text{for some } R_i, S_i \in \mathrm{U}(2). \tag{3.9}$$

The second result involves hermitian binary-controlled gates, bipartite unitaries of any dimension $d_A \otimes d_B$ where an hermitian unitary V is applied to the target system if the control system lies in the subspace defined by a projector Π. $\mathrm{QPV}_{\mathrm{BB84}}$ falls under this case, for $d_A = d_B = 2$ (promoting the basis bit b to a quantum state), $\Pi = |1\rangle\langle 1|$ and $V = H$. Gonzales and Chitambar show that one ebit is sufficient for an exact INQC implementation, for any d_A and d_B.

[21] Interestingly, Wakakuwa *et al.* [WSM19] show a separation in terms of entanglement cost between 2 and 3 rounds of interactive communication. While beyond the scope of PV, it shows that the study of these tradeoffs is of wider theoretical interest.

However, the most interesting result (to us) is the following theorem:

Theorem 7 ([GC20, theorem 1])**.** *There exist approximate* INQC *implementations of any two-qubit unitary, parametrized by N, which consume $8N+1$ ebits and succeed with probability $(1-\frac{1}{2^N})^3$.*

The INQC protocol which enables a proof of Theorem 7 implies the existence of an attack to $\mathrm{QPV}_{\mathrm{full}}$ for fixed $n=2$ and empty x, y, where the attacker's failure probability (equivalently, the approximation error ε) drops *exponentially* with the number of EPR pairs N, i.e. $N = O(\log \frac{1}{\varepsilon})$. For comparison, Beigi and König port-based INQC gives $O(\frac{1}{\varepsilon^2})$, i.e. the error only drops with the square root of N. The protocol QPV_θ we consider in Section 3.3 is also subject to this attack.

At the heart of their results lies the decomposition of U in a local and nonlocal part M, namely $U \simeq M$ up to pre- and post-processing with local unitaries like in eq. (3.9). It turns out [KC01] that M can be particularly simple for two-qubit unitaries: it is always diagonal in the so-called *magic basis*, a set of four maximally entangled states closely related to the Bell basis. In this basis M can be completely described by three angles, which correspond to two local rotations around z and an Ising ZZ two-qubit rotation:

$$I_{zz}(\beta) = \begin{pmatrix} R_z(-\beta) & & & \\ & & & \\ & & & R_z(\beta) \end{pmatrix} \tag{3.10}$$

namely $M(\alpha,\beta,\gamma) \simeq I_{zz}(\beta)(R_z(\alpha)\otimes R_z(\gamma))$ (up to a change of basis which only involves Clifford operations). Therefore, they just have to show how to implement $M(\alpha,\beta,\gamma)$ via INQC, which they do by exploiting the commutation relations between the Pauli corrections and the rotations in M.

For some angles (multiples of $\pi/2^{N-1}$), their protocols implements U exactly (i.e. with unit probability) using a finite number of ebits. We compare its efficiency in this case with our exact attacks in Section 3.4.4 for a much more restricted class of unitaries.

Lower bounds

In parallel with lowering the requirements for attacks, an arguably more important (and more difficult) task was being researched: proving *lower bounds* on the resources needed for breaking QPV protocols. A lower bound is needed to formalize the claim that a protocol which uses n qubits is secure against adversaries sharing at most a certain amount of entanglement $f(n)$. The security of $\mathrm{QPV}_{\mathrm{BB84}}$ in the no-PE model (Section 3.2.6) can be considered the first kind of such lower bounds for a protocol, in this case with $f(n) = 0$. While bounds against exact attacks are interesting, usually security is argued in a more realistic scenario where the prover is allowed to fail with probability ε: the protocol is secure if and only if the verifiers can reliably distinguish between an imperfect honest prover and (perfect) dishonest adversaries.

Unfortunately, strong lower bounds have been hard to come by: with the possible exception of [Unr14] and more recently [Jun+21], an exponential gap in resources still stands between the best generic attack and the best lower bound. In our work we don't prove any new lower bound, instead arguing for the optimality of our efficient attacks by means of numerical evidence. In this light, we will only quickly summarize this ongoing line of work in the next sections.

3.2.12 Linear lower bounds

This Sections contains entanglement consumption lower bounds which are linear in the number of qubits n used in the honest protocol.

Beigi and König

To our knowledge, the first result to improve on the no-PE bound is in [BK11]. The protocol they consider is based on *mutually unbiased bases* (MUB). A set of bases $\{\{|x_a\rangle\}_x\}_a$ of $\mathbb{C}^d$ is called mutually unbiased if the squared overlap between any two basis elements picked from different bases is $1/d$.[22] They define a PV protocol—which we call $\mathrm{QPV}_{\mathrm{MUB}}$—with

[22] We already encountered these bases for $n = 2$: they are the eigenvectors of the Pauli operators X, Y, Z (the set S in eq. (3.1)). There are always $d + 1$ such bases if d is a prime power, in particular

one-sided quantum information where the prover is expected to recover x from the n-qubit state $U_a |x\rangle = |x_a\rangle$, where U_a is the basis change to the basis $\{|x_a\rangle\}_x$. QPV_{MUB} can be considered a d-dimensional generalization of $\text{QPV}_{\text{six-states}}$. Beigi and König prove QPV_{MUB} secure against adversaries sharing less than $n/2$ ebits, but the honest prover has to manipulate n qubits at a time. Via a general result relating the tolerable error in the no-PE case with the dimension of the adversaries' entangled system (used for example in [Tom+13]), they show that a more practical sequential composition of l rounds of QPV_{MUB} with fixed d is secure with a similar scaling in l.

Tomamichel *et al.*

The main result of [Tom+13] is a *monogamy-of-entanglement* game, with bounds on its winning probability. Roughly, monogamy of entanglement (MOE) is a property of entangled states which forbids two maximally-entangled systems A and B to be also maximally entangled to a third system C. In the MOE game, a referee receives a subsystem of a tripartite state prepared by two players and performs a measurement on it in a random basis. Then the two players, not allowed to collaborate from now on, receive the choice of basis from the referee and are both asked to correctly guess the output of her measurement. The authors apply the result to various cryptographic tasks, including PV. They show that a 1-round parallel repetition of n instances of QPV_{BB84} is secure against adversaries sharing an entangled state of at most αn qubits, with $\alpha = -\log_2(\cos^2(\pi/8)) \simeq 0.23$. Here quantum communication is also allowed among the adversaries,[23] and the protocol only requires that the honest prover manipulate single qubits. However, the price to pay is a smaller coefficient than in Beigi and König's bound ($\alpha = 0.5$).

for $d = 2^n$ in the n-qubit case [Ban+02].

[23] This is not a small detail, as the restriction to classical communication might be difficult to justify in a realistic setting if the adversaries are supposed to be able to share entanglement before the start of the protocol. Most of the other results we discuss require the classical communication restriction. Notice however that quantum communication can be simulated by sharing additional EPR pairs and communicating classical outcomes via teleportation, with a linear entanglement overhead. The interplay between classical and quantum communication among the adversaries has been recently investigated by Allerstorfer *et al.* [All+21].

Ribeiro and Grosshans

Can we keep the experimental advantages of the n-qubit QPV_{BB84} protocol while giving a lower bound matching the known best attack of n ebits? In [RG15] Ribeiro and Grosshans bound the *max relative entropy of entanglement* of the resource shared by the adversaries, finding $E_{\max} \geq n - O(\log_2(n))$. To this aim, they adapt a security proof of a cryptographic task—the *weak string erasure* (WSE) protocol—which was given in [DFW15] in the noisy storage model (NSM), to the noisy *entanglement* model (NEM). In NSM, the adversary quantum memory decoheres after a certain time, while in NEM the adversary is split in multiple spatially separated parties which communicate via classical channels and share a (possibly mixed) quantum state. They complete the proof by showing that an attack to QPV_{BB84} translates to an attack to WSE, thus requiring an essentially linear amount of entangled resources. At the cost of one ebit per qubit, Kent's teleportation attack ([KMS11], Section 3.2.4) saturates the bound, which is thus tight for QPV_{BB84}.

Gonzales and Chitambar

In addition to the one-ebit attack for hermitian binary controlled gates (Section 3.2.11), in [GC20] Gonzales and Chitambar prove a linear lower bound in the non-hermitian case, when the control system is a qubit ($d_A \times 2$). They show the optimal exact INQC strategy requires $\log_2(d_A)$ ebits, which is linear in n with $\alpha = 1$ for an n-qubit unitary ($d_A = 2^n$). It should be noted (as they do) that their result is more difficult to convert to a security claim for a QPV protocol, as approximate implementations are not taken into account. With respect to the bound for QPV_{BB84} in [RG15], here we lose the practicality of qubit-wise operations; an important improvement however is that the bound is given in terms of required ebits, which is more robust than bounding the max relative entropy or the dimension of the entangled state.[24]

[24] For example, they show a family of d-dimensional states with $E_{\max}(d) \to \infty$ when $d \to \infty$, while the entanglement entropy $E(d) \to 0$.

Unruh: Quantum Random Oracles

With the possible exception of [Tom+13], all results presented up to now use some sort of restriction of the attack model (mainly classical communication) to achieve a better bound. While removing these restriction remains the primary challenge, an equally interesting question is to identify under which (potentially strong) assumption other than linear entanglement it is possible to achieve computational security, matching the best known universal attack [BK11]. Unruh finds a positive answer [Unr14] in the (quantum) random oracle model. In classical cryptography, a *random oracle* is a black box which is accessible by all parties and returns the output of a function f chosen uniformly at random. While they give all parties access to shared randomness, hidden via the exponentially many inputs $x \in \{0,1\}^n$, ideal random oracles cannot be simulated efficiently. Nonetheless, they are used as a proof device in many protocols [KM15], in hope that they can be replaced with the weaker assumption of cryptographic hash functions. Unruh's protocol is similar to the n-qubit parallel $\mathrm{QPV}_{\mathrm{BB84}}$ but the basis information $b \in \{0,1\}^n$ is distributed via two classical inputs x_1, x_2 coming from V_1 and V_2, such that $b = f(x_1 \oplus x_2)$. Unruh's proof—which can be extended to protocols taking place in general curved spacetime—uses the *reprogrammability* of random oracles[25] to hide the choice of b until *after* x_1 and x_2 cross the region P to be authenticated. At this point, he shows that the adversaries have to win a version of the monogamy of entanglement game in [Tom+13] in order to break the protocol, which they can do with at most exponentially small probability in n. Notice that the security in this case is of computational nature, meaning adversaries which have access to exponential (in n) computational power are able to break the protocol without sharing any entanglement.

[25] Roughly, reprogramming a random oracle means changing the random function f on the fly in order to choose the output of a queried input at a desired time. The idea behind it is that for a true random function, reprogramming is undetectable by the oracle users except with vanishingly small probability (some care has to be taken in the quantum case, when the oracle can be queried in a superposition of all inputs). This property can seem particularly unphysical if the oracle is instantiated with a specific hash function; for example, the reprogramming might happen "nonlocally" for two users, i.e. on a spacelike surface.

3.2.13 Recent results

After a 3-year gap (not including our 2020 paper [Oli+20]), during the redaction of this thesis a series of new preprints [Jun+21; BCS21; All+21; LLQ21] appeared on the arXiv. All of them provide a substantial leap forward in the design of secure QPV protocol, both from the theoretical and practical perspective. We select the two of them which we deem more relevant to our review and summarize their results in the following.

An exponentially secure protocol

In March, Junge *et al.* [Jun+21] looked at the connections between quantum games and geometric functional analysis and their applications to PV. They study a protocol which is quite more involved than the ones above, using many of the features allowed in $\mathrm{QPV}_{\mathrm{full}}$ such as the distribution of a tripartite state additionally entangled with a register local to the verifiers and an acceptance condition which requires them to later regroup the received state and apply a joint measurement. Under certain regularity assumptions (which are satisfied by known attacks), they show that their protocol requires an exponential amount of entanglement to break for dishonest adversaries. Additionally, they provide conjectures in the theory of Banach spaces which, if proven true, would help lift the regularity assumption and provide security of their protocol for all practical purposes.

A secure qubit routing protocol

In April 2021, Bluhm, Christandl and Speelman [BCS21] turn their attention to a particularly simple kind of protocol of the qubit routing type (we briefly talked about them in the introduction to Section 3.2). In their protocol, the prover receives two n-bit strings x, y from V_1 and V_2 as well as a qubit from V_1 which is maximally entangled with a verifiers' local register. The prover has to simply reroute the qubit to V_1 or V_2 depending on the output of a boolean function $f(x, y)$. This kind of protocol is arguably one the most practical and simplest possible for the honest prover, as he only has to manipulate a single qubit irrespective of n. A universal attack consuming

2^n EPR pairs against routing protocols of this kind has been known since [KMS11]. The appeal of an entanglement lower bound in this case is that in the honest case the only resource which scales is classical, i.e. the size n of f's inputs. Bluhm *et al.* manage to prove precisely that: adversaries sharing less than $(n/2 - 3)$ entangled qubits[26] can be caught with at least 10^{-2} probability, showing a potentially unbounded gap between the *quantum* resources in the honest vs. adversarial case. Improving the chances of detecting the adversaries can be done via sequential repetition, meaning that unfortunately one loses the desirable one-round nature of other schemes. Choosing explicitly the function f incurs in similar issues as the ones discussed for the random oracle model (Section 3.2.12), as there is no guarantee that it can be efficiently computed. In a subsequent version of their paper they give examples of explicit efficient functions, at the cost of a logarithmic (instead of linear) entanglement lower bound.

[26] Notice that here, too, only the dimension of the quantum resource is bounded, not the ebit content.

3.3 Defining QPV_θ and its attack model

3.3.1 The protocol

The family of protocols we chose to analyze is not new: indeed, similar versions of it are already present in early work [KMS11; LL11]. QPV_θ is a 1D protocol using 1 qubit per round, which extends the more common QPV_{BB84} by allowing for an arbitrary angle θ between the two measurement bases in which the verifiers' bit is encoded. Generic remarks about timings and verification from Section 3.1.1 apply. We choose QPV_θ for multiple reasons:

- On the practical side, QPV_θ shares much of the experimental benefits with QPV_{BB84}, especially when implemented through linear optics. It can leverage the current implementations of free-space QKD [Pug+17; Lia+17; Ave+21], with added timing constraints—albeit losses are a significant problem for QPV [QS15].

- As shown in [LL11] and discussed in Section 3.2.5, for almost all angles QPV_θ is resistant to exact attacks from adversaries sharing up to a maximally entangled pair of qutrits. It is natural to ask how well can attackers do with bigger (while still relatively small) entangled states, and if relaxing the requirements to approximate attacks changes the picture significantly.

- From the theoretical standpoint, some of the INQC-based attacks we reviewed in the literature point to an increased attack complexity for unitaries at higher levels of the Clifford hierarchy (Sections 3.2.9 and 3.2.10). By tuning the parameter θ we have quick access to any level of the hierarchy, as well as unitaries outside of it.

- A further reason to explore low-dimensional protocols is that they are open to be analyzed by numerical methods, on a similar vein as the work we did in Chapter 2.

We detail our conventions for QPV_θ here as reference for the rest of the thesis.

Protocol 6 (QPV$_\theta$)**.** The verifiers set up stations V_1 and V_2, collinear with P (Fig. 3.1). During one round of the protocol:

1. V_1 and V_2 agree on two random bits $x, b \in \{0,1\}$ by means of pre-shared randomness or through a secure classical channel.

2. V_1 prepares the state $|\psi\rangle = (R_\theta)^b |x\rangle$, where

$$R_\theta = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \tag{3.11}$$

   is a real rotation matrix[27] defining the encoding basis for $|x\rangle$. Then, $|\psi\rangle$ is sent towards P through a public quantum channel.

3. V_2 sends b towards P through a public classical channel, carefully timed such that the quantum state and the classical bit arrive simultaneously at P.

4. Upon receiving $|\psi\rangle$ and b, the prover applies $(R_\theta^\dagger)^b = R_{-\theta}^b$ to $|\psi\rangle$ and measures in the computational basis, recovering x. He immediately broadcasts x to V_1 and V_2.

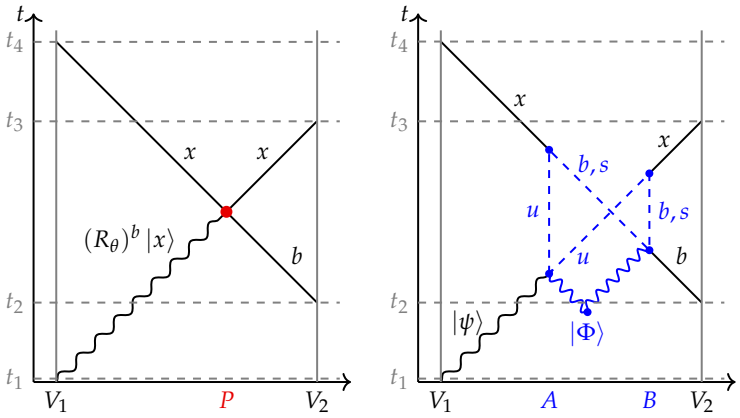
5. The verifiers receive the results, check their correctness and that the timestamps of the received signals are consistent with the honest prover being at P.

The above steps are repeated for N rounds. The protocol terminates successfully if the answers to the challenges have been accepted often enough. We will not account for losses in the following,[28] but we allow the honest prover to output the wrong x with probability ε. According to the precision of their clock, the verifiers bound the prover's position to a neighborhood of P.

Remark. We can choose the computational basis for $b = 0$ without loss of generality if $b \in \{0,1\}$. Indeed, for any pair of single-qubit unitaries B_0 and B_1 which the verifier chooses to apply to the secret bit $|x\rangle$, we can

[27] We choose θ to be akin to the polarization angle, at variance with the convention for a σ_y rotation in the Bloch sphere where the corresponding angle would have been $\theta/2$.

[28] Loss-tolerant protocols have been explored for example in [QS15].



(a) honest prover at P (b) colluding attackers at A, B

Figure 3.5: Spacetime diagrams of QPV_θ protocol and attack model. Lines at 45° represent lightspeed quantum (ondulated) and classical (straight, solid and dashed) channels. (a) When a prover is present at P, they measure the quantum input in the correct basis and broadcasts the measurement result x back to V_1 and V_2. (b) Attackers have access to locations A and B and to the quantum resource $|\Phi\rangle$. They share the classical outcomes of their measurements and attempt to reconstruct x in time to be broadcast back to the verifier.

always find an equivalent protocol with $B'_0 = I$ and $B'_1 = R_\theta$ by setting $\cos(\theta) = \langle 0|B_0^\dagger B_1|0\rangle$. Up to a global rotation (which can be included in the protocol), the four quantum inputs can then be described by $b, x \in \{0,1\}$ as $|\psi\rangle = (R_\theta)^b |x\rangle$.

It can be easily seen from the definition that $\text{QPV}_{\pi/4}$ is just the old QPV_{BB84}. If $\theta \equiv 0 \pmod{\frac{\pi}{2}}$ then R_θ is trivial (i.e. $= I$ or XZ): we call the resulting QPV_θ protocol *classical*.

3.3.2 Attack model

On top of the common features in Section 3.1.1, we have to specify our own choice of attack framework. Our model is directly motivated by the way we chose to implement the numerical search of the attack space. For both

analytic and numerical results, there is certainly a tradeoff here: considering more general attackers leads to stronger security evidence but involves harder proofs or heavier computation, limiting the searchable range. In the decision, we were primarily inspired by the teleportation attack for the $\mathrm{QPV}_{\mathrm{BB84}}$ protocol [KMS11; LL11], which we recalled in Section 3.2.4. In the following we will study similar attacks but for different values of θ.

As usual, the attackers Alice and Bob have no access to the location P to be authenticated, but control two stations A and B respectively located between V_1 and P and between P and V_2. A resource quantum state $|\Phi\rangle$ is pre-shared between the two stations at the start of each round of the protocol. QPV_θ requires the prover to output a classical message, so we constrain internal communication to be classical as well. This is in line with many of the works in the literature, save for a couple of results [Tom+13; BCS21]; in [GC20] it is called LOBC model (Section 3.2.11).

The focus of our analysis is the dimension d^2 of the shared entangled state $|\Phi\rangle$. For this and other technical reason, we make somewhat restrictive assumptions:

- The quantum operations of Alice and Bob are unitary evolutions or projective measurements.
- We fix $|\Phi\rangle$ as the maximally entangled qu$d$it pair in order to exploit some of its properties and simplify the analysis.
- Alice and Bob act identically and independently on each round.

We do not look at the (bigger) space of general quantum maps acting on $|\Phi\rangle$. Recall that in general non-unitary operations can be extended to a unitary one through a Stinespring dilation [Wil13], using only resources local to Alice and Bob. However we do not factor in these extra local resources when we look for attacks in the following. It should be noted that this model includes most of the known attacks, and it is only really relevant in this work as establishing the security domain of the theorems in Section 3.4.1 and when interpreting the results of our computer search as numerical evidence for the optimality of our attacks. Moreover, the choice of this model leads to an optimal attack for $\mathrm{QPV}_{\mathrm{BB84}}$ [RG15].

From the nonlocal computation point of view, attacking QPV_θ involves the INQC implementation of the following two-qubit unitary:

$$U_\theta = \mathrm{c_A}\text{-}X_\mathrm{B} \cdot \mathrm{c_B}\text{-}(R_{-\theta})_\mathrm{A}, \tag{3.12}$$

where $\mathrm{c_A}\text{-}X_\mathrm{B}$ is a CNOT controlled on Alice's side and $\mathrm{c_B}\text{-}(R_{-\theta})_\mathrm{A}$ is a rotation controlled on Bob's. U_θ is applied on the incoming state $R_\theta^b |x\rangle \otimes |b\rangle$:

$$U_\theta \left(R_\theta^b |x\rangle \otimes |b\rangle \right) = |x\rangle \otimes |x \oplus b\rangle . \tag{3.13}$$

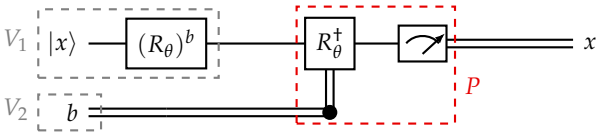
This embedding is useful to compare known attacks to ours, in particular the efficient ones in [GC20]. We will give more details about this comparison in Section 3.4.4.

3.3.3 Circuit picture

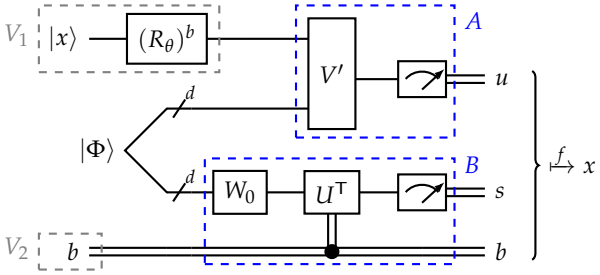
The spacetime diagrams in Fig. 3.5 are useful to visualize the timing constraints, but they don't encode information about which operations are carried out at each point (*event*) in spacetime. We can upgrade the representation to a spacetime *circuit*, defined in [Unr14], in which we can give a detailed picture of QPV_θ, both in the case of honest prover and cheating adversaries (Fig. 3.6). In the following, we argue that under our model the adversarial circuit in Fig. 3.6b describes any attack to QPV_θ in full generality.

Alice and Bob's strategy consists in obtaining clonable classical information by interacting the inputs they receive from the verifiers with their local share of the resource state $|\Phi\rangle$. In the end, their goal is to deduce x. As usual, Alice is unaware of the basis b in which the incoming qubit $(R_\theta)^b |x\rangle$ is encoded. Her actions are modeled by a unitary operation V' acting on the joint system of the verifier's qubit $|\psi\rangle$ and her half of the entangled qudit pair, followed by a measurement in the computational basis. The outcome she obtains is $u \in \mathbb{Z}_{2d}$, which can be forwarded to Bob.

Bob initially only receives the basis bit $b \in \{0, 1\}$. His only quantum resource is his half of the qudit pair, to which he can apply a unitary W_b followed by a measurement in the computational basis. He obtains the outcome $s \in \mathbb{Z}_d$, which he forwards to Alice along with b. Now, without



(a) honest protocol



(b) colluding attackers

Figure 3.6: Circuit representation of the spacetime diagrams in Fig. 3.5, where the actions of the verifier, prover and attackers correspond here to the dashed boxes. The causal relations are enforced by the wires between the boxes; the final broadcasting of x is not represented.

loss of generality, we define:

$$U^\mathsf{T} := W_1 W_0^\dagger, \text{ where } W_b = (U^\mathsf{T})^b W_0. \tag{3.14}$$

This allows to rewrite Bob's unitary as a fixed gate W_0 followed by a gate U^T classically conditioned on $b = 1$. In general, the value of x sought can be a function of all the classical information they have obtained—as they can share it freely during the protocol. Therefore, the attack is completed by a classical map $f(b, s, u) = x$ that they each independently compute after exchanging their measurement results.

Now, we make use of the fact that $|\Phi\rangle$ is maximally entangled. We have in this case:

$$(I \otimes W_b^\mathsf{T}) |\Phi\rangle = (W_b \otimes I) |\Phi\rangle , \tag{3.15}$$

which lets us "shift" Bob's unitaries W_0 and $(U^\mathsf{T})^b$ to Alice's side. This

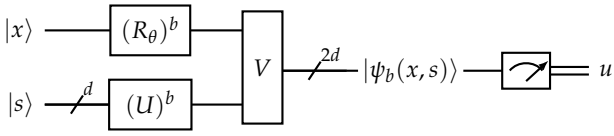


Figure 3.7: The reduced circuit. It is no longer a spacetime circuit, hence we dropped the actors' labels. It is nonetheless equivalent to Fig. 3.6b when $|s\rangle$ is chosen uniformly at random and $V \coloneqq V'(I \otimes W_0^{\mathsf{T}})$. Bob's measurement of his share of the entangled qudit has been omitted.

leads to a formally equivalent circuit for the attack, shown in Fig. 3.7. In this version the unitary $W_0^{\mathsf{T}} U^b$ is performed by Alice on *her* half of the entangled state. The actions of Bob are not shown: the only operation left to him is to measure right away his half in the computational basis, obtaining outcome s, which collapses Alice's half into the qudit state $|s\rangle$. We can then group Alice's operations in two unitaries V and U, respectively acting on her "input" qudit $|s\rangle$ and on the whole $2d$-dimensional space:

$$V \coloneqq V'(I \otimes W_0^{\mathsf{T}}) \tag{3.16}$$

This simplified circuit gives a leaner description of the problem, and reduces the description of an attack to two unitaries and a classical postprocessing function. It should be noted that Fig. 3.7 is no longer a spacetime circuit, as it does not preserve the spacetime locality of the operations. Indeed, in the real world Alice has no access to b and cannot decide when to apply the correction U. The sense in which the reduced circuit is useful is that a specific (U, V, f) can be converted to an attack to QPV_θ, and viceversa all attacks under our model are parametrized by a tuple of that kind.

We denote with $|\psi_b(x,s)\rangle$ the output state of our reduced circuit before the final measurement:

$$|\psi_b(x,s)\rangle \coloneqq V(R_\theta \otimes U)^b(|x\rangle \otimes |s\rangle). \tag{3.17}$$

A pair of unitaries (U, V) defines a set of such states for all $s \in \mathbb{Z}_d$ and $b, x \in \{0,1\}$, of cardinality $4d$. As will be clear in the following, the choice

of f is essentially unique for the exact attacks and suitably chosen (i.e. part of the definition) for the approximate attacks; we will then omit it from the attack description. The rest of the Chapter is dedicated to work out various properties of these states and uncover some symmetries for various values of d. For now, we can start by characterizing the inner products between the states in eq. (3.17): for all $b, x, y \in \{0,1\}$ and $s, t \in \mathbb{Z}_d$,

$$\langle \psi_b(x,s) | \psi_b(y,t) \rangle = \delta_{xy}\delta_{st}, \tag{3.18}$$

$$\langle \psi_0(x,s) | \psi_1(y,t) \rangle = \langle x | R_\theta | y \rangle \langle s | U | t \rangle . \tag{3.19}$$

Equation (3.18) shows that all states with the same b are orthonormal, while eq. (3.19) shows that the inner product of states of different b is independent of V. From eq. (3.17) we can additionally see that the unitarity of R_θ, U, V implies that each of the two subsets of states $|\psi_b(x,s)\rangle$ corresponding to fixed b have to span the whole output space ($\simeq \mathbb{C}^{2d}$), i.e. they form an orthonormal basis.

3.4 Exact attacks against QPV$_\theta$

For Alice and Bob to carry out an exact attack, they need a function $f(b,s,u)$ which correctly predicts x with unit probability. We can see the consequences of this requirement from Alice's point of view: her measurement of $|\psi_b(x,s)\rangle$ in the computational basis $\{|u\rangle\}$ has to be such that, when learning b and s, her outcome u unambiguously selects one of the subsets indexed by $x = 0$ or $x = 1$. This imposes the first strong restriction on the states in eq. (3.17). We call it the *deterministic distinguishability condition* (DDC):

Definition 5 (DDC)**.** The set of states $|\psi_b(x,s)\rangle$ for some (U,V) satisfies the DDC if and only if, for all $u \in \mathbb{Z}_{2d}$, $s \in \mathbb{Z}_d$ and $b \in \{0,1\}$:

$$\langle u | \psi_b(0,s) \rangle = 0 \quad \text{or} \quad \langle u | \psi_b(1,s) \rangle = 0. \tag{3.20}$$

Equivalently:

$$\langle u | \psi_b(0,s) \rangle \langle u | \psi_b(1,s) \rangle = 0. \tag{3.21}$$

When the DDC is satisfied, an f such that $f(b,s,u) = x$ is naturally constructed by assigning the value of x which has a nonzero probability of occurring. Notice that f might only need to be partially defined: an outcome u may not occur at all for some s, b. The DDC has a nice geometrical interpretation: it says that the states $|\psi_b(0,s)\rangle$ and $|\psi_b(1,s)\rangle$ need to have disjoint supports in the computational basis. Furthermore, the DDC has another technical consequence on the allowed amplitudes which we prove below.

Theorem 8. *Either QPV_θ is classical,*[29] *or the states $|\psi_b(x,s)\rangle$ have to satisfy:*

$$\sum_s |\langle u|\psi_b(x,s)\rangle|^2 = \frac{1}{2}, \tag{3.22}$$

for all $b, x \in \{0,1\}$ and $u \in \mathbb{Z}_{2d}$.

Proof. As noticed above, the families of vectors $\{|u\rangle\}$, $\{|\psi_0(x,s)\rangle\}$ and $\{|\psi_1(y,t)\rangle\}$ form three orthonormal bases of the same space of dimension $2d$. We can expand $|\psi_1(y,t)\rangle$ in the $\{|\psi_0(x,s)\rangle\}$ basis, obtaining:

$$\begin{aligned} |\psi_1(y,t)\rangle &= \sum_{x,s} \langle\psi_0(x,s)|\psi_1(y,t)\rangle\, |\psi_0(x,s)\rangle \\ &= \sum_{x,s} \left(\langle x| \otimes \langle s|\right) V^\dagger V (R_\theta |y\rangle \otimes U|t\rangle)\, |\psi_0(x,s)\rangle \\ &= \sum_{x,s} \langle x|R_\theta|y\rangle\, \langle s|U|t\rangle\, |\psi_0(x,s)\rangle, \end{aligned} \tag{3.23}$$

which is valid $\forall y \in \{0,1\}, \forall t \in \mathbb{Z}_d$. For the second step, we used eq. (3.19). For brevity, we define the scalar $\psi_{u,b}(x,s) \coloneqq \langle u|\psi_b(x,s)\rangle$. The DDC can thus be seen as imposing

$$\psi_{u,b}(0,s)\, \psi^*_{u,b}(1,s) = 0, \tag{3.24}$$

$\forall u \in \mathbb{Z}_{2d}, \forall s \in \mathbb{Z}_d$ and $\forall b \in \{0,1\}$. Projecting eq. (3.23) onto $|u\rangle$ we have,

[29] Recall that this means θ is a multiple of $\frac{\pi}{2}$.

for all $y \in \{0,1\}$:

$$\begin{aligned}\psi_{u,1}(y,t) &= \sum_{x,s} \langle x|R_\theta|y\rangle \, \langle s|U|t\rangle \, \psi_{u,0}(x,s) \\ &= \sum_{s} \langle s|U|t\rangle \, \Big[\langle 0|R_\theta|y\rangle \, \psi_{u,0}(0,s) + \langle 1|R_\theta|y\rangle \, \psi_{u,0}(1,s)\Big],\end{aligned} \tag{3.25}$$

substituting $y = 0$ and $y = 1$:

$$\psi_{u,1}(0,t) = \sum_{s} \langle s|U|t\rangle \, \Big[\cos(\theta)\psi_{u,0}(0,s) + \sin(\theta)\psi_{u,0}(1,s)\Big], \tag{3.26}$$

$$\psi_{u,1}(1,t) = \sum_{s} \langle s|U|t\rangle \, \Big[\cos(\theta)\psi_{u,0}(1,s) - \sin(\theta)\psi_{u,0}(0,s)\Big]. \tag{3.27}$$

Using the DDC (in form 3.24) for $b = 1$, i.e. $\psi_{u,1}(0,s)\,\psi^*_{u,1}(1,s) = 0$, we obtain along with eqs. (3.26) and (3.27):

$$\begin{aligned}\sum_{s,s'} \langle s|U|t\rangle \, \langle s'|U|t\rangle^* \cdot \big[\cos(\theta)\psi_{u,0}(0,s) + \sin(\theta)\psi_{u,0}(1,s)\big] \\ \cdot \big[\cos(\theta)\psi^*_{u,0}(1,s') - \sin(\theta)\psi^*_{u,0}(0,s')\big] = 0.\end{aligned} \tag{3.28}$$

Summing over t gives:

$$\begin{aligned}\sum_{s,s'}\Big(\sum_{t} \langle s|U|t\rangle \, \langle t|U^\dagger|s\rangle\Big) \cdot \big[\cos(\theta)\psi_{u,0}(0,s) + \sin(\theta)\psi_{u,0}(1,s)\big] \\ \cdot \big[\cos(\theta)\psi^*_{u,0}(1,s') - \sin(\theta)\psi^*_{u,0}(0,s')\big] = 0,\end{aligned} \tag{3.29}$$

and since $\sum_t |t\rangle\langle t| = I$ and $\langle s|s'\rangle = \delta_{ss'}$, we have

$$\begin{aligned}\sum_{s} \big[\cos(\theta)\psi_{u,0}(0,s) + \sin(\theta)\psi_{u,0}(1,s)\big] \\ \cdot \big[\cos(\theta)\psi^*_{u,0}(1,s) - \sin(\theta)\psi^*_{u,0}(0,s)\big] = 0.\end{aligned} \tag{3.30}$$

With the DDC for $b = 0$, this finally simplifies to:

$$\cos(\theta)\sin(\theta)\Big(\sum_{s} |\psi_{u,0}(1,s)|^2 - \sum_{s} |\psi_{u,0}(0,s)|^2\Big) = 0\,. \tag{3.31}$$

Now we make use of the assumption of non-classicality of the protocol, namely $\theta \not\equiv 0 \pmod{\frac{\pi}{2}}$, which ensures $\cos(\theta)\sin(\theta) \neq 0$. Then, using the

following "normalization relation" of $|u\rangle$ in the $\{|\psi_0(x,s)\rangle\}$ basis:

$$\sum_s |\psi_{u,0}(0,s)|^2 + \sum_s |\psi_{u,0}(1,s)|^2 = 1, \tag{3.32}$$

eq. (3.31) implies that:

$$\sum_s |\psi_{u,0}(0,s)|^2 = \sum_s |\psi_{u,0}(1,s)|^2 = \frac{1}{2}, \tag{3.33}$$

which is the part of the result we seek, for $b = 0$. The argument above can be repeated step by step, this time writing $|\psi_0(x,s)\rangle$ in the $\{|\psi_1(y,t)\rangle\}$ basis. Eventually, we obtain for a nonclassical protocol:

$$\sum_s |\psi_{u,b}(x,s)|^2 = \frac{1}{2} \tag{3.34}$$

for all $u \in \mathbb{Z}_{2d}$ and $b, x \in \{0,1\}$. □

Theorem 8 has a number of consequences on the supports of the $|\psi_b(x,s)\rangle$ states, which will be exploited in the next Section.

3.4.1 Graphical language

The restrictions imposed by the DDC and by Theorem 8 suggest that, for small Hilbert spaces, we can recover enough structure on the states

$$|\psi_b(x,s)\rangle = V(R_\theta \otimes U)^b(|x\rangle \otimes |s\rangle) \tag{3.17}$$

to constrain the allowed attacks. To this aim, we introduce a custom representation of the $2d$-dimensional Hilbert space, in which the states $|\psi_b(x,s)\rangle$ are embedded. The visualization is loosely based on *hypergraphs*, a generalization of graphs where an edge is allowed to join any number of vertices. However, we need an additional feature: while in regular hypergraphs the edges are only described by the vertices they join, we allow our edges a label (color-coded). This way, two edges can join the same subset of vertices.

Figure 3.8, as well as pairs (3.35) to (3.39) in the proof below, are examples of the visualization. Each element of the hypergraph has a counterpart:

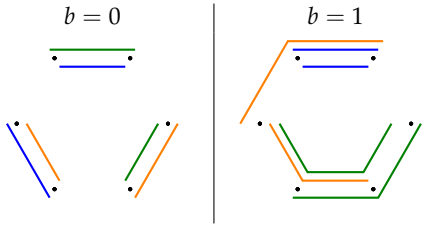


Figure 3.8: An hypergraph pair, describing an (unsuccessful) attack for $d = 3$. Each hyperedge marks the support of a $|\psi_b(x,s)\rangle$, with its color encoding the index s and its position (whether inside or outside the vertices "ring") encoding the index x.

- Each vertex corresponds to an element $|u\rangle$ of the computational basis.
- To each edge joining a specific subset of vertices is assigned a state $|\psi_b(x,s)\rangle$ having support in the corresponding subspace.
- Each hypergraph has $2d$ vertices and d edges in the inner region, where $x = 0$, and d in the outer region, where $x = 1$.
- The color of the edges encodes the different values of the index s.
- A full attack strategy is represented by a pair of graphs drawn side-by-side, one for $b = 0$ and one for $b = 1$.

Remark. The representation is not one-to-one. Important information about the state is lost: for example, the specific amplitudes $\langle u|\psi_b(x,s)\rangle$ are not encoded in the hypergraph.

As we envisioned above, the restrictions imposed by eqs. (3.18) to (3.22) can be captured by imposing some structure on the allowed graphs. We have derived the following necessary (but not sufficient) properties, valid for all d:

(I) The disjointness of $|\psi_b(0,s)\rangle$'s and $|\psi_b(1,s)\rangle$'s supports implies that any vertex joined by an s-colored inner edge cannot *also* be joined by the corresponding s-colored outer edge.

(II) Equation (3.22), which gives the total "probability budget" for all inner (outer) edges joining a given vertex, has several graphical implications:

  (a) All vertices have to be part of at least one inner and one outer edge;

  (b) Vertices joined by an inner (outer) edge of length 2 cannot be part of other inner (outer) edges;

  (c) Each edge has to join at least two vertices and, due to Property I, cannot join more than $2d - 2$ vertices.

(III) Due to Properties I and II, no vertex can be covered by all inner edges or by all outer edges.

(IV) According to eq. (3.17), all states $|\psi_b(x,s)\rangle\rangle$ with the same b are orthogonal to each other. This forbids any two edges from having only one vertex in common.

(V) Finally, while a bit trickier to visualize, eq. (3.19) imposes that if an s-colored edge on the $b = 0$ hypergraph does not share any vertex with a t-colored edge on the $b = 1$ one, then all four edges of that color combination (s,t) represent orthogonal states.

Remark. The order of the computational basis elements $|u\rangle$, $|x\rangle$ and $|s\rangle$, as well as the basis bit b, is irrelevant; clearly, any permutation can be taken into account in the classical postprocessing step by redefining $f(b,s,u)$ accordingly. Therefore all hypergraph pairs that can be obtained by swapping colors, inner edges with outer edges, left graph with right graph and vertex order[30] are equivalent to each other, as shown in Fig. 3.9.

Properties I to V are powerful enough to completely characterize the exact attacks for $d = 2$ and $d = 3$. We prove the two cases in Theorem 9 for adversaries sharing an ebit and Theorem 10 for a maximally entangled three-level system. These theorems are nothing new: the same characterization

[30]Once a vertex assignment is chosen for one of the hypergraphs in a pair, the other cannot be freely rearranged anymore. Nonetheless, this sometimes allows to "prettify" the graphs by drawing as many edges as possible which only join consecutive vertices.

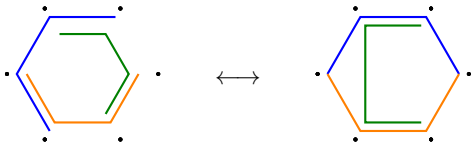
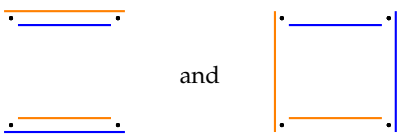


Figure 3.9: Two equivalent $d = 3$ hypergraphs.

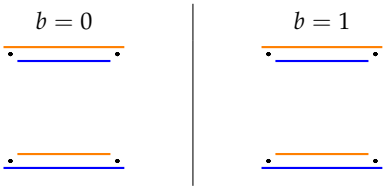
has already been carried out in [LL11], as the reader might recall from Section 3.2.5. However, our hypergraph objects enable an arguably easier proof and could be used in the future as an avenue to the $d = 4$ case, which we could only partially characterize by means of a computer enumeration.

Theorem 9. *Under the assumptions of our attack model (Section 3.3.2), adversaries sharing a maximally entangled qubit cannot perfectly break QPV_θ unless θ is a multiple of $\pi/4$.*

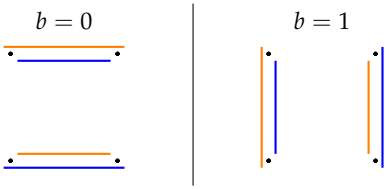
Proof. For $d = 2$, our hypergraphs have four vertices, two inner and two outer edges. Before even building the attack pair, we can start by ruling out the allowed hypergraphs which can be part of it for fixed b. Indeed, Property IIc alone is sufficient to reduce them to only two possibilities:



but an application of Property I immediately rules out the second one. Then, up to vertex reordering, two hypergraph pairs are possible:



(3.35)



(3.36)

but Property V on pair (3.35) implies that the four top edges (inner blue and outer orange) would correspond to four orthogonal states defined on the same support of size 2, which is impossible. Therefore, only pair (3.36) is allowed. Using Theorem 8 we can even recover the absolute value of the amplitudes for this configuration, which has to be $\frac{1}{\sqrt{2}}$ everywhere. Then:

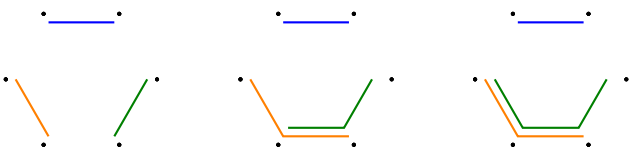
$$|\langle\psi_0(x,s)|\psi_1(y,t)\rangle| = |\langle x|R_\theta|y\rangle\,\langle s|U|t\rangle| = \frac{1}{2} \tag{3.37}$$

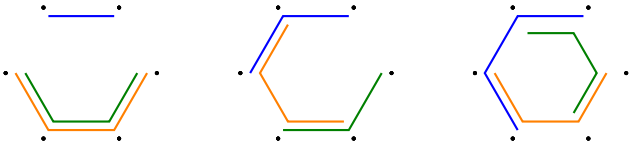
for all x, y and s, t. This implies $\theta = \frac{n\pi}{4}$ and $U = R_{\pi/4}$ up to phases. □

Surprisingly, a maximally entangled qutrit ($d = 3$) gives the adversaries even less power than in the case above, as already noticed in [LL11]: it turns out that they can only break *classical* protocols. In Section 3.5 we will see that this is not just a quirk of exact attacks, as this "inversion" also occurs (for θ in some range) in the approximate case.

Theorem 10. *Under the assumptions of our attack model (Section 3.3.2), adversaries sharing a maximally entangled qutrit cannot perfectly break QPV_θ unless θ is a multiple of $\pi/2$.*

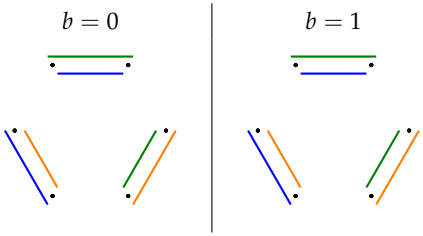
Proof. For $d = 3$ we start sifting through the legal hypergraphs by focusing on the inner edges. We have to place three edges of length ranging from 2 to 4, while satisfying Properties I to IV. We are left with six possibilities, which we can fortunately all draw using "contiguous" hyperedges:



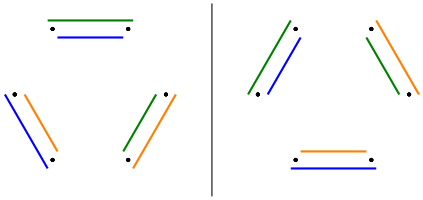


The outer edges are subject to the same rules, meaning that they can only be picked from the same six cases above. However, not all combinations work: when placing them, we have to be careful to respect Properties II and IV. We are left with just two non-trivial cases:

We can rule out the second hypergraph because it requires that three states—for example the ones represented by the inner green, inner orange and outer blue edges—are all orthogonal on the 2-dimensional intersection of their supports. Therefore, only the first case can be used to build an attack pair. Depending on vertex reordering, we obtain two possibilities:



(3.38)



(3.39)

but Property V applied to pairs (3.38) and (3.39) shows that either we have again too many orthogonal states on a support of size two, or that the supports of two orthogonal states have a 1-dimensional intersection. In both cases we have a contradiction, so there is *no exact attack* to QPV_θ for any nonclassical θ for adversaries sharing an entangled qutrit, confirming the result in [LL11] and proving the theorem. □

When applied to the $d = 4$ case, Properties I to IV are unfortunately not powerful enough to prove such strong theorems. With the help of a systematic computer search which we first used to verify our proofs, we enumerated all compliant $d = 4$ hypergraphs. Adding a couple of more refined conditions (which are difficult to check by inspection), we were able to get the configurations of inner edges down to about a thousand. From them, we could single out 17 hypergraphs that admit at least one legal set of outer edges. They are reported in Table 3.1. From them, we tried to construct attack pairs and use Property V to sift through all the possibilities, but the legal pairs ended up too numerous to be handled manually. A more careful analysis of the DDC could give tighter rules, allowing to reduce them to a manageable number, but we did not pursue further this avenue. Instead, in the next Section we present an alternative route to find new attacks.

3.4.2 Numerical optimization: a comeback

Exploiting our circuit simplification of Section 3.3.3, we reduced the problem of looking for exact attacks which use a pair of maximally entangled qudits to finding a pair of matrices (U, V) such that the states $|\psi_b(x,s)\rangle$ in eq. (3.17) satisfy the DDC. Our hypergraph approach is an attempt to constrain the allowed attacks without having to directly look at the matrix entries (i.e. the amplitudes $\langle u|\psi_b(x,s)\rangle$).

In analogy with our numerical approach to linear optical Bell measurement, we can pose our search of attacks to QPV_θ problem as a nonlinear optimization over unitary matrices. For exact attacks however, we do not have to set up a constrained optimization. In the following we will exploit

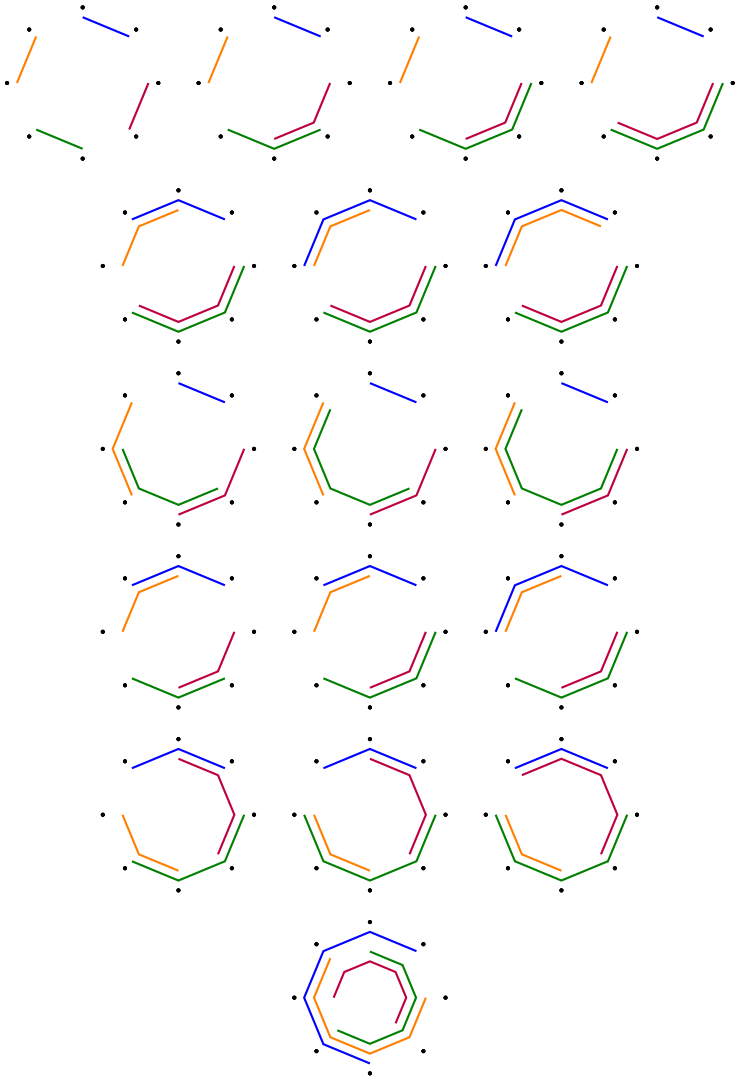


Table 3.1: The 17 valid sets of inner edges for $d = 4$ which can be extended to at least one set of outer edges without contradicting Properties I to IV. Due to the symmetry between $x = 0$ and $x = 1$, the outer edges have to be chosen among the above too.

more of its structure, phrasing it as finding the solution of a system of real polynomial equations in the entries of U, V and R_θ.

Methods

The DDC can be written (eq. (3.21)) as:

$$\langle u|\psi_b(0,s)\rangle \langle \psi_b(1,s)|u\rangle = 0, \tag{3.40}$$

for all $s \in \mathbb{Z}_d$, $u \in \mathbb{Z}_{2d}$ and $b \in \{0,1\}$. We can obtain polynomial equations from eq. (3.40) using the definition of the states $|\psi_b(x,s)\rangle$, which we recall here:

$$|\psi_b(x,s)\rangle := V(R_\theta \otimes U)^b(|x\rangle \otimes |s\rangle). \tag{3.17}$$

By writing U_{st} for $\langle s|U|t\rangle$, $V_{u,xs}$ for $\langle u|\, V\big(|x\rangle \otimes |s\rangle\big)$ and R_{xy} for $\langle x|R_\theta|y\rangle$, expanding the matrix product leads us to:

$$V^*_{u,0s} V_{u,1s} = 0 \tag{3.41}$$

for $b = 0$, and:

$$\Big(\sum_{ij} V_{u,ij} R_{i0} U_{js}\Big)\Big(\sum_{kl} V^*_{u,kl} R_{k1} U^*_{ls}\Big) = 0 \tag{3.42}$$

for $b = 1$. Additionally, we have to enforce unitarity of U, V:

$$\sum_{k=1}^{d} U^*_{k,i} U_{k,j} = \delta_{ij} \qquad \forall i,j \in \mathbb{Z}_d, i \geq j, \tag{3.43}$$

$$\sum_{k=1}^{2d} V^*_{k,i} V_{k,j} = \delta_{ij} \qquad \forall i,j \in \mathbb{Z}_{2d}, i \geq j. \tag{3.44}$$

These constraints are not polynomial equations *as is*, due to the presence of the complex conjugate. They can nonetheless always be written as real polynomials of real variables, in the real and imaginary parts of U's and V's entries.

It is convenient at this point to properly define what exactly we want to achieve with a numerical search.

1. In order to find *new* attacks, we can search over whatever subset of the domain we want; in particular, we can restrict U, V to be real matrices, approximately saving a factor of 2 in the number of variables.

2. In order to gain (numerical) *evidence* of the nonexistence of attacks to QPV_θ under our model for a specific θ, we have to be as general as possible. However, for security purposes it makes little sense to analyze exact attacks; this case is better covered by our search of approximate attacks in Section 3.5.

We will consider the first scenario here. Numerical tests surprisingly suggest that this is not restrictive, i.e. even in the approximate case we find the exact same results when considering real orthogonal matrices vs. general unitary. Assuming real variables then, eqs. (3.41) to (3.44) become

$$2d^2 + 2d^2 + \frac{d(d+1)}{2} + \frac{2d(2d+1)}{2} = \frac{13d^2 + 3d}{2} \tag{3.45}$$

equations of the form $f_i(U, V) = 0$, in $d^2 + (2d)^2 = 5d^2$ scalar variables. If treating θ as a variable instead of a parameter, e.g. in order to scan for potentially weak angles without prior assumptions on their form, we need to add two variables $R_{00} = \cos\theta$ and $R_{10} = \sin\theta$ as well as the constraint $R_{00}^2 + R_{10}^2 = 1$.[31] It should be noted that not all these constraints are independent: as a matter of fact, they cannot be if solutions exist—their number quickly outgrows the number of independent variables. We can see hints of their interdependence already: for example, eq. (3.41) already implies the orthogonality of at least d among the $2d$ columns of V. At variance with the linear optical case, the degree of the system is constant and does not grow with d. The conditions in eq. (3.42) have the highest total degree, which is 4 or 6 depending if θ is treated as a parameter or as a variable; the rest of the equations are at most quadratic.

Even if the system looks heavily overdetermined, we know it has at least a (trivial) solution for all d, namely when the protocol is classical. In the following, we will tacitly ignore those. In order to look for the nontrivial

[31] Technically, the numerical method of our choice works with nonlinear functions; in case we do not strictly require polynomials we only need to add one variable (θ) and no additional constraints.

solutions, we define:

$$F_\theta(U, V) = \sum_i f_i^2, \tag{3.46}$$

i.e. the sum of the squares of all polynomials. The zeros of the function in eq. (3.46) are also simultaneous zeros of all the polynomials f_i. We can minimize F_θ with a numerical method such as gradient descent. If we find zero as minimum, we have found an exact attack for a specific θ. In order to look for zeros of F, we leveraged a nonlinear least-squares method provided by the Python library SciPy [Vir+20]. The method is similar to the one we used for linear optics (Section 2.5.5), with some more details explained in Section 3.5.

Sum of Squares

While computational algebraic tools for working with symbolic polynomial equations are available, their inherent exponential scaling makes them challenging to apply directly to our system. For small(ish) systems it is in fact possible to obtain a proof of unsolvability, using *sum of squares* (SOS) techniques developed by [Par03] in the context of global polynomial optimization [Las01]. Broadly speaking, these proofs work by providing a hierarchy of increasingly complex SDPs, such that the existence of a feasible point of any of them can be turned into a certificate of unsolvability. An important result in real algebraic geometry, *Putinar's Positivstellensatz* [Put93], guarantees that inconsistent systems will produce a certificate at some level of the hierarchy; however the resulting SDP may be too large to be solvable in practice. This was indeed the case for us, where the smallest interesting case where our hypergraph strategy fails (real matrices, $d = 4$) involves 80 variables and 174 polynomials. Inputting this system in a "black-box" solver with only minor optimizations[32] resulted in no certificate being found for the levels of the hierarchy we could reasonably reach. This does not rule out a future role of such proofs for QPV_θ: these approaches can be unsuccessful when applied *as is*, and may prove more effective if coupled with a more careful analysis which takes more of our problem's symmetries

[32] We used the Python's library PICOS [SS22] and SOSTOOLS [Pap+13].

into account.

Results

From the behavior of $d = 2$ and $d = 3$ and the growing number of constraints, Lau and Lo [LL11] expected a bigger entangled state not to help much for other θs than $\pi/4$—putting aside the huge dimensions required for the generic attack, which they could not know at the time. We find a different story: starting from $d = 4$, every even dimension within reach of our computation produce new attacks for more and more θs.

For $d = 4$ we quickly find solutions for all θ multiples of $\pi/8$, showing an attack to a non-Clifford operation ($R_{\pi/8} \in \mathcal{C}_3$ is equivalent to a T gate). In general, we have $R_{\pi/2^n} \in \mathcal{C}_n$. Therefore, sharing two ebits per round is strictly more powerful for the adversaries than sharing just one—or an entangled qutrit pair, for that matter. Based on known attacks to other QPV protocols (notably Speelman's [Spe16] and Chakraborty-Leverrier's [CL15]), we could conjecture a link between the level of the Clifford hierarchy to which R_θ belong and the dimension of the entangled state needed to break it. While this might be the case, we are not constrained to qubits: indeed, we find that a pair of maximally entangled six-level systems gives an attack to $\mathrm{QPV}_{\pi/6}$, despite the corresponding rotation being completely *outside* of the Clifford hierarchy on qubits.

With our program we could tackle pretty big instances: we proceeded to raise the dimension up to $d = 12$, which involves finding a solution of a system of 954 equations in 720 variables. Our findings are collected in Table 3.2. An interesting pattern emerges: for *even* d we find an attack for, among others, all θ multiples of $\pi/2d$. Sometimes, we also get extra angles. For example, $\pi/8$ is broken by $d = 6$ even if a maximally entangled 6-level system does not necessarily provide two ebits. In line with the $d = 3$ case, when the dimension is *odd* it seems the adversaries get less power, as they appear to only be capable to break angles which could already be broken by much smaller d.

A drawback of the numerical strategy is that direct inspection of the solution matrices (U, V) have not offered us a straightforward generalization,

d	2	3	4	5	6	7	8	9	10	11	12	16^*
k	4	2	8	4	8, 12	4	16	4, 6	20	4	24	32^*

Table 3.2: Exact attacks for QPV_θ. Depending on the entangled state dimension d available to the attackers, we list the values of k for which a valid pair (U, V) breaking $\theta = \frac{n\pi}{k}$ is found $\forall n$.
$(*)$ For $d = 16$ we found an explicit attack to $\theta = \frac{\pi}{32}$, but we could not reliably explore the rest of the θ range.

from which an analytic attack strategy for all d could be derived. There are a variety of discrete symmetries that are difficult to remove; furthermore, for $d > 4$ we find that solutions retain continuous degrees of freedom, which makes extracting a "nice" matrix out of them quite difficult. We were able to get some structure for some attacks at small d, which we show below.

Explicit solutions

The following pairs (U, V) are examples of explicit exact attacks we found via inspection of the results of the numerical optimization. We present them in a "tidied" form, by swapping rows and columns and multiplying by phases when these operations lead to an equivalent attack. The matrices are written in terms of 2×2 blocks for readability. For $d = 4$, the solutions split into two types, both of which attack $\mathrm{QPV}_{\pi/8}$:

$$V = \frac{1}{2}\begin{pmatrix} X & I & -Z & ZX \\ ZX & X & I & Z \\ X & -I & -Z & -ZX \\ ZX & -X & I & -Z \end{pmatrix} \tag{3.47}$$

$$U = \frac{1}{\sqrt{2}}\begin{pmatrix} R_{-\pi/8} & R_{\pi/8}\,Z \\ -ZR_{\pi/8} & R_{-3\pi/8} \end{pmatrix},$$

and

$$V = \frac{1}{\sqrt{2}} \begin{pmatrix} XHX & 0 & 0 & I \\ 0 & -XHX & -I & 0 \\ ZX & 0 & 0 & -H \\ 0 & -ZX & H & 0 \end{pmatrix} \tag{3.48}$$

$$U = \frac{1}{\sqrt{2}} \begin{pmatrix} R_{\pi/8} & -R_{-\pi/8} Z \\ Z R_{-\pi/8} & R_{3\pi/8} \end{pmatrix}.$$

For $d = 6$, the (U, V) pair below attack $\mathrm{QPV}_{\pi/12}$:

$$V = \frac{1}{2} \begin{pmatrix} I \otimes \sqrt{2} R_{\pi/6} & \begin{matrix} 0 \\ 0 \end{matrix} & X \otimes \sqrt{2} R_{-\pi/3} & \begin{matrix} 0 \\ 0 \end{matrix} \\ H \otimes I & \begin{matrix} -Z \\ X \end{matrix} & -ZH \otimes X & \begin{matrix} -X \\ Z \end{matrix} \\ H \otimes I & \begin{matrix} Z \\ -X \end{matrix} & -ZH \otimes X & \begin{matrix} X \\ -Z \end{matrix} \end{pmatrix} \tag{3.49}$$

$$U = \begin{pmatrix} A & -\frac{1}{2} ZH & \frac{1}{3-\sqrt{3}} ZX \\ B & \frac{1}{2} ZH & \frac{1}{3+\sqrt{3}} ZX \\ \frac{1}{\sqrt{6}} R_{\pi/12} & \frac{1}{\sqrt{2}} R_{\pi/12} & \frac{1}{\sqrt{3}} ZH \end{pmatrix},$$

where A, B are defined as:

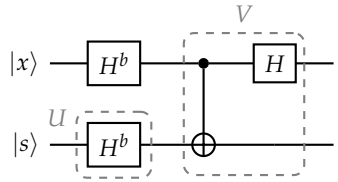
$$A = \frac{2 - \sqrt{3}}{2\sqrt{6}} ZX - \frac{1}{2\sqrt{2}} I, \tag{3.50}$$

$$B = \frac{1}{2\sqrt{2}} XZ - \frac{2 + \sqrt{3}}{2\sqrt{6}} I. \tag{3.51}$$

The above is a special case of a continuum of solutions with one real degree of freedom.

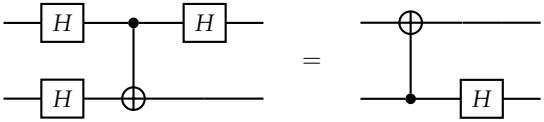
3.4.3 Circuit solutions

Kent's attack for $d = 2$ also corresponds to a (U, V) pair:[33]



(3.52)

We can show a manifestation of the symmetry which allows the attack to work—without referring directly to the teleportation protocol. Suppose the attack works for $b = 0$, namely $H^b = I$. Then, it must work for $b = 1$ too, through the circuit identity [Lom03]:



(3.53)

which makes it clear that the two cases with $b = 0$ and $b = 1$ are the same up to a permutation of the inputs (which does not change the validity of the DDC). Can we find an analogous explicit circuit for the $d = 4$ strategy, such that the actions of Alice and Bob could be explained as a specific quantum algorithm instead of a (much more opaque) unitary? A way to do this is to find a useful decomposition of the (U, V) pair in terms of single qubit gates and controlled-operations. To this aim, we implemented the *quantum Shannon decomposition* (QSD), which is based on the recursive application of a linear algebra matrix decomposition routine called *cosine-sine decomposition* (CSD).

[33] For the sake of clarity in this example, we used an Hadamard instead of the protocol's rotation $R_{\pi/4} = HZ$. The main point still stand, as the extra Z can always be absorbed into V.

Cosine-sine decomposition

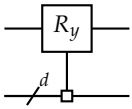
We briefly introduce the decomposition, following [SBM06]. Any $2d \times 2d$ unitary can be decomposed into six $d \times d$ matrices this way:

$$U = \begin{pmatrix} A_1 & \\ & B_1 \end{pmatrix} \begin{pmatrix} C & -S \\ S & C \end{pmatrix} \begin{pmatrix} A_2 & \\ & B_2 \end{pmatrix} \tag{3.54}$$

where the A_i, B_i are unitaries and C, S are diagonal and real. The name of the decomposition comes from the relation $C^2 + S^2 = 1$, which implies C and S are of the form:

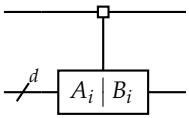
$$C = \begin{pmatrix} \cos\beta_1 & & \\ & \ddots & \\ & & \cos\beta_d \end{pmatrix}, \qquad S = \begin{pmatrix} \sin\beta_1 & & \\ & \ddots & \\ & & \sin\beta_d \end{pmatrix}. \tag{3.55}$$

Each of the three terms in eq. (3.54) corresponds to a specific circuit element. The central term is a multiplexed single-qubit rotation around the y axis by one of the angles β_i, applied to the most significant qubit (i.e. in the usual circuit representation, the first from the top). *Multiplexed* refers to a generalization of a controlled gate where a different unitary is applied for each value of the control system, which may be more than 2-dimensional. The corresponding circuit element is written as:



(3.56)

The left and right block-diagonal terms are also multiplexed gates, which apply A_i or B_i depending on the value of the first qubit:

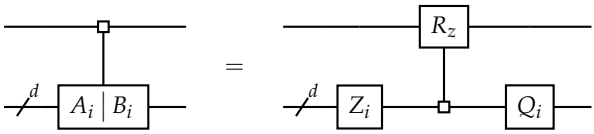


(3.57)

The cosine-sine decomposition can be worked out from the generalized singular value decomposition. We used the LAPACK CSD routine exposed by Scipy [Vir+20].

Intermediate step

While we could directly apply the CSD again on each of the A_i, B_i, [SBM06] provides an intermediate step which "demultiplexes" them, namely it factors out the controlled part:



(3.58)

through the relation:

$$\begin{pmatrix} A_i & \\ & B_i \end{pmatrix} = \begin{pmatrix} Q_i & \\ & Q_i \end{pmatrix} \begin{pmatrix} D_i & \\ & D_i^\dagger \end{pmatrix} \begin{pmatrix} Z_i & \\ & Z_i \end{pmatrix}, \tag{3.59}$$

where Q_i, Z_i are $d \times d$ unitaries and D_i is diagonal with d rotation angles (i.e. phases) around the z axis.[34] They can be obtained by diagonalizing $A_i B_i^\dagger = Q_i D_i^2 Q_i^\dagger$ and working out $Z_i = D_i Q_i^\dagger B_i$. However, this factorization is not unique and this route (suggested in the paper) did not work well numerically for us when $A_i B_i^\dagger$ has degenerate eigenvectors. Luckily, the *generalized Schur* (also known as QZ) *decomposition*—conveniently exposed by Scipy—can directly give Q_i, Z_i and D_i from A_i, B_i.[35]

[34] Remember that the order of the gates is flipped in the circuit vs. matrix representation, a peculiar quirk of the quantum computation conventions which systematic self-observational studies have shown to be the primary source of errors and typos in the field. *[citation needed]*

[35] Technically, the Schur decomposition of a pair of matrices (A, B) returns Q, Z unitary and D_A, D_B upper triangular such that $A = Q D_A Z$ and $B = Q D_B Z$. When A and B are unitary, D_A, D_B are unitary too and thus diagonal. However, the requirement $D_A = D_B^\dagger$ is not guaranteed, and has to be enforced by pulling out phase factors from the columns of Q.

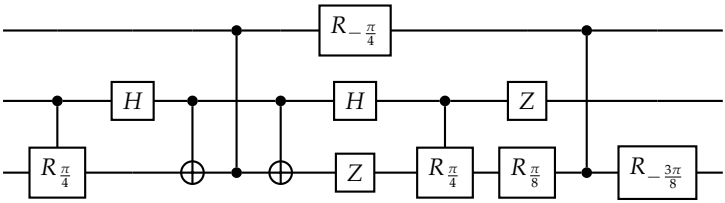
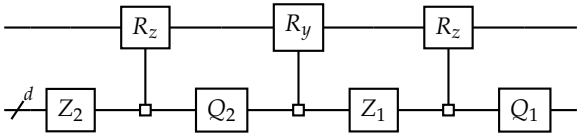


Figure 3.10: A compact circuit which implements V in eq. (3.48). We chose to stick with our convention for R_θ here instead of the one commonly used in the circuit model, i.e. $R_\theta = \sigma_y(2\theta)$.

Quantum Shannon decomposition

One iteration of QSD on $V_{2d\times 2d}$ then looks like the following circuit [SBM06, Theorem 13]:

$V =$  (3.60)

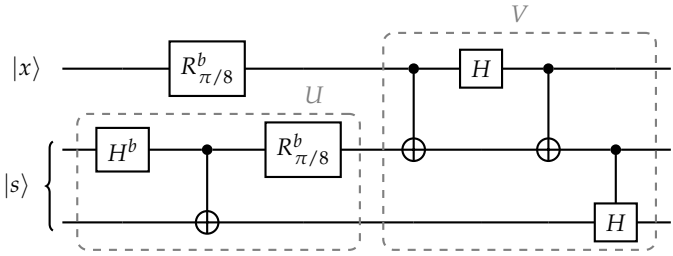
The procedure can then be iterated until the only non-controlled operations are single-qubit gates.

Circuit for $\pi/8$ attack

Unfortunately, while these kind of black-box techniques provide the required decomposition in terms of simple gates, it often happens that significant postprocessing is required in order to get the circuit down to a manageable size. In our case, we applied QSD to the (U, V) pair in eq. (3.48), obtaining around 80 gates for the 8×8 unitary V and a dozen for U. We employed the full arsenal of circuit simplifications we could find, ranging from lists of known circuit identities [Lom03] to the use of ZX calculus [CD11; Wet20]. For now, we focus on the simplification of V, which we could significantly reduce to a handful of gates (Fig. 3.10).

However it is still significantly more complex than in the $d = 2$ case,

which hinders an analogous clean interpretation. We can do better by remembering that the DDC is unaffected by local phases and operations which permute the computational basis. This way, we obtain an extremely barebone circuit for a (different) (U, V) pair which still breaks $\mathrm{QPV}_{\pi/8}$:



(3.61)

3.4.4 QPV_θ in the INQC picture

Our definition of the attack model is tailored to the circuit picture that we propose and exploit. However, as it is clear from our review in Section 3.2, an extensive part of the literature [Spe16; Buh+14; BK11; GC20] characterizes an attack as the much more general INQC implementation of a suitable unitary U_{AB} on a bipartite quantum input ρ_{AB}. In the following, we justify the mapping from our model to INQC that we gave in eq. (3.13). In particular, we show how the linear attacks in [GC20] (Section 3.2.11) compare to ours.

During a round of QPV_θ, the adversaries receive the quantum-classical [Wil13] state:

$$|\Psi_b(x)\rangle_{\mathrm{AB}} = (R_\theta)^b |x\rangle \otimes |b\rangle\,, \tag{3.62}$$

where $x = 0, 1$ and $b = 0, 1$ with equal probability $1/4$. At first glance it seems that in order to get x, they only need to apply via INQC a rotation $R_{-\theta}$ controlled on system B:

$$\mathrm{c_B}\text{-}(R_{-\theta})_{\mathrm{A}} = I \otimes |0\rangle\langle 0| + R_{-\theta} \otimes |1\rangle\langle 1|\,, \tag{3.63}$$

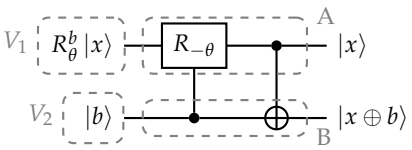
which leaves them with the state $|x\rangle_{\mathrm{A}} \otimes |b\rangle_{\mathrm{B}}$. However, at this point only Alice knows x and cannot send it to Bob: the allowed round of communication has already been used up by the nonlocal protocol. Indeed, eq. (3.63) can be applied on $|\Psi_b(x)\rangle$ without using any entanglement, by just asking

Bob to forward b to Alice.

The unitary we want to implement then should output enough information to retrieve x on both sides. This is realized through a CNOT gate controlled on Alice's side, giving the embedding in eq. (3.12)

$$U_\theta = \mathrm{c_A}\text{-}X_\mathrm{B} \cdot \mathrm{c_B}\text{-}(R_{-\theta})_\mathrm{A} \tag{3.12}$$

which leaves them with the state $|x\rangle_\mathrm{A} \otimes |x \oplus b\rangle_\mathrm{B}$. In circuit form:



(3.64)

Now Bob is also able to retrieve x, by measuring in the computational basis and XORing the result with b—which is also available as a classical bit. The INQC implementation of the two-qubit unitary U_θ is then an attack to QPV$_\theta$.

The protocol defined in [GC20] gives an INQC implementation of all two-qubit unitaries, consuming a linear amount of ebits in the desired approximation accuracy. Through the embedding defined above, all QPV$_\theta$ protocols can be attacked in this way. In order to compare their ebit requirement to ours, we first need to decompose U_θ. Matching their notation, their strategy is based on the decomposition, valid for all two-qubit unitaries U [KC01]:

$$U = (R_1 \otimes S_1)\Omega(R_2 \otimes S_2), \tag{3.65}$$

where R_i, S_i are single-qubit unitaries. The matrix

$$\Omega = \exp\left\{i\left(\alpha\,\sigma_x \otimes \sigma_x + \beta\,\sigma_y \otimes \sigma_y + \gamma\,\sigma_z \otimes \sigma_z\right)\right\} \tag{3.66}$$

describes the nonlocal part of U, and is always diagonal in a basis of maximally entangled states called *magic basis*. For a generic U, their strategy implements Ω with some vanishing probability of failure; comparing it to our exact attacks would only make sense if we defined a (somewhat arbitrary) cutoff error ε. However, they give in two special cases a perfect

implementation of U, provided that α, β, γ are all integer multiples of $\pi/2^n$:

- If $n = 2$, consuming 2 ebits [GC20, Proposition 1];
- If $n > 2$, consuming a finite number of ebits [GC20, Corollary 1].

It turns out that in our case, the angles $\alpha_\theta, \beta_\theta, \gamma_\theta$ corresponding to U_θ are particularly simple. We can obtain their values through the Cartan (also known as KAK) decomposition [DL08]. A solution is:

$$\begin{aligned} R_1 &= \frac{I - iZ}{\sqrt{2}}, & S_1 &= R_{\pi/4}, \\ R_2 &= R_{-\theta/2}, & S_2 &= \frac{Z - iI}{\sqrt{2}} H, \end{aligned} \tag{3.67}$$

which gives a factorization of U_θ in the form of eq. (3.65), with

$$\alpha_\theta = 0, \qquad \beta_\theta = \theta/2, \qquad \gamma_\theta = \pi/4. \tag{3.68}$$

We fall in the exact case then for θ multiple of $\pi/2^n$. An explicit count of the resources used does not seem to be provided in the paper. Going through their INQC protocol we conclude that a direct application of their strategy gives an exact attack consuming $4n + 15$ ebits, which might be further optimizable in the specific case at hand. How does this compare to the attacks we found in Section 3.4.2? From Table 3.2 we can see that for $n = 2, 3, 4, 5$ (i.e. $k = 4, 8, 16, 32$) we need respectively $1, 2, 3, 4$ ebits. If our conjectures on exact attacks holds, there exist an attack for all n requiring just $n - 1$ ebits, a fourfold efficiency improvement over [GC20]. We emphasize though that the gain in ebit consumption is likely due to the large amount of structure in the family U_θ that we consider, which is reflected for example by how α_θ and γ_θ are independent of θ.

Remark. In Section 3.5.4 we will define $\mathrm{QPV}_{(n)}$, a protocol which uses n bases distributed in the interval $[0, \frac{\pi}{2})$ to encode $|x\rangle$ instead of just two. We can analogously embed $\mathrm{QPV}_{(n)}$ into a INQC unitary U_n, which now acts on a $(2 \otimes n)$-dimensional space (as $b \in \mathbb{Z}_n$). The input state is $|\Psi_b(x)\rangle_{\mathrm{AB}} = R_{\theta_b} |x\rangle \otimes |b\rangle$, where the θ_b are defined in Protocol 7. Due to the higher dimensionality, Gonzales and Chitambar's linear attack does not apply to

$\mathrm{QPV}_{(n)}$, *provided that the embedding above is optimal*. We do not prove this, and as a matter of fact the highly structured nature of U_n might provide a direct mapping to [GC20]'s attacks.

3.5 Approximate attacks

We already pointed out in the previous Section that, while exact attacks are useful to highlight the weaknesses in the specific structure of QPV_θ, it is essential for a practical protocol to analyze its security in the presence of imperfect provers. Roughly speaking, the main assumption is that the adversaries could have better equipment than the imperfect prover: this lets them exploit this gap in capabilities to "hide" the unavoidable mark of their presence—which is due to the intrinsic security of the ideal protocol. Imperfections arise in two main ways:

1. As *losses* during the communication among the honest parties, which from the verifier's perspective look like a missed answer to the challenge

2. As *errors* during the honest prover's measurement of $|\psi\rangle$, which result in the wrong secret bit $\tilde{x}$ being sent back to the verifiers.

Losses are an important concern in protocols implemented with discrete-variable photons. Qi and Siopsis [QS15] analyze the resilience of a class of protocols which includes QPV_θ, showing that some modifications are necessary to enable loss-resistance. We instead focus in the next Section on the other main source of imperfections, measurement errors.

3.5.1 Figure of merit

In order to numerically search for optimal attacks, we have to modify the strategy used in Section 3.4.2, by accommodating for the presence of errors. We therefore relax the DDC requirement, and only asks the adversaries to output their best guess for x, i.e. the one which they deem more probable given their measurement results.

Using the notation of Section 3.3.3, we have that for all x, b a measurement result of s (by Bob) and u (by Alice) occurs with probability:

$$p(x,b,s,u) = |\langle u|\psi_b(x,s)\rangle|^2\, p(x)\, p(b)\, p(s). \tag{3.69}$$

Their best guess for the value of x is thus

$$p_{\text{succ}}(b,s,u) = \max_x\, p(x,b,s,u). \tag{3.70}$$

We focus in the following on the probability of error, $p_{\text{err}} = 1 - p_{\text{succ}}$. Our protocol only involves qubits, i.e. $x \in \{0,1\}$, therefore

$$p_{\text{err}}(b,s,u) = \min\{p(0,b,s,u), p(1,b,s,u)\}. \tag{3.71}$$

Recall that in QPV_θ x, b are uniformly distributed (with $p = \frac{1}{2}$) and s is too (with $p = \frac{1}{d}$) in our attack model, as it comes from measuring half of a maximally entangled pair. Then, the overall error probability for an attack strategy can be obtained by summing over b, s, u:

$$p_{\text{err}}(U,V,\theta) = \frac{1}{2\cdot 2\cdot d}\sum_{b,s,u} \min\left\{|\langle u|\psi_b(0,s)\rangle|^2, |\langle u|\psi_b(1,s)\rangle|^2\right\}. \tag{3.72}$$

We can check for consistency that imposing the DDC (eq. 3.20) gives $p_{\text{err}} = 0$, as expected of exact attacks.

3.5.2 Methods

Our new figure of merit we seek to minimize over all attack strategies is, therefore, $p_{\text{err}}(U,V,\theta)$. While the techniques are similar, this case is somewhat different from the search for exact attacks in Section 3.4.2. For starter, at variance with the previous case, we *have* to explore the complex unitary space, if we are to obtain credible evidence for bounds on p_{err}. Only looking among orthogonal matrices like we used to do makes little sense from a security standpoint. Nonetheless, we noticed with some surprise that when restricting to the orthogonal group, which is much faster to search, we obtain the exact same results and curves that are presented below. A possible reason for this phenomenon are symmetries in our attack

model: for example, (U, V) and (U^*, V^*) are both attacks with the same error probability.[36]

Once minimized at various fixed θ, we can obtain for each size d of the adversaries' entangled state a curve $p_{\text{err}}(\theta)$, showing the best minima found:

$$p_{\text{err}}(\theta) = \min_{U,V} p_{\text{err}}(U, V, \theta). \tag{3.73}$$

Remark. We only need to explore a small range of θ, as symmetries allow to restrict the relevant values to $[0, \frac{\pi}{4}]$ through the relations:

$$R_{\frac{\pi}{2}-\theta} = XR_\theta Z \quad \text{and} \quad R_{-\theta} = XR_\theta X. \tag{3.74}$$

The other quadrants are covered by similar relations. The extra X and Z can then be either absorbed into V or taken into account by the adversaries by flipping their prediction for x.

Here, we ideally want to obtain evidence about the *global* minimum of the continuous function (3.72) subject to unitarity constraints, instead of a solution of $p_{\text{err}} = 0$ like in the exact case. This situation is really analogous to the linear optics optimization discussed in Section 2.5.5, and we are similarly faced with the choice between a constrained method and a parametrization of the search space. In general, the shape of the search space can heavily affect the effectiveness of the multistart method we use. To this aim, we separately employ a variety of algorithms and parametrizations:

- Instead of SLSQP, which implements the constraints as lagrangian multipliers without exploiting sparsity, we use IPOPT, a constrained sparse interior point method [WB06]. Here, the search space is big: two arbitrary complex $d \times d$ and $2d \times 2d$ matrices.

- L-BFGS, an unconstrained quasi-Newton method [Byr+95]. Here, U and V are parametrized via skew-hermitian matrices. This reduces the size of the search space a lot, but has the potential to introduce

[36] A result by Rudolph and Grover [RG02] that there is a real gate universal for quantum computation does not help here directly (but may be part of the solution), because it requires an overhead of one qubit. However, it certainly implies that a unitary attack at dimension d can always be simulated by an orthogonal one in dimension $2d$.

unwanted additional structure to it. We explore two choices for the mapping from skew-hermitian to unitary:

- the *Cayley transform* [Cay46; Zhu17]: $U = (I + A)^{-1}(I - A)$,
- the usual exponential map: $U = e^A$.

In Section 2.5.5, we noticed that a major drawback of parametrizing the space through the exponential map is that we lose the convenience of a straightforward analytical gradient. This is where the Cayley transform helps, as its algebraic structure allows for an easy propagation of the gradient. Indeed, if our variables are the entries a_{ij} of A:

$$\begin{aligned}\frac{\partial U}{\partial a_{ij}} &= \frac{\partial}{\partial a_{ij}}\left[(I+A)^{-1}(I-A)\right] \\ &= -\left[(I+A)^{-1}\frac{\partial A}{\partial a_{ij}}(I+A)^{-1}(I-A) + (I+A)^{-1}\frac{\partial A}{\partial a_{ij}}\right] \\ &= -(I+A)^{-1}\frac{\partial A}{\partial a_{ij}}(U+I),\end{aligned} \tag{3.75}$$

where we used in the second line the following relation, valid for an invertible matrix K which depends on a parameter x:

$$\frac{\partial K^{-1}}{\partial x} = -K^{-1}\frac{\partial K}{\partial x}K^{-1}. \tag{3.76}$$

The term $\frac{\partial A}{\partial a_{ij}}$ is just the projector $|i\rangle\langle j|$, which means the derivative ends up being the outer product between the i-th column of $(I + A)^{-1}$ and the j-th row of $(U + I)$. Another nice feature of this mapping is the ability to re-use the result of its most expensive operation (matrix inversion) for the gradient.

Nonetheless, we found the exponential map implementation (with no analytical gradient) to be slightly faster for the small cases ($d < 4$), with IPOPT and Cayley catching up and outperforming for larger d. Interestingly, IPOPT manages similar performances than the Cayley method, despite the bigger search space: we think this is due to the effective exploitation of the constraints' sparsity. In terms of number of starting points needed to

reach the lowest valley, the three methods seem to give comparable results. For example, for $d = 4$ we observe convergence to the same (hopefully global) optimum after between 10^4 and 10^5 randomly sampled starting points. Unfortunately we could not go much further than $d = 5$ for the approximate attacks, at variance than with the exact case. At that point the computation time needed to get reliable curves is beyond the amount we considered reasonable to invest.

3.5.3 Results

Our results are plotted in Fig. 3.11 for $d \leq 5$. Already at first, one can notice a much richer structure than what could be expected given the fairly regular behavior of the exact attacks' angles in Table 3.2. As expected we find consistent results: p_{err} drops to 0 where the exact attack suggest. The shapes of the $p_{\text{err}}(\theta)$ curves we found in Fig. 3.11 merit a consideration, as they appear to be composed of different sections $\{p_1(\theta), p_2(\theta), \ldots, p_n(\theta)\}$. If they represent optimal strategies then this suggests that Alice and Bob's best attacks can be very different depending on the protocol's parameter θ: at various points in the range $[0, \frac{\pi}{4}]$, a certain strategy becomes more effective and overcomes the previous one. In the following, we analyze our results for each dimension d.

$d = 1, 2$ When the adversaries share just one ebit ($d = 2$), the curve we found numerically is reproduced by:

$$p_{\text{err}}(\theta) = \begin{cases} \sin\left(\frac{\theta}{2}\right)^2 & 0 \leq \theta \leq \frac{\pi}{8}\,, \\ \sin\left(\frac{\theta}{2} - \frac{\pi}{8}\right)^2 & \frac{\pi}{8} \leq \theta \leq \frac{\pi}{4}\,. \end{cases} \tag{3.77}$$

The probability in the first half of the range, the region $0 \leq \theta \leq \frac{\pi}{8}$, is the same that can be attained with no entanglement at all ($d = 1$ in Fig. 3.11). As a matter of fact, there is a simple strategy matching the probability for this section of the curve: it is the ubiquitous *pretty good measurement* (PGM) [HW94], which instructs Alice to just measure the unknown $|\psi\rangle = (R_\theta)^b |x\rangle$ in the "intermediate" basis $R_{\theta/2}$ and send the classical result to Bob in the

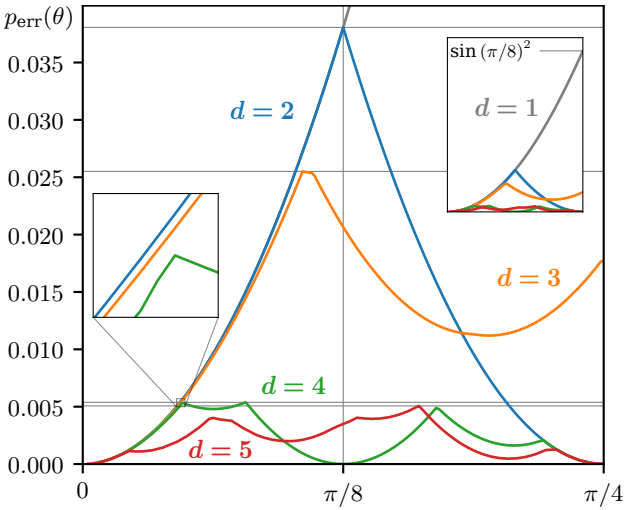
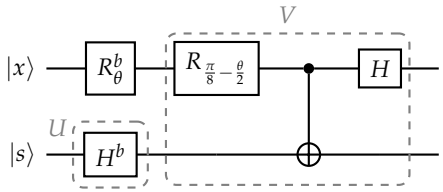


Figure 3.11: The numerically minimized $p_{\mathrm{err}}(\theta)$ for $\theta \in [0, \frac{\pi}{4}]$, the other values of θ being deduced by symmetry. Horizontal lines mark the maximum value of p_{err} attained by each curve. In the inset, the $d = 1$ curve corresponding to the best attack for no pre-shared entanglement is traced analytically.

broadcasting phase.

In the second half of the range ($\frac{\pi}{8} \leq \theta \leq \frac{\pi}{4}$), the ebit starts to help. We can find an explicit strategy for this region of the curve too: the idea is to modify the teleportation-based exact attack for $\mathrm{QPV}_{\pi/4}$ in eq. (3.52), by prepending a rotation of half the angle separating θ from $\frac{\pi}{4}$:



(3.78)

d = 3 Theorem 10 tells us that in this case we do not have exact attacks. This quirk of entangled qutrits is better detailed in the approximate context, where it becomes clear that they help lowering the error probability with respect to qubits in a small region of θ while being less useful around $\pi/4$. In this case, the piecewise function $p_{\mathrm{err}}(\theta)$ appears to be much more complex: we can identify six separate curves, with some strategies prevailing only in tiny parameter regions—namely, the ones flattening the "cusp" around $\theta/\pi \simeq 0.11$. Unfortunately, we could not find a clear analytical fit to any of the sections.

d = 4 This case corresponds to two ebits per round. We can identify five distinct regions, four of which (all but the first rising curve around $\theta = 0$) fit to an expression of the type:

$$(1-t)\sin\left(\frac{\theta}{2}-\phi\right)^2+\frac{t}{2}. \tag{3.79}$$

Around $\theta \simeq \frac{\pi}{8}$ and $\theta \simeq \frac{\pi}{4}$, where $p_{\mathrm{err}}(\theta)$ touches the x axis, we have $t = 0$ and respectively $\phi = \frac{\pi}{16}$, $\phi = \frac{\pi}{8}$. Similarly than in the $d = 2$ case, the regions of the curve around the zeros of $p_{\mathrm{err}}(\theta)$ have a matching strategy consisting of an appropriately rotated version of the corresponding exact attack. The region around $\theta \simeq 0$ is interesting, as both $d = 3$ and $d = 4$ manage to slightly beat the non-entangled PGM strategy (while $d = 2$

does not). We could not find a simple analytical formula reproducing this behavior, which might be due to numerical inaccuracy; we notice however that the difference in the value objective function is well above the observed convergence accuracy (by about three orders of magnitude), hinting at it being a real effect.

In analogy with the circuit in eq. (3.78) for the $d = 2$ case, in the region around $\frac{\pi}{8}$ we can find a matching explicit strategy by modifying the exact attack described in eq. (3.61), prepending a rotation of $\frac{1}{2}(\theta - \frac{\pi}{8})$.

$\mathbf{d = 5}$ Continuing the trend of $d = 3$, in this case too we observe a pair of maximally entangled five-level systems to yield worse attacks than two ebits in the region around $\frac{\pi}{8}$.

Comments on security

If we were to only look at the exact attacks in Table 3.2, we could think that QPV_θ is broken for a limited set of angles, and secure otherwise. But as already noted, if QPV_θ is going to be used with non-ideal (honest) provers, we need guarantees that perfect adversaries can be caught reliably. Even before looking at the results of this Section, if the conjectured pattern $\theta = \frac{k\pi}{2d}\ \forall k$ for the new "weak" angles holds for all d it is not looking well for QPV_θ security—adversaries can indeed get arbitrarily close to an attack to any angle fairly quickly in d. The analytical and numerical results above paint an even grimmer picture: From the data in Fig. 3.11 we can see that just sharing one ebit lets the adversaries get away with a small $\sin(\pi/16)^2 \simeq 3.8\%$ error probability, even around the best $\theta = \pi/8$. For two ebits per round, allowing an honest error of just $\sim 0.5\%$ is already enough to nullify any security claim across the entire parameter space!

3.5.4 $\text{QPV}_{(n)}$: a better protocol with little effort?

Can we modify QPV_θ in such a way that the experimental implementation does not suffer much, with the hopes of making it harder for adversaries to attack? We chose to analyze $\text{QPV}_{(n)}$, a simplified version the $\text{QPV}_{\text{Bloch}}$ protocol proposed by Kent and analyzed by Lau and Lo (Section 3.2.5). In

$\text{QPV}_{(n)}$, the basis in which $|x\rangle$ is encoded is chosen from n possibilities around a Bloch circle, instead of just two. The idea is that during an attack, only the adversary that receives the basis information b can adapt his quantum strategy (represented by U_b in our circuit reduction), while the other (the unitary V) has to be the same for all b. We would expect thus to find better error tolerance than what we found for QPV_θ if the adversaries are given the same entangled state.

Protocol 7 ($\text{QPV}_{(n)}$)**.** The setting is the same as QPV_θ, with the following differences:

1. The verifiers now choose a rotation R_{θ_b}, where the basis angle θ_b is picked uniformly at random from the set:

$$S_n = \left\{ \frac{b\pi}{2n}, \quad \forall b \in \mathbb{Z}_n \right\}; \tag{3.80}$$

2. They send $|\psi\rangle = R_{\theta_b} |x\rangle$ from V_1 and the basis index $b \in \mathbb{Z}_n$ from V_2;

3. The parties then follow the same actions as in QPV_θ.

Remark. The restriction to a Bloch circle instead of the entire Bloch sphere is entirely technical, and is due the possibility of a fairly substantial optimization speedup when using a real input state. We expect that a similar analysis for $\text{QPV}_{\text{Bloch}}$ could yield tighter constraints on the adversaries (which is the case for non-entangled adversaries), to which our results should be regarded as lower bounds.

The set S_n is composed of n equally-spaced angles in the range $[0, \frac{\pi}{2})$, e.g. $n = 3$ results in $\{0, \frac{\pi}{6}, \frac{\pi}{3}\}$ and $n = 7$ is $\{0, \frac{\pi}{14}, \frac{\pi}{7}, \frac{3\pi}{14}, \frac{2\pi}{7}, \frac{5\pi}{14}, \frac{3\pi}{7}\}$. As n grows, the set covers the angle range better and better. The definition of S_n is chosen so that $\text{QPV}_{(2)}$ reduces to $\text{QPV}_{\pi/4}$. An interesting feature of $\text{QPV}_{(n)}$ is that it fails to be embedded (at least, trivially) in the Gonzales and Chitambar's linear attack [GC20] which we discussed in Section 3.4.4, as b cannot be stored in a qubit anymore.

No preshared entanglement

In this case ($d = 1$) we can compute analytically the optimal error probability. Similarly to the analogous case discussed for QPV$_\theta$, all Alice can do is measure $|\psi\rangle$ at an angle $\tilde{\theta}$ as soon as it arrives, as Bob has no resource state to act on. Then they try to guess x from sharing b and Alice's measurement outcome $\tilde{x}$. The probability of error coming from a measurement of $|\psi\rangle = R_\theta |x\rangle$ in the $R_{\tilde{\theta}}$ basis is:

$$p_{\text{err}}(\tilde{\theta} \,|\, \theta_b) = \frac{1}{2}|\langle 1| R_{\tilde{\theta}} R_\theta |0\rangle|^2 + \frac{1}{2}|\langle 0| R_{\tilde{\theta}} R_\theta |1\rangle|^2 = \sin(\tilde{\theta} - \theta)^2. \tag{3.81}$$

We can immediately look at the asymptotic case for $n \to \infty$, where θ is picked from the uniform distribution over the interval $[0, \frac{\pi}{2})$. The average probability of error can be computed by the integral:

$$p_{\text{err}}(\tilde{\theta}) = \frac{2}{\pi} \int_0^{\frac{\pi}{2}} \sin(\tilde{\theta} - \theta)^2 \, \mathrm{d}\theta = \frac{1}{2} - \frac{1}{\pi} \sin(2\tilde{\theta}). \tag{3.82}$$

Minimizing over $\tilde{\theta}$, we recover the intuitive result that Alice's best measurement angle is the interval's midpoint $\pi/4$, which gives $p_{\text{err}} = (\frac{1}{2} - \frac{1}{\pi})$.

In order to get the error probability for a finite value of n, we need to average eq. (3.81) over the angles θ_b in S_n:

$$\begin{aligned} p_{\text{err}}(n, \tilde{\theta}) &= \frac{1}{n} \sum_{b=0}^{n-1} \sin\left(\tilde{\theta} - \frac{b\pi}{2n}\right)^2 \\ &= \frac{1}{n} \sum_{b=0}^{n-1} \left[\frac{1 - \cos\left(2\tilde{\theta} - \frac{b\pi}{n}\right)}{2}\right] \\ &= \frac{1}{2n} \left[n - \sum_{b=0}^{n-1} \cos\left(2\tilde{\theta} - \frac{b\pi}{n}\right)\right]. \end{aligned} \tag{3.83}$$

The angles inside the cosine are in arithmetic progression. We can evaluate the sum using the following trigonometric identity:

$$\sum_{b=0}^{n-1} \cos(x \pm by) = \frac{\sin\left(\frac{ny}{2}\right) \cos\left(x \pm \frac{(n-1)y}{2}\right)}{\sin\left(\frac{y}{2}\right)}, \tag{3.84}$$

which leads to:

$$\begin{aligned} p_{\text{err}}(n,\tilde{\theta}) &= \frac{1}{2n}\left[n - \frac{\sin\left(\frac{\pi}{2}\right)\cos\left(2\tilde{\theta} - \frac{\pi}{2} + \frac{\pi}{2n}\right)}{\sin\left(\frac{\pi}{2n}\right)}\right] \\ &= \frac{1}{2n}\left[n - \frac{\sin\left(2\tilde{\theta} + \frac{\pi}{2n}\right)}{\sin\left(\frac{\pi}{2n}\right)}\right]. \end{aligned} \tag{3.85}$$

For $n \to \infty$ we get the correct limit, matching the integral in eq. (3.82). The optimal error probability $p_{\text{err}}(n)$ for the adversaries is obtained by minimizing $p_{\text{err}}(n,\tilde{\theta})$ over $\tilde{\theta} \in [0, \frac{\pi}{2})$. Equation (3.85) reaches its minimum when $\sin(2\tilde{\theta} + \frac{\pi}{2n}) = 1$, which occurs for $\tilde{\theta} = (\frac{\pi}{4} - \frac{\pi}{4n})$. Substituting, we finally get:

$$p_{\text{err}}(n) = \min_{\tilde{\theta}\in[0,\frac{\pi}{2})} \left\{ p_{\text{err}}(n,\tilde{\theta}) \right\} = \frac{1}{2}\left[1 - \frac{1}{n}\csc\left(\frac{\pi}{2n}\right)\right]. \tag{3.86}$$

As above, for large n we get $p_{\text{err}} \to (\frac{1}{2} - \frac{1}{\pi}) \simeq 18\%$, which is therefore the highest error that $\text{QPV}_{(n)}$ can tolerate against unentangled adversaries. While the convergence is pretty fast in n, the additional implementation cost might be unjustified, when $\text{QPV}_{\pi/4}$ already achieves $p_{\text{err}} = \sin(\pi/8)^2 \simeq 14.5\%$.

Numerical results for $d > 2$

In Fig. 3.12 we collected the $p_{\text{err}}(n)$ of our optimized attacks for $d = 2, 3, 4$. Keeping in mind that the numerical minimization only provides upper bounds, assuming our results are not far from optimal we can see that the relative gain in error probability grows with d. Indeed, at $d = 4$ we already have that, for large n, the tolerable p_{err} is more than double the one provided by the best parameter for QPV_{θ}.

3.6 Summary and conclusion

We conclude the thesis with a summary of this Chapter, where we analyzed the resources needed to attack a class of position verification protocols.

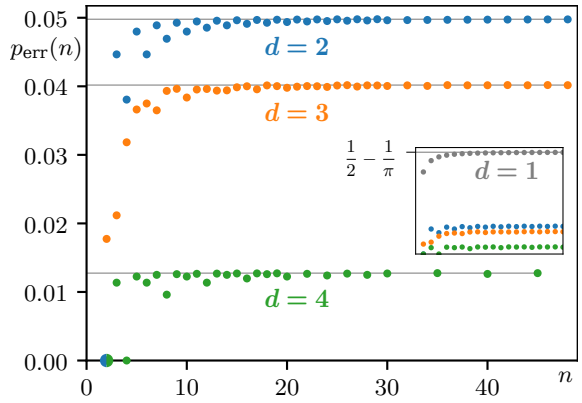


Figure 3.12: The numerically minimized $p_{\mathrm{err}}(n)$ representing attacks to the $\mathrm{QPV}_{(n)}$ protocol. For large n, the best attacks found are weaker (that is, they succeed with lower probability) than in QPV_θ, at a mild implementation cost.

Our findings

We started the Chapter with an introduction to the task of position verification, and a review of the recent literature about the topic. We then define in Section 3.3.1 the QPV_θ family, a simple class of protocols already used in previous work, due to its useful properties. After clarifying our attack model, we reformulate the protocol and the available attack strategies through the spacetime quantum circuit in Section 3.3.3. This allows us to reduce it to an equivalent, smaller regular circuit, which we use to parametrize all possible attacks in our model via a pair of unitaries (U, V).

Armed with these tools, we look for characteristics of exact attacks. We formulate the *deterministic distinguishability condition* (Definition 5), which we use to prove necessary conditions for an attack to be exact, mainly in the form of Theorem 8. Through a hypergraph-like representation of states in Hilbert space, we investigate the consequences of the DDC on exact attacks with an entangled qubit and qutrit, re-discovering with a simple proof the result in [LL11] that there exist secure protocols in this case. We try to extend the result to two ebits, which results in a reduced space of allowed graphs.

However, this is not enough to lead to a complete characterization.

In Section 3.4.2 we explain how it is possible to numerically search for exact attacks, by translating the DDC to a set of simultaneous polynomial equations. We explore the space up to $d = 12$, finding some regularities in the expression of the angles which can be broken with a d-level entangled pair. Notably, we observe a linear relation between d and the level of the Clifford hierarchy in which the protocol's rotation R_θ resides. For $d = 4$, we obtain a simple attacking circuit by reverse-engineering the numerical result. We then compare the entanglement consumption of our attacks to the strategy in [GC20].

Finally, we remove the constraint of exact attacks and show how to optimize the attackers' error probability in Section 3.5, in order to ascertain the error-tolerance properties of QPV_θ. We discuss some subtleties, as well as caveats, of the optimization method and find two ebits to be sufficient to attack any fixed angle with failure probability $< 0.5\%$. $\mathrm{QPV}_{(n)}$, a variant of QPV_θ with modest additional experimental requirements, is observed to have a slightly better error tolerance.

Future directions

For us, the main motivation for the treatment of QPV_θ was the lack of evidence that adversaries with access to near-term technology could be effective against a practical near-term protocol. Despite QPV_θ being one of the first protocols to have ever been proposed, even the specialized attacks present in the literature required the simultaneous manipulation of an impractical number of ebits for some values of θ—the closest being the protocols in [GC20]. Some interesting questions are left open, and may point to future research directions. We (only numerically) found a trend associating an exact attack of dimension d to all θ multiples of $\frac{\pi}{2d}$ up to $d \sim 12$, which suggests a generic explicit strategy continuing the trend for all d could be found. Given the similarities between the gate teleportation techniques used in measurement-based quantum computation [RHG07] and the attacks we consider, they could lead to improvements in those areas too.

Finally, the situation in quantum position verification at the moment of writing is much different than when this work started. As we noted in our literature review, 2021 has been a golden year, with strong theoretical results on lower bounds for both near-term [BCS21; LLQ21] and future [All+21; Jun+21] protocols. We hope that the relevance of position-based cryptography grows more and more in the coming years, providing a valuable addition to the already long list of reasons to accelerate the development of quantum technologies.

Bibliography

[AA13] Scott Aaronson and Alex Arkhipov. "The Computational Complexity of Linear Optics". In: *Theory of Computing* 9.4 (2013), pp. 143–252. DOI: 10.4086/toc.2013.v009a004. arXiv: 1011.3245 [quant-ph].

[AA80] Yakir Aharonov and David Z. Albert. "States and observables in relativistic quantum field theories". In: *Physical Review D* 21 (12 June 1980), pp. 3316–3324. DOI: 10.1103/PhysRevD.21.3316.

[AA81] Yakir Aharonov and David Z. Albert. "Can we make sense out of the measurement process in relativistic quantum mechanics?" In: *Physical Review D* 24 (2 July 1981), pp. 359–370. DOI: 10.1103/PhysRevD.24.359.

[AA84a] Yakir Aharonov and David Z. Albert. "Is the usual notion of time evolution adequate for quantum-mechanical systems? I". In: *Physical Review D* 29 (2 Jan. 1984), pp. 223–227. DOI: 10.1103/PhysRevD.29.223.

[AA84b] Yakir Aharonov and David Z. Albert. "Is the usual notion of time evolution adequate for quantum-mechanical systems? II. Relativistic considerations". In: *Phys. Rev. D* 29 (2 Jan. 1984), pp. 228–234. DOI: 10.1103/PhysRevD.29.228.

[AAV86] Yakir Aharonov, David Z. Albert, and Lev Vaidman. "Measurement process in relativistic quantum theory". In: *Physical Review D* 34 (6 Sept. 1986), pp. 1805–1813. DOI: 10.1103/PhysRevD.34.1805.

[ADR82] Alain Aspect, Jean Dalibard, and Gérard Roger. "Experimental Test of Bell's Inequalities Using Time-Varying Analyzers". In: *Physical Review Letters* 49 (25 Dec. 1982), pp. 1804–1807. DOI: 10.1103/PhysRevLett.49.1804.

[AG04] Scott Aaronson and Daniel Gottesman. "Improved simulation of stabilizer circuits". In: *Physical Review A* 70.5 (2004), p. 052328. DOI: 10.1103/PhysRevA.70.052328. arXiv: quant-ph/0406196.

[Ala+19] Gorjan Alagic, Gorjan Alagic, Jacob Alperin-Sheriff, Daniel Apon, David Cooper, Quynh Dang, Yi-Kai Liu, Carl Miller, Dustin Moody, Rene Peralta, et al. *Status report on the first round of the NIST post-quantum cryptography standardization process*. US Department of Commerce, National Institute of Standards and Technology, 2019. DOI: 10.6028/nist.ir.8240.

[All+21] Rene Allerstorfer, Harry Buhrman, Florian Speelman, and Philip Verduyn Lunel. *New Protocols and Ideas for Practical Quantum Position Verification*. 2021. arXiv: 2106.12911 [quant-ph].

[ARL14] Gerardo Adesso, Sammy Ragy, and Antony R. Lee. "Continuous Variable Quantum Information: Gaussian States and Beyond". In: *Open Systems & Information Dynamics* 21.01n02 (2014), p. 47. DOI: 10.1142/S1230161214400010. arXiv: 1401.4679 [quant-ph].

[Ave+21] Marco Avesani, Luca Calderaro, Matteo Schiavon, Andrea Stanco, Costatino Agnesi, Alberto Santamato, Mujtaba Zahidy, Alessia Scriminich, Giulio Foletto, Giampiero Contestabile, Marco Chiesa, Davide Rotta, Massimo Artiglia, Angela Montanaro, Manuela Romagnoli, Vito Sorianello, Francesco Vedovato, Giuseppe Vallone, and Paolo Villoresi. "Full daylight quantum-key-distribution at 1550 nm enabled by integrated silicon photonics". In: *npj Quantum Information* 7.1 (June 2021). DOI: 10.1038/s41534-021-00421-2. arXiv: 1907.10039 [quant-ph].

[Avo+18] Gildas Avoine, Muhammed Ali Bingöl, Ioana Boureanu, Srdjan čapkun, Gerhard Hancke, Süleyman Kardaş, Chong Hee Kim, Cédric Lauradoux, Benjamin Martin, Jorge Munilla, Alberto Peinado, Kasper Bonne Rasmussen, Dave Singelée, Aslan Tchamkerten, Rolando Trujillo-Rasua, and Serge Vaudenay. "Security of Distance-Bounding: A Survey". In: *ACM Comput. Surv.* 51.5 (Sept. 2018). ISSN: 0360-0300. DOI: 10.1145/3264628.

[Ban+02] Somshubhro Bandyopadhyay, P. Oscar Boykin, Vwani Roychowdhury, and Farrokh Vatan. "A new proof for the existence of mutually unbiased bases". In: *Algorithmica* 34.4 (2002), pp. 512–528. DOI: 10.1007/s00453-002-0980-7. arXiv: quant-ph/0103162.

[Bar+21] Sara Bartolucci, Patrick Birchall, Hector Bombin, Hugo Cable, Chris Dawson, Mercedes Gimeno-Segovia, Eric Johnston, Konrad Kieling, Naomi Nickerson, Mihir Pant, Fernando Pastawski, Terry Rudolph, and Chris Sparrow. *Fusion-based quantum computation*. 2021. arXiv: 2101.09310 [quant-ph].

[BB84] Charles H. Bennett and Gilles Brassard. "Quantum cryptography: Public key distribution and coin tossing". In: *IEEE International Conference on Computers, Systems and Signal Processing*. Vol. 175. Bangalore, India, 1984, p. 8. DOI: 10.1016/j.tcs.2014.05.025. arXiv: 2003.06557 [quant-ph].

[BC09] Stephen M. Barnett and Sarah Croke. "Quantum state discrimination". In: *Advances in Optics and Photonics* 1.2 (Apr. 2009), pp. 238–278. DOI: 10.1364/AOP.1.000238.

[BC94] Stefan Brands and David Chaum. "Distance-Bounding Protocols". In: *Advances in Cryptology – EUROCRYPT '93*. Ed. by Tor Helleseth. Vol. 765. Lecture Notes in Computer Science. Springer Berlin Heidelberg, 1994, pp. 344–359. ISBN: 978-3-540-57600-6. DOI: 10.1007/3-540-48285-7_30.

[BCS21] Andreas Bluhm, Matthias Christandl, and Florian Speelman. *Position-based cryptography: Single-qubit protocol secure against multi-qubit attacks*. 2021. arXiv: 2104.06301 [quant-ph].

[Bel64] J. S. Bell. "On the Einstein Podolsky Rosen paradox". In: *Physics Physique Fizika* 1 (3 Nov. 1964), pp. 195–200. DOI: 10.1103/PhysicsPhysiqueFizika.1.195.

[Ben+93] Charles H. Bennett, Gilles Brassard, Claude Crépeau, Richard Jozsa, Asher Peres, and William K. Wootters. "Teleporting an unknown quantum state via dual classical and Einstein-Podolsky-Rosen channels". In: *Physical review letters* 70.13 (1993), p. 1895. DOI: 10.1103/PhysRevLett.70.1895.

[Ben+96] Charles H. Bennett, Herbert J. Bernstein, Sandu Popescu, and Benjamin Schumacher. "Concentrating partial entanglement by local operations". In: *Physical Review A* 53 (4 Apr. 1996), pp. 2046–2052. DOI: 10.1103/PhysRevA.53.2046. arXiv: quant-ph/9511030.

[Ben80] Paul Benioff. "The computer as a physical system: A microscopic quantum mechanical Hamiltonian model of computers as represented by Turing machines". In: *Journal of statistical physics* 22.5 (1980), pp. 563–591. DOI: 10.1007/bf01011339.

[Bev+20] Michael Beverland, Earl Campbell, Mark Howard, and Vadym Kliuchnikov. "Lower bounds on the non-Clifford resources for quantum computations". In: *Quantum Science and Technology* 5.3 (June 2020), p. 035009. DOI: 10.1088/2058-9565/ab8963. arXiv: 1904.01124 [quant-ph].

[BK05] Sergey Bravyi and Alexei Kitaev. "Universal quantum computation with ideal Clifford gates and noisy ancillas". In: *Physical Review A* 71 (2 Feb. 2005), p. 022316. DOI: 10.1103/PhysRevA.71.022316. arXiv: quant-ph/0403025.

[BK11] Salman Beigi and Robert König. "Simplified instantaneous non-local quantum computation with applications to position-based cryptography". In: *New Journal of Physics* 13 (2011), p. 093036.

DOI: 10.1088/1367-2630/13/9/093036. arXiv: 1101.1065 [quant-ph].

[BK15] Joonwoo Bae and Leong-Chuan Kwek. "Quantum state discrimination and its applications". In: *Journal of Physics A: Mathematical and Theoretical* 48.8 (Jan. 2015), p. 083001. DOI: 10.1088/1751-8113/48/8/083001. arXiv: 1707.02571 [quant-ph].

[BL17] Daniel J. Bernstein and Tanja Lange. "Post-quantum cryptography". In: *Nature* 549.7671 (2017), pp. 188–194. DOI: 10.1038/nature23461.

[BM95] Samuel L. Braunstein and A. Mann. "Measurement of the Bell operator and quantum teleportation". In: *Phys. Rev. A* 51 (3 Mar. 1995). Erratum in [BM96]., R1727–R1730. DOI: 10.1103/PhysRevA.51.R1727.

[BM96] Samuel L. Braunstein and A. Mann. "Erratum: Measurement of the Bell operator and quantum teleportation [Phys. Rev. A 51, R1727 (1995)]". In: *Physical Review A* 53 (1 Jan. 1996). Erratum of [BM95], expanding the reference list., pp. 630–630. DOI: 10.1103/PhysRevA.53.630.2.

[Bog58] N. N. Bogoljubov. "On a new method in the theory of superconductivity". In: *Il Nuovo Cimento* 7.6 (Mar. 1958), pp. 794–805. DOI: 10.1007/BF02745585.

[Boh49] Niels Bohr. "Discussion with Einstein on Epistemological Problems in Atomic Physics". In: *Albert Einstein: Philosopher–Scientist*. Ed. by P. A. Schilpp. Vol. 7. The Library of Living Philosophers. Evanston, 1949, pp. 201–241. ISBN: 0-87548-286-4. DOI: 10.1016/S1876-0503(08)70379-7.

[Bou+97] Dik Bouwmeester, Jian-Wei Pan, Klaus Mattle, Manfred Eibl, Harald Weinfurter, and Anton Zeilinger. "Experimental quantum teleportation". In: *Nature* 390.6660 (1997), pp. 575–579. DOI: 10.1038/37539. arXiv: 1901.11004 [quant-ph].

[BP08] R. W. Boyd and D. Prato. *Nonlinear Optics*. Elsevier Science & Techn., 2008. 640 pp. ISBN: 9780080485966.

[Bro16] Anne Broadbent. "Popescu-Rohrlich correlations imply efficient instantaneous nonlocal quantum computation". In: *Physical Review A* 94.2 (2016), p. 022318. DOI: 10.1103/physreva.94.022318. arXiv: 1512.04930 [quant-ph].

[Buh+13] Harry Buhrman, Serge Fehr, Christian Schaffner, and Florian Speelman. "The Garden-Hose Model". In: *Proceedings of the 4th Conference on Innovations in Theoretical Computer Science*. ITCS '13. Berkeley, California, USA: Association for Computing Machinery, 2013, pp. 145–158. ISBN: 9781450318594. DOI: 10.1145/2422436.2422455. arXiv: 1109.2563 [quant-ph].

[Buh+14] Harry Buhrman, Nishanth Chandran, Serge Fehr, Ran Gelles, Vipul Goyal, Rafail Ostrovsky, and Christian Schaffner. "Position-Based Quantum Cryptography: Impossibility and Constructions". In: *SIAM Journal on Computing* 43.1 (2014), pp. 150–178. DOI: 10.1137/130913687. arXiv: 1009.2490 [quant-ph].

[BV97] Ethan Bernstein and Umesh Vazirani. "Quantum complexity theory". In: *SIAM Journal on computing* 26.5 (1997), pp. 1411–1473. DOI: 10.1145/167088.167097.

[Byr+95] Richard H. Byrd, Peihuang Lu, Jorge Nocedal, and Ciyou Zhu. "A Limited Memory Algorithm for Bound Constrained Optimization". In: *SIAM Journal on Scientific Computing* 16.5 (1995), pp. 1190–1208. DOI: 10.1137/0916069.

[Cay46] A. Cayley. "Sur quelques propriétés des déterminants gauches". In: *Journal für die reine und angewandte Mathematik* 32 (1846), pp. 119–123. DOI: 10.1017/cbo9780511703676.053.

[CD11] Bob Coecke and Ross Duncan. "Interacting quantum observables: categorical algebra and diagrammatics". In: *New Journal of Physics* 13.4 (Apr. 2011), p. 043016. DOI: 10.1088/1367-2630/13/4/043016. arXiv: 0906.4725 [quant-ph].

[Cha+09] Nishanth Chandran, Vipul Goyal, Ryan Moriarty, and Rafail Ostrovsky. "Position Based Cryptography". In: *Advances in Cryptology - CRYPTO 2009*. Ed. by Shai Halevi. Vol. 5677. Lecture

Notes in Computer Science. Springer Berlin Heidelberg, 2009, pp. 391–407. ISBN: 978-3-642-03355-1. DOI: 10.1007/978-3-642-03356-8_23. Cryptology ePrint Archive: 2009/364.

[Cha+10] Nishanth Chandran, Serge Fehr, Ran Gelles, Vipul Goyal, and Rafail Ostrovsky. *Position-Based Quantum Cryptography*. Withdrawn and replaced by [Buh+14]. 2010. arXiv: 1005.1750 [quant-ph].

[Cha+18] Ulysse Chabaud, Eleni Diamanti, Damian Markham, Elham Kashefi, and Antoine Joux. "Optimal quantum-programmable projective measurement with linear optics". In: *Physical Review A* 98.6 (Dec. 2018), p. 062318. DOI: 10.1103/PhysRevA.98.062318. arXiv: 1805.02546 [quant-ph].

[Cha21] Ulysse Chabaud. *Continuous Variable Quantum Advantages and Applications in Quantum Optics*. 2021. arXiv: 2102.05227 [quant-ph].

[Chi+14] Eric Chitambar, Debbie Leung, Laura Mančinska, Maris Ozols, and Andreas Winter. "Everything You Always Wanted to Know About LOCC (But Were Afraid to Ask)". In: *Communications in Mathematical Physics* 328.1 (May 1, 2014), pp. 303–326. ISSN: 1432-0916. DOI: 10.1007/s00220-014-1953-9. arXiv: 1210.4583 [quant-ph].

[Chr+21] Matthias Christandl, Felix Leditzky, Christian Majenz, Graeme Smith, Florian Speelman, and Michael Walter. "Asymptotic performance of port-based teleportation". In: *Communications in Mathematical Physics* 381.1 (2021), pp. 379–451. DOI: 10.1007/s00220-020-03884-0. arXiv: 1809.10751 [quant-ph].

[CL01] J. Calsamiglia and N. Lütkenhaus. "Maximum efficiency of a linear-optical Bell-state analyzer". In: *Applied Physics B* 72.1 (Jan. 2001), pp. 67–71. ISSN: 0946-2171. DOI: 10.1007/s003400000484. arXiv: quant-ph/0007058.

[CL15] Kaushik Chakraborty and Anthony Leverrier. "Practical position-based quantum cryptography". In: *Physical Review A* 92 (5 Nov. 2015), p. 052304. DOI: 10.1103/PhysRevA.92.052304. arXiv: 1507.00626 [quant-ph].

[Cla+69] John F. Clauser, Michael A. Horne, Abner Shimony, and Richard A. Holt. "Proposed Experiment to Test Local Hidden-Variable Theories". In: *Physical Review Letters* 23 (15 Oct. 1969), pp. 880–884. DOI: 10.1103/PhysRevLett.23.880.

[Cle+16] William R. Clements, Peter C. Humphreys, Benjamin J. Metcalf, W. Steven Kolthammer, and Ian A. Walmsley. "Optimal design for universal multiport interferometers". In: *Optica* 3.12 (Dec. 2016), pp. 1460–1465. DOI: 10.1364/OPTICA.3.001460.

[Cyb01] George Cybenko. "Reducing quantum computations to elementary unitary operations". In: *Computing in science & engineering* 3.2 (2001), pp. 27–32. DOI: 10.1109/5992.908999.

[Dar12] O. Darrigol. *A History of Optics from Greek Antiquity to the Nineteenth Century*. OUP Oxford, 2012. ISBN: 9780199644377.

[Deu85] David Deutsch. "Quantum theory, the Church–Turing principle and the universal quantum computer". In: *Proceedings of the Royal Society of London. A. Mathematical and Physical Sciences* 400.1818 (July 1985), pp. 97–117. DOI: 10.1098/rspa.1985.0070.

[DFW15] Frédéric Dupuis, Omar Fawzi, and Stephanie Wehner. "Entanglement sampling and applications". In: *IEEE Transactions on Information Theory* 61.2 (Feb. 2015), pp. 1093–1112. ISSN: 0018-9448. DOI: 10.1109/TIT.2014.2371464. arXiv: 1305.1316 [quant-ph].

[Die88] D. Dieks. "Overlap and distinguishability of quantum states". In: *Physics Letters A* 126.5-6 (Jan. 1988), pp. 303–306. ISSN: 0375-9601. DOI: 10.1016/0375-9601(88)90840-7.

[DiV00] David P. DiVincenzo. "The Physical Implementation of Quantum Computation". In: *Fortschritte Der Physik* 48.9-11 (Jan. 2000), pp. 771–783. DOI: 10.1002/1521-3978(200009)48:9/11<771::AID-PROP771>3.0.CO;2-E. arXiv: quant-ph/0002077 [quant-ph].

[DJ92] David Deutsch and Richard Jozsa. "Rapid solution of problems by quantum computation". In: *Proceedings of the Royal Society of London. Series A: Mathematical and Physical Sciences* 439.1907 (1992), pp. 553–558. DOI: 10.1098/rspa.1992.0167.

[DL08] Byron Drury and Peter Love. "Constructive quantum Shannon decomposition from Cartan involutions". In: *Journal of Physics A: Mathematical and Theoretical* 41.39 (Sept. 2008), p. 395305. DOI: 10.1088/1751-8113/41/39/395305. arXiv: 0806.4015 [quant-ph].

[Dor07] Andreas Dorsel. "Linear Optics". In: *The Optics Encyclopedia: Basic Foundations and Practical Applications* (2007). DOI: 10.1002/9783527600441.oe046.

[DRC17] Christian L. Degen, F. Reinhard, and Paola Cappellaro. "Quantum sensing". In: *Reviews of modern physics* 89.3 (2017), p. 035002. DOI: 10.1103/revmodphys.89.035002. arXiv: 1611.02427 [quant-ph].

[DS21] Siddhartha Das and George Siopsis. "Practically secure quantum position verification". In: *New Journal of Physics* 23.6 (June 2021), p. 063069. DOI: 10.1088/1367-2630/ac0755. arXiv: 1711.03392 [quant-ph].

[Eis+11] M. D. Eisaman, J. Fan, A. Migdall, and S. V. Polyakov. "Invited Review Article: Single-photon sources and detectors". In: *Review of Scientific Instruments* 82.7 (2011), p. 071101. DOI: 10.1063/1.3610677.

[Eke91] Artur K. Ekert. "Quantum cryptography based on Bell's theorem". In: *Physical Review Letters* 67 (6 Aug. 1991), pp. 661–663. DOI: 10.1103/PhysRevLett.67.661.

[EL14] Fabian Ewert and Peter van Loock. "3/4–Efficient Bell Measurement with Passive Linear Optics and Unentangled Ancillae". In: *Phys. Rev. Lett.* 113 (14 Sept. 2014), p. 140403. DOI: 10.1103/PhysRevLett.113.140403. arXiv: 1502.07437 [quant-ph].

[EPR35] Albert Einstein, Boris Podolsky, and Nathan Rosen. "Can Quantum Mechanical Description of Physical Reality Be Considered Complete?" In: *Physical Review* 47 (10 May 1935), pp. 777–780. DOI: 10.1103/PhysRev.47.777.

[Fey82] Richard P. Feynman. "Simulating physics with computers". In: *International Journal of Theoretical Physics* 21.6-7 (June 1982), pp. 467–488. ISSN: 1572-9575. DOI: 10.1007/BF02650179.

[FFS88] Uriel Feige, Amos Fiat, and Adi Shamir. "Zero-knowledge proofs of identity". In: *Journal of Cryptology* 1.2 (June 1988), pp. 77–94. DOI: 10.1007/BF02351717.

[Fou53] Léon Foucault. "Sur les vitesses relatives de la lumière dans l'air et dans l'eau". PhD thesis. 1853.

[Gal89] Francis Galton. *Natural Inheritance*. Macmillan, 1889. DOI: 10.5962/bhl.title.61710.

[GAN14] I. M. Georgescu, S. Ashhab, and Franco Nori. "Quantum simulation". In: *Reviews of Modern Physics* 86 (1 Mar. 2014), pp. 153–185. DOI: 10.1103/RevModPhys.86.153. arXiv: 1308.6253 [quant-ph].

[GC20] A. Gonzales and E. Chitambar. "Bounds on Instantaneous Nonlocal Quantum Computation". In: *IEEE Transactions on Information Theory* 66.5 (2020), pp. 2951–2963. DOI: 10.1109/TIT.2019.2950190. arXiv: 1810.00994 [quant-ph].

[GC99] Daniel Gottesman and Isaac L. Chuang. "Demonstrating the viability of universal quantum computation using teleportation and single-qubit operations". In: *Nature* 402.6760 (Nov. 1999), pp. 390–393. DOI: 10.1038/46503. arXiv: quant-ph/9908010.

[GEW21] Kenneth Goodenough, David Elkouss, and Stephanie Wehner. "Optimizing repeater schemes for the quantum internet". In: *Physical Review A* 103 (3 Mar. 2021), p. 032610. DOI: 10.1103/PhysRevA.103.032610. arXiv: 2006.12221 [quant-ph].

[GGM21] Juan Carlos Garcia-Escartin, Vicent Gimeno, and Julio José Moyano-Fernández. "Optimal approximation to unitary quantum operators with linear optics". In: *Quantum Information Processing* 20.9 (2021), p. 314. DOI: 10.1007/s11128-021-03254-2. arXiv: 2011.15048 [quant-ph].

[Gim15] Mercedes Gimeno-Segovia. "Towards practical linear optical quantum computing". PhD thesis. Imperial College London, 2015. DOI: 10.25560/43936.

[GLM11] Vittorio Giovannetti, Seth Lloyd, and Lorenzo Maccone. "Advances in quantum metrology". In: *Nature photonics* 5.4 (2011), pp. 222–229. DOI: 10.1038/nphoton.2011.35. arXiv: 1102.2318 [quant-ph].

[GLW13] Fei Gao, Bin Liu, and Qiao-Yan Wen. *Enhanced No-Go Theorem for Quantum Position Verification*. 2013. arXiv: 1305.4254 [quant-ph].

[Got98] Daniel Gottesman. "The Heisenberg representation of quantum computers". In: *22nd International Colloquium on Group Theoretical Methods in Physics*. July 1998. arXiv: quant-ph/9807006.

[Gri11] W. P. Grice. "Arbitrarily complete Bell-state measurement using only linear optical elements". In: *Physical Review A* 84.4 (Oct. 2011), p. 042331. ISSN: 1050-2947. DOI: 10.1103/PhysRevA.84.042331.

[GS18] David J. Griffiths and Darrell F. Schroeter. *Introduction to Quantum Mechanics*. 3rd ed. Cambridge University Press, 2018. ISBN: 9781316995433. DOI: 10.1017/9781316995433.

[GV01] Berry Groisman and Lev Vaidman. "Nonlocal variables with product-state eigenstates". In: *Journal of Physics A: Mathematical and General* 34.35 (Aug. 2001), pp. 6881–6889. DOI: 10.1088/0305-4470/34/35/313. arXiv: quant-ph/0103084.

[Har+16] Nicholas C. Harris, Darius Bunandar, Mihir Pant, Greg R. Steinbrecher, Jacob Mower, Mihika Prabhu, Tom Baehr-Jones, Michael Hochberg, and Dirk Englund. "Large-scale quantum photonic circuits in silicon:" in: *Nanophotonics* 5.3 (2016), pp. 456–468. DOI: 10.1515/nanoph-2015-0146.

[HBB99] Mark Hillery, Vladimír Bužek, and André Berthiaume. "Quantum secret sharing". In: *Physical Review A* 59 (3 Mar. 1999), pp. 1829–1834. DOI: 10.1103/PhysRevA.59.1829. arXiv: quant-ph/9806063.

[HBE21] Paul Hilaire, Edwin Barnes, and Sophia E. Economou. "Resource requirements for efficient quantum communication using all-photonic graph states generated from a few matter qubits". In: *Quantum* 5 (2021), p. 397. DOI: 10.22331/q-2021-02-15-397. arXiv: 2005.07198 [quant-ph].

[Hel69] Carl W. Helstrom. "Quantum detection and estimation theory". In: *Journal of Statistical Physics* 1.2 (June 1, 1969), pp. 231–252. ISSN: 1572-9613. DOI: 10.1007/BF01007479.

[Hen+15] B. Hensen, H. Bernien, A. E. Dréau, A. Reiserer, N. Kalb, M. S. Blok, J. Ruitenberg, R. F. L. Vermeulen, R. N. Schouten, C. Abellán, W. Amaya, V. Pruneri, M. W. Mitchell, M. Markham, D. J. Twitchen, D. Elkouss, S. Wehner, T. H. Taminiau, and R. Hanson. "Loophole-free Bell inequality violation using electron spins separated by 1.3 kilometres". In: *Nature* 526.7575 (Oct. 1, 2015), pp. 682–686. ISSN: 1476-4687. DOI: 10.1038/nature15759. arXiv: 1508.05949 [quant-ph].

[HOM87] C. K. Hong, Z. Y. Ou, and L. Mandel. "Measurement of subpicosecond time intervals between two photons by interference". In: *Physical Review Letters* 59.18 (Nov. 1987), pp. 2044–2046. ISSN: 0031-9007. DOI: 10.1103/PhysRevLett.59.2044.

[HR18] Godfrey H. Hardy and Srinivasa Ramanujan. "Asymptotic formulaæ in combinatory analysis". In: *Proceedings of the London*

Mathematical Society 2.1 (1918), pp. 75–115. DOI: 10.1112/plms/s2-17.1.75.

[HW94] Paul Hausladen and William K. Wootters. "A 'Pretty Good' Measurement for Distinguishing Quantum States". In: *Journal of Modern Optics* 41.12 (1994), pp. 2385–2390. DOI: 10.1080/09500349414552221.

[IBM] IBM. *IBM Quantum*. URL: https://quantum-computing.ibm.com/.

[IH09] Satoshi Ishizaka and Tohya Hiroshima. "Quantum teleportation scheme by selecting one of multiple output ports". In: *Physical Review A* 79 (4 Apr. 2009), p. 042306. DOI: 10.1103/PhysRevA.79.042306. arXiv: 0901.2975 [quant-ph].

[Iva87] I. D. Ivanovic. "How to differentiate between non-orthogonal states". In: *Physics Letters A* 123.6 (Aug. 1987), pp. 257–259. ISSN: 0375-9601. DOI: 10.1016/0375-9601(87)90222-2.

[JK09] Rahul Jain and Hartmut Klauck. "New Results in the Simultaneous Message Passing Model via Information Theoretic Techniques". In: *2009 24th Annual IEEE Conference on Computational Complexity*. 2009, pp. 369–378. DOI: 10.1109/CCC.2009.28.

[Jun+21] Marius Junge, Aleksander M. Kubicki, Carlos Palazuelos, and David Pérez-García. *Geometry of Banach spaces: a new route towards Position Based Cryptography*. 2021. arXiv: 2103.16357 [quant-ph].

[KC01] B. Kraus and J. I. Cirac. "Optimal creation of entanglement using a two-qubit gate". In: *Physical Review A* 63 (6 May 2001), p. 062309. DOI: 10.1103/PhysRevA.63.062309. arXiv: quant-ph/0011050.

[Ken+06] Adrian P. Kent, William J. Munro, Thimoty P. Spiller, and Raymond G. Beausoleil. "Quantum Tagging". US7075438B2. 2006.

[Ker75] John Kerr. "XL. A new relation between electricity and light: Dielectrified media birefringent". In: *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science* 50.332 (Nov. 1875), pp. 337–348. DOI: 10.1080/14786447508641302.

[KL10] P. Kok and B. W. Lovett. *Introduction to Optical Quantum Information Processing*. Cambridge University Press, 2010. ISBN: 9781139486439. DOI: 10.1017/cbo9781139193658.

[KLM01] E. Knill, R. Laflamme, and G. J. Milburn. "A scheme for efficient quantum computation with linear optics". In: *Nature* 409.6816 (Jan. 2001), pp. 46–52. ISSN: 0028-0836. DOI: 10.1038/35051009.

[KM15] Neal Koblitz and Alfred J. Menezes. "The random oracle model: a twenty-year retrospective". In: *Designs, Codes and Cryptography* 77.2 (2015), pp. 587–610. DOI: 10.1007/s10623-015-0094-2. Cryptology ePrint Archive: 2015/140.

[KMS11] Adrian Kent, William J. Munro, and Timothy P. Spiller. "Quantum tagging: Authenticating location via quantum information and relativistic signaling constraints". In: *Physical Review A* 84 (1 July 2011), p. 012326. DOI: 10.1103/PhysRevA.84.012326. arXiv: 1008.2147 [quant-ph].

[Kni04] Emanuel Knill. *Fault-tolerant postselected quantum computation: Schemes*. 2004. arXiv: quant-ph/0402171.

[Kni95] Emanuel Knill. *Approximation by quantum circuits*. 1995. arXiv: quant-ph/9508006.

[Knu74] Donald E. Knuth. "Structured Programming with Go to Statements". In: *ACM Comput. Surv.* 6.4 (Dec. 1974), pp. 261–301. ISSN: 0360-0300. DOI: 10.1145/356635.356640.

[Kok+07] Pieter Kok, W. J. Munro, Kae Nemoto, T. C. Ralph, Jonathan P. Dowling, and G. J. Milburn. "Linear optical quantum computing with photonic qubits". In: *Reviews of Modern Physics* 79.1 (Jan. 2007), pp. 135–174. ISSN: 0034-6861. DOI: 10.1103/RevModPhys.79.135. arXiv: quant-ph/0512071.

[Kra88] Dieter Kraft. "A software package for sequential quadratic programming". In: *Forschungsbericht- Deutsche Forschungs- und Versuchsanstalt fur Luft- und Raumfahrt* (1988).

[Las01] Jean B. Lasserre. "Global optimization with polynomials and the problem of moments". In: *SIAM Journal on optimization* 11.3 (2001), pp. 796–817. DOI: 10.1137/s1052623400366802.

[LCS99] N. Lütkenhaus, J. Calsamiglia, and K.-A. Suominen. "Bell measurements for teleportation". In: *Physical Review A* 59.5 (May 1999), pp. 3295–3300. ISSN: 1050-2947. DOI: 10.1103/PhysRevA.59.3295. arXiv: quant-ph/9809063.

[LGG10] Anthony Leverrier, Frédéric Grosshans, and Philippe Grangier. "Finite-size analysis of a continuous-variable quantum key distribution". In: *Physical Review A* 81 (6 June 2010), p. 062343. DOI: 10.1103/PhysRevA.81.062343. arXiv: 1005.0339 [quant-ph].

[Lia+17] Sheng-Kai Liao, Wen-Qi Cai, Wei-Yue Liu, Liang Zhang, Yang Li, Ji-Gang Ren, Juan Yin, Qi Shen, Yuan Cao, Zheng-Ping Li, Feng-Zhi Li, Xia-Wei Chen, Li-Hua Sun, Jian-Jun Jia, Jin-Cai Wu, Xiao-Jun Jiang, Jian-Feng Wang, Yong-Mei Huang, Qiang Wang, Yi-Lin Zhou, Lei Deng, Tao Xi, Lu Ma, Tai Hu, Qiang Zhang, Yu-Ao Chen, Nai-Le Liu, Xiang-Bin Wang, Zhen-Cai Zhu, Chao-Yang Lu, Rong Shu, Cheng-Zhi Peng, Jian-Yu Wang, and Jian-Wei Pan. "Satellite-to-ground quantum key distribution". In: *Nature* 549 (Aug. 2017), pp. 43–47. DOI: 10.1038/nature23655. arXiv: 1707.00542 [quant-ph].

[Lim+16] Charles Ci Wen Lim, Feihu Xu, George Siopsis, Eric Chitambar, Philip G. Evans, and Bing Qi. "Loss-tolerant quantum secure positioning with weak laser sources". In: *Physical Review A* 94 (3 Sept. 2016), p. 032315. DOI: 10.1103/PhysRevA.94.032315. arXiv: 1607.08193 [quant-ph].

[LL04] Peter van Loock and Norbert Lütkenhaus. "Simple criteria for the implementation of projective measurements with linear optics". In: *Physical Review A* 69.1 (Jan. 2004), p. 012302. ISSN:

1050-2947. DOI: 10.1103/PhysRevA.69.012302. arXiv: quant-ph/0304057.

[LL11] Hoi-Kwan Lau and Hoi-Kwong Lo. "Insecurity of position-based quantum-cryptography protocols against entanglement attacks". In: *Physical Review A* 83 (1 Jan. 2011), p. 012322. DOI: 10.1103/PhysRevA.83.012322. arXiv: 1009.2256 [quant-ph].

[LL82] L. D. Landau and E. M. Lifshitz. *Mechanics: Volume 1*. v. 1. Elsevier Science, 1982. ISBN: 9780080503479.

[LLQ21] Jiahui Liu, Qipeng Liu, and Luowen Qian. *Beating Classical Impossibility of Position Verification*. 2021. arXiv: 2109.07517 [quant-ph].

[Lom03] Chris Lomont. *Quantum circuit identities*. 2003. arXiv: quant-ph/0307111.

[Loo17] Peter van Loock. "Implementations and Protocols for the Quantum Repeater". In: *presentation at "Secure Communication via Quantum Channels"*. Bielefeld, Germany, 2017.

[LP31] Lev Landau and Rudolf Peierls. "Erweiterung des Unbestimmtheitsprinzips für die relativistische Quantentheorie". In: *Z. Physik* 69 (1931), pp. 56–69. DOI: 10.1007/BF01391513.

[Mac92] Ludwig Mach. "Über einen Interferenzrefraktor". In: *Zeitschrift für Instrumentenkunde* 12 (1892), pp. 89–93.

[Mal10a] Robert A. Malaney. "Location-dependent communications using quantum entanglement". In: *Phys. Rev. A* 81 (4 Apr. 2010), p. 042319. DOI: 10.1103/PhysRevA.81.042319. arXiv: 1003.0949 [quant-ph].

[Mal10b] Robert A. Malaney. "Quantum Location Verification in Noisy Channels". In: *Global Telecommunications Conference (GLOBECOM 2010), 2010 IEEE*. Dec. 2010, pp. 1–6. DOI: 10.1109/GLOCOM.2010.5684009. arXiv: 1004.4689 [quant-ph].

[Més] MésoLUM. *Mésocentre LUmière Matière (MésoLUM)*. URL: https://web.archive.org/web/20191203162341/http://www.mesolum.lumat.u-psud.fr/.

[Meu+17] Aaron Meurer, Christopher P. Smith, Mateusz Paprocki, Ondřej Čertík, Sergey B. Kirpichev, Matthew Rocklin, AMiT Kumar, Sergiu Ivanov, Jason K. Moore, Sartaj Singh, Thilina Rathnayake, Sean Vig, Brian E. Granger, Richard P. Muller, Francesco Bonazzi, Harsh Gupta, Shivam Vats, Fredrik Johansson, Fabian Pedregosa, Matthew J. Curry, Andy R. Terrel, Stěpán Roučka, Ashutosh Saboo, Isuru Fernando, Sumith Kulal, Robert Cimrman, and Anthony Scopatz. "SymPy: symbolic computing in Python". In: *PeerJ Computer Science* 3 (Jan. 2017), p. 27. DOI: 10.7717/peerj-cs.103.

[Mez07] Francesco Mezzadri. "How to generate random matrices from the classical compact groups". In: *Notices of the American Mathematical Society* 54.5 (2007), pp. 592–604. arXiv: math-ph/0609050.

[Mil89] Gerard J. Milburn. "Quantum optical Fredkin gate". In: *Physical Review Letters* 62.18 (1989), p. 2124. DOI: 10.1103/physrevlett.62.2124.

[MR18] Patrick Kofod Mogensen and Asbjørn Nilsen Riseth. "Optim: A mathematical optimization package for Julia". In: *Journal of Open Source Software* 3.24 (2018), p. 615. DOI: 10.21105/joss.00615.

[MT18] Alexandra E. Moylett and Peter S. Turner. "Quantum simulation of partially distinguishable boson sampling". In: *Physical Review A* 97 (6 June 2018), p. 062329. DOI: 10.1103/PhysRevA.97.062329. arXiv: 1803.03657 [quant-ph].

[MY21] Filippo Miatto and Yuan Yao. *Bolt: a blazing fast optimizer for quantum interferometers*. https://github.com/Miatto-research-group/bolt. 2021.

[NC10] Michael A. Nielsen and Isaac L. Chuang. *Quantum Computation and Quantum Information: 10th Anniversary Edition*. Cambridge University Press, 2010. ISBN: 9781107002173. DOI: 10.1017/CBO9780511976667.

[NW06] Jorge. Nocedal and Stephen J. Wright. *Numerical optimization*. Springer, 2006, p. 664. ISBN: 9780387400655.

[OG18] Andrea Olivo and Frédéric Grosshans. "Ancilla-assisted linear optical Bell measurements and their optimality". In: *Physical Review A* 98 (4 Oct. 2018), p. 042323. DOI: 10.1103/PhysRevA.98.042323. arXiv: 1806.01243 [quant-ph].

[Oli+20] Andrea Olivo, Ulysse Chabaud, André Chailloux, and Frédéric Grosshans. *Breaking simple quantum position verification protocols with little entanglement*. 2020. arXiv: 2007.15808 [quant-ph].

[Ozo09] Maris Ozols. "How to generate a random unitary matrix". 2009. URL: http://home.lu.lv/~sd20008/papers/essays/Random%20unitary%20[paper].pdf.

[Pap+13] A. Papachristodoulou, J. Anderson, G. Valmorbida, S. Prajna, P. Seiler, and P. A. Parrilo. *SOSTOOLS: Sum of squares optimization toolbox for MATLAB*. Available from http://www.eng.ox.ac.uk/control/sostools. 2013. arXiv: 1310.4716 [math.OC].

[Par00] Matteo G. A. Paris. "Optical qubit by conditional interferometry". In: *Physical Review A* 62 (3 Aug. 2000), p. 033813. DOI: 10.1103/PhysRevA.62.033813. arXiv: quant-ph/9909075.

[Par03] Pablo A. Parrilo. "Semidefinite programming relaxations for semialgebraic problems". In: *Mathematical programming* 96.2 (2003), pp. 293–320. DOI: 10.1007/s10107-003-0387-5.

[Pau40] W. Pauli. "The Connection Between Spin and Statistics". In: *Physical Review* 58 (8 Oct. 1940), pp. 716–722. DOI: 10.1103/PhysRev.58.716.

[PBR12] Matthew F. Pusey, Jonathan Barrett, and Terry Rudolph. "On the reality of the quantum state". In: *Nature Physics* 8.6 (June 1, 2012), pp. 475–478. ISSN: 1745-2481. DOI: 10.1038/nphys2309. arXiv: 1111.3328 [quant-ph].

[Ped] Samuele Pedroni. *Goals and Architecture Overview – PyPy documentation*. URL: http://doc.pypy.org/en/latest/architecture.html.

[Per88] Asher Peres. "How to differentiate between non-orthogonal states". In: *Physics Letters A* 128.1-2 (Mar. 1988), p. 19. ISSN: 0375-9601. DOI: 10.1016/0375-9601(88)91034-1.

[Pey+12] Thibault Peyronel, Ofer Firstenberg, Qi-Yu Liang, Sebastian Hofferberth, Alexey V. Gorshkov, Thomas Pohl, Mikhail D. Lukin, and Vladan Vuletić. "Quantum nonlinear optics with single photons enabled by strongly interacting atoms". In: *Nature* 488.7409 (2012), pp. 57–60. DOI: 10.1038/nature11361.

[Pir+15] Stefano Pirandola, Jens Eisert, Christian Weedbrook, Akira Furusawa, and Samuel L. Braunstein. "Advances in quantum teleportation". In: *Nature photonics* 9.10 (2015), pp. 641–652. DOI: 10.1038/nphoton.2015.154. arXiv: 1505.07831 [quant-ph].

[PR94] Sandu Popescu and Daniel Rohrlich. "Quantum nonlocality as an axiom". In: *Foundations of Physics* 24.3 (1994), pp. 379–385. DOI: 10.1007/bf02058098. arXiv: quant-ph/9508009.

[Pre18] John Preskill. "Quantum Computing in the NISQ era and beyond". In: *Quantum* 2 (Aug. 2018), p. 79. ISSN: 2521-327X. DOI: 10.22331/q-2018-08-06-79. arXiv: 1801.00862 [quant-ph].

[PS88] P. M. Pardalos and G. Schnitger. "Checking local optimality in constrained quadratic programming is NP-hard". In: *Operations Research Letters* 7.1 (1988), pp. 33–35. ISSN: 0167-6377. DOI: 10.1016/0167-6377(88)90049-1.

[Pug+17] Christopher J. Pugh, Sarah Kaiser, Jean-Philippe Bourgoin, Jeongwan Jin, Nigar Sultana, Sascha Agne, Elena Anisimova, Vadim Makarov, Eric Choi, Brendon L. Higgins, and Thomas Jennewein. "Airborne demonstration of a quantum key distribution receiver payload". In: *Quantum Science and Technology* 2.2 (June 2017), p. 024009. DOI: 10.1088/2058-9565/aa701f. arXiv: 1612.06396 [quant-ph].

[Put93] Mihai Putinar. "Positive polynomials on compact semi-algebraic sets". In: *Indiana University Mathematics Journal* 42.3 (1993), pp. 969–984.

[PV94] Sandu Popescu and Lev Vaidman. "Causality constraints on nonlocal quantum measurements". In: *Physical Review A* 49 (6 June 1994), pp. 4331–4338. DOI: 10.1103/PhysRevA.49.4331. arXiv: hep-th/9306087.

[QS15] Bing Qi and George Siopsis. "Loss-tolerant position-based quantum cryptography". In: *Physical Review A* 91.4 (2015), p. 042337. DOI: 10.1103/physreva.91.042337. arXiv: 1502.02020 [quant-ph].

[RB09] Joseph M. Renes and Jean-Christian Boileau. "Conjectured Strong Complementary Information Tradeoff". In: *Physical Review Letters* 103 (2 July 2009), p. 020402. DOI: 10.1103/PhysRevLett.103.020402. arXiv: 0806.3984 [quant-ph].

[RBB03] Robert Raussendorf, Daniel E. Browne, and Hans J. Briegel. "Measurement-based quantum computation on cluster states". In: *Physical Review A* 68 (2 Aug. 2003), p. 022312. DOI: 10.1103/PhysRevA.68.022312. arXiv: quant-ph/0301052.

[RBH01] J. M. Raimond, M. Brune, and S. Haroche. "Manipulating quantum entanglement with atoms and photons in a cavity". In: *Reviews of Modern Physics* 73 (3 Aug. 2001), pp. 565–582. DOI: 10.1103/RevModPhys.73.565.

[Rec+94] Michael Reck, Anton Zeilinger, Herbert J. Bernstein, and Philip Bertani. "Experimental realization of any discrete unitary oper-

ator". In: *Physical Review Letters* 73.1 (1994), p. 58. DOI: 10.1103/physrevlett.73.58.

[Ren+18] J. J. Renema, A. Menssen, W. R. Clements, G. Triginer, W. S. Kolthammer, and I. A. Walmsley. "Efficient Classical Algorithm for Boson Sampling with Partially Distinguishable Photons". In: *Physical Review Letters* 120 (22 May 2018), p. 220502. DOI: 10.1103/PhysRevLett.120.220502. arXiv: 1707.02793 [quant-ph].

[RG02] Terry Rudolph and Lov Grover. *A 2 rebit gate universal for quantum computing*. 2002. arXiv: quant-ph/0210187.

[RG15] Jérémy Ribeiro and Frédéric Grosshans. *A tight lower bound for the BB84-states quantum-position-verification protocol*. 2015. arXiv: 1504.07171 [quant-ph].

[RHG07] R. Raussendorf, J. Harrington, and K. Goyal. "Topological fault-tolerance in cluster state quantum computation". In: *New Journal of Physics* 9.6 (June 2007), pp. 199–199. DOI: 10.1088/1367-2630/9/6/199.

[Rob55] Herbert Robbins. "A Remark on Stirling's Formula". In: *The American Mathematical Monthly* 62.1 (Jan. 1955), p. 26. ISSN: 0002-9890. DOI: 10.2307/2308012.

[Rud17] Terry Rudolph. "Why I am optimistic about the silicon-photonic route to quantum computing". In: *APL Photonics* 2.3 (2017), p. 030901. DOI: 10.1063/1.4976737. arXiv: 1607.08535 [quant-ph].

[San+11] Nicolas Sangouard, Christoph Simon, Hugues de Riedmatten, and Nicolas Gisin. "Quantum repeaters based on atomic ensembles and linear optics". In: *Reviews of Modern Physics* 83 (1 Mar. 2011), pp. 33–80. DOI: 10.1103/RevModPhys.83.33. arXiv: 0906.2699 [quant-ph].

[ŠB20] Ivan Šupić and Joseph Bowles. "Self-testing of quantum systems: a review". In: *Quantum* 4 (2020), p. 337. DOI: 10.22331/q-2020-09-30-337. arXiv: 1904.10042 [quant-ph].

[SBM06] V. V. Shende, S. S. Bullock, and I. L. Markov. "Synthesis of quantum-logic circuits". In: *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems* 25.6 (2006), pp. 1000–1010. DOI: 10.1109/TCAD.2005.855930. arXiv: quant-ph/0406176.

[Sch11] Christian Schaffner. *Position Based Quantum Cryptography*. Personal homepage of Christian Schaffner. 2011. URL: http://homepages.cwi.nl/~schaffne/positionbasedqcrypto.php.

[SD18] Gleb Struchalin and Ivan Dyakonov. *Linopt: linear optics circuit calculator*. https://github.com/qotlabs/linopt. 2018.

[Shc13] V. S. Shchesnovich. "Asymptotic Evaluation Of Bosonic Probability Amplitudes In Linear Unitary Networks In The Case Of Large Number Of Bosons". In: *International Journal of Quantum Information* 11.05 (Aug. 2013), p. 1350045. DOI: 10.1142/S0219749913500457. arXiv: 1304.6675 [quant-ph].

[Sho94] Peter W. Shor. "Algorithms for quantum computation: discrete logarithms and factoring". In: *Proceedings 35th annual symposium on foundations of computer science*. Ieee. 1994, pp. 124–134. DOI: 10.1109/sfcs.1994.365700.

[Sho96] Peter W. Shor. "Fault-tolerant quantum computation". In: *Proceedings of 37th Conference on Foundations of Computer Science*. IEEE. 1996, pp. 56–65. DOI: 10.1109/sfcs.1996.548464.

[Sim97] Daniel R. Simon. "On the power of quantum computation". In: *SIAM journal on computing* 26.5 (1997), pp. 1474–1483. DOI: 10.1137/s0097539796298637.

[SK18] Jake A. Smith and Lev Kaplan. *Approaching near-perfect state discrimination of photonic Bell states through the use of unentangled ancilla photons*. 2018. arXiv: 1802.10527 [quant-ph].

[Smi17] Jake A. Smith. private communication. June 2017.

[Spe16] Florian Speelman. "Instantaneous Non-Local Computation of Low T-Depth Quantum Circuits". In: *11th Conference on the Theory of Quantum Computation, Communication and Cryptography (TQC 2016)*. Ed. by Anne Broadbent. Vol. 61. Leibniz International Proceedings in Informatics (LIPIcs). Dagstuhl, Germany: Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik, 2016, 9:1–9:24. ISBN: 978-3-95977-019-4. DOI: 10.4230/LIPIcs.TQC.2016.9.

[SS22] Guillaume Sangol and Maximilian Stahlberg. "PICOS: a Python interface to conic optimization solvers". In: *Journal of Open Source Software* 7.70 (Feb. 2022), p. 3915. DOI: 10.21105/joss.03915.

[SSW17] Pascale Senellart, Glenn Solomon, and Andrew White. "High-performance semiconductor quantum-dot single-photon sources". In: *Nature Nanotechnology* 12.11 (Nov. 2017), pp. 1026–1039. ISSN: 1748-3387. DOI: 10.1038/nnano.2017.218.

[Tea16] Theano Development Team. *Theano: A Python framework for fast computation of mathematical expressions*. 2016. arXiv: 1605.02688 [cs.SC].

[Tic11] Malte Christopher Tichy. "Entanglement and interference of identical particles". PhD thesis. Freiburg University, 2011.

[Tip+11] Nils Ole Tippenhauer, Christina Pöpper, Kasper Bonne Rasmussen, and Srdjan Capkun. "On the Requirements for Successful GPS Spoofing Attacks". In: *Proceedings of the 18th ACM Conference on Computer and Communications Security*. CCS '11. Chicago, Illinois, USA: Association for Computing Machinery, 2011, pp. 75–86. ISBN: 9781450309486. DOI: 10.1145/2046707.2046719.

[Tom+13] Marco Tomamichel, Serge Fehr, Jędrzej Kaniewski, and Stephanie Wehner. "A Monogamy-of-Entanglement Game With Applications to Device-Independent Quantum Cryptography". In: *New*

Journal of Physics 15 (Oct. 2013), p. 103002. DOI: 10.1088/1367-2630/15/10/103002. arXiv: 1210.4359 [quant-ph].

[Unr14] Dominique Unruh. "Quantum Position Verification in the Random Oracle Model". In: *Advances in Cryptology – CRYPTO 2014*. Ed. by Juan A. Garay and Rosario Gennaro. Vol. 8617. Lecture Notes in Computer Science. Springer Berlin Heidelberg, Aug. 2014, pp. 1–18. ISBN: 978-3-662-44380-4. DOI: 10.1007/978-3-662-44381-1_1. Cryptology ePrint Archive: 2014/118.

[Vai03] Lev Vaidman. "Instantaneous Measurement of Nonlocal Variables". In: *Phys. Rev. Lett.* 90 (1 Jan. 2003), p. 010402. DOI: 10.1103/PhysRevLett.90.010402. arXiv: quant-ph/0111124.

[Val58] J. G. Valatin. "Comments on the theory of superconductivity". In: *Il Nuovo Cimento* 7.6 (Mar. 1958), pp. 843–857. DOI: 10.1007/BF02745589.

[Val79] L. G. Valiant. "The complexity of computing the permanent". In: *Theoretical Computer Science* 8.2 (Jan. 1979), pp. 189–201. ISSN: 0304-3975. DOI: 10.1016/0304-3975(79)90044-6.

[Van13] Wim Van Dam. "Implausible consequences of superstrong nonlocality". In: *Natural Computing* 12.1 (2013), pp. 9–12. DOI: 10.1007/s11047-012-9353-6. arXiv: quant-ph/0501159.

[Vir+20] Pauli Virtanen, Ralf Gommers, Travis E. Oliphant, Matt Haberland, Tyler Reddy, David Cournapeau, Evgeni Burovski, Pearu Peterson, Warren Weckesser, Jonathan Bright, Stéfan J. van der Walt, Matthew Brett, Joshua Wilson, K. Jarrod Millman, Nikolay Mayorov, Andrew R. J. Nelson, Eric Jones, Robert Kern, Eric Larson, C. J. Carey, İlhan Polat, Yu Feng, Eric W. Moore, Jake VanderPlas, Denis Laxalde, Josef Perktold, Robert Cimrman, Ian Henriksen, E. A. Quintero, Charles R. Harris, Anne M. Archibald, Antônio H. Ribeiro, Fabian Pedregosa, Paul van Mulbregt, and SciPy 1.0 Contributors. "SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python". In: *Nature*

Methods 17 (2020), pp. 261–272. DOI: 10.1038/s41592-019-0686-2.

[VY99] Lev Vaidman and Nadav Yoran. "Methods for reliable teleportation". In: *Physical Review A* 59 (1 Jan. 1999), pp. 116–125. DOI: 10.1103/PhysRevA.59.116. arXiv: quant-ph/9808040.

[WB06] Andreas Wächter and Lorenz T. Biegler. "On the implementation of an interior-point filter line-search algorithm for large-scale nonlinear programming". In: *Mathematical programming* 106.1 (2006), pp. 25–57. DOI: 10.1007/s10107-004-0559-y.

[Wee+12] Christian Weedbrook, Stefano Pirandola, Raúl García-Patrón, Nicolas J. Cerf, Timothy C. Ralph, Jeffrey H. Shapiro, and Seth Lloyd. "Gaussian quantum information". In: *Reviews of Modern Physics* 84 (2 May 2012), pp. 621–669. DOI: 10.1103/RevModPhys.84.621. arXiv: 1110.3234 [quant-ph].

[WEH18] Stephanie Wehner, David Elkouss, and Ronald Hanson. "Quantum internet: A vision for the road ahead". In: *Science* 362.6412 (2018). DOI: 10.1126/science.aam9288.

[Wei94] Harald Weinfurter. "Experimental Bell-state analysis". In: *EPL (Europhysics Letters)* 25.8 (1994), p. 559. DOI: 10.1209/0295-5075/25/8/001.

[Wet20] John van de Wetering. *ZX-calculus for the working quantum computer scientist*. 2020. arXiv: 2012.13966 [quant-ph].

[Wie17] K. Wiesner. *The careless use of language in quantum information*. 2017. arXiv: 1705.06768 [physics.soc-ph].

[Wil13] Mark M. Wilde. *Quantum Information Theory*. Cambridge University Press, 2013. DOI: 10.1017/CBO9781139525343. arXiv: 1106.1445 [quant-ph].

[Wil82] Frank Wilczek. "Quantum Mechanics of Fractional-Spin Particles". In: *Physical Review Letters* 49 (14 Oct. 1982), pp. 957–959. DOI: 10.1103/PhysRevLett.49.957.

[WSM19] Eyuri Wakakuwa, Akihito Soeda, and Mio Murao. "Complexity of Causal Order Structure in Distributed Quantum Information Processing: More Rounds of Classical Communication Reduce Entanglement Cost". In: *Physical Review Letters* 122 (19 May 2019), p. 190502. DOI: 10.1103/PhysRevLett.122.190502. arXiv: 1810.08447 [quant-ph].

[Xan] Xanadu. *Xanadu Quantum Cloud*. URL: https://xanadu.ai/cloud.

[ZCC08] Bei Zeng, Xie Chen, and Isaac L. Chuang. "Semi-Clifford operations, structure of C k hierarchy, and gate complexity for fault-tolerant quantum computation". In: *Physical Review A* 77.4 (2008), p. 042313. DOI: 10.1103/PhysRevA.77.042313. arXiv: 0712.2084 [quant-ph].

[Zeh91] Ludwig Zehnder. "Ein neuer Interferenzrefraktor". In: *Zeitschrift für Instrumentenkunde* 11 (1891), pp. 275–285.

[Zho+21] Han-Sen Zhong, Yu-Hao Deng, Jian Qin, Hui Wang, Ming-Cheng Chen, Li-Chao Peng, Yi-Han Luo, Dian Wu, Si-Qiu Gong, Hao Su, Yi Hu, Peng Hu, Xiao-Yan Yang, Wei-Jun Zhang, Hao Li, Yuxuan Li, Xiao Jiang, Lin Gan, Guangwen Yang, Lixing You, Zhen Wang, Li Li, Nai-Le Liu, Jelmer J. Renema, Chao-Yang Lu, and Jian-Wei Pan. "Phase-Programmable Gaussian Boson Sampling Using Stimulated Squeezed Light". In: *Physical Review Letters* 127 (18 Oct. 2021), p. 180502. DOI: 10.1103/PhysRevLett.127.180502. arXiv: 2106.15534 [quant-ph].

[Zhu17] Xiaojing Zhu. "A Riemannian conjugate gradient method for optimization on the Stiefel manifold". In: *Computational optimization and Applications* 67.1 (2017), pp. 73–110. DOI: 10.1007/s10589-016-9883-4.

[ZL13] Hussain A. Zaidi and Peter van Loock. "Beating the One-Half Limit of Ancilla-Free Linear Optics Bell Measurements". In: *Physical Review Letters* 110.26 (June 2013), p. 260501. ISSN: 0031-

9007. DOI: 10.1103/PhysRevLett.110.260501. arXiv: 1301.2749 [quant-ph].