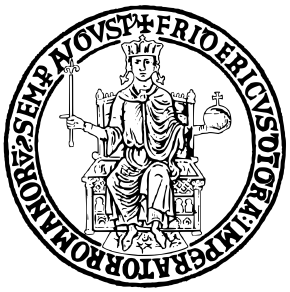

Università degli Studi di Napoli Federico II

DOTTORATO DI RICERCA IN QUANTUM TECHNOLOGIES

Ciclo XXXIV

Coordinatore: Prof. Dr. Francesco Tafuri

Quantum Resources in Quantum Technologies: Identification, Verification, and Application

---

Settore Scientifico Disciplinare FIS/03

Dottorando
Valentin Gebhart

Tutori
Dr. Augusto Smerzi
Dr. Luca Pezzè

Anni Accademici 2018/2021

Abstract

Quantum technologies employ the laws and phenomena of quantum mechanics to address different tasks in various technological fields such as, e.g., computation, cryptography, or sensing. Specific quantum features empower quantum technologies to achieve technological advantages over classical technologies. These features are often called quantum resources. In this thesis, we study quantum resources from different viewpoints. We delve into the identification of the central quantum resources in quantum computation, we address the verification of specific quantum resources from experimental data, and we apply quantum resources to develop new quantum technologies. The results of this thesis are published in the papers [1, 2, 3, 4, 5].

The crucial quantum resources that enable a quantum advantage of quantum computers over classical ones for specific computational problems have been intriguing scientists since the introduction of quantum computation. We focus on a specific quantum algorithm, Grover's algorithm, and examine quantum resources that are necessary for its computational advantage over classical technologies. We find that the maximal trace speed of the quantum state during the algorithm can be used to bound the quantum advantage in different noisy versions of Grover's algorithm. The trace speed can be interpreted as a measure of the quantum resources of coherence or entanglement.

Verifying the presence of a certain quantum resource in an experimental setup is generally a difficult problem and requires specific techniques and strategies that depend on the quantum resource in question. For instance, the verification of the quantum resource of Bell nonlocality requires violations of Bell inequalities. These violations can be forged by means of the selection bias if the observed data are postselected collaboratively by the different experimental parties. We prove conditions for partially-collaborative postselection strategies that are valid for the verification of genuine multipartite nonlocality.

In the field of continuous-variable quantum technologies, a central quantum resource is nonclassicality. The verification of nonclassicality is generally cumbersome and requires large amounts of experimental data. We develop and train neural-network-based nonclassicality indicators that predict nonclassicality directly from small amounts of experimental data produced in different standard quantum-optical measurement schemes. The method is applied to real experimental data from homodyne measurements.

Finally, we employ the toolbox of quantum mechanics to develop a quantum algorithm that performs Bayesian multiphase estimation at the optimal precision scaling, where we take into account all physical resources that are used in the estimation protocol. The algorithm can be implemented in state-of-the-art quantum optical architectures and represents a potential subroutine for technologies in quantum sensing and quantum computation.

Acknowledgements

First of all, I would like to thank Augusto Smerzi for giving me the opportunity to realize my PhD project in his group in Florence, for his supervision and for many fascinating discussions about various topics of quantum information. I very much enjoyed the broad spectrum of our digressions and I feel that I have learned many new aspects and mysteries of quantum physics thanks to them.

Many thanks to Luca Pezzè for his co-supervision of my projects, for taking his time for many interesting discussions and for his steady availability when I was stuck.

A big thanks also to Martin Bohmann for several interesting discussions and his support that have eventually led to a very fruitful collaboration. Also thank you for proofreading preliminary versions of my thesis.

Thanks to Thomas Körber for delving into the methods of machine learning for quantum physics with me, even though eventually we could not gain any publishable insights.

I would like to thank all members of the Quantum Science and Technology group and my fellow office mates in Arcetri for brightening up the daily office life and for giving me beautiful memories: Martin, Eliza, Riccardo, Karsten, Laura, Sam, Elise, Beatrice, Vasilis, Robin, Ghofran, Arthur, and Marco.

Thanks to Natalia and Lucky for their company and support during the difficult times of lockdown.

Lastly, I want to thank my parents Johanna and Urban for supporting me during all of my studies and for offering me a foothold I can always rely on.

List of publications

[1] V. Gebhart, L. Pezzè and A. Smerzi. *Quantifying computational advantage of Grover's algorithm with the trace speed*. Sci. Rep. **11**, 1 (2021).

[2] V. Gebhart, L. Pezzè and A. Smerzi. *Genuine Multipartite Nonlocality with Causal-Diagram Postselection*. Phys. Rev. Lett **127** 140401 (2021).

[3] V. Gebhart, M. Bohmann, K. Weiher, N. Biagi, A. Zavatta, M. Bellini and E. Agudelo. *Identifying nonclassicality from experimental data using artificial neural networks*. Phys. Rev. Research **3**, 023229 (2021).

[4] V. Gebhart and M. Bohmann. *Neural-network approach for identifying nonclassicality from click-counting data*. Phys. Rev. Research **2**, 023150 (2020).

[5] V. Gebhart, A. Smerzi and L. Pezzè. *Bayesian Quantum Multiphase Estimation Algorithm*. Phys. Rev. Applied **16**, 014035 (2021).

Contents

Abbreviations

BQMEA	Bayesian quantum multiphase estimation algorithm
BS	beam splitter
CHSH	Clauser-Horne-Shimony-Holt
CV	continuous-variable
DAG	directed acyclic graph
DQC1	deterministic quantum computation with one clean qubit
EPR	Einstein-Podolsky-Rosen
FD	frequency doubled
GHZ	Greenberger-Horne-Zeilinger
GMN	genuine multipartite nonlocality
HD	homodyne detection
LHV	local hidden variable
LO	local oscillator
LOCC	local operations and classical communication
LOSR	local operations and shared randomness
ML	machine learning
NN	neural network
NPATS	n-photon-added thermal state
OPA	optical parametric amplifier
PDC	parametric down-conversion
POVM	positive operator-valued measure
QFI	quantum Fisher information
QKD	quantum key distribution
QPE	quantum phase estimation
QS	quantum statistical speed
RFPE	rejection-filtering phase estimation
RV	random variable
SPACS	single-photon-added coherent state
TS	trace speed
YS	Yurke-Stoler

Chapter 1

Introduction

Contents

1.1 Quantum technologies

The development of the intriguing and counterintuitive concepts of quantum physics and its applications can be coarsely divided into two sections [6]. The *first quantum revolution*, describing the insights and applications of quantum physics since its discovery until the end of the twentieth century, places research in a rather passive role: the ideas of quantum physics were mostly used to explain phenomena that were observed, such as, e.g., black body radiation, the photoelectric effect, the periodic table of elements, wave-particle duality, etc. The thereby-understood effects could then be used to obtain central technological advances such as the transistor or the laser.

Currently, research is experiencing the *second quantum revolution* that is characterised by an active modeling or engineering of systems that are described by quantum mechanics. This motivated the widely used term *quantum technologies*, i.e., technologies that explicitly control and use quantum properties at the elementary level in order to achieve a certain technological task. The sparkling insight of this field was the realization that, by engineering and manipulating quantum systems, one can achieve an advantage over any classical technology for resolving specific technological tasks, the so-called *quantum advantage*.

The initial step toward quantum technologies was the realization by Einstein, Podolsky and Rosen in 1935 that a perfectly prepared quantum system shows certain nonlocal correlations that escape any possible classical and local description [7], an observation that is since called the EPR paradox. This paradox was formalized and made testable by the famous Bell inequalities introduced in 1964 [8, 9]. The first technological advantage of these nonlocal phenomena was realized about two decades later, with the beginning of quantum cryptography [10, 11, 12]. Around the same time, quantum technologies were conjectured [13] (and later shown) to also yield computational advantages [14, 15], culminating in Shor's algorithm in 1994 [16] that gave the field of quantum technologies its arguably largest boost. Independently, an increased performance of precision measurements by the use of quantum features (squeezed light) was proposed in 1981 [17].

Today, quantum technologies can be roughly divided into four subfields.

- *Quantum cryptography* employs quantum systems to share unbreakable key distributions (see, e.g., Ref. [18]) between distant parties.

- *Quantum computation* replaces classical bits by quantum bits (qubits) to perform certain computations faster than classical computers (see, e.g., Ref. [19]). Depending on the algorithm, the speed-up can be polynomial or even exponential.

- *Quantum metrology* builds on quantum systems as sensors to achieve a measurement precision that is inaccessible when using classical measurement techniques (see, e.g., Ref. [20, 21]).

- *Quantum simulation* enables the classically inaccessible simulation of complex quantum systems by engineering and controlling other quantum systems (see, e.g., Ref. [22]).

Various questions concern the field of quantum technologies and quantum advantages since their introduction. The central and most practical challenge is how to use the toolbox of quantum mechanics and its resources to develop new quantum technologies that show advantages over classical technologies for specific tasks. Furthermore, ever since quantum advantages were found, there has been a large and on-going discussion among scientists about where these advantages stem from. What characteristics of quantum systems that are inaccessible for classical technologies, make the former more powerful for certain technological tasks? These characteristics are often referred to as *quantum resources*. Finally, from an experimental or engineering point of view, once having defined an interesting quantum resource, how can one detect and certify this resource from real experimental data of measurements of a quantum system?

In this thesis, we will address and partially contribute to an answer (for specific problem settings) to the above questions. The following general introduction serves as a rough

orientation for the basic ideas and conceptual difficulties that underlie the field of quantum resources.

1.2 Origins of quantum advantage

Quantum systems behave differently than classical ones. From the mathematical point of view, the possible states of a quantum system are described as unit-trace density operators acting on complex Hilbert spaces [19], a description that is in contrast to the phase space description of classical mechanics. As being described by different fundamental building blocks, quite naturally, technologies that rely on quantum and classical mechanics will be of dissimilar shape. However, it came as a surprise for many researchers when specific technological tasks were found that can be performed more efficiently with quantum systems, and, still today, there exists a large community of "sceptics" that question the usefulness of (some) quantum technologies. The big question arose about which features of quantum mechanical systems and laws enable such an advantage. What is its resource? Some commonly stated candidates for these resources are, to name a few, the exponential size of the Hilbert space, or the phenomena of superposition and interference. The most famous quantum feature that is widely thought to be responsible for the quantum advantage, and that constantly amazes researchers is the quantum phenomenon named *entanglement*.

1.2.1 Multipartite entanglement

Imagine a quantum system that consists of two subsystems. Recall that according to the laws of quantum mechanics, the total Hilbert space $\mathcal{H}$ of the system is the tensor product of the subsystem Hilbert spaces $\mathcal{H}_i$, $\mathcal{H} = \mathcal{H}_1 \otimes \mathcal{H}_2$. Pure quantum states are described by unit vectors $|\psi\rangle \in \mathcal{H}$, and general (mixed) states by (unit-trace Hermitian) density operators $\rho : \mathcal{H} \to \mathcal{H}$ (see, e.g., Ref. [19] for an extensive introduction to the axioms and definitions of quantum mechanics). A pure quantum state $|\psi\rangle$ is called separable if it can be written as $|\psi\rangle = |\psi_1\rangle \otimes |\psi_2\rangle$ with $|\psi_i\rangle \in \mathcal{H}_i$. In other words, a pure state is separable if and only if there are no correlations between the subsystems. If $|\psi\rangle$ is not separable, it is called entangled. A general mixed entangled state shows correlations between the subsystems that cannot be described by classical mixtures of separable states.

In the case of a quantum system consisting of n subsystems, entangled states show further structure. A general pure state $|\psi\rangle$ can be written as

$$|\psi\rangle = \bigotimes_j |\psi_j\rangle \,, \tag{1.1}$$

where for each j, $|\psi_j\rangle$ is a pure state of a group of subsystems. If $|\psi\rangle$ can be written according to Eq. (1.1) with a maximal group size of k, but not with a maximal group size of $k-1$,

it is called k-partite entangled (for $k = 1$, it is called separable). A general state ρ is called k-partite entangled if it can be written as a mixture $\rho = \sum_l p_l |\psi_l\rangle \langle\psi_l|$, where each $|\psi_l\rangle$ is at most k-partite entangled, but not as a mixture of $(k-1)$-partite entangled states. Here, $|\psi\rangle \langle\psi|$ denotes the density operator corresponding to the pure state $|\psi\rangle$. An n-partite state that is n-partite entanglement is also called genuinely multipartite entangled. For a conclusive review of entanglement we refer to the literature, see, e.g., Ref. [23].

Entanglement represents the key difference between pure quantum and pure classical systems in terms of the description of composite systems. In fact, while the degrees of freedom (the number of parameters needed to uniquely describe a pure physical state) of two classical systems equals the sum of the degrees of freedom of each subsystem, in composite quantum systems, the degrees of freedom multiply. This results in the terminology of an exponential size of the Hilbert space (as a function of the number of subsystems). The overhead possibilities for pure quantum composite systems are precisely the entangled states.

Many key developments of quantum technologies rely on (multipartite) entanglement. For instance, in quantum communication tasks, one needs an entangled state to violate Bell inequalities [24], and to perform superdense coding [25] or quantum teleportation [26]. All of the above represent basic building blocks for current advanced quantum communication technologies. In quantum metrology, multipartite entanglement is known to be the key resource to surpass the classical shot-noise limit [21, 27, 28, 29, 30], i.e., to yield a quantum advantage with respect to the possible precision of a sensing protocol. Many quantum algorithms rely on multipartite entanglement [19] and, in the specific case of an exponential classical complexity of pure state quantum computations, multipartite entanglement was proven to be necessary [31].

Yet, the general role of entanglement in quantum advantage is still unclear [32]. In particular, in quantum computation, apparent quantum advantages have been found that do not seem to rely on entanglement [15, 33]. We will discuss these observations in detail in Sec. 2.1. A crucial obstacle in the resolution of the role of entanglement and other quantum resources are the varying definitions of quantum advantage: in short, defining a quantum advantage in different ways can lead to different answers about the origin and the resources of the advantage, a problem that we will depict in the following.

1.2.2 Different contexts of quantum advantage

In order to meaningfully talk about a quantum advantage, it is necessary to precisely define its context, i.e., the precise technological task that we want to accomplish together with all its boundary conditions. Changing these contexts might transform a quantum technology from yielding a quantum advantage to not showing any benefits compared to a classical technology. Furthermore, since advantageous and optimal quantum technologies change

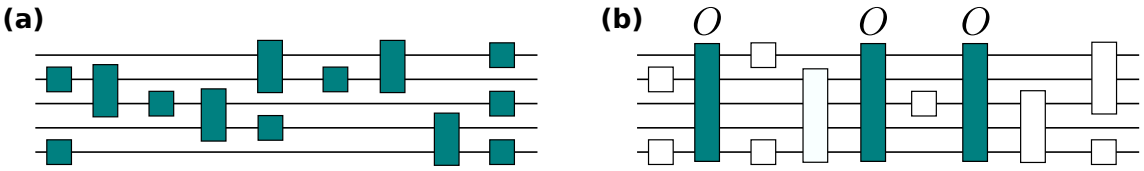

Figure 1.1: (a) To specify the gate complexity of a quantum algorithm, one has to take account of all basic quantum gates that together implement the algorithm. (b) In query complexity, one simply counts the number of applications of an oracle unitary O, while additional gates of the quantum algorithm are ignored.

when we vary between these contexts, also the key quantum features of these technologies might differ.

The best known example of this ambiguity is the definition of the quantum advantage in quantum computation. In theoretical quantum computation, there are two main definitions of the complexity of a certain (quantum) algorithm [19]. On the one hand, *gate complexity* is a measure of how many "basic" gates have to applied to implement a quantum algorithm (cf. Fig. 1.1(a)). The set of basic gates also represents some freedom to choose, but commonly it is some universal gate set of quantum computation consisting of single- and two-qubit gates. The other complexity is *query complexity* (also called oracle complexity). Here, the complexity is simply defined as the number of applications of one specific gate, the oracle, even though other gates are needed to implement the complete algorithm (cf. Fig. 1.1(b)). The number and complexity of the remaining gates as well as the complexity of the oracle gate are not taken into account (the oracle is said to act as a "black box"). The two complexities generally differ significantly. For instance, Grover's algorithm yields an advantage in terms of query complexity [34]. Furthermore, while the importance of Shor's algorithm [16] is due to its exponential advantage with respect to gate complexity, it also yields an exponential advantage in query complexity[1]. The crucial insight in Shor's algorithm is that the number of additional gates as well as the number of gates to implement the oracle is so small that an exponential advantage persists also in terms of gate complexity. With respect to gate complexity, entanglement is generally believed to play the key role of the quantum advantage. In contrast, as we discuss in detail in Sec. 2.1, taking only into account query complexity and not specifying any other boundary conditions, entanglement cannot represent the resource for quantum advantage because all quantum algorithms can be equivalently implemented in a single multilevel system.

A second field in which different definitions of quantum advantage led to disputes is quantum metrology. Let us consider optical metrology as an example. The standard definition of the technological task is the following: we are given an optical element that induces an unknown phase shift θ that we want to measure. To measure the phase shift

[1] In particular, Shor's algorithm to factor a number N relies on the quantum algorithm for period finding that is successful for $\mathcal{O}(1)$ oracle calls [19]. The best known classical algorithm for period finding requires $\mathcal{O}(N^{1/4})$ oracle calls [35].

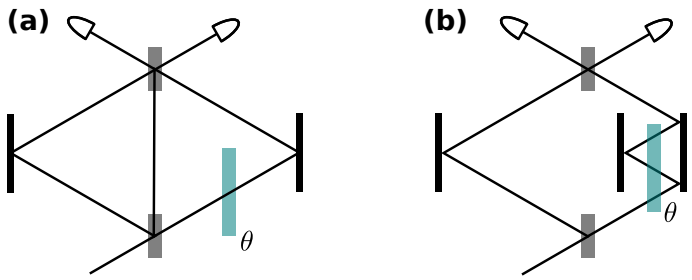

Figure 1.2: (a) In the standard setting of optical quantum metrology, N photons pass through an interferometer that imprints an unknown phase shift θ. The optimal achievable phase sensitivity requires multiphoton entanglement. (b) Counting only the number of times the phase shift is penetrated by a photon, one can achieve the optimal sensitivity also by multipass interferometers and single photons.

we have to place the element in one arm of an interferometer (cf. Fig. 1.2(a)). Now we are given N photons that can pass through the interferometer. Which N-photon state in the interferometer results in the highest sensitivity to the phase shift? In this case, the optimal sensitivity (the so-called Heisenberg limit [36, 37]) is shown to require multipartite entanglement [27]. However, instead of assuming that N photons pass through the interferometer once, we can just assume that the unknown optical element can be penetrated N times by a photon without further constraints. Then one can observe that the sensitivity of the above strategy using multipartite entanglement can also be achieved by using just a single photon that is traversing the phase shift multiple times [37, 38] (cf. Fig. 1.2(b)). Since now we are using only a single photon, multipartite entanglement cannot be the resource of the advantage. Entanglement is however still present [39] because the highly nonclassical single photon results in an entangled state in the interferometer (see also Sec. 4.1). A different quantum resource that is considered in this implementation is coherence. We will revisit this formal equivalence between different these settings of quantum metrology in Ch. 5, where we also add a third equivalent approach that is used in the quantum phase estimation algorithm [40]. The different quantum resources will be discussed in Sec. 5.1.4.

To summarise, different definitions of a technological task (and its boundary conditions) lead to different optimal strategies and quantum advantages, and possibly alter the quantum resource that is necessary for the quantum advantage.

1.2.3 Quantum resources

After having specified the precise context of a quantum advantage, one can ask the question of the quantum resource that is necessary for the advantage. In many quantum technologies, several quantum features are present, such that a unique identification of the necessary resource is not straightforward [32]. Furthermore, resources can be largely dependent such as, for instance, the quantum features of entanglement and quantum discord [41, 42, 43] that generally represent different quantum correlations but coincide for the set

of pure states. Finally, quantum resources can often be converted into each other. An example is the conversion between nonclassicality and entanglement in continuous-variable systems [44, 45]: a single-mode nonclassical state (that is not entangled) generates mode entanglement after getting mixed with vacuum in a balanced beam splitter.

In contrast, as mentioned above, a clear identification of the quantum resource was possible in the field of quantum metrology and precision measurements when considering the standard setting of single-pass interferometers. According to the quantum Cramér-Rao bound [46, 47], the ultimate precision limit of a measurement is closely connected to the quantum Fisher information (QFI) [48, 49], see Ch. 5 and in particular Eq. (5.1) for more details. The QFI is one among many quantum statistical speeds that quantify the susceptibility of quantum states to parameter displacements [50], and will be introduced in detail in Sec. 2.1.1. This quantum statistical speed, that is central to quantum-enhanced precision measurements, serves furthermore as a detection of multipartite entanglement [27, 28, 29]. Therefore, the necessary quantum resource in quantum metrology (in the standard setting, see Sec. 1.2.2) is multipartite entanglement. In Ch. 2, we will discuss the role of a different quantum statistical speed, the trace speed, in Grover's search algorithm [34].

Resource theories

In recent years, an operational approach to quantum resources has been developed [51]. Here, the main idea is that one defines some subset of all possible quantum operations as the set of free operations. For instance, the free operations can represent all available quantum operations implementable in a specific experimental setup. All states that can be generated using free operations from some fixed initial state are called free states. A state that is not part of the free states is thus called resourceful. Furthermore, the amount of resourcefulness can be measured by a non-negative and convex function on the set of states, that must vanish for free states and not increase under the free operations [51, 52].

For the traditional operational resource theory of entanglement, the set of free operations consists of local operations and classical communication (LOCC) [23, 53]. With this choice, the set of free states coincides with the set of separable states, such that all entangled states are resourceful. Different measures of entanglement can be chosen that are compatible with this resource theory, for instance the relative entropy distance [51]. We note that recently it was argued that in resource theories of entanglement, the set of free operations should be chosen depending on the context [54]. For instance, in the standard Bell scenario that we will further discuss in Ch. 3, a natural candidate for free operations are local operations and shared randomness (LOSR). This is because, in the standard scenario, several measurements are performed at a spacelike distance, such that the laws of special relativity forbid any communication during the measurement procedure (the so-

called no-signalling principle). While these different resource theories provide the same classification of states into separable and entangled states, notions of multipartite entanglement (and nonlocality) differ [54], see Sec. 3.1.2 for more details. Resource theories have also been widely applied to other quantum features such as coherence [55, 56, 57], asymmetry [58, 59], purity [60], nonclassicality [61], and more. For an overview of different resource theories, we refer to Ref. [62].

1.3 Detection of quantum resources

After having identified a quantum resource of interest, the question arises of how to certify that this resource is present in an experiment. In general, the detection of different quantum resources may require completely dissimilar methods. Often, due to the dependence of different resources, detecting one resource directly implies the presence of a second one. The prime example is the classification of bipartite quantum states and correlations that we want to explain in the following.

Consider many copies of a bipartite quantum state ρ that are shared between two parties, Alice and Bob. By measuring their respective parts of the states and processing the results, Alice and Bob want to decide what kind of quantum states or correlations they share. Depending on the quantum resource they want to detect, they need to follow different strategies. For instance, if they want to decide if the state is separable or entangled, one possibility is to perform full state tomography: they perform a series of measurements that correspond to a basis for all possible states, and after collecting sufficient statistics, they can estimate the state ρ. Now they can test different conditions for entanglement on their estimated ρ. As an example, in the case of small local dimensions of the Hilbert spaces ($d_1 = d_2 = 2$ and $d_1 = 2$, $d_2 = 3$), the positive partial transpose criterion is a necessary and sufficient criterion for entanglement [63]. For larger Hilbert space dimensions or more than two parties, there are no efficient general conditions to decide if the state is separable or entangled. It was even shown that the decision problem of whether a state is separable or entangled is NP-hard if the state is close to the boundary between separable and entangled states [64]. Alternatively, theoretical conditions that require full quantum state tomography may be out of reach for experimental tests, such that experimentally-accessible entanglement detection methods have to be developed, see, e.g., Refs. [65, 66, 67]. Commonly, conditions are applied that, when fulfilled, allow the experimenters to conclude that the state is entangled. These conditions are called entanglement witnesses [23, 68]. As mentioned above, entanglement is central to many quantum technologies, such that the detection of its presence is crucial to decide if an experimental setup can produce quantum-technologically useful states.

A second quantum feature that the parties might want to demonstrate is Bell nonlo-

cality[2] [8, 9]. Here, Alice and Bob locally measure their subsystems, where in each round, they have to choose their measurement randomly from a few measurement options. The demonstration of nonlocality can then be achieved by a violation of Bell inequalities [24] and will be discussed in more detail in Ch. 3. Nonlocality represents the central quantum resource in security and cryptography [12, 18].

As a last example, there is a third class of quantum correlations that goes back to an idea of Schrödinger in 1936 [70] (as a response to the EPR paradox [7]). In short, by Alice's choice of measurement she can steer the ensemble of possible states at Bob's side. This phenomenon is called EPR steering. In contrast to nonlocality, there is an asymmetry between the parties in the sense of who steers whom. An exemplary problem setting that is solved by steerable states is the task that Alice has to convince Bob that they share entanglement, while Bob distrusts her. Therefore, Bob receives classical information from Alice about her measurement result, and can then measure his state to check certain steering conditions [71, 72]. EPR steering was also identified as the necessary resource in several quantum technologies such as, e.g., one-sided quantum key distribution (QKD) [73].

The different classes of quantum states that provide the above quantum resources (i.e., entanglement, nonlocality and steerability) have simple inclusion relations [71]. We exemplify this by comparing separable states with local states (i.e., states that cannot give rise to nonlocal correlations). A bipartite state ρ is called local if for any local measurement performed by Alice and Bob, the resulting correlations can be described by a so-called local hidden variable (LHV) model, see Ch. 3 for more details. Formally, ρ is local if, for any local measurement, there exist probability distributions P_λ, $P_{a|x\lambda}$ and $P_{b|y\lambda}$ such that

$$P_{ab|xy} = \mathrm{tr}[E_a \otimes F_b \rho] = \sum_\lambda P_\lambda P_{a|x\lambda} P_{b|y\lambda}. \tag{1.2}$$

Here, a (b) and x (y) label Alice's (Bob's) measurement outcome and setting, respectively, and E_a (F_b) is a positive semi-definite operator ($\sum_a E_a = \mathbb{1}$) on the Alice's (Bob's) Hilbert space corresponding to the measurement choice x (y) and outcome a (b), see, e.g., Ref. [19]. The discrete variable λ is called a local hidden variable. For a separable state $\rho_s = \sum_l p_l \sigma_l \otimes \sigma'_l$, one immediately obtains

$$P_{ab|xy} = \sum_l p_l \,\mathrm{tr}[E_a \sigma_l]\,\mathrm{tr}[F_b \sigma'_l] \tag{1.3}$$

for any measurement operators E_a and F_b, which is of the form of Eq. (1.2). Thus, we see that the set of separable states is included in the set of local states.

The converse is not true: there exist entangled states that cannot give rise to nonlocal correlations for any local measurements [74]. Thus, the separable states form a strict subset of the local states. Furthermore, similar considerations show that that the set of

[2] For simplicity, throughout this thesis, we will use the term nonlocality for this phenomenon. The right designation and interpretation is still highly disputed, see, e.g., Ref. [69].

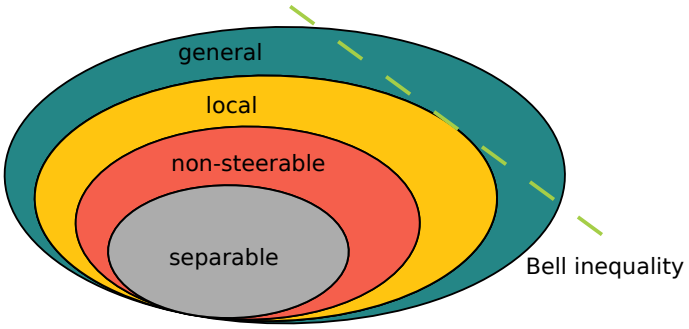

Figure 1.3: Sketch of different convex classes of bipartite quantum states. The set of separable states (grey) is included in the set of non-steerable states (red), i.e., states that cannot be used for EPR steering. The set of non-steerable states is included in the set of local states (yellow), which are all the states for which the correlations for any local measurement can be described by a local-hidden-variable model, cf. Eq. (1.2). All classes of states are included in the set of general bipartite quantum states (blue). We also indicate that the nonlocality of a state (i.e., the nonlocality of the state's measurement correlations for some local measurement) is detected by a Bell inequality.

non-steerable states includes the separable states but is included in the local states[3], where, again, the inclusions are strict [71].

All inclusion relations are sketched in Fig. 1.3, where we also indicate that the different classes of "resourseless" states (i.e., separable, non-steerable, and local) are strict subsets of the set of general bipartite quantum states. States outside the local set are detected when the resulting measurement correlations violate Bell inequalities, as we will discuss in detail in Ch. 3. Finally, all different sets are convex because, for any two states that are separable, non-steerable, or local, the mixture of the states has the same feature as well.

The inclusion relations between these different classes of quantum resources highlight that, while their demonstration might be possible using completely different strategies, demonstrating the presence of one quantum resource often directly demonstrates the presence of a second one. As seen above, entanglement is necessary for steerability, and entanglement and steerability are necessary for states that can show nonlocality. Therefore, a demonstration of, say, nonlocality automatically demonstrates the presence of entanglement and steerability.

[3]For completeness, a state ρ does not show EPR steering from Alice to Bob if for any measurement of Alice there exists probabilities P_λ and $P_{a|x\lambda}$ and states ρ_λ such that Bob's state after Alice's measurement with measurement choice x and outcome a is given by

$$\rho_{xa} = \sum_\lambda P_\lambda P_{a|x\lambda} \rho_\lambda .$$

This model is called a local-hidden-state model. Clearly, any separable state $\rho_s = \sum_l p_l \sigma_l \otimes \sigma'_l$ results in a local-hidden-state model for any measurement performed by Alice, and if Bob further performs measurements on ρ_{xa}, the final correlations are of the form of Eq. 1.2.

1.4 Outline

In this thesis, we will address the three topics of identification, verification and application of quantum resources for different specific problem settings. As each situation requires different theoretical backgrounds and literature overviews, we include extensive introductions to the respective topics at the beginning of each chapter. The different subjects of the remainder of this thesis are summarized in the following.

Identification

In Ch. 2, we consider one of the first quantum algorithms, the Grover search algorithm [34], and identify a necessary quantum resource for its quantum advantage. Grover's algorithm offers a provable quadratic speed-up for searching a large database for a marked element, and represents one of the key building blocks for many modern quantum algorithms. Different quantum resources have been considered to be crucial for this quantum advantage, but no general consensus was reached. We show that in the pure-state version of Grover's algorithm as well as its pseudo-pure-state generalization, we can identify a specific quantum statistical speed, the trace speed, as a necessary quantum resource. The trace speed can be further interpreted as demonstrating coherence or, assuming an implementation with several qubits, multipartite entanglement. The results of this chapter are published in Ref. [1].

Verification

An ubiquitous method in physics is the postselection of observed measurement data. Generally, postselection can corrupt the information of the data and lead to false conclusions, by means of the postselection bias. In Ch. 3, we address this problem in view of the certification of the quantum resource of genuine multipartite nonlocality. By the postselection bias, a general postselection can mimic nonlocal behaviour even if the complete correlations show no nonlocality. It was long believed that the only "safe" postselection that avoids the postselection bias is a locally decided postselection, e.g., that each party knows whether to keep or discard their result without the need for communication with the other parties. We show that even certain postselection strategies that require communication between several experimental parties can be used to verify genuine multipartite nonlocality, by providing conditions on the postselection to exclude the postselection bias. To show our results, we employ the statistical-inference tools of causal diagrams [75]. Furthermore, the postselection strategy is used for an experimental proposal to create genuine multipartite nonlocality with independent particle sources [76]. The results are published in Ref. [2].

In Ch. 4, we develop machine learning methods to detect the quantum resource of nonclassicality for different quantum optical measurement processes. Nonclassicality, defined

as the negativity of a quasiprobability distribution representing the quantum state, is one of the central quantum resources in continuous-variable quantum technologies. Typically, the detection of nonclassicality employs full quantum state tomography as a first step, and consequently requires large amounts of experimental data as well as advanced post processing strategies. A direct and fast detection of nonclassicality from experimental data that is applicable also for immediate amounts of measurement data has not been developed so far. We address this problem using machine learning methods. We train artificial neural networks by means of supervised learning with simulated data from (i) single-mode homodyne measurements and (ii) multiplexed click-counting measurements of different quantum optical states. The trained networks then operate as fast nonclassicality identifiers that can also be applied to small sets of measurement data. In case of homodyne measurements, the performance of the network is tested on several types of experimental data. The results are published in Ref. [3] for the homodyne measurements and in Ref. [4] for the click-counting measurements.

Application

Finally, in Ch. 5, we apply the available quantum resources of quantum mechanics to develop a new technique to perform quantum multiphase estimation. We introduce the Bayesian quantum multiphase estimation algorithm that has potential applications in both quantum computation and quantum metrology. The algorithm is shown to perform multiphase estimation, i.e., the parallel estimation of several phases, at the optimal scaling limit, the Heisenberg limit[4] with respect to the total number of resources used in the estimation protocol. The algorithm consists of a basic quantum measurement circuit that simultaneously measures the different phases as well as a classical post-processing strategy that is based on Bayesian inference. We show that, for a fixed number of total resources, the Bayesian quantum multiphase algorithm can outperform optimal sequential single-phase estimation protocols for several multiphase estimation tasks and can be further extended to the presence of phase damping noise. Finally, we propose two experimental realizations of the algorithm by either using generalized N00N states, or a multipass interferometer that can be implemented in state-of-the-art quantum optical experiments. The results are published in Ref. [5].

[4]The termination Heisenberg limit is usually used only in the standard quantum optical setting that does not allow for multiple passes of the phase shifts, see Sec. 1.2.2. Due to the formal equivalence of the standard setting to a multipass setting, we will use the term Heisenberg limit independently of the specific implementation of the algorithm, while the different implementations together with their respective advantages and disadvantages will be discussed in detail in Ch. 5.

Chapter 2

Identification: Quantum resources in Grover's algorithm

Contents

In this chapter, we will discuss necessary quantum resources of Grover's search algorithm and demonstrate the crucial role of the trace speed in the algorithm's pure-state and pseudo-pure-state implementations. First, in Sec. 2.1, we provide a detailed introduction to the quest of the identification of quantum resources of quantum computation and give an overview of the literature. In Sec. 2.2, we focus on Grover's algorithm and describe its operating principle. Finally, in Sec. 2.3, we prove that the trace speed describes a necessary quantum resource in Grover's algorithm.

2.1 Quantum resources in quantum computation

As being the arguably most famous quantum technology, quantum computation enjoys large attention from academics, industry and even public news coverage. An unavoidable question asked by people with any level of expertise has been where the apparent power of quantum computers comes from. As already indicated in Sec. 1.2, this problem has not yet

been completely resolved. In this section, we will sketch the main difficulties and partial resolutions to this question.

The question of where a quantum computational advantage comes from can be asked in two depths: first, one can ask which quantum resource has to be present in a given quantum algorithm, or to generally achieve a specific quantum advantage. This question corresponds to a rather explanatory role of the quantum resource for the advantage. Second, one can further ask for a quantitative attribution to the quantum resource: here, a quantitative relation between the amount of the quantum resource and the amount of the quantum advantage is sought-after. Most research has been attempting to answer the second (quantitative) problem, and we will roughly overview it in Sec. 2.1.1. We will briefly touch upon considerations about the first (explanatory) question in Sec. 2.1.2.

The quest for a quantitative attribution of a quantum resource to a quantum advantage is highly complicated by the following obstacles. Most importantly, the situation that the quantum advantage itself can be quantified is the exception. Even more, many quantum advantages rely on complexity-theoretical conjectures that have not been proven. Take the problem of integer factorization as an example. The classical hardness of factoring (see, e.g., Ref. [77]) results in an apparent exponential advantage of Shor's algorithm [16]. However, it has not been proven that no "efficient" classical algorithm exists that would make the quantum advantage disappear [77]. A different computational task that was impacted by a similar problem was discussed in Ref. [78], where a quantum algorithm for the recommendation problem, that was thought to yield an exponential quantum advantage over classical algorithms, was "dequantized" to yield a classical algorithm with a similar performance as its quantum relative. Furthermore, more efficient classical methods for factoring probably might be found that, while not touching the exponential advantage of Shor's algorithm, still change the exact quantitative advantage. In summary, one could say that the main difficulty of proving quantum computational advantage is not to prove why a quantum computer is fast (and of quantizing how much so) but to prove why a classical one is slow and to what extend[1].

There is an even more fundamental complexity-theoretic obstacle in the proof of an exponential advantage. The proof of an exponential gate-complexity advantage corresponds to a proof that

$$\mathbf{P} \neq \mathbf{BQP}, \tag{2.1}$$

where **P** is the complexity class of decision problems solvable by a classical computer (Turing machine) in polynomial time, and **BQP** is the class of decision problems solvable by a universal quantum computer in polynomial time. Since it can be proven that **BQP** is contained in the class **PSPACE** of decision problems solvable by a classical computer us-

[1] For a nice newspaper article that addresses this point, see https://www.quantamagazine.org/why-is-quantum-computing-so-hard-to-explain-20210608/.

ing a polynomial amount of space [15], a proof of Eq. (2.1) would represent a proof of $\mathbf{P} \neq \mathbf{PSPACE}$. This latter conjecture, similar to the famous $\mathbf{P} = \mathbf{NP}$ problem, has not been solved even despite decades of attempts of theoretical computer science, see, e.g., Ref. [79]. Therefore, without assuming complexity conjectures, an exponential advantage of quantum computers cannot be proven.

The situation looks somewhat better for quantum advantages with respect to oracle complexity, cf. Sec. 1.2.2. Even though the advantages are typically polynomial and not exponential, they often can be proven. For instance, the quadratic advantage of Grover's search algorithm that we will discuss below is provable and even provably optimal [80, 81]. Recently, even an oracle separation between **BQP** and the polynomial hierarchy **PH**, a superset of **NP**, has been proven [82].

In this context, we also want to note that the recent quantum computational advantage results [83, 84] rely on sophisticated complexity-theoretical reductions. Essentially, besides the fact that no efficient classical solutions to the corresponding computational problems are known, the hardness of the corresponding computational problems was reduced to reasonable hardness assumptions similar to $\mathbf{P} \neq \mathbf{NP}$. For the quantum advantage experiment using random circuit sampling [83], see Refs. [85, 86], and for the quantum advantage experiment using boson sampling [84], see Refs. [85, 87, 88].

On the other hand, a second obstacle is the quantification of a given quantum resource. For instance, there is an immense variety of entanglement measures or quantifiers that are non-equivalent [23]. As we will see in Sec. 2.1.1, different entanglement measures yield different answers to the question of the necessity of entanglement, which can easily lead to confusion. The answer of the necessity of a quantum resource thus depend on which measure of the resource is used, a fact that we have already seen in Sec. 1.2.3, where we described entanglement as detected by the QFI as the key resource for metrology, and we will see a similar result in Sec. 2.3.

2.1.1 Previous results on necessary quantum resources

In the following, we provide an overview of the key insights of the quest for quantum resources for quantum computation.

Entanglement

Among the first attempts to answer the question of the origin of quantum computational advantage was a result by Jozsa in 1997 [89] where he points out an essential role of entanglement. In particular, he discusses the (entanglement-requiring) exponential advantage of the quantum Fourier transform with respect to the classical (fast) Fourier transform. The important quantum algorithms of Deutsch–Jozsa [90], Simon [91] and Shor [16] are

all based on the quantum Fourier transform (over different groups). This line of reasoning was further deepened in Ref. [92].

Further insights were reached by a result of Linden and Popescu [93] who considered noisy quantum algorithms instead of pure ones. In particular, the pure quantum state is assumed to be mixed with uniform white noise, resulting in a so-called pseudo-pure state (for a detailed definition and description, see Sec. 2.3 and Eq. (2.22)), that naturally arises in nuclear magnetic resonance proposals for quantum computing [94, 95, 96]. Exponentially-advantageous quantum algorithms implemented with pseudo-pure states are then shown to only keep their advantage if the noise is not strong enough to remove the entanglement. Also, using the same reasoning, they show that entanglement is not a sufficient resource of the advantage: pseudo-pure states that are still entangled may not be sufficient for the exponential advantage anymore.

The first rigorous proof concerning the necessity of entanglement in pure-state algorithms yielding an exponential advantage was given by Jozsa and Linden in 2003 [31]. If the quantum state during the algorithm admits a factorization in which the factors consist of a bounded number of qubits, the final probability distribution of the algorithm can be simulated classically in polynomial time. Therefore, for an exponential advantage in a pure-state algorithm (assuming the usual complexity conjectures, see above), multipartite entanglement has to be present. Furthermore, they show why the state during Shor's algorithm generally is multipartite entangled, and they explicitly state that the situation for mixed-state advantages is not resolved by their results. Finally, they also argue why claiming a necessity of entanglement for quantum advantage is misleading because which resource is crucial depends on the formalism used to describe the state. For instance, by using the stabilizer formalism (see, e.g., Ref. [19]), a different quantum resource is important that is called magic [97].

A similar but quantitative result was shown by Vidal in 2003 [98]. From the maximal Schmidt rank[2] χ with respect to all bipartitions of the qubits, one can construct the entanglement measure $E_\chi = \log_2 \chi$. If E_χ scales as $\ln n$ (n is the number of qubits) during the pure-state quantum algorithm, it can be simulated on a classical computer. This result was then explicitly discussed for Shor's algorithm [99]. Furthermore, Ref. [99] considered an instance of a quantum adiabatic algorithm [100, 101] to solve an NP-complete problem and gave numerical evidence that it shows a maximal scaling of entanglement.

[2] For a bipartite system with Hilbert space $\mathcal{H} = \mathcal{H}_1 \otimes \mathcal{H}_2$, one can write any composite state $|\psi\rangle$ as

$$|\psi\rangle = \sum_{j=1}^{d} \alpha_j |\psi_j\rangle \otimes |\phi_j\rangle ,$$

where $\{|\psi_j\rangle\}_j$ ($\{|\phi_j\rangle\}_j$) is a set of orthonormal vectors in $\mathcal{H}_1$ ($\mathcal{H}_2$), and the real and positive coefficients α_j are unique up to reordering [19]. The decomposition is called the Schmidt decomposition, and d is called the Schmidt rank. For separable states along the bipartition, $d = 1$.

On the other side, a few results were derived that question a general role of entanglement in quantum computational advantage. First, the pure-state quantum algorithm proposed by Bernstein and Vazirani in 1997 [15] (solving a special case of the Deutsch–Jozsa decision problem), that achieves a polynomial quantum advantage in oracle complexity, makes no use of entanglement[3] [103]. This result was extended to the maximal subset of Deutsch–Jozsa-type problems that allow for a quantum advantage without entanglement [104]. Also, in Ref. [105], by using pseudo-pure separable states in the Deutsch–Jozsa and Simon algorithm, it was argued that while a classical computer learns no information of the answer after one step of the algorithm, a quantum computer learns a positive amount (although exponentially small in the number of qubits). Here, the gained information is quantified by the mutual information. See also Ref. [32] for a similar discussion.

Furthermore, in 1998, Knill and Laflamme proposed a computational model that only requires a single pure qubit and n further qubits in the completely mixed state [106] (often called the power of one qubit or deterministic quantum computation with one quantum bit, DQC1) and, while not representing a universal quantum computational model, achieves an exponential advantage in the estimation the trace of a unitary acting on the n qubits. This model was then shown to only host very little amounts of bipartite entanglement [33], and other appropriate quantum resources were proposed [107], see below.

Finally, a general result by Van den Nest in 2013 questioned the quantitative role played by entanglement in quantum computational advantage [108]: he shows that for many continuous measures of entanglement, an exponential advantage in pure-state quantum computations can be achieved with polynomially small amounts (in the number of qubits) of entanglement. This speaks against a quantitative role of entanglement measures in quantum computational advantage. Here, we want to mention another result about the quantitative role of entanglement in a specific computational model: in measurement-based quantum computation [109], states can be too entangled (with respect to the geometric measure of entanglement) to offer a quantum computational advantage [110].

In this context, we should also mention the Gottesmann–Knill theorem that showed that if a quantum computer is restricted to stabilizer circuits, the computation can be efficiently simulated on a classical computer [111]. Stabilizer circuits consist of Hadamard gates, controlled-NOT gates and phase gates applied on (and measured in) the computational basis. They include highly entangled states and are sufficient for quantum advantages in other quantum technologies such as quantum communication. Thus, we see that entanglement cannot be a sufficient quantum resource of quantum computational advantage, since a wide class of highly entangled quantum circuits can be efficiently simulated on a classical computer.

[3] Here, we want to note that the advantage of the Bernstein–Vazirani algorithm was questioned in Ref. [102] where a different classical oracle complexity of the problem is claimed.

Other quantum resources

Due to the obscure significance of entanglement for quantum computational advantage, other possible quantum resources have been extensively considered. As we discussed in Sec. 1.3, different quantum resources are often dependent or are simultaneously present in a quantum technology, so the following proposals do not rule out the importance of entanglement.

The vanishing amount of entanglement in the DQC1 computational model [33, 106] led to the investigation of the importance of a different quantum correlation, the quantum discord [107]. Quantum discord describes nonclassical correlations of separable states that arise in different quantum generalizations of the classical mutual information [41]. For instance, take the bipartite mixed state.

$$\rho = \frac{1}{2}\left(|0\rangle\langle 0| \otimes |0\rangle\langle 0| + |1\rangle\langle 1| \otimes |+\rangle\langle +|\right), \tag{2.2}$$

where, $|0\rangle$ and $|1\rangle$ are orthogonal states for both subsystems, and $|+\rangle = (|0\rangle + |1\rangle)/\sqrt{2}$. The state of Eq. (2.2) is separable but does not correspond to a classical mixture since $|0\rangle$ and $|+\rangle$ are not orthogonal, and it thus contains a positive amount of quantum discord. It was shown that almost all randomly picked states from the Hilbert space show a positive quantum discord [112]. Later, the role of quantum discord in DQC1 was questioned when considering specific unitaries that result in zero quantum discord during the computation but show the same apparent classical computational complexity (and thus quantum advantage) [113]. Finally, the necessary role of entanglement in the DQC1 model was shown by constructing an efficient simulation of the computation whenever no entanglement is present [114].

A different and recently much considered quantum resource is coherence [57]. In Ref. [115], it was shown that, in the Deutsch–Jozsa algorithm, less coherence leads to a larger error probability of the quantum algorithm and thus less quantum advantage. In the DQC1 algorithm, coherence is consumed to generate the quantum discord [116] and the precision of the computation can be quantified to the coherence measure of recoverable coherence [117]. The role of coherence in Grover's algorithm was investigated in Refs. [118, 119, 120] and will be further discussed in Sec. 2.3.

Recently, it was shown that nonlocality offers computational advantages in shallow (low-depth) quantum circuits [121]: a specific computational problem for binary quadratic forms (that resembles the Bernstein–Vazirani problem [15] but does not require a black-box oracle) can be solved with a constant-depth quantum circuit and local quantum gates, independent of the problem size (i.e., the number of qubits). Being restricted to Bell-inequality-type constraints, a classical circuit cannot solve the problem in constant depth. Due to the focus on shallow quantum circuits, this result is interesting in view of near term implementations that do not allow for arbitrary-depth circuits. Furthermore, in contrast

to many other results on quantum computational resources, it gives a clear answer to where the quantum advantage (for this specific problem) originates: correlations in quantum circuits can spread faster due to entanglement and nonlocality.

Further quantum resources have been considered as well. Reference [32] highlights the role of distinguishability (as measured by the relative entropy) for mixed-state quantum computations such as the DQC1 or the Shor and Deutsch–Jozsa algorithm using pseudo-pure states, but generally suggests that the origin of quantum computational advantage is problem dependent. In Ref. [122], the phenomenon of quantum contextuality [123, 124] was shown as the crucial quantum resource in the computational model of magic state distillation [97]. In the same computational model, the negativity of a Wigner representation of the quantum algorithm's state is necessary for quantum advantage [125, 126, 127]. In Ref. [128], quantum algorithms with little amount of interference (measured by the capacity of the algorithm's quantum gates to generate interference) are shown to be efficiently simulatable. In Ref. [129], it was shown that by using the quantum resource of indefinite causal order for the arrangement of the gates in the quantum circuit, an advantage in oracle complexity can be achieved for certain computational tasks. Finally, it was shown that in the measurement-based quantum computational model, an exponential complexity of the quantum states (measured by the so-called tree size) is essential for the quantum advantage [130].

Quantum statistical speeds

A different approach to quantum resources that we will further pursue in Sec. 2.3, are quantum statistical speeds [48, 50, 131, 132]. Quantum statistical speeds quantify the susceptibility of a parameter-dependent quantum state ρ_θ to variations of the parameter θ. As being directly built upon the statistical character of quantum measurements, they represent an experimentally attainable measure of the state's susceptibility. Different quantum statistical speeds were shown to be crucial for different quantum technological tasks, which highlights their potential as quantum resources. In the following, we briefly sketch the main ideas of quantum statistical speeds and then discuss their potential application in quantifying quantum resources.

A general measurement of a quantum state ρ is described by a positive-operator-valued measure (POVM) consisting of elements $\{E_x\}_x$ where each E_x is positive semi-definite operator and $\sum_x E_x = \mathbb{1}$ [19]. Corresponding to a specific POVM, the quantum state ρ induces a classical probability distribution

$$p_x = \mathrm{tr}[\rho E_x]. \tag{2.3}$$

Furthermore, two quantum states ρ and σ generally give rise to different probability distributions $\{p_x\}_x$ and $\{q_x\}_x$, respectively. A distance defined on the space of classical probability distributions is called a classical statistical distance, and can measure the difference

between the two distributions. Since the classical statistical distance between the distributions induced by ρ and σ depends on the chosen POVM, a measurement-independent quantum statistical distance D is obtained by maximizing the classical statistical distance over all possible POVMs (i.e., measurements). The quantum statistical distance then fulfils the axioms of a distance (non-negativity, symmetry, and the triangle inequality) on the space of quantum states.

For a given quantum statistical distance D, the corresponding quantum statistical speed QS of a parameter-dependent quantum state ρ_θ at θ_0 is defined as

$$\mathrm{QS}(\rho_{\theta_0}) = \partial_\theta D(\rho_{\theta_0}, \rho_{\theta_0+\theta})\big|_{\theta=0}. \tag{2.4}$$

Intriguingly, if the state's parameter dependence is induced by a unitary evolution, $\rho_\theta = U_\theta \rho_0 U_\theta^\dagger$, the possible quantum statistical speed is limited depending on the amount of entanglement of ρ_{θ_0}: generally, the speed that can be achieved by entangled states is strictly larger than the possible speed that can be achieved by separable states. In other words, if the state ρ_θ surpasses the quantum speed limit of separable states, entanglement of ρ_θ is demonstrated [27, 50]. Thus, the QS represents an entanglement witness as described in Sec. 1.3.

The most prominent example of a quantum statistical speed is the quantum Fisher information (QFI) [48, 49], also mentioned in Sec. 1.2.3. The QFI is the quantum statistical speed of the Bures distance [49] that corresponds to the classical Hellinger distance [133] defined as $D_\mathrm{H}(p_x, q_x) = \sqrt{\sum_x (\sqrt{p_x} - \sqrt{q_x})^2/2}$. For a unitary evolution $U_\theta = e^{-iH\theta}$ that is generated by a Hamiltonian H, the QFI of a quantum state ρ is given by [49]

$$\mathrm{QFI}(\rho, H) = 2 \sum_{k,k'} \frac{(\lambda_k - \lambda_{k'})^2}{\lambda_k + \lambda_{k'}} \left|\langle k| H |k'\rangle\right|^2, \tag{2.5}$$

where $|k\rangle$ label the eigenvectors of ρ with corresponding eigenvalues λ_k, and the sum is taken over terms for which $\lambda_k + \lambda_{k'} > 0$. The QFI has a direct operational importance in quantum metrology by means of the Quantum Cramér–Rao bound [46, 47]

$$(\Delta\theta)^2 \propto \frac{1}{\mathrm{QFI}}, \tag{2.6}$$

where $(\Delta\theta)^2$ is the frequentist variance of the estimated parameter θ, see, e.g., Ref. [134] for a conclusive review, and also Ch. 5 and Eq. (5.1) for further details. Furthermore, the QFI was shown to induce a hierarchy of speed limits that can be used as a witness of multipartite entanglement [27, 28, 29]: for a system consisting of n qubits and a Hamiltonian $H = \sum_i H_i$, where H_i acts locally on the ith qubit with a spectrum $\mathrm{spec}(H_i) = \{1/2, 1/2\}$, if one finds $\mathrm{QFI}(\rho, H) > nk$, the state ρ has to be at least $(k+1)$-partite entangled. Here, recall that the state ρ is called k-partite entangled if it is k-producible but not $(k-1)$-producible, where ρ is k-producible if $\rho = \sum_l \lambda_l |\psi_l\rangle \langle\psi_l|$ and each $|\psi_l\rangle$ is a tensor product

of different subsystems containing maximally k qubits. In this way, maximal sensitivities in quantum metrology require multipartite entanglement that is captured by the QFI. In other words, the key quantum resource of quantum metrology is multipartite entanglement that is detected by the QFI, also called metrologically-useful entanglement. In this context, we want to mention that the connection between the QFI and the multipartite entanglement can be further refined by examining the Young diagrams corresponding to the entanglement structure of the state [135]. In this way, the crucial resource of metrological advantage is the Dyson's rank that combines the notions of k-partite entanglement (here referred to as entanglement depth) and k-separability (i.e., the maximal number of separable blocks in any partition of the state) [136].

A second and less-known example of a quantum statistical speed is the trace speed that is obtained from the trace distance which corresponds to the classical Kolmogorov distance $D_{\mathrm{K}}(p_x, q_x) = \sum |p_x - q_x| /2$ [19, 137]. The trace speed (TS) of a state $\rho_\theta = e^{-iHt}\rho_0 e^{iHt}$ at θ_0 is given by [50, 138]

$$\mathrm{TS}(\rho_{\theta_0}, H) = \|\partial_\theta \rho_\theta\|_1\big|_{\theta=\theta_0} = \|[\rho_{\theta_0}, H]\|_1, \tag{2.7}$$

where $[\cdot,\cdot]$ is the commutator and $\|\cdot\|_1$ is the l_1-norm defined as $\|X\|_1 = \mathrm{tr}\left[\sqrt{X^\dagger X}\right]$. The TS can be bounded by the QFI, $\mathrm{TS}(\rho, H) \leq \sqrt{\mathrm{QFI}(\rho, H)}$ [139]. Consequently, as the QFI acts as an entanglement witness, the TS can detect multipartite entanglement as well. In particular, finding $\mathrm{TS}(\rho, H) > \sqrt{nk}$ (again, $H = \sum_i H_i$ with local Hamiltonians H_i and $\mathrm{spec}(H_i) = \{1/2, 1/2\}$) is a witness of at least $(k+1)$-partite entanglement. Furthermore, similar to the QFI, the TS represents an experimentally relevant measure of coherence [57, 138]: a state that has no coherence with respect on a Hamiltonian H, i.e., a classical mixture of the Hamiltonian's eigenstates, does not change under displacements generated by H, while larger coherences (off-diagonal entries) result in larger susceptibilities to the displacements.

The TS plays a central quantum technological role in distinguishing quantum states [19, 50, 140]. A single-shot measurement can only perfectly distinguish two quantum states ρ and σ if the latter are orthogonal. In general, an optimal measurement can be found such that the states can be correctly distinguished with probability $P = [1 + D_{\mathrm{tr}}(\rho, \sigma)]/2$ [19, 140], where D_{tr} is the trace distance. Therefore, if the task is to find most-distinguishable ρ_0 and ρ_θ for a fixed small θ, states with large TS lead to higher performances [50]. As above, we see again that multipartite entanglement gives advantages for this specific task, but this time it is the multipartite entanglement detected by the TS. In Sec. 2.3, we will show that the TS can be used to quantify the speed-up of different generalizations of Grover's search algorithm.

A key advantage of quantum statistical speeds with respect to many other candidates of quantum resources is that the former can be measured or efficiently bounded in experiments [141, 142]: for instance, by measuring ρ_θ for a given measurement observable for

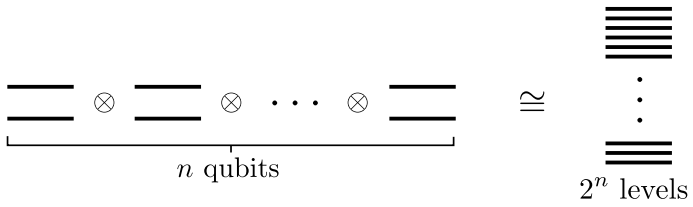

Figure 2.1: Formally, the Hilbert space describing a system that consists of n qubits is isomorphic to the Hilbert space of a single 2^n-level quantum system.

several (small) values of θ, one can estimate the resulting classical probability distributions. A quadratic series expansion of the classical statistical distance of these distributions to the initial one yields an estimate of the classical statistical speed (corresponding to the chosen measurement observable). The classical statistical speed then serves as a lower bound on the quantum statistical speed [50]. Depending on the quantum statistical speed of interest, different classical statistical distances have to be used (e.g., the Hellinger distance for the QFI, and the Kolmogorov distance for the TS).

2.1.2 Hilbert space equivalence

In this section, we want to briefly discuss the qualitative, or explanatory, origins of quantum computational advantage. We first recall that the Hilbert space describing n qubits is isomorphic to the Hilbert space describing a single 2^n-level system (cf. Fig. 2.1),

$$\bigotimes_{i=1}^{n} \mathcal{H}_2 \cong \mathcal{H}_{2^n}, \tag{2.8}$$

by means of the isomorphic mapping $|x_1\rangle \otimes |x_2\rangle \otimes \cdots |x_n\rangle \mapsto |x_1 x_2 \cdots x_n\rangle$. By this isomorphism, each state and each unitary in the n-qubit system correspond to a state and a unitary in the single multilevel system, so all possible quantum algorithms formally also work in the single multilevel system. Therefore, the answer to the question of why any quantum algorithm works, and why it gives the correct answer, cannot be entanglement but has to be a feature that also describes the single multilevel system. This feature is constructive and destructive interference: in order to have a large probability to find the correct answer, the amplitudes of this answer have to constructively interfere, while the amplitudes of the wrong answers should destructively interfere.

The above reasoning answers the question about how a quantum algorithm works, but not how efficient it can be implemented in the two respective platforms. Crucially, this opens a large efficiency difference between the two: a 2^n-level system naturally requires some exponentially large physical resource to be implemented. If the levels have a constant energy spacing, the total energy needed to implement a general unitary (say, to map the lowest to the highest energy level), requires an exponential amount of energy (with respect to the level spacing). If the energy levels are allowed to have a spacing that decreases

exponentially with n, a finite amount of energy suffices, however, the resolution of the energy levels requires an exponential precision (e.g., resulting in an exponential measurement time or a necessary exponential spatial resolution). A second difference is that the basic (easy-implementable) gates for both systems are very different. Say, the basic gates of the n-qubit system are single and two-qubit gates, while the basic gates of the multimode system act on a single or two modes (e.g., phase shifts and beam splitters, respectively, in the case of a quantum-optical architecture). Then, the application of a basic qubit gate on the first qubit is easily-implementable but the action of the corresponding gate in the multilevel system requires exponentially many basic gates because it has to act on all modes [12]. Therefore, while quantum algorithms simply require interference as a working principle, the n-qubit model offers an exponentially more efficient implementation of states and operations.

The consideration of all physical resources underlines that the quantum advantage with respect to gate complexity originates from the tensor-product structure and (multipartite) entanglement. However, as we discussed in Sec. 1.2.2, many algorithms offer a quantum advantage in query complexity. Here, only the oracle unitaries, that are taken as black boxes, are contributing to the complexity of the algorithm, and the number of additional gates, even if exponential, does not matter. Furthermore, strictly speaking, no other physical resource such as, e.g., energy, is taken into account. This observation has led to proposals about quantum advantage without entanglement [143] that were later implemented using a Rydberg atom [144] and classical light waves [145] (the later is possible because interference can also be present in some "classical" systems, e.g., in systems that are described by classical electromagnetism). However, all results require certain exponential physical resources as discussed in Refs. [103, 146].

In summary, we have seen that quantum advantages in query complexity formally do not require entanglement since they can be implemented with a single quantum system. However, an efficient (and thus realistic) implementation generally makes use of the tensor-product structure and entanglement. We will observe this situation for Grover's algorithm in Sec. 2.3.

2.2 Grover's algorithm

In this section, we give an in-depth description of Grover's search algorithm [34]. Grover's algorithm is one of the first and most famous quantum algorithms, which is mainly due to the ubiquitous need for a fast search subroutine in computations. Furthermore, it is not part of the "hidden subgroup" algorithms such as the Simon, Deutsch-Jozsa and Shor algorithm (see, e.g., Ref. [19]). It serves as a key subroutine in many subsequent quantum algorithms such as, e.g., quantum amplitude estimation [147], quantum counting [148],

and the quantum singular value transformation [149], see also Refs. [150, 151, 152, 153] for more applications. Its main critique is that its advantage is usually only discussed in query complexity (cf. Secs. 1.2.2 and 2.1.2), and that the mere quadratic advantage cannot compensate for realistic implementations using quantum error correction [154]. We also want to note that Grover's algorithm was conjectured to occur naturally in fermionic systems with topological defects [155], and it was even used to motivate why the genetic code uses four nucleotide bases to encode information about 20 amino acids [156]. Also note a surprising analogy of Grover's algorithm to a problem of classical mechanics ("bouncing billiard balls"), by which it provides a (extremely inefficient and fragile) way of measuring the digits of π [157]. In the following, we first sketch the algorithm's steps and then discuss its quantum advantage. Finally, we briefly overview previous results on quantum resources that play a role in Grover's algorithm.

The algorithm

The computational task for which Grover's algorithm was developed is the search of a marked element in a large unstructured database. We consider a database of size $N = 2^n$ and label each entry by $x \in \{1, \dots, N\}$. One of the elements, ω, is the marked element that is also called target. We don't know which element is marked but we are given an oracle function f that, given an input entry x, tells us if x is the marked element, $f(x) = 1$, or not, $f(x) = 0$. As we explain below, Grover's algorithm offers a quadratic advantage in this task in terms of query complexity, when counting the average number of applications of the oracle function.

In the quantum setting, the database corresponds to the computational basis vectors of n qubits, where each vector is denoted as $|x\rangle$. Here, the oracle function f can be used in two different ways. First, it can be used as a measurement observable M_f acting on the algorithm's quantum state $|\psi\rangle$ with outcomes -1 if $|\psi\rangle = |\omega\rangle$ and 1 if $|\psi\rangle = |x\rangle$ for $x \neq \omega$. For a quantum state $|\psi\rangle$ that has a finite overlap with the target, measuring M_f yields -1 with probability $p = |\langle\omega|\psi\rangle|^2 < 1$, after which we can read out the target entry. On the other hand, the oracle can be used as a quantum gate U_f. Generally, the answer of an oracle call is imprinted in an additional (ancilla) quantum system by $\tilde{U}_f(|x\rangle \otimes |y\rangle) = |x\rangle \otimes |f(x) \oplus y\rangle$, where $\oplus$ denotes the addition modulo the dimension of the ancilla system. For Boolean functions f (i.e., $f(x) = 0$ or $f(x) = 1$ for all x), the oracle unitary can be equivalently expressed as a phase oracle[4]

$$U_f(|x\rangle) = (-1)^{f(x)} |x\rangle , \tag{2.9}$$

[4]Explicitly, by preparing the ancilla state in $|-\rangle = (|0\rangle - |1\rangle)/\sqrt{2}$, one has

$$\tilde{U}_f(|x\rangle \otimes |-\rangle) = \frac{1}{\sqrt{2}} |x\rangle \otimes (|f(x)\rangle - |f(x) \oplus 1\rangle) = \frac{1}{\sqrt{2}} (-1)^{f(x)} |x\rangle \otimes (|0\rangle - |1\rangle) = U_f(|x\rangle) \otimes |-\rangle .$$

without the need of an ancillary quantum system.

Grover's algorithm consists of preparing the initial quantum state in an equal superposition of all entries,

$$|\psi_{\text{in}}\rangle = \frac{1}{\sqrt{N}} \sum_x |x\rangle \,, \tag{2.10}$$

and then alternatingly apply the phase oracle and the so-called Grover diffusion operator

$$U_d = 2\,|\psi_{\text{in}}\rangle\,\langle\psi_{\text{in}}| - \mathbb{1}. \tag{2.11}$$

The corresponding quantum circuit of the n-qubit implementation is shown in Fig. 2.2[5]. We use n Hadamard gates H ($H_{ij} = (-1)^{ij}/\sqrt{2}$ in the computational basis) to prepare the initial state, $|\psi_{\text{in}}\rangle = H^{\otimes n}\,|0\rangle^{\otimes n}$. During the algorithm, all amplitudes are real, so we can easily sketch the first steps in Fig. 2.3. The initial state ($k = 0$) corresponds to an equal superposition of all amplitudes. The action of the phase oracle U_f simply reverts the sign of the target state's amplitude. On the other hand, the action of U_d results in a reflection of all amplitudes about the mean amplitude (in the computational basis). This can be seen by writing the algorithm's state as $|\psi\rangle = \sum_x \alpha_x\,|x\rangle$ with a mean $\bar{\alpha}_x = \sum_x \alpha_x/N = \langle\psi_{\text{in}}|\psi\rangle/\sqrt{N}$. Then,

$$\begin{aligned} U_d(|\psi\rangle) &= 2\langle\psi_{\text{in}}|\psi\rangle\,|\psi_{\text{in}}\rangle - |\psi\rangle && (2.12)\\ &= \sum_x \left(\frac{2\langle\psi_{\text{in}}|\psi\rangle}{\sqrt{N}} - \alpha_x\right)|x\rangle && (2.13)\\ &= \sum_x (2\bar{\alpha}_x - \alpha_x)\,|x\rangle \,, && (2.14) \end{aligned}$$

and the function $r(x) = 2b - x$ describes a reflection about the value b. In this way, the amplification during the algorithm can be qualitatively understood, see Fig. 2.3. Explicitly, the state after k iterations of $U_d U_f$ is given by (see Refs. [19, 158] for a detailed derivation)

$$|\psi_k\rangle = (U_d U_f)^k\,|\psi_{\text{in}}\rangle = \sin[(2k+1)\theta_N]\,|\omega\rangle + \cos[(2k+1)\theta_N]\,|\perp\rangle \,, \tag{2.15}$$

where we defined $\theta_N = \arcsin\left(1/\sqrt{N}\right)$ and $|\perp\rangle = \sum_{x\neq\omega} |x\rangle\,/\sqrt{N-1}$. $|\perp\rangle$ is the normalized projection of the initial state $|\psi_{\text{in}}\rangle$ onto the subspace orthogonal to the target $|\omega\rangle$. In short, the evolution during the algorithm can be shown to be restricted to the two-dimensional subspace spanned by $\{|\omega\rangle\,, |\perp\rangle\}$. In this subspace, the action of U_f acts as a reflection along $|\perp\rangle$, while U_d acts as a reflection along $|\psi_{\text{in}}\rangle$. The concatenation of the reflections results in a rotation about an angle $2\theta_N$.

From Eq. (2.15), we see that the probability of an outcome -1 when measuring the observable M_f, corresponding to a successful search, approaches 1 after a number of k_{Gr}

[5] During this thesis, we draw quantum circuits with the help of the Tikz subpackage Qunatikz, see https://ctan.org/pkg/quantikz.

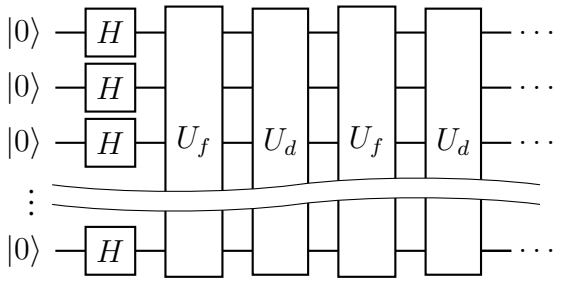

Figure 2.2: The quantum circuit of Grover's algorithm in the n-qubit implementation. The initial layer of Hadamard gates H prepares the initial state $|\psi_{\text{in}}\rangle$. Then, the oracle unitary U_f and the Grover diffusion operator U_d are alternatingly applied.

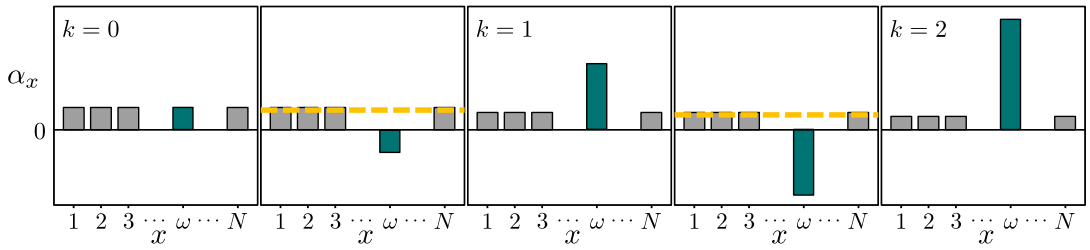

Figure 2.3: Sketch of the quantum state's amplitudes α_x during the first two iterations of Grover's algorithm. In each iteration, the amplitude of the target (blue) is flipped, after which all amplitudes are reflected about the mean amplitude $\bar{\alpha}_x$ (yellow).

iterations[6] with

$$k_{\text{Gr}} = \frac{\pi}{4\theta_N} \approx \frac{\pi}{4}\sqrt{N}, \tag{2.16}$$

where we have used that $\theta_N = \arcsin\left(1/\sqrt{N}\right) \approx 1/\sqrt{N}$ for large N. Thus, Grover's search is successful after $\mathcal{O}(\sqrt{N})$ oracle calls. Equation (2.15) also directly gives a qualitative hint of why Grover's algorithm outperforms a classical search algorithm: while in a classical search (assuming large N), each oracle call initially increases the probability of measuring the target by roughly $1/N$ (so $\mathcal{O}(N)$ oracle calls are needed), in the quantum algorithm, each oracle call (initially) increases the target's amplitude by $2/\sqrt{N}$. Since probabilities are squared amplitudes, only $\mathcal{O}(\sqrt{N})$ oracle calls are needed to find the target.

We want to note that Grover's algorithm also can be applied if $M > 1$ database entries are marked and the task is to find any of the marked elements. In this case, one has $\theta_N = \sqrt{M/N}$ and, thus, $\mathcal{O}(\sqrt{N/M})$ oracle calls are needed. This still offers a quadratic advantage to a classical search. For more details, see, e.g., Ref. [158]. In the following, we will focus on the case that only one element is marked.

[6] Here and in the following, we assume that n and thus N are sufficiently large such that the discreteness of the algorithm's iterations are negligible. For instance, for $n = 30$, we have $k_{\text{Gr}} \approx 2.6 \times 10^4$. The exact probability for a successful search after $\lfloor k_{\text{Gr}} \rfloor$ iterations is $p = 1 - \mathcal{O}(1/N)$ [158], where $\lfloor \cdot \rfloor$ is the floor function.

Quantum advantage of Grover's algorithm

In order to consistently quantify the advantage offered by Grover's algorithm, one first needs to define the complexity, or the cost, with respect to which there is an advantage, see Sec. 1.2.2 for a discussion. Grover's algorithm offers an advantage in query complexity, so the consistent definition of the cost C is the average number of oracle calls (taken as a measurement M_f or as a unitary U_f) to find the marked element [159].

In the classical search problem, an oracle call can be thought of as opening one of N boxes to find a target object. By randomly opening the boxes, the probability of finding the target is $p = 1/N$, so the average number of oracle calls is $C_{\mathrm{cl}} = N$. By remembering which boxes have been already opened, the average number of calls can be reduced to an optimal[7]

$$C_{\mathrm{cl}} = \frac{N}{2} + \mathcal{O}(1). \tag{2.17}$$

In the quantum setup, the oracle can be applied both as a measurement (M_f) or as a unitary (U_f). If we use k unitary oracle calls and then measure the state, we have used a total of $k+1$ oracle calls. If the probability of finding the target after this sequence is p_k, the average number of oracle calls needed to find the target is then $(k+1)/p_k$. The optimal cost of any quantum search algorithm is therefore

$$C_{\mathrm{qu}} = \min_k \frac{k+1}{p_k}. \tag{2.18}$$

In Grover's algorithm, we have $p_k = \sin^2[(2k+1)\theta_N]$ (cf. Eq. (2.15)) after k Grover iterations. The optimal number of iterations $\tilde{k}_{\mathrm{Gr}}$ can be calculated by solving $\partial_k[(k+1)/p_k] = 0$, yielding[8]

$$\tilde{k}_{\mathrm{Gr}} \approx 0.74 \times k_{\mathrm{Gr}}. \tag{2.19}$$

[7] After opening $k-1$ boxes, there are $N-k+1$ remaining boxes, so the probability to find the target is $q_k = 1/(N-k+1)$. The probability of finding the target in the kth oracle calls is thus $p_k = q_k \prod_{l=1}^{k-1}(1-q_l)$, and average number of oracle calls is given by

$$C_{\mathrm{cl}} = \sum_{k=1}^{N} k p_k = \frac{N+1}{2}.$$

[8] One has

$$\partial_k \frac{k+1}{\sin^2[(2k+1)\theta_N]} = \frac{1}{\sin^2[(2k+1)\theta_N]} \left\{1 - 4\theta_N(k+1)\cot[(2k+1)\theta_N]\right\}.$$

We assume large N, so we have $\theta_N \ll 1$. Furthermore, note that for $k\theta_N \ll 1$ and using $\cot x \approx 1/x$ for small x, a solution to the above equation corresponds to solving $1 - 4(k+1)/(2k+1) = 0$, which has no solution for positive k. Therefore, $\tilde{k}_{\mathrm{Gr}}$ has to scale at least as θ_N, such that $\tilde{k}_{\mathrm{Gr}}$ can be defined by the first positive solution of

$$\tan(2\theta_N k) = 2\theta_N k,$$

which we numerically solve to find $\tilde{k}_{\mathrm{Gr}} \approx 0.58 \times \sqrt{N}$.

We see that, according to the cost definition of Eq. (2.18), it is more efficient (on average) to measure the algorithm's state before reaching the highest success probability $p_{k_{\mathrm{Gr}}} = 1$ [81]. This is because the final maximal success probability is reached only very slowly with an increasing number of iterations. The corresponding minimal cost is then given by

$$C_{\mathrm{qu}} = \frac{\tilde{k}_{\mathrm{Gr}} + 1}{\sin^2[(2\tilde{k}_{\mathrm{Gr}} + 1)\theta_N]} = K_{\mathrm{Gr}}\sqrt{N} \tag{2.20}$$

with $K_{\mathrm{Gr}} \approx 0.69$.

In summary, we see that the pure-state Grover algorithm leads to a quadratic quantum speed-up S of

$$S = \frac{C_{\mathrm{cl}}}{C_{\mathrm{qu}}} = \frac{\sqrt{N}}{2K_{\mathrm{Gr}}}. \tag{2.21}$$

This speed-up was proven to be optimal: no quantum algorithm can perform the search with less than $\mathcal{O}(\sqrt{N})$ oracle calls on average [80, 81]. Note that if there existed a quantum algorithm that performs the search in $\mathcal{O}(\ln N) = \mathcal{O}(n)$ time, this would imply that **NP** $\subset$ **BQP**, i.e., that NP-complete problems can be solved efficiently on a quantum computer: a straight-forward search among all possible solutions using the quantum search algorithm would be efficient (since problems in **NP** can be efficiently checked).

We want to note that Grover's algorithm not only offers a provable quantum advantage in terms of query complexity, but is efficient in terms of memory and gate complexities: it only uses $n = \log_2 N$ qubits, so the necessary memory space is small, and the total gate complexity is of the order $\mathcal{O}[\sqrt{N}(\ln N)^2]$. The latter result comes from the fact that an operation like the $(n-1)$-qubit-controlled phase gate $\mathbb{1} - 2\,|0^{\otimes n}\rangle\langle 0^{\otimes n}|$ (that can be converted to U_d and U_f by applying $\mathcal{O}(n)$ single-qubit Hadamard and Pauli gates), can be implemented by $\mathcal{O}(n^2)$ basic gates [160].

Previous results on quantum resources of Grover's algorithm

The question of which quantum resource gives rise to the advantage of Grover's algorithm has been the subject of several investigations during the last years. The mere presence of genuine multipartite entanglement (see Sec. 1.2.1 for a definition) in Grover's algorithm after only the first step was noted in Ref. [161], and a specific efficiently-computable entanglement measure during the algorithm's evolution was considered in Ref. [162]. No quantitative discussion of the role of entanglement in the quantum advantage was given. In Ref. [163], the bipartite entanglement measure called concurrence [164] was examined during the algorithm (for the reduced density matrix of any two qubits), and an analytical dependence between the concurrence and the success probability was found. Reference [165] considered the concurrence with respect to a bipartition of the qubits and found a relation between the change of the success probability and the concurrence. Finally, in

Ref. [166], entanglement during Grover's algorithm was quantified by the multipartite geometric measure of entanglement [167] and was seen to be independent of the number of qubits. In contrast, increasing amounts of entanglement in the initial state of the algorithm were shown to decrease the success probability [168].

Further work has investigated the role of coherence during Grover's algorithm. In Refs. [118, 119], the success probability and the optimal measurement iteration were shown to be related to the consumption of coherence during the algorithm, where coherence is quantified in terms of the relative entropy of coherence and the l_1-norm of coherence [57]. Here, also the dynamics of entanglement and quantum discord were further studied. In Ref. [120], it was shown that during each single iteration of the algorithm, coherence is both first decreased and then increased, or vice versa.

While all above results shine light on the presence of different quantum resources during the pure-state Grover algorithm, they do not touch on noisy mixed-state generalizations. Conversely, the (loss of the) quantum advantage in different noisy versions of Grover's algorithm was investigated, using the noise models of Gaussian noise [169], systematic phase errors [170], depolarization noise [159, 171], bit-flip noise [172], unitary noise [173], oracle errors [174, 175, 176] and decoherence between subspaces [177]. Furthermore, in Ref. [178], it was shown that any constant noise rate per iteration (assuming depolarization noise) inhibits the asymptotic quadratic quantum advantage, even in a general quantum search algorithm. However, these results for noisy search algorithms did not discuss the role of quantum resources.

2.3 Trace speed as a quantum resource in Grover's algorithm

In this section, we approach the question of quantum resources in mixed-state versions of Grover's algorithm from the viewpoint of quantum statistical speeds, see 2.1.1 for an introduction. We first investigate a pseudo-pure initial state and a (noiseless) unitary algorithm in Sec. 2.3.1 and connect the quantum advantage to a wide class of quantum statistical speeds. In Sec. 2.3.2, we then consider a general pseudo-pure-state version of Grover's algorithm and observe that the maximal trace speed during the evolution represents a necessary resource for the quantum advantage. The results of this section are published in Ref. [1].

2.3.1 Unitary version of Grover's algorithm

We first assume that the quantum algorithm can be implemented in a unitary way, i.e., without any noise, so the quantum state's evolution is the one described for the pure-state

algorithm in Sec. 2.2. However, we do not assume that the initial state is a pure state as before, but rather that it is mixed with uniform noise. This initial state corresponds to a so-called pseudo-pure state

$$\rho_{\epsilon,\psi} = \epsilon \left|\psi\right\rangle \left|\psi\right\rangle + \frac{1-\epsilon}{2^n}\mathbb{1}, \tag{2.22}$$

where $\epsilon \in (0, 1]$ is the polarization of the pseudo-pure state. Note that from now on, we will focus on the n-qubit implementation of Grover's algorithm, so we identify the labels $x \in \{1, \ldots, N\}$ used Sec. 2.2 with the labels $x \in \{0, 1\}^n$. Note that, for $\epsilon < 1/(1+2^{2n-1})$, the pseudo-pure state is always separable independent of the entanglement of $|\psi\rangle$ [179].

The pseudo-pure state is widely used to model a noisy state of a quantum computer: it is a simple noise model such that it does not render the computation intractable. Further, it describes the quantum state during a noisy algorithm in presence of depolarising noise, see Sec. 2.3.2. Finally, it is used to describe the quantum state in the computational model of ensemble quantum computing using nuclear magnetic resonance [94, 95, 96]. Here, Eq. (2.22) approximates the high-temperature Gibbs ensemble, where $|\psi\rangle$ corresponds to the lowest-energy state.

Due to the linearity of quantum mechanics and the trivial action of the algorithm on the noise part, the algorithm's state after k iterations is given by

$$\rho_{\epsilon,\psi_k} = \epsilon \left|\psi_k\right\rangle \left|\psi_k\right\rangle + \frac{1-\epsilon}{2^n}\mathbb{1}, \tag{2.23}$$

where $|\psi_k\rangle$ is the state of the pure-state version of the algorithm, cf. Eq. (2.15). Therefore, the probability of a successful measurement after k iterations is now $p_k = \epsilon \sin^2[(2k+1)\theta_N] + (1-\epsilon)/2^n$. We see that for $\epsilon \gg 1/2^n$, the probability of finding the target state during any step of the algorithm is simply reduced by a factor of ϵ. In particular, the cost C_{qu} (Eq. (2.18)) is minimized after the same number $\tilde{k}_{\mathrm{Gr}}$ of iterations as in pure-state algorithm, and takes the value

$$C_{\mathrm{qu}} = \frac{K_{\mathrm{Gr}}\sqrt{2^n}}{\epsilon}, \tag{2.24}$$

cf. Eq. (2.20).

For $\epsilon \sim 1/\sqrt{2^n}$, the state is so noisy that a classical search is advantageous. We observe numerically that the optimal step to measure the state is either $k = 0$ (for $\epsilon \sim 1/\sqrt{2^n}$) or $k = \tilde{k}_{\mathrm{Gr}}$ (for $\epsilon \gg 1/\sqrt{2^n}$). More precisely, the critical amount of noise ϵ_{crit} is defined by the equality of quantum and classical costs, $(K_{\mathrm{Gr}}\sqrt{2^n})/\epsilon_{\mathrm{crit}} = 2^n/2$, and is thus given by

$$\epsilon_{\mathrm{crit}} = \frac{2K_{\mathrm{Gr}}}{\sqrt{2^n}}, \tag{2.25}$$

where $K_{\mathrm{Gr}} \approx 0.69$. In the following, we will assume that the polarization is much larger $\epsilon \gg \epsilon_{\mathrm{crit}}$, such that we are in the regime of a quantum advantage.

By the above considerations, we can easily quantify the quantum speed-up S in terms of the polarization ϵ,

$$S = \frac{\epsilon\sqrt{2^n}}{2K_{\mathrm{Gr}}}. \tag{2.26}$$

At this point, the crucial parameter to quantify the speed-up is ϵ, and any quantum feature or quantum resource that can be connected to ϵ can thus be used to quantify the speed-up. Note that this will no longer hold true in the case of depolarization noise during the algorithm, as we will see in Sec. 2.3.2.

In view of later results, we now want to focus on the role of quantum statistical speeds. For a wide class of quantum statistical speeds, one can show that for $\epsilon \gg 1/2^n$,

$$\mathrm{QS}(\rho_{\epsilon,\psi}, H) = \epsilon\, \mathrm{QS}(|\psi\rangle\langle\psi|, H) \tag{2.27}$$

for any Hamiltonian H. This is fulfilled, e.g., by the families of generalized quantum Fisher information and the Schatten speeds [50] (of which the trace speed is a special case, see below). For instance, one easily checks using the definition of the QFI, cf. Eq. (2.5), that

$$\begin{aligned}
\mathrm{QFI}(\rho_{\epsilon,\psi}, H) &= 2\sum_{k,k'} \frac{(\lambda_k - \lambda_{k'})^2}{\lambda_k + \lambda_{k'}} |\langle k| H |k'\rangle|^2 \\
&= 4\sum_{k\neq\psi} \frac{\epsilon^2}{\epsilon + 2(1-\epsilon)/2^n} |\langle k| H |\psi\rangle|^2 \\
&= 4\sum_{k\neq\psi} (\epsilon + \mathcal{O}[(1-\epsilon)/2^n]) |\langle k| H |\psi\rangle|^2 \\
&= \epsilon\, \mathrm{QFI}(|\psi\rangle\langle\psi|, H) + \mathcal{O}[(1-\epsilon)/2^n].
\end{aligned} \tag{2.28}$$

Here, recall that $|k\rangle$ label the eigenvectors of $\rho_{\epsilon,\psi}$ with corresponding eigenvalues λ_k. In the first line, we used that all eigenvalues but one (λ_ψ) of $\rho_{\epsilon,\psi}$ are degenerate.

By means of the polarization ϵ, any quantum statistical speed QS fulfilling Eq. (2.27) can be used to quantify the algorithm's speed-up S. As we discussed in Sec. 2.1.1, for a local Hamiltonian H, a large $\mathrm{QS}(\rho, H)$ offers a witness of multipartite entanglement and coherence of ρ. Therefore, we want to quantify the speed-up by the maximal statistical speed $\mathrm{QS}_{\max}$ that occurs during the algorithm, maximized over all possible local H. We will explicitly perform this optimization for the trace speed below. Since the maximal QS during the algorithm is reduced by a factor of ϵ with respect to the pure-state algorithm,

$$\mathrm{QS}_{\max} = \epsilon\, \mathrm{QS}^{\mathrm{pure}}_{\max}, \tag{2.29}$$

we can quantify the quantum speed-up of the noiseless Grover algorithm with a pseudo-pure initial state as

$$S = \frac{\sqrt{2^n}}{2K} \frac{\mathrm{QS}_{\max}}{\mathrm{QS}^{\mathrm{pure}}_{\max}}. \tag{2.30}$$

At this point, a wide class of quantum statistical speeds (again, and any other quantum feature connected to the polarization ϵ) can be used to quantify the speed-up. In view of

its role in Grover's algorithm subject to depolarising noise that we will discuss in Sec. 2.3.2, we now explicitly consider the trace speed as an example.

Trace speed in the unitary algorithm

The trace speed TS (cf. Sec. 2.1.1) is one possible candidate to quantify the speed-up in the unitary version of Grover's algorithm with a pseudo-pure initial state since, by using its definition (Eq. (2.7)), it exactly fulfills Eq. (2.27),

$$\mathrm{TS}(\rho_{\epsilon,\psi}, H) = \|[\rho_{\epsilon,\psi}, H]\|_1 = \left\|\epsilon[|\psi\rangle\langle\psi|, H] + \frac{1-\epsilon}{2^n}[\mathbb{1}, H]\right\|_1 = \epsilon\,\mathrm{TS}(|\psi\rangle\langle\psi|, H). \tag{2.31}$$

In order to maximize the TS over all possible local Hamiltonians H (with restricted spectrum), we will use the fact for pure states, the trace speed is the square root of the QFI, $\mathrm{TS}(|\psi\rangle\langle\psi|, H) = \sqrt{\mathrm{QFI}(|\psi\rangle\langle\psi|, H)}$ [50]. Furthermore, if the state is restricted to the completely symmetric subspace of n qubits[9], the maximal QFI (and thus TS) is obtained for collective local Hamiltonians $H_i = \mathbf{n}\cdot\boldsymbol{\sigma}^{(i)}/2$, where $\mathbf{n}$ is a three-dimensional unit vector, $\boldsymbol{\sigma}^{(i)} = (\sigma_x^{(i)}, \sigma_y^{(i)}, \sigma_z^{(i)})$ are the Pauli matrices[10] of the ith qubit and $H = \sum_{i=1}^n H_i$ [28]. In this case, the maximal QFI for pure states $|\psi\rangle$ is given by the maximal eigenvalue of the matrix $\boldsymbol{\Gamma}$ with entries [28]

$$\Gamma_{ij} = 4\left(\mathrm{Re}\left[\langle J_i J_j\rangle\right] - \langle J_i\rangle\langle J_j\rangle\right), \tag{2.33}$$

where $J_k = \sum_{i=1}^n \mathbf{e}_k\cdot\boldsymbol{\sigma}^{(i)}/2$ is the coherent spin operator in k direction ($\mathbf{e}_k$ is the unit vector in k direction, $k = x, y, z$), and $\langle\cdot\rangle$ is the expectation value with respect to $|\psi\rangle$.

In the following, we will assume that the target state $|\omega\rangle$ is $|\omega\rangle = |0\rangle^{\otimes n}$ which is completely symmetric. This can be seen as a simple relabeling of the basis vectors (if the ith binary component of ω is 1, exchange the computational basis vectors of the ith qubit). More formally, if the computational basis labeling is fixed (e.g., $|0\rangle$ should label the state

[9] A state $|\psi\rangle$ is in the completely symmetric subspace if permuting the qubits does not alter the state. Formally, a state

$$|\psi\rangle = \sum_l \alpha_l \left|\psi_1^{(l)}\right\rangle \otimes \cdots \otimes \left|\psi_n^{(l)}\right\rangle$$

is in the completely symmetric subspace if

$$|\psi\rangle = \sum_l \alpha_l \left|\psi_{\pi(1)}^{(l)}\right\rangle \otimes \cdots \otimes \left|\psi_{\pi(n)}^{(l)}\right\rangle,$$

for all permutations $\pi : \{1,\ldots,n\} \to \{1,\ldots,n\}$. For instance, the state $(|01\rangle + |10\rangle)/\sqrt{2}$ is completely symmetric, but the state $|01\rangle \neq |10\rangle$ is not.

[10] In the computational basis, the Pauli matrices are defined as

$$\sigma_x = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \qquad \sigma_y = \begin{bmatrix} 0 & -i \\ i & 0 \end{bmatrix}, \quad \text{and} \quad \sigma_z = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}. \tag{2.32}$$

with lower energy), the state $|\omega\rangle$ can be transformed to the state $|0\rangle^{\otimes n}$ by applying σ_x (the spin-flip operator) to some of the qubits. Since we maximize the TS over all possible local Hamiltonians, this operation does not change the maximal TS[11], so the dynamics of the maximized TS coincide in Grover's algorithm for any target state. Hence, we can assume that the initial state $|\psi_{\text{in}}\rangle$ is completely symmetric, as well as the Grover operations U_d and U_f (using again that $|\omega\rangle = |0\rangle^{\otimes n}$). Therefore, the state $|\psi_k\rangle$ during any iteration of the pure-state algorithm is completely symmetric, and thus, at any step k, we can use Eq. (2.33) to maximize the TS.

By using the exact expression for $|\psi_k\rangle$, cf. Eq. (2.15), the matrix elements Γ_{ij} can be straightforwardly calculated. For instance, we have

$$\begin{aligned}\langle J_x\rangle &= \frac{1}{2}\sum_{i=1}^{n}\langle\psi_k|\,\sigma_x^{(i)}\,|\psi_k\rangle \\ &= \frac{n}{2}\,\langle\psi_k|\,\sigma_x^{(1)}\,|\psi_k\rangle \\ &= \frac{n}{2}\cos^2[(2k+1)\theta_N] + \mathcal{O}(1/\sqrt{2^n}).\end{aligned} \tag{2.34}$$

In the first line, we used that $|\psi_k\rangle$ is completely symmetric. In the second line, we used $\langle\omega|\,\sigma_x^{(1)}\,|\omega\rangle = 0$, $\langle\omega|\,\sigma_x^{(1)}\,|\perp\rangle = 1/\sqrt{2^n-1}$ and $\langle\perp|\,\sigma_x^{(1)}\,|\perp\rangle = (2^n-2)/(2^n-1)$.

The quadratic terms of Γ_{ij} require a slightly longer calculation: for instance, to calculate $\langle J_x^2\rangle$, we first observe that

$$\langle J_x^2\rangle = \frac{1}{4}\sum_{i,j=1}^{n}\langle\sigma_x^{(i)}\sigma_x^{(j)}\rangle = \frac{1}{4}\left(n\langle\sigma_x^{(1)}\sigma_x^{(1)}\rangle + n(n-1)\langle\sigma_x^{(1)}\sigma_x^{(2)}\rangle\right), \tag{2.35}$$

because $|\psi_k\rangle$ is completely symmetric. We then use that $\sigma_x^2 = \mathbb{1}$, and calculate $\langle\omega|\,\sigma_x^{(1)}\sigma_x^{(2)}\,|\omega\rangle = 0$, $\langle\omega|\,\sigma_x^{(1)}\sigma_x^{(2)}\,|\perp\rangle = 1/\sqrt{2^n-1}$ and (after a few lines of calculation) $\langle\perp|\,\sigma_x^{(1)}\sigma_x^{(2)}\,|\perp\rangle = (2^n-2)/(2^n-1)$, and find

$$\langle J_x^2\rangle = \frac{1}{4}\left(n + n(n-1)\cos^2[(2k+1)\theta_N]\right) + \mathcal{O}(1/\sqrt{2^n}). \tag{2.36}$$

After similar calculations, one obtains

$$\mathbf{\Gamma} = \begin{bmatrix} n+n(n-1)\cos^2\theta_k - n^2\cos^4\theta_k & 0 & -n^2\sin^2\theta_k\cos^2\theta_k \\ 0 & n & 0 \\ -n^2\sin^2\theta_k\cos^2\theta_k & 0 & n+n(n-1)\sin^2\theta_k - n^2\sin^4\theta_k \end{bmatrix} + \mathcal{O}(1/\sqrt{2^n}) \tag{2.37}$$

where we defined $\theta_k = (2k+1)\theta_N$.

[11] Explicitly, note that

$$\max_{\text{local } H}\text{TS}(\sigma_x^{(i)}\rho\sigma_x^{(i)}, H) = \max_{\text{local } H}\|[\sigma_x^{(i)}\rho\sigma_x^{(i)}, H]\|_1 = \max_{\text{local } H}\|[\rho, \sigma_x^{(i)}H\sigma_x^{(i)}]\|_1 = \max_{\text{local } H}\|[\rho, H]\|_1,$$

where we have used that $[\sigma_x^{(i)}\rho\sigma_x^{(i)}, H] = \sigma_x^{(i)}[\rho, \sigma_x^{(i)}H\sigma_x^{(i)}]\sigma_x^{(i)}$, and as unitary transformations (here $\sigma_x^{(i)}$) are basis transformations, they do not change the (basis-independent) trace norm of an operator.

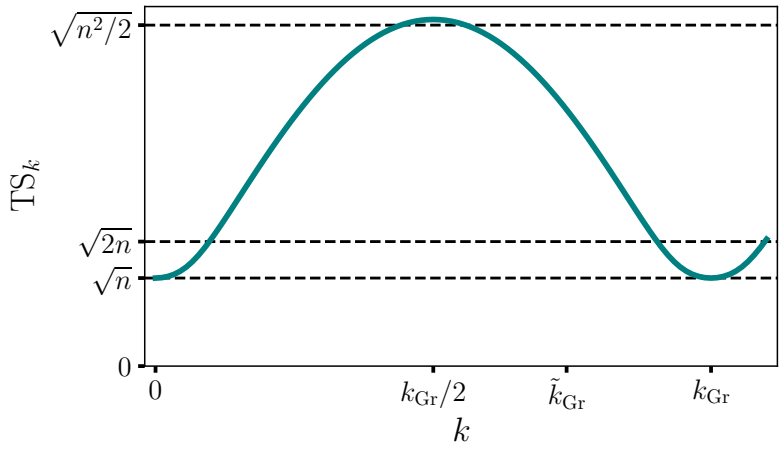

Figure 2.4: The trace speed TS_k of the state $|\psi_k\rangle$ after k iterations of the pure-state Grover algorithm, maximized over all local Hamiltonians with restricted spectrum, see text. The dashed lines correspond to speed limits imposed by separable states ($\sqrt{n}$), bipartite entangled states ($\sqrt{2n}$) and $n/2$-partite entangled states ($\sqrt{n^2/2}$). $n = 30$, $k_{\mathrm{Gr}} = (\pi\sqrt{2^n})/4$, $\tilde{k}_{\mathrm{Gr}} \approx 0.74 k_{\mathrm{Gr}}$.

Finally, the square root of the largest eigenvalue of $\mathbf{\Gamma}$ is given by

$$\begin{aligned}\mathrm{TS}_k^{\mathrm{pure}} &= \sqrt{\frac{1}{8}n\left\{4+n-f(k)n+\sqrt{8[1+f(k)]+n^2[1-f(k)]^2}\right\}} + \mathcal{O}(1/\sqrt{2^n}) \\ &\xrightarrow{n\to\infty} \sqrt{n+\frac{n(n-1)}{2}\sin^2[(4k+2)\theta_N]},\end{aligned} \tag{2.38}$$

where we defined $\mathrm{TS}_k^{\mathrm{pure}} = \max_{\text{local } H} \mathrm{TS}(|\psi_k\rangle\langle\psi_k|, H)$ (the maximization over local H is restricted to $\mathrm{spec}(H_i) = \{-1/2, 1/2\}$, see above) and the auxiliary function $f(k) = \cos 4\theta_k = \cos[4(2k+1)\theta_N]$.

The TS during the pure-state algorithm is shown in Fig. 2.4. Initially ($k = 0$), the trace speed TS_k is $\sqrt{n}$, since the initial state $|\psi_{\mathrm{in}}\rangle$ is separable. Then, TS_k increases and detects entanglement by surpassing the speed limit $\sqrt{n}$ imposed by separable states (cf. Sec. 2.1.1 for how to use the TS as an entanglement witness). At $k = k_{\mathrm{Gr}}/2$, it reaches its maximum

$$\mathrm{TS}_{\mathrm{max}}^{\mathrm{pure}} = \sqrt{\frac{n(n+1)}{2}}, \tag{2.39}$$

witnessing $(n+1)/2$-partite entanglement. Then, TS_k decreases until reaching $\sqrt{n}$ for $k = k_{\mathrm{Gr}}$, since also the target state $|\omega\rangle$ is separable. Note that this behaviour is similar to what was seen in Refs. [162, 163, 165, 166] for different entanglement monotones during Grover's algorithm.

For the unitary algorithm with a pseudo-pure initial state, the maximal TS during the algorithm can be used to quantify the speed-up, $S \propto 1/\mathrm{TS}_{\mathrm{max}}$, cf. Eq. (2.30). In this case, we have $\mathrm{TS}_{\mathrm{max}} = \epsilon\sqrt{n(n+1)/2}$. Thus, $\mathrm{TS}_{\mathrm{max}}$ does not detect entanglement anymore for $\epsilon < \sqrt{2/(n+1)}$. Note that by using the QFI instead of the TS, entanglement is witnessed until $\epsilon < 2/(n+1)$ and thus for smaller polarizations. However, as we discussed before,

Grover's algorithm offers a quantum advantage for all $\epsilon > \epsilon_{\text{crit}} = 2K_{\text{Gr}}/\sqrt{2^n}$. Therefore, the entanglement witnessed by the TS (or the QFI) is not a necessary quantum resource for small quantum advantages of mixed-state versions of Grover's algorithm. Note that it was already noted in Refs. [32, 180, 181] that the advantage persists until $\epsilon \sim 1/\sqrt{2^n}$. However, even for these exponentially small polarizations ϵ, the algorithm's state is entangled (even though the entanglement measure of multiplicative negativity in any bipartition is exponentially small) [181].

We want to emphasize again that to derive our results, we have assumed both that $\epsilon > \epsilon_{\text{crit}}$ (discussed above) and that n is sufficiently large such that the series expansions for small $1/\sqrt{2^n}$ are good approximations of the corresponding quantities. For small n, the analysis has to be performed case by case. Take for instance Grover's algorithm for $n = 2$ qubits. It reaches the target state after only one iteration and, thus, the state $|\psi_k\rangle$ (cf. Eq. (2.15)) is always separable. However, after the application of the oracle unitary U_f and before the application of the Grover diffusion operator U_d, the state is (maximally) entangled.

2.3.2 Grover's algorithm under partial depolarization

We now turn to the more realistic setting of a noisy version of Grover's algorithm. For this purpose, we consider the noise model of the totally depolarizing channel [19], such that the state during the algorithm is always a pseudo-pure state, cf. Eq. (2.22), but the polarization ϵ_k decreases during the algorithm. The algorithm's state after k steps is then

$$\rho_{\epsilon_k,\psi_k} = \epsilon_k \left|\psi_k\right\rangle \left|\psi_k\right\rangle + \frac{1-\epsilon_k}{2^n}\mathbb{1}, \tag{2.40}$$

where $|\psi_k\rangle$ is the pure-state algorithm's state, cf. Eq. (2.15), and ϵ_k encodes both initial polarization and depolarization during the algorithm. This noise model describes the situation that, during each step of the algorithm, the complete information about the quantum state is lost with some probability. It thus represents a worst case scenario and is commonly used whenever the form of the noise is not known [182]. We also want to note that the pseudo-pure noise model plays an important role in the analysis of quantum random circuit sampling (see supplementary information of Ref. [83]).

As before, we assume that $\epsilon_k \gg \epsilon_{\text{crit}} \sim 1/\sqrt{2^n}$ such that we are in the regime of quantum advantage. Thus, we can still use that $\text{TS}(\rho_{\epsilon_k,\psi_k}, H) = \epsilon_k \,\text{TS}(|\psi_k\rangle \langle\psi_k|, H)$ and $p_k = \sin^2[(2k+1)\theta_N]/\epsilon_k$. Note that we again focus on the special case of the trace speed, for reasons that we discuss shortly. Here, as we will see, the decreasing depolarization complicates the computation of both the quantum speed-up and its quantification in terms of resources.

First, one observes that due to the decreasing ϵ_k, the optimal iteration step k_{opt} to measure the state (defined by minimizing $C_{\text{qu}} = \min_k(k+1)/p_k$, cf. Eq. (2.18)) is smaller than

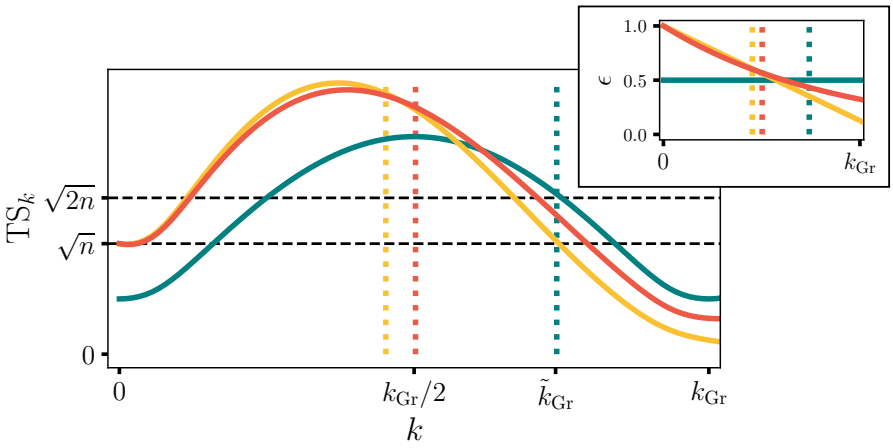

Figure 2.5: The trace speed TS_k of the state ρ_{ϵ_k,ψ_k} after k iterations of the mixed-state Grover algorithm, maximized over all local Hamiltonians. We consider depolarization models leading to a constant polarization (blue), a linearly decaying polarization (yellow) and an exponentially decaying polarization (red), plotted in the inset, that all lead to the same quantum cost and quantum speed-up. For each model, a dotted line indicates the optimal iteration step k_{opt} to measure the algorithm's state. The dashed lines correspond to speed limits imposed by separable states ($\sqrt{n}$) and bipartite entangled states ($\sqrt{2n}$). $n = 30$, $k_{\mathrm{Gr}} = (\pi\sqrt{2^n})/4$, $\tilde{k}_{\mathrm{Gr}} \approx 0.74 k_{\mathrm{Gr}}$.

$k_{\mathrm{opt}} = \tilde{k}_{\mathrm{Gr}}$ in the unitary algorithm [171]. This behaviour can be seen in Fig. 2.5. We show the maximized trace speed TS_k during the algorithm for three different polarization dynamics that all result in the same cost C_{qu} (and thus quantum advantage S). We consider a unitary algorithm with initial pseudo-pure state as in Sec. 2.3.1 (blue), as well as an initial pure state with a linearly decreasing polarization (yellow) and an exponentially decaying polarization (red). The exponential decay of the polarization might best model realistic noise because it corresponds to a constant depolarization rate during the algorithm. All polarization dependencies are shown in the inset. For each model, the optimal step k_{opt} is marked as a dotted vertical line in the corresponding color and clearly varies for the different models. Furthermore, observe that the maximal trace speed $\mathrm{TS}_{\mathrm{max}}$ differs for the three models, even though they provide the same quantum speed-up. Thus, a direct one-to-one connection of the speed-up S with $\mathrm{TS}_{\mathrm{max}}$, as in Eq. (2.30), is not possible.

Even though there is no one-to-one connection between the speed-up S and $\mathrm{TS}_{\mathrm{max}}$, we show in the following that $\mathrm{TS}_{\mathrm{max}}$ can be used to bound the speed-up, thus representing a necessary quantum resource. In particular, we can prove the following theorem.

Theorem 1. *For a mixed-state version of Grover's algorithm that is described by a pseudo-pure state with dynamical polarization, cf. Eq. (2.40), the quantum speed-up S is bounded by the maximal trace speed $\mathrm{TS}_{\mathrm{max}}$ of the state during the algorithm,*

$$S \leq \frac{\sqrt{2^n}}{2K_{\mathrm{Gr}}} \frac{\mathrm{TS}_{\mathrm{max}}}{\mathrm{TS}^{\mathrm{pure}}_{\mathrm{max}}}. \tag{2.41}$$

Here, the trace speed $\mathrm{TS}_{\mathrm{max}}$ is maximized over all local Hamiltonians with restricted spectrum and all steps of the algorithm, n is the number of qubits, $\mathrm{TS}^{\mathrm{pure}}_{\mathrm{max}} = \sqrt{n(n+1)/2}$ and

$K_{\mathrm{Gr}} \approx 0.69$.

Proof. We divide the proof into two cases, $k_{\mathrm{opt}} \geq k_{\mathrm{Gr}}$ and $k_{\mathrm{opt}} < k_{\mathrm{Gr}}$, corresponding to slow and fast depolarization dynamics, respectively. Note that we actually never use that k_{opt} is minimizing the quantum cost C_{qu}, so the results also hold true for any (suboptimal) interruption of the algorithm at any step.

Case 1: $k_{\mathrm{opt}} \geq k_{\mathrm{Gr}}/2$

First, consider the case that the polarization during the algorithm is slow enough such that $k_{\mathrm{opt}} \geq k_{\mathrm{Gr}}/2$ (see, e.g., the exponentially decaying polarization (red) in Fig. 2.5). Here, note that after k_{opt} steps, the corresponding pure-state version of the algorithm has already reached its maximal trace speed $\mathrm{TS}^{\mathrm{pure}}_{\mathrm{max}}$ (which is reached at $k = k_{\mathrm{Gr}}/2$), cf. Fig. 2.4. As ϵ_k is monotonically decreasing, we have $\epsilon_{k_{\mathrm{opt}}} \leq \epsilon_{k_{\mathrm{Gr}}/2}$. Further, the TS_k at any step of the algorithm is bounded by the maximal TS during the algorithm $\mathrm{TS}_k \leq \mathrm{TS}_{\mathrm{max}}$. Finally, we know that $\epsilon_{k_{\mathrm{Gr}}/2} = \mathrm{TS}_{k_{\mathrm{Gr}}/2} \,/\, \mathrm{TS}^{\mathrm{pure}}_{k_{\mathrm{Gr}}/2} = \mathrm{TS}_{k_{\mathrm{Gr}}/2} \,/\, \mathrm{TS}^{\mathrm{pure}}_{\mathrm{max}}$, and thus we find

$$\epsilon_{k_{\mathrm{opt}}} \leq \epsilon_{k_{\mathrm{Gr}}/2} = \frac{\mathrm{TS}_{k_{\mathrm{Gr}}/2}}{\mathrm{TS}^{\mathrm{pure}}_{\mathrm{max}}} \leq \frac{\mathrm{TS}_{\mathrm{max}}}{\mathrm{TS}^{\mathrm{pure}}_{\mathrm{max}}}. \tag{2.42}$$

On the other hand, if one could stop all depolarization at the step k_{opt}, such that ϵ_k is constant for $k > k_{\mathrm{opt}}$, the cost could be further reduced until reaching a minimal cost of $(K_{\mathrm{Gr}}\sqrt{2^n})/\epsilon_{k_{\mathrm{opt}}}$ at $k = \tilde{k}_{\mathrm{Gr}}$, cf. Eq. (2.20). Thus, the quantum cost when measuring the state at step k_{opt} is bounded as $C_{\mathrm{qu}} \geq (K_{\mathrm{Gr}}\sqrt{2^n})/\epsilon_{k_{\mathrm{opt}}}$. Finally, we obtain

$$S = \frac{C_{\mathrm{cl}}}{C_{\mathrm{qu}}} \leq \frac{C_{\mathrm{cl}}}{K_{\mathrm{Gr}}\sqrt{2^n}}\epsilon_{k_{\mathrm{opt}}} \leq \frac{C_{\mathrm{cl}}}{K_{\mathrm{Gr}}\sqrt{2^n}}\frac{\mathrm{TS}_{\mathrm{max}}}{\mathrm{TS}^{\mathrm{pure}}_{\mathrm{max}}} = \frac{\sqrt{2^n}}{2K_{\mathrm{Gr}}}\frac{\mathrm{TS}_{\mathrm{max}}}{\mathrm{TS}^{\mathrm{pure}}_{\mathrm{max}}}. \tag{2.43}$$

Case 2: $k_{\mathrm{opt}} < k_{\mathrm{Gr}}/2$

Now consider the case that the depolarization is so strong that it is advantageous to measure the state at $k_{\mathrm{opt}} < k_{\mathrm{Gr}}/2$ (or that we prematurely interrupt the algorithm). The first step of the above reasoning generally does not hold, i.e., $\mathrm{TS}_{\mathrm{max}}$ generally cannot be bounded by $\epsilon_{k_{\mathrm{opt}}}\,\mathrm{TS}^{\mathrm{pure}}_{\mathrm{max}}$. Here, we have to make use of the explicit form of the TS during the first half of the algorithm, cf. Eq. (2.38). We first recall that we have $\mathrm{TS}_k = \epsilon_k\,\mathrm{TS}^{\mathrm{pure}}_k$ for any step k, and that the quantum cost of stopping the algorithm at step k is $C_k = (k+1)/(\epsilon_k \sin^2[(2k+1)\theta_N])$, cf. Eq. (2.18). We can then bound C_k according to

$$\begin{aligned} C_k &= \frac{k+1}{\sin^2[(2k+1)\theta_N]}\frac{\mathrm{TS}^{\mathrm{pure}}_k}{\mathrm{TS}_k} \\ &\geq \frac{(k+1)\,\mathrm{TS}^{\mathrm{pure}}_k}{\sin^2[(2k+1)\theta_N]}\frac{1}{\mathrm{TS}_{\mathrm{max}}} \\ &= \frac{f(k)}{\mathrm{TS}_{\mathrm{max}}}, \end{aligned} \tag{2.44}$$

where $\mathrm{TS}_{\max}$ is the maximal TS of the algorithm's state until the step k and we denote the factors of the second line by $f(k)$. To show Eq. (2.41), we have to show that $f(k)$ is lower bounded by the constant $b = K_{\mathrm{Gr}}\sqrt{2^n}\sqrt{n(n+1)/2}$. We use the explicit form of $\mathrm{TS}_k^{\mathrm{pure}}$, cf. Eq. (2.38) and, after introducing $r = k/\sqrt{2^n}$, we obtain for large n (and $k \gg 1$)

$$\begin{aligned} f(\sqrt{2^n}r) - b \xrightarrow{n\to\infty} & \frac{\sqrt{2^n}r\sin[4r]n}{\sin^2[2r]} - K_{\mathrm{Gr}}\sqrt{2^{n-1}}n \\ = & \frac{\sqrt{2^{n-1}}n}{\sin^2(2r)}\left(r\sin[4r] - K_{\mathrm{Gr}}\sin^2[2r]\right). \end{aligned} \tag{2.45}$$

Using $K_{\mathrm{Gr}} \approx 0.69$, one quickly checks that $r\sin[4r] - K_{\mathrm{Gr}}\sin^2[2r] > 0$ for $r \in (0, \pi/8]$, which corresponds to $k \in (0, k_{\mathrm{Gr}}/2]$. Thus, $f(k) > K_{\mathrm{Gr}}\sqrt{2^n}\,\mathrm{TS}_{\max}^{\mathrm{pure}}$, and we finally obtain

$$S = \frac{C_{\mathrm{cl}}}{C_k} \leq \frac{C_{\mathrm{cl}}}{f(k)}\,\mathrm{TS}_{\max} \leq \frac{\sqrt{2^n}}{2K_{\mathrm{Gr}}}\frac{\mathrm{TS}_{\max}}{\mathrm{TS}_{\max}^{\mathrm{pure}}}. \tag{2.46}$$

□

We want to make a few remarks about our results.

- Theorem 1 implies that the TS is a necessary and quantitative resource for the quantum advantage in pseudo-pure-state versions of Grover's algorithm: limiting the TS below any fixed value automatically limits the possible quantum speed-up S. As we discussed in Sec. 2.1.1, the TS offers an entanglement witness by surpassing certain speed limits [27, 28, 29, 50], and also represents a widely used measure of coherence [57, 138], see the next bullet point for a further discussion. In this context, we want to recall the discussion of Sec. 2.1.2. Strictly speaking, considering *only* query complexity (as we do with the average number of oracle calls), any quantum algorithm can be implemented in a single multilevel system [92, 143], where no entanglement is present. Note that by considering all physical resources, which corresponds either to changing the definition of the cost, Eq. (2.18), or to restricting any quantum algorithm's implementation to multiple qubits (or qudits), entanglement was shown to be present in the pseudo-pure Grover algorithm offering an advantage (see, e.g., Ref. [181]). Furthermore, as we have seen in the case of a unitary algorithm with pseudo-pure initial states, the quantum advantage still remains for small polarizations for which the TS does not detect entanglement anymore (even though it is present). We conclude that while the TS is a necessary resource for the quantum advantage in the pseudo-pure version of Grover's algorithm, the entanglement detected by the TS is not.

- To discuss the amount of coherence in Grover's algorithm detected by the TS, we want to compare our results to previous work. In Refs. [118, 119], the coherence was shown to decay from a maximally coherent initial state to no coherence in the

target state, where coherence was measured in terms of the relative entropy of coherence and the l_1-norm of coherence. In contrast, the TS detects little coherence for the initial and the target state, but large amounts of coherence halfway during the algorithm, cf. Fig. 2.4. This contrasting behaviour is due to the use of different coherence measures. For instance, the l_1-norm of coherence is maximized by the initial state $|\psi_{\text{in}}\rangle = (\sum_x |x\rangle)/\sqrt{2^n}$, while the TS is maximized for the state $|\psi\rangle = (|0\rangle^{\otimes n} + |1\rangle^{\otimes n})/\sqrt{2}$. The two types of coherence are called speakable and unspeakable coherence, respectively [183, 184]. Speakable coherence is insensitive to which degrees of freedom (i.e., which computational basis vectors) encode the coherence. For instance, the state $|\psi_1\rangle = (|00\rangle + |01\rangle)/\sqrt{2}$ holds the same amount of speakable coherence as the state $|\psi_2\rangle = (|00\rangle + |11\rangle)/\sqrt{2}$. In contrast, unspeakable coherence is encoding-dependent, and distinguishes different degrees of freedom of the Hilbert space regarding their coherence. This coherence measure is appropriate for many quantum technological applications [185]. E.g., in quantum metrology, the state $|\psi_2\rangle$ offers a higher measurement precision than the state $|\psi_1\rangle$. See Ref. [185] for a discussion of speakable and unspeakable coherence. To summarize, our results show that the amount of unspeakable coherence detected by the TS is a necessary resource for the quantum advantage.

- The bound of Theorem 1 is tight. We have seen this explicitly in the case of a unitary algorithm and a pseudo-pure initial state, cf. Eq. (2.30).

- In principle, Theorem 1 might hold also for other quantum statistical speeds QS, as suggested by the fact that the unitary algorithm can be quantified by a wide class of quantum statistical speeds, cf. Eq. (2.30). Furthermore, the reasoning of the theorem in the case $k_{\text{opt}} \geq k_{\text{Gr}}$ holds true for a general quantum statistical speed fulfilling Eq. (2.27). However, in the case of an early measurement of the algorithm ($k_{\text{opt}} < k_{\text{Gr}}$), we have had to make use of the explicit form of the TS. Similar results do not hold for a general QS. In particular, the QFI cannot be used in Theorem 1 in this form.

- We want to briefly put our results in the context of previous work on noisy versions of Grover's algorithm. Commonly, it is seen that a constant depolarization (or noise) rate (independent of n) resulting in an exponentially decaying polarization (cf. Fig. 2.5) leads to a loss of the quadratic speed-up [171, 173, 174, 175, 176, 178]. This loss is also present in our results: for a constant (n-independent) depolarization rate, TS_{max} decreases exponentially with n, and by Theorem 1 so does the quantum speed-up S.

- An open question is whether Theorem 1 can be generalized to other forms of noise or even a general (noisy) search algorithm. We want to briefly comment why con-

sidering a different noise model or even a general noise model renders the analysis very cumbersome. First, most noise models render the calculation of the exact quantum state during the algorithm intractable. Second, for large n, there may be no efficient computation methods of quantities like quantum statistical speeds (see, e.g., Ref. [134]). A common approach is to consider noise such that the quantum state is always completely symmetric, as we also did in the pure-state algorithm in order to make use the coherent spin operators to maximize the TS, cf. Eq. (2.33). However, even though describing pure states and their dynamics only requires an $(n+1)$-dimensional Hilbert space (instead of a 2^n-dimensional Hilbert space for general pure states), completely symmetric mixed states cannot be described only in this Hilbert space. E.g., for two qubits, the completely mixed state ρ_{mix} takes the form $\mathbb{1}/4$ in any orthonormal basis and is thus completely symmetric, and thus $\langle\psi|\,\rho_{\mathrm{mix}}\,|\psi\rangle = 1/4$ for $|\psi\rangle = (|01\rangle - |10\rangle)/\sqrt{2}$ which is orthogonal to the (three-dimensional) completely symmetric subspace of pure two-qubit states. Thus, even a general symmetric noise model can strongly complicate the calculation of the state's evolution and the calculation of, e.g., quantum statistical speeds. See, for instance, Ref. [171] for a treatment of the "simple" symmetric noise model of local depolarization during Grover's algorithm (instead of a totally depolarizing noise model that was assumed in our analysis).

2.4 Conclusions and outlook

In this chapter, we have addressed and contributed to the topic of necessary resources for quantum computational advantage. First, we have summarized previous results and common difficulties in the identification of central quantum resources in quantum computation. We have then focused on Grover's search algorithm which is one of the rare quantum algorithms that offers a provable and quantifiable quantum advantage (in terms of query complexity).

For a general pseudo-pure-state version of Grover's algorithm, we have found that the quantum speed-up with respect to a classical search is bounded by the maximal trace speed (TS) that occurs during the algorithm, where the TS is optimized over all local Hamiltonians with bounded spectrum. The TS can further be used both as an entanglement witness and a coherence measure. The main result given in Theorem 1 states that for any pseudo-pure-state version of Grover's algorithm, the quantum speed-up S is bounded by the maximal trace speed $\mathrm{TS}_{\mathrm{max}}$ that occurs during the algorithm according to

$$S \leq \frac{\sqrt{2^{n-1}n(n+1)}}{2K_{\mathrm{Gr}}}\,\mathrm{TS}_{\mathrm{max}},$$

where n is the number of qubits and K_{Gr} is a constant. Thus, the TS is a necessary resource

for the quantum speed-up in pseudo-pure-state versions of Grover's algorithm. Note that, as the TS is an example of a quantum statistical speed, it can be measured or bounded in experiments, see Sec. 2.1.1. For small amounts of noise, the TS witnesses multipartite entanglement. However, the speed-up persists even when the TS does not detect entanglement anymore. Also, implementing the algorithm in a single multilevel quantum system does not use any entanglement (but requires exponential amounts of other physical resources that are ignored in query complexity). On the other hand, for any implementation, the TS represents a measure of coherence. Thus, our results can be interpreted in that the coherence measured by the TS is a necessary quantum resource for the quantum speed-up of pseudo-pure-state versions of Grover's algorithm (when *only* considering query complexity).

Different possible directions for future research are suggested by our results. First, can one generalize our results to other noise models that cannot be described by pseudo-pure states, or even to a general noisy quantum search algorithm? Second, can one prove similar connections of quantum statistical speeds to the quantum advantage of other quantum algorithms? A restricting factor is that the quantum advantage should be quantifiable which is commonly not the case, as the best classical algorithm to solve the computational problem might not be known (e.g., in the case Shor's algorithm [16]). Quantifiable quantum advantages are known for other oracle-based algorithms like, e.g., Bernstein–Vazirani [15], Deutsch–Jozsa [90] or Simon [91] algorithms.

To conclude, the quest for the origin of quantum computational advantage remains to appear highly problem dependent [32]. The approach via quantum statistical speeds represents a new point of view that can potentially create new insights to this unresolved topic. The importance of the TS in Grover's algorithm, together with the role of quantum statistical speeds in other quantum technologies, shows that quantum statistical speeds represent a promising candidate to recognize useful quantum features and quantum resources in quantum technologies.

Chapter 3

Verification: Postselection strategies for genuine multipartite nonlocality

Contents

In this chapter, we address the question of under what conditions one can postselect measurement correlation data to verify genuine multipartite nonlocality (GMN). For this purpose, we first give an extensive introduction to nonlocality in Sec. 3.1 covering a brief introduction, the main definitions for the bipartite and multipartite Bell scenario, some applications of nonlocality and different loopholes that hinder its validation in experiments. In Sec. 3.2, we focus on the problem of postselecting measured data before using them in a Bell nonlocality test, and prove conditions under which even a collective postselection can be used to verify GMN. The proof relies on the methods of causal diagrams which we will thus first introduce in Sec. 3.2.1. Finally, the result is applied in a specific scenario to show

that one can create genuine three-partite nonlocality from independent particle sources in Sec. 3.2.3.

3.1 Bell nonlocality

The phenomenon of Bell nonlocality (which we will simply refer to as nonlocality in the following) and its implications [8, 9] have been among the most disputed topics in quantum physics [69]. Nonlocality is demonstrated by the violation of Bell inequalities that can be derived using a few assumptions that hold for "classical" physics. Due to the strong conceptual consequences that arise from the violation of Bell inequalities, but also in view of its practical applications, two main questions have been driving discussions and research about nonlocality:

1. Was a Bell inequality really violated in an experiment?

2. Assuming that Bell inequalities are violated, which of the initial assumptions was wrong?

The first question (that constitutes the main focus of nonlocality research in recent years) opens the field of so-called loopholes for nonlocality and loophole-free Bell tests. It is mostly motivated from two different viewpoints. From a philosophical perspective, one might not want to drop any of the initial assumptions that describe our intuitive understanding of (classical) reality, and use these loopholes as an escape route to keep them. More importantly, from a technological and application perspective, nonlocality represents a central ingredient for many provably secure quantum technologies such as, e.g., quantum key distribution (QKD). Loopholes create possibilities for malicious adversaries or eavesdroppers to corrupt the security promises of these quantum technologies. In Sec. 3.2, we will address such a loophole, and we thus provide a brief overview of the different loopholes in Sec. 3.1.4, see also Refs. [24, 186].

The second question illustrates that physicists have different preferences about which of the classically fundamental principles is not correct in our universe. This decision is strongly influenced by which "interpretation of quantum mechanics" is favoured. One can roughly group the opinions in two camps [69]. The "operationalist" camp (see, e.g., Ref. [187]) identifies determinism and locality as crucial assumptions (besides the free will assumption). Since quantum mechanics respects the no-signalling principle that can be seen as an operational form of a locality condition, they usually conclude that determinism is the faulty assumption. On the other side, there is the "realist" camp (see, e.g., Ref. [188]) that, demanding from physical theories to also give explanations of reality by means of Reichenbach's principle [189], argues that the only assumption (besides free will) is local causality (confusingly often also termed locality). Consequently, this camp concludes that

one necessarily has to discard the principle of local causality. A detailed comparison and discussion of the two viewpoints as well as a possible agreement is given in Ref. [69]. In the following, we will simply speak of nonlocality (as short for Bell nonlocality) whenever Bell inequalities are violated.

3.1.1 Bipartite nonlocality

Both sets of initial assumptions mentioned above eventually lead to a local hidden variable (LHV) model that should describe all experimentally observed correlations. We first consider the simplest Bell scenario consisting of two measurement parties, Alice and Bob, where we assume that each party can choose between two measurement settings that each correspond to a two-outcome measurement. A preparation device creates two particles that are then separated and sent to the different parties. On her particle, Alice (Bob) can choose between measuring observables x_1 and x_2 (y_1 and y_2) and, without loss of generality, obtains an outcome $a = \pm 1$ ($b = \pm 1$) [1]. Note that the assumption that each party has two measurement choices with two possible outcomes per measurement is not the most general bipartite Bell scenario [24], as we will also briefly mention below. A general LHV model for the joint probability for the outcomes is given by (compare Eq. (1.2))

$$P_{ab|xy} = \sum_{\lambda \in \Lambda} P_\lambda P_{a|x\lambda} P_{b|y\lambda}, \tag{3.1}$$

where λ is a discrete LHV and $\sum_\lambda P_\lambda = 1$. The LHV λ is called hidden because, in each run of the measurement, its value is hidden from the experimental parties either through experimental limitations or even in a fundamental sense. It can be thought of as a label of the initially prepared two-particle state that is created with probability P_λ. This state, together with the measurement choice x of Alice, fully determines the probability that Alice observes a, and similarly for Bob. This represents the locality assumption, as both parties are supposed to measure simultaneously, and thus, assuming locality, Bob's measurement choice and his outcome cannot influence Alice's outcome. By means of λ, Alice's and Bob's outcomes can be correlated. However, this correlation is due to a local interaction in the past (when the particles were prepared), and there are no direct nonlocal influences from Alice's measurement choice and outcome to Bob's outcome and vice versa. Note that to write Eq. (3.1), it was also assumed that Alice's and Bob's measurement choices are independent from the LHV, $P_{\lambda|xy} = P_\lambda$, which is usually referred to as the free will assumption.

[1] Note that during this thesis, random variables will be denoted as capital letters (e.g., X as Alice's choice of measurement setting, or A as her outcome) while their possible values as lowercase letters (e.g., Alice measures observable x_1 and observes the outcome a), independently of whether the variable is operator-valued or real-valued. Furthermore, we write

$$P_{a|x} \equiv P(a|x) \equiv P(A = a|X = x)$$

because, later, this more compact notation proves advantageous.

Correlations that can be described by the LHV model of Eq. (3.1) fulfill certain restrictions, the so-called Bell inequalities. For the bipartite case described above, the best known and widely used example of such a restriction is the CHSH inequality [190]

$$|\langle A_1 B_1\rangle + \langle A_2 B_1\rangle + \langle A_1 B_2\rangle - \langle A_2 B_2\rangle| \leq 2, \tag{3.2}$$

where $\langle A_i B_j\rangle = \sum_{a,b} ab P_{ab|x_i y_j}$ is shorthand for the expectation value of the product of Alice's and Bob's outcomes when they measure x_i and y_j, respectively. To show that the correlations given by Eq. (3.1) fulfill the CHSH inequality, note that by inserting $\langle A_i B_j\rangle = \sum_{\lambda,a,b} ab P_\lambda P_{a|x\lambda} P_{b|y\lambda}$ and using the triangle inequality, it is enough to prove Eq. (3.2) for each λ separately. Furthermore, the set of possible probabilities is convex, and thus, for any λ, maximal values are obtained with extremal probabilities $P_{a|x\lambda} \in \{0, 1\}$ [24]. Therefore, for a given λ, one can set $A_i = \pm 1$ and $B_j = \pm 1$, and with $|A_1(B_1 + B_2) + A_2(B_1 - B_2)| \leq 2$, Eq. (3.2) follows.

We want to mention that, in the above case that both parties have two measurement choices with each two possible outcomes, the CHSH inequality (3.2) is the only nontrivial Bell inequality (trivial inequalities correspond to positivity conditions such as $P_{a|x\lambda} \geq 0$) [24]. If the measurement parties have more than two measurement choices, or if the different measurements have more than two possible outcomes, many non-equivalent Bell inequalities can be constructed. Furthermore, we note that every inequality can be "lifted" to a Bell inequality for more parties, measurements, or outcomes by merging the latter [191], a trick that is an important tool when dealing for finite detection efficiencies, cf. Sec. 3.1.4. All non-equivalent inequalities have to be checked if one wants to show that the correlations can be described by a LHV model. To the extreme, the problem of deciding whether correlations are local or not in a bipartite setup with two outputs but many measurement choices was shown to be a NP complete problem [192] (compare to the hardness-result of the separability decision [64] mentioned in Sec. 1.3). For an overview of several bipartite Bell inequalities we refer to Ref. [24].

If the observed correlations of Alice's and Bob's results violate a Bell inequality such as Eq. (3.2), they can conclude that there is no LHV model that describes their system. In this case, we say that they could demonstrate nonlocality. This conclusion can build the foundation for several quantum technological tasks, cf. Sec. 3.1.3. Furthermore, as discussed above, the demonstration of nonlocality has strong conceptual consequences and also implies the presence of other quantum resources such as entanglement or EPR-steerability, cf. Sec. 1.3.

3.1.2 Multipartite nonlocality

In a Bell scenario with more than two parties, there is a richer diversity of Bell tests that can be performed by the experimental parties. For simplicity, here we consider the three-

partite case, but all definitions can be easily generalized to the n-partite scenario. First, the parties can ask whether their observed statistics show nonlocality in any form by the exclusion of a LHV model, similar to the bipartite case. The general LHV model for three parties is given by

$$P_{abc|xyz} = \sum_{\lambda\in\Lambda} P_\lambda P_{a|x\lambda} P_{b|y\lambda} P_{c|z\lambda}, \tag{3.3}$$

where the third party Charlie measures z and obtains outcome c. Assuming Eq. (3.3), different three-partite Bell inequalities can be derived. As an example, the CHSH inequality can be generalized to the Mermin inequality [193]

$$|\langle A_1 B_1 C_2\rangle + \langle A_1 B_2 C_1\rangle + \langle A_2 B_1 C_1\rangle - \langle A_2 B_2 C_2\rangle| \leq 2. \tag{3.4}$$

If the measured statistics violate Eq. (3.4), the experimental parties can again exclude a description of their correlations in terms of a LHV model (3.3). In the multipartite case though, there are different levels of possible nonlocal structures, in contrast to the bipartite case: further distinctions about the amount of multipartite nonlocality are possible (similar to the discussion of k-partite entanglement, cf. Sec. 1.2.1). Say, for instance, that Alice and Bob share a nonlocal state which is correlated to Charlie's state only by a LHV. This behaviour is surely nonlocal but not maximally so, and is called *two-way local*. If three-partite correlations are more nonlocal than two-local, they are said to be *genuinely multipartite nonlocal*. Formally, genuinely multipartite nonlocal correlations are correlations that cannot be written as a hybrid local-nonlocal hidden variable model

$$P_{abc|xyz} = \sum_{\lambda_1\in\Lambda_1} P_{\lambda_1} P_{bc|yz\lambda_1} P_{a|x\lambda_1} + \sum_{\lambda_2\in\Lambda_2} P_{\lambda_2} P_{ac|xz\lambda_2} P_{b|y\lambda_2} + \sum_{\lambda_3\in\Lambda_3} P_{\lambda_3} P_{ab|xy\lambda_3} P_{c|z\lambda_3}. \tag{3.5}$$

Here, the complete set of LHV Λ is divided into three disjoint subsets Λ_j ($\Lambda = \Lambda_1 \cup \Lambda_2 \cup \Lambda_3$). The set Λ_j dictates which two parties can share bipartite nonlocal correlations, e.g., Bob and Charlie for Λ_1 in Eq. (3.5). If the probabilities cannot be written in this way, the correlations cannot be described by any mixtures of two-way local correlations. We want to mention that the bipartite nonlocal terms are often restricted to the no-signalling principle to make the definition consistent in an operational framework [194]. For more details see below.

When assuming the hybrid local-nonlocal model (3.5), one can derive new Bell inequalities that test for genuine multipartite nonlocality (GMN). One example that we will employ in Sec. 3.2.3 is the Svetlichny inequality [195]

$$\Big|\langle A_1 B_1 C_1\rangle + \langle A_1 B_1 C_2\rangle + \langle A_1 B_2 C_1\rangle + \langle A_2 B_1 C_1\rangle - \langle A_2 B_2 C_1\rangle - \langle A_2 B_1 C_2\rangle - \langle A_1 B_2 C_2\rangle - \langle A_2 B_2 C_2\rangle\Big| \leq 4. \tag{3.6}$$

It can be proven by noting that it represents a combination of two CHSH inequalities (3.2) conditioned on the third party's measurement, see Ref. [24]. The violation of Eq. (3.6)

demonstrates the maximal amount of nonlocality for this three-partite setup. GMN represents the building blocks for many proposals of multipartite quantum technologies such as quantum networks and quantum conference key agreement, cf. Sec. 3.1.3 for more details.

Three-partite nonlocality can be classified even further [194, 196, 197]. In the hybrid local-nonlocal model of Eq. (3.5), no restrictions on the bipartite nonlocal terms were given. In particular, this allows correlations in which both parties not only share nonlocal resources but also arbitrarily communicate (after their measurements). If each party has to produce its output immediately after its measurement, one has to consider that either (i) the parties' measurements represent spacelike separated events, or (ii) the parties perform their measurements in a specific (not necessarily predetermined) timelike sequence. In the case of spacelike-separated measurements, the resulting correlations have to fulfill the no-signalling principle [198]. For instance, the bipartite nonlocal correlations $P_{bc|yz\lambda}$ in Eq. (3.5) now need to fulfill[2]

$$P_{b|yz\lambda} = P_{b|y\lambda} \qquad \text{and} \qquad P_{c|yz\lambda} = P_{c|z\lambda}, \tag{3.7}$$

where $P_{b|yz\lambda} = \sum_c P_{bc|yz\lambda}$ etc. Due to these restrictions, the hybrid local-nonlocal model can describe less correlations and, therefore, GMN can be demonstrated in more cases. In the case that the measurements are performed in a timelike sequence, signalling from the earlier to the later experiments is generally allowed. Thus, the correlations are only restricted by one of the no-signalling conditions in Eq. (3.7), for each two parties. As imposing a weaker restriction on the possible correlations, its describing capacity is thus between the general hybrid model Eq. (3.5) (that allows for two-way signalling) and the model restricted by the no-signalling conditions.

It can be shown that the differences between the classes of no-signalling, one-way signalling and two-way signalling correlations can be observed in the data: the no-signalling conditions can be employed to derive stricter Bell inequalities [197]. On the other hand, a violation of a Bell inequality derived from the general hybrid model Eq. (3.5) implies GMN in all the above definitions. In Sec. 3.2, we will use the Svetlichny inequality, cf. Eq. (3.6), to demonstrate GMN but we will use the no-signalling principle explicitly to prove results on how experimental data may be postselected.

We want to mention that while a violation of the Svetlichny inequality (3.6) (or its no-signalling generalizations) is needed to demonstrate genuine three-partite nonlocality (implying also genuine three-partite entanglement), a demonstration of genuine three-partite entanglement can also be achieved by the simpler Mermin inequality (3.4) [199]. This is

[2] This typically-used definition of no-signalling [194, 196, 197] actually says that, even if the LHV λ was known, the no-signalling principle would be fulfilled. It implies the more operational no-signalling condition

$$P_{b|yz} = P_{b|y} \qquad \text{and} \qquad P_{c|yz} = P_{c|z}$$

that does not include hidden variables.

because in the demonstration of multipartite entanglement, it is assumed that all correlations are created by measurements of quantum states. In particular, the nonlocal terms of Eq. (3.5) have to be of the form, e.g., $P_{bc|yz\lambda_1} = \mathrm{tr}[E_b F_c \rho_{\lambda_1}]$, where E_b (F_c) is a POVM element corresponding to the y-measurement of Bob (z-measurement of Charlie), and ρ_{λ_1} is a quantum state. This represents an even stronger restriction to the hybrid model than the signalling conditions, making it possible to prove a Mermin-type inequality. Here, the violation of the Mermin inequality (3.4) has to be stronger ($2\sqrt{2}$ instead of 2) [199]. This insight has been used to create a device-independent verification of genuine multipartite entanglement in Ref. [200]. In summary, the Svetlichny inequality is only needed when all possible two-way local distributions have to be ruled out, even those that cannot be generated by local measurements on two-partite entangled quantum states. For a discussion and further details, see Ref. [201].

Alternative definitions of genuine multipartite nonlocality

Before we move on, we want to mention recent developments towards an alternative definition of GMN (and genuine multipartite entanglement) [54]. This line of reasoning has originated from inconsistencies in the above (in the following called "original") definitions when considering a generalized Bell scenario. Here, the parties can share several independent states among different subgroups of the parties, such that, eventually, the total quantum state contains $M > n$ subsystems which are distributed among the n parties. Each party can then perform joint measurements on its subsystems. In contrast, in the standard framework of the Bell scenario, there is one central source that distributes a quantum state consisting of n fundamental subsystems among the n parties. In the generalized scenario, it was found that one can generate the original GMN (i.e., violations of inequalities such as Eq. (3.6)) from sharing merely bipartite entangled states between the parties [202]. Thus, when allowing for the generalized scenario, the original definition of GMN does not correspond to a resource that is closed under tensor products: a single maximally entangled Bell state, $|\psi\rangle = (|00\rangle + |11\rangle)/\sqrt{2}$, shows no genuine three-partite nonlocality, but two copies of it can be used to activate it, where one of the three parties holds a qubit of both maximally entangled states in its lab [202]. Further, it was shown that any network that distributes bipartite pure entangled states shows GMN [203]. We note that the same weakness applies for the definition of genuine multipartite entanglement (cf. Sec. 1.2.1), which again can be activated by distributing bipartite entangled states in a network [204].

The above weakness of the original definitions of GMN and entanglement led to the proposal of alternative definitions, resolving some of the apparent paradoxes or inconsistencies in the original definitions [54]. The crucial insight is that, in the original approach to quantify entanglement [23], the class of free operations (i.e., the operations that do not create resourceful objects; cf. Sec. 1.2.3) was taken to be the set of Local Operations and

Classical Communication (LOCC). This is because the first investigations of entanglement as a resource were made in quantum communication tasks where classical communication is usually allowed and often necessary. In contrast, in Bell scenarios, classical communication should not be a free operation because, by using only classical communication and without sharing any quantum states, it is possible to violate Bell inequalities such as Eq. (3.2). Also, classical communication is even forbidden by relativity if the measurement results are produced in spacelike separation. Therefore, in a Bell scenario, the natural free operations are Local Operations and Shared Randomness (LOSR). In order to consistently combine the notions of entanglement and locality, the classification of entanglement should then only treat LOSR operations as free. These considerations have motivated the definition of the so-called genuine LOSR three-partite nonlocality (entanglement), which are those correlations (states) that cannot be created by bipartite nonlocal correlations (bipartite entangled states) and shared randomness [54]. Note that the original definition and classification of entanglement in terms of LOCC operations is still perfectly appropriate for other important quantum technologies that allow for classical communication between the parties, such as, e.g., superdense coding [25] and quantum teleportation [26].

By means of the alternative LOSR definitions, several paradoxical situations can be clarified. One example is the nonlocality demonstration by Hardy [205] that requires that the quantum state is entangled but not maximally entangled. This constitutes an apparent paradox since a maximally entangled state can be converted to a non-maximally entangled state by means of LOCC. The paradox is resolved by noticing that maximally entangled states cannot be converted to non-maximally entangled ones by means of LOSR, and thus they represent nonconvertible resources. A different example of the same type is that less entangled states allow for smaller detection inefficiencies in Bell experiments [206], see Sec. 3.1.4, or that less entangled states can generate higher production rates of certified random numbers [207].

A second clarification from the alternative definitions addresses the situation encountered above: in the original definitions, bipartite entangled states suffice to create genuine multipartite entanglement and nonlocality. A specific paradoxical example is that some correlations that can be created from measurements of the GHZ state, $|\psi\rangle = (|000\rangle + |111\rangle)/\sqrt{2}$, cannot be created from measurements of two maximally entangled states[3] [208]. This is in apparent contradiction to the fact that, by means of LOCC, the two pairs could first be converted to the GHZ state[4] and then measured. This process

[3]Here we actually meet another class of operations, the Local Operations (denoted as LO; not to confuse with the local oscillator), that are usually assumed to convert the quantum state in a Bell scenario to outcome probability distributions. Thus, the above statement is that by means of LO, the GHZ can be converted to certain correlations (termed the Mermin box [54, 208]), while two maximally entangled states can not.

[4]Say Alice shares a maximally entangled state with Bob and a second one with Charlie. After she locally creates a GHZ state (allowed by LOCC and LOSR), she can use each entangled state to teleport one of the

is not possible in terms of LOSR operations, and thus the paradox is resolved.

In the remainder of this chapter, we will return to the original setups and definitions of GMN and entanglement, corresponding to one central source that distributes n fundamental subsystems among the n parties. A generalization of our results to the new definition of GMN is an interesting direction for future investigation.

Network nonlocality

In this context, we also want to mention a second alternative and recently widely pursued notion of nonlocality, so-called network nonlocality [209]. Also here we consider the generalized Bell scenario, i.e., several independent sources that are shared between different subgroups of the parties (see, e.g., Fig. 3.12 for the triangle scenario discussed later), and each party can perform a joint measurement on its subsystems. If the observed correlations cannot be described in a LHV model, such that each independent source is described by a independent LHV (see Eq. (3.40) for an example below), the correlations are termed network nonlocal. This kind of nonlocality was first considered in the entanglement swapping protocol [210, 211], and then also examined in several other network structures, such as, e.g., the triangle scenario [209, 212, 213] that we will discuss in more detail in Sec. 3.2.3, or the instrumental scenario [214, 215]. Network nonlocality has been experimentally verified in specific networks [214, 216, 217, 218]. Network nonlocality has also led to intriguing foundational results, e.g., that the Hilbert spaces that describe quantum mechanics need to be complex Hilbert spaces [219].

The assumptions of independent sources is more restrictive than that of a central source, such that network nonlocality can be proven for some correlations that can be described with a LHV model in the standard Bell setting. For instance, some correlations that do not take any measurement inputs can be shown to be network nonlocal [209, 212], while they would always be describable by a LHV model in the standard setting. However, while the notion of network nonlocality naturally captures the structure of many standard experimental setups in quantum technologies such as quantum networks [220, 221], the characterization and detection of network nonlocality has proven challenging. The main difficulty is that, in contrast to the standard version of nonlocality, the set of network local correlations is not convex. Therefore, proving the network nonlocality of a specific scenario requires advanced methods such as, e.g., the identification of logical contradictions of the probabilities with the network LHV model [212, 213]. For more methods and introductions of network nonlocality, see Refs. [208, 222].

GHZ state's qubits to a distant party by means of quantum teleportation [26]. This teleportation process makes use of classical communication, so it is not a free operation in LOSR.

3.1.3 Applications of nonlocality

In this section, we provide a rough overview of the importance of nonlocality and GMN for quantum technologies. For a more detailed review we refer to Ref. [24].

The first application that comes to mind when considering nonlocal correlations is an instantaneous transmission of information by means of these correlations. This application is however excluded by the no-signalling principle (cf. Eq. (3.7), for a formal proof see Ref. [198]). Note that a violation of the no-signalling principle in quantum mechanics would represent an incompatibility with the concept of causality in special relativity. Nevertheless, nonlocality can be used to reduce the communication complexity of certain protocols. The standard scenario is that Alice and Bob receive two n-bit strings x and y, respectively, after which Bob is asked to compute a function $c(x, y)$. In the classical communication setting, it is generally necessary that Alice has to send all her information x (i.e., n bits) to Bob. In Ref. [223], it was shown that the amount of information that Alice has to send to Bob can be reduced if the parties share nonlocal resources. This is achieved by translating a thought experiment by Greenberger, Horne and Zeilinger (GHZ) [224] to a communication problem of the above form. Building on the EPR paradox [7], the GHZ thought experiment offers an alternative to Bell inequalities for demonstrating nonlocality [224, 225] and will be described in detail in Sec. 3.2.3. It was even shown that the violation of any (multipartite) Bell inequality offers a quantum advantage in communication complexity, and thus nonlocality is the necessary and sufficient quantum resource for these tasks [226]. We also mention that nonlocality enables one to obtain a finite zero-noise channel capacity from a noisy classical channel [227], a capability that can also be achieved with the use of the quantum resource of indefinite causal order [228] (here, the channel can even have a zero capacity to transmit quantum information [229]). For a review of the role of nonlocality in communication complexity, see Ref. [230].

The most prominent applications of nonlocality are found in the field of quantum cryptography. Here, the establishment of a provably secure communication is reduced to the task of sharing a secret key between the communication parties that cannot be corrupted or intercepted by an eavesdropper. The seminal proposal that made use of nonlocality for this task was provided by Ekert in 1991 [12]: in a slight generalization of the CHSH setup, two parties share many copies of the maximally entangled state $|\psi\rangle = (|00\rangle + |11\rangle)/\sqrt{2}$, and in each round each party can choose from three different measurement settings. After many measurement rounds on their quantum states, the parties publicly share all measurement settings. By comparing their settings, they observe which measurement rounds they can use for a violation of the CHSH inequality, cf. Eq. (3.2), and publicly share the outcomes of the corresponding measurements. A violation of the CHSH inequality then proves that their shared quantum states were not intercepted by an eavesdropper, a conclusion that is based on the no-cloning theorem [231]. Finally, they can use the measurement rounds for

which they have chosen the same measurement setting to establish the secret key (if the parties use the same measurement setting, their outcomes are perfectly correlated).

Later it was realized that the violation of Bell inequalities implies an even stronger statement: not only can one exclude the presence of an eavesdropper, but one can even deduce the state which was shared by the parties from only observing the correlations. This insight was called self testing [232] and was first noted by the observation that a maximal violation of the CHSH inequality proves that the measuring parties share a maximally entangled state (up to local isometries) [233], see also Ref. [54] for a reformulation of self testing in the context of LOSR resource theories. This result was later made robust (i.e., tolerant to experimental imperfections) [234] and represents the central building block of device-independent quantum technologies. Most prominently, device-independent QKD that requires no assumptions on the physical devices (other than that they behave according to quantum mechanics) was developed [235, 236, 237] and made fully general and noise resilient in Ref. [238]. A second famous application of this line of research is the creation of quantum random number generators in which the randomness is certified by Bell inequality violations [239].

In this context, we want to mention another surprising result in device-independent quantum technologies. The field of delegated quantum computation addresses the problem of how a classical client can ever verify that a quantum computer (e.g., on the cloud) correctly performs a given quantum computation that it is asked to do. Furthermore, one can even ask for "blind" quantum computation in which the quantum computer cannot deduce which quantum computation it is performing. These seemingly impossible problems can be solved by making use of nonlocality [240]: the classical client interacts with two entangled quantum computers (that are not allowed to communicate, as usual in LOSR), asking both of them for the (encrypted) quantum computation of interest and, at the same time, performing a Bell test to verify that the quantum computers really use the correct quantum states. By hiding which of the two tasks is demanded in a specific interaction, the client forces the quantum computers to honestly perform the quantum computation, as otherwise the client detects that the Bell inequality is not violated. Note that, eventually, a direct verification of a single quantum computer by a classical client was proven based on post-quantum cryptography (classical cryptography that cannot be broken by quantum computers) [241].

There are also direct applications of nonlocality in the computational power of quantum computers. In the measurement-based model of quantum computation [109], all computational power is based on the nonlocal correlations of the initial cluster state [242]. In particular, the initial state is defined by means of eigenvalue equations of correlation operators between the different cluster sites. Note that these defining equations resemble (but generally differ from) the eigenvalue equations used in the GHZ-type verification of nonlocality [224]. Also recall the role played by nonlocality in quantum advantage of

shallow quantum circuits [121], cf. Sec. 2.1.1.

Applications of multipartite nonlocality

While multipartite nonlocality is ubiquitous in modern quantum technologies (similar to multipartite entanglement), there are only few results that explicitly highlight its role in quantum advantage. This might be due to its relatively short history, as well as the above mentioned debates about its definition. Here, we give a short overview of quantum technologies that make explicit use of GMN.

The first example of an application is the protocol of quantum secret sharing [243], where a classical secret is encoded in an n-partite GHZ state, $|\psi\rangle = (|0^{\otimes n}\rangle + |1^{\otimes n}\rangle)/\sqrt{2}$, that is then shared between n parties. Note that a shared GHZ state always indicates the presence of GMN for all definitions of GMN (see Ref. [201] for the original definition and Ref. [208] for the LOSR definition). If all parties collaborate they can reveal the secret, however, if any subset of parties tries to access the secret without consulting the remaining parties, they will not gain any information about the secret. Furthermore, the presence of an eavesdropping attack is made visible by comparing the final measurements. The problem of sharing a quantum secret is addressed in Ref. [244].

Similar to the bipartite case, multipartite nonlocality can be used to share a secret key between n parties (see, e.g., Ref [245] for a device-dependent and Ref. [246] for a device-independent protocol). This multipartite quantum key sharing is also called quantum conference key agreement. While quantum conference key agreement can also be established using only bipartite entanglement and nonlocality [247], a protocol that makes use of genuine multipartite entanglement can achieve higher key rates [248]. Reference [249] achieves a quantum conference key agreement that uses multipartite entangled states with large local Hilbert space dimension, in contrast to the GHZ state approach of Refs. [246, 248]. See Ref. [250] for a review of quantum conference key agreement.

Another class of future quantum technologies that are recently attracting large attention and rely on multipartite entanglement and nonlocality are quantum networks and the quantum internet [220]. Similar to quantum conference key agreement, many different parties (also called nodes) share entangled quantum states in order to implement quantum communication or quantum computation protocols. For realistic large-scale quantum networks, two different building blocks are crucial. The first ingredient are quantum repeaters (based on entanglement swapping) that enable the sharing of highly entangled quantum states among many parties and along large distances, see Ref. [251] for the bipartite and Ref. [252] for the multipartite proposal. The second ingredient are quantum memories that represent the local quantum system at each node and that have to be connected by means of quantum channels (usually photons) with the other nodes, see Refs. [253, 254] for reviews of the state-of-the-art technology. For an overview of recent results and open

questions about quantum networks see, e.g., Ref. [221].

Finally, similar to the discussion in Sec. 1.3, the detection of GMN is a direct witness of genuine multipartite entanglement. This can easily be seen by observing that the set of correlations that can be created by bipartite entangled states and local measurements can be described by a hybrid local-nonlocal hidden variable model, cf. Eq. (3.5) (all probabilities now need to stem from measurements of quantum states). We also want to note that in many investigations of quantum resources, the main focus was placed on entanglement. However, finding that entanglement is a necessary resource does not rule out that also nonlocality is a necessary resource (since the set of nonlocal states is included in the set of entangled states). It might just be that the explicit role of nonlocality was not considered. One example for an explicit consideration of multipartite nonlocality as a resource was offered in Ref. [255], where it is shown that a certain class of genuinely nonlocal states can be used to achieve the optimal Heisenberg limit of quantum metrology. By the large attention experienced by multipartite nonlocality recently, one can expect that other explicit applications of multipartite nonlocality will be identified in the future.

3.1.4 Loopholes in nonlocality experiments

In this section, we provide a brief overview of different loopholes that can affect nonlocality experiments. Loopholes have been shaping much of theoretical and experimental research in Bell nonlocality in the last decades. In the following, we first address the two main loopholes of locality and detection inefficiency that have to be checked in any loophole-free nonlocality demonstration. Then, we focus on the specific loophole of postselection that, while in theory only affecting a subclass of nonlocality experiments, practically is always present in experiments, even beyond the ones that test for nonlocality. In Sec. 3.2, we will then discuss new results that address the postselection loophole. For more extensive reviews and for discussions of other loopholes in Bell experiments, we refer to Refs. [24, 186].

Locality loophole

One central ingredient of derivation and discussion of Bell inequalities is that the hidden variable model (cf. Eq. (3.1)) is "local": if the measurement parties are far distant from each other, special relativity dictates that their measurement choices and observed measurement outcomes can have no causal influence on the measurement outcomes of the other parties, since any future event can only be influenced by earlier events that have happened in the event's past light cone. If the measurement parties are not sufficiently separated (the experiments have to be spacelike separated[5]), the measurement devices of

[5]Two points in spacetime are spacelike separated if neither of them is in the past light cone of the other, or, in other words, if they cannot be connected by timelike (or null) paths. This means that any signal from

the different parties could communicate (possibly in a hidden way) to produce any desired correlation of measurement results, without violating causality and special relativity [256]. Furthermore, not only must the different experiments be performed at spacelike distance, but the decision of measurement setting also has to be done at spacelike distance. Note that this further condition is connected to the assumption that measurement settings are chosen independently of the LHV, called the free-will assumption [257] or the measurement-independence assumption [258, 259].

The very first experiments on Bell inequalities [260] did not insist on distant measurement parties and thus were affected by the locality loophole. The first experiment to address the locality loophole was performed in 1982 by the group of Aspect [261] who used fast-switching measurement settings to measure two entangled photons. This method, however, did not represent a truly random setting and was vulnerable to the so-called memory loophole, i.e., that due to the periodicity of the settings, detectors with memory could predict future measurement settings [256]. Later, measurement settings were chosen both locally and randomly by means of quantum random number generators[6] [262], cf. Sec. 3.1.3. Finally, both the measurement choices and the photon generators were put in spacelike separation [257], addressing also the free-choice loophole. In this context, we also want to mention recent Bell experiments that used random numbers generated by humans [263] or by cosmic light coming from distant stars in the Milky Way [264] or even from quasars [265]. Strictly speaking, the free-choice loophole can never be closed completely but one has to rely on ever more complex machineries to explain why the settings are not chosen randomly.

Detection loophole

The loophole that has been most extensively discussed is the so-called detection or efficiency loophole. In all experiments, some parts of the measurement process of a quantum system are not perfectly efficient, meaning that with some probability, the physical system can either remain undetected or can be detected with a false output (e.g., in photon-counting detectors, a single photon has a finite probability to be detected as no photon or as two photons). Commonly, if only one of the parties produces an outcome of its detector, the event is discarded. This kind of postselection of observed events opens the possibility for LHV models that are able to describe the complete correlation of all data, even though the postselected data violate a Bell inequality. This loophole was observed already in the early days of nonlocality [266, 267]. While there are rigorous methods to close the detection loophole as we discuss below, the main approach in experiments is to assume that the

one point to the other would necessarily have to travel faster than light.

[6] Note that this approach actually relies on circular reasoning: the Bell inequality is violated if one assumes that the measurement settings are chosen randomly, and the certification of randomness of a quantum random number generator is based on the violation of Bell inequalities.

detectors satisfy "fair sampling": the postselected data that are observed by the experimental parties should fairly represent (i.e., show the same correlations etc.) the data that would be observed when using perfectly efficient detectors [268, 269]. If, on the other hand, the detection probability depends on the chosen detector setting, fair sampling is not fulfilled and apparent Bell inequality violations can be described by a LHV model [266].

However, due to the key role of nonlocality for secure tasks and cryptography, closing every loophole in nonlocality tests is of crucial importance. In particular, fair sampling is not a condition that is necessarily fulfilled in real experiments: by implementing detection schemes that do not satisfy fair sampling, fake violations of Bell inequalities have been created in experiments, see, e.g., Refs. [270, 271] for violations of Bell inequalities using classical optics. As a variation of the detection loophole, we also mention Ref. [272] for a direct hack of commercial QKD devices using the implementation loophole, i.e., that the physical implementation of the detection differs from the physical model used in the security proof.

The only way to rigorously close the detection loophole is to somehow take into account the no-detection events. Two different approaches have been developed for this purpose. First, one can include all measured data in the analysis of the Bell inequality, without neglecting any undesired data, see, e.g., Ref. [273]. Here, the easiest approach is to merge the no-detection events with one of the two initial outcomes and test the initial (two-outcome) Bell inequality [191]. In this way, a detection efficiency of $\eta > 2/(1+\sqrt{2}) \approx 0.83$ is required [274] (here, η is the probability of the detection of the incoming particle). Interestingly, it was shown that non-maximally entangled states (i.e., two-qubits states whose single-qubit reduced density matrices are not completely mixed) allow to close the detection loophole even for smaller detection efficiencies ($\eta > 2/3$) [206]. We also note that by using higher dimensional states and more measurement settings, lower detection efficiencies can be used (see, e.g., Ref. [275]). For multipartite Bell scenarios, even arbitrarily low detection efficiencies allow for closing the detection loophole [276]. The second approach consists of postselecting double detections but, at the same time, sharpening the Bell inequality (i.e., increasing the threshold value) accordingly. The threshold efficiency for a demonstration of nonlocality in this approach is again $\eta = 2/(1+\sqrt{2})$ (this threshold was derived assuming symmetric detection efficiencies [267] or conditional detection efficiencies [277]).

Experimentally, the only way to close the detection loophole is thus to use detectors with sufficiently high detection efficiencies. While experiments with ions typically enjoy high detection efficiencies and could be used to close the detection loophole earlier [278, 279], this was achieved only in the last decade for photonic implementations [280, 281]. These works were able to close the detection loophole without addressing the locality loophole. Finally, in 2015, a series of experiments was able to simultaneously close locality and detection loopholes, using experiments based on photons [282, 283] and on nitrogen va-

cancy centers [284].

Postselection loophole

The final loophole that we touch upon in this short overview is the postselection loophole. It strongly resembles the detection loophole that we have discussed above. In contrast to the detection loophole though, it is not present in all experiments. Furthermore, this loophole is not originating in technological or implementation limitations, but remains even for a hypothetical ideal noiseless experiment (or, for the theoretical proposal), and thus represents a different kind of loophole.

The postselection loophole principally affects any experiment in which some of the collected data are neglected and thus the remaining data are postselected. Note that the postselection that is performed due to inefficient detection is naturally included in this loophole, however, in context of Bell experiments, one usually refers to the postselection loophole whenever the postselection is not due to experimental imperfections. In the field of statistical inference, postselecting data is long known to create the possibility to change data structures and correlations. This is known as the selection bias. A famous example of selection bias is Berkson's paradox where two random variables become negatively correlated after postselecting the complete sample [285]. To give a contemporary example of Berkson's paradox, we highlight its role in false conclusions concerning possible influences of COVID-19 risk [286]. For instance, there have been claims that smoking can protect against COVID-19 infection. Obviously, both smoking and COVID-19 infection can lead to hospitalization, see Fig. 3.1. However, by only considering hospitalized patients for the study or the statistic, COVID-19 and smoking seem to be anti-correlated even though they may have been independent in the full population. The reason is that if a hospitalized patient does not have COVID-19, she necessarily has to suffer from a different disease, increasing the probability for a smoking-related disease in comparison to the probability in the total population. This highlights the importance of the selection bias in statistics and why samples for medical studies have to be chosen carefully to correctly represent the full population.

Since violations of Bell inequalities originate from correlations of the observed data, Bell tests are especially susceptible to postselection. In particular, the selection bias opens

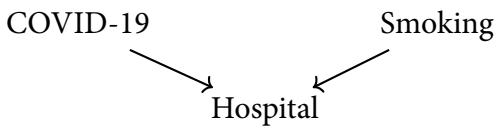

Figure 3.1: Illustration of Berkson's paradox: both COVID-19 and smoking can lead to hospitalization. By considering only hospitalized patients, COVID-19 and smoking thus seem anti-correlated even if they are independent in the full population.

the possibility that postselected measurement data violate a Bell inequality even though the complete set of experimental observations can be described by a LHV model. In this case, the conclusion that the experimental system demonstrates nonlocality is false. Until recently, the only postselection that was considered to be provable safe from the selection bias was a locally decidable postselection [273]: if each party can decide postselection by observing its measurement result only (and not awaiting confirmation from the other experimental parties), and every party's decision on postselection coincides, the postselected statistics is valid for Bell tests. This is because if, say, Alice is able to decide the postselection on her own, Bob's measurement choices can have no influence on the postselection due to relativity, as otherwise this would correspond to superluminal signalling. Recently, less restrictive postselection strategies have been proven valid for the demonstration of multipartite (but not genuine multipartite) nonlocality [287]. In particular, it was shown that in the multipartite Bell scenario, the postselection strategy is safe if it can be equivalently decided by excluding any party, or, in other words, if it can be equivalently decided by any all-but-one parties. We will further comment on these results in Sec. 3.2.2.

An important special case of the postselection loophole is the so-called time-coincidence loophole [186, 288]. This loophole affects nonlocality demonstrations that are created by so called time-bin entanglement originating from a proposal by Franson [289]. Here, two particles are sent to two experimental parties but the exact emission time of the particles is unknown. Therefore, when the parties detect the particles in coincidence, there are different possibilities for the exact time that the particles have been created (and for which path they have traveled). The corresponding probability amplitudes of these possibilities have to be added up enabling interference and also violation of Bell inequalities. However, the violation is only visible if the events with a coincident detection are postselected, a set that only represents a fraction of the complete set of observed events, even if perfect detection efficiencies and a noiseless implementation are assumed. Leveraging on the selection bias, a LHV model was developed that could describe all measurement statistics of the initial Franson experiment [290]. Here, the trick was that the postselection of an event was both dependent on the LHV and the local measurement settings of the parties.

Since explicit LHV models have been found that describe Franson's proposal, one has to conclude that it cannot demonstrate nonlocality. However, different variations of the proposal have been developed that are not affected by the postselection loophole. For instance, a hugged version of Franson's original setup was proposed [291] and experimentally implemented [292, 293] that, by means of causality, can rule out that the detector settings influence the postselection, so a LHV as in Ref. [290] is forbidden. This new proposal is very similar to an old idea by Yurke and Stoler [294] that has inspired the development of the widely used methods of entanglement swapping. We will introduce and discuss the setup by Yurke and Stoler in detail in Sec. 3.2.3.

Finally, we want to note that postselection is an ubiquitous tool in physics and science in general. In almost all experiments, undesired data or unsuccessful runs are neglected before proceeding to the post-processing of data. In principle, each postselection should be carefully analyzed to conclude that no selection bias alters the data. An example can be given in the field of quantum metrology. In the last decade, much attention has been enjoyed by the field of weak values, a phenomenon that creates paradoxically large measurement outcomes by means of postselecting measurements [295]. This phenomenon was then argued to be beneficial for amplification in experiments (so-called weak-value amplification) [296, 297, 298, 299]. However, it was later shown that the average precision (in particular, the average Fisher information) of a measurement cannot be increased with the help of postselection [300, 301]. Exceptions to this result might arise if detectors are subject to saturation limitations [298] or if the postselection is performed during the measurement process and not in the post-processing after the complete measurement procedure [302]. In the latter example, postselection can yield advantages if the final measurement is costly and can be circumvented by postselection.

3.2 Postselecting statistics for genuine multipartite nonlocality certification

We now address the postselection loophole for the specific task of demonstrating genuine multipartite nonlocality (GMN). For this purpose, we first specify the problem setting. Then, in Sec. 3.2.1, we insert a small digression to causal inference and causal diagrams, methods that we build upon to show our later results. In Sec. 3.2.2, we provide the main results of this chapter, showing that in a multipartite Bell scenario, even certain postselection strategies that require communication between the parties are valid to demonstrate GMN. Finally, in Sec. 3.2.3, we apply the results to an old proposal by Yurke and Stoler to show that one can generate genuine three-partite nonlocality from independent particle sources. The results of this chapter are published in Ref. [2].

As we have seen in Sec. 3.1.2, GMN is a phenomenon that requires that the quantum resource of nonlocality is shared in a collective way between all parties of a multipartite Bell scenario. Furthermore, we have discussed that certain postselection strategies that make use of communication between the parties can be used to fake nonlocality – rendering the postselection strategy invalid for any nonlocality experiments. One is tempted to conclude that, in order to demonstrate GMN, all kinds of postselection that require any amount of communication should be invalid. Surprisingly, this is not true: as we show in this section, there are postselection strategies that build on communication between (parts of) the experimental parties which are valid to demonstrate GMN.

As an illustrative example, consider a group of experimental parties who have shared

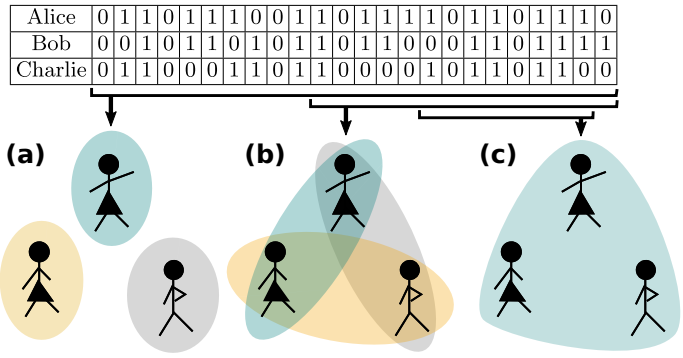

Figure 3.2: After three parties have performed a series of measurements, they come together to analyze their results. Even if the whole set of data does not violate any Bell inequality (a), different kinds of postselection can transform the data to violate Bell inequalities demonstrating general (non-genuine) three-partite nonlocality (b) or genuine three-partite nonlocality (c). Which kinds of postselection are safe for nonlocality demonstration and cannot be affected by selection bias? The Figure is taken from Ref. [2].

a quantum state and who have performed a series of local measurements. After having finished their measurements, they come together to analyze the results and search for correlations, in order to eventually violate a Bell inequality that demonstrates GMN. Unfortunately, from the complete set of measurement results, the corresponding Bell inequalities are not violated. Can they do more?

As we have sketched in Sec. 3.1.4, they know that they are allowed to postselect parts of the data if this postselection can be decided by each party only examining its own measurement results and without consulting other parties about theirs. Of course, the measurement runs that are postselected by each party in this way must coincide for all parties (otherwise, correlations between different measurement runs, i.e., different independent quantum states, are analyzed, so Bell inequalities cannot be violated). As mentioned above, this kind of postselection is called locally decidable (since the postselection could also have been decided directly after recording the measurement, in a spacelike distance to all other parties) and is known to be safe [273]. If locally postselected data still do not violate the Bell inequalities, does it imply that their correlations will never demonstrate GMN?

Next, assume that a different postselection strategy that cannot be decided locally by every party, but that requires some amount of communication between the parties after the measurements, and that results in a set of data that violates the Bell inequality. This situation is plausible since, generally, more postselection (and more complicated postselection) creates or uncovers more extreme correlations in the data. However, as we discussed above, postselection potentially opens the postselection loophole, i.e., the loophole that nonlocal correlations are mimiced by means of a selection bias. In Sec. 3.2.3, we provide a class of valid postselection strategies for GMN tests, where even partial communication between the parties is allowed to decide the postselection.

The problem at hand is sketched in Fig. 3.2, for the smallest situation of three parties,

Alice, Bob and Charlie. The parties perform measurements and obtain a series of outcomes (here depicted as binary outcomes). By using different postselection procedures, their joint data may demonstrate different levels of nonlocality. For instance, by considering all data, no Bell inequality is violated, cf. Fig. 3.2(a). Furthermore, by postselecting same parts of the data, a Bell inequality that separates local from nonlocal behaviour (e.g., Eq. (3.4)) may be violated Fig. 3.2(b). However, as we discussed in Sec. 3.1.2, a violation of such a Bell inequality does not demonstrate the strongest form of nonlocality, GMN, as hybrid local-nonlocal models such as Eq. (3.5) can also lead to violations. Finally, the parties postselect the data further such that a Bell inequality that tests for GMN is violated (e.g., Eq. (3.6)), see Fig. 3.2(c). In Sec. 3.2.2, we introduce valid postselection strategies for the demonstration of GMN and also mention previous results in the literature about valid postselection strategies for general (non-genuine) multipartite nonlocality.

As can be seen from the introduction to Bell inequalities in Sec. 3.1.1 and Sec. 3.1.2, the crucial ingredient for Bell inequalities are LHV models or hybrid local-nonlocal hidden variable models. These models can be conveniently illustrated as causal diagrams to highlight the causal connections between different variables. Furthermore, the effect of postselection on the statistics and causal structures can be analyzed by means of the causal diagrams and the tool set of d-separation rules [75]. These rules constitute the methods we use to prove the validity of postselection strategies in Sec. 3.2.2, and thus we first introduce the central notions of and the rationale behind causal diagrams in the following.

3.2.1 Digression: Causal diagrams

In this section, we will briefly introduce the concepts of causal inference and causal diagrams. For an extensive introduction we refer to Ref. [75]. Causal inference is the task of inferring causal structures from the observation of correlation data. It is built on Reichenbach's principle which states that if two random variables are dependent (meaning that their joint probability distribution does not factorize), then one is a cause of the other or they share a common cause. The result of a causal analysis of the data is a directed acyclic graph (DAG) that describes the causal influences between the different random variables. Furthermore, having established a DAG, one can use the tool set of d-separation (not to confuse with k-separability, cf. Sec. 2.1.1) to quickly infer conditional independence even for complicated causal structures, a method that we will utilize in Sec. 3.2.2.

Consider the case of three ordered random variables (RVs), A, B and C. We want to find a possible causal explanation of the joint probability distribution P_{abc}, meaning that we want to understand all causal influences between the variables. For the causal analysis, it is important to assume that the distribution P_{abc} is stable, meaning that by slightly changing the parameters describing the causal influences we do not alter the causal structure [75]. In other words, we impose that if some joint probability distribution factorizes, this is not

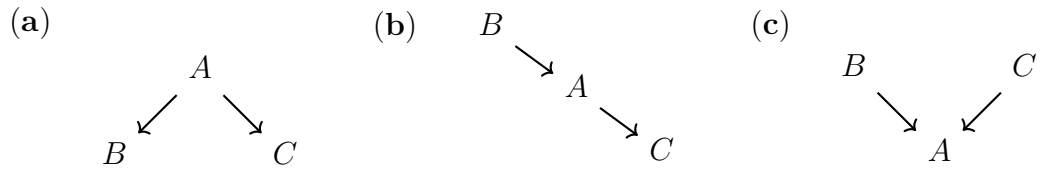

Figure 3.3: Different directed acyclic graphs (DAGs) for three random variables (RVs) A, B and C. A joint probability distribution that factorizes as Eq. (3.8) is compatible with both DAGs in (a) and (b), and merely by the joint distribution one cannot tell which is describing the true causal influences. The DAGs in (a) and (b) are equivalent because they have the same connected pairs of nodes and show the same v-structure (in particular, they have no v-structure). This is in contrast to the DAG in (c) that has the same connected pairs but a different v-structure, making it nonequivalent to the others and incompatible to the distribution of Eq. (3.8).

due to a fine tuning of the causal influences.

To construct the DAG corresponding to P_{abc}, we need to introduce the notion of (Markovian) parents. In short, the parents of the RV X_j are a minimal set of the predecessor RVs (in the given order of the RVs) that render X_j conditionally independent of the other predecessors. For instance, if we have that $P_{c|ab} = P_{c|a}$ and $P_{c|a} \neq P_c$, the set $\{A\}$ constitutes the parents of C. Finally, to draw a DAG, one represents each RV as a node and draws an arrow from the parent nodes to the corresponding RV. As an example, if the joint distribution P_{abc} factorizes as

$$P_{abc} = P_a P_{b|a} P_{c|a}. \tag{3.8}$$

we draw the DAG shown in Fig. 3.3(a).

If the joint distribution P_{abc} is strictly positive (so all different configurations of RVs have non-zero probability), the DAG is unique for a given ordering of the RVs [75]. But how should one choose the correct ordering of the RVs? In fact, also the DAG shown in Fig. 3.3(b) describes the same factorization as Eq. (3.8). This can be seen directly by using Bayes theorem, $P_a P_{b|a} = P_b P_{a|b}$. In Ref. [303], it is shown that reversing an arrow always leads to an equivalent DAG whenever the "v-structures" of the DAG remain the same, where v-structures refer to nodes where two arrows end and the arrows' beginning nodes are not directly connected. For instance, the DAG shown in Fig. 3.3(c) is not equivalent to the ones in Fig. 3.3(a,b) because it has one v-structure, whereas the former ones have none. To summarize, given only the correlation data of RVs, the DAG can be uniquely determined up to directions of some causal influences.

d-separation rules

The above motivation and application of DAGs is to find a (possible) causal explanation and a corresponding visualization of correlated data. However, DAGs are also useful in the inverse direction: if one is given a DAG that describes the causal influences between

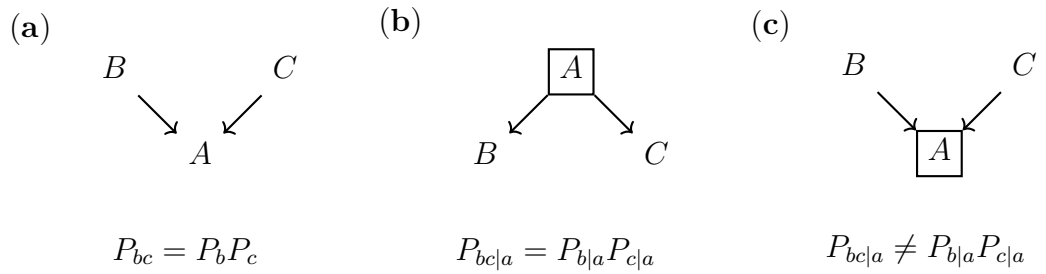

Figure 3.4: Illustration of the three d-separation rules for three random variables. The rules state that (a) a collider along a path blocks the path, (b) conditioning on a non-collider blocks the path, and (c) conditioning on a collider unblocks a path (selection bias).

different RVs, one can easily read off which RVs are conditionally independent of others. This is formalized by a small set of rules known as the d-separation rules ("d" stands for "directional"). In the following, we explain the rules with the simple set of three RVs as above. For general proofs and more information, see Refs. [75, 303].

In short, to check if two nodes of a DAG are independent (also when conditioned on other nodes) one has to consider all different paths connecting the two nodes. A path consists of a series of arrows, for arbitrary directions of the arrows. The d-separation rules then dictate whether a given path is blocked or not. If all paths connecting the two nodes of interest are blocked, the nodes are independent. To understand the rules, we need to introduce the notion of a collider, i.e., a node along the path where two arrows end head to head. The d-separation rules are as follows.

Rule 1: A collider (that is not conditioned on) along a path blocks the path. We illustrate this rule with the DAG of Fig. 3.4(a), corresponding to the joint probability distribution

$$P_{abc} = P_b P_c P_{a|bc}. \tag{3.9}$$

We can directly see that the path connecting B and C over the (unconditioned) collider A is blocked since

$$P_{bc} = \sum_a P_{abc} = \sum_a P_b P_c P_{a|bc} = P_b P_c \sum_a P_{a|bc} = P_b P_c. \tag{3.10}$$

Rule 2: Conditioning on a non-collider blocks a path. Note that a non-collider is any node along the path that is not a collider, e.g., a node where the two arrows originate (fork, $B \leftarrow A \rightarrow C$) or a node where one arrow ends and a second originates (chain, $B \rightarrow A \rightarrow C$). For instance, by conditioning on the fork A in Fig. 3.4(b) (a DAG corresponding to Eq. (3.8)) blocks the path connection B and C, as can be seen by

$$P_{bc|a} = \frac{P_{abc}}{P_a} = \frac{P_a P_{b|a} P_{c|a}}{P_a} = P_{b|a} P_{c|a}. \tag{3.11}$$

Note that, commonly, conditioning on a node of a DAG is marked by drawing a box around the node, see the node A in Fig. 3.4(b).

Rule 3: Conditioning on a collider unblocks a path. The easiest example is shown in Fig. 3.4(c) and exactly describes the selection bias discussed before, see Fig. 3.1. An easy example of this rule are two independent binary RVs B and C that determine the third RV A as $a = b \oplus c$, where $\oplus$ represents addition modulo 2 or "exclusive or". Conditioned on A, B and C are now perfectly correlated ($a = 0$) or anti-correlated ($a = 1$).

Before we move on to address postselection in Bell scenarios with the use of causal diagrams and d-separation rules, we want to mention that causal diagrams have been used to derive several results about Bell scenarios. For instance, they have been used to quantify the amount of measurement independence and locality that have to be given up to describe Bell inequality violations with classical causal diagrams [304, 305, 306]. Causal networks also represent the central tool in examinations of network nonlocality [209], see the introduction in Sec. 3.1.2. For instance, they have served to examine instrumental inequalities that build on so-called intervention and can be violated by quantum systems [214, 215], an effect that can be used to certify randomness [217] and that shines light on the interplay of quantum mechanics and our traditional notions of causality.

3.2.2 Valid postselection strategies for genuine multipartite nonlocality certification

Equipped with the tools of causal diagrams and d-separation rules, we are now able to provide valid postselection strategies for the demonstration of GMN. Before we proceed to the exact statements, we first sketch how directed acyclic graphs (DAGs) occur naturally in the study of nonlocality and how they immediately highlight how postselection can lead to problems. Then, we proceed by proving the results about valid postselection strategies for GMN in the three-partite and, finally, in the general n-partite Bell scenario.

DAGs in the bipartite Bell scenario

For simplicity, before considering a multipartite Bell scenario, we briefly discuss causal diagrams and postselection in the bipartite scenario. Recall that, according to the local hidden variable (LHV) model, the joint probability distribution for Alice and Bob recording outcomes x and y when measuring y and z is given by,

$$P_{ab|xy} = \sum_{\lambda \in \Lambda} P_\lambda P_{a|x\lambda} P_{b|y\lambda}, \tag{3.12}$$

where Λ is the set of LHVs, cf. Sec. 3.1.1; Eq. (3.1). Furthermore, the variables have causal influences as depicted in Fig. 3.5(a). These causal influences originate from the structure and the assumptions of the Bell scenario: first, both parties' measurement choices should be free, so nothing can causally influence X and Y. Second, the parties might share some common LHV Λ that can have causal influences on the parties' measurement outcomes

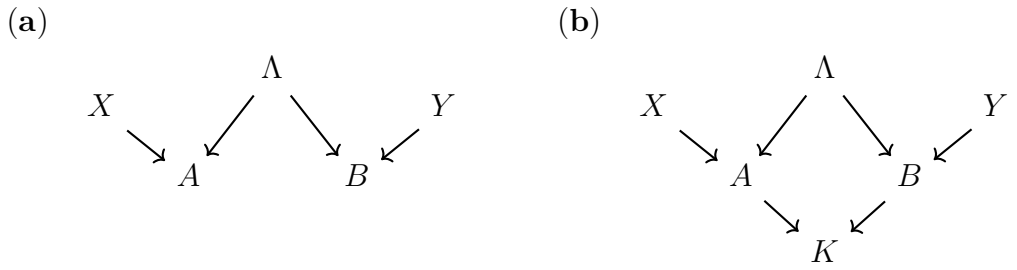

Figure 3.5: DAGs describing the causal structure of the LHV model of the bipartite Bell scenario (a) without postselection, and (b) including postselection K that is decided after communication between the parties, rendering it an invalid postselection strategy for the demonstration of nonlocality.

A and B. Finally, each party's measurement choice can influence its outcome but not the outcome of the other party because the corresponding events are spacelike separated. From this causal structure, one directly derives the joint probability of Eq. (3.12).

We want to emphasize that, similar to the discussion in Sec. 3.2.1, starting from the joint distribution of Eq. (3.12), the DAG in Fig. 3.5(a) is not the only possible causal explanation. However, the DAG follows from the remaining assumptions and the structure of the Bell scenario: first of all, to uniquely construct a DAG, recall that one has to fix the ordering of the RVs. Here, the outcome variables (A, B) must be subsequent to the measurement choice variables (X, Y) and the LHV variable Λ. Further, the free choice assumptions (cf. Sec. 3.1.4) essentially say that one has $P_{xy\lambda} = P_x P_y P_\lambda$, i.e., the LHV and the measurement choices are all independent. Finally, with the joint probability distribution of Eq. (3.12), the DAG of Fig. 3.5(a) follows.

After having understood the intrinsic causal structure of the bipartite Bell scenario, we can now easily discuss the influence of postselection. In the remainder of this section, we will denote postselection by the binary RV K taking the values $k = 1$ for a positive postselection and $k = 0$ if the experimental run is not postselected. As we discussed above, different postselection strategies are possible that can either be valid for nonlocality demonstration or not. In Fig. 3.5(b), we sketch a postselection that is decided when both parties compare their results and then, together, decide if they postselect or not. Therefore, the postselection K is influenced by both measurement outcomes A and B at the same time (in general, they could even postselect due to different measurement choices, which we do not assume here). To see that this kind of postselection is not valid for the demonstration of nonlocality, note that, e.g., due to the postselection on K, there is an open path from X to B: the path $X \to A \to K \leftarrow B$ is open since A is a non-collider (which is not conditioned on) and K is a collider that we condition on. Therefore, the postselection opens the possibility of signalling between the parties. If the parties are allowed to send signals, it is known that without using any quantum effect, the CHSH inequality, cf. Eq. (3.2), can be maximally violated (i.e., it can be violated even stronger than what is allowed by quantum mechanics), see, e.g., Ref. [24]. Thus, a violation of the CHSH inequality after such a

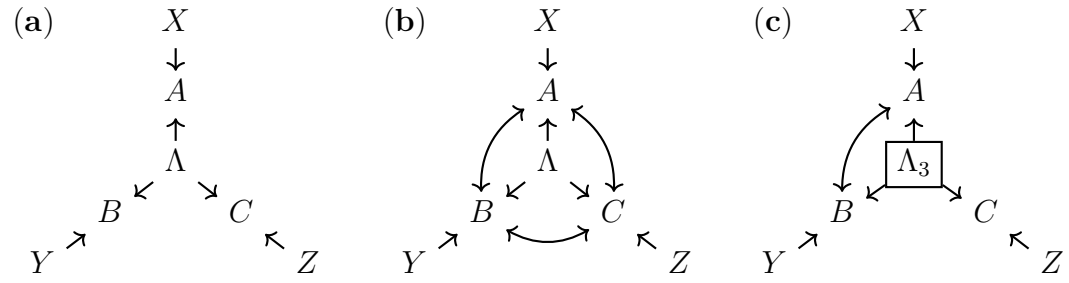

Figure 3.6: Causal diagrams showing the structure of the different hidden variable models in the three-partite Bell scenario, for (a) the LHV model of Eq. (3.3), and (b,c) for the hybrid local-nonlocal hidden variable model of Eq. (3.13). In (b,c), the direct causal influences between the parties' measurement outcomes are drawn as bidirected arrows and are subject to the fine-tuning condition of no-signalling. In (c), by conditioning on the LHV Λ (indicated by the box), the causal diagram further simplifies because now only two parties can share nonlocal correlations.

postselection does not demonstrate nonlocality.

In contrast, we can now easily see why locally decidable postselection is safe: the postselection can now be decided only from information about the outcome A, or, equivalently, from information about B. Hence, there are two valid causal diagrams that describe the situation and that can be used to infer conditional independence: one where K is decided only by A (so there is just the arrow from A to K and not the one from B), and the second one where K is decided by B. In this case, there is no open path that allows for signalling, and one can also rigorously proof with the d-separation rules that the postselected data is a valid candidate to demonstrate nonlocality. The proof is analogous to (and much simpler than) the proofs we provide later, so we do not give full particulars here. In short, by means of the d-separation rules, the postselected statistics $P_{ab|xyk}$ factorizes in the same way as the original LHV model (without postselection), Eq. (3.12), such that it can also be used to derive the Bell inequality.

DAGs in the multipartite Bell scenario

We now consider the three-partite Bell scenario. The corresponding DAGs discussed here can be directly generalized to the case of $n > 3$ parties. First, consider the three-partite LHV model of Eq. (3.3). Here, the DAG is very similar to the DAG of the bipartite scenario and is shown in Fig. 3.6(a): Each measurement party freely chooses its measurement setting (X, Y, Z) which can influence the corresponding measurement outcome (A, B, C). Furthermore, the measurement outcomes can be influenced by the shared LHV Λ.

Similar to the bipartite case, a local postselection is valid to demonstrate nonlocality, while a postselection that requires communication between all measurement parties can mimic any correlations including nonlocality even if no quantum state is shared. However, since we have more than two parties, other postselection strategies can be considered. In particular, in Ref. [287], it was shown that a postselection that can be equivalently de-

cided by any all-but-one parties (by two parties in the three-partite scenario and by $n-1$ parties in the n-partite scenario), without consulting the remaining party for its result, is a valid postselection to demonstrate multipartite nonlocality. Note that this implies that any postselection strategy that can be decided by $m < n$ parties is a valid postselection to demonstrate nonlocality, since it can be trivially extended to a postselection that is decided by $n-1$ parties. The proof of this result is very similar to the proof we provide later for the demonstration of GMN, where we explicitly highlight the differences. We again emphasize that by means of the LHV model described by the DAG in Fig. 3.6(a), one can only demonstrate general multipartite nonlocality, but not the stronger form of GMN.

To demonstrate GMN, the LHV model above must be relaxed to a hybrid nonlocal-local hidden variable model, see Sec. 3.1.2 for a detailed discussion. For clarity, recall from Eq. (3.5) that the joint probability distribution according to the hybrid model is given by

$$P_{abc|xyz} = \sum_{\lambda_1 \in \Lambda_1} P_{\lambda_1} P_{bc|yz\lambda_1} P_{a|x\lambda_1} + \sum_{\lambda_2 \in \Lambda_2} P_{\lambda_2} P_{ac|xz\lambda_2} P_{b|y\lambda_2} + \sum_{\lambda_3 \in \Lambda_3} P_{\lambda_3} P_{ab|xy\lambda_3} P_{c|z\lambda_3}, \tag{3.13}$$

where the LHV Λ is divided into disjoint subsets dictating which two parties can share nonlocal correlations. Note that, additionally, we have the no-signalling conditions, Eq. (3.7), which will play an important role later. The diagram that describes the causal influences in this model is depicted in Fig. 3.6(b). We include the possible nonlocal influences between the parties' outcomes by bidirected arrows (since nonlocal correlations do not have a direction), so strictly speaking, we do not have a directed acyclic graph anymore. We emphasize that this bidirected connection cannot be described by a common cause between the outcomes (this is exactly the definition of nonlocal correlations), in contrast to the usual classical causal diagrams [75]. Even more, these correlations have characteristics that make them hard to incorporate in classical causal diagrams as we discuss below. We also recall that by only assuming Eq. (3.7) without the no-signalling principle, the bipartite contributions generally allow for signalling, which has led to critique of and alternative definitions to the original definition by Svetlichny [194, 196, 197]. In the original definition by Svetlichny [195], causal influences from, e.g., X to B are allowed.

We now touch upon the difficulty of describing the model of Eq. (3.13) with a classical causal diagram. In Fig. 3.6(b), the no-signalling conditions represent a fine-tuning condition on the causal influences: for instance, while there are causal influences $X \to A$ and, due to the nonlocal correlations, there may be influences $A \to B$, there can be no causal influences $X \to B$. This follows from the no-signalling condition $P_{b|xy\lambda} = P_{b|y\lambda}$, cf. Eq. (3.7). Therefore, the causal influences of any classical causal model have to be fine tuned. In fact, it was shown that to describe nonlocal correlations in a classical causal model, it is necessary to use fine tuning, independent of which explanation of nonlocality is used (e.g., superluminal causal influences, or superdeterminism) [304, 307]. As mentioned

above, fine tuning is usually forbidden in classical causal models [75]. This challenges the use of classical causal models and causal inference for the description of nonlocal correlations in general, and for the hybrid nonlocal-local model of Eq. (3.13) in particular. In this context, we want to mention recent ideas to study quantum causal models that naturally include nonlocality without the use of fine tuning [208, 307], an approach that is still in its beginnings. Nonetheless, in the following, we will work with the fine-tuned causal diagram of Fig. 3.6(b) because, while general methods of causal inference should be checked in the presence of fine tuning, we can still meaningfully use the d-separation rules to prove the conditional independence between random variables and thus prove our the results on safe postselection strategies.

If we condition on a specific LHV λ, the DAG further simplifies to show the deeper structure of Eq. (3.13). For instance, if we condition on $\lambda \in \Lambda_3$, we know that the outcome C of Charlie's measurement is only correlated to the outcomes of the other parties via the LHV, cf. Fig. 3.6(c). Alice and Bob still share nonlocal correlations as indicated by the bidirected arrow that, as we discussed above, are subject to the fine-tuning conditions of no-signalling.

Postselection strategies in the three-partite Bell scenario

After having understood the causal structure and the causal diagram for the hybrid local-nonlocal model of Eq. (3.13), we can now examine which kinds of postselection are safe for the demonstration of GMN. For this purpose, we focus here on the Svetlichny inequality, cf. Eq. (3.6), that demonstrates GMN, or equivalently any Bell inequality that is proven by means of the hybrid model, Eq. (3.13). We emphasize that to derive the inequality, the only requirement is that the joint probability distribution $P_{abc|xyz}$ factorizes as in Eq. (3.13). Therefore, if the postselected distribution $P_{abc|xyzk}$ (note that we have conditioned on the postselection variable K) factorizes in the same way, the postselected distribution is also valid to show the inequality. Consequently, if the postselected data violated the inequality, the postselected distribution cannot be written in the hybrid form, implying that also the original joint distribution without postselection does not factorize according to Eq. (3.13), demonstrating GMN.

But how can we show that a given postselection preserves the factorization of the joint distribution? First, we use the general chain rule to write

$$P_{abc|xyzk} = \sum_{\lambda} P_{\lambda|xyzk} P_{abc|xyz\lambda k}. \tag{3.14}$$

Now it is clear that the postselected distribution $P_{abc|xyzk}$ factorizes in the correct way if one can show the conditions

$$\textbf{C1} \qquad P_{\lambda|xyzk} = P_{\lambda|k} \qquad \forall \lambda \in \Lambda,$$

$$\textbf{C2c} \qquad P_{abc|xyz\lambda_3 k} = P_{ab|xy\lambda_3 k} P_{c|z\lambda_3 k} \qquad \forall \lambda_3 \in \Lambda_3,$$

with analogous conditions **C2a** and **C2b** for conditioning on Λ_1 and Λ_2, respectively. If these conditions are fulfilled and the original joint distribution $P_{abc|xyz}$ factorizes according to Eq. (3.13), also the postselection distribution $P_{abc|xyzk}$ factorizes in the same way, i.e., the postselection is valid for Bell tests based on the hybrid local-nonlocal model. Here, we want to mention that if the Bell inequality of interest (see, e.g., Ref. [197]) is derived by explicitly using the no-signalling conditions, Eq. (3.7), a safe postselection strategy must also fulfill conditions such as

$$\mathbf{C3} \qquad P_{b|yz\lambda k} = P_{b|y\lambda k} \qquad \forall \lambda \in \Lambda. \tag{3.15}$$

This assures that also the postselected model fulfills the no-signalling conditions that are needed to derive the Bell inequality. If the conditions are not required to derive the inequality, as is the case, e.g., for Svetlichny's inequality, Eq. (3.6), the condition **C3** is not necessary for a valid postselection.

In the following, we focus on a class of postselection strategies that can be equivalently decided by any subgroup of the experimental parties of minimal size. For instance, in the three-partite case, we consider a postselection that can be equivalently decided by any two parties, meaning that one can write the postselection K as a function of any pair of measurement outcomes, i.e.,

$$k = k_{AB}(a, b) = k_{BC}(b, c) = k_{CA}(c, a). \tag{3.16}$$

Note that this is exactly the all-but-one principle that was shown to lead to valid postselection strategies to demonstrate general (non-genuine) multipartite nonlocality [287]. For more than three parties, we will later consider postselection strategies that do not follow the all-but-one principle. We can now prove that a postselection following the all-but-one principle is valid to demonstrate genuine three-partite nonlocality. The case of the n-partite Bell scenario for $n > 3$ is discussed in the next subsection.

Theorem 2. *In the three-partite Bell scenario, a postselection strategy that can be equivalently decided by any two parties is valid to demonstrate genuine three-partite nonlocality.*

Proof. In the following, we use the d-separation rules introduced in Sec. 3.2.1 to show why a postselection that can be equivalently decided by any two parties fulfills the conditions **C1** and **C2**. First, observe that due to the structure of the postselection, we can produce several different causal diagrams that we can use in the proof. For instance, in Fig. 3.7(a), the postselection K is decided by Alice and Bob. Two further diagrams of this type are valid where the postselection is decided by other pairs of parties. Furthermore, by conditioning on the LHV Λ, there are nine more possible causal diagrams that hold. For instance, in Fig. 3.7(b), we condition on Λ_3 and the postselection is again decided by Alice and Bob. By conditioning on other subsets of Λ and by other postselection groups, we find the other valid diagrams.

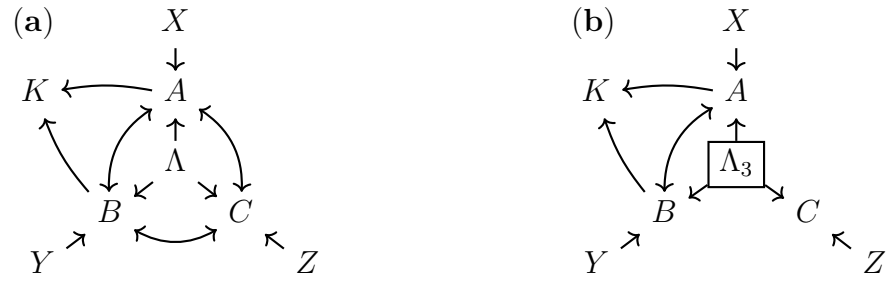

Figure 3.7: Different causal diagrams of the hybrid model (cf. Eq (3.13) and Fig. 3.6) that include a postselection K that can be decided by any two parties. (a) Causal diagram for the unconditioned model where the postselection is decided by Alice and Bob. Two similar diagrams are valid where the other pairs of parties decide the postselection. (b) Conditioned on the LHV, one can construct further causal diagrams. Here, we obtain the shown diagram when conditioning on Λ_3 and for a postselection that is decided by Alice and Bob. There are eight other possible combinations when conditioning on Λ_1 and Λ_2 and where other pairs of parties decide the postselection.

Now consider the condition **C1**. In Fig. 3.8(a), we sketch the causal diagram to prove that

$$P_{\lambda|xyzk} = P_{\lambda|xyk}. \tag{3.17}$$

To show this conditional independence by means of the d-separation rules, we have to check for all possible paths that connect Z and Λ when conditioning on X, Y and K. As before, the conditioning is indicated by boxes. First, observe that the path $Z \to C \leftarrow \Lambda$ is blocked because there is a collider (C, not conditioned on) along the path (d-separation rule 1). Next, consider that path $Z \to C \to A \to K \leftarrow B \leftarrow \Lambda$. Note that here we write $C \to A$ for the causal influence from C to A that is possible by the nonlocal correlations shared by Alice and Charlie, even though in the diagram, we use a bidirected arrow. At first sight, this path appears to be open because all chains along the path (C, A and B) are not conditioned on (d-separation rule 2), and the collider K is conditioned on (d-separation rule 3). However, we are helped by the no-signalling condition (that we sketch as a dotted line in Fig. 3.8(a)): while the causal influences $Z \to C$ and $C \to A$ are possible, causal influences from Z to A are forbidden by the no-signalling principle, cf. Eq. (3.7). Therefore, the complete path is blocked. In the same way, one argues that the remaining path $Z \to C \to B \to K \leftarrow A \leftarrow \Lambda$ is blocked. Notice the crucial role played by the no-signalling condition: if we had only assumed the hybrid model of Eq. (3.13) without the no-signalling conditions (as was done originally by Svetlichny [195]), we could not have proven condition **C1**. Therefore, even if the Bell inequality of interest might not require the no-signalling assumption, it is necessary to prove that the postselection strategy that we discuss is valid.

In a similar way we can use the other causal diagrams, where the K is decided by other combinations of parties, to show that $P_{\lambda|xyk} = P_{\lambda|xk}$ and $P_{\lambda|xk} = P_{\lambda|k}$, and we obtain condition **C1**.

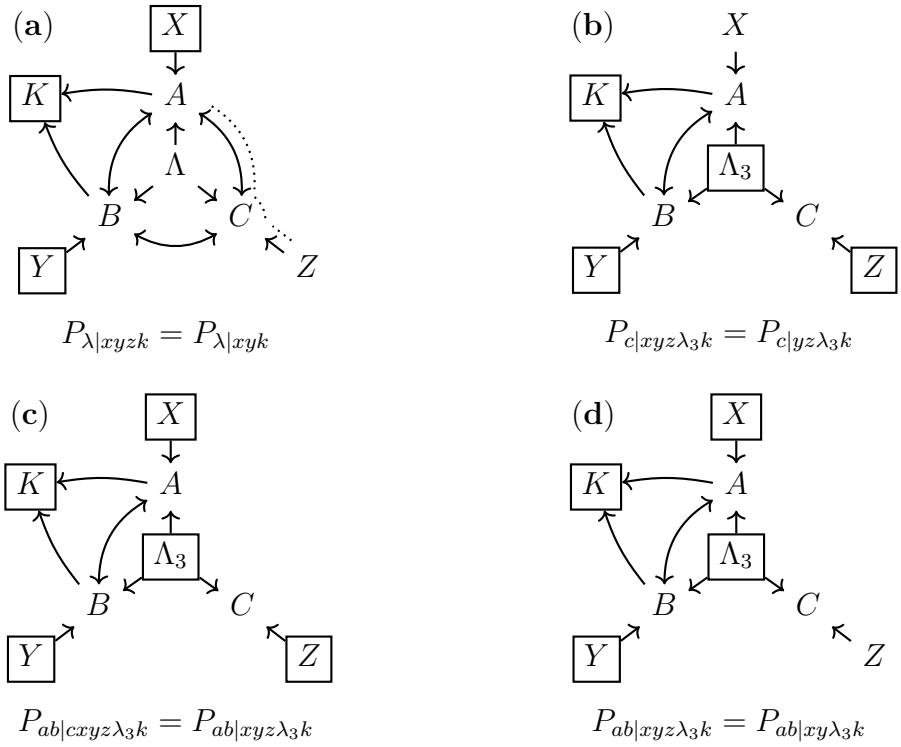

Figure 3.8: Causal diagrams that are used in the proof of condition **C1** (a) and condition **C2c** (b-d). Below the diagrams, we note the conditional independence relation that is demonstrated. In (a), we indicate one of the no-signalling conditions with a dotted line.

Now let's turn to condition **C2c**. First, we use the chain rule to write

$$P_{abc|xyz\lambda_3 k} = P_{ab|cxyz\lambda_3 k} P_{c|xyz\lambda_3 k}. \tag{3.18}$$

Then, we use the causal diagram of Fig. 3.8(b) to prove that $P_{c|xyz\lambda_3 k} = P_{c|yz\lambda_3 k}$. All paths that connect X and C pass through the collider Λ_3 that is conditioned on and are thus blocked. From the same diagram (but removing the condition on Y), one also obtains $P_{c|yz\lambda_3 k} = P_{c|z\lambda_3 k}$.

Next, we use the diagram shown in Fig. 3.8(c) to show that $P_{ab|cxyz\lambda_3 k} = P_{ab|xyz\lambda_3 k}$. As before, every path from C to A or B passes through the collider Λ_3 that is conditioned on and is therefore blocked. We want to emphasize that, for this step, the factorization of the hybrid model is crucial: If there was the bidirected connection between C and A or C and B, the conditional independence would not hold. In contrast, one quickly observes that all other steps required to obtain condition **C2c** can also be proven merely by the no-signalling principle, without using the factorization structure of the hybrid model. Finally, we use Fig. 3.8(d) to show that $P_{ab|xyz\lambda_3 k} = P_{ab|xy\lambda_3 k}$ because all paths connecting Z with A or B go through the conditioned collider Λ_3 (and the unconditional chain C). Combining all steps, we find that condition **C2c** is fulfilled. By using the corresponding diagrams, one also shows conditions **C2a** and **C2b** (or simply note that the hybrid model is symmetric).

□

A few comments are in order.

- We first want to emphasize that, if we want to keep the postselection strategy completely general, Theorem 2 provides the most collaborative kind of postselection strategy that is still safe for the demonstration of GMN: if the postselection strategy is more demanding, i.e., it can only be decided by communication between all parties, the postselected data can mimic nonlocal behaviour even when the latter is not present in the original data, similar to what we have discussed in Sec. 3.1.4 for the bipartite case[7]. To see this, note that already the proof of condition **C1** breaks down because now there are open paths that hinder its demonstration: for instance, the path $Z \rightarrow C \rightarrow K \leftarrow A \leftarrow \Lambda$ (cf. Fig. 3.8) is open since the collider K is conditioned on and the chains C and A not.

- As we mentioned above, one might also want to prove that the postselected model still fulfills the no-signalling conditions, e.g., condition **C3**: $P_{b|yz\lambda k} = P_{b|y\lambda k}$. This can be proven easily in a similar fashion as above: by conditioning on K, Y and Λ in Fig. 3.8, one observes that all paths connecting Z and B are blocked due to the no-signalling conditions between Z and B or between Z and A.

- Now we want to compare Theorem 2 to previous results on safe postselection strategies for the demonstration of (non-genuine) three-partite nonlocality. In Ref. [287], the above techniques of causal diagrams and d-separation rules have been used to prove that postselection strategies are safe for (non-genuine) three-partite nonlocality if they can be decided by any two parties (the all-but-one principle). This condition coincides with our result above, which is surprising since Theorem 2 addresses postselection for demonstrating the stronger form of genuine three-partite nonlocality. In particular, the original model to derive the Bell inequality is the LHV model of Eq. (3.3) instead of the hybrid model of Eq. (3.13). Furthermore, since the LHV explicitly implies the no-signalling condition (in contrast to the hybrid model), the direct role of the no-signalling principle was hidden in Ref. [287]. Thus, it is remarkable that in the three-partite case, safe postselection strategies coincide for the demonstration of multipartite nonlocality and the stronger GMN. As we will see in the next subsection, this is no longer the case for the n-partite Bell scenario with $n > 3$, as here the safe postselection strategies for GMN are more restricted than the ones found in Ref. [287] for general multipartite nonlocality.

- We want to briefly discuss possible applications of Theorem 2. Similar to what was found in Ref. [287] (and for a specific experimental setups also in Ref. [309]), Theo-

[7] An example of such a postselection strategy in the bipartite case is the proposal by Franson (cf. Sec. 3.1.4). For the multipartite case, see Ref. [308] for a discussion of fake (non-genuine multipartite) nonlocality by postselection in the n-partite generalization of Franson's proposal.

rem 2 can be applied in situations where each party has to obtain a certain number of particles (e.g., one) and the total number of particles is conserved. Then, any two parties can communicate if they received the correct number of particles and conclude whether also the remaining party has received the correct number or not. This kind of postselection is ubiquitous in current quantum optical experiments, see, e.g., Refs. [310, 311]. We will discuss an intriguing example in detail in Sec. 3.2.3, where we use Theorem 2 to show that genuine three-partite nonlocality can be created from independent particle sources.

- Finally, we have to emphasize that in all real experiments, one faces the additional problem of finite efficiencies of the detectors (see the detection loophole discussed in Sec. 3.1.4). This impedes the possibilities for postselection that can be decided when excluding one party. In the example above (where each party should receive a certain number of particles), even if two parties obtain the correct number of particles, they cannot be sure that the remaining party does so as well. Therefore, the postselection again has to be decided by all parties together. As we sketched in Sec. 3.1.4, rigorous treatments of finite detection efficiencies are the inclusion of all experimental data (so no postselection) or a sharpening of the Bell inequalities. Commonly though, these rigorous approaches do not yield a violation of the Bell inequality, such that one is forced to rely on the fair-sampling assumption (i.e., that the correlations that were detected correctly represent the correlations that would occur when using perfect detectors). A generalization of the above postselection strategies to situations with finite detection efficiencies is an interesting direction for future work.

Postselection strategies in the n-partite Bell scenario

We now consider postselection strategies for demonstrating GMN in the n-partite Bell scenario. In this case, many of the steps in the proof of safe postselection in the three-partite case remain valid, with one important exception. Consider first the easiest case of $n = 4$ parties. Here, the hybrid local-nonlocal model consists of a mixtures of subensembles in which either (i) three parties share nonlocal correlations and are classically correlated to the last party, or (ii), two pairs of parties share bipartite nonlocal correlations, or (iii) all parties are only correlated by a shared LHV. For the subensembles of type (i) and (iii), one can easily show that a postselection of the previous type, i.e., a postselection that can be equivalently decided by any three parties (all-but-one), preserves the factorization of the respective subensemble. However, when there are two pairs of nonlocally correlated parties, a postselection that is decided by three parties necessarily opens a path between the two groups, and thus the postselected subensemble generally does not factor in the same way.

This situation is sketched in Fig. 3.9. The jth party chooses measurement setting x_j

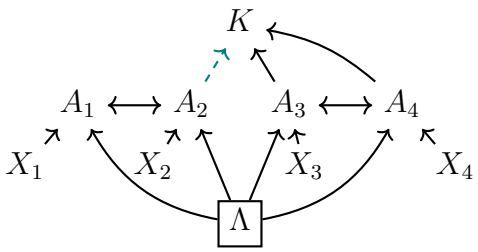

Figure 3.9: Causal diagram of the four-partite Bell scenario, corresponding to the subensemble of the hybrid local-nonlocal model where two pairs share nonlocal correlations (A_1 with A_2, and A_3 with A_4). A postselection K that can only be decided by three parties does not preserve this pairwise structure, but a postselection that can be decided by any two parties does. The critical causal influence from A_2 to K is marked with a dashed arrow.

and observes outcome a_j. In the shown subensemble (that makes part of the hybrid model because it does not show genuine four-partite nonlocality), the pair (A_1, A_2) and the pair (A_3, A_4) exhibit nonlocal correlations, which we again sketch as bidirected arrows. Note that we focus on this subensemble, so we condition on the corresponding LHV set Λ. Since the postselection K is decided by three parties (here, e.g., A_2, A_3 and A_4), conditioning on K must open a path between the two pairs, see the dashed arrow connecting to groups via the conditioned collider K (all paths through Λ are blocked because it is a conditioned fork). Note that if the postselection could be decided by any two parties (so one could delete the dashed arrow from A_2 to K), the groups could be again separated.

We see that safe postselection strategies are more restricted in the n-partite Bell scenario. However, surprisingly, we can still show that certain postselection strategies that require communication between subgroups of parties are valid for the verification of GMN:

Theorem 3. *In the n-partite Bell scenario, a postselection strategy that can be equivalently decided by any $\lceil n/2 \rceil$ ($\lceil \cdot \rceil$ is the ceiling function) parties ("all-but-$\lfloor n/2 \rfloor$") is valid to demonstrate genuine n-partite nonlocality.*

Proof. In the n-partite Bell scenario, the hybrid local-nonlocal hidden variable model of Eq. (3.13) has to be extended to include all possible subensembles of combinations of nonlocal correlations that do not show GMN. Say that the kth measurement party can choose the measurement setting X_k to observe an outcome A_k. Then, the hybrid model states that the conditional probability distribution that the parties observe the outcomes $(a_1, \ldots, a_n)$ when measuring $(x_1, \ldots, x_n)$ is given by

$$P_{a_1 \ldots a_n | x_1 \ldots x_n} = \sum_j \sum_{\lambda_j \in \Lambda_j} P_{\lambda_j} P_{a_1 \ldots a_n | x_1 \ldots x_n \lambda_j}, \tag{3.19}$$

where for each j, $P_{a_1 \ldots a_n | x_1 \ldots x_n \lambda_j}$ factorizes in some way, e.g., as

$$P_{a_1 \ldots a_n | x_1 \ldots x_n \lambda_j} = P_{a_1 \ldots a_k | x_1 \ldots x_k \lambda_j} P_{a_{k+1} \ldots a_n | x_{k+1} \ldots x_n \lambda_j}. \tag{3.20}$$

As above, we furthermore demand the no-signalling conditions $P_{a_l|x_1\ldots x_n\lambda} = P_{a_l|x_l\lambda}$ for all l to be fulfilled.

Again, to show that a given postselection strategy is valid to demonstrate GMN, we first write

$$P_{a_1\ldots a_n|x_1\ldots x_n k} = \sum_j \sum_{\lambda_j\in\Lambda_j} P_{\lambda_j|x_1\ldots x_n k} P_{a_1\ldots a_n|x_1\ldots x_n\lambda_j k}, \tag{3.21}$$

to see that a valid postslection strategy must fulfill the conditions

$$\textbf{C1n} \qquad P_{\lambda|x_1\ldots x_n k} = P_{\lambda|k} \qquad \forall\lambda\in\Lambda,$$

$$\textbf{C2n} \qquad P_{a_1\ldots a_n|x_1\ldots x_n\lambda_j k} = P_{a_1\ldots a_k|x_1\ldots x_k\lambda_j k} P_{a_{k+1}\ldots a_n|x_{k+1}\ldots x_n\lambda_j k} \qquad \forall\lambda_l\in\Lambda_l,$$

and analogous expressions to condition **C2n** for all other subensembles j.

Now assume that the postselection K can be decided by $\lceil n/2\rceil$ parties, where $\lceil\cdot\rceil$ is the ceiling function. We first consider condition **C1n**. The causal diagram for the first step of the proof is shown in Fig. 3.10(a). Here, no specific subensemble is considered, so there may be nonlocal correlations between any subgroup of the A_k's (compare Fig. 3.7(a)). For the sake of visualization, this is sketched as a dashed round box around the A_k's. In this diagram, the postselection K is decided by the last $\lceil n/2\rceil$ parties (recall that it can be equivalently decided by any $\lceil n/2\rceil$ parties, so other diagrams are also valid). We want to show that $P_{\lambda|x_1x_2\ldots x_n k} = P_{\lambda|x_2x_3\ldots x_n k}$, so we have to check all possible paths connecting Λ to X_1 given the conditioning. First note that the direct path $\Lambda \to A_1 \leftarrow X_1$ is blocked because A_1 is a collider that is not conditioned on. Second, note that all remaining paths are of the form $X_1 \to A_1 \to A_l \to \ldots$ for some $l > 1$ and are blocked because of the no-signalling conditions (from the first to the lth party). The conditional independence of Λ to the other X_l is proven in a similar way (for each l, a causal diagram has to be used in which the postselection K is not influenced by A_l), and we obtain condition **C1n**. We want to note again that this condition could also be proven in case of a postselection that is decided by all-but-one parties (i.e., $n-1$ parties), in contrast to the next condition.

Also the proof of condition **C2n** is similar to the one of the conditions **C2** (cf. Theorem 2), with one important difference, as we see in the following. We now first consider the subensemble Λ_{j_0} for which the joint probability factorizes as in Eq. (3.20) with $k = \lfloor n/2\rfloor$: the first k parties may share arbitrary nonlocal (but no-signalling) correlations, as well as the remaining parties. This is again sketched as the dashed boxed in Fig. 3.10(b). Note that permutations of this factorization can be treated in the same way. Again, we use the chain rule to write

$$P_{a_1\ldots a_n|x_1\ldots x_n\lambda_{j_0} k} = P_{a_1\ldots a_k|a_{k+1}\ldots a_n x_1\ldots x_n\lambda_{j_0} k} P_{a_{k+1}\ldots a_n|x_1\ldots x_n\lambda_{j_0} k}. \tag{3.22}$$

Then, we use the causal diagram shown in Fig. 3.10(b) where the postselection K is decided by the last $\lceil n/2\rceil$ parties. The diagram implies that

$$P_{a_1\ldots a_k|a_{k+1}\ldots a_n x_1\ldots x_n\lambda_{j_0} k} = P_{a_1\ldots a_k|x_1\ldots x_n\lambda_{j_0} k}. \tag{3.23}$$

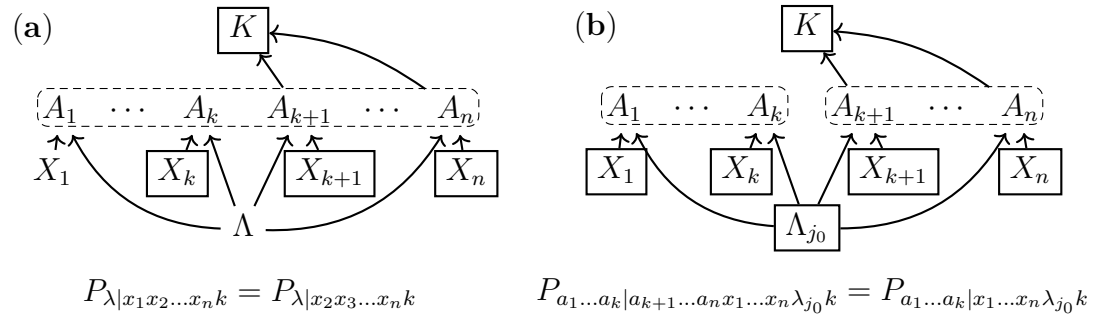

Figure 3.10: Causal diagrams of the n-partite Bell scenario that are used in the proof of Theorem 3. (a) No specific subensemble of Λ is selected, so there may be any kind of multipartite nonlocal (but no-signalling) correlations in any subgroup of $\{A_1, \ldots, A_n\}$, as indicated with the dashed box. (b) By postselecting on the specific subensemble Λ_{j_0}, there can be only nonlocal correlations in the group $\{A_1, \ldots, A_k\}$ and in the group $\{A_{k+1}, \ldots, A_n\}$. In all diagrams, the no-signalling fine-tuning conditions are implied.

For this purpose, we need to check all paths from any A_i for $i \leq k$ to any A_j for $j > k$, conditioned on all the measurement settings, Λ_{j_0} and K. It is clear from Fig. 3.10(b) that every such path must pass through the fork Λ_{j_0} that is conditioned on, so it is blocked. Finally, we can use the same causal diagram (with different conditioned nodes), or, equivalently, the no-signalling principle, to show that $P_{a_1\ldots a_k|x_1\ldots x_n\lambda_{j_0}k} = P_{a_1\ldots a_k|x_1\ldots x_k\lambda_{j_0}k}$ and $P_{a_{k+1}\ldots a_n|x_1\ldots x_n\lambda_{j_0}k} = P_{a_{k+1}\ldots a_n|x_{k+1}\ldots x_n\lambda_{j_0}k}$, and we obtain condition **C2n**.

At this point, we want to highlight why a postselection that has to be decided by more than $\lceil n/2 \rceil$ parties (e.g., by all-but-one parties) is not valid to show condition **C2n**. If any other party needs to be included to decide the postselection K in Fig. 3.10(b), there has to be at least one arrow connecting one of the $A_1, \ldots, A_k$ to K. Thus, Eq. (3.23) does not hold in general, since now there are open paths between the two groups $\{A_1, \ldots, A_k\}$ and $\{A_{k+1}, \ldots, A_n\}$ via the conditioned collider K.

Finally, what about the remaining subensembles Λ_j with $k < \lfloor n/2 \rfloor$ or $k > \lfloor n/2 \rfloor$? If $k < \lfloor n/2 \rfloor$, the information contained in $\{A_{k+1}, \ldots, A_n\}$ is actually redundant to decide the postselection K because it contains more than $\lceil n/2 \rceil$ measurement outcomes. Therefore, in Fig. 3.10(b), one can even cancel some of the arrows from $\{A_{k+1}, \ldots, A_n\}$ to K, and condition **C2n** can be shown in the same way as above. If $k > \lfloor n/2 \rfloor$, we have $k \geq \lceil n/2 \rceil$, so the postselection can be decided by (parts of) the group $\{A_1, \ldots, A_k\}$, and condition **C2n** can be shown as above when reversing the roles of the groups $\{A_1, \ldots, A_k\}$ and $\{A_{k+1}, \ldots, A_n\}$. □

We discuss Theorem 3 with the following comments.

- We first want to emphasize again why the result of Theorem 3 is surprising. As discussed in Sec. 3.1.4 and later, a postselection of bipartite measurement data that can only be decided by a collaboration between the two parties is not safe to demonstrate nonlocality. Furthermore, GMN requires nonlocal correlations between all of

the parties, and not just between subgroups of parties. It is thus surprising that even highly collaborative postselection strategies that require communication between half of the parties are valid for the demonstration of GMN.

- The conditions on safe postselection strategies for GMN are more restrictive than the ones found in Ref. [287] for the demonstration of general (non-genuine) multipartite nonlocality. This is because, in Ref. [287], the model to proof the corresponding Bell inequality was the LHV model of Eq. (3.3) instead of the hybrid local-nonlocal model of Eq. (3.13). Therefore, the proof of the factorization condition similar to condition **C2n** in Theorem 3 allows less restrictive postslection strategies since there are no nonlocal correlations between the different A_i's (cf. Fig. 3.6(a) and (b)). Furthermore, the LHV model automatically fulfills the no-signalling condition, so it does not have to be required additionally as above.

  A different way to see why the two hidden variable models allow for different postselection strategies is the following: The crucial property that decides which kind of postselection is valid for a given hidden variable model is the size of the smallest group that shares nonlocal correlations, maximized over all subensembles of the hidden variable model. This is because one has to use causal diagrams where the postselection is decided by excluding this smallest group. In the LHV model, this size is always one (since there are no nonlocal correlations), so an all-but-one postselection is safe. On the other hand, in the n-partite hybrid model, the size of the smallest group sharing nonlocal correlation is $\lfloor n/2 \rfloor$ for some subensembles, so postselection is only safe if it can be decided by all-but-$\lfloor n/2 \rfloor$ parties (which is the same as $\lceil n/2 \rceil$ parties). Finally, this reasoning shows why in the three-partite case, the valid postselection strategies coincide: in all possible subensembles of the three-partite hybrid model of Eq. (3.13), there is always one party that shares only classical correlations with the others, so the smallest group sharing nonlocal correlations is again one. Thus, the all-but-one postselection strategies are also valid in this case.

- An interesting open question for future research is whether the conditions on safe postselection of Theorem 3 are strict, or whether one can even prove than an all-but-one postselection is safe to demonstrate GMN (by means of new or refined methods). In particular, to prove that Theorem 3 represents strict limitations on safe postselection, one must find a hybrid nonlocal-local hidden variable model with a postselection the can only be decided by $m > \lceil n/2 \rceil$ parties that, after the postselection, violates a GMN Bell inequality. Finding models that use the selection bias to fake nonlocality is a difficult problem already in the smallest Bell scenario of two parties, see Refs. [290, 291, 308]. On the other hand, different approaches to demonstrate conditions for safe postselection might even prove all-but-one postselection strate-

gies valid.

- Applications of Theorem 3 are more elusive than the ones of Theorem 2: previously, for an all-but-one postselection, one needs one conserved quantity such as the total number of particles such that (parts of) the information of the measurement outcomes becomes redundant and one can exclude one of parties from the postselection decision. For a more restrictive postselection, say one that is decided by all-but-two parties, one must exclude two parties, so there have to be two conserved quantities in the measurement outcomes. Finding an all-but-m postselection strategy for $m > 1$ is an interesting open question for future considerations.

  Finally, similar to the discussion of Theorem 2, Theorem 3 cannot take into account realistic finite detection efficiencies. See the final comment below Theorem 2 for a discussion of this point.

3.2.3 Generating genuine three-partite nonlocality from independent particle sources

In this section, we apply our results on safe postselection strategies to a proposal by Yurke and Stoler (YS) to generate multipartite nonlocality from independent particle sources [76]. For this purpose, we first sketch the setup of the YS proposal. Then, we discuss its original demonstration of (non-genuine) three-partite nonlocality by means of the GHZ effect [224]. Finally, we use Theorem 2 to show that the YS proposal also creates genuine three-partite nonlocality.

Yurke–Stoler proposal

The YS proposal of Ref. [76] is sketched in Fig. 3.11. Three independent (but identical) particle sources (S1, S2, and S3) each emit a single particle, a boson (e.g., a photon) or a fermion (e.g., an electron). Each particle passes through a balanced beam splitter (BS) whose outgoing modes are directed to two of the three measuring parties, Alice, Bob and Charlie. As usual, the three parties are assumed to perform their measurement (i.e., freely decide their measurement choice and record their measurement outcome) at a spacelike distance to all other measurements. The measurement process consists of imprinting a known control phase (ϕ_A, ϕ_B, and ϕ_C, respectively) in one of the two incoming modes, placing a second BS to interfere the incoming modes, and finally, measuring the outgoing modes with a number-resolving detector.

The final probability distribution of outcomes is calculated by summing the probability amplitudes of all possible paths that the particles can take. Each particle passes two BS and has thus four possible paths, so in the end the state $|\psi_{\mathrm{f}}\rangle$ is a sum of 64 different amplitudes (we will see shortly that some of the amplitudes add up constructively or cancel out

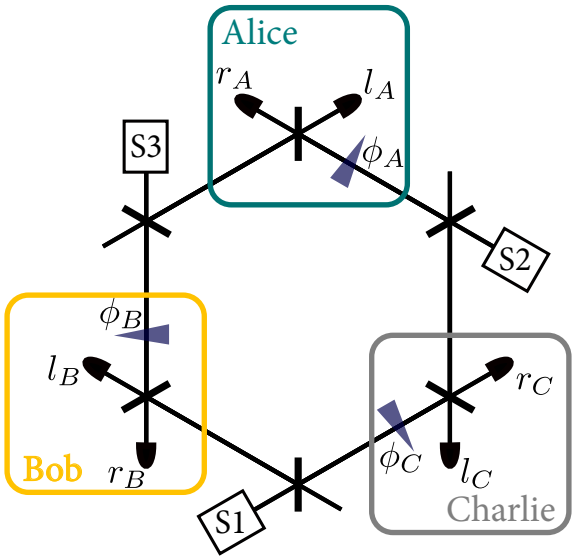

Figure 3.11: Proposal by Yurke and Stoler (YS) to generate three-partite nonlocality from independent particle sources [76]. The three identical particle sources S1, S2, and S3 each emit a particle that each enter a beam splitter (BS) whose outgoing arms are directed to two of the experimental parties. The three parties, Alice, Bob, and Charlie, combine their incoming modes in second BSs after having imprinted a control phase (ϕ_A, ϕ_B, and ϕ_C) in one of the incoming arms. Finally, each party measures its outgoing modes with number-resolving detectors. All events in which one of the parties detects two particles in one detector, and a second party detects no particle, will be later sorted out in the postselection. The interesting events for the demonstration of (genuine) three-partite nonlocality are the ones where each party detects only one particle, which we call r_β (l_β) if party β obtains a detector click in its right (left) detector. The figure is taken from Ref. [2].

destructively). The initial state is given by

$$|\psi_0\rangle = a_3^\dagger a_2^\dagger a_1^\dagger |0\rangle \,, \tag{3.24}$$

where $a_j^\dagger$ is the creation operator corresponding to the particle source Sj. The first BS at Sj transforms the modes according to

$$\begin{bmatrix} b_{\beta l}^\dagger \\ b_{\alpha r}^\dagger \end{bmatrix} = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & -i \\ -i & 1 \end{bmatrix} \begin{bmatrix} a_j^\dagger \\ \tilde{a}_j^\dagger \end{bmatrix} , \tag{3.25}$$

with $j\beta\alpha = 1CB, 2AB, 3BA$. Here, we use the notation that the operator $b_{\beta l}^\dagger$ ($b_{\beta r}^\dagger$) creates a particle entering the apparatus of party β from the left (from the right), as seen by the receiving party in the plane of Fig. 3.11. For instance, a particle created at source S1 results in a superposition of a particle entering Charlie's apparatus from the left ($b_{Cl}^\dagger$) and a particle entering Bob's apparatus from the right ($b_{Br}^\dagger$). The reflected mode at each BS experiences a phase shift (represented by the off-diagonal entries of Eq. (3.25)). The operator $\tilde{a}_j^\dagger$ is a creation operator for the other incoming mode of the initial BSs which is always assumed to be in the vacuum state.

Next, at each measurement apparatus, the left incoming mode $b_{\beta l}^\dagger$ experiences a phase

shift,

$$c^\dagger_{\beta l} = e^{-i\phi_\beta} b^\dagger_{\beta l}, \tag{3.26}$$

while the right incoming mode does not, $c^\dagger_{\beta r} = b^\dagger_{\beta r}$. Finally, the particles pass a second BS, so the final creation operators for party β are given by

$$\begin{bmatrix} d^\dagger_{\beta r} \\ d^\dagger_{\beta l} \end{bmatrix} = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & -i \\ -i & 1 \end{bmatrix} \begin{bmatrix} c^\dagger_{\beta l} \\ c^\dagger_{\beta r} \end{bmatrix}, \tag{3.27}$$

where, for the operators d, the label l (r) labels the outgoing directions of the BS.

By combining Eqs. (3.25-3.27), we find (after simple matrix inversions), for instance,

$$\begin{aligned} a^\dagger_1 &= \frac{1}{\sqrt{2}} \left(b^\dagger_{Cl} + i b^\dagger_{Br} \right) && (3.28)\\ &= \frac{1}{\sqrt{2}} \left(e^{-i\phi_C} c^\dagger_{Cl} + i c^\dagger_{Br} \right) && (3.29)\\ &= \frac{1}{2} \left[e^{-i\phi_C} \left(d^\dagger_{Cr} + i d^\dagger_{Cl} \right) c^\dagger_{Cl} + i \left(i d^\dagger_{Br} + d^\dagger_{Bl} \right) \right]. && (3.30) \end{aligned}$$

After similar calculations for $a^\dagger_2$ and $a^\dagger_3$, and by inserting the results into Eq. (3.24), one obtains the complete expression for $|\psi_{\rm f}\rangle$. To find the probability for a specific outcome, say that each party receives a particle in its right detector ($r_A r_B r_C$), one has to compute

$$P(r_A r_B r_C) = \left\| \langle 0 | \, d_{Ar} d_{Br} d_{Cr} \, |\psi_{\rm f}\rangle \right\|^2. \tag{3.31}$$

Notice that the only terms of $|\psi_{\rm f}\rangle$ that contribute to Eq. (3.31) are the ones that include the three creation operators $d^\dagger_{Ar}$, $d^\dagger_{Br}$, and $d^\dagger_{Cr}$. From Eq. (3.30) and the analogous expressions for $a^\dagger_2$ and $a^\dagger_3$, we see that we have only two terms that contribute, yielding

$$\begin{aligned} P(r_A r_B r_C) &= \frac{1}{64} \left\| \langle 0 | \, d_{Ar} d_{Br} d_{Cr} \left[e^{-i(\phi_A+\phi_B+\phi_C)} d^\dagger_{Br} d^\dagger_{Ar} d^\dagger_{Cr} - d^\dagger_{Ar} d^\dagger_{Cr} d^\dagger_{Br} \right] |0\rangle \right\|^2 && (3.32)\\ &= \frac{1}{64} \left| e^{-i(\phi_A+\phi_B+\phi_C)} - 1 \right|^2 && (3.33)\\ &= \frac{1}{16} \sin^2 \phi, && (3.34) \end{aligned}$$

where we have defined $\phi = (\phi_A + \phi_B + \phi_C)/2$ and we have used that for both bosonic and fermionic (anti-)commutation relations, we have $d^\dagger_{Br} d^\dagger_{Ar} d^\dagger_{Cr} = d^\dagger_{Ar} d^\dagger_{Cr} d^\dagger_{Br}$. Similar derivations hold for probabilities for the other events with one particle per party.

Now consider the events in which one party receives two particles and a second party receives none. Say, for instance, we want to calculate the probability that Alice receives two particles in the right detector, Bob detects one in the right detector, and Charlie detects no particle ($r_A^2 r_B 0$). For this event, only one term of $|\psi_{\rm f}\rangle$ contributes:

$$\begin{aligned} P(r_A^2 r_B 0) &= \frac{1}{64} \left\| \langle 0 | \, d_{Ar} d_{Ar} d_{Br} \left[-e^{-i\phi_A} d^\dagger_{Ar} d^\dagger_{Ar} d^\dagger_{Br} \right] |0\rangle \right\|^2 && (3.35)\\ &= \frac{1}{64} \left\| \langle 0 | \, d_{Ar} d_{Ar} d^\dagger_{Ar} d^\dagger_{Ar} \, |0\rangle \right\|^2 && (3.36) \end{aligned}$$

If the particles are fermions, we have $d^\dagger_{Ar}d^\dagger_{Ar} = 0$, so the event will never happen. This is the fermionic generalization of the Hong–Ou–Mandel effect [312]. If the particles are bosons, we have $d^\dagger_{Ar}d^\dagger_{Ar}\left|0\right\rangle = \sqrt{2}\left|2\right\rangle$ ($\left|2\right\rangle$ is the two-boson Fock state), so we find $P(r_A^2 r_B 0) = 1/32$.

Finally, a similar analysis shows that any event for which one party receives a particle in both its detectors (e.g., $B_A r_B 0$, where B_A means that both Alice's detectors click) happens for fermions with probability $1/16$. For bosons, the Hong–Ou–Mandel effect dictates that these events never occur [312].

In total, we can group the possible events into four groups: either

- one party observes two particles in one detector (D; for double), or
- one party observes one particle in each of its detectors (B; for both), or
- every party receives a particle and the total number of right detector clicks is even (E), or
- every party receives a particle and the total number of right detector clicks is odd (O).

We find the probability for an event e as $P(e) = \sin^2(\phi)/16$ for $e \in O$ and $P(e) = \cos^2(\phi)/16$ for $e \in E$. For fermions, we have $P(e) = 0$ for $e \in D$ and $P(e) = 1/16$ for $e \in B$, while for bosons, we have $P(e) = 1/32$ for $e \in D$ and $P(e) = 0$ for $e \in B$.

The final step before discussing the nonlocality properties of the YS setup is to recognize the similarity of the above probabilities (for coincident particle detection between all parties) to the ones produced by the GHZ state $\left|\psi_{\mathrm{GHZ}}\right\rangle = (\left|000\right\rangle + \left|111\right\rangle)/\sqrt{2}$. Here, observe that when restricting to the events of each party receiving one particle, the probabilities are proportional to (reduced by a factor of $1/4$) the ones produced by the GHZ state, when party β measures the local observable $\cos(\phi_\beta)\sigma_x^{(\beta)} + \sin(\phi_\beta)\sigma_y^{(\beta)}$ [224]. Here, $\sigma_{x/y}^{(\beta)}$ are the x and y Pauli matrices of party β (cf. Eq. (2.32) for a definition).

To see the analogy even clearer, note that when removing each party's apparatus (the phase imprinting and the second BS), the state (when postselected to the events when each party receives one photon) is exactly $\left|\psi_{\mathrm{GHZ}}\right\rangle$, where 0 corresponds to the particle in the mode from the right and 1 corresponds to the mode from the left. We can now discuss the original reasoning of YS [76] that the YS setup shows nonlocality, which is based on the GHZ effect [224].

Nonlocality in the GHZ thought experiment and in the YS setup

We first sketch the thought experiment of the original GHZ demonstration of nonlocality [224], and then see how this effect can be used for the YS proposal. Let us say that three parties prepare a GHZ state $\left|\psi_{\mathrm{GHZ}}\right\rangle$ and measure it, where the party β locally measures

either $\sigma_x^{(\beta)}$ or $\sigma_y^{(\beta)}$ by choosing the measurement setting $\phi_\beta = 0$ or $\phi_\beta = \pi/2$, respectively. After many measurements, the parties find that their results are perfectly correlated when they measure specific combinations, e.g.,

$$\left\langle \sigma_x^{(A)}\sigma_y^{(B)}\sigma_y^{(C)}\right\rangle = \left\langle \sigma_y^{(A)}\sigma_x^{(B)}\sigma_y^{(C)}\right\rangle = \left\langle \sigma_y^{(A)}\sigma_y^{(B)}\sigma_x^{(C)}\right\rangle = -1. \tag{3.37}$$

Each of these equations can be derived by noticing that for measuring, e.g., $\sigma_x^{(A)}\sigma_y^{(B)}\sigma_y^{(C)}$, the parties need to choose $\phi_A = 0$, $\phi_B = \pi/2$, and $\phi_C = \pi/2$, such that we have $P(E) = \sin^2\phi = 1$ and $P(O) = \cos^2\phi = 0$. Here, E (O) are the set of events with an even (odd) number of 1's.

A LHV model that aims to describe these perfect correlations needs to contain instructions about which outcome each local measurement has to produce, given any combination of measurement settings [225]. Equation (3.37) implies that for the given settings, this instruction has to be deterministic (otherwise, the expectation values would not be one). Thus, for a given LHV λ, the instruction must be either 1 or -1 for each measurement setting (i) at each party (β), and we can label the instruction as $s_i^{(\beta)} \in \{-1, 1\}$ (the dependence on λ is implied). Crucially, the instructions need to fulfill Eq. (3.37), e.g., $s_x^{(A)}s_y^{(B)}s_y^{(C)} = -1$. Finally, we multiply the different instructions corresponding to Eq. (3.37) for the given λ to find

$$-1 = s_x^{(A)}s_y^{(B)}s_y^{(C)}s_y^{(A)}s_x^{(B)}s_y^{(C)}s_y^{(A)}s_y^{(B)}s_x^{(C)} = s_x^{(A)}s_x^{(B)}s_x^{(C)}, \tag{3.38}$$

where we have used that $\left(s_i^{(\beta)}\right)^2 = 1$. However, this is in conflict with what the parties will observe whenever they measure $\sigma_x^{(A)}\sigma_x^{(B)}\sigma_x^{(C)}$ because then one has $\phi_A = \phi_B = \phi_C = 0$ and thus $P(O) = \cos^2\phi = 1$, yielding

$$\left\langle \sigma_x^{(A)}\sigma_x^{(B)}\sigma_x^{(C)}\right\rangle = 1. \tag{3.39}$$

In this way, the parties can conclude that their correlations cannot be described by a LHV model, a method that is often called "nonlocality without inequalities", and that nonlocality is demonstrated "with a single experiment" (the single measurement of $\sigma_x^{(A)}\sigma_x^{(B)}\sigma_x^{(C)}$). However, we want to note that, if the parties want to establish that share some perfect correlations, e.g., $\left\langle \sigma_x^{(A)}\sigma_y^{(B)}\sigma_y^{(C)}\right\rangle = -1$, they need to perform a large number of measurements (and even then they only know the correlations up to some confidence). Since each party is required to randomly choose their measurement settings, actually, the parties have already measured $\sigma_x^{(A)}\sigma_x^{(B)}\sigma_x^{(C)}$ many times. Finally, in real experiments, noise necessarily destroys perfect correlations, so a valid nonlocality demonstration again must rely on (Bell-type) inequalities. For more details, see the discussion in the original proposal [224] or its experimental realization [313]. In summary, we can say that the GHZ demonstration of nonlocality is a logical demonstration in the ideal thought experiment, while an experimental demonstration of nonlocality requires the violation of Bell inequalities.

Building on the GHZ effect [224] and its elaboration in Ref. [225], it is merely a small step to see that the YS proposal also demonstrates nonlocality. Even though the probability of the events of interest (every party receives one particle) is reduced with respect to the GHZ thought experiment, the coincident outcomes for the measurement settings of Eq. (3.37) are still perfectly correlated and lead to the same contradiction [76]. But isn't it dangerous to neglect all events in which one party detects two particles for the logical contradiction? No, because in a LHV model, this postselection cannot depend on the local measurement settings. This can be understood (i) by observing that, if the local measurement setting had influence on the double detection, it would require superluminal signalling between the measurement devices [273, 308], or (ii), by arguing that the paths taken by the particles are EPR-elements of reality (since they can be remotely predicted) and as such must be fixed with the LHV λ [314]. Again, as in the GHZ thought experiment above, the nonlocality demonstration of the original YS proposal [76] corresponds to a logical contradiction of the ideal probabilities with LHV models. For an experimental demonstration, a Bell inequality such as Mermin's inequality (cf. Eq. (3.4)) has to be violated.

At this point, we also want to mention that the YS setup of Fig. 3.11 is closely related to the triangle scenario that is often examined in network nonlocality [209, 212], cf. Sec. 3.1.2. Recall that for network nonlocality, several independent sources are shared between different subgroups of the parties, and the observed correlations are termed network nonlocal if at least one of the sources cannot be described by a LHV. In the YS setup, the LHV model thus allows that each pair of parties shares a LHV to produce their respective measurement outcomes. Furthermore, as network nonlocality can be observed in experiments without measurement inputs, we can now even fix the measurements for each party, such that we do not include the measurement choice variables X, Y, and Z anymore. This scenario corresponds to the causal diagram shown in Fig. 3.12, and any network-local correlation can be described by the LHV model

$$P_{abc} = \sum_{\lambda_1,\lambda_2,\lambda_3} P_{\lambda_1} P_{\lambda_2} P_{\lambda_3} P_{a|\lambda_2\lambda_3} P_{b|\lambda_1\lambda_3} P_{c|\lambda_1\lambda_2}. \tag{3.40}$$

As we have discussed before, the techniques to prove that a specific correlation is network nonlocal are rather advanced. However, the ideal (loss-free) YS setup was shown to create network nonlocality in a wide range of transmissivity parameters of the BSs in the setup[8] [213]. Here, the measurement settings were fixed such that $\phi = (\phi_A + \phi_B + \phi_C) = 0$, yielding correlations that were shown to be in logical contradiction with the LHV model of Eq. (3.40), similar to the methods used in Ref. [212]. Furthermore, in Ref. [213], it was

[8] Actually, for a transmissivity of $t = 1/2$ as was considered in the YS setup and our analysis above, there is a network local model to describe the correlations [212]. This is not a contradiction because we have chosen all measurement settings, so the produced correlations have more facility to be describable by a LHV model.

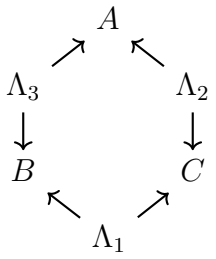

Figure 3.12: Causal diagram for the definition of network locality. In this model, the different measurement results A, B, and C can only be correlated by the pairwise sharing of the LHVs Λ_j, $j = 1, 2, 3$. The most general correlation that can be modeled by this diagram is given in Eq. (3.40).

argued by means of a machine learning oracle (cf. Ref. [315]; and a short explanation in Sec. 4.2.2 for how the network-nonlocality oracle works) that the network nonlocality is robust to small amounts of losses and detection inefficiencies.

Genuine three-partite nonlocality in the YS proposal

By means of the GHZ thought experiment, we have seen that the YS proposal creates multipartite nonlocality. What about GMN? There are a few subtleties when trying to argue for GMN in the YS proposal along the lines of the GHZ thought experiment. First, the assumption of a LHV model is central to the GHZ contradiction. If one allows for a hybrid local-nonlocal model instead, this reasoning cannot be applied anymore. Second, the postselection used in the YS proposal was safe for detecting (non-genuine) multipartite nonlocality because, here, the measurement settings cannot influence the postselection, see the discussion above. For hybrid models, this reasoning cannot be applied anymore because there may be (bipartite) nonlocal influences, and no EPR elements of reality can be identified because the latter rely on locality.

Third, the measurement settings used in the GHZ thought experiment correspond to the ones used in Mermin's inequality, Eq. (3.4), that detects (non-genuine) nonlocality. Now, we need to consider an inequality that tests for GMN such as Svetlichny's inequality, Eq. (3.6). Indeed, the GHZ correlations and thus the postselected YS correlations violate Eq. (3.6) when changing the possible measurement settings [316]. To see this, say that each party chooses from the following measurement settings: Alice measures A_1 ($\phi_A = 0$) or A_2 ($\phi_A = -\pi/2$), Bob measures B_1 ($\phi_B = \pi/4$) or B_2 ($\phi_B = -\pi/4$), and Charlie measures C_1 ($\phi_C = 0$) or C_2 ($\phi_C = -\pi/2$). We rename the outcomes r_β and l_β to 1 and -1, respectively. Then, given the postselected data (i.e., renormalizing the probabilities in the YS proposal to the events in O and E), one has for instance

$$\langle A_1 B_1 C_1 \rangle = 4 \left(\sum_{e \in O} P(e) - \sum_{e \in E} P(e) \right) = \sin^2 \phi - \cos^2 \phi = -\frac{1}{\sqrt{2}}, \tag{3.41}$$

where $\phi = (\phi_A + \phi_B + \phi_C)/2 = \pi/8$ and the factor 4 stems from the renormalization of the postselected probabilities ($P(k = 1) = P(O \cup E) = 1/4$). The other terms of Svetlichny's

inequality, Eq. (3.6), can be computed in a similar way, and we finally obtain a value of $4\sqrt{2}$ (each summand contributes $1/\sqrt{2}$) for the left hand side of Eq. (3.6), which demonstrates genuine three-partite nonlocality for the postselected data of the YS proposal.

Finally, we can use Theorem 2 to show that the postselection strategy that we used to violate Svetlichny's inequality is valid. Indeed, an event that is sorted out in the postselection procedure can be recognized by two different measurement parties independently: one party receives two particles (so either a double detection in one detector or a coincident detection in both the party's detectors) and a second party detects no particle. Therefore, the postselection can always be decided by excluding any party, and by Theorem 2, the postselection is valid to demonstrate GMN.

We want to briefly discuss the influence of experimental imperfections. As discussed below Theorems 2 and 3, in the case of non-ideal detectors (i.e., detectors with finite efficiencies or producing false detections) and finite transmission efficiencies of the channels in the setup, the postselection cannot be decided by two parties anymore: for instance, if Alice and Bob receive both a single particle, they cannot be sure that Charlie has also received one. Thus, for realistic settings, one has to rely on the fair-sampling assumption or one has to include all data without postselection[9], see Sec. 3.1.4 for a discussion. Finally, the three particle sources may be not perfectly identical, reducing the Hong–Ou–Mandel effect. This imperfection does not hinder the postselection because the latter can still be decided by any two parties. However, high distinguishability of the sources reduces the interference such that at a critical distinguishability, the Svetlichny inequality is not violated anymore. One could thus say that the indistinguishability of the single particle sources is a critical quantum resource to generate GMN from independent particle sources.

3.3 Conclusions and outlook

In this chapter, we have entered the field of verifying the quantum resource of nonlocality. Verifying nonlocality is a task that is addressed by means of testing the violation of Bell inequalities. First, we have given an extensive introduction to the notions of bipartite, multipartite, and genuine multipartite nonlocality, their corresponding Bell inequalities, their applications and possible difficulties in their verification. Then, we have focused on the question of under which conditions it is allowed to postselect data in the demonstration of genuine multipartite nonlocality, meaning that we can exclude any postselection bias that mimics nonlocal correlations.

The main result of the chapter is given by Theorem 3, stating that

[9] We want to note that the more limiting source of losses for applications of nonlocality, e.g., in device-independent quantum key distribution, is a finite detection efficiency because finite transmission efficiencies can be usually taken care of by heralding strategies, see, e.g., Ref. [317].

In the n-partite Bell scenario, a postselection strategy that can equivalently be decided by any $\lceil n/2 \rceil$ parties is valid to demonstrate genuine n-partite nonlocality.

This means that certain postselection strategies in which half of the parties have to communicate to decide the postselection are valid to demonstrate genuine multipartite nonlocality. This is surprising because collective postselection is known to potentially mimic nonlocal correlations even when the complete data are local (e.g., in Franson's original proposal [289]), and the fact that genuine multipartite nonlocal correlations require that all parties share common nonlocal correlations. Our results offer new and valid postselection strategies that go beyond locally decidable postselection which has been known to be safe in nonlocality demonstrations. Furthermore, the probability of postselection can be arbitrarily small as long as the above condition is fulfilled. We have proven the result with the help of the methods of causal inference and causal diagrams that are common tools in statistical analysis.

In the case of three measurement parties, our result can be applied to scenarios where there exists some conserved quantity such as, e.g., the conservation of the total number of particles. Then, partial information of one of the parties is redundant and the party may be excluded from the postselection decision, rendering the postselection valid. An example of such a situation is a proposal by Yurke and Stoler [76] that, before, could only demonstrate (non-genuine) multipartite nonlocality. By postselecting the observed data (in a way that fulfills the above condition), one can maximally violate the Svetlichny inequality that demonstrates genuine three-partite nonlocality. For more than three measurement parties, more than one party has to be excluded from the postselection. Thus, one needs two conserved quantities and possible applications of our results are more difficult to identify.

There are several directions for future investigations originating from our results. First, our results only hold in ideal situations with no experimental imperfections. In particular, if the detectors have finite efficiencies such that some particles are not detected, a helpful postselection may not be decidable when excluding one measurement party. It is an interesting and challenging task to extend our results to these realistic situations and then finally apply them for experimental demonstrations of genuine multipartite nonlocality. Also, an extension of our results to new definitions of (LOSR) genuine multipartite nonlocality [54] is a clear direction for future research. Furthermore, we want to note that postselection is an ubiquitous method in physics and in general science, and thus, in principle, a rigorous account of the postselection should be demanded. Finally, the tools of causal inference and causal diagrams that are mostly used in mathematics and statistics, represent valuable techniques that have the potential to advance all fields of physics such as, e.g., quantum information.

Chapter 4

Verification: Detection of nonclassicality by neural networks

Contents

In this chapter, we enter the field of continuous-variable (CV) quantum technologies. In particular, we address the verification of one specific CV quantum resource, the nonclassicality defined by the negativity of the Glauber–Sudarshan P function in different measurement schemes of quantum optical experiments. For this purpose, we train artificial neural networks to classify experimental data as classical or nonclassical.

In Sec. 4.1, we briefly review different quantum resources in CV systems and their applications, with a focus on quasiprobability distributions. Then, in Sec. 4.2, we give a short

overview of machine learning (ML) methods that are applied in quantum information and technology. Finally, we present our results of ML nonclassicality classifiers for experimental data from homodyne detection (Sec. 4.3) and for simulated data of multiplexed click-counting detection (Sec. 4.4).

4.1 Quantum resources in continuous-variable quantum technologies

CV quantum systems differ from finite-dimensional (or discrete-variable) quantum systems such as qubits and qudits by the fact that their Hilbert space is infinite dimensional[1]. The best known example of an infinite dimensional system is the quantum harmonic oscillator that naturally describes several systems in different areas of quantum physics such as, e.g., quantum optics (i.e., the single-mode electromagnetic field) or superconducting quantum technologies (i.e., the LC circuit). For extensive introductions to the quantum harmonic oscillator in the context of quantum optics see Ref. [318], to CV quasiprobability distributions see Ref. [319], and to CV quantum technologies see Ref. [320].

The framework of CV quantum technologies differs from the discrete variable one at many levels: an infinite dimensional Hilbert space contains many more degrees of freedom (infinitely many so) and thus offers a larger playground for possible technologies. On the other hand, the type of operations that can be implemented in experiments also differ significantly. For instance, while general (and entangling) operations between two and more different qubits are efficiently implementable in many discrete-variable quantum architectures, general operations between different CV quantum systems (and also within a single CV system) can represent a significant challenge to implement in an experiment. In CV systems, the natural operations can be grouped into passive elements like, e.g., beam splitters and phase shifts, and active elements such as, e.g., nonlinear crystals, all of which can be harnessed to approach the technological tasks.

Due to these differences, quantum technologies and characteristics in one description do not necessarily find an equivalent counterpart in the other. Also, technologies that correspond to the same task may require very different logical operations and descriptions (see, e.g., the protocols for quantum teleportation in discrete variables [26] and in CV [321]). Consequently, the role and identification of possible quantum resources for the different technological tasks can disagree. As in the discrete variable case, one of the most

[1]Strictly speaking, infinite dimensional Hilbert spaces also allow a description by means of a discrete (and infinite) basis, the Fock basis, while for many tasks, the description by CVs is more useful. On the other hand, while mostly discussed in terms of discrete variables, finite dimensional systems can also be understood by CV descriptions. Here, we simply identify finite dimensional systems with discrete variable quantum technologies and infinite dimensional systems with CV quantum technologies. These (oversimplified) identifications are common in the literature.

important quantum resources is entanglement[2]. Here, many notions like definitions or different entanglement tests, as well as the general difficulty of deciding whether a given state is entangled or not, resemble the corresponding notions in the discrete-variable case, see, e.g., Ref. [23] for a review.

However, due to the richer structure and information capacity of merely a single (i.e., non-composite) CV system compared to, say, a single qubit, there are also several (single-mode) quantum resources that play important roles in CV quantum technologies. In the following, we will focus on the role of the quantum resources of nonclassicality defined by negativity of quasiprobability distributions. In Sec. 4.3 and Sec. 4.4, we will address the verification of one of these notions of nonclassicality.

4.1.1 Quasiprobability distributions

First, we want to briefly recall the notions of the quantum harmonic oscillator. For a detailed technical introduction, see, e.g., Ref. [318]. The Hilbert space of the quantum harmonic oscillator has a discrete orthonormal basis called number or Fock state basis that is denoted as $\{|n\rangle \, | n \in \mathbb{N}_{\geq 0}\}$. The state $|0\rangle$ is the ground state or vacuum state. Furthermore, there are annihilation operators a and creation operators $a^\dagger$ acting on the Fock basis as $a\,|n\rangle = \sqrt{n}\,|n-1\rangle$ for $n > 0$ and $a\,|0\rangle = 0$, and $a^\dagger\,|n\rangle = \sqrt{n+1}\,|n+1\rangle$ for $n \geq 0$. These operators fulfill the bosonic commutation relation $\left[a, a^\dagger\right] = 1$.

The creation and annihilation operators play a central role in CV quantum technologies because both the system's Hamiltonians and measurement observables can be expressed as functions of low-degree polynomials in a and $a^\dagger$. Therefore, a widely used CV basis is given by the coherent states $|\alpha\rangle$ defined as eigenstates of the annihilation operator, $a\,|\alpha\rangle = \alpha\,|\alpha\rangle$ for $\alpha \in \mathbb{C}$. This now continuous basis is not orthogonal anymore, $|\langle\beta|\alpha\rangle|^2 = e^{-|\alpha-\beta|^2}$, and is often called overcomplete, $\int_{\mathbb{C}} |\alpha\rangle\langle\alpha|\,\mathrm{d}^2\alpha/\pi = \mathbb{1}$. In this basis, any system density operator or measurement observable can be represented as a distribution of the two-dimensional space (over $\mathbb{R}$), the quantum-optical phase space. This strongly resembles classical notions of probability distributions over the phase space of a one-dimensional system, spanned by position and momentum. Given this analogy, the idea of quasiprobability distribution is

[2]Note that in CV systems, one often speaks of "mode" entanglement when different modes, each described by a quantum harmonic oscillator, are entangled. Here, the system is described in second quantization, i.e., the basic objects are the different modes of the CV system. In contrast, distinguishable discrete-variable systems (such as qubits or atoms at different spatial locations) are usually described in first quantization, where one speaks of different particles that are simply entangled, even though this entanglement can also be thought of as entanglement of the spatial modes of the qubits. Finally, in the situation of indistinguishable discrete-variable systems (such as cold atoms in a cloud), the notion of particle entanglement is commonly used, a notion that also includes entanglement due to symmetrization of the particles that are described in the first quantization picture. For a review and practical applications of entanglement of identical particles, see, e.g., Refs. [322, 323, 324].

to write the expectation value of a measurement observable $O(a, a^\dagger)$ with respect to a state ρ as [325, 326, 327, 328]

$$\langle O \rangle = \mathrm{tr}[\rho O] = \int_{\mathbb{C}} \tilde{P}_\rho(\alpha)\tilde{Q}_O(\alpha)\mathrm{d}^2\alpha, \tag{4.1}$$

for some distributions $\tilde{P}_\rho(\alpha)$ and $\tilde{Q}_O(\alpha)$. Furthermore, for the specific observable $O(a, a^\dagger)$, a perfect classical analogy would be to simply use $\tilde{Q}_O(\alpha) = O(\alpha, \alpha^*)$ (α^* is the complex conjugate of α), i.e., one would like to simply insert the complex numbers α and α^* instead of the operators a and $a^\dagger$, respectively. Here, one can already see why different definitions of quasiprobability distributions are possible: for instance, the operator $O = a^\dagger a$ can be written equivalently as $O = aa^\dagger - 1$ due to the bosonic commutation relations. However, by simply inserting α for a and α^* for $a^\dagger$, both expressions lead to different functions over the state space. Thus, to generally describe the same expectation value in Eq. (4.1), also the distributions $\tilde{P}_\rho(\alpha)$describing the state ρ must differ. Note that, in contrast to $\tilde{P}_\rho(\alpha)$ and $\tilde{Q}_O(\alpha)$, only the expectation value is really observable in experiments.

There are three major choices for these different possible function assignments. First, one can use the bosonic commutation relations to write the operator $O(a, a^\dagger)$ in its normally-ordered form, meaning that one uses the commutation relations to bring all operators a to the right of the operators $a^\dagger$[3], and then inserts $a \mapsto \alpha$. In the above example, the operator $O = a^\dagger a$ is already in normally ordered form. The corresponding function $\tilde{P}_\rho(\alpha)$ is called the Glauber–Sudarshan P function [329, 330] that, as we explain below, implies the definition of nonclassicality that our later results address. There are different possible expressions for how to calculate the P function of a state ρ, for which we refer to Ref. [318]. Here, we only want to note that the state ρ can be written in terms of its P function $P(\alpha)$ as[4]

$$\rho = \int_{\mathbb{C}} P(\alpha) \left|\alpha\right\rangle \left\langle\alpha\right| \mathrm{d}^2\alpha. \tag{4.2}$$

[3]Note that this reordering of the operator $O(a, a^\dagger)$ should not be confused with a different formal reordering that we will meet later and that we denote as $: O(a, a^\dagger) :$. Here, one also brings all operators a to the right of the operators $a^\dagger$, but this time without respecting the commutation relations. Thus, while using the commutation relations to reorder the operators of $O(a, a^\dagger)$ results in the same operator, the operator $: O(a, a^\dagger) :$ is generally different to $O(a, a^\dagger)$.

[4]This can be seen easily by inserting Eq. (4.2) into Eq. (4.1),

$$\begin{aligned}\mathrm{tr}[\rho O] &= \int_{\mathbb{C}} P(\alpha)\mathrm{tr}[\left|\alpha\right\rangle \left\langle\alpha\right| O]\mathrm{d}^2\alpha \\ &= \int_{\mathbb{C}} P(\alpha) \left\langle\alpha\right| O \left|\alpha\right\rangle \mathrm{d}^2\alpha \\ &= \int_{\mathbb{C}} P(\alpha) O_{\mathrm{N}}(\alpha)\mathrm{d}^2\alpha,\end{aligned}$$

where we have used that $\left\langle\alpha\right| O \left|\alpha\right\rangle = O_{\mathrm{N}}(\alpha)$, where O_{N} is the normally ordered version of O with inputs $a \mapsto \alpha$ and $a^\dagger \mapsto \alpha^*$.

When the $P(\alpha)$ is non-negative, Eq. (4.2) describes ρ as a classical mixture of coherent states. We note that the P function is generally not well behaved, for instance, the P function of a coherent state is a Dirac delta distribution.

The second and most famous example of a quasiprobability distribution is the Wigner function [331] that was introduced to construct a phase-space description for the spatial wave function $\psi(x)$ and its evolution. The Wigner function is the appropriate quasiprobability function if the operator $O(a, a^\dagger)$ is written in its symmetrized form with respect to a and $a^\dagger$, before inserting $a \mapsto \alpha$. For instance, in the above example, one writes $O = a^\dagger a = (a^\dagger a + aa^\dagger - 1)/2$. The Wigner function can be obtained from the P function by a convolution with a Gaussian kernel, and is well behaved. In fact, it is bounded but still can take negative values [318]. The Wigner function exhibits several useful characteristics that let it stand out with respect to the other quasiprobability functions. For instance, taking the marginal distribution along any line through the origin, one obtains the (quadrature) probability distribution of a specific balanced homodyne measurement setting [318]. These characteristics are the reason why for defining quasiprobability distributions that describe discrete-variable systems, usually the Wigner function is generalized [332].

Finally, we want to mention the Husimi Q function [333] that corresponds to an antinormal ordering of the operators in $O(a, a^\dagger)$ (i.e., all operators $a^\dagger$ should be to the right of the operators a). Above, this corresponds to $O = aa^\dagger - 1$. The Q function is a further convolution of the Wigner function with a Gaussian kernel, and shows no negative values anymore. Still, it cannot be seen as a classical probability distribution over the coherent amplitude α because states corresponding to different α are not orthogonal. The Q function is directly measurable in different quantum optical measurements. We also want to note that there is actually a continuous spectrum of possible quasiprobability distributions between the P and the Q functions, which trade off mathematical difficulties in the state's representation $\tilde{P}_\rho(\alpha)$ (for the P function) or the observable's representation $\tilde{Q}_O(\alpha)$ (for the Q function) in Eq. (4.1). We refer to Ref. [318] for more details.

There are further possible quasiprobability distributions considered in quantum information, especially in the case of discrete variable systems, that do not fall in the above family of distributions between the P and the Q function, see Ref. [319] for the CV case and Ref. [332] for the discrete case. We will mention some of these alternative quasiprobability distributions below when we discuss applications.

Nonclassicality

As discussed above, the quasiprobability description of quantum systems is an attempt to make an analogy to classical phase-space descriptions. In the latter though, the distributions on the phase space fulfill all conditions of probability distributions such as, e.g., pos-

itivity. Therefore, a natural indicator of "quantumness" or "nonclassicality" of a quantum system is the negativity of a quasiprobability distribution. Of course, this indicator crucially depends on the type of quasiprobability distribution that we employ. For instance, the Husimi Q function is always positive (since the negativity and all singularities are transferred to the distribution $\tilde{Q}_O(\alpha)$ that describes the observable), and thus itself does not indicate nonclassicality. A commonly used indicator of nonclassicality is the negativity of the Wigner function, as we will also discuss below for different applications. However, since the Wigner function is a smoothed version of the Glauber–Sudarshan P function, there are states for which the P function is negative but the Wigner function is not. An important example with many applications (see below) are squeezed states that have a positive Wigner function but a negative and highly irregular P function.

In any definition of a quantum feature, some choice has to be made about which states do not offer this feature. For instance, in the definition of entanglement, these "free" states are the separable states (cf. Sec. 1.2.1). In the context of single-mode CV systems, coherent states are usually defined as the free or "classical" states. This is because coherent states most closely resemble the states of classical electrodynamics because they exactly reproduce the expectation values and variances of electromagnetic field operators in classical electrodynamics [334]. Then, in analogy to entanglement, a general classical state is defined as an arbitrary convex combination of coherent states,

$$\rho = \int_{\mathbb{C}} P(\alpha) \left|\alpha\right\rangle \left\langle\alpha\right| \mathrm{d}^2\alpha, \tag{4.3}$$

with $P(\alpha) \geq 0$ describing the distribution of the mixture of the coherent states $\left|\alpha\right\rangle \left\langle\alpha\right|$. As nonclassicality corresponds to the impossibility to express a quantum state by such a mixture (i.e., by a convex combination of the "classical" coherent states), we see that it is precisely corresponds to the negativity of P function in Eq. (4.3).

The definition of nonclassicality by means of the P function is the one we will use later in this chapter. We again want to emphasize that nonclassicality can also be considered in terms of negativity of the Wigner function, which is the common approach in the context of discrete variable systems. However, we note that since the Wigner function is a convolution of the P function with a Gaussian kernel, nonclassicality by means of the Wigner function implies nonclassicality by means of the P function. This is because a positive P function convoluted with a Gaussian kernel will necessarily lead to a positive Wigner function, so the negativity of the latter implies the negativity of the former.

Finally, before we discuss different applications, we note that similar to quantum resources in finite dimensional systems, the detection and verification of nonclassicality is a difficult task. Since we address this point with ML methods in Sec. 4.3 and Sec. 4.4, we will discuss and compare current methods for the verification of nonclassicality in the respective sections.

Applications

We now give a brief overview about situations where nonclassicality is necessary for specific quantum technologies. We note that many results are derived for nonclassicality defined as negativity of the Wigner function, which, as discussed above, implies negativity of the P function, and thus our definition of nonclassicality as well. In particular, if it is found that the negativity of the Wigner function is necessary to reach some technological task, also the negativity of the P function is necessary. In the following, we will thus only speak of nonclassicality as the negativity of the P function, even though many results are derived for the Wigner function.

In several quantum technologies, nonclassicality has been found to be central to quantum advantages. For instance, as we discussed in Sec. 2.1.1, in the magic state distillation model of quantum computation [97], nonclassicality is necessary for a quantum computational advantage [125, 126, 127]. Also, nonclassicality is necessary for quantum advantages for distributed quantum computing [335]. Furthermore, in distributed (nonlocal) boson sampling, even though input and output correlations may be classically correlated, nonclassicality of the output state can render the output distribution intractable for classical computers, thus representing a quantum advantage [336]. The power of quantum networks (in the field of distributed sensing) was shown to require nonclassical states when only passive optical elements are available [61]. Finally, nonclassicality as a quantum resource for metrology was quantified by means of a connection to the QFI [337]. Note that this is a very compressed list of applications and many more can be found in the literature.

A central family of nonclassical states that many quantum technologies build upon are squeezed states [320], see also Ref. [338] for a specific review of Gaussian quantum information processing with different applications. In particular, we want to mention that squeezed states have advanced the current technology offering the highest sensitivities in precision measurements. For instance, they have been used in the detection of gravitational waves [339, 340].

As discussed in Sec. 1.3, different quantum resources can be converted to each other. Here, the single-mode resource of nonclassicality was shown to be convertible to entanglement merely by means of passive optical elements such as beam splitters [44, 45, 341]. In particular, when mixing an incoming quantum state with the vacuum state on a balanced beam splitter, the outgoing modes are entangled if and only if the incoming state is nonclassical: for specific measures of nonclassicality and entanglement, the amount of nonclassicality of the incoming mode is equal to the amount of entanglement of the outgoing ones [44].

Notions of nonclassicality have also been found to be deeply connected to quantum foundations. For instance, nonlocality and Bell inequalities have been derived in terms of quasiprobability distributions [342], and in Ref. [343], it was shown that nonclassicality

and noncontextuality in measurements are equivalent. The negativity of quasiprobability distributions has further been connected to the incompatibility of quantum measurements [344].

We also want to mention that different quasiprobability distributions that are generalizations of the above introduced P, Wigner, and Q functions and do not fit in their family, have been identified as crucial in quantum resources and quantum technologies. For instance, the Kirkwood–Dirac distribution [345, 346] is a generalization of the Wigner function that cannot only take negative values but also complex ones. This distribution is deeply connected to the explanation of weak values [295, 347] and is central for different quantum technologies such as weak value amplification [296, 297, 298, 299], direct measurements of wave functions [348, 349, 350], and the phenomenon of information scrambling [351].

Similar to all other quantum resources, the task of detecting nonlocality from experimental data is difficult. We will explicitly discuss different verification possibilities of nonclassicality in Secs. 4.3 and 4.4 for specific measurement schemes of single-mode CV systems. We will then approach this task by means of machine learning techniques. Therefore, in the following, we provide an overview of the methods of machine learning and their applications in different problems of quantum physics.

4.2 Machine learning methods in quantum physics

The increasingly active field of machine learning (ML) is becoming ever more important in quantum physics and quantum technologies. In the following, we will list different branches of modern ML and overview how they have been applied in quantum information. For a detailed overview of the topic, we refer to Ref. [352]. We want to emphasize that our introduction only comprises the applications of "classical" ML methods to quantum physics (i.e., we use ML that is performed on a classical computer), whereas, recently, much attention is given to the field of quantum ML, see, e.g., Refs. [353, 354].

The field of ML or artificial intelligence considers problems that are seemingly hard to solve by computers but that are easily solved by humans. In particular, this includes several learning tasks in which humans naturally learn to understand or deduce underlying information or patterns. There are many different methods and approaches to ML, see Refs. [355, 356] for detailed introductions. In the following overview, we will restrict ourselves to the topic of artificial neural networks due to their prevalence in modern ML applications.

4.2.1 Artificial neural networks

Artificial neural networks (which we now simply call NNs) are inspired by the complex structure of deeply connected neural networks in biological brains. The central idea is

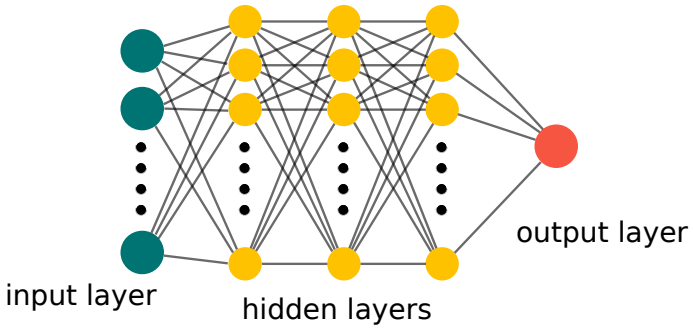

Figure 4.1: Sketch of a deep neural network (NN) with an input layer (blue), three hidden layers (yellow) and an output layer (red) consisting of only one neuron.

that, by modelling a biological neural network by means of a mathematical network with simple functional dependencies between the nodes, the artificial network can be trained with large amounts of data to learn in a similar fashion as real brains. In particular, different parts of the network should correspond to regions that receive information, parts that process information and compare it to previously seen situations, and parts that describe the network's answer to the learning task.

The most common model of NNs is that of a deep feedforward NN consisting of different layers that are interconnected, see Fig. 4.1. Here, an input layer $\mathbf{v}_0$ consists of the data that are given to the NN like, e.g., the pixels of a picture. The input layer is then connected to the first hidden layer $\mathbf{v}_1$, after which some number of hidden layers may follow. The last hidden layer is connected to the output layer $\mathbf{v}_f$ that describes the output of the NN (in Fig. 4.1, the output is merely a real number and thus the output layer consists of a single neuron, but generally there might be many output neurons). Each layer is connected to its previous layer as

$$\mathbf{v}_i = f\left(\mathbf{W}_i \mathbf{v}_{i-1} + \mathbf{b}_i\right), \tag{4.4}$$

where $\mathbf{W}_i$ is a weight matrix, $\mathbf{b}_i$ is a bias, and f is some nonlinear function that is evaluated element-wise. The set of all weights and biases constitute the parameters of the NN and will be changed during the training of the NN. The nonlinear function is fixed and is usually chosen as a simple sigmoid function or a rectified linear unit [355]. It is crucial for the expressive power of the NN: without the nonlinearities, all elements of the NN would be linear, such that such a NN could only describe linear functions. On the other hand, nonlinear NNs are universal in the sense that they can approximate any nonlinear function between input and output layer to arbitrary precision if the number of hidden layers and hidden neurons is chosen sufficiently large [357]. We want to note that the information of the network's parameters can be strongly compressed if the weights and biases fulfill some symmetry conditions and act only locally on the next layer. In this case, one speaks of convolutional NNs that are also inspired by biological neurons. In the following, we describe different ways to train the parameters of the NN that are appropriate in different learning

situations.

Supervised learning

Supervised learning of the NN can be applied when one possesses a large number of input data vectors $\mathbf{v}_0$ that each are equipped with a label $\mathbf{y}$ that should be learned by the NN. Say, the output of the NN is $\mathbf{r}(\mathbf{v}_0)$ when given the input $\mathbf{v}_0$. After choosing an appropriate distance measure d, the goal is thus to minimize the average cost

$$C = \langle d[\mathbf{r}(\mathbf{v}_0), \mathbf{y}] \rangle_{\text{data}} , \tag{4.5}$$

where $\langle \cdot \rangle_{\text{data}}$ indicates the average over the whole data set. By means of the functional structure of the NN, one can calculate the gradient of C with respect to its parameters and update the parameters using a simple gradient descent. If the size of the data is large (which is usually the case), this learning method is very slow. Instead, one approximates the gradient using only a subset of the data, a so-called batch. The computation of these approximate gradients can be performed very fast on computers by means of the so-called backpropagation algorithm which essentially consists of fast matrix multiplications [355]. Therefore, in each training epoch, one splits the data into several batches each of which is used to approximate the gradient and to update the NN's parameters. This is called stochastic gradient descent. There are a few more details of how to optimize the supervised training such as, e.g., how fast one moves along the gradient, or whether one includes some memory of previous gradient approximations, and more [355]. Finally, we note that this type of supervised learning is easy to implement due to a wide range of openly-available programs and libraries, see, e.g., the python libraries *keras* and *tensorflow*.

An important obstacle that must be avoided in supervised learning algorithms is "overfitting". Overfitting occurs when the size of the NN (i.e., the number of hidden layers and the number of hidden neurons) is chosen too large for the specific task, and the NN is trained for a very long time. In this case, since the NN is a universal function approximator, it perfectly learns by heart all the input-output relations of the data set. This is not what the NN was trained for, since it should recognize general patterns or structure in the data, and an overfitted NN does not generalize well to data that were never shown to it during the training. An easy way to circumvent overfitting is to split the complete data set into training data and validation data (e.g., in a ratio 80% to 20%). The NN will then be trained with the training data and gets better over time. However, after each training epoch, one checks the performance of the NN on the validation data set. Initially, the NN will have an increasing performance on the validation data because it starts to see essential structures in the data. Then, as soon as the performance on the validation data decreases, one stops the training of the NN because this means that now the NN is overfitting as it is just learning by heart the training data. Note that there are also other methods to fight overfitting such

as, e.g., including a dropout probability of neural connections [355]. Later, we will simply work with the validation data method. There are other general difficulties concerning the learning phase and the generalization performance of NNs that are addressed by computational learning theory, a topic that is beyond the scope of this introduction and we refer to Refs. [352, 356] for more details.

We want to mention that in supervised learning, apart from artificial NNs, a second widely applied ML technique are support vector machines that aim to find a linear high-dimensional hyperplane that correctly separates binary-labeled data. This method constitutes one of the approaches to implement ML on quantum computers [358], entering the field of quantum ML. Furthermore, in supervised learning of NNs, we note that not all networks are simple feedforward networks as described in Eq. (4.4). For instance, information can also travel backwards in NNs in recurrent neural networks (modeling memory effects) or restricted Boltzmann machines which are used to approximate probability distributions. For more ML models and approaches in supervised learning, see Ref. [352].

Further ML approaches

A second important and commonly used ML technique is reinforcement learning. This type of ML might be the one that most closely resembles artificial intelligence and human learning. In short, the ML agent interacts with an environment by means of observing the environmental state and then executing a set of possible actions that may influence the state of the environment. The agent is either rewarded for specific steps during this interaction, or is evaluated at the end of some fixed number of actions according to some figure of merit. By trial and error, the agent then learns to adjust its actions for given states of the environment, until it optimizes its final reward. Reinforcement learning is the common ML method when computers teach themselves how to play games such as Tetris or car racing, or when some computer model should learn a specific task such as walking or jumping. We again refer to Refs. [352, 355] for a detailed introduction to reinforcement learning.

Finally, we also want to mention the last major ML method, unsupervised learning. Here, the NN is asked to find hidden structure in data, where the data are not labeled. This method is important for classification and clustering analysis of large data sets with no prior information. See Ref. [356] for details and specific applications.

4.2.2 Applications of ML in quantum physics

Various (if not all) fields of quantum information and technology have been tackled by means of ML methods. Here, we give a brief (and very incomplete) overview of different applications. For more extensive overviews, see Refs. [352, 356].

The task of quantum state tomography that asks to construct the underlying density matrix of a quantum state from many measurement results, has been very successfully addressed by ML. In Ref. [359], restricted Boltzmann machines were used for highly efficient quantum state tomography of systems consisting of many qubits for which known tomography methods are computationally demanding. The efficiency of the method is due to a specific ansatz that restricts its general expressive power (even though it has proved sufficient in the common models of real physical systems), similar to the expressive power matrix product states [360]. This method has been used for tomography in many body systems [361] and has even been generalized to homodyne detection schemes of quantum optical states [362]. In the latter case though, the training of the NN could not be accelerated sufficiently (which was possible in the case of many-body systems) such that it cannot outperform standard tomography methods up to now[5]. For quantum optical states, supervised learning of NNs using so-called auto encoders has been used for reconstruction and classification [363]. Also, feedforward NNs were used to optimally include preparations and measurement noise in state tomography [364], and to observe and apply results of computational learning theory in the tomography of systems consisting of few qubits [365].

The study of quantum many-body physics and quantum phases of matter suffers from exponential computational complexity in classical descriptions. Here, different ML approaches have been used to yield new insights or to offer powerful alternatives to standard methods. For instance, reinforcement learning of restricted Boltzmann machines has been used to find the ground state and the time evolution in complex quantum systems [361]. Furthermore, unsupervised learning, that requires no physical model of prior information, could successfully identify different phases of a quantum system [366], and it was used in the form of anomaly detection to search for interesting parameter ranges of possible new phases of a quantum system [367]. Also, convolutional NNs have been trained with supervised learning to recognize topological phases of matter [368].

The field of quantum error correction has also benefited from the use of ML methods: NNs were trained by supervised learning to quickly decode error correcting codes [369], reinforcement learning [370] and recurrent NNs [371] have been used to optimize specific sets of surface codes, and new quantum error correction codes can be found from scratch by reinforcement learning [372].

ML has also been widely applied in quantum control. In particular, feedback ML methods have uncovered previously unknown ways to prepare highly entangled states [373], an approach that was strengthened by the use of reinforcement learning [374]. Reinforcement learning has also been used to find fast, high-fidelity driving protocols from an initial to

[5]The author has been working on optimizing the training of restricted Boltzmann machines for homodyne tomography, with open code available at https://github.com/thomaskoerber/RBM-quantum-tomography-for-continuous-variable-systems.

a target state in many-body systems [375]. Furthermore, photonic architectures for universal computation have been designed with the help of ML methods [376]. Also, the certification of quantum simulations [377] and of boson sampling [378] has been addressed with ML.

ML methods have even proved helpful for quantum foundation tasks: using reinforcement learning of restricted Boltzmann machines, maximal violations of Bell inequalities for a given set of measurement observables have been found [379], and ML methods could even discover new kinds of nonlocality that previously have been inaccessible [380]. Finally, neural networks resembling the structure of local-hidden-variable models have been used as an oracle for network nonlocality (cf. Sec. 3.1.2), in the sense that if the observed behaviour cannot be learned by the NN, it is suspected to be nonlocal [315].

In the field of quantum metrology, ML has been applied in a wide variety of ways. ML has been used to automatically adjust control parameters in adaptive measurement protocols [381], which has been applied in experimental quantum metrology with single photons [382]. NN estimators have been used as a fast subroutine in variational quantum eigensolvers [383]. Furthermore, the supervised learning of NNs has been applied to frequentist parameter estimation [384] and to Bayesian parameter estimation [385] (see Ref. [386] for a comparison of frequentist and Bayesian parameter estimation). In this context, we want to mention Ref. [387] which discusses that, while the ML task of classification is a naturally Bayesian problem, the ML task of regression (that includes frequentist parameter estimation) is a frequentist problem that is solved by a Bayesian estimator.

Finally, we want to mention that the supervised learning of NNs has been applied to the detection of nonclassicality from multimode homodyne detection data [388], an approach that is similar to the one we will employ in the following. In Ref. [388], the authors focus on the negativity of multimode Wigner functions. We will now discuss results that contribute to the detection of single-mode nonclassicality in different quantum optical measurement protocols.

4.3 Nonclassicality detection with homodyne measurements

We now focus on the task of detecting nonclassicality, defined as the negativity of the Glauber–Sudarshan P function, in typical quantum optical measurement schemes. In this section, we consider the widely used detection method of balanced homodyne detection. In Sec. 4.3.1, we first give a brief background on single-mode homodyne detection and the difficulty to verify nonclassicality. Then, in Sec. 4.3.2, we address the task with the help of NNs. We present the efficiency of this ML approach to identify nonclassicality for real experimental homodyne data of different quantum states. The results of this section are

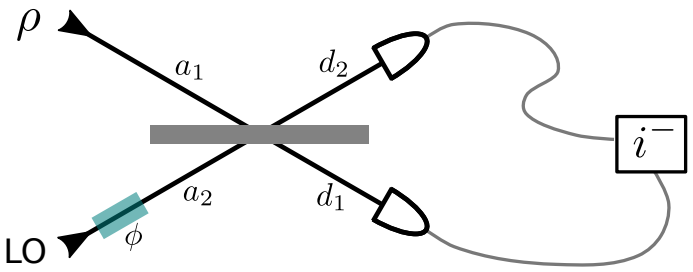

Figure 4.2: Homodyne detection (HD) scheme. The quantum state ρ in mode a_1 is mixed with the local oscillator (LO) in mode a_2 with a balanced beam splitter (BS). The phase shift ϕ acting on the LO is called the quadrature angle and represents the measurement setting. The BSs outgoing modes (d_1 and d_2) are measured using proportional photodetectors whose signals are subtracted and then amplified.

published in Ref. [3].

4.3.1 Homodyne measurements

Similar to what we have seen in the case of finite (discrete-variable) quantum systems in Sec. 1.3, to obtain different types of insights about a quantum state at hand, different measurement strategies are required. For instance, an ideal number-resolving detector measures the number operator $n = a^\dagger a$ of the optical field operator a. This measurement cannot extract all information from the quantum state because it is insensitive to the state's phase. In fact, phase-sensitive measurements must always rely on interference. The easiest and widely used phase-sensitive detection scheme of a quantum-optical state is balanced homodyne detection (HD).

In balanced HD, the quantum-optical state ρ to be measured is mixed with a reference beam, the so-called local oscillator (LO), on a balanced beam splitter (BS), see Fig. 4.2. The balanced BS converts the (annihilation operators of the) incoming modes a_1 and a_2 to the outgoing modes d_1 and d_2 as

$$\begin{bmatrix} d_1 \\ d_2 \end{bmatrix} = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & i \\ i & 1 \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \end{bmatrix}, \tag{4.6}$$

cf. also Eq. (3.25). The two outgoing modes are measured by proportional photodetectors (i.e., detectors that output a signal proportional to $\left\langle a^\dagger a \right\rangle$) whose output is subtracted and amplified. Thus, the measured operator is the difference of the number operators of the two outgoing modes,

$$\begin{aligned} I^- &= d_1^\dagger d_1 - d_2^\dagger d_2 && (4.7)\\ &= \frac{1}{2}(a_1^\dagger - i a_2^\dagger)(a_1 + i a_2) - \frac{1}{2}(-i a_1^\dagger + a_2^\dagger)(i a_1 + a_2) && (4.8)\\ &= i(a_1^\dagger a_2 - a_1 a_2^\dagger). && (4.9) \end{aligned}$$

The quantum state ρ that we want to measure is entering the measurement apparatus in mode a_1, while the LO entering in a_2 is prepared in the coherent state $\left|\alpha_{LO} e^{-i(\phi+\pi/2)}\right\rangle$

(note that this already includes the phase setting ϕ that is depicted as a phase shift in Fig. 4.2). Then, the measured homodyne current $\langle I^- \rangle$ is given by

$$\left\langle I^- \right\rangle = i \left\langle a_1^\dagger \alpha_{LO} e^{-i(\phi+\pi/2)} - a_1 \alpha_{LO} e^{i(\phi+\pi/2)} \right\rangle = \alpha_{LO} \left\langle x(\phi) \right\rangle , \tag{4.10}$$

where $x(\phi) = a_1 e^{i\phi} + a_1^\dagger e^{-i\phi}$ is the phase-rotated quadrature operator. If the mean photon number of the reference mode a_2 (i.e., $|\alpha_{LO}|^2$) is much larger than the state's mean photon number, the HD scheme yields a measurement of the quadrature distributions described by marginal distributions of the Wigner function [318]. Here, the control phase ϕ, also called the quadrature angle, dictates which marginal distribution is measured.

After performing a large number of HD measurements with different phase settings ϕ, one obtains the complete information of the state ρ. The measurement results can be combined using different advanced post-processing techniques (that we briefly discuss below) to reconstruct the full Wigner function and thus the state ρ. Quantum tomography based on balanced homodyne detection and Wigner-function reconstruction has been among the standard tomography techniques in different quantum systems such as light [389], molecules [390] or trapped atoms [391, 392].

Detecting nonclassicality in homodyne measurements

How can one experimentally verify that a quantum state is nonclassical in the HD scheme? The straightforward method that comes to mind is to perform a full quantum state tomography and then check for negativity of the P or Wigner functions. This approach is possible in principle and has also been used [389, 390, 391, 392]. However, quantum state tomography with HD both demands a very large amount of measurement data and advanced post-processing techniques to correctly reconstruct the Wigner function. This difficulty is due to the fact that reconstructing the Wigner function from its marginals is a very intricate operation [393]. Different regularization tools have been proposed such as, e.g., the inverse Radon transform, pattern functions, or maximum-likelihood reconstructions, see Ref. [393] for an overview.

An alternative approach to full quantum state tomography has been proposed in Ref. [394], where nonclassicality of phase-randomized states (i.e., states that are not phase dependent) was certified with the help of semidefinite programming, requiring significantly less measurement HD data than a full quantum state tomography.

Recently, the decomposition of a classical state into a mixture of coherent states, cf. Eq. (4.2), has been used to derive phase-space inequalities that hold for different quasiprobability distributions at a specific point α in phase space [395, 396]. One example is the inequality $W(\alpha) - 2\pi Q(\alpha)^2 \geq 0$ (W and Q are the Wigner and Husimi Q functions, respectively) that holds for classical states [395, 396]. For experimental data from single-photon-added thermal states (that are nonclassical), the inequality was violated resulting

in the detection of nonclassicality in a larger detection-efficiency domain than previous methods [397]. However, to calculate the Wigner function, also here quantum state tomography has to be performed that requires a large amount of measurement data.

We also want to note that the nonclassicality of a specific subset of nonclassical states, the states that show a reduced quadrature variance for a HD measurement with a fixed quadrature angle, can be easily detected by direct sub-shot-noise conditions on the variance of the quadrature distributions [318], as we will see in more detail in Sec. 4.3.2. However, we emphasize that the sub-shot-noise condition only detects nonclassicality for a strict subset of the nonclassical states, as many important nonclassical states such as, e.g., single photon states, do not show a sub-shot-noise variance in HD measurements.

To summarize, we see that the identification of nonclassicality from quadrature data usually relies on large amounts of measurement data and advanced post-processing techniques such as, e.g., the reconstruction of quasiprobability distributions. In the following, we provide a direct and fast nonclassicality identifier based on a trained NN that also works for comparatively small amounts of measurement data.

4.3.2 Identifying nonclassicality with NNs from homodyne measurement data

We now examine the potential of NNs as fast nonclassicality identifiers. We first describe in detail how to accordingly train NNs for this task, and then discuss their performance in various settings.

Training the NN

To create a NN nonclassicality identifier, the main idea is to train a NN that, when given experimental data from balanced HD along some fixed quadrature angle ϕ as input, outputs a nonclassicality prediction. As described above, the HD measurement current can be used to infer the expectation value of the quadrature operator $x(\phi)$,

$$\langle x(\phi)\rangle = \left\langle a e^{i\phi} + a^\dagger e^{-i\phi} \right\rangle , \tag{4.11}$$

where the measurement setting ϕ is called the quadrature angle. In general, the result of this measurement can be any (unbounded) real number. After performing (or simulating) many HD measurements with a given ϕ, one obtains a list of outcomes x that are sampled according to the quadrature distribution $p(x)$ of the state, see Appendix A for details and Figs. 4.5, 4.7, and 4.9 for exemplary quadrature distributions. In order to standardize the input of the NN, we bin the measured or simulated HD data into 160 equally-sized intervals that cover the HD outcome range $x \in [-8, 8]$, and then convert the measurement results to a normalized histogram on this grid. A normalized histogram is useful because it enables

the NN to process inputs from HD measurements of various sample sizes. The input layer thus consists of 160 neurons and is given the HD measurement histogram. The output layer is a single neuron that serves as the nonclassicality prediction. In our situation, we have found optimal performances for a fully-connected NN with three hidden layers of sizes 64, 32, and 16, cf. Fig. 4.1. The hidden-layer neurons are activated with rectified linear units, and the output neuron is activated with a softmax function that bounds the possible output r to the interval $r \in [0, 1]$.

We note that by binning the HD data in the above manner, one has to restrict the quantum states' parameter ranges (in simulations and experiments) such that the probability of an event outside the histogram grid is negligible. Here, we have chosen the parameters such that $P(|x| > 8) < 10^{-6}$. The corresponding parameter ranges to simulate the different states is given in Appendix A. If the NN is applied to states that create results outside of this range, the histogram grid has to be adjusted accordingly.

The NN is trained by means of supervised learning, cf. Sec. 4.2. For this purpose, we simulate HD measurement data for different kinds of classical and nonclassical states and equip the simulated data with a label "0" for classical and "1" for nonclassical states. As classical states, we simulate measurements of coherent states, mixtures of coherent states[6], and thermal states. As nonclassical states, we simulate Fock states, squeezed-coherent states, and single-photon-added coherent states (SPACS). The states' parameters are randomly chosen in specific ranges, cf. Appendix A. Furthermore, we only use squeezed states that are squeezed along the optimal direction such that their quadrature distribution shows a sub-shot-noise variance, see below. This is because in other quadrature directions, the quadrature distribution of squeezed states coincide with or is even broader than the one of coherent states (here one also speaks of anti-squeezed states), such that a correct nonclassicality classification of the NN is impossible. Finally, also the range of SPACS is reduced because, for large amplitudes of the initial coherent state, SPACS strongly resemble coherent states, see below. To realistically simulate quantum-optical experiments, we include a finite total detection efficiency[7] of $\eta = 0.6$ [397]. The finite detection efficiency can be

[6]Here, there are different possibilities of which mixtures of coherent states to use. In the following, we have focused on mixtures of the form

$$\rho_{\mathrm{mix}} = \frac{1}{2}(|\alpha\rangle\langle\alpha| + |-\alpha\rangle\langle-\alpha|).$$

The reason is that both coherent and thermal states have single-peaked quadrature distributions, such that it was necessary to include a classical multi-peak distribution in the training for the NN not to identify many peaks as a nonclassical feature. A different choice that may naturally occur in experiments are phase-averaged coherent states,

$$\rho_{\mathrm{av}} = \frac{1}{2\pi}\int |\alpha e^{i\phi}\rangle\langle\alpha e^{i\phi}|\,\mathrm{d}\phi.$$

Therefore, we see that specific experimental conditions must be considered when optimally choosing the simulated states for the training data. We emphasize and discuss this point again below.

[7]In the literature, η is also called the quantum or system efficiency. We will simply use the term detection

modeled by a beam splitter mixing the ideal quantum state (with a transmission coefficient $\sqrt{\eta}$) with vacuum noise. It can describe realistic detectors, lossy channels and noisy state preparations. For a detailed discussion of how the HD measurement of the different states is simulated, also including finite detection efficiencies, we refer to Appendix A.

For each family of states, we simulate 2×10^4 input data vectors, each consisting of a histogram of HD measurements with a fixed quadrature angle ϕ. Here, the histogram of each input vector is constructed by data from 16000 single HD measurement results. The simulated data are then shuffled and divided into training data (80%) and validation data (20%). In each training epoch, the training data are split into batches of size 20 which are then used one after the other to train the NNs parameters. The NN is implemented using the python libraries *keras* and *tensorflow*, where we minimize the mean squared error using the optimization algorithm ADAM [398]. After each training epoch, we check the performance of the NN on the validation data. If it decreases for more than 10 epochs, the training is halted. As discussed above, this early-stopping strategy avoids overfitting [355]. For instance, the NN we discuss below (cf. Fig. 4.3) has been trained with 26 training epochs.

NNs performance on training data

The output of the NN for a given input is a real number $r \in [0, 1]$. In order to use the NN as a binary nonclassicality indicator, one must choose a threshold value t above which we say that the NN predicts nonclassicality. Higher values of t will result in less states that are identified as nonclassical but, importantly, also in less states that are *falsely* identified as nonclassical. As for entanglement witnesses, cf. Sec. 1.3, a false negative (i.e., identifying a nonclassical state as classical) is acceptable (and, in our case, unavoidable) while a false positive should not occur. Therefore, in the following, we choose the threshold value $t = 0.9$, a choice that can be adapted according to specific situations.

In this context, we want to emphasize that the NN nonclassicality prediction only provides an indication of whether a state is classical or nonclassical. A robust certification and verification of nonclassicality must necessarily rely on proven nonclassicality conditions and an appropriate discussion of statistical errors (similar to violations of Bell inequalities discussed in Ch. 3). Furthermore, even if the NN performs well on the states used in the training, there is no guarantee that it correctly classifies a state that has not been seen in the training. Here, again, a "pessimistic" NN is more valuable: false negatives are acceptable while false positives are not. This behaviour can either be achieved by including more types of classical states in the training (so more types of them must be recognized as classical), or by increasing the threshold value t.

We can now discuss the performance of the NN nonclassicality prediction in and be-

efficiency in accordance with Ch. 3.

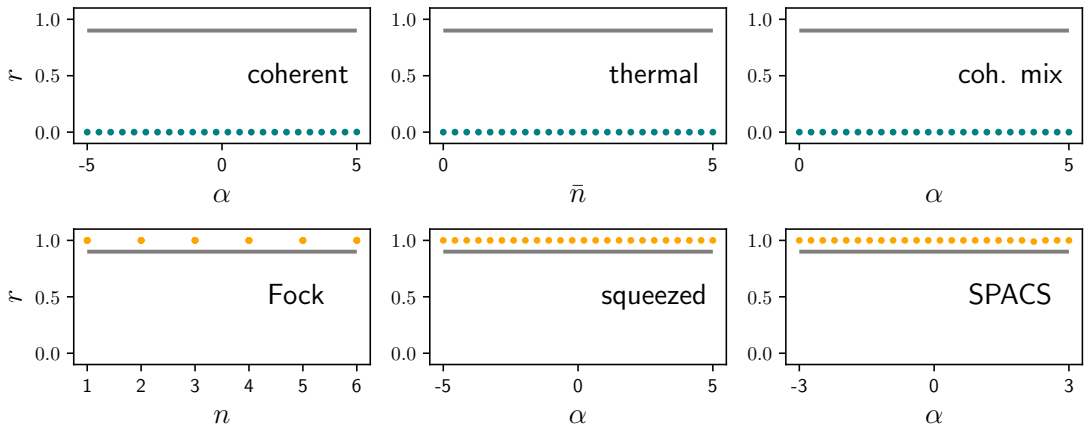

Figure 4.3: Nonclassicality prediction r of the neural network (NN) for the different families of states in the training data, consisting of classical states (top; coherent, mixtures of coherent, thermal) and nonclassical states (bottom; Fock, squeezed-coherent, single-photon-added coherent). α is the coherent amplitude, $\bar{n}$ is the average photon number, and n is the number of photons. The grey line is the threshold value $t = 0.9$ above which the NN predicts nonclassicality. For the Fock states, each n is tested four times. For the squeezed-coherent states, the randomly chosen squeezing parameter $\xi \in [0.5, 1]$ is not shown in the plot.

yond the training set. In Fig. 4.3, we show the NN's nonclassicality prediction r for the different states that have been used in the training. For each family of states, we plot r for 24 different simulations of the HD measurement for different parameters of the states. We see that the NN succeeds in correctly classifying all states used in the training phase.

NNs performance on experimental data from squeezed states

We now test the trained NN on data from real quantum-optical experiments. We first consider the detection of nonclassicality for squeezed states. The experiment was conducted in the group of B. Hage at the University of Rostock[8].

A brief sketch of the experimental setup to generate and measure squeezed states is shown in Fig. 4.4. A laser generates a coherent beam (532nm) that is split into a LO beam and a frequency-doubled (FD) pump beam (1064nm). The frequency-doubled pump beam enters an optical parametric amplifier (OPA). The OPA consists of a nonlinear optical crystal (7%MgO:LiNbO3) that, when pumped in the strong regime, effectively acts as the squeezing operator $S(\xi) = e^{(\xi a^{\dagger 2} + \xi^* a^2)/2}$ on the initial (red, not frequency-doubled) mode that is in its vacuum state in the OPA [318]. For more technical details of the experimental setup, we refer to Ref. [399].

The squeezed-state experimental data were recorded while continuously changing the quadrature angle $\phi \in [0, 2\pi]$. We have divided the data into 125 groups such that each group consists of $\sim$ 16000 measurement points that were taken in the interval

[8]We thank B. Hage for kindly providing the experimental quadrature data for the measurement of the squeezed states.

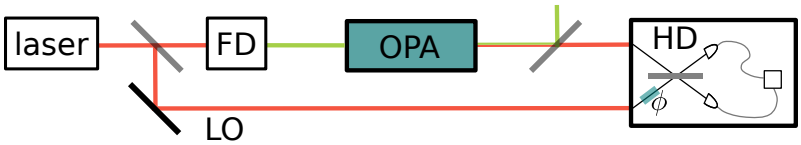

Figure 4.4: Experimental setup to generate and measure squeezed states with an optical parametric amplifier (OPA). A laser generates a local oscialltor (LO; red) beam and a frequency-doubled pump beam (green). The pump beam enters the OPA, a nonlinear crystal, resulting in a squeezing operation acting on the initial mode (red, that is prepared in the vacuum state) [318]. Finally, using the LO beam, the outcoming mode is measured in homodyne detection (HD). For technical details see Ref. [399].

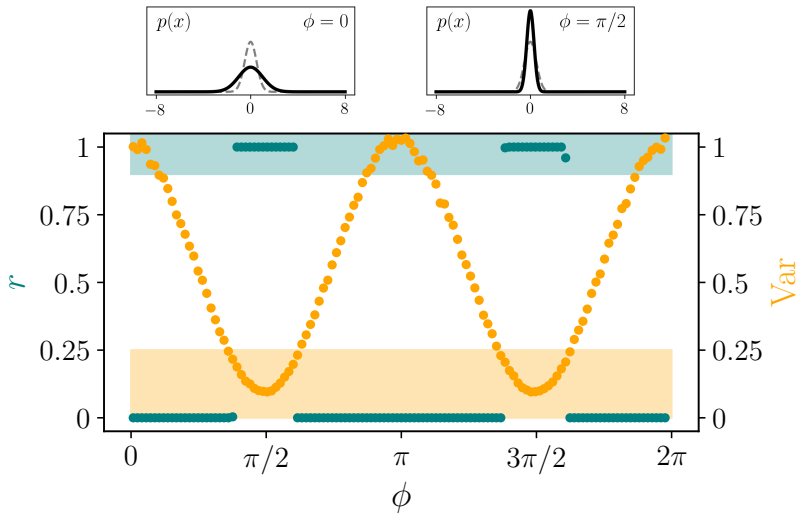

Figure 4.5: Bottom: The nonclassicality prediction r of the trained neural network (NN) for experimental homodyne detection (HD) data of squeezed states for measurements along different quadrature angles ϕ (blue), compared to the variance of the same data (yellow). Shaded areas indicate where the NN and the sub-shot-noise condition for the variance predict nonclassicality. Top: Exemplary quadrature distributions that are broader (left; $\phi = 0$) or narrower (right; $\phi = \pi/2$) than the vacuum quadrature distribution (dashed line). The figure is taken from Ref. [3].

$\Delta\phi = 2\pi/125$ (the small phase difference within each group is negligible with respect to experimental phase fluctuations). From each group, we construct an input vector for the NN and test its nonclassicality prediction r. The prediction r is shown as the blue points in Fig. 4.4. We also include two exemplary homodyne distributions on top of Fig. 4.4 for quadrature angles ϕ that result in a broader (left; solid line; $\phi = 0$) and a narrower (right; solid line; $\phi = \pi/2$) distribution than that of the vacuum state (dashed lines).

Furthermore, we include the well-known nonclassicality criterion of sub-shot-noise variance as yellow points in Fig. 4.4. This criterion says that any quadrature distribution with a variance smaller than the vacuum variance, $\mathrm{Var}[x(\phi)] < 1/4$, can only stem from a measurement of a nonclassical state. We observe that the NN prediction and the sub-shot-noise criterion detect nonclassicality in the same parameter domain.

To summarize, we confirm that the NN (that was trained with squeezed states) is able to correctly learn the sub-shot-noise criterion of nonclassicality and is able to correctly pre-

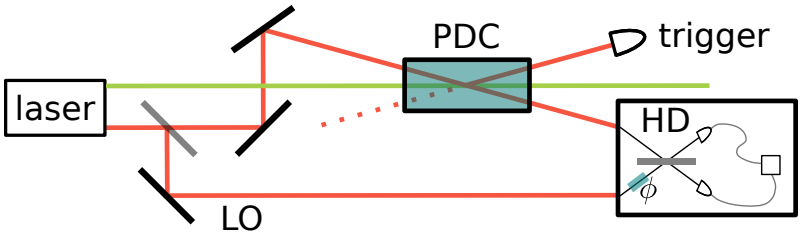

Figure 4.6: Experimental setup to generate and measure single-photon-added coherent states (SPACS). A laser produces a signal beam and a frequency-doubled pump beam. After branching off the local oscillator (LO) with a beam splitter, the signal beam enters a parametric-down-conversion (PDC) crystal together with a vacuum mode (dotted line). Heralded by the detection of a single photon in the trigger detector, the signal beam is prepared in a SPACS and measured by homodyne detection (HD).

dict nonclassicality of experimental data from squeezed states. If one was only interested in the sub-shot-noise feature of nonclassicality, the well-known sub-shot-noise condition of the variance would suffice. However, as we see below, the NN also recognizes other nonclassicality features beyond sub-shot-noise variance, making it advantageous for other nonclassical states or for situations where the type of the state is not known.

We also note that for the NN nonclassicality prediction, there is no need to know the exact parameters of the experiment such as the squeezing parameter $|\xi|$ or the detection efficiency η. This shows the easy applicability and practicability of the NN nonclassicality identification, for instance when searching for optimal experimental parameters.

NNs performance on experimental data from SPACS

Next, we test the NN on experimental HD data from a second family of nonclassical states, single-photon-added coherent states (SPACS). The experimental data were recorded in the group of M. Bellini and A. Zavatta in the University of Florence.

SPACS represent an interesting class of states to test the performance of the NN because they exhibit a variety of different nonclassical features. First, note that for any seed coherent state $|\alpha\rangle$, SPACS are nonclassical. However, for large seed coherent amplitudes α, they strongly resemble slightly displaced coherent states (see below). For small seed amplitudes, they show the nonclassical features of single photon Fock states such as the negativity of the Wigner function. Finally, for intermediate amplitudes, they show quadrature squeezing in optimal quadrature angles, as we also discuss below.

To create SPACS, one has to apply the creation operator $a^\dagger$ on a coherent state $|\alpha\rangle$. The experimental setup to create and measure SPACS in HD is shown in Fig. 4.6. Again, a laser produces a signal beam and a frequency-doubled pump beam. A BS is used to branch off the LO reference beam. Then, the signal beam enters a type-I beta-barium borate crystal that induces parametric down conversion (PDC), when pumped with an intense frequency-doubled pump beam. The PDC process acts on the two incoming modes a_1 and a_2 (of which one is the signal beam prepared in the coherent state $|\alpha\rangle$, and the

second is in the vacuum state $|0\rangle$ indicated as the dotted line) by means of [318, 400]

$$U_{\mathrm{PDC}} |\alpha\rangle |0\rangle = e^{g(a_1^\dagger a_2^\dagger - a_1 a_2)} |\alpha\rangle |0\rangle \tag{4.12}$$

$$= (1 + g a_1^\dagger a_2^\dagger) |\alpha\rangle |0\rangle + \mathcal{O}(g^2) \tag{4.13}$$

$$= |\alpha\rangle |0\rangle + g a_1^\dagger |\alpha\rangle |1\rangle + \mathcal{O}(g^2), \tag{4.14}$$

where, in the second equality, we have assumed a small parametric gain, $g \ll 1$. Hence, we see that if we detect a single photon in the additional outgoing mode, the signal outgoing mode is prepared as a SPACS. We thus include a trigger photo detector after the PDC crystal to herald the creation of a SPACS which is finally measured using HD. For more technical details of the experiment, see Refs. [3, 400].

The experimental data consist of HD measurements of SPACS with 14 different seed coherent amplitudes α. For each α, the SPACS are measured along 11 different quadrature angles ϕ, and data from 15963 measurements per quadrature angle and α are collected. Each of these sets of results creates an input vector of the NN. Different sources of noise such as a beam mismatch, optical losses, electronic noise, and detector efficiency all result in an overall detection efficiency of $\eta \approx 0.6$ (this is why we have simulated the training data using a similar η).

In Fig. 4.7, we show the performance of the NN on the experimental HD data of SPACS. In Fig. 4.7(a; bottom), we plot the binary prediction of the NN as a function of the seed coherent amplitude α and the quadrature angle ϕ. Here, experimental data that are classified as nonclassical by the NN (i.e., data resulting in an output $r > 0.9$) are marked yellow, while data that are classified as classical are colored in blue. Furthermore, in Fig. 4.7(b), we explicitly plot the output of the NN, r, along $\sin\phi \approx 0$. We can identify different parameter ranges: For small α, the NN correctly classifies the data as nonclassical. For intermediate α, SPACS are only identified as nonclassical along $\sin\phi \approx 0$. For larger α, the NN does not recognize the nonclassicality of SPACS anymore. For comparison, we show the prediction of the NN for experimental data of (classical) coherent states for the same parameters as in Fig. 4.7(c).

To better understand the performance and the difficulties of the nonclassicality identification by the NN, we include a few exemplary homodyne distributions $p(x)$ on top of Fig. 4.7(a). For small $\alpha < 0.5$, the homodyne distributions resemble the ones of a single-photon Fock state and are recognized as nonclassical (see the homodyne distributions for $\alpha = 0.32$). Furthermore, for intermediate α, the quadrature distributions differ from the ones of coherent states in the optimal directions ($\sin\phi = 0$), while in other directions, the distribution is very similar to the ones of coherent states. Finally, for larger $\alpha > 3$, SPACS start to resemble coherent states with a slightly displaced coherent amplitude [400, 401, 402] and are thus classified as classical in any direction. In the shown homodyne distributions for $\alpha = 1.9$ and $\alpha = 3.2$, we see the resemblance to coherent

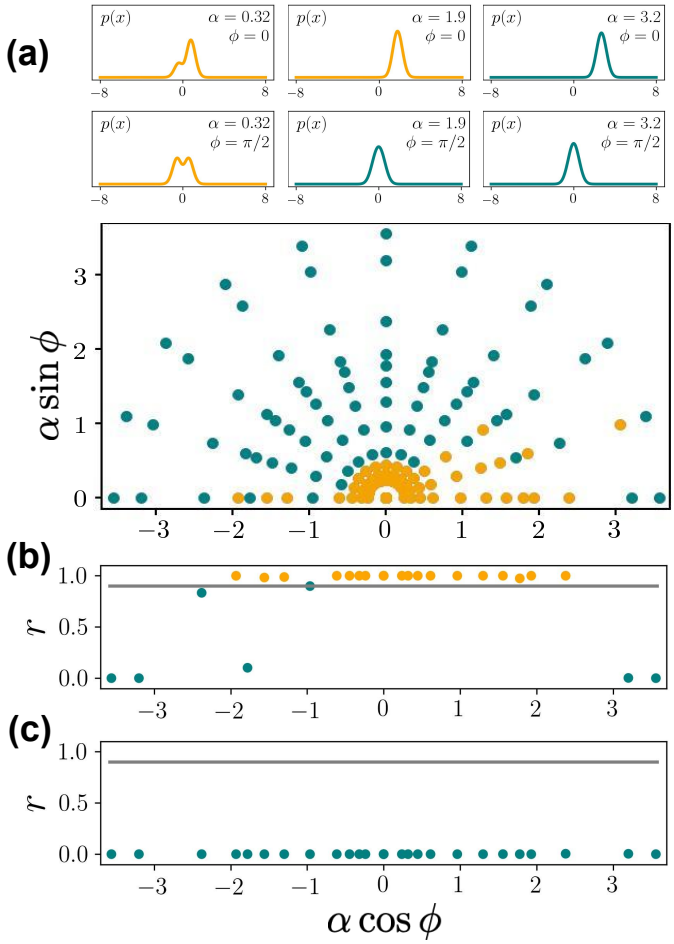

Figure 4.7: (a) bottom: Binary nonclassicality prediction of the neural network (NN) for experimental homodyne detection (HD) data from single-photon-added coherent states (SPACS) with different seed coherent amplitudes α and measured along different quadrature angles ϕ. Data classified as nonclassical (classical) are shown as yellow (blue) points. top: Exemplary quadrature distributions for different combinations of α and ϕ. (b) Nonclassicality prediction r of the NN for SPACS measured along $\sin\phi = 0$. (c) Nonclassicality prediction of the NN for coherent states with the same parameters and measurement settings as in (b). The figure is taken from Ref. [3].

state distributions and thus the difficulty of the classification task in this regime. Actually, due to these similarities, we have only used SPACS for $\alpha < 3$ and $\sin\phi = 0$ in the training because, otherwise, the NN is trained with similar distributions with different labels and thus might falsely classify coherent states as nonclassical.

To summarize, we see that the NN is able to recognize different nonclassical features of SPACS. In particular, in the optimal direction $\sin\phi \approx 0$, nonclassicality is identified for a wide range of seed coherent amplitudes. This highlights the strength and flexibility of the NN to identify optimal parameter settings for multiple nonclassical features. In the following, we will see in detail that the NN learns to recognize different features as nonclassical, even beyond sub-shot-noise variance or similarity to Fock states.

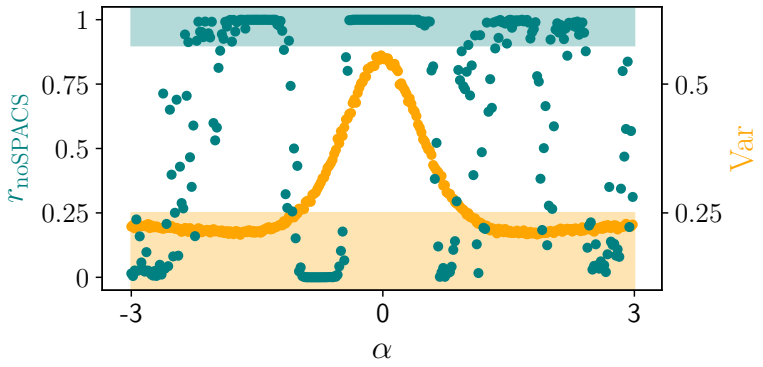

Figure 4.8: Nonclassicality prediction r_{noSPACS} of a neural network (NN) that was trained without single-photon-added coherent states (SPACS), on simulated measurement data of SPACS (blue) as a function of the seed coherent amplitude α. For comparison, we show the variance of the same data in yellow. Shaded regions indicate where the NN and the sub-shot-noise variance condition recognize nonclassicality. The figure is taken from Ref. [3].

Beyond single-feature recognition

To examine which nonclassical features are recognized by the NN, we compare the output r of the NN that we have discussed above, with the output r_{noSPACS} of a NN that is trained with the same training states as above but excluding SPACS. We compare the two NNs on simulated measurement data of SPACS along a quadrature angle $\phi = 0$. The prediction of the NN trained with SPACS was shown above in Fig. 4.3, where all simulated data in the considered parameter regime was classified as nonclassical ($r \approx 1$).

In contrast, the prediction r_{noSPACS} of the NN trained without SPACS in shown in blue in Fig. 4.8. We see that the NN recognizes SPACS as nonclassical for small seed coherent amplitudes $|\alpha| < 0.5$ and again for intermediate amplitudes $|\alpha| \in [1, 2]$. To understand this behaviour, we also show the variance of the simulated data as yellow points in Fig. 4.8. We see that, for $|\alpha| > 1$, SPACS measured along $\phi = 0$ show a sub-shot-noise variance, $\mathrm{Var}[x] < 1/4$. However, for increasing $|\alpha|$, the variance approaches the variance of a vacuum or coherent state ($\mathrm{Var}[x] = 1/4$). Since coherent states are (and should be) classified as classical, the NN only identifies the sub-shot-noise variance in the intermediate regime. Furthermore, for small α, SPACS resemble single-photon Fock states, as we discussed above, and are thus also recognized as nonclassical.

This analysis shows two points. First, as commonly seen in the training of NNs, the NN exhibits some (but limited) generalizability beyond the data that were used in the training. Here, this is due to similarities of the different families of states. Second, we see that the NN trained with SPACS must learn to recognize different features as nonclassical: Beyond sub-shot-noise and Fock state features, it also learns nonclassical features that cannot be recognized by the latter. The nonclassicality feature in this case would be "similarity to SPACS". Note that this discussion also highlights the necessity of a deep NN architecture

for the task of learning different nonclassicality features: simple models such as, e.g., linear regression, are not able to recognize these different nonclassical characteristics.

To conclude, we see that the NN effectively learns to recognize multiple features of the quadrature distributions as nonclassical at the same time. This can be advantageous because, usually, different tests must be conducted to see the different features. Furthermore, we see that also states that were not used in the training phase can be correctly recognized as nonclassical. Next, we will further elaborate and examine this second point.

NNs performance beyond training data

Now, we test the NN (that was trained with SPACS) on a further class of states that was not used in the training. Due to their popularity in foundational questions and quantum technological applications, we consider the class of the so-called (odd) cat states

$$|\alpha_-\rangle = \frac{1}{\sqrt{2-2e^{-2|\alpha|^2}}}(|\alpha\rangle - |-\alpha\rangle). \tag{4.15}$$

As a superposition (and not a mixture) of two coherent states, cat states are nonclassical for all α.

In Fig. 4.9, we show the NN's prediction r as a function of α for the quadrature angles $\phi = \pi/2$ (a) and $\phi = \pi/4$ (a). For both quadrature angles, cat states are recognized as nonclassical in a wide range of α. In particular, cat states are seen as nonclassical in the regions of small α. This behaviour can be understood by the fact that, in this parameter range, odd cat states resemble single-photon Fock states (in fact, one can show that $\lim_{\alpha\to 0}|\alpha_-\rangle = |1\rangle$). For larger α, cat states are not classified as nonclassical anymore. Surprisingly, the NN detects nonclassicality in a larger domain when measured in the quadrature direction $\phi = \pi/4$ (note that, for perfect efficiencies, cat states show the most nonclassical features, i.e., fast oscillations, in their quadrature distribution along $\phi = \pi/2$, while along $\phi = 0$, the distribution coincides with the one of a mixture of coherent states).

Again, to better understand the prediction of the NN, we show exemplary homodyne distributions above the plots in Fig. 4.9. Here, we include the quadrature distribution with a realistic detection efficiency $\eta = 0.6$ (solid lines; that was used in the simulation of the measurement data) and the one of an ideal noiseless situation with detection efficiency $\eta = 1$ (dashed lines). In the noiseless case, the quadrature distribution shows fast oscillations (indicating nonclassicality) for any α when measured in the optimal direction $\phi = \pi/2$. However, the finite efficiency effectively smooths away the interference pattern such that the distributions approach the ones of a coherent state. In contrast, when measured along $\phi = \pi/4$, the distribution longer shows a double-peak structure that partially resembles the distribution generated by a multi-photon Fock state. Thus, for $\phi = \pi/4$, cat states are recognized as nonclassical in a larger parameter domain.

At this point, we again want to emphasize an important caveat of the NN approach

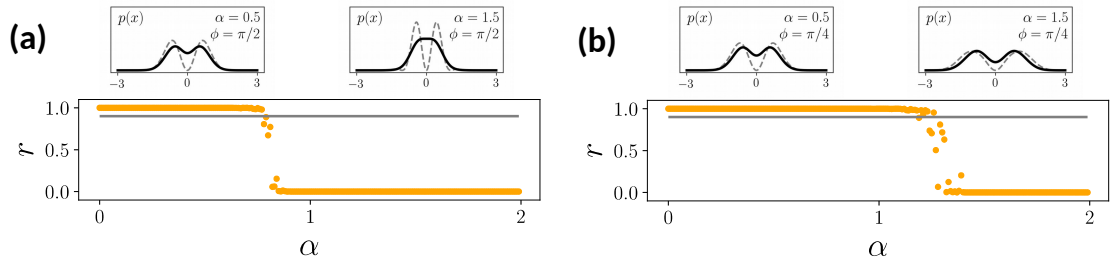

Figure 4.9: Nonclassicality prediction r of the neural network (NN) for simulated measurement data of odd cat states $|\alpha_-\rangle$ as a function of α, when measured along the quadrature angles $\phi = \pi/2$ (a) and $\phi/4$ (b). Above the plots, we show exemplary quadrature distributions for a realistic detection efficiency $\eta = 0.6$ (solid lines; which was used in the simulation of the measurement data) and an ideal detection efficiency $\eta = 1$ (dashed lines). The figure is taken from Ref. [3].

to identify nonclassicality. An optimal choice of the training data is crucial for a correct operation of the NN nonclassicality prediction. This point can be easily understood by means of an example: as we mentioned above, we have trained the NN with the mixture of coherent states of the form $\rho_{\text{mix}} = (|\alpha\rangle\langle\alpha| + |-\alpha\rangle\langle-\alpha|)/2$. If we instead choose to use only phase-averaged coherent states in the training, the NN classifies the states ρ_{mix} and thus also cat states measured along $\phi = 0$ as nonclassical for larger values of α. This false classification of the classical states ρ_{mix} highlights the limit of the NN nonclassicality identification: generally, there could always exist a classical state that is falsely classified as nonclassical by the NN. This shows that the exact experimental conditions have to be taken into account when choosing the optimal set of training states of the NN. Furthermore, it highlights that the NN cannot be seen as a universal nonclassicality certification. For such a certification of nonclassicality, proven nonclassicality conditions have to be checked including a proper discussion of statistical errors.

Influence of the sample size and practical advantage

Finally, we briefly want to mention a last beneficial characteristic of the NN approach that shows its applicability as a fast nonclassicality indicator requiring only small amounts of measurement data. We again test the performance of the NN trained above on experimental data of SPACS – but now we reduce the sample size that is used to construct the normalized histogram used as an input. In Fig. 4.10, we show the NN's prediction r for experimental data of SPACS and coherent states for the parameters $\alpha = 0.32$ and $\phi = 0$, as a function of the sample size. We see that the NN (that was trained using a sample size of 16000, as indicated by the vertical line) correctly classifies both states even for largely reduced sample sizes. Only for very small sample sizes ($\leq$ 800), the NN shows a finite error rate such that it cannot further be used as a NN classifier.

This analysis demonstrates that the NN is very flexible when applied to experimental data, even once trained. Furthermore, its prediction holds valid until the sample size is

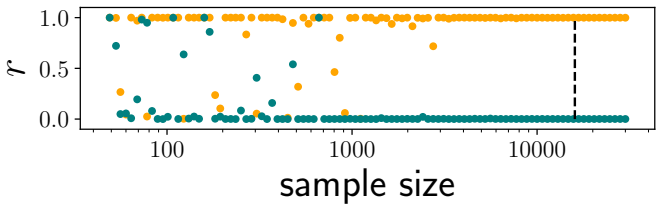

Figure 4.10: Nonclassicality prediction r of the neural network (NN) for experimental data of single-photon-added coherent states (SPACS; yellow) and coherent states (blue) for the parameters $\alpha = 0.32$ and $\phi = 0$, as a function of the sample size that is used to construct the input vector of the NN. The vertical line indicates the sample size that was used in the training phase. The figure is taken from Ref. [3].

significantly reduced. We thus see that the NN offers a very practical tool for fast online classification during measurements or fast (pre-)sorting of measurement data. Finally, we note that the performance of the NN on small sample sizes can be further improved by training it with measurement data that are simulated using the same sample size.

4.4 Nonclassicality detection with click-counting measurements

In this section, we consider a second commonly applied measurement method in quantum optics, the technique of multiplexed click-counting measurements. In Sec. 4.4.1, we first introduce this measurement protocol and briefly discuss how it can be used to certify nonclassicality. Then, in Sec. 4.4.2, we again train a NN via supervised learning to identify nonclassicality for different quantum-optical states from their simulated click-counting measurement data. The results of this section are published in Ref. [4].

4.4.1 Multiplexed click-counting measurements

The technique of multiplexed click-counting measurements is similar to photon-counting measurements that measure the observable $n = a^\dagger a$. However, instead of using a number-resolving photodetector, the incoming mode is divided into several modes (i.e., it is multiplexed) that are each measured with an on-off detector, see, e.g., Refs. [403, 404, 405]. An on-off detector is a photodetector that has a binary output, indicating either the presence of (any number of) photons or the absence of photons. Ideally, an on-off detector is described by the POVM elements $m = |0\rangle\langle 0|$ (no photon detected; no click) and $m^\perp = 1 - m$ (click).

There are different ways to split the incoming mode into multiple detection modes. Three possibilities are shown in Fig. 4.11: (a) the spatial profile of the incoming beam can be widened by lenses and fed into different detectors, (b) the beam can be divided into several temporal modes (in so-called time-bin multiplexing), or (c) the beam can be divided into different spatial modes by means of beam splitters. Finally, each multiplexed

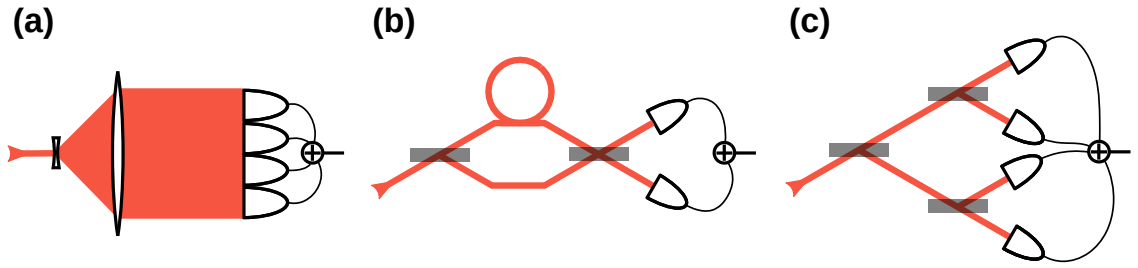

Figure 4.11: Different techniques of multiplexing an incoming mode, by means of an (a) array detector, (b) temporal multiplexing, or (c) spatial multiplexing. The multiplexed modes are measured with on-off detectors. Here, the incoming beam is multiplexed into four modes.

mode is detected in an on-off detector. While in Fig. 4.11, we show these multiplexing techniques for $N = 4$ multiplexed modes, similar constructions can produce $N = 2^n$ ($n \in \mathbb{N}$) multiplexed modes. Later in this section, we will focus on multiplexing with $N = 16$ measurement modes.

The outcome of the multiplexed click-counting detection is the total number of recorded clicks (indicated as the $\oplus$-element in Fig. 4.11). In other words, we do not record which detectors have produced the clicks. Thus, the measurement with N multiplexed modes has $N + 1$ different outcomes $k \in \{0, \ldots, N\}$. If we have balanced multiplexing (i.e., the incoming beam is divided with equal intensities into the multiplexed modes), the probability p_k to observe k clicks is given by the generalized binomial distribution[9] [406]

$$p_k = \left\langle : \binom{N}{k} \left(e^{-\eta n/N}\right)^{N-k} \left(1 - e^{-\eta n/N}\right)^k : \right\rangle , \tag{4.16}$$

where $\langle \cdot \rangle = \mathrm{tr}[\rho \cdot]$ is the expectation with respect to the state ρ, $n = a^\dagger a$ is the number operator and η is the detection efficiency of the on-off detectors. Furthermore, $: \cdot :$ is the normal-ordering operation that moves all creation operators a on the right of the annihilation operators $a^\dagger$ *without* respecting their commutation relations[10] [318].

The multiplexed click-counting measurement technique represents a simplification of the photon-number-resolving measurement because it only requires the simpler on-off

[9] To make the analogy to the binomial distribution even clearer, we note that the "no-click" POVM element of the on-off detector is [406]

$$m = |0\rangle \langle 0| =: e^{-\eta n} :,$$

where, as in Sec. 4.3.2, the finite detection efficiency η is modeled by a beam splitter with a transmission coefficient $\sqrt{\eta}$ before the on-off detectors, effectively resulting in the transformation $a \mapsto \sqrt{\eta} a$. Thus, Eq. (4.16) can be informally understood as the binomial distribution $\binom{N}{k} p^k (1-p)^{N-k}$ with $p = 1 - e^{-\eta n/N}$ representing the probability of a click, and the number operator n is divided by N due to the balanced multiplexing into N modes. For a rigorous derivation and a discussion we refer to Ref. [406].

[10] We emphasize that the normal-ordering operation $: \cdot :$ does *not* take into account bosonic commutation relations and, therefore, results in a different operator. This is different to simply rewriting an operator in its normally ordered form by means of the bosonic commutation relations, a technique that we have used in Sec. 4.1.1.

detectors (but, importantly, does *not* give access to the complete number distribution of the state [406]). Furthermore, in contrast to the homodyne detection method discussed in Sec. 4.3, it does not require a reference mode. This results in a phase insensitivity of the click-counting measurements that shows that click-counting measurements do not form a complete measurement basis that can be used for quantum state tomography. However, similar to photon-number-resolving measurements, even the restricted information from click-counting measurements can be used to certify nonclassicality, as we briefly discuss in the following.

Certifying nonclassicality in click-counting measurements

Different nonclassicality-certification techniques for multiplexed click-counting measurement have been developed. First, the Mandel Q_M parameter, that detects sub-Poissonian statistics as a nonclassical feature of quantum states of light [407], has been generalized to multiplexed click-counting measurements [406, 408], an approach that has proven successful to demonstrate nonclassicality in experiments [409]. Furthermore, nonclassicality conditions including higher moments of the number distribution have been derived [410]: the so-called matrix of moments $\mathbf{M}$ that is defined as

$$M_{ij} = \left\langle : (1 - e^{-\eta n/N})^{i+j} : \right\rangle , \tag{4.17}$$

with $i, j \in \{0, \ldots, \lfloor N/2 \rfloor\}$ ($\lfloor \cdot \rfloor$ is the floor function), has to be positive semidefinite if it stems from the measurement of a classical state of light [410]. Furthermore, the elements of $\mathbf{M}$ can be directly sampled in multiplexed click-counting measurements [410, 411]. The matrix-of-moments method has also been successfully applied in experiments to demonstrate nonclassicality of several quantum-optical states [411, 412, 413].

Various nonclassicality conditions can be derived from the condition that Eq. (4.17) is positive semidefinite. The easiest condition is the positivity the upper left submatrix of $\mathbf{M}$,

$$\mathbf{M}^{(2)} = \begin{bmatrix} 1 & \left\langle : 1 - e^{-\eta n/N} : \right\rangle \\ \left\langle : 1 - e^{-\eta n/N} : \right\rangle & \left\langle : (1 - e^{-\eta n/N})^2 : \right\rangle \end{bmatrix} . \tag{4.18}$$

If the smallest eigenvalue x_{mom} of $\mathbf{M}^{(2)}$ is significantly negative, the measured state has to be nonclassical. Later, in Sec. 4.4.2, we will compare the nonclassicality prediction of a NN to this matrix-of-moments condition.

Finally, we note that nonclassicality conditions have also been derived for different generalizations of balanced multiplexed click-counting measurement techniques, such as a combination of click-counting and unbalanced homodyne measurements [414, 415], or unbalanced multiplexed click-counting measurements [416], leading to a detector-device-independent verification of nonclassicality in multiplexed measurements [417].

4.4.2 Identifying nonclassicality with NNs from click-counting measurement data

We now train an artificial neural network (NN) to recognize nonclassicality of typical quantum-optical states from their multiplexed click-counting data. We will focus on the regime of small sample sizes for which the NN approach outperforms other nonclassicality conditions. Furthermore, we focus on detecting nonclassicality in the few-photon regime ($\bar{n} < 16$, where $\bar{n}$ is the mean photon number of the measured states) and thus choose a multiplexing strategy with $N = 16$ final detection modes, cf. Fig. 4.11. We note that for states that contain more photons, the state must be multiplexed into more detection modes because, otherwise, the on-off detectors saturate and simply always register a click, independent of the quantum-optical state. If the mean photon number in experiment is further restricted, also a multiplexing in fewer modes can be sufficient.

Training the NN

As in Sec. 4.3, we train a NN using supervised learning with different classical and nonclassical states to learn nonclassical features of typical quantum-optical states. The input data of the NN consists of a sampled estimation of the multiplexed click-counting probabilities, Eq. (4.16). Recall that the outcome of a single multiplexed experiment is the total number of clicks $k \in \{0, \ldots, 16\}$. Thus, the input layers of the NN contains 17 neurons, and an input data vector is a normalized histogram constructed with (simulated) data from m repetitions of the multiplexed click-counting measurement. We want to primarily focus on small sample sizes that highlight the advantages of the NN approach and hence consider $m = 1000$ and $m = 100$. Furthermore, we first discuss the case of perfect detection efficiency of the detectors ($\eta = 1$) and then investigate the case of a realistic detection efficiency $\eta = 0.6$ (see also Sec. 4.3).

The optimal architecture of the NN slightly varies with the quality of the measurement data (such as sample size m and efficiency η). We have found a good performance of the NN for three hidden layers containing each 50 neurons which are activated by the rectified linear unit. The output is again only one neuron (here activated by a sigmoid function), serving as the NN nonclassicality prediction. The NN is implemented with the python libraries *keras* and *tensorflow*.

In the training phase, we have used simulated measurement data of coherent and thermal states for the classical states (equipped with the label "0") and Fock and squeezed states for the nonclassical states (label "1"). The training data consist of 1000 measurement histograms per type of state, each of which is generated by the simulation of m single click-counting measurements. For each histogram, the state's parameters are chosen randomly such that the mean photon number $\bar{n}$ is distributed uniformly in $\bar{n} \in [1, \ldots, 16]$ (with a discrete distribution for Fock states). For more details of the simulation and the calculation

of the click probabilities for the different types of states using Eq. (4.16), see Appendix B. The simulated data are shuffled and split into training data (80%) and validation data (20%) to avoid overfitting (cf. Secs. 4.2 and 4.3 for discussions of overfitting). The NN is trained by minimizing the mean squared error via the optimization algorithm ADAM [398]. After each training epoch, the performance of the NN on the validation data is computed. The training is halted after the validation error increases for 10 training epochs.

We finally want to note that, for the task of recognizing nonclassicality from multiplexed click-counting measurements, the complexity and curve-fitting potential of deep NNs is required: taking a simple linear regression model as a baseline model for the same task, the classification of nonclassicality in the training data is inaccessible (see Fig. 3 in Ref. [4]).

NNs performance with perfect detectors

We now first consider a multiplexed click-counting measurement with ideal detection efficiency $\eta = 1$. In Fig. 4.12, we show the performance of the NN nonclassicality identification for sample sizes $m = 1000$ (a) and $m = 100$ (b). For this purpose, we simulate 100 input histograms for each family of states with mean photon numbers in $\bar{n} \in [1, 16]$ (note that for Fock states, we use $n = \lfloor \bar{n} \rfloor$). As in Sec. 4.3, we choose a threshold value of $t = 0.9$, indicated as the gray horizontal line. If the NN's output r is larger than the threshold, $r > t$, we say that the NN predicts nonclassicality. Recall that the choice of this threshold is arbitrary. However, a high threshold value is sensible because, ideally, there should be no false positives (i.e., classical states that are identified as nonclassical).

We see that for both sample sizes, the NN correctly learns to separate data from classical and nonclassical states of the training families. Note that for the small sample sizes $m = 100$, the NN's output begins to fluctuate. This shows that even smaller sample sizes result in a finite probability for a false nonclassicality detection of a classical state, such that the NN prediction is not reliable anymore. However, we want to emphasize that the sample size of $m = 100$ is already very small and it is surprising how well the NN can still separate the classical and nonclassical states. Furthermore, we note that due to the normalized form of the NN's input (the normalized histogram of sampled frequencies), a NN that is trained with a specific sample size also shows a very good performance for classifying data generated with a different sample size.

Now, we want to compare the nonclassicality prediction of the NN to the well-established nonclassicality condition of the matrix of moments, see Sec. 4.4.1 for an introduction and Ref. [410] for details. For this purpose, for each input measurement data, we compute the smallest eigenvalue x_{mom} of the 2×2-submatrix $\mathbf{M}^{(2)}$ of the matrix of moments, cf. Eq. (4.18)[11]. To obtain a significant nonclassicality detection without false

[11] Actually, in the following analysis, we do not compute the smallest eigenvalue of Eq. (4.18) but the

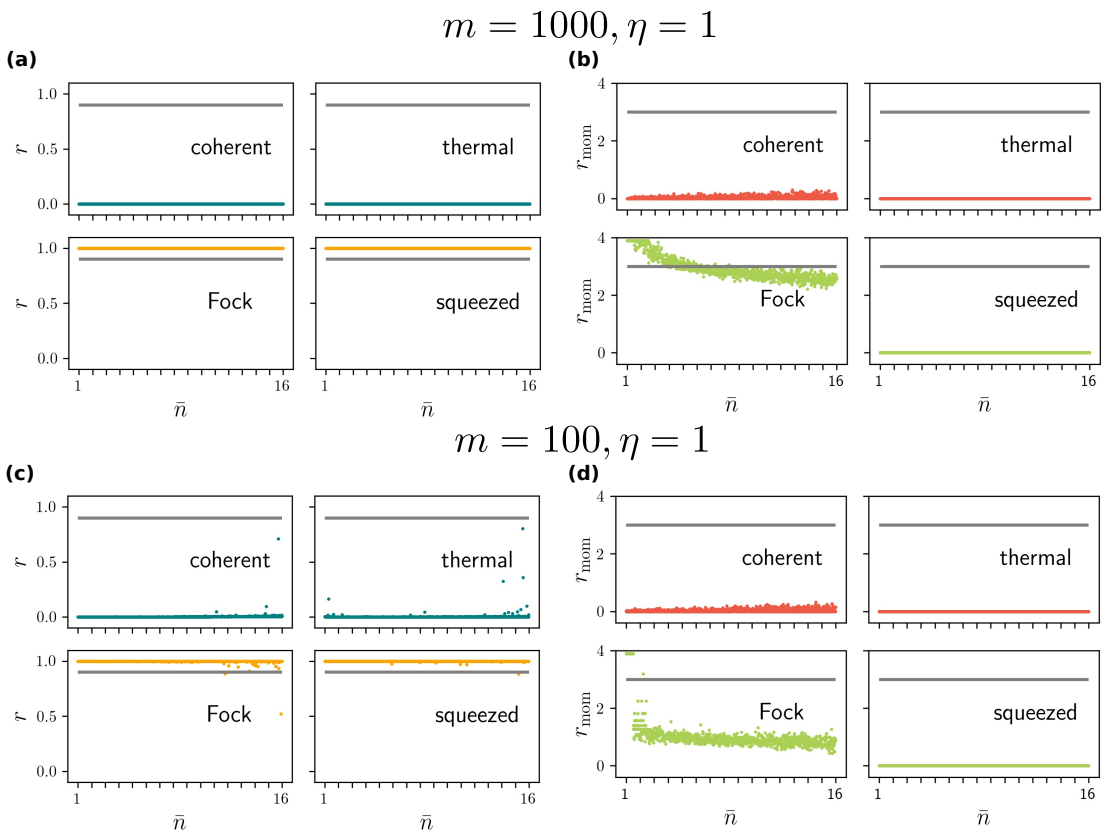

Figure 4.12: (a,c) Nonclassicality prediction r of the neural network (NN) for simulated multiplexed click-counting measurement data with perfect detectors, $\eta = 1$, generated from the training states (coherent and thermal states as classical and squeezed and Fock states as nonclassical) for input histograms constructed from sample sizes $m = 1000$ (a) and $m = 100$ (c), depending on the mean number of photons $\bar{n}$. For the simulation of Fock states, we used a photon number $n = \lfloor \bar{n} \rfloor$. For each state family, we have simulated 1000 input vectors. (b,d) Nonclassicality detection r_{mom} of the matrix-of-moments method (cf. Eq. (4.19)) for the same measurement data as in (a) and (c). Gray horizontal lines indicate the (chosen) threshold value above which nonclassicality is predicted.

positives, a proper investigation of error bars is necessary. Thus, we also compute the error of x_{mom}, Δ_{mom}, and define the nonclassicality prediction using the matrix of moments as

$$r_{\mathrm{mom}} = -\frac{x_{\mathrm{mom}}}{\Delta_{\mathrm{mom}}}. \tag{4.19}$$

The calculation of the error Δ_{mom} from measurement data is explained in detail in the supplementary information of Ref. [411]. We say that the matrix-of-moments method shows nonclassicality if $r_{\mathrm{mom}} > 3$, corresponding to an x_{mom} that is negative by three

smallest eigenvalue of the matrix

$$\tilde{\mathbf{M}}^{(2)} = \begin{bmatrix} 1 & \langle : e^{-\eta n/N} : \rangle \\ \langle : e^{-\eta n/N} : \rangle & \langle : e^{-2\eta n/N} : \rangle \end{bmatrix}.$$

The corresponding nonclassicality criterion is shown to be equivalent, i.e., the above matrix fails to be positive semidefinite whenever Eq. (4.18) does, see Appendix A of Ref. [418].

standard deviations. Again, this threshold value can be chosen differently, depending on how much one wants to suppress a false-positive rate.

The matrix-of-moment prediction r_{mom} of nonclassicality is shown in Fig. 4.12 for $m = 1000$ (b) and $m = 100$ (d). Note that if $r_{\mathrm{mom}} < 0$, we plot $r_{\mathrm{mom}} = 0$ for better visualization. We see that the matrix-of-moments method is insensitive to the nonclassicality of squeezed states and never detects them as nonclassical. For $m = 1000$, Fock states are recognized for small photon numbers $n < 8$, while for $m = 100$, the negativity of x_{mom} is not significant enough to detect nonclassicality. We also checked that the nonclassicality detection of the matrix-of-moments methods only slightly increases when considering higher moments (up to the full 8×8 matrix of moments, cf. Eq. (4.17)), and also here squeezed states are never detected as nonclassical. Thus, we see that the NN nonclassicality identification is more sensitive than the matrix-of-moments conditions (Fock states) and even predicts nonclassicality for states for which the matrix-of-moments nonclassicality detection fails completely (squeezed states).

Next, we test the nonclassicality prediction of the NN for different kinds of states that were not included in the training phase. In Fig. 4.13, we show the predictions of the NNs trained above for two new kinds of states: we first consider the nonclassical states of n-photon-added thermal states (NPATS), defined as $\rho_{\mathrm{NPATS}} = \mathcal{N} a^{\dagger n} \rho_{\mathrm{th}} a^n$ ($\mathcal{N}$ is a normalization constant), that are thermal states acted upon n times with the creation operator (compare with single-photon-added coherent states in Sec. 4.3). We simulate measurement data from NPATS for $n = 1$ (gray) and $n = 2$ (yellow). Second, we test the NN on even coherent states $|\alpha_+\rangle = \tilde{\mathcal{N}}(|\alpha\rangle + |-\alpha\rangle)$ ($\tilde{\mathcal{N}}$ is a normalization constant) that are similar to the cat states or odd coherent states tested in Sec. 4.3 and are also nonclassical for all $|\alpha| > 0$.

We see that NPATS are only recognized as nonclassical for small initial seed mean photon numbers of the thermal state. In this regime, they strongly resemble single-photon and two-photon Fock states. For larger seed mean photon numbers, NPATS are very similar to thermal states and are thus seen as classical. On the other hand, even coherent states are recognized in a significant parameter range ($|\alpha| < 4$), highlighting that the NN also generalizes beyond data from its training phase. For larger values of $|\alpha|$, the two coherent states in the superposition of $|\alpha_+\rangle$ become increasingly orthogonal, such that the resulting distribution approaches a mixture of the two coherent states. We also note that the performance is similar for the sample sizes $m = 1000$ (a) and $m = 100$ (c), showing again the surprising performance of the NN for small sample sizes. In comparison, we show the methods-of-moments nonclassicality prediction r_{mom} for the same data as in parts (a,c). This method does only recognize nonclassicality in the regime where the NPATS coincide with Fock states, and never recognizes nonclassicality for even coherent states.

Finally, we test the prediction of the NN for a mixture of a classical and a nonclassical state. In particular, we consider a mixture of a Fock state $|n\rangle$ and a coherent state $|\alpha\rangle$ with

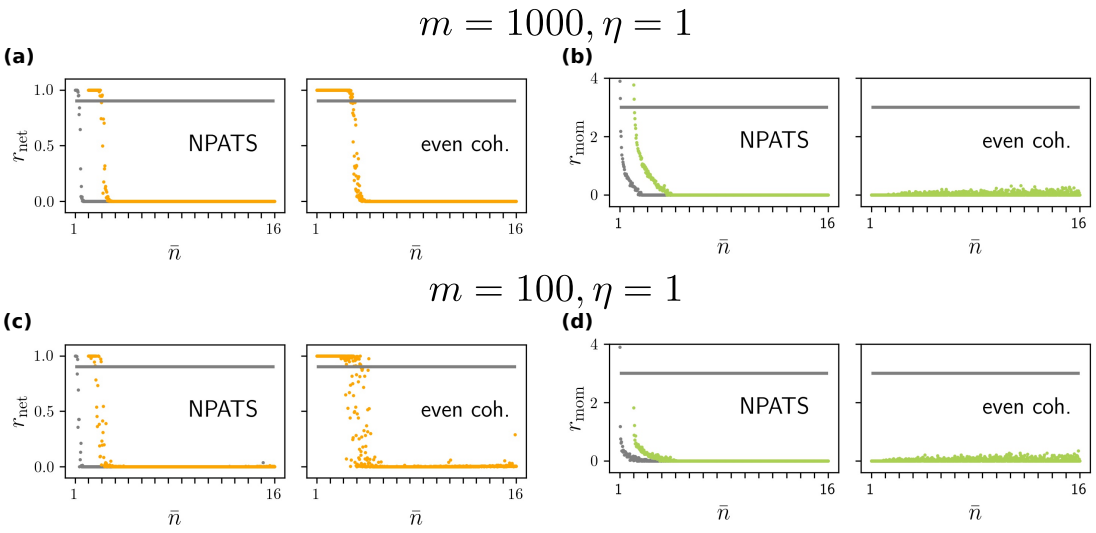

Figure 4.13: (a,c) Nonclassicality prediction r of the neural network (NN) for simulated multiplexed click-counting measurement data with perfect detectors, $\eta = 1$, generated from nonclassical n-photon-added thermal states (NPATS; $n = 1$ in gray and $n = 2$ in yellow) and even coherent states (both of which were not included in the training of the NN) for sample sizes $m = 1000$ (a) and $m = 100$ (c), depending on the mean number of photons $\bar{n}$. For each state family, we have simulated 1000 input vectors. (b,d) Nonclassicality detection r_{mom} of the matrix-of-moments method for the same measurement data as in (a) and (c). Gray horizontal lines indicate the (chosen) threshold value above which nonclassicality is predicted.

the smae mean photon number, $|\alpha|^2 = n$, given by

$$\rho = p\,|\alpha\rangle\langle\alpha| + (1-p)\,|n\rangle\langle n|\,, \tag{4.20}$$

where $p \in [0,1]$ parametrizes the weights of the mixture. In Fig. 4.14(a), we show the nonclassicality prediction r of the NN as a function of the parameter p for three different mean photon numbers, $n = 3$, 8, and 15. Here we only consider a sample size of $m = 1000$. The NN prediction monotonically decreases when the mixture is tuned from nonclassical Fock states ($p = 0$) to classical coherent states ($p = 1$). Furthermore, nonclassicality is recognized in a significant range $p < 0.4$, slightly dependent on the mean photon number. This quite balanced transition of the NN prediction shows that the NN is not biased towards a classification of classicality or nonclassicality after the training phase. In comparison, the matrix-of-moments nonclassicality prediction r_{mom} only significantly identifies nonclassicality of the mixed state of Eq. (4.20) for $p \approx 0$.

We also want to note that due to the similarity of the NPATS to thermal states and the similarity of even coherent state to mixtures of coherent states for large mean photon numbers $\bar{n}$, an inclusion of these nonclassical states in the training phase would lead to worse performances of the NN: the NN would essentially see the same input vector with two different labels (classical and nonclassical), such that the probability of a false nonclassicality classification of, e.g., thermal states increases.

To summarize, we have seen that the NN successfully learns to separate classical from

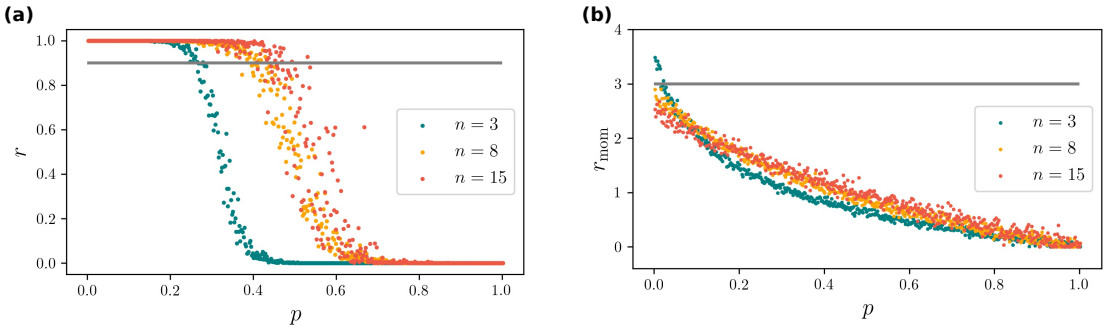

Figure 4.14: (a) Nonclassicality prediction r of the neural network (NN) for simulated multiplexed click-counting measurement data with perfect detectors, $\eta = 1$, generated from a mixture of a nonclassical Fock state $|n\rangle$ ($p = 0$; $n = 3, 8, 15$) and a classical coherent state $|\alpha\rangle$ ($p = 1$; $|\alpha|^2 = n$), cf. Eq. (4.20), depending on the mixture weight p. We use a sample size of $m = 1000$. (b) Nonclassicality detection r_{mom} of the matrix-of-moments method for the same measurement data as in (a).

nonclassical features for several typical quantum-optical states that are measured with multiplexing strategies using perfect detectors, even for significantly small sample sizes. In this regime, the NN nonclassicality prediction clearly outperforms the method-of-moments nonclassicality condition. Furthermore, the NN also achieves a good performance for measurement data from different (mixtures of) states that have not been included in the training phase. Similar to what was seen in Sec. 4.3, these results highlight the potential of NNs to be used as a direct, fast, and easy-to-implement nonclassicality indicator in multiplexed click-counting measurements, which can be adventurous in real-time monitoring or (online) presorting or pre-selecting measurement data. In this context, we note again that for a proper certification of nonclassicality, a proven nonclassicality condition (such as the matrix-of-moments condition) has to be violated with an appropriate analysis of statistical errors.

NNs performance with realistic detectors

Finally, we train and test the NN with measurement data from multiplexed click-counting measurements that make use of realistic detection efficiencies $\eta = 0.6$. As in Sec. 4.3, we consider a detection efficiency $\eta = 0.6$ which can be modeled as a beam splitter with a transmission coefficient $\sqrt{\eta}$ in each outgoing detection mode. The inclusion of η in the detection probabilities p_k is shown in Eq. (4.16), for more details we refer to Appendix B. We first note that due the finite detection efficiency η, the NNs trained in the previous section do not result in a good classification of measurement data with finite detection efficiencies. This is different from the good generalization of the NN to data generated from different sample sizes and highlights that, if the NN approach is to be applied in an experiment, the experimental details should be taken into account. Therefore, in the following, we show the performance of NNs that were trained with data from finite detection efficiencies. The

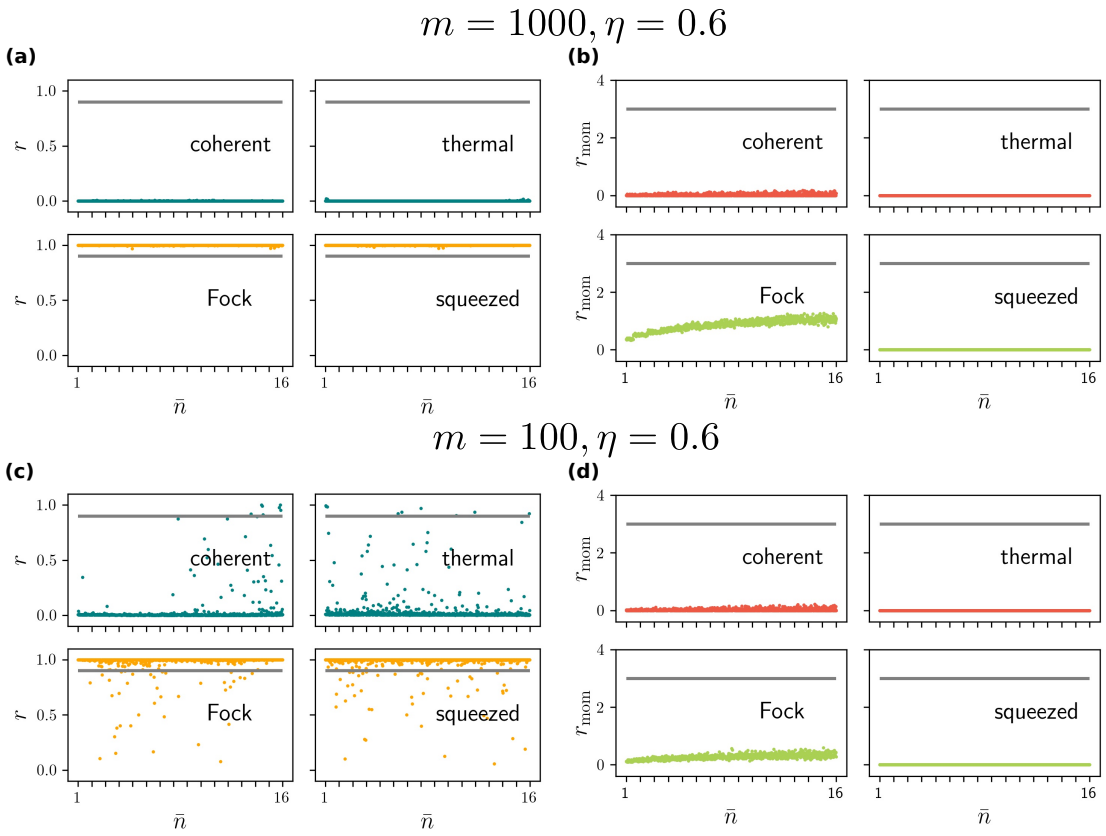

Figure 4.15: (a,c) Nonclassicality prediction r of the neural network (NN) for simulated multiplexed click-counting measurement data with realistic detectors, $\eta = 0.6$, generated from the training states (coherent and thermal states as classical and squeezed and Fock states as nonclassical) for sample sizes $m = 1000$ (a) and $m = 100$ (c), depending on the mean number of photons $\bar{n}$. For the simulation of Fock states, we used a photon number $n = \lfloor \bar{n} \rfloor$. For each state family, we have simulated 1000 input vectors. (b,d) Nonclassicality detection r_{mom} of the matrix-of-moments method for the same measurement data as in (a) and (c). Gray horizontal lines indicate the (chosen) threshold value above which nonclassicality is predicted.

training settings and NN architecture is the same as in the previous section. The training of a single NN with data from different detection efficiencies, such that, eventually, data from several detection efficiencies are correctly classified, is an interesting direction for future investigations.

The performance of the NN trained with a realistic detection efficiency of $\eta = 0.6$ on the different training states is shown in Fig. 4.15 for sample sizes of $m = 1000$ (a) and $m = 100$ (b). We see that in both cases, the NN correctly learns to classify the different training states. However, while the prediction of the NN for $m = 1000$ is solid, for $m = 100$, the prediction is largely fluctuating. While it is remarkable that the NN prediction still works in the regime of noisy detections and very small sample sizes, we clearly see the limits of the NN approach: when further decreasing the sample size m, the nonclassicality prediction of the NN is not reliable anymore because there is a finite probability of false positives. The conditions of Fig. 4.15(c) are just at the limit of the applicability of the NN nonclassicality identification. Note that the false-positive rate could not be removed

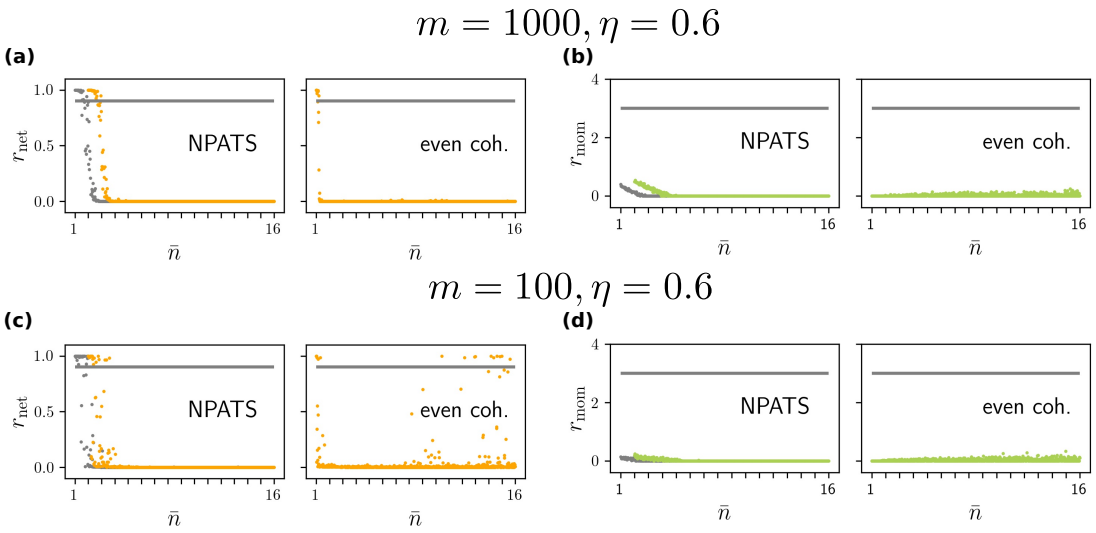

Figure 4.16: (a,c) Nonclassicality prediction r of the neural network (NN) for simulated multiplexed click-counting measurement data with realistic detectors, $\eta = 0.6$, generated from nonclassical n-photon-added thermal states (NPATS; $n = 1$ in gray and $n = 2$ in yellow) and even coherent states (both of which were not included in the training of the NN) for sample sizes $m = 1000$ (a) and $m = 100$ (c), depending on the mean number of photons $\bar{n}$. For each state family, we have simulated 1000 input vectors. (b,d) Nonclassicality detection r_{mom} of the matrix-of-moments method for the same measurement data as in (a) and (c). Gray horizontal lines indicate the (chosen) threshold value above which nonclassicality is predicted.

by simply increasing the threshold value t of the NN output. We also see that, for finite detection efficiencies, the NN nonclassicality identification requires a larger (but still small) minimal sample size m than for ideal detection efficiencies. In comparison, we again show the methods-of-moments nonclassicality prediction r_{mom} (cf. Eq. (4.19)) in Fig. 4.15(b,d) and see that the finite detection efficiency η results in no nonclassicality detection for all the training states.

Finally, in Fig. 4.16(a,c), we show the nonclassicality prediction of the NN with finite detection efficiency for simulated measurement data from NPATS and even coherent states, which both have not been used in the training phase. While the performance on NPATS is similar to the one for perfect detection efficiency (cf. Fig. 4.13), even coherent states are only recognized as nonclassical for $|\alpha| \approx 1$. This is because the interference pattern of the superposition $|\alpha_+\rangle$ is very fragile to noise for larger η, as we have already seen in Sec. 4.3.2 for odd cat states $|\alpha_-\rangle$ (cf. Fig. 4.9). We also see in Fig. 4.16(b,d) that for all considered states, neither does the matrix-of-moments nonclassicality condition detect nonclassicality.

4.5 Conclusions and outlook

In this chapter, we have employed artificial neural networks (NNs) to act as a nonclassicality indicator for different single-mode quantum optical measurement techniques. Both for

the measurement method of balanced homodyne detection (Sec. 4.3) and for multiplexed click-counting measurements (Sec. 4.4), the NNs were successfully trained by supervised learning to recognize the (non)classicality several types of classical and nonclassical states. These include coherent, thermal, and mixtures of coherent states as classical states, and Fock, squeezed, photon-added coherent states, photon-added thermal states, and cat states (even and odd coherent states) as nonclassical states. For homodyne measurements, the correct operation and the pratical usefulness of the NN was confirmed on experimental measurement data from coherent, squeezed and single-photon-added coherent states. In both measurement settings, the NNs were able to recognize nonclassicality from different features (e.g., sub-shot-noise variance, or similarity to Fock states), and could partially also recognize the nonclassicality of states that were not seen in the training phase. The main advantage of the NN approach to predict nonclassicality is the regime of small sample sizes: the NNs perform remarkably well even for very small sample sizes, such that they require much less measurements than conventional nonclassicality-certification methods such as, e.g., methods based on full quantum state tomography or the moments of the click distribution.

The NN nonclassicality prediction does not constitute a certification or verification of nonclassicality but can merely act as an indicator. For a proper certification, a nonclassicality test or witness has to be employed, together with an analysis of statistical errors. Furthermore, to maximize the performance and the applicability of the NN nonclassicality identification, the specific experimental conditions should be taken into account. These include knowledge about the different states that could possibly be produced in the experiment, or different noise factors such phase noise, amplitude noise, or finite detection efficiencies.

The high performance of the NN approach for small sample sizes and its easy implementation points to several possible applications: large amounts of experimental data can be quickly presorted, or the NN prediction can even be applied online during the measurement acquisition, e.g., to search for optimal experimental parameter settings. The advantages become significant especially when the measurement rate is low. Furthermore, in contrast to different advanced methods of nonclassicality certification, e.g., requiring sophisticated post-processing techniques such as quantum state tomography, the NN approach is easy to implement with openly-available software and does not require profound knowledge of the underlying physical theory.

Finally, we briefly want to mention possible future research directions that originate from our findings. First, to increase the NNs general performance and applicability also beyond training data, further classical and nonclassical states can be included in the training, as well as a fluctuating detection efficiency. Also, further machine learning methods such as convolutional NNs or regularization, or even other learning models such as reinforcement learning, can be examined to increase the generalizability of the NNs. Sec-

ond, the NN approach can be applied to multimode quantum optical measurements (both with homodyne and click-counting measurements), in order to train the NNs to identify cross-correlations such as entanglement between the incoming modes, cf. also Ref. [388]. Finally, one can consider other physical systems such as, e.g., cold atoms, that go beyond single-mode continuous-variable systems.

Chapter 5

Application: Bayesian quantum multiphase estimation algorithm

Contents

In this chapter, we address to the field of applications of quantum resources by describing a new quantum technology, the Bayesian quantum multiphase estimation algorithm. This algorithm enables one to measure multiple phases simultaneously with a precision that shows Heisenberg scaling with respect to the total number of resources[1] used in the

[1] In this context, the term "resources" should not be confused with the term "quantum resources" that has been central to the previous chapters and that entitles this thesis. As we will further discuss in more detail below, estimation tasks are specified by counting the number of some resource, or "physical" resource, that is used in the estimation, in order to consistently compare different estimation strategies. Depending on the specific estimation task, these resources are commonly taken as, e.g., the total number of particles, the total number of applications of the phase shift, or the total number of applications of specific gates. In contrast, after having fixed the number of physical resources, one can additionally make use of quantum resources to achieve a better performance to what would be possible without quantum resources, i.e., with only "classical" resources. Thus, quantum resources can lead to a quantum metrological advantage that is quantified after

estimation procedure and provides a subroutine for quantum metrological and quantum computational tasks.

In Sec. 5.1, we will first give a brief overview over the state of the art and the basic notions of quantum single-phase estimation, quantum phase estimation algorithms, and quantum multiphase estimation. Then, in Sec. 5.2, we introduce the Bayesian quantum multiphase estimation algorithm, analyze its performance compared to existing multiphase estimation strategies and discuss different possible implementations.

5.1 Quantum phase estimation

The field of quantum phase estimation (QPE) aims to develop quantum technologies that enable one to achieve a measurement precision that is out of reach for measurements that make only use of classical resources. Here, as always when identifying any quantum advantage (cf. Sec. 1.2), the available resources for the measurement task are fixed such as, e.g., the number of measurements or the physical resources that are used in each measurement such as the (mean) number of particles or the number of applications of a phase shift, see below. In the general investigation of precision measurements, the measurement task is commonly associated with a phase estimation task because, eventually, all precision measurements of a continuous variable reduce to a phase estimation measurement of some kind. In both quantum and classical phase estimation measurements, the phase can never be directly measured but has to be inferred by means of some kind of interferometer: phases always represent relative quantities that are only defined with respect to some phase reference. The fact that there is no direct quantum mechanical phase measurement, is also fundamentally explained by the insight that there is no phase observable [419].

It has been known for a long time that a clever use of quantum resources can result in increased measurement precision with respect to classical measurement strategies [17, 420, 421, 422]. These pioneering proposals mostly focus on the use of squeezed states to reduce the amount of intrinsic quantum noise in the measurement output distribution. In recent years, the metrological quantum advantage has been demonstrated in several physical systems such as, e.g., optics [423, 424] or cold atoms [141, 425]. In the following, we will sketch the main ideas of the theory of phase estimation.

5.1.1 Quantum single-phase estimation

The most common approach to study the precision in QPE is the use of the quantum Cramér–Rao bound [46, 47, 426, 427]. For an extensive review of the notions and detailed proofs of the methods that we describe here, see Ref. [134]. Let us consider an interferometer in which we have prepared a quantum state ρ_0 (i.e., ρ_0 is the quantum state inside

fixing the physical resources that are used in the protocol.

the interferometer directly before the phase imprinting). The phase shift θ that should be estimated is then imprinted by the unitary $e^{-iH\theta}$, where H is the Hamiltonian generating the phase shift. Then, the quantum Cramér–Rao bound bounds the frequentist variance $(\Delta\theta)^2$ of the estimated phase around the point θ with the quantum Fisher information (QFI) according to

$$(\Delta\theta)^2 \geq \frac{1}{m\,\mathrm{QFI}(\rho_\theta, H)}, \tag{5.1}$$

where we have defined $\rho_\theta = e^{-iH\theta}\rho_0 e^{iH\theta}$ and m is the number of repetitions of the measurement.

The central role of the QFI in phase estimation and why quantum resources can lead to an increased precision can be informally understood by the following observations. As we have discussed in detail in Sec. 2.1.1, the QFI is a quantum statistical speed, meaning that it is a measure of how fast a quantum state becomes distinguishable from its initial state when displaced by a phase. This phase difference can only be resolved if it has led to a significant distinguishability of the initial and the displaced state. Therefore, a larger QFI implies more distinguishability, which results in a smaller minimal phase difference that can be resolved. Furthermore, as we have also discussed in Sec. 2.1.1, the QFI is subject to different speed limits for different classes of states [27, 28, 29]. Roughly speaking, for separable states of M qubits, the QFI is bounded as $\mathrm{QFI} \leq M$, leading to the well-known shot-noise limit $(\Delta\theta)^2 \geq 1/M$ [36, 37]. On the other side, entangled states of M qubits can exhibit values up to the maximum of $\mathrm{QFI} = M^2$, leading to the optimal Heisenberg limit $(\Delta\theta)^2 \geq 1/M^2$ [27, 28, 29, 37]. The most prominent example of an entangled state of M qubits that leads to an optimal sensitivity is the N00N state[2] $|\psi_{\mathrm{N00N}}\rangle = (|0\rangle^{\otimes M} + |1\rangle^{\otimes M})/\sqrt{2}$.

We want to comment on a few important caveats of the quantum Cramér–Rao bound and the QFI-approach to phase estimation. First, the bound of Eq. (5.1) is an asymptotic result: it generally holds only in the limit of large m, the number of repetitions of the measurement. The optimal precision is scaling as $(\Delta\theta)^2 \propto 1/(mM^2)$ but the total number of used resources (here the number of qubits) is actually $N_T = m \times M$ and not M. Thus, when taking into account the total number of resources that are used in the measurement, the quantum Cramér–Rao bound does not imply $(\Delta\theta)^2 \propto 1/(N_T^2)$. To prove that a measurement shows a precision scaling that scales at the Heisenberg-limit with respect to the total number of resources, methods beyond the quantum Cramér–Rao bound have to be

[2] In the literature, the N00N state can be often found as

$$|\psi_{\mathrm{N00N}}\rangle = \frac{1}{\sqrt{2}}(|M,0\rangle + |0,M\rangle),$$

which is just the above state written in second quantization: the state is a superposition of the state in which all particles are in the first mode (called 0 above) and the state in which all particles are in the second one (called 1 above). In photonic experiments, the two modes often correspond to two spatially separated modes (the two arms of the interferometer).

applied. More about this below.

Second, the quantum Cramér–Rao bound only characterizes the phase sensitivity of the state ρ_θ at the point θ. Therefore, an optimal scaling might only be possible for a specific value of θ (i.e., at an optimal point). Furthermore, states that show the highest precision, such as the N00N state from above, show a high local phase sensitivity but only provide a highly ambiguous (M-fold) phase estimate in the total phase domain. Thus, to really make use of the high precision of, e.g., N00N states, the prior knowledge of the phase must have been already reduced to a range with a size proportional to $1/M$ [428].

One resolution of this phase ambiguity is to perform several measurements using different N00N states with varying photon numbers M. The measurements with larger M provide a high local phase resolution while the measurements with small M resolve the phase ambiguity. A common method to combine these different measurements is the Bayesian approach to phase estimation [134], see also Ref. [386] for a comparison between Bayesian and frequentist phase estimation. Here, the central object is the Bayesian probability distribution $P(\vartheta)$ that encodes the belief that $\theta = \vartheta$, where θ is the unknown phase. Starting with a flat initial distribution (corresponding to the complete absence or ignorance of knowledge of the phase $\theta \in [0, 2\pi]$ before the first measurement), each measurement results in an updated distribution by means of Bayes' theorem, $P(\vartheta) \propto P(o|\vartheta)P_{\text{in}}(\vartheta)$, where $P_{\text{in}}(\vartheta)$ is the Bayesian distribution before the measurement, o is the measurement outcome, and $P(o|\vartheta)$ is the likelihood (i.e., the probability) of observing o when the phase is ϑ.

We show an exemplary case of the Bayesian approach that combines the information from several measurements with different M in Fig. 5.1. In Fig. 5.1(a), we plot the probability distribution [see also Eq. (5.4)] after a first measurement using $M = 4$. It is clear that the measurement has a good local precision (as can be seen in the fast oscillations of the Bayesian distribution), but by itself any estimator will be highly ambiguous. On the other hand, in Fig. 5.1(b), we show the Bayesian distribution after performing three measurements with different M. We see that the final Bayesian distribution (black) shows the same resolution as the single measurement with the largest M ($M = 4$), while the measurements with smaller M ($M = 1$ and $M = 2$) are used to resolve the phase ambiguity.

Finally, we want to recall an important observation that was already touched upon in Sec. 1.2.1. There are different ways to count the physical resources in phase estimation. The traditional resource counting assumes a fixed interferometer including a single phase shift, and counts how many particles (e.g., photons) have been used in the measurement protocol, i.e., one counts the total number of particles that have traversed the interferometer. On the other hand, in the QPE algorithm [40, 429], the counted resource is the number of controlled-U operations that are applied in the quantum circuits of the measurement, where the phase to be estimated is an eigenphase of U (see Sec. 5.1.2 for more details). Similar to the second way of counting resources, without fixing the exact shape of

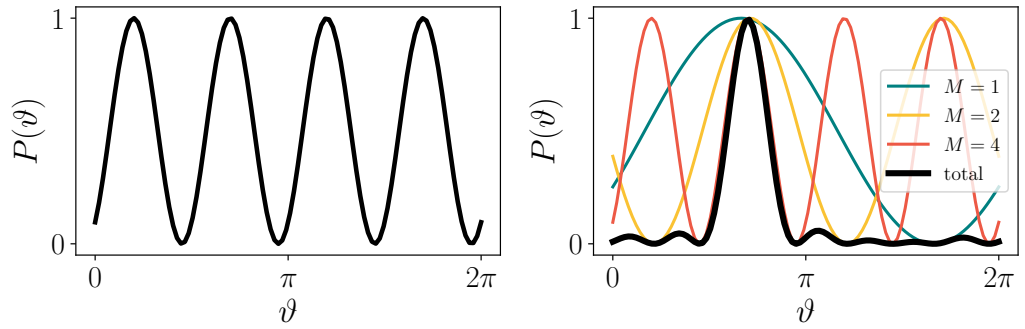

Figure 5.1: Different Bayesian probability distributions $P(\vartheta)$ highlighting the high local sensitivity and the phase ambiguity for the measurements of N00N state with large M (a), and showing a strategy to resolve the phase ambiguity (b). (a) The Bayesian distribution after a measurement of a N00N state with $M = 4$. (b) The final Bayesian distribution (black) after performing three measurements with N00N states consisting of $M = 1$ (blue), $M = 2$ (yellow), and $M = 4$ particles, respectively.

the interferometer, one can also count the number of times that the unknown phase shift is penetrated by a particle, which gives rise to a so-called multipass protocol, in which a single particle passes the phase shift multiple times.

All above ways of counting resources formally produce the same measurement probabilities (see Sec. 5.1.2) and thus represent equivalent footings for the estimation strategy that might reach the Heisenberg limit. However, they have different advantages and disadvantages. For instance, a multipass implementation requires that each measurement (and its phase imprinting) can take an arbitrary amount of time (that is increasing with the precision), while in an estimation protocol that makes use of N00N states, the phase imprinting can be applied in a constant, short time. On the other hand, N00N states of a large number of particles are harder to prepare in comparison to, say, single photon states, and N00N-state implementations also require photon-number resolving detectors. For a technical discussion of the formal equivalence of the above resources, see, e.g., the Appendix of Ref. [430]. Finally, as discussed in Sec. 2.1.1, we note that the different ways of counting physical resources eventually also result in different quantum features (or quantum resources) that are necessary to achieve the quantum advantage. We will briefly discuss the quantum resources in the different setting in Sec. 5.1.4.

5.1.2 Quantum phase estimation algorithm

Before we introduce the QPE algorithm, we first explain the formal equivalence of the basic building block of the QPE algorithm and an interferometer (here, we focus on Mach–Zehnder interferometers for simplicity). Imagine a single photon that enters the interferometer by passing through as beam splitter (BS), cf. Fig. 5.2(a), such that the photon is in a superposition of the two arms of the interferometer, $|\psi\rangle = (|0\rangle + |1\rangle)/\sqrt{2}$. Then, a control phase shift ϕ (i.e., a known phase ϕ) is applied to one of the arms, while the other is im-

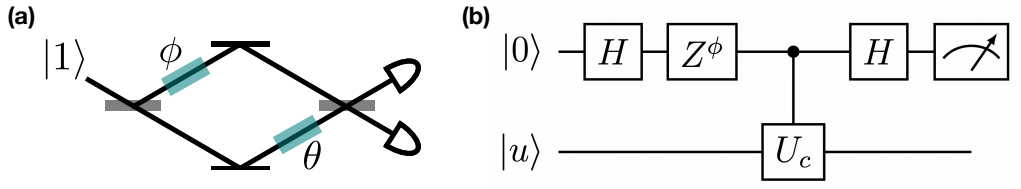

Figure 5.2: A Mach–Zehnder interferometer (a) and the basic circuit of the quantum phase estimation (QPE) algorithm (b) which result in formally equivalent phase estimation protocols. (a) A single photon ($|1\rangle$) enters the interferometer via a beam splitter. The control phase ϕ and the unknown phase θ are imprinted in the two arms of the interferometer. After passing a second beam splitter, the photon is measured. (b) The basic circuit of the QPE algorithm. The ancilla and register qubits are prepared in the state $|0\rangle \otimes |u\rangle$. Then, a Hadamard gate H and a phase gate Z^ϕ are applied to the ancilla qubit, after which the controlled unitary U_c is applied (imprinting the unknown (eigen)phase θ into the ancilla qubit's state). After the application of a second Hadamard gate to the ancilla, the ancilla qubit is measured in the computational basis.

printed with the unknown phase θ. After passing a second BS, the probability of measuring the photon in the arm o is given by

$$P(o|\theta,\phi) = \frac{1}{2}[\langle 0| + (-1)^o \langle 1|](e^{-i\phi}|0\rangle + e^{-i\theta}|1\rangle) = \frac{1}{2}[1 + (-1)^o \cos(\theta - \phi)], \quad (5.2)$$

where $o = 0$ and $o = 1$ represent a photon detection in the upper and lower arm, respectively. By repeating the measurement several times, the value of θ can be estimated by means of Eq. (5.2) and the observed statistics.

In QPE algorithms, one assumes that one can prepare the eigenstate $|u\rangle$ of a unitary U with a corresponding (unknown) eigenvalue $e^{i\theta}$ that encodes the phase that we want to measure. Furthermore, it is assumed that one can apply any number of controlled-U gates U_c [19, 40, 429, 431, 432]. The basic circuit of the QPE algorithm is shown in Fig. 5.2(b). The eigenstate $|u\rangle$ (also called the register state) is prepared together with an ancilla qubit in the state $|0\rangle$. Then, a Hadamard gate H [$\langle m|H|n\rangle = (-1)^{mn}/\sqrt{2}$] and a phase gate Z^ϕ ($\langle m|Z^\phi|n\rangle = \delta_{mn}e^{-i\phi_m}$ where we identify $\phi_1 = \phi$ and set $\phi_0 = 0$) are applied to the ancilla state, resulting in the state $(|0\rangle + e^{-i\phi}|1\rangle)/\sqrt{2} \otimes |u\rangle$. The central operation is the controlled-U gate U_c that applies U on the register state if the ancilla is in the state $|1\rangle$, $U_c = |0\rangle\langle 0| \otimes \mathbb{1} + |1\rangle\langle 1| \otimes U$. Given the intial register-ancilla state, the application of U_c effectively imprints the eigenphase θ into the ancilla state,

$$\begin{aligned}\frac{1}{\sqrt{2}}U_c(|0\rangle + e^{-i\phi}|1\rangle) \otimes |u\rangle &= \frac{1}{\sqrt{2}}\left[|0\rangle \otimes |u\rangle + e^{i(\theta-\phi)}|1\rangle \otimes |u\rangle\right] \\ &= \frac{1}{\sqrt{2}}\left[|0\rangle + e^{i(\theta-\phi)}|1\rangle\right] \otimes |u\rangle\,. \end{aligned} \quad (5.3)$$

Finally, when applying a second Hadamard gate to the ancilla and measuring the ancilla in the computational basis, one finds the same outcome probabilities as Eq. (5.2) and we thus see that both basic measurement steps are formally equivalent.

Furthermore, one directly sees that when applying the phase shift M times in the interferometer of Fig. 5.2(a), or when applying U_c M times in the quantum circuit of Fig. 5.2(b),

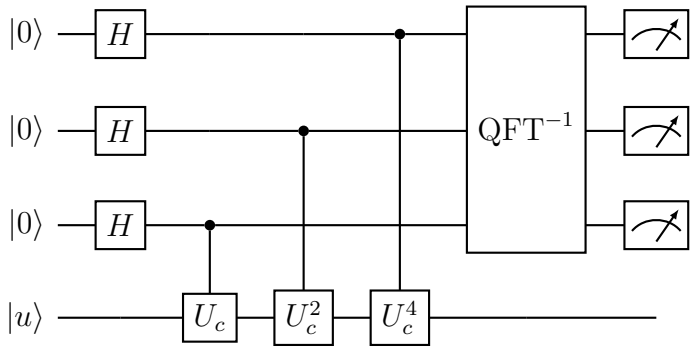

Figure 5.3: Quantum circuit of the QPE algorithm for the estimation of the phase θ up to three binary digits ($\theta = 2\pi \times 0.t_0t_1t_2$). Three ancilla qubits are prepared in the state $|0\rangle^{\otimes 3}$, while the register is prepared in the eigenstate $|u\rangle$. After each ancilla passes a Hadamard gate H, the kth ancilla is used as a control qubit for the controlled gate U_c^{K-1-k} acting on the register. Then, the ancilla qubits are acted upon by the inverse quantum Fourier transform QFT^{-1} and are finally measured in the computational basis.

the corresponding output probability is given by

$$P(o|\theta,\phi,M) = \frac{1}{2}\{1 + (-1)^o \cos[M(\theta-\phi)]\}. \tag{5.4}$$

Equivalently, one can apply the unknown phase shift θ only once (together with the control phase ϕ/M) to a N00N state with M photons, $|\psi_{\mathrm{N00N}}\rangle = (|M,0\rangle + |0,M\rangle)/\sqrt{2}$, to obtain the state $(e^{-i\phi}|M,0\rangle + e^{-iM\theta}|0,M\rangle)/\sqrt{2}$. When measuring this state by the projectors $|\psi_{\mathrm{N00N}}\rangle\langle\psi_{\mathrm{N00N}}|$ and $\mathbb{1} - |\psi_{\mathrm{N00N}}\rangle\langle\psi_{\mathrm{N00N}}|$, one finds the probability of Eq. (5.4) as well. As discussed above, we see that when considering the different resources, i.e., the number of applications of the phase gate, the number of applications of U_c, and the number of photons used in a single-pass interferometer, as equivalent, all different implementations result in the same performance for the estimation of θ.

After having understood that the basic measurement circuit of Fig. 5.2(b) resembles a simple interferometric measurement, we can now continue to see how different such circuits are combined in the QPE algorithm. As mentioned above, one of the problems of the highly sensitive N00N states with large M (or, equivalently, M applications of U_c or of the phase shift) is that the phase θ can only be resolved in an interval of size $2\pi/M$. In other words, for a completely unknown phase $\theta \in [0, 2\pi]$, the estimator is M-fold multivalued. In the QPE algorithm, this is circumvented by a combination of the basic circuit, cf. Fig. 5.2(b), with different values of M. In particular, it is a combination of K basic circuits, where, in the kth circuit ($k = 0, \ldots, K-1$), we set $M_k = 2^{K-1-k}$ to estimate the $(K-1-k)$th binary digit of $\theta/(2\pi)$, see below. The standard version of the QPE algorithm given in Ref. [40] is shown in Fig. 5.3 for $K = 3$ and explained in the following.

Let us assume for simplicity that $\theta = 2\pi \times 0.t_0t_1\ldots t_{K-1}$, i.e., that the binary representation of $\theta/(2\pi)$ consists of at most K digits. In the QPE algorithm, one then prepares K ancilla qubits in the state $|0\rangle^{\otimes K}$, acted upon by K Hadamard gates $H^{\otimes K}$. Then, the kth

ancilla qubit acts as the control qubit for an application of $U_c^{2^{K-1-k}}$ on the register state $|u\rangle$, see Fig. 5.3. Similar to Eq. (5.3), this results in phase shifts on the ancilla states,

$$\frac{1}{\sqrt{2^K}}\bigotimes_{k=0}^{K-1}(|0\rangle+|1\rangle)\otimes|u\rangle \mapsto \frac{1}{\sqrt{2^K}}\bigotimes_{k=0}^{K-1}(|0\rangle+e^{i\theta 2^{K-1-k}}|1\rangle)\otimes|u\rangle \tag{5.5}$$

$$=\frac{1}{\sqrt{2^K}}\sum_{x=0}^{2^K-1}e^{i\theta x}|x\rangle\otimes|u\rangle\,, \tag{5.6}$$

where, in the second line, we have written $x = x_0 \dots x_{K-1}$ to label the combined state of the ancilla qubits, and we have expanded the tensor product to simplify $\Pi_k e^{i\theta 2^{K-1-k}x_k} = e^{i\theta\sum_k 2^{K-1-k}x_k} = e^{i\theta x}$. Next, we apply the inverse quantum Fourier transform QFT^{-1} to the ancilla qubits that is defined as $\mathrm{QFT}^{-1}|x\rangle = 1/\sqrt{2^K}\sum_y e^{-ixy2\pi/2^K}|y\rangle$ [19]. This results in a state of the ancilla qubits given by

$$\frac{1}{2^K}\sum_y\sum_x e^{2\pi i x/2^K(2^K\theta/2\pi-y)}|y\rangle\,. \tag{5.7}$$

Now we use that if $n \in \mathbb{Z}\setminus\{0\}$, it holds that $\sum_{x=0}^{2^K-1}e^{2\pi i x n/2^K} = 0$. Finally, we have that $2^K\theta/2\pi - y \in \mathbb{Z}$, and $2^K\theta/2\pi - y = 0$ if $y = t_0t_1\dots t_{K-1}$. This final step only holds true if θ has a binary representation with at most K digits, see below if this is not the case. When measuring the final state of the ancilla qubits, Eq. (5.7), in the computational basis, we measure the outcome $y = t_0t_1\dots t_{K-1}$ with unit probability and can thus directly read off the phase θ from the measurement results.

If θ does not admit a binary representation with K digits, the measurement result does not yield $y = t_0t_1\dots t_{K-1}$ with unit probability. However, the probability of a correct estimation of the first K (rounded) digits of θ is still large, $P > 4/\pi^2 \approx 0.4$ [40]. Furthermore, by increasing the number of qubits used in the algorithm as $\mathcal{O}(\ln 1/\epsilon)$, the probability of a correct estimation of the first K digits can be increased to $1-\epsilon$ [40].

We want to note that, while the QPE algorithm described above uses all different basic circuits in parallel, the initial Kitaev algorithm (that was developed to solve the Abelian stabilizer problem, of which phase estimation is just a special case) makes uses of the different circuits (starting from $M = 2^{K-1}$) sequentially, by properly adjusting the control phase for the next circuit [429]. This sequential adjustment of the control phases for the next circuit is related to a "semiclassical" implementation of the inverse quantum Fourier transform [433], an approach that is used in most applications of the QPE algorithm and we will refer to as the "iterative" or "adaptive" version.

Applications and implementations

The QPE algorithm represents a central subroutine for a wide class of quantum algorithms [19], such as, e.g., Shor's factoring algorithm[3] [16]. The iterative version of the QPE al-

[3] Here, we want to emphasize that while the QPE algorithm offers a quadratic quantum advantage, Shor's algorithm achieves an exponential one. This is because Shor's algorithm is a special case of the QPE algo-

gorithm was used in first proof-of-principle implementations of Shor's algorithm, which used two photons and feedforward control to factor the number 21 [434], and an ion trap quantum computer to factor the number 15 [435].

A variation of the QPE algorithm has been used to find the eigenvalues and eigenvectors of Hamiltonians [431], an approach that can be used in the computation of ground state energies of molecular Hamiltonians [432]. This algorithm was implemented on quantum computers using superconducting qubits to simulate the ground state energy of molecular hydrogen [436]. Furthermore, the QPE algorithm is used as a subroutine in a quantum generalization of the Metropolis algorithm [437]. Other quantum technologies that build upon the QPE algorithm are used for reference-frame sharing [438], clock (and atomic-clock) synchronization [439, 440], and frequency estimation [441, 442].

The most obvious application of the QPE algorithm is QPE [21, 30, 443]. Here, the ultimate goal is to reach the Heisenberg limit of the precision scaling for the estimation of a completely unknown phase $\theta \in [0, 2\pi]$, taking into account the total number of resources. Even though the error probability of the QPE algorithm can be made arbitrarily small by increasing the number of ancilla qubits, as we have mentioned above, the original QPE algorithm can only reach a shot-noise-limited precision scaling of the variance $(\Delta\theta)^2$. This is due to the fact that to compute $(\Delta\theta)^2$, the (squared) errors have to be taken into account[4]. Repeating each round (i.e., the basic circuit for each M) only once results in significant tails of probability distribution of the phase such that a Heisenberg-limited precision is not achievable [444].

This obstacle was resolved in Ref. [38] by repeating each round at least three times. Here, the iterative version of the QPE algorithm was implemented in the optical multipass version to experimentally observe a Heisenberg–limited precision scaling[5], using a Bayesian post-processing technique. Note that in Ref. [445], single repetitions of the rounds of the QPE algorithm were sufficient to reach the Heisenberg limit because it was conducted at an optimal point ($\theta = 0$, meaning the θ can be written exactly as a finite binary string, see the discussion above). Later, it was shown that by using more repetitions of each round (resulting in a larger constant overhead of the precision scaling with respect to the Heisenberg limit), the QPE protocol could be rendered nonadaptive [430].

---

rithm in which the gates U^{2^k} can be implemented efficiently, i.e., using only a polynomial number of basic gates in terms of k and in terms of the number of qubits, which is achieved by a method called modular exponentiation [19]. In contrast, in the general QPE algorithm, the application of U^{2^k} requires 2^k applications of the gate U that cannot be circumvented.

[4]In order to calculate $(\Delta\theta)^2$, the magnitude of the error is crucial, and not just its probability. For instance, in QPE, an error in the estimation of the first digit of θ is much more costly than an error for the last digit.

[5]More precisely, the first six rounds of the algorithm were implemented (until a maximal $M = 2^5 = 32$), corresponding to a total number of resources of $N_T \approx 300$. In this range, the precision showed a Heisenberg scaling.

For an overview of the adaptive, nonadaptive and hybrid versions of the QPE algorithm, see Ref. [444].

A similar QPE protocol as in Ref. [38] that makes use of N00N states instead of a multipass single-photon protocol, was proposed in Ref. [428]. See also Ref. [446] for a similar proposal using the computational model DQC1 (cf. Sec. 2.1.1 and Ref. [106]). Recently, an adaptive forward version of the QPE algorithm has been proposed that makes use of Gaussian spin states [447]. Finally, a significant simplification of the QPE algorithm was proposed in Ref. [448]. Here, the basic circuit of Fig. 5.2(b) is used sequentially, with adaptively chosen values of M. Furthermore, the Bayesian probability distribution is modeled by a Gaussian which renders the computational post-processing much more efficient. However, as we further discuss below, this simplification results in a significant probability of an erroneous estimation.

Quantum-enhanced adaptive phase estimation has been experimentally demonstrated with squeezed states [449], single and few-photon states [38, 430, 450], integrated photonic devices [451, 452] and NV centers [453].

5.1.3 Quantum multiphase estimation

We now give a brief overview of the field of quantum multiphase estimation (QME) that recently has been experiencing an increased interest. Here, most of the literature employs a QFI approach (cf. Sec. 5.1.1 for the QFI approach to single-phase estimation). In QME, the central object is the QFI matrix $\mathbf{QFI}$ that generalizes the QFI that corresponds to the estimation of only one phase. Say that the initial quantum state is given by ρ_0, which is displaced by multiple phases according to $\rho = e^{-i\sum_{l=1}^{d} H_l\theta_l}\rho_0 e^{i\sum_{l=1}^{d} H_l\theta_l}$, where H_l is the Hamiltonian generation the lth phase shift θ_l. The elements of $\mathbf{QFI}$ are then given by [46]

$$\mathrm{QFI}_{ab}(\rho, \{H_l\}_l) = 2\sum_{k,k'} \frac{(\lambda_k - \lambda_{k'})^2}{\lambda_k + \lambda_{k'}} \langle k| H_a |k'\rangle \langle k'| H_b |k\rangle , \tag{5.8}$$

where $|k\rangle$ are the eigenvectors of ρ with corresponding eigenvalues λ_k and the sum is taken over terms for which $\lambda_k + \lambda_{k'} > 0$. In multiphase case, the quantum Cramér–Rao bound, Eq. (5.1), is replaced by the matrix inequality

$$\mathbf{\Sigma} \geq \frac{1}{m}\mathbf{QFI}^{-1}(\rho, \{H_l\}_l), \tag{5.9}$$

where $\mathbf{\Sigma}$ is the (frequentist) covariance matrix of the estimated phases that is defined as $\Sigma_{ab} = \mathrm{Cov}(\theta_a, \theta_b) = \langle\theta_a\theta_b\rangle - \langle\theta_a\rangle\langle\theta_b\rangle$. The inequality of Eq. (5.9) means that the matrix $\mathbf{\Sigma} - \mathbf{QFI}^{-1}(\rho, \{H_l\}_l)/m$ is positive semidefinite, or, in other words, that one has

$$\mathbf{n}^T\mathbf{\Sigma}\mathbf{n} \geq \frac{\mathbf{n}^T\mathbf{QFI}^{-1}(\rho, \{H_l\}_l)\mathbf{n}}{m} \tag{5.10}$$

for any $\mathbf{n} \in \mathbb{R}^d$, where $\mathbf{n}^T\mathbf{\Sigma}\mathbf{n} = \sum_{a,b=1}^{d} n_a n_b \mathrm{Cov}(\theta_a, \theta_b)$ and n_l are the components of $\mathbf{n}$. For instance, taking $n_l = 1$ for all l, Eq. (5.10) implies a bound on the trace of $\mathbf{\Sigma}$.

Based on the QFI matrix, Eq. (5.8), and the Cramér–Rao matrix bound, Eq. (5.9), the performance of different multiphase estimation protocols has been examined. In Ref. [454], generalized N00N states [see Eq. (5.39) for a definition below] have been employed to argue for an advantage between a simultaneous estimation of all d phases with respect to a sequential one in the case of commuting H_l: while generalized N00N states show a scaling of the trace of the QFI matrix as d^2/M^2 (d is the number of phases to estimate and M is the number of particles of the generalized N00N state), a sequential protocol that uses standard N00N states to estimate each phase separately shows a scaling of d^3/M^2. We note that, recently, it was argued that this parallel-over-sequential advantage is actually only constant (and not scaling with d) when taking into account the total number of resources, because the saturability of the bound Eq. (5.9), i.e., the value of m that is needed to reach the Cramér–Rao bound, depends on d [455].

If the different phase-displacement generating Hamiltonians H_l do not commute, there is a trade-off between the precision of the estimation of the different parameters [456, 457, 458]. This can be understood by the fact that if the H_l do not commute, the optimal measurement to estimate each phase is different, so they cannot be optimally measured simultaneously. In specific cases, multiple phases can be estimated simultaneously even if the H_l do not commute, if certain compatibility conditions are fulfilled [459].

As for the necessary quantum resources in QME in the traditional setting (i.e., when counting the number of particles and resources and allowing only for single-pass interferometers), the role of useful (particle) entanglement has been discussed in Ref. [460]. Furthermore, in Ref. [461], it was found that for QME in quantum networks, a parallel estimation protocol using entangled states only offers an advantage for specific estimation tasks, while separable states are optimal in other estimation problems. The role of nonclassical states that are not entangled for QME was further considered in Ref. [462]. These results highlight that a clear identification of appropriate figures of merit in QME estimation tasks, and how to count the used physical resources, are crucial to discuss about the quantum advantage in QME and its necessary quantum resources, see Refs. [463, 464] for a discussion of this point. In Ref. [463], a systematic analysis of the quantum advantage offered by particle entanglement and/or mode entanglement has been conducted, showing that to achieve the optimal scaling of the QFI matrix, both particle and mode entanglement have to be present. Building on these results, a QME protocol using entangled Gaussian states has been developed in Ref. [465], and the concept of multiparameter squeezing parameters has been introduced in Ref. [466]. In Ref. [464], it was shown that in QME, there are different nonequivalent choices of a reference mode that lead to different figures of merit for the final precision scaling.

All above approaches to QME are based on the QFI matrix and the quantum Cramér–Rao bound. In contrast to single-phase QPE, the quantum Cramér–Rao bound for multiple phases is not saturable in general [445]. In Ref. [445], conditions are derived that imply

that the bound can be saturated. Furthermore, as we have already discussed in Sec. 5.1.1, any QFI-based approach suffers from the drawback that it only quantifies the precision scaling at a specific point, and that the quantum Cramér–Rao bound might only be saturable asymptotically, such that a Heisenberg-limited precision scaling with respect to the total amount of physical resources cannot be proven. Here, we want to note that by optimizing single-shot measurements, estimation strategies have been found that outperform the QFI-based strategies in the non-asymptotic regime of finite measurement data [467]. Furthermore, advanced post-processing strategies that are based on a time-series analysis have been introduced [468, 469]. Here, the same basic measurement circuit of the QPE algorithm, Fig. 5.2, is used but the system is prepared in a superposition of eigenstates of U corresponding to different eigenvalues, instead of a single such eigenstate. In this case, a combination of phases are imprinted in the ancilla qubits state. By this technique, a small number of eigenphases of U could be resolved [468], or a large number of eigenphases can be categorized into several bins that cover the full range of possible eigenvalues [469].

In Sec. 5.2, we want to examine the total amount of resources that are used in a QME protocol and thus focus on Bayesian QME protocols instead of a QFI matrix approach. We will discuss an QME algorithm that shows a Heisenberg-limited precision scaling with respect to the total number of physical resources, for the estimation of d completely unknown phases $\boldsymbol{\theta} \in [0, 2\pi]^d$.

Applications and implementations

Applications of QME have been suggested in several different quantum metrology subfields. First, we want to emphasize that while in a specific estimation task, a parallel (simultaneous) estimation might be advantageous with respect to a sequential estimation or not, there can be spatial or temporal constraints that only allow for a parallel estimation strategy. This is why QME strategies have been proposed in the field of quantum imaging [470], where the quantum Cramér–Rao bound has been discussed for different imaging tasks [471, 472, 473]. Furthermore, QME protocols have been considered for optimal sensing of the three spatial components of a magnetic field [456], for the estimation of field gradients [474, 475], for distributed sensing using sensor networks [461, 462, 476], and for the synchronization of atomic clocks [440]. Ideas of QME also promise potential applications in biological sensing [477] or in the analysis of chemical processes [478].

QME strategies that simultaneously estimate two phases have been implemented using integrated photonic chips to observe a quantum enhancement of the precision in both frequentist estimation tasks [479] and Bayesian approaches [480] of QME. Furthermore, the estimation of the average of the phase shifts experienced by four nodes of a quantum optical network has been performed using continuous-variable entangled states, resulting in a precision that is beyond any achievable precision with separable states [476]. A quan-

tum advantage for the estimation of a specific linear combination of phases at an optimal point has also been observed in distributed discrete-variable quantum sensing networks [481, 482].

5.1.4 Quantum resources in QPE

To place the current topic of QPE in the context of quantum resources that we treat in this thesis, we want to briefly elaborate on the quantum resources that are required in the different implementations of the QPE protocols. As we discussed before, in the traditional setting of using a single application of the phase shift in the interferometer and counting the number of particles that enter the interferometer as the physical resource, multipartite entanglement is a necessary quantum resource for the quantum metrological advantage [27, 28, 29, 37]. Similar insights have been derived in multiphase estimation strategies [460, 463].

In contrast, in a single-photon multipass implementation, where we count the number of applied phase shifts as the physical resource, the single-photon state is not multipartite entangled (because there is only one photon). Since a single photon is a nonclassical state though, there is mode entanglement between the different modes, and one might conclude that entanglement is still necessary for the quantum advantage [39]. However, it is easy to see that by merely using a coherent state as one input of the interferometer, the phase sensitivity shows the same optimal scaling with the number of applied phase shifts. Thus, one could say that, in a multipass protocol, coherence is the crucial quantum resource [483]. On the other hand, as we have discussed in detail in Ch. 4, in quantum optics, coherent states are usually considered as "classical" resources. From this point of view, one could argue that a multipass implementation can scale at the optimal scaling limit by only using classical resources, which would imply that there is no quantum advantage (in the multipass setting). This dilemma eventually boils down to the question of whether the interferometric coherence of light is a quantum or a classical phenomenon.

A different viewpoint is the following. In the multipass implementation with M applications of the phase shift, the $(M \times \theta)$-dependence of the final probabilities stems from the M-fold application of the phase shift, independent of the type of input state used (e.g., single photons or coherent states). In contrast, in the traditional setting, the dependence stems from the M-partite entanglement between the incoming particles. As we mentioned earlier, the M-fold application of the phase shift requires that the experiments can take an increasing amount of time for larger M. Thus, one could say that the quantum resource of entanglement is replaced by the resource of time [37].

Finally, in the quantum-circuit implementation where we count the number of controlled-U gates as the physical resource, we again have to use an increasing amount of

time for each measurement when increasing M[6]. Coherence is also necessary here, both in the ancilla state and in the register eigenstates, such that the phases are correctly imprinted in the ancilla state. Furthermore, the quantum state during the different steps of the algorithm is separable. However, we note that controlled-U gates usually create entanglement (if we do not prepare an eigenstate of U), and specific realizations of the controlled unitary gate, in terms of basic gates from a universal set, can also produce intermediate entangled states during the execution of the gate.

5.2 Bayesian quantum multiphase estimation algorithm

In this section, we introduce the Bayesian quantum multiphase estimation algorithm (BQMEA) for the simultaneous (parallel) estimation of $d \geq 1$ (eigen)phases $\boldsymbol{\theta}$. In Sec. 5.2.1, we describe the basic quantum circuit and the corresponding Bayesian post-processing techniques that are used in the BQMEA. Then, in Sec. 5.2.2, we analyze the precision scaling of the algorithm and its error rate, compare its performance to different sequential multiphase estimation protocols, and discuss a generalization of the BQMEA in the presence of noise. Finally, in Sec. 5.2.3, we propose different implementations of the BQMEA in optical experiments. The results of this section are published in Ref. [5].

5.2.1 Estimation protocol of the BQMEA

The idea of the BQMEA is to generalize the key concepts of the single-phase QPE algorithm to an estimation protocol to estimate $d \geq 1$ phases. In particular, the aim is to estimate d independent and completely unknown phases $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_d) \in [0, 2\pi]^d$ with optimal precision scaling with respect to the total number of used resources N_T, where N_T is quantified as the total number of controlled-U applications. In Sec. 5.2.3, we will discuss optical implementations for which N_T corresponds to the total number of particles used in a single-pass implementation, or the total number of applied phase shifts used in a multipass implementation.

As we have discussed above, the standard assumptions in QPE algorithms is that one is able to prepare an eigenstate $|u\rangle$ of a unitary U with (unknown) eigenphase $e^{i\theta}$, and that one is able to apply controlled-U^M gates with $M \in \mathbb{N}_{>0}$ [19, 40, 429, 431, 432]. Similarly, we assume here we can prepare the eigenstates $|u_l\rangle$ of a unitary U corresponding to the (unknown) eigenvalues $e^{i\theta_l}$, $l \in \{1, \ldots, d\}$. Furthermore, we can apply a generalized controlled-U^M operation ($M \in \mathbb{N}_{>0}$) that we specify below. In the following, we first

[6]Note that, as we discussed above, the exponential advantage of Shor's algorithm is exactly due to the fact that for the specific unitary of Shor's algorithm, the time (or the number of gates we need) to implement M controlled unitary gates only increases polynomially in $\ln M$.

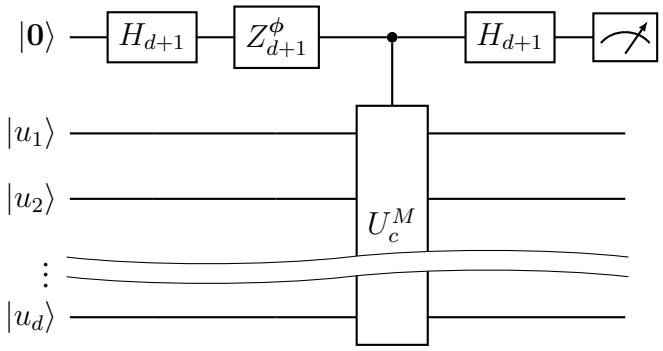

Figure 5.4: The basic quantum circuit of the Bayesian quantum multiphase estimation algorithm (BQMEA). An ancilla qudit is prepared in the state $|\mathbf{0}\rangle$, while the register is prepared in the state $\bigotimes_{l=1}^{d}|u_l\rangle$. The ancilla is acted then upon by a $(d+1)$-dimensional Hadamard gate H_{d+1} [Eq. (5.11)] and a $(d+1)$-dimensional phase gate $Z^{\boldsymbol{\phi}}_{d+1}$ [Eq. (5.12)] that imprints the control phases $\boldsymbol{\phi}$. Then, we apply the generalized controlled-U gate U_c [Eq. (5.14)] M times, imprinting the unknown phases $\boldsymbol{\theta}$ onto the ancilla state. After a second application of H_{d+1}, the ancilla qudit is measured in the computational basis.

discuss the basic quantum circuit of the BQMEA and then describe the corresponding Bayesian post-processing technique.

Basic quantum circuit of the BQMEA

The basic quantum circuit of the BQMEA, shown in Fig. 5.4, is inspired by the basic circuit of the single-phase QPE algorithm, cf. Fig. 5.2(b). Instead of a register consisting of a single state prepared in the eigenstate $|u\rangle$, the register of the BQMEA consists of d eigenstates of U, $\bigotimes_{l=1}^{d}|u_l\rangle$. Furthermore, the ancilla qubit is replaced by a $(d+1)$-dimensional ancilla qudit prepared in the state $|\mathbf{0}\rangle$ (we label the states of the ancilla qudits with bold letters for better visualization). Note that in a practical implementation, the ancilla qudit might be composed by several single qubits.

After the preparation, the ancilla qubit is acted upon by a generalized $(d+1)$-dimensional Hadamard gate H_{d+1} defined as

$$H_{d+1}=\frac{1}{\sqrt{d+1}}\begin{bmatrix}1 & 1 & 1 & \cdots & 1\\ 1 & \omega_d & \omega_d^d & \cdots & \omega_d^d\\ 1 & \omega_d^2 & \omega_d^4 & \cdots & \omega_d^{2d}\\ \vdots & \vdots & \vdots & \ddots & \vdots\\ 1 & \omega_d^d & \omega_d^{2d} & \cdots & \omega_d^{d^2},\end{bmatrix} \tag{5.11}$$

where $\omega_d = e^{i2\pi/(d+1)}$ is the $(d+1)$th root of unity. After the application of H_{d+1}, the ancilla state $|\psi_{\mathrm{A}}\rangle$ is given as $|\psi_{\mathrm{A}}\rangle = (|\mathbf{0}\rangle+\cdots+|\mathbf{d}\rangle)/\sqrt{d+1}$. We note that for $d=2^n-1$ ($n\in\mathbf{N}_{>0}$), H_{d+1} corresponds to the quantum Fourier transform QFT acting on n qubits[7] [19].

[7] Actually, in the n-qubit implementation, we could replace the initial application of QFT by an operation of n single-qubit Hadamard gates, $H^{\otimes n}$, since both gates transorm the initial ancilla state to

Next, we apply the generalized phase gate $Z_{d+1}^{\boldsymbol{\phi}}$ given by

$$Z_{d+1}^{\boldsymbol{\phi}} = \begin{bmatrix} 1 & & & \\ & e^{-i\phi_1} & & \\ & & \ddots & \\ & & & e^{-i\phi_d} \end{bmatrix} \tag{5.12}$$

onto the ancilla qudit to imprint the control phases $\boldsymbol{\phi} = (\phi_1, \ldots, \phi_d)$,

$$|\psi_{\mathrm{A}}\rangle = \frac{1}{\sqrt{d+1}} \sum_{n=0}^{d} e^{-i\phi_n} |\mathbf{n}\rangle , \tag{5.13}$$

where we have set $\phi_0 = 0$.

We now apply the generalized controlled-U^M gate U_c^M, where U_c is given by

$$U_c = |\mathbf{0}\rangle\langle\mathbf{0}| \otimes \mathbb{1} \otimes \cdots \otimes \mathbb{1} + |\mathbf{1}\rangle\langle\mathbf{1}| \otimes U \otimes \mathbb{1} \otimes \cdots \otimes \mathbb{1} + \ldots + |\mathbf{d}\rangle\langle\mathbf{d}| \otimes \mathbb{1} \otimes \cdots \otimes \mathbb{1} \otimes U. \tag{5.14}$$

If the ancilla is prepared in the state $|\mathbf{0}\rangle$, the gate U_c acts as the identity on the register states. If the ancilla qudit is instead prepared in the state $|\mathbf{l}\rangle$ for $l > 1$, U_c applies U to the lth register state that is prepared in $|u_l\rangle$, imprinting the eigenphase $e^{i\theta_l}$ as a relative phase onto the ancilla state, similar to Eq. 5.3. Given the prepared ancilla state Eq. (5.13), the application of U_c^M thus effectively imprints the different eigenphases $e^{iM\theta_l}$ onto the ancilla state without altering the states of the register,

$$U_c \sum_{n=0}^{d} e^{i\phi_n} |\mathbf{n}\rangle \otimes \bigotimes_{l=1}^{d} |u_l\rangle = \sum_{n=0}^{d} e^{i(M\theta_n - \phi_n)} |\mathbf{n}\rangle \otimes \bigotimes_{l=1}^{d} |u_l\rangle , \tag{5.15}$$

where we have set $\theta_0 = 0$.

Finally, we apply a second generalized Hadamard gate H_{d+1}, such that the final state of the ancilla qudit is given by

$$|\psi_{\mathrm{A}}\rangle = \frac{1}{d+1} \sum_{m,n=0}^{d} e^{i[M\theta_n - \phi_n + 2\pi mn/(d+1)]} |\mathbf{m}\rangle . \tag{5.16}$$

When measuring the ancilla qudit in the computational basis, we find that the probability of observing the outcome o ($o \in \{0, \ldots, d\}$) is given by

$$\begin{aligned} P(o|\boldsymbol{\theta}, \boldsymbol{\phi}, M) &= \frac{1}{(d+1)^2} \left| \sum_{n=0}^{d} e^{i[\theta_n - \phi_n - 2\pi mn/(d+1)]} \right|^2 \\ &= \frac{1}{(d+1)^2} \left\{ d + 1 + 2 \sum_{n} \cos[M\theta_n + \beta_n(o)] + 2 \sum_{m<n} \cos[M(\theta_n - \theta_m) + \gamma_{nm}(o)] \right\} , \end{aligned} \tag{5.17}$$

where we have defined $\beta_n(o) = 2\pi on/(d+1) - \phi_n$ and $\gamma_{nm}(o) = 2\pi o(n-m)/(d+1) - \phi_n + \phi_m$.

$|\psi_{\mathrm{A}}\rangle = (|\mathbf{0}\rangle + \cdots + |\mathbf{d}\rangle)/\sqrt{d+1}$. However, the second QFT gate cannot be replaced by $H^{\otimes n}$ without changing the algorithm's final state and the corresponding outcome probabilities.

Bayesian post-processing of the BQMEA

The BQMEA consists of several rounds of repetitions of the basic quantum circuit that we have described in the previous paragraph and in Fig. 5.3. In the kth round, we use $M_k = 2^k$ ($k = 0, 1, \ldots,$) such that the each phase θ_l can be resolved in an interval of size $2\pi/2^k$, see Fig. 5.1(b), and the local sensitivity of the measurement increases each round. Similar to the single-phase QPE algorithm, a single measurement per round results in large tails of the Bayesian probability distribution such that a Heisenberg-limited precision cannot be reached. In the following, we describe how to adaptively decide when to proceed to the next round.

Since our algorithm should be able to estimate arbitrary phases $\boldsymbol{\theta} \in [0, 2\pi]^d$, we use a uniform prior Bayesian distribution, $P(\boldsymbol{\vartheta}) = 1/(2\pi)^d$ before the first measurement. Recall that $P(\boldsymbol{\vartheta})$ represents the belief that $\boldsymbol{\theta} = \boldsymbol{\vartheta}$[8]. Then, after we have performed a measurement using M applications of U_c and the control phases $\boldsymbol{\phi}$ (see below for how to choose the control phases), we observe the measurement outcome o and update the posterior Bayesian probability distribution according to Bayes' theorem,

$$P(\boldsymbol{\vartheta}|o, \boldsymbol{\phi}, M) = \frac{P(o|\boldsymbol{\vartheta}, \boldsymbol{\phi}, M)P(\boldsymbol{\vartheta})}{\int \mathrm{d}^d\boldsymbol{\vartheta} P(o|\boldsymbol{\vartheta}, \boldsymbol{\phi}, M)P(\boldsymbol{\vartheta})}, \tag{5.18}$$

where $P(\boldsymbol{\vartheta})$ is the prior Bayesian distribution before the measurement, and $P(o|\boldsymbol{\vartheta}, \boldsymbol{\phi}, M)$ is the probability (or likelihood) of the outcome o given by Eq. (5.17). The denominator corresponds to a renormalization of the posterior distribution. After each measurement, we

- update the Bayesian distribution $P(\boldsymbol{\vartheta})$ according to $P(\boldsymbol{\vartheta}) \leftarrow P(\boldsymbol{\vartheta}|o, \boldsymbol{\phi}, M)$,
- produce an estimate for each phase θ_l according to

$$\bar{\theta}_l = \arg\left[\int \mathrm{d}^d\boldsymbol{\vartheta} e^{i\vartheta_l} P(\boldsymbol{\vartheta})\right], \tag{5.19}$$

- and calculate the Bayesian probability

$$P_{\text{half}} = \int_C \mathrm{d}^d\boldsymbol{\vartheta} P(\boldsymbol{\vartheta}) \tag{5.20}$$

  corresponding to the (estimated) probability to find the true phase $\boldsymbol{\theta}$ in the hypercube $C = \bigtimes_{l=1}^{d} \left[\bar{\theta}_l - \pi/2^{k+1}, \bar{\theta}_l + \pi/2^{k+1}\right]$ centered around $\bar{\boldsymbol{\theta}}$, for a measurement in the kth round of the algorithm.

[8] Strictly speaking, $P(\boldsymbol{\vartheta})$ is a probability density, meaning that the (estimated) probability $P(\boldsymbol{\theta} \in \Omega)$ for finding $\boldsymbol{\theta}$ in some subset $\Omega \subset [0, 2\pi]^d$ is given by

$$P(\boldsymbol{\theta} \in \Omega) = \int_\Omega \mathrm{d}^d\boldsymbol{\vartheta} P(\boldsymbol{\vartheta}).$$

Input: $\{\epsilon, K\}$
$P(\boldsymbol{\vartheta}) = 1/(2\pi)^d$
for $k = 0, \ldots, K-1$:
  $M = 2^k$; $P_{\text{half}} = 0$
  while $P_{\text{half}} < 1-\epsilon$:
    $\boldsymbol{\phi} = \text{generate\_random}()$
    $o = \text{measurement}(M, \boldsymbol{\phi})$
    $P(\boldsymbol{\vartheta}) = \text{Bay\_update}(P(\boldsymbol{\vartheta}), o, M, \boldsymbol{\phi})$
    $P_{\text{half}} = \text{compute}(P(\boldsymbol{\vartheta}))$
  $P(\boldsymbol{\vartheta}) = \text{cut\_grid}(P(\boldsymbol{\vartheta}))$
  $P(\boldsymbol{\vartheta}) = \text{normalize}(P(\boldsymbol{\vartheta}))$
return: $P(\boldsymbol{\vartheta})$

Table 5.1: The pseudocode of the Bayesian quantum multiphase estimation algorithm (BQMEA). The algorithm consists of K rounds, where in the kth round, we repeatedly run the basic quantum circuit, cf. Fig. 5.4, with $M = 2^k$ and random control phases $\boldsymbol{\phi}$. After each measurement, the (initially flat) Bayesian distribution $P(\boldsymbol{\vartheta})$ is updated according to Bayes' theorem, and the measurements are repeated until $P_{\text{half}} > 1-\epsilon$, i.e., until the Bayesian distribution is localized enough to restrict the phase domain.

Next, if we find that $P_{\text{half}} \leq 1-\epsilon$ for some small ϵ (we call ϵ the "decision" parameter), we choose new control phases $\boldsymbol{\phi}$ and repeat the measurement with the same M_k. On the other hand, if $P_{\text{half}} > 1-\epsilon$, we restrict the possible estimation domain of $\boldsymbol{\theta}$ to the hypercube C[9] and renormalize the Bayesian distribution such that $\int_C \mathrm{d}^d\boldsymbol{\vartheta} P(\boldsymbol{\vartheta}) = 1$. We then proceed to the next $(k+1)$th round, i.e., we set $M = 2^{k+1}$. In this way, we restrict the phase domain of the estimation protocol whenever the Bayesian probability is sufficiently localized. The parameter ϵ is called the decision parameter because it dictates how fast we proceed with the next round of the algorithm: as we see below, smaller values of ϵ result in more measurements per round, such that the error probability is reduced but the overhead constant of the precision scaling with respect to the Heisenberg limit is enlarged. The pseudocode of the algorithm (using random control phases $\boldsymbol{\phi}$, see below) is shown in Tab. 5.1.

The restriction of the phase domain whenever the Bayesian distribution $P(\boldsymbol{\vartheta})$ is localized enough is important to limit to memory requirement of recording $\boldsymbol{\phi}$: for increasing k, the distribution becomes very narrow such that a high resolution of the variable $\boldsymbol{\vartheta}$ is required. In particular, one usually records the Bayesian distribution on a grid over the phase domain such that, if we were not restricting the phase domain during the algorithm, the number of grid points has to be chosen very large if a high precision is required. Note also that, for a given number of grid points in one dimension, the total number of grid

[9] Note that in the first phase restriction, we may have to use the periodicity of $P(\boldsymbol{\vartheta})$ to compute P_{half} and to define the Bayesian distribution on the domain C. For instance, if $\bar{\theta}_l = 0.9\pi$, the lth dimension of C covers the interval $[0.4\pi, 1.4\pi]$. For $\vartheta_l > \pi$, we then use that $P(\vartheta_1, \ldots, \vartheta_l + 2\pi, \ldots, \vartheta_d) = P(\vartheta_1, \ldots, \vartheta_l, \ldots, \vartheta_d)$.

points increases exponentially with d, the number of phases that we want to estimate. Furthermore, we have observed numerically that a small grid size used to record $P(\boldsymbol{\vartheta})$ in the according phase domain is not only advantageous with respect to memory requirements but also serves as a coarse-graining and thus a smoothening of $P(\boldsymbol{\vartheta})$ that results in better performances than larger grid sizes. In the following, we use a grid of size of 21^d for the recording of $P(\boldsymbol{\vartheta})$. On the other hand, the restriction of the phase domain after each round introduces an error rate $P_{\text{err}}^{(d)}(\epsilon)$ that the true phase $\boldsymbol{\theta}$ is outside of the restricted domain, $\boldsymbol{\theta} \notin C$. We will discuss and analyze $P_{\text{err}}^{(d)}(\epsilon)$ in detail in Sec. 5.2.2.

In this context, we want to mention alternative approaches to record the Bayesian distribution. In the single-phase estimation protocols of Refs. [38, 430, 444], it was used that in each update of the Bayesian distribution, Eq. (5.18), $P(\boldsymbol{\vartheta})$ is multiplied by terms of the form $e^{\pm i2^k\vartheta}$, such that all information of $P(\boldsymbol{\vartheta})$ can be elegantly stored in finitely many Fourier coefficients of $P(\boldsymbol{\vartheta})$. This recording is efficient for a "backward" estimation protocol, i.e., a protocol where K, the number of rounds, and the maximal number of measurements per round is fixed, and the measurements start with $M = 2^{K-1}$. This efficient description is not possible for our "forward" protocol (that starts with $M = 1$, and K does not have to be fixed in advance), and, furthermore, the update rule of the Fourier coefficients, see Eq. (5.4) in Ref. [444], becomes increasingly complex for increasing $d > 1$. A second and more recent approach to store the Bayesian distribution $P(\boldsymbol{\vartheta})$ is the use of resampling algorithms [480, 484, 485]. Here, the phase domain is represented by a fixed number of points $\boldsymbol{\vartheta}_l$ that move to areas of high Bayesian probability by means of Monte-Carlo sampling. Crucially, this approach is independent of d, the number of phases to estimate, but a proper error rate analysis of this recording method must be conducted. Applying resampling-type post-processing to the BQMEA is an interesting direction for future research. Finally, the Bayesian distribution $P(\boldsymbol{\vartheta})$ can be simply approximated by a (multivariate) Gaussian distribution [448, 468]. As we discuss further in Sec. 5.2.2, while being very efficient in terms of memory space, this method introduces significant error probabilities of biased estimation.

Choice of the control phases

The control phases $\boldsymbol{\phi}$ that we have used in the BQMEA as described above are chosen randomly in $\boldsymbol{\phi} \in [0, 2\pi]^d$. Varying the control phases between the single measurements is crucial to avoid any proliferation of significant tails in the Bayesian distribution $P(\boldsymbol{\vartheta})$. This effect is shown in Fig. 5.5 for the BQMEA for $d = 1$ (i.e., for the estimation of a single phase). If we use a constant control phase, say $\phi = 0$, Eq. (5.17) shows that the final Bayesian distribution after the first round of measurements will be a product of powers of $\cos^2(\vartheta/2)$ and $\sin^2(\vartheta/2)$, depending on the measurement outcomes. This leads to a double-peak structure of $P(\vartheta)$ that cannot be resolved by later measurements with higher

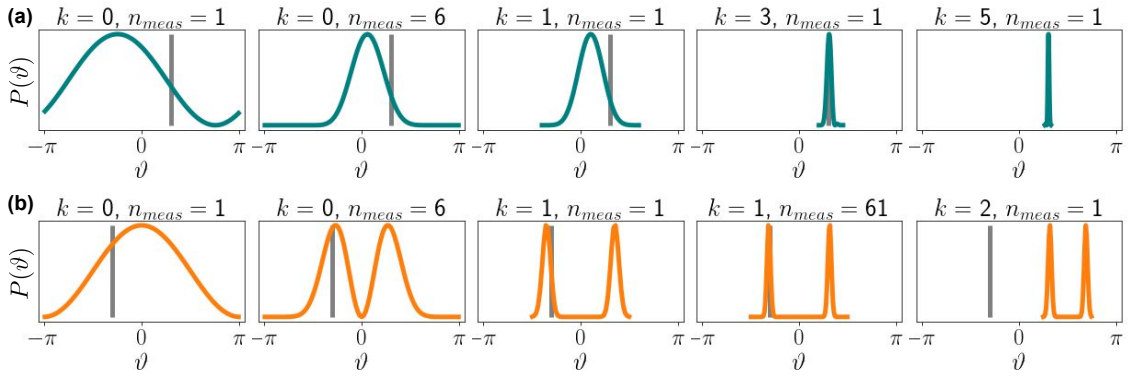

Figure 5.5: Different snapshots of the Bayesian distribution $P(\vartheta)$ during the Bayesian quantum multiphase estimation algorithm (BQMEA) after the n_{meas}th measurement of the kth round, in the case $d = 1$ and using $\epsilon = 10^{-4}$, for (a) randomly chosen control phases $\phi \in [0, 2\pi]$, and (b) a constant control phase $\phi = 0$. The true phase θ is indicated by a vertical line. The figure is taken from Ref. [5].

M, resulting in a large probability of an erroneous estimation. This effect can be seen in Fig. 5.5(b) where we show different snapshots of $P(\vartheta)$ during the BQMEA after the n_{meas}th measurement of the kth round for $d = 1$ and $\phi = 0$. In contrast, in Fig. 5.5(a), we show different snapshots of $P(\vartheta)$ during the BQMEA with randomly chosen control phases, where no double-peak structure evolves.

In single-phase QPE algorithms, analytical strategies to optimally and adaptively choose the control phase for each measurement have been developed [38, 444]. However, these approaches cannot straightforwardly be extended to an estimation algorithm for $d \geq 2$ phases, and advanced calculations of optimal control phases would also represent a computational slowdown during the algorithm. How to efficiently choose optimal control phases for the estimation of $d \geq 2$ phases is an interesting open question to improve the performance of the BQMEA. We note that, for $d = 1$, an alternation of the measurement settings between $\phi_1 = 0$ and $\phi_2 = \pi/2$ also prevents the proliferation of large tails in $P(\vartheta)$ [430].

In the case $d = 2$, we have tested an alternative of random control phases by setting the control phases $\boldsymbol{\phi}$ in the nth measurement of each round as

$$\boldsymbol{\phi} = \begin{cases} (0,0) & \text{for} \quad n = 0 \mod 4 \\ (\pi/2,0) & \text{for} \quad n = 1 \mod 4 \\ (0,\pi/2) & \text{for} \quad n = 2 \mod 4 \\ (\pi/2,\pi/2) & \text{for} \quad n = 3 \mod 4 \end{cases} \tag{5.21}$$

This choice of control phases leads to a similar performance of the BQMEA as using random control phases. For instance, for $\epsilon = 10^{-3}$, the control phases of Eq. (5.21) result in a variance scaling of each phase as $(\Delta\theta_l)^2 \approx 113/N_T^2$ with an error rate of $P_{\text{err}}^{(2)} \approx 7.1 \times 10^{-4}$, while for random control phases, we find $(\Delta\theta_l)^2 \approx 105/N_T^2$ and $P_{\text{err}}^{(2)} \approx 7.7 \times 10^{-4}$, see Sec. 5.2.2.

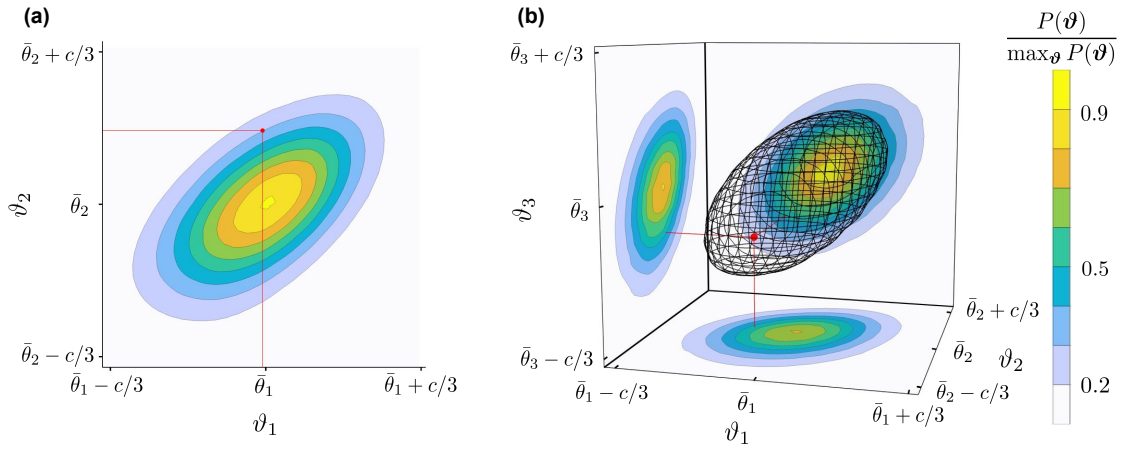

Figure 5.6: Exemplary snapshots of the Bayesian distribution $P(\boldsymbol{\vartheta})$ after completing five rounds of the Bayesian quantum multiphase estimation algorithm (BQMEA) for estimating (a) $d = 2$ and (b) $d = 3$ phases, where we use $\epsilon = 10^{-4}$. The red point indicates the true phases $\boldsymbol{\theta}$, and the phase domains are shown in a fraction of the hypercube $C = \bigtimes_{l=1}^{d} \left[\bar{\theta}_l - c, \bar{\theta}_l + c\right]$, where $c = \pi/2^4$ and $\bar{\theta}_l$ is the estimated value of θ_l. (a) The color scale shows $P(\boldsymbol{\vartheta})$ normalized to the maximal value $\max_{\boldsymbol{\vartheta}} P(\boldsymbol{\vartheta})$. (b) The three color-scale plots show each marginal Bayesian distribution, $\int \mathrm{d}\theta_l P(\boldsymbol{\vartheta})$, normalized to their respective maximal value. An equiprobable surface is shown as the ellipsoid. The figure is taken from Ref. [5].

A second alternative for $d \geq 2$ is the use of quasirandom control phases. Quasirandom numbers in dimension d are sequences of tuples $\mathbf{x}_l \in [0,1]^d$ such that for each L, the set $\{\mathbf{x}_1, \ldots, \mathbf{x}_L\}$ optimally covers the interval $[0,1]^d$, i.e., the set covers $[0,1]^d$ more uniformly than L random numbers on average for finite L, and cover $[0,1]^d$ densely in the limit $L \to \infty$ [486]. Using, e.g., the Halton sequence in the BQMEA for $d = 2$ and $\epsilon = 10^{-3}$, we find $(\Delta\theta_l)^2 \approx 101/N_T^2$ and $P_{\mathrm{err}}^{(2)} \approx 8.5 \times 10^{-4}$, showing again a similar performance to random control phases.

Running the BQMEA

In Fig. 5.6, we show two exemplary snapshots of the Bayesian distribution $P(\boldsymbol{\vartheta})$ of the BQMEA with $\epsilon = 10^{-4}$ for (a) $d = 2$ and (b) $d = 3$. Both snapshots show $P(\boldsymbol{\vartheta})$ after completing five rounds ($k = 4$) of the BQMEA. We see that, in both cases, $P(\boldsymbol{\vartheta})$ is localized in a subset of the hypercube $C = \bigtimes_{l=1}^{d} \left[\bar{\theta}_l - c, \bar{\theta}_l + c\right]$, where $c = \pi/2^4$ (note that we plot only a sector of C), such that we can further restrict the phase domain and continue with the next round. We also see that $P(\boldsymbol{\vartheta})$ shows correlations between the different phases (i.e., the equiprobable surfaces of $P(\boldsymbol{\vartheta})$ resemble ellipsoids instead of spheres), stemming from the cross terms in Eq. (5.17). The consequences of these correlations will be discussed in Sec. 5.2.2.

To gain more insight about how the BQMEA proceeds, we provide a few exemplary details about the algorithm for the estimation of $d = 2$ phases in Fig. 5.7. In Fig. 5.7(a), we show the average number of measurements n_{meas} per round k for a simulation of 1000 runs

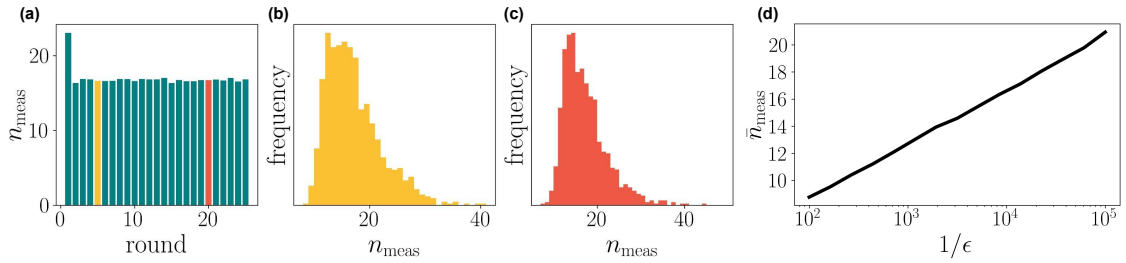

Figure 5.7: (a) Average number of measurements per round, n_{meas}, during the different rounds of the Bayesian quantum multiphase estimation algorithm (BQMEA) for $d = 2$ and $\epsilon = 10^{-4}$, estimated from 1000 simulations of the algorithm. (b,c) Histograms of the number of measurements used (b) in the fifth and (c) in the 20th round of the algorithm, from the same 1000 simulations as in (a). (d) Average asymptotic number of measurements per round, $\bar{n}_{\mathrm{meas}}$, depending on the decision parameter ϵ. For 15 different values of ϵ, $\bar{n}_{\mathrm{meas}}$ was estimated from 100 simulations of the algorithm for 25 rounds.

of 25 rounds of the BQMEA with $\epsilon = 10^{-4}$. Note that n_{meas} is related to the average total number of measurements after K rounds, $N_{\mathrm{meas}}(K)$, as $N_{\mathrm{meas}}(K) = \sum_{k=0}^{K} n_{\mathrm{meas}}(k)$. We see that after completing the first round, the average number of measurements remains constant in the remaining rounds. In the first round, more measurements are required because the initially flat Bayesian distribution $P(\boldsymbol{\vartheta})$ has to be localized sufficiently before one is able to restrict the phase domain and proceed with the second round. In contrast, in the remaining rounds, the initial distribution at the beginning of the rounds is already significantly localized such that less measurements are needed.

In Fig. 5.7(b,c) we show a histogram (from the same 1000 runs of the algorithm) of how many measurements have been used (b) in the fifth round and (c) in the 20th round of the BQMEA. We see that the distribution of the number of measurements remains similar during the different rounds of the algorithm. Furthermore, due to the tail of the distributions, we can see that using a fixed number of measurements per round (which would render the algorithm nonadaptive) could be problematic: when using only a small number of measurements per round, $P(\boldsymbol{\vartheta})$ might not be localized enough to restrict the phase domain, resulting in a large error rate. On the other hand, using a large number of measurements per round increases the constant overhead of precision scaling with respect to the Heisenberg limit.

Finally, in Fig. 5.7(d), we plot the average asymptotic number of measurements per round, $\bar{n}_{\mathrm{meas}}$, for different values of the decision parameter ϵ. Here, we approximate $\bar{n}_{\mathrm{meas}}$ using rounds two to 25 of the algorithm, $\bar{n}_{\mathrm{meas}} \approx \sum_{k=1}^{24} n_{\mathrm{meas}}(k)/24$, for 100 runs of the BQMEA ($d = 2$) for each ϵ. We see that, as expected, decreasing ϵ results in more measurements per round. In particular, we observe that $\bar{n}_{\mathrm{meas}} = O(\ln 1/\epsilon)$, or, as we see below, $\bar{n}_{\mathrm{meas}} = O(\ln 1/P_{\mathrm{err}})$. This scaling is similar to what is seen in the original single-phase QPE algorithm [429], where the phase θ is estimated up to an error of ε (with high probability, see discussion in Sec. 5.1.2) using $O(\ln 1/\varepsilon)$ ancilla qubits.

5.2.2 Precision scaling of the BQMEA

In this section, we discuss the performance of the BQMEA in the presence and the absence of noise, and compare it to several single-phase and sequential multiphase estimation protocols.

Figure of merit and resource counting

Before we can meaningfully discuss the performance scaling of the BQMEA and other estimation algorithms, we must define an appropriate figure of merit and a correct account of the physical resources used during the estimation. In single-phase estimation, the Bayesian variance of the estimation protocol is given by [487]

$$V = 4 \int \mathrm{d}\vartheta P(\vartheta) \sin^2[(\vartheta - \bar{\theta})/2], \tag{5.22}$$

where $P(\vartheta)$ is the Bayesian distribution and $\bar{\theta}$ is the estimate for the true phase θ. Note that for a sufficiently sharp $P(\vartheta)$ (and by correctly regarding the periodicity of the phase), V simplifies to the Bayesian mean-squared-error $\int \mathrm{d}\vartheta P(\vartheta)(\vartheta - \bar{\theta})^2$. We note also that the Bayesian variance V does not necessarily coincide with the frequentist variance $V_{\mathrm{fre}} = 4\langle \sin^2[(\bar{\theta} - \theta)/2]\rangle$, where $\langle\cdot\rangle$ is an average over many realizations of the estimation. Both variances are known to coincide asymptotically for many repetitions of the estimation protocol [134, 386], but the frequentist variance V_{fre} can only be computed when the true phase is known, which one could argue is never the case in real experiments. In our case, since we perform simulations, we have access to the true phase θ and thus can compare both variances. However, since we restrict the phase domain between the different rounds, we introduce an error probability $P_{\mathrm{err}}^{(d)}$ that does not affect the Bayesian variance but, whenever the phase restriction is erroneous, can be seen in the frequentist variance. We observe numerically that when we only consider the realizations that do not show an error (representing the vast majority of the realizations, see below), both variances coincide up to statistical fluctuations. Since our estimation protocol is Bayesian, we will focus on the Bayesian variance V in the following, but below also analyze in detail the error rates $P_{\mathrm{err}}^{(d)}$.

In a multiphase estimation protocol, we generalize the expression of V to obtain the $d \times d$ Bayesian covariance matrix $\mathbf{V}$ with elements

$$V_{ab} = 4 \int \mathrm{d}^d\boldsymbol{\vartheta} P(\boldsymbol{\vartheta}) \sin[(\vartheta_a - \bar{\theta}_a)/2] \sin[(\vartheta_b - \bar{\theta}_b)/2]. \tag{5.23}$$

The diagonal element V_{ll} of the covariance matrix is the single-phase Bayesian variance V, Eq. (5.22), of the phase θ_l, while the off-diagonal elements quantify the correlations of the distribution $P(\boldsymbol{\vartheta})$. According to Eq. (5.23), an arbitrary linear combination $\mathbf{n} \cdot \boldsymbol{\theta}$ ($\mathbf{n} \in \mathbb{R}^d$) is estimated with Bayesian variance $V_{\mathbf{n}\cdot\boldsymbol{\theta}} = \sum_{a,b} n_a n_b V_{ab}$.

To consistently account for the physical resources used in the BQMEA, we apply the standard way of counting resources in QPE algorithms [19, 40, 429, 431, 432], and count

the total number of applications of the controlled-U gate U_c, Eq. (5.14). We note that the gate U_c of Eq. (5.14) is generally more difficult to implement (i.e., it requires more elementary gates) than the standard controlled-U gate that is used in single-phase QPE algorithms, see Eq. (5.3). However, as in single-phase QPE algorithms, the complexity to implement U_c in terms of elementary gates is not included in resource counting. Furthermore, from the form of Eq. (5.14), one might think that the application of U_c should correspond to d applications of the unitary U. However, in each term of Eq. (5.14), the unitary U appears only once such that, for any initial state, U_c can only imprint the single phases θ_l once and never the phase $d \times \theta_l$. This is in contrast to, e.g., the gate $U \otimes \cdots \otimes U$ that should be counted as d applications of U.

A second way to understand that the gate U_c should be counted as one application of U is by means of the formal equivalence to optical implementations, see the discussion in the end of Sec. 5.1.1. As we will see in detail in Sec. 5.2.3, the M-fold application of U_c is formally equivalent to a single application of the phase-imprinting unitary on a generalized N00N state with M particles,

$$|\psi\rangle = \frac{1}{\sqrt{d+1}} \left(|M, 0, \ldots, 0\rangle + |0, M, 0, \ldots, 0\rangle + |0, \ldots, 0, M\rangle \right). \tag{5.24}$$

Here, the accounted-for number of resources is the number of particles M, according to the standard resource counting used in multiparameter quantum metrology. For several recent experimental results that use this resource counting for multiphase estimation, see Refs. [476, 479, 480, 481, 482, 488]. Therefore, via the formal equivalence, the M-fold application of U_c should be counted as M applications of U (and not, e.g., $M \times d$).

Consequently, in the BQMEA, if we use m_k single measurements in the kth round of the algorithm, the total number of resources used in K rounds of the algorithm is given by $N_T = \sum_{k=0}^{K-1} m_k M_k = \sum_{k=0}^{K-1} m_k 2^k$. The total number of measurements (that sometimes is used as the accounted-for resource in estimation protocols, see, e.g., Ref. [448]) is given by $N_{\text{meas}} = \sum_{k=0}^{K-1} m_k$.

Sensitivity of the BQMEA in the absence of noise

We now discuss the performance of the BQMEA for $d = 1, 2$, and 3. Recall that in Fig. 5.6, we have shown two exemplary snapshots of the Bayesian distributions $P(\boldsymbol{\vartheta})$ during the algorithm for $d = 2$ and $d = 3$. In Fig. 5.8, we show numerical results of the scaling of the elements of the Bayesian covariance matrix, Eq. (5.23), during the algorithm. For each d, we have performed 100 simulations of the BQMEA for 25 rounds, corresponding to a total resource count of $N_T \approx 10^9$ and about $N_{\text{meas}} \approx 500$ measurements, and use a decision parameter of $\epsilon = 10^{-4}$. In each simulation, a different uniformly-generated random $\boldsymbol{\theta} \in [0, 2\pi]^d$ is estimated. In Fig. 5.8, we plot the Bayesian variances V_{ii} and covariances V_{ij} $(i \neq j)$ for (a) $d = 1$, (b) $d = 2$, and (c) $d = 3$. The points show the numerical results

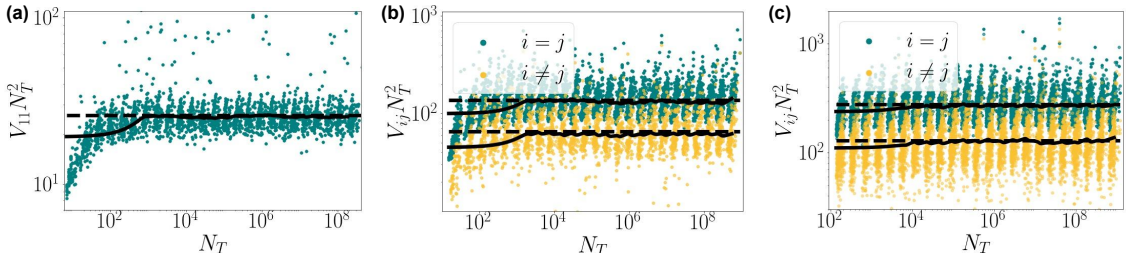

Figure 5.8: Scaling of the Bayesian covariance matrix $\mathbf{V}$ (multiplied by N_T^2) as a function of the used resources N_T during the BQMEA for (a) $d = 1$, (b) $d = 2$, and (c) $d = 3$. Numerical simulations of the Bayesian variances V_{ii} and covariances V_{ij} $(i \neq j)$ are shown as blue dots and yellow dots, respectively. Each dot corresponds to the simulated (co)variance after a specific round of the algorithm during the first 25 rounds of the algorithm. For each d, we show the results of 100 simulations. The solid lines show average values of the components of $\mathbf{V}$ that, for $N_T \gg 1$, yield the corresponding asymptotic limit, shown as horizontal dashed lines, that indicate a precision scaling as $\mathbf{V} \propto 1/N_T^2$ and are used to estimate the fitting parameters $C_{\mathrm{H}}^{(d)}(\epsilon)$, cf. Eqs. (5.25-5.27). In each simulation, we estimate a randomly chosen $\boldsymbol{\theta} \in [0, 2\pi]^d$ using the decision parameter $\epsilon = 10^{-4}$. The figure is taken from Ref. [5].

of $\mathbf{V}$ after each round of the algorithm (i.e., after the last measurement of each round). We emphasize that in Fig. 5.8, we have not ignored any erroneous runs of the algorithm. Also, we indicate the averages and asymptotic limits of the Bayesian variances and covariances as solid and dashed lines, respectively.

We see that, after an initial transient, the covariance matrix $\mathbf{V}$ scales as $\mathbf{V} \propto 1/N_T^2$, so it shows a Heisenberg-limited scaling behaviour. The reason for the transient is that in the beginning of the estimation protocol, the initially flat Bayesian distribution $P(\boldsymbol{\vartheta})$ requires more measurements with fixed M to become localized before one can restrict the phase domain and continue with the second round (cf. Fig. 5.7). Thus, initially, $\mathbf{V}$ does not scale as $1/N_T^2$ but shows a shot-noise scaling $1/N_T$. The different simulated results of the elements of $\mathbf{V}$ fluctuate around the average mostly due to varying numbers of measurements m_k that are used before proceeding to the next round of the algorithm. For $N_T \gg 1$, we can use the average covariances to estimate the asymptotic behaviour of the precision scaling of the BQMEA. The exact numerical scaling depends on the chosen value of the decision parameter ϵ and can be fitted by

$$\mathbf{V} = \frac{C_{\mathrm{H}}^{(1)}(\epsilon)}{N_T^2} \qquad \text{for} \quad d = 1; \tag{5.25}$$

$$\mathbf{V} = \frac{C_{\mathrm{H}}^{(2)}(\epsilon)}{N_T^2} \begin{bmatrix} 1 & 0.47 \\ 0.47 & 1 \end{bmatrix} \qquad \text{for} \quad d = 2; \tag{5.26}$$

$$\mathbf{V} = \frac{C_{\mathrm{H}}^{(3)}(\epsilon)}{N_T^2} \begin{bmatrix} 1 & 0.45 & 0.45 \\ 0.45 & 1 & 0.45 \\ 0.45 & 0.45 & 1 \end{bmatrix} \qquad \text{for} \quad d = 3, \tag{5.27}$$

with a fitting parameter $C_{\mathrm{H}}^{(d)}$ that depends on ϵ. In a second set of simulations with varying

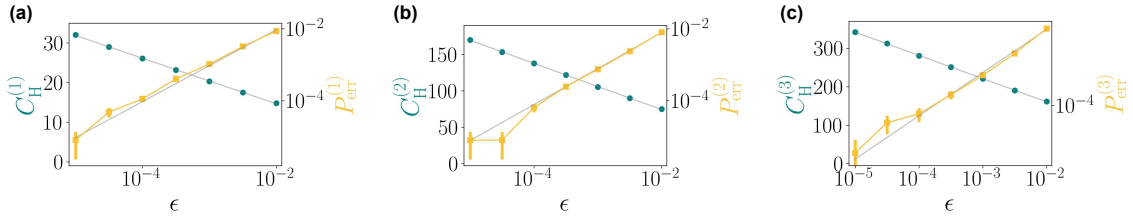

Figure 5.9: The fitting parameters $C_{\mathrm{H}}^{(d)}$ (blue dots) of the scaling of the Bayesian covariance matrix $\mathbf{V}$, cf. Eqs. (5.25-5.27), and the error rates $P_{\mathrm{err}}^{(d)}$ (orange squares), as a function of the decision parameter ϵ for (a) $d = 1$, (b) $d = 2$, and (c) $d = 3$. Each value is estimated by an average of 10^5 repetitions ($d = 1$ and $d = 2$), and 5×10^4 repetitions ($d = 3$), of the first 25 rounds of the BQMEA. The gray lines are fits that correspond to Eqs. (5.28-5.30) for $C_{\mathrm{H}}^{(d)}(\epsilon)$ and Eqs. (5.32-5.34) for $P_{\mathrm{err}}^{(d)}(\epsilon)$. The error bars are estimated standard deviations of the error rate, see the text. The figure is taken from Ref. [5].

ϵ (here we simulate the first 25 rounds of the BQMEA with 10^5 repetitions for $d = 1$ and $d = 2$, and 5×10^4 repetitions for $d = 3$), we estimate the dependence of $C_{\mathrm{H}}^{(d)}$ on ϵ, see the blue dots in Fig. 5.9, to obtain

$$C_{\mathrm{H}}^{(1)}(\epsilon) = 3.13 + 2.50 \ln 1/\epsilon \qquad \text{for} \quad d = 1; \tag{5.28}$$

$$C_{\mathrm{H}}^{(2)}(\epsilon) = 10.8 + 13.8 \ln 1/\epsilon \qquad \text{for} \quad d = 2; \tag{5.29}$$

$$C_{\mathrm{H}}^{(3)}(\epsilon) = 40.1 + 26.2 \ln 1/\epsilon \qquad \text{for} \quad d = 3. \tag{5.30}$$

Note that for $\epsilon = 10^{-4}$, we obtain the values $C_{\mathrm{H}}^{(2)} \approx 138$ and $C_{\mathrm{H}}^{(3)} \approx 281$ that can be seen in Fig. 5.8.

Next, we analyze and estimate the error rate $P_{\mathrm{err}}^{(d)}$ depending on the decision parameter ϵ. We say that an error has occurred whenever the phase restriction has been erroneous, i.e., whenever the true phase $\boldsymbol{\theta}$ lies outside the restricted phase hypercube C, $\boldsymbol{\theta} \notin C$. Then, the error rate is $P_{\mathrm{err}}^{(d)}$ is defined as the probability of an error per round. Before we estimate $P_{\mathrm{err}}^{(d)}$, we observe that, after a larger error probability after the first round of the BQMEA, $P_{\mathrm{err}}^{(d)}$ remains constant (or even slightly decreases) for the remaining rounds, see Fig. 5.10. This is due to the fact that, before the first round, the Bayesian distribution $P(\boldsymbol{\vartheta})$ is flat, while in later rounds, $P(\boldsymbol{\vartheta})$ is already localized in the phase domain (cf. Fig. 5.6), such that errors in the estimation are less probable.

To find a simple estimate of the asymptotic error rate $P_{\mathrm{err}}^{(d)}$, we assume a constant error probability for each round. As the error probability is higher in the first round, this assumption results in a slight overestimation of the asymptotic rate $P_{\mathrm{err}}^{(d)}$. We count the number of errors N_{err} that have occurred in the simulation of N_{sim} runs of the BQMEA, meaning the number of times that we find $\boldsymbol{\theta} \notin C$ after completing the 25 rounds of the algorithm. This results in an estimated total error probability of $P_{\mathrm{err,T}}^{(d)} = N_{\mathrm{err}}/N_{\mathrm{sim}}$ after $k = 25$ rounds of the algorithm, that can be used to estimate the error rate $P_{\mathrm{err}}^{(d)}$ by means of the relation

$$1 - P_{\mathrm{err,T}}^{(d)} = (1 - P_{\mathrm{err}}^{(d)})^k \tag{5.31}$$

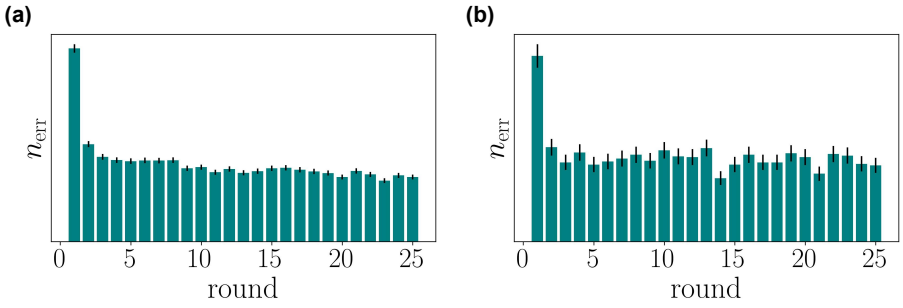

Figure 5.10: The error distribution n_{err} during the first 25 rounds of the BQMEA for $d = 1$ using a decision parameter (a) $\epsilon = 10^{-2}$ and (b) $\epsilon = 10^{-3}$. Each distribution is estimated using 10^6 simulations of the algorithm. The shown error bars are estimated standard deviations of the error counts according to the Poissonian distribution, $\Delta_{n_{\mathrm{err}}} = \sqrt{n_{\mathrm{err}}}$. The figure is taken from Ref. [5].

corresponding the probability for no error in the first k rounds of the algorithm. In this way, we find the error rate estimates that are shown as yellow squares in Fig. (5.9). The estimated standard deviations of the error rates $P_{\mathrm{err}}^{(d)}$ are obtained by error propagation and assuming a Poissonian distribution of the error counts, $\Delta_{P_{\mathrm{err,T}}^{(d)}} = \sqrt{N_{\mathrm{err}}}/N_{\mathrm{sim}}$. We note that in Fig. 5.9, the values of $P_{\mathrm{err}}^{(d)}$ are estimated using the same simulations that were used to estimate the values of $C_{\mathrm{H}}^{(d)}$.

After fitting the error rates $P_{\mathrm{err}}^{(d)}$ by the simple model $c \times \epsilon$, shown as the gray lines in Fig. 5.9, we can estimate the error rates as

$$P_{\mathrm{err}}^{(1)} = 0.94\epsilon \qquad \text{for} \quad d = 1; \tag{5.32}$$

$$P_{\mathrm{err}}^{(2)} = 0.78\epsilon \qquad \text{for} \quad d = 2; \tag{5.33}$$

$$P_{\mathrm{err}}^{(3)} = 0.58\epsilon \qquad \text{for} \quad d = 3. \tag{5.34}$$

We note that for a more exact analysis of $P_{\mathrm{err}}^{(d)}$, one might use a more accurate model. For instance, when fitting $P_{\mathrm{err}}^{(d)}$ with the model $c_1 \times \epsilon^{c_2}$, we find the parameters $c_1 = 0.57$ and $c_2 = 0.91$ for $d = 1$, $c_1 = 0.94$ and $c_2 = 1.04$ for $d = 2$, and $c_1 = 0.64$ and $c_2 = 1.02$ for $d = 3$. Furthermore, a proper error rate analysis should also account for the larger error probability of the first round of the algorithm (cf. Fig. 5.10). For our qualitative error rate analysis that we will use to compare the performance of the BQMEA to other existing estimation strategies in the literature, we will content ourselves with the simple estimations of Eqs. (5.32-5.34).

Finally, we briefly mention that the precision of the BQMEA scales exponentially with the number of total measurements N_{meas}. This scaling can be seen for $d = 2$ in Fig. 5.11, where we show results from 100 simulations of the algorithm for the first 25 rounds, using a decision parameter $\epsilon = 10^{-4}$. The solid line indicates an exponential fit of V_{ii} given by $V_{ii} \sim 0.61e^{-0.082N_{\mathrm{meas}}}$. For $d = 1$ and $d = 3$, one can also observe an exponential dependence of the covariance matrix $\mathbf{V}$ with respect to N_{meas}. We note that in the above description of the BQMEA, N_{meas} corresponds to the number of measurements (or the

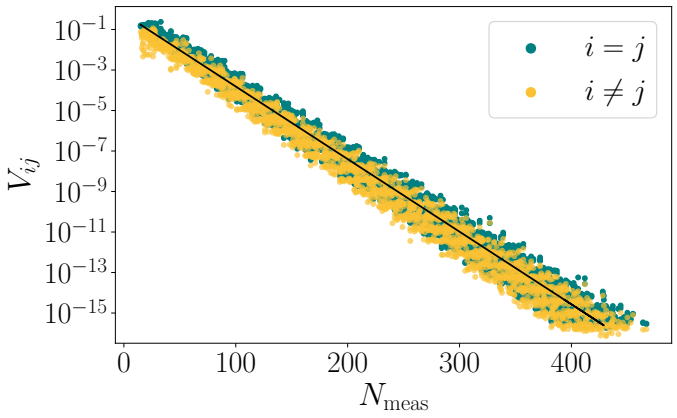

Figure 5.11: Components V_{ij} of the Bayesian covariance matrix as a function of the total number of measurements N_{meas}. The dots indicate results of the BQMEA for $d = 2$ during the first 25 rounds of the algorithm using $\epsilon = 10^{-4}$. The solid line represents an exponential fit of V_{ii}. The figure is taken from Ref. [5].

number of ancilla qudits) that are used during the estimation protocol, while in the optical implementations that we will discuss in Sec. 5.2.3, N_{meas} corresponds to the total number of applications of the unknown phase shifts in the N00N-state protocol, or the total number of single photons used in the single-photon multipass protocol.

Comparison of the BQMEA for $d = 1$ with existing single-phase estimation protocols

For the estimation of a single phase, $d = 1$, we can directly compare the BQMEA to different single-phase estimation protocols in the literature. The protocols can be divided into the groups of "forward" protocols (initiating with $M = 1$ and thereupon increasing M) and "backward" protocols (initiating with $M = 2^{K-1}$ and thereupon decreasing M). Direct applications of the original QPE algorithm (that we have discussed in detail in Sec. 5.1.2) make use of the semiclassical implementation of the quantum Fourier transform [433] and thus consist of a adaptive backwards measurement protocol [38, 444]: the measurement protocol starts with the measurements using the largest value of $M = 2^{K-1}$ (recall that K is the number of rounds of the estimation protocol) and then proceeds backwards until reaching $M = 1$, while always adaptively changing the control phase ϕ based on the recorded measurement results. The adaptive strategy of Ref. [38] was used to reach a Heisenberg-limited precision scaling of the Bayesian variance, $V \sim C_s/N_T^2$, with $C_s = 23$. Due to the optimized adaptive adjustment of the control phases for each measurement, the error rate is negligibly small. However, one cannot simply increase the precision of the estimation procedure by adding more rounds: since one necessarily has to start with the measurements of the round with the largest M, the number of rounds K has to be known or chosen in advance because adding more rounds during the algorithm requires a complete restart of the estimation protocol.

In contrast, the BQMEA is part of the class of forward estimation protocols [430, 447,

448, 451, 468] that initiate with the measurements using $M = 1$ and generally do not require fixing the number of rounds K in advance. Forward estimation protocols can be adaptive, where different parameters such as number of measurements per round, control phases, or the chosen values of M may depend on previous measurement outcomes[10]. In Ref. [430], a fully nonadaptive Heisenberg-limited measurement protocol was discussed that increases the number of repetitions for the rounds with small M, to obtain a negligible error probability. The small error probability comes at the price of an increased constant overhead in the Heisenberg-limited scaling, showing a precision scaling as $V \sim C_s/N_T^2$ with $C_s = 40.5$.

A widely applied Bayesian forward estimation protocol was introduced in Ref. [448]. In this so-called rejection-filtering phase estimation (RFPE), the Bayesian distribution $P(\vartheta)$ after each measurement is approximated as a Gaussian distribution, reaching a precision scaling of $V \sim C_s/N_T^2$ with $C_s = 22$. The restrictive simplification of approximating $P(\vartheta)$ with by Gaussian distribution introduces a finite error probability, similar to the BQMEA. However, the phase restrictions used during the BQMEA allow for the recording of more information about $P(\vartheta)$, such that for similar precision scaling, the error probability is below the one of Ref. [448]. Indeed, from Fig. 5.9(a), we see that the BQMEA reaches $C_{\mathrm{H}}^{(1)}$ for a decision parameter $\epsilon \approx 5.3 \times 10^{-4}$ that corresponds to an error rate of $P_{\mathrm{err}}^{(1)} \approx 5 \times 10^{-4}$. For this ϵ, we compare the cumulative error distribution of the single-phase BQMEA to the error distribution that was reached in the RFPE protocol of Ref. [448]. Both distributions are shown in Fig. 5.12, where they are estimated from 1000 simulations of the estimation protocols each consisting of $N_{\mathrm{meas}} = 200$ single measurements. We see that, in the BQMEA, larger errors are less probable in comparison to the RFPE protocol.

Comparison of the BQMEA to sequential multiphase estimation strategies

We now compare the precision scaling of the BQMEA for $d = 2$ and $d = 3$ to respective sequential measurement protocols of the phases. A sequential multiphase estimation protocol interrogates each of the d phases individually, making use of single-phase estimation schemes like the one shown in Fig. 5.2(b), where one prepares a single register state in one of the d eigenstates and performs the measurement of each phase independently. We first want to note that due to spatial or temporal boundary conditions of the multiphase estimation task, such as time-varying signals, one might be forced to use a parallel sensing technique such as the BQMEA. However, as we see now, even if a sequential protocol is possible, a parallel measurement scheme may still outperform the sequential one for the estimation of specific linear combinations of the phases.

[10] In contrast, any nonadaptive measurement protocol can be equivalently implemented in a forward or backward fashion because the are no adaptive parameters that require knowledge of previous measurement results.

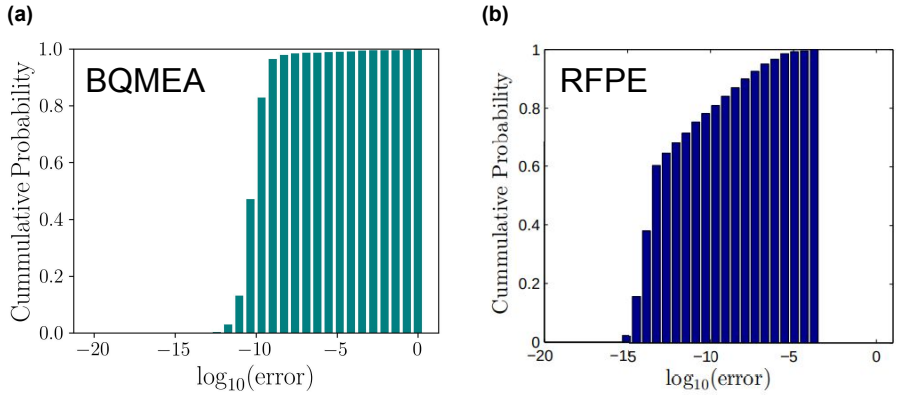

Figure 5.12: Cumulative error distributions for single-phase estimation using (a) the BQMEA and (b) the rejection-filtering phase estimation (RFPE) after 200 measurements of the protocols. For each protocol, the error distributions are estimated using 1000 simulated phase estimations. For the BQMEA, we use $\epsilon = 5.3 \times 10^{-4}$ such that both protocols show a precision scaling of $V \sim C_s/N_T^2$ with $C_s = 22$. The figure of part (b) is taken from the Supplementary Material of Ref. [448] (Fig. 2; right).

We assume here for simplicity that we use the same amount of resources to measure each of the d phases. Note that depending on the specific linear combination $\mathbf{n} \cdot \boldsymbol{\theta}$ that we want to measure, one can optimize the precision by accordingly distributing the total number of available resources, N_T, over the individual measurements. Thus, we simply perform d separate single estimation protocols each using the N_T/d resources (for simplicity, we assume that d divides N_T). Since the precision scaling of a single-phase BQMEA using N_T resources is $V_{\theta_l} = C_{\mathrm{H}}^{(1)}/N_T^2$, a sequential measurement scheme making use of N_T resources estimates the linear combination $\mathbf{n} \cdot \boldsymbol{\theta}$ with a precision of

$$V_{\mathbf{n}\cdot\boldsymbol{\theta}}^{(\mathrm{seq})} = \sum_{l=1}^{d} n_l^2 V_{\theta_l} = \frac{C_{\mathrm{H}}^{(1)} d^2 \sum_{l=1}^{d} n_l^2}{N_T^2}. \tag{5.35}$$

In order to meaningfully compare the sequential and parallel estimation strategies, we have to fix the error probability P_{err} for both cases. For this purpose, we express the scaling constants $C_{\mathrm{H}}^{(d)}$ explicitly in terms of the error rates $P_{\mathrm{err}}^{(d)}$ using Eqs. (5.28-5.30) and Eqs. (5.32-5.34). For instance, we have that $C_{\mathrm{H}}^{(1)}(P_{\mathrm{err}}) = 3.13 + 2.50 \ln 0.94/P_{\mathrm{err}}$.

As an example, we consider the measurement of the phase difference $\theta_1 - \theta_2$ for $d = 2$. According to Eq. (5.35), a sequential scheme measuring one phase after the other results in a precision scaling of $V_{\theta_1-\theta_2}^{(\mathrm{seq})} = 8C_{\mathrm{H}}^{(1)}(P_{\mathrm{err}})/N_T^2$. In contrast, with Eq. (5.26), a parallel estimation of both phases results in a precision scaling of

$$V_{\theta_1-\theta_2}^{(\mathrm{par})} = \frac{C_{\mathrm{H}}^{(2)}(P_{\mathrm{err}})}{N_T^2} \begin{bmatrix} 1 & -1 \end{bmatrix} \begin{bmatrix} 1 & 0.47 \\ 0.47 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ -1 \end{bmatrix} = 2(1 - 0.47)\frac{C_{\mathrm{H}}^{(2)}(P_{\mathrm{err}})}{N_T^2} \tag{5.36}$$

with $C_{\mathrm{H}}^{(2)}(P_{\mathrm{err}}) = 10.8 + 13.8 \ln 0.78/P_{\mathrm{err}}$. Taking, e.g, an error rate of $P_{\mathrm{err}} = 10^{-3}$, the sequential protocol offers a sensitivity of $V_{\theta_1-\theta_2}^{(\mathrm{seq})} = 162/N_T^2$ while the parallel protocol achieves a sensitivity of $V_{\theta_1-\theta_2}^{(\mathrm{par})} = 109/N_T^2$. Analogously, one obtains that for $d = 3$ and

$P_{\text{err}} = 10^{-3}$, all mutual differences $\theta_l - \theta_k$ between two of the three phases are measured with a sensitivity of $V^{(\text{seq})}_{\theta_l-\theta_k} = 364/N_T^2$ using the sequential protocol, and with a sensitivity of $V^{(\text{par})}_{\theta_l-\theta_k} = 227/N_T^2$ using the parallel protocol.

To summarize, we see that for a comparable error rate and a fixed amount of resources N_T, a parallel estimation offers a better sensitivity to measure specific linear combinations $\mathbf{n}\cdot\boldsymbol{\theta}$ such as the differences $\theta_l - \theta_k$[11]. This advantage stems from the correlations of the Bayesian distribution $P(\boldsymbol{\vartheta})$, see Fig. 5.6: for instance, in Eq. (5.36), we see that the overhead factor of the variance is reduced by the correlations of $P(\boldsymbol{\vartheta})$. Note that this correlation also results in a decreased performance for the parallel estimation of other linear combinations, such as, e.g., the sum $\theta_1 + \theta_2$.

Sensitivity of the BQMEA in the presence of noise

We now introduce an extension of the BQMEA in the presence of noise. As a noise model, we assume that the phase imprinting of the phase θ_l is accompanied by decoherence characterized by a decoherence rate Γ_l. In particular, if we apply the basic circuit of the BQMEA, Fig. 5.4, with M applications of controlled-U gate U_c, the state of the ancilla qudit after the phase imprinting is given by

$$E(\rho) = \sum_{l=0}^{d} K_l \rho K_l. \tag{5.37}$$

Here, $\rho = |\psi_{\text{A}}\rangle\langle\psi_{\text{A}}|$ is the ancilla state of the BQMEA after the application of U_c in the absence of noise [$|\psi_{\text{A}}\rangle$ is given in Eq. (5.15)], and E denotes the dephasing channel that is given in terms of the Kraus operators K_l that are defined as $\langle\mathbf{m}|K_0|\mathbf{n}\rangle = e^{-\Gamma_l M}\delta_{mn}$ (we have defined $\Gamma_0 = 0$ and δ_{mn} denotes the Kronecker delta) and $\langle\mathbf{m}|K_l|\mathbf{n}\rangle = \sqrt{1-e^{-2\Gamma_l M}}\delta_{ml}\delta_{nl}$ for $l > 0$. This noise model generalizes the phase damping channel for qubits [19] to a qudit phase damping channel. In the optical implementation that we discuss in Sec. 5.2.3, the noise model corresponds to a mode-dependent dephasing that occurs during the phase imprinting. Furthermore, in the case $d = 1$, the noise model leads to a decay of the measurement probabilities towards white noise, a model that is often considered in single-phase estimation protocols, see, e.g., Ref. [448]. The final probabilities [see Eq. (5.17) for the final probabilities in the absence of noise] are now given

[11] Note that, actually, in the above comparison, the error probability of the parallel estimation is even smaller than the one for the sequential protocol because we have simply equated P_{err}, the error probability per round. However, since the sequential protocol has to perform d separate estimations, more rounds (with less single measurements per round) are required, resulting in a larger total error probability. Thus, equating the total error probability would only increase the scaling advantage of the parallel scheme with respect to the sequential one.

Input: $\{\epsilon, K, \Gamma_l\}$
$P(\boldsymbol{\vartheta}) = 1/(2\pi)^d$
for $k = 0, \ldots, K-1$:
  $M = \min_l[2^k, 1/\Gamma_l]$; $P_{\text{half}} = 0$
  while $P_{\text{half}} < 1 - \epsilon$:
    $\boldsymbol{\phi} = \text{generate\_random}()$
    $o = \text{measurement}(M, \boldsymbol{\phi})$
    $P(\boldsymbol{\vartheta}) = \text{Bay\_update}(P(\boldsymbol{\vartheta}), o, M, \boldsymbol{\phi})$
    $P_{\text{half}} = \text{compute}(P(\boldsymbol{\vartheta}))$
  $P(\boldsymbol{\vartheta}) = \text{cut\_grid}(P(\boldsymbol{\vartheta}))$
  $P(\boldsymbol{\vartheta}) = \text{normalize}(P(\boldsymbol{\vartheta}))$
return: $P(\boldsymbol{\vartheta})$

Table 5.2: The pseudocode of the Bayesian quantum multiphase estimation algorithm (BQMEA) in the presence of noise. In contrast to the algorithm in the absence of noise, cf. Tab. 5.1, we bound the maximal number M of applications of U_c by $\min_l[1/\Gamma_l]$ according to the dephasing rates Γ_l.

by

$$P(o|\boldsymbol{\theta}, \boldsymbol{\phi}, M) = \frac{1}{(d+1)^2}\Big\{d + 1 + 2\sum_n e^{-\Gamma_n M}\cos[M\theta_n + \beta_n(o)] + \\ + 2\sum_{m<n} e^{-(\Gamma_n+\Gamma_m)M}\cos[M(\theta_n - \theta_m) + \gamma_{nm}(o)]\Big\}. \quad (5.38)$$

The exponential increase of the noise with M effectively erases the information about θ_l in measurements with $M > 1/\Gamma_l$. Thus, as soon as the noise becomes dominant, we stop further increasing M. In particular, in the kth round, we choose $M = \min[2^k, 1/\Gamma_l]$, while all other parts of the BQMEA remain the same as for the BQMEA in the absence of noise. For a pseudocode of the BQMEA in the presence of noise, see Tab. 5.2. Note that the strategy to halt increasing M as soon as the noise becomes dominant is similar to what was proposed in Ref. [448].

We now examine the performance of the BQMEA in the presence of noise for the exemplary case of $d = 2$, where we consider the dephasing rates $\Gamma_1 = 0.02$ and $\Gamma_2 = 0.01$. In Fig. 5.13, we plot the components of the covariance matrix $\mathbf{V}$ (times N_T) as a function of N_T. While the precision of the BQMEA initially shows a scaling as $1/N_T^2$ (here the noise still has a little effect), for $N_T \geq 10^4$, the scaling approaches the shot-noise scaling $1/N_T$ as expected for measurements using constant M. However, we emphasize that even though $\mathbf{V}$ is scaling as $1/N_T$, the precision is still far below the shot-noise limit that is indicated as a dashed line. This is because during the initial part of the BQMEA that is not affected by the noise, the precision is already strongly increased with respect to the shot-noise limit. Furthermore, we also see that the precision of estimating θ_2 is larger than the one of esti-

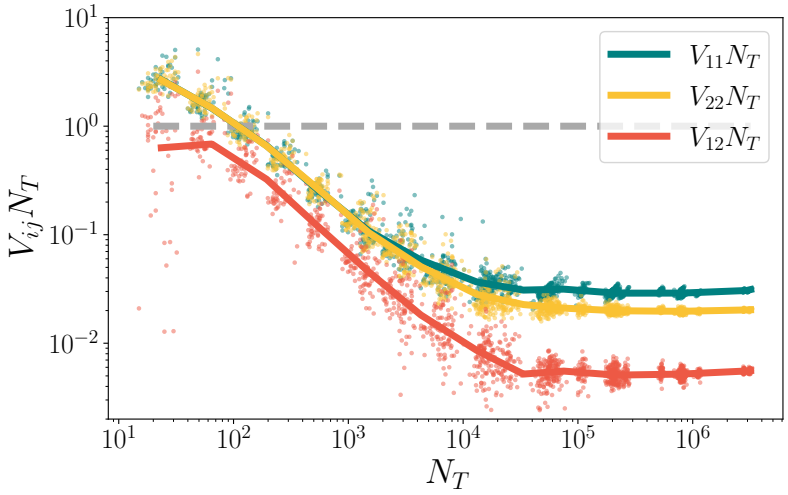

Figure 5.13: Components of the Bayesian covariance matrix $\mathbf{V}$ during the BQMEA for $d = 2$ in the presence of noise. The dots represent single results after each round of 100 simulations of the BQMEA using $\epsilon = 10^{-4}$, $\Gamma_1 = 0.01$ and $\Gamma_2 = 0.02$, and the solid lines are averages. The dashed line indicates the shot-noise limit N_T^{-1}. The figure is taken from Ref. [5].

mating θ_1 because the latter is accompanied by the larger dephasing rate. Finally, note that the correlations decrease in the final stage of the algorithm with respect to the initial stage. This is because the correlations actually decay with the rate $\Gamma_1 + \Gamma_2$, see Eq. (5.38), and are thus more sensitive to the dephasing noise.

We want to mention that an asymmetric noise such as the one above can be counteracted by replacing the first Hadamard gate H_{d+1} by an (M-dependent) unbalanced Hadamard gate $\tilde{H}_{d+1}$ such that the ancilla state's components that correspond to the largest dephasing rates have larger amplitudes.

5.2.3 Optical implementations

In this section, we will discuss two possible optical implementations of the BQMEA. Here, instead of the estimation of d eigenvalues of a unitary matrix, we consider the usual multiphase estimation problem in quantum metrology. We restrict ourselves to optical quantum interferometry, where the unknown phases $\boldsymbol{\theta}$ correspond to the optical phase shifts acting on the different modes of an interferometer. We first discuss an implementation using generalized N00N states, and then focus on a single-photon multipass protocol that can be implemented with start-of-the-art optical elements.

N00N-state implementation

The implementation of the BQMEA with generalized N00N states corresponds to what one could call the "traditional" interferometric setting of multiphase estimation. The problem setting is depicted in Fig. 5.14. A preparation device prepares a state consisting of M parti-

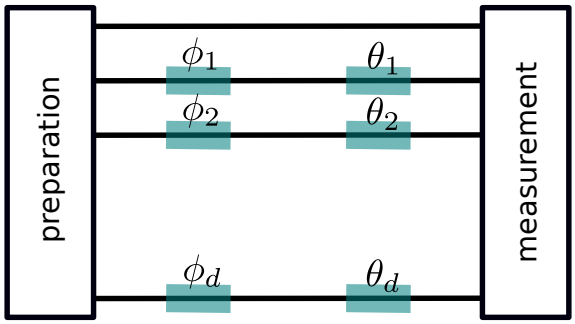

Figure 5.14: The "traditional" interferometric setting of multiphase estimation. After a $(d+1)$ mode state that contains M particles is prepared, the lth mode experiences a control phase shift ϕ_l and the unknown phase shift θ_l. The zeroth mode serves as a reference mode. Finally, the state of the interferometer is measured.

cles distributed among $d+1$ modes. The zeroth mode serves as the reference mode, while the lth mode experiences the known control phase shift ϕ_l followed by the unknown phase shift θ_l. The phases are imprinted by the unitaries $e^{i(\phi_l+\theta_l)n_l}$, where $n_l = a_l^\dagger a_l$ is the number operator of the lth mode. Finally, the state is measured in a measurement device. If we use preparation and measurement devices that reproduce the probabilities of the BQMEA, Eq. (5.17), we know from the above analysis that the setup performs Heisenberg-limited multiphase estimation, where now the total number of resources, N_T, corresponds to the total number of particles that are used in the measurement protocol. As we discussed in Sec. 5.1.1, this is the "traditional" way of counting resources in (multi)phase estimation, that requires the quantum resource of (multipartite) entanglement, cf. Sec. 5.1.4.

To reproduce the probabilities of Eq. (5.17), we prepare the initial state as the generalized multimode N00N state [454, 461, 462, 463, 489]

$$|\psi_{\text{in}}\rangle = \frac{1}{\sqrt{d+1}} \left(|M,0,\ldots,0\rangle + |0,M,\ldots,0\rangle + \cdots + |0,\ldots,0,M\rangle\right), \tag{5.39}$$

where M particles are distributed among the $d+1$ modes (note that we write $|\psi_{\text{in}}\rangle$ in second quantization). After the phases are imprinted according to $e^{i(\phi_l+\theta_l)n_l}$, we project the final state in the interferometer onto the states

$$|\psi_o\rangle = \frac{1}{\sqrt{d+1}} \left(|M,0,\ldots,0\rangle + e^{i\frac{2\pi o}{d+1}} |0,M,\ldots,0\rangle + \cdots + e^{i\frac{2\pi do}{d+1}} |0,\ldots,0,M\rangle\right), \tag{5.40}$$

where $o \in \{0,\ldots,d\}$ is the measurement outcome. The resulting probabilities for the outcome coincide with the main probability of the BQMEA, Eq. (5.17) upon replacing the random control phases ϕ_l with ϕ_l/M. We finally want to note that N00N states are highly susceptible to noisy environments [490, 491] and that, while generalized multimode N00N state have been long proposed [454, 461, 462, 463] and there is recent progress towards their experimental realization [489, 492], generating them as well as projecting the final state according to Eq. (5.40) requires advanced, non-linear optical elements and is thus difficult to implement, especially for larger values of M.

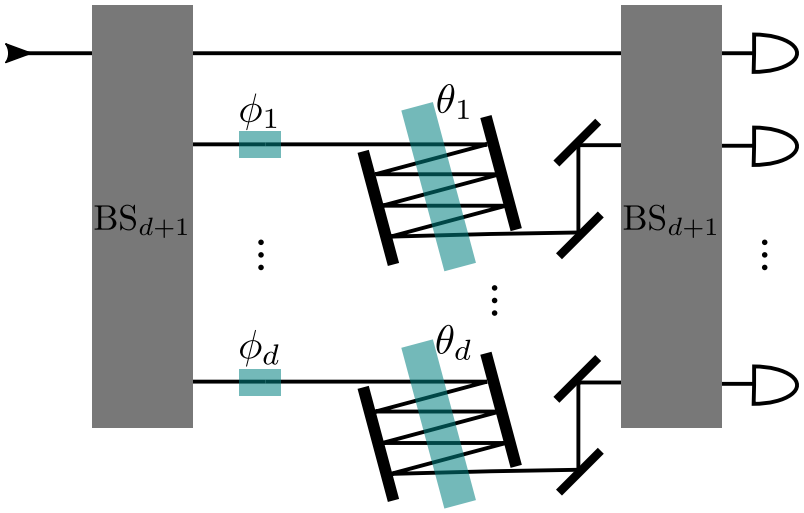

Figure 5.15: Single-photon multipass implementation of the BQMEA. A single photon enters a multiport beam splitter BS_{d+1}. After the imprinting of the control phases $\boldsymbol{\psi}$, the unknown phase shift $\boldsymbol{\theta}$ is applied M times. The zeroth mode of the interferometer serves as a reference mode. After a second multiport beam splitter, the photon is measured in one of the outgoing modes. The figure is taken from Ref. [5].

Single-photon multipass implementation

Now, we discuss a single-photon multipass implementation of the BQMEA that can be implemented with standard state-of-the-art optical elements [479, 493, 494]. This proposal is a generalization of the multipass single-phase estimation protocol that was implemented in Ref. [38]. The implementation is sketched in Fig. 5.15. A single photon enters a balanced multiport beam splitter BS_{d+1} [493, 495] that corresponds to the generalized Hadamard gate H_{d+1} in the BQMEA, see Eq. (5.11). For $d = 2$ and for $d = 3$, the optical elements are called tritters and quarters, respectively, and have been recently implemented with integrated optical circuits [479, 480, 494, 496]. Alternatively, the balanced multiport beam splitter can be implemented using a cascade of balanced two-mode beam splitters [476].

The unitary operation of a tritter is given by [493, 496]

$$\mathrm{BS}_3 = \frac{1}{\sqrt{3}} \begin{bmatrix} 1 & 1 & 1 \\ 1 & e^{i2\pi/3} & e^{i4\pi/3} \\ 1 & e^{i4\pi/3} & e^{i2\pi/3} \end{bmatrix}, \tag{5.41}$$

which coincides with the expression for the generalized Hadamard gate H_3, see Eq. (5.11). For $d = 3$, the quarter is parametrized by a parameter γ that depends on the specific implementation [495],

$$\mathrm{BS}_3 = \frac{1}{2} \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & e^{i\gamma} & -1 & -e^{i\gamma} \\ 1 & -1 & 1 & -1 \\ 1 & -e^{i\gamma} & -1 & e^{i\gamma} \end{bmatrix}. \tag{5.42}$$

For $\gamma = \pi/2$, we obtain the generalized Hadamard gate H_4 of Eq. (5.11). For other values of γ, one only needs to slightly adjust the final probability of Eq. (5.17).

In contrast to the single-pass N00N-state implementation, in the single-photon implementation, the unknown phases θ_l are imprinted using multiple passes of the interferometric modes through the optical element, see Fig. 5.15. Together with the imprinting of the control phases ϕ_l, the lth mode gets acted upon by the operator $e^{i(M\theta_l+\phi_l)}$, where, as above, n_l is the number operator of the lth mode. Again, the zeroth mode serves as a reference. Thus, in the multipass implementation, the phase sensitivity $M \times \theta_l$ in the final probability is achieved by means of M applications of the phase shift θ_l, cf. Sec. 5.1.4. This is similar to the initial description of the BQMEA, where the sensitivity is due to M applications of the controlled-U gate U_c, and in contrast to the implementation with generalized N00N states, where the increased sensitivity is due to the multipartite entanglement of the interferometric state [460, 463].

Finally, the interferometric modes enter a second balanced multiport beam splitter BS_{d+1}, and the photon is measured in one of the $d+1$ outgoing arms. This measurement scheme again reproduces the outcome probabilities of the BQMEA, Eq. (5.17). Therefore, it achieves a (multipass) Heisenberg-limited precision scaling $\mathbf{V} \propto 1/N_T^2$, where N_T is quantified as the total number of applications of the phase shift $\boldsymbol{\theta}$.

5.3 Conclusions and outlook

In this chapter, we have entered the domain of quantum phase estimation algorithms that play central roles in quantum computations and quantum metrology. After an extensive introduction to single-phase estimation, the original quantum phase estimation algorithm and its variants, and an overview of recent literature about multiphase estimation protocols, we have introduced the Bayesian quantum multiphase estimation algorithm (BQMEA). The BQMEA combines different measurements consisting of a basic multiphase-estimation quantum circuit by means of an appropriate Bayesian post-processing. It is able to simultaneously estimate d different and completely unknown phases $\boldsymbol{\theta} \in [0, 2\pi]^d$ with a Heisenberg-limited precision scaling, $\mathbf{V} \propto 1/N_T^2$, where $\mathbf{V}$ is the covariance matrix of the Bayesian posterior distribution and N_T is the total number of resources used in the estimation protocol. Here, N_T corresponds to the number of applications of a generalized controlled-U operation U_c, where the unitary U encodes the unknown phases $\boldsymbol{\theta}$ as eigenphases.

The BQMEA enables one to estimate any linear combination of the unknown phases, $\mathbf{n}\cdot\boldsymbol{\theta}$, with a Heisenberg-limited precision, $\Delta(\mathbf{n}\cdot\boldsymbol{\theta}) \propto 1/N_T$. Furthermore, due to correlations that build up in the Bayesian distribution during the estimation, certain linear combinations can be estimated with better precision than sequential multiphase estimation

protocols, i.e., protocols, that estimate each of the d phases separately using a single-phase estimation protocol. We have also shown that the single-phase version of the BQMEA performs similar to different current single-phase estimation strategies. Finally, we have extended the BQMEA to a multiphase estimation protocol that is noise tolerant in the presence of phase damping noise.

While the BQMEA is described in the standard quantum-phase-estimation-algorithm language of quantum circuits and controlled unitary gates, we have also discussed two possible quantum optical implementations. In the standard multiphase estimation scheme of quantum metrology, the BQMEA can be implemented by using generalized multimode N00N states [454, 461, 462, 463]. Here, the total number of resources corresponds to the number of particles that have traversed the interferometric setup, and the crucial quantum resource of the quantum metrological advantage is multipartite entanglement. Furthermore, a single-photon multipass version of the BQMEA can be implemented with standard state-of-the-art optical elements [479, 493, 494] and is thus in reach of current experimental setups.

The Bayesian post-processing of the BQMEA makes use of a d-dimensional grid and is thus only efficient for estimation protocols that estimate a small number of phases. If a larger number of phases is to be measured, new post-processing strategies should be considered. For instance, one can approximate the Bayesian posterior distribution during the algorithm with a multivariate normal distribution [448, 468], or one can adaptively update the grid points by means of resampling algorithms and Monte–Carlo methods [480, 484, 485].

The BQMEA is implementable in current experimental platforms. However, any implementation is accompanied with different types of noise that most probably differ from the simple phase damping noise model considered above. Thus, for any experimental realization of the BQMEA, an adaptation of the algorithm to the experiment-specific noise conditions is required.

Chapter 6

Conclusions

The central objects of study in this thesis have been quantum resources. Quantum resources are specific features and characteristics of quantum systems that are not present in classical systems, and that enable an advantage of quantum technologies over classical ones for different technological tasks, leading to the so-called quantum advantage. We have touched upon several key questions that arise in the discussion of quantum resources, and we have attempted to contribute to partial answers to these questions.

Identification

The first question about quantum resources, that we have addressed in Ch. 2, is how to identify the key features of quantum systems that are responsible for specific quantum advantages. In this chapter, we have focused on the field of quantum computation and we have seen that the question of which quantum resource is necessary for a quantum advantage is highly dependent (i) on the particular computational problem that we have to solve and (ii) on the specific model of quantum computation that we employ. Due to these obstacles, the role of different quantum resources for a quantum computational advantage is still highly disputed.

We have then focused on a specific quantum algorithm, Grover's search algorithm. As one of the first and most important quantum algorithms, Grover's algorithm achieves a provable quadratic speed-up in terms of query complexity for the task of searching a large database. Inspired from the field of quantum metrology, we have then examined different quantum resources that are captured by so-called quantum statistical speeds. We have found that in the noiseless version and several noisy generalizations of Grover's algorithm, the maximal trace speed that occurs during the algorithm represents a necessary resource for the quantum advantage. The trace speed can further be interpreted as an indicator of entanglement or coherence.

Our results demonstrate that, similar to the role of the quantum Fisher information in quantum metrology, quantum statistical speeds open a new perspective to the identifi-

cation of central quantum resources in quantum computational advantage. In the future, it would be interesting to further generalize our results to more general noise models in search algorithms, or to other quantum algorithms that offer a provable quantum advantage.

Verification

The second question that we have addressed in Ch. 3 and Ch. 4 is the question of how to verify the presence of specific quantum resources from experimental data. This verification is crucial in order to confirm the correct operation of experimental setups to provide the quantum resources, such that the setups can be further used to create quantum technologies. Alas, the methods needed for the verification vary significantly for different quantum resources.

In Ch. 3, we have focused on the verification of (genuine multipartite) Bell nonlocality. Nonlocality is a central quantum resource for several quantum technologies, most prominently in the field of quantum cryptography. To detect nonlocality, the observed experimental data needs to violate the famous Bell inequalities. Here, we address the commonly used practise of postselecting the experimental data in order to purify or amplify nonlocal correlations. When performed incautiously, this postselection can distort the data via the selection bias, such that the verification of nonlocality with the postselected data is invalid. By using the tool box of causal inference and causal diagrams, we have provided valid postselection strategies for the verification of genuine multipartite nonlocality. Intriguingly, we have found that even postselection methods that require partial collaboration between the measurement parties can be used to verify genuine multipartite nonlocality. Finally, the result is applied to the creation of genuine three-partite nonlocality from independent particle sources.

The most direct direction for future research is the inclusion of finite detection probabilities and noise, such that the results can be applied to real experimental. Furthermore, there have been new approaches to nonlocality recently, such as, e.g., network nonlocality, and it would be interesting to check for similar results about valid postselection strategies in these new scenarios.

In Ch. 4, we have discussed the verification of nonclassicality (defined as the negativity of the Glauber P function). This quantum resource of a single-mode continuous-variable quantum system is central for several quantum technologies that build on continuous-variable quantum systems. The verification of nonlocality is known to require a significant amount of experimental data and large post-processing efforts. Here, we have developed a neural-network nonclassicality indicator that is trained via supervised learning from simulated experimental data of different standard quantum-optical measurement schemes. In particular, we have considered the common detection methods of balanced homodyne de-

tection and multiplexed click-counting measurements for several typical quantum optical states. We have found the neural-network prediction to perform well even for small sample sizes of experimental data, and to partially generalize beyond the states that are used in the training phase. Finally, for homodyne detection, we have confirmed the correct functioning of the neural-network approach on data from real experiments with squeezed, coherent, and single-photon-added coherent states.

The neural-network prediction does not represent a complete verification of nonclassicality, for which one has to rely on proven nonclassicality conditions that include an analysis of statistical errors. The neural network however serves a fast and easy-to-implement nonclassicality indicator that can be used even during the data acquisition. In the future, one could consider further experimental setups such as, e.g., multimode continuous-variable systems, and one could strengthen the performance of the neural networks by exploiting advanced machine learning techniques.

Application

Finally, in Ch. 5, we have addressed the third (and most common) question of how to employ quantum resources to build new quantum technologies. For this purpose, we have entered the field of quantum metrology that employs quantum resources such as entanglement and coherence to achieve a high precision in measurement procedures, a precision that is inaccessible for classical measurement techniques. We have focused on the task of the simultaneous estimation of multiple parameters when taking into account the total number of physical resources (e.g., the total number of particles) used in the estimation protocol. For this task, we have developed the Bayesian quantum multiphase estimation algorithm that combines different measurements of a basic quantum circuit by means of Bayesian post-processing techniques. We have found that the algorithm scales with the optimal precision scaling (the Heisenberg scaling) and that, for certain estimation tasks, it outperforms sequential estimation protocols that rely on optimized single-phase estimation subroutines. Furthermore, the algorithm can be made noise resilient and can be implemented in state-of-the-art experimental setups with standard quantum-optical elements.

In the future, to apply the multiphase estimation algorithm to specific experiments, one should consider the specific type of noise that is present in the given experimental setup to adapt the algorithm accordingly. Furthermore, inspired by the importance of the single-phase estimation algorithm in quantum computation, one could search for new quantum algorithms that use the multiphase estimation algorithm as a subroutine.

Appendix A

Measurement distributions in homodyne detection

In this appendix, we provide a detailed description of how to calculate the homodyne-detection outcome distributions $p(x)$ for several classes of quantum optical states that were used in training and testing the neural-network nonclassicality identifiers in Sec. 4.3.

Calculation of quadrature distributions

As discussed in Sec. 4.3.1, when preparing the local oscillator in an intense coherent state with relative phase $-(\phi + \pi/2)$ (with respect to the state of interest ρ), the homodyne detection scheme shown in Fig. 4.2 results in a measurement of $\langle I^- \rangle \propto \langle x(\phi) \rangle$, where $x(\phi) = a_1 e^{i\phi} + a_1^\dagger e^{-i\phi}$ is the phase-rotated quadrature operator. The corresponding quadrature distribution $p(x, \phi)$ is given by the marginal of the state's Wigner function W along a line through the origin with angle ϕ, defined as $x \sin\phi - y \cos\phi = 0$ [318]. For instance, choosing $\phi = 0$ results in the marginal of the Wigner function along the x-axis (i.e., the y-direction is integrated over).

To include a finite detection efficiency η in the simulation of the quadrature distributions $p(x)$, we recall the characteristic function of the state ρ that is given by [497]

$$\Phi(\beta) = \mathrm{tr}\left[\rho e^{\beta a^\dagger} e^{-\beta^* a}\right], \tag{A.1}$$

representing the complex Fourier transform of the Glauber–Sudarshan P function. A finite detection inefficiency η can be modeled as a simple beam splitter (BS) [318], where the input mode a_{in} is mixed with a loss mode a_{loss} (that is prepared in the vacuum state) to yield the output mode $a_{\mathrm{out}} = \sqrt{\eta} a_{\mathrm{in}} + \sqrt{1-\eta} a_{\mathrm{loss}}$. This noise model results in a finite-efficiency-including characteristic function $\Phi_\eta(\beta) = \Phi(\sqrt{\eta}\beta)$ [497], by means of which one can calculate the quadrature distributions in the presence of noise. In particular, the quadrature distribution is obtained from the characteristic function as [318]

$$p(x, \phi) = \frac{1}{\pi} \int e^{-2ixy - y^2/2} \Phi(i\sqrt{\eta} y e^{-i\phi}) \mathrm{d}y. \tag{A.2}$$

State	Probability $p(x,\phi)$
coherent	$\sqrt{\frac{2}{\pi}}e^{-2(x-\sqrt{\eta}\alpha\cos\phi)^2}$
thermal	$\sqrt{\frac{2}{\pi(1+2\eta\bar{n})}}e^{-\frac{2x^2}{1+2\eta\bar{n}}}$
Fock	$\sqrt{\frac{2}{\pi}}\sum_{k=0}^{n}\binom{n}{k}\frac{\eta^k}{2^k k!}e^{-2x^2}H_{2k}(\sqrt{2}x)$
squeezed-coherent	$\sqrt{\frac{2}{\pi[1-\eta(1-e^{2\xi}\cos^2\phi/2-e^{-2\xi}\sin^2\phi/2)]}}e^{-\frac{2(x-\sqrt{\eta}\alpha\cos\phi)^2}{1-\eta(1-e^{2\xi}\cos^2\phi/2-e^{-2\xi}\sin^2\phi/2)}}$
SPACS	$\frac{1}{1+\alpha^2}\sqrt{\frac{2}{\pi}}e^{-2(x-\sqrt{\eta}\alpha\cos\phi)^2}$
	$\times\left[\eta\left(2x\cos\phi-\frac{2\eta-1}{\sqrt{\eta}}\alpha\right)^2+4\eta x^2\sin^2\phi+(1-\eta)(1+4\eta\alpha^2\sin^2\phi)\right]$
cat	$\sqrt{\frac{2}{\pi}}\frac{1}{2-2e^{-2\alpha^2}}\Big\{e^{-2(x-\sqrt{\eta}\alpha\cos\phi)^2}$
	$-2e^{-2\alpha^2}\,\mathrm{Re}[e^{-2(x+i\sqrt{\eta}\alpha\sin\phi)^2}]+e^{-2(x+\sqrt{\eta}\alpha\cos\phi)^2}\Big\}$

Table A.1: The homodyne-detection quadrature distributions $p(x,\phi)$ are for each of the different classes of simulated states, where x is the measured quadrature value and ϕ is the quadrature angle that is used in the balanced homodyne detection measurement. η is the detection efficiency, α is the coherent amplitude, $\bar{n}$ is the mean photon number, n is the photon number, and ξ is the squeezing parameter.

The final quadrature probabilities for the different states that are used in the simulation of homodyne measurement data are summarized in Tab. A.1. Below, we provide exemplary derivations for how to calculate the distributions $p(x,\phi)$ for coherent, thermal, and Fock states.

Coherent states

For coherent states $|\alpha\rangle$ ($\alpha\in\mathbb{R}$[1]), the characteristic function (including the detection efficiency η) is directly obtained as $\Phi(\beta)=e^{\sqrt{\eta}\alpha(\beta-\beta^*)}=e^{2i\sqrt{\eta}\alpha\,\mathrm{Im}[\beta]}$. By Eq. (A.2), this yields

$$p(x,\phi)=\frac{1}{\pi}\int e^{-2ixy-y^2/2}e^{2i\sqrt{\eta}\alpha\,\mathrm{Im}[iye^{-i\phi}]}\mathrm{d}y \tag{A.3}$$

$$=\frac{1}{\pi}\int e^{-2iy(x-\sqrt{\eta}\alpha\cos\phi)-y^2/2}\mathrm{d}y \tag{A.4}$$

$$=\sqrt{\frac{2}{\pi}}e^{-2(x-\sqrt{\eta}\alpha\cos\phi)^2}. \tag{A.5}$$

Note that we use unit of our quadrature data that results in a vacuum quadrature variance of $\mathrm{Var}[x]=1/4$.

[1]Note that taking real amplitudes does not restrict generality because a coherent state with amplitude $|\alpha|e^{i\phi_\alpha}$ measured along the quadrature angle ϕ results in the same quadrature distribution as a coherent state with amplitude $|\alpha|$ measured along the quadrature angle $\phi-\phi_\alpha$.

Thermal states

For thermal states, the characteristic function is given by $\Phi(\beta) = e^{-\eta\bar{n}|\beta|^2}$ [498], such that we find

$$p(x,\phi) = \frac{1}{\pi}\int e^{-2ixy-y^2/2}e^{-\eta\bar{n}y^2}\mathrm{d}y \tag{A.6}$$

$$= \frac{1}{\pi}\int e^{-2ixy-y^2/2(1+2\eta\bar{n})}\mathrm{d}y \tag{A.7}$$

$$= \sqrt{\frac{2}{\pi(1+2\eta\bar{n})}}e^{-\frac{2x^2}{1+2\eta\bar{n}}}. \tag{A.8}$$

Fock states

The characteristic function of a Fock state $|n\rangle$ is given by $\Phi(\beta) = \mathrm{L}_n(\eta|\beta|^2)$ [498], where L_n is the nth Laguerre polynomial. By Eq. (A.2), one then obtains

$$p(x,\phi) = \frac{1}{\pi}\int e^{-2ixy-y^2/2}\,\mathrm{L}_n(\eta y^2)\mathrm{d}y \tag{A.9}$$

$$= \frac{1}{\pi}\sum_{k=0}^{n}\binom{n}{k}\frac{(-1)^k}{k!}\eta^k\int e^{-2ixy-y^2/2}y^{2k}\mathrm{d}y \tag{A.10}$$

$$= \sqrt{\frac{2}{\pi}}\sum_{k=0}^{n}\binom{n}{k}\frac{\eta^k}{2^k k!}e^{-2x^2}\,\mathrm{H}_{2k}(\sqrt{2}x), \tag{A.11}$$

where H_{2k} is the Hermite polynomial of degree $2k$ and in the third line, we have used the relation [499]

$$\int y^n e^{-py^2-qy}\mathrm{d}y = \left(\frac{i}{2}\right)^n\sqrt{\pi}p^{-\frac{n+1}{2}}e^{\frac{q^2}{4p}}\,\mathrm{H}_n(iq/2\sqrt{p}) \tag{A.12}$$

for $\mathrm{Re}\,[p] > 0$.

Squeezed states

For the squeezed-coherent states $D(\alpha)S(\xi)\,|0\rangle$ (recall the squeezing operator $S(\xi) = e^{(\xi{a^\dagger}^2+\xi^*a^2)/2}$ and the displacement operator $D(\alpha) = e^{\alpha a^\dagger-\alpha^*a}$), we first calculate the quadrature distribution of squeezed vacuum $S(\xi)\,|0\rangle$ for $\xi > 0$. Using the Baker–Campbell–Hausdorff formula, one can derive [318]

$$S^\dagger(\xi)D(\beta)S(\xi) = D(\beta') \tag{A.13}$$

with $\beta' = \beta(e^\xi + e^{-\xi})/2 - \beta^*(e^\xi - e^{-\xi})/2$. Then, after using $e^{\beta a^\dagger}e^{-\beta^*a} = D(\beta)e^{|\beta|^2/2}$, one finds after a few lines of calculation

$$\Phi(\beta) = \mathrm{tr}\left[\rho e^{\beta a^\dagger}e^{-\beta^*a}\right] \tag{A.14}$$

$$= e^{|\beta|^2/2}\,\langle 0|\,S^\dagger(\xi)D(\beta)S(\xi)\,|0\rangle \tag{A.15}$$

$$= e^{|\beta|^2/2+(\beta-\beta^*)^2e^{2\xi}/8-(\beta+\beta^*)^2e^{-2\xi}/8}. \tag{A.16}$$

Next, using Eq. (A.2), we obtain the quadrature distribution of a squeezed vacuum state as

$$p(x,\phi) = \frac{1}{\pi}\int e^{-2ixy-y^2/2}\Phi(i\sqrt{\eta}ye^{-i\phi})\mathrm{d}y \tag{A.17}$$

$$= \frac{1}{\pi}\int e^{-2ixy-y^2/2(1-\eta(1-e^{2\xi}\cos^2\phi/2-e^{-2\xi}\sin^2\phi/2)}\mathrm{d}y \tag{A.18}$$

$$= \sqrt{\frac{2}{\pi[1-\eta(1-e^{2\xi}\cos^2\phi/2-e^{-2\xi}\sin^2\phi/2)]}}e^{-\frac{2x^2}{1-\eta(1-e^{2\xi}\cos^2\phi/2-e^{-2\xi}\sin^2\phi/2)}}. \tag{A.19}$$

Finally, to obtain the quadrature distribution of a squeezed-coherent state $D(\alpha)S(\xi)\left|0\right\rangle$ with $\xi > 0$ and $\alpha \in \mathbb{R}$, we simply replace x with $x - \sqrt{\eta}\alpha\cos\phi$, see the calculation for coherent states above, resulting in a distribution as shown in Tab. A.1.

SPACS and cat states

The quadrature distribution of a SPACS is given in Ref. [400], and the quadrature distribution of a cat state can be computed similarly to the one of coherent states (here, the characteristic function consists of four summands that each resemble the one of a coherent state).

State parameters for simulations

In the simulation of the training and test data for Sec. 4.3, we fix a quadrature angle $\phi = 0$. This setting is still general: Fock states and thermal states are phase-independent, such that their quadrature distributions are independent of ϕ. Furthermore, the quadrature distribution of a coherent state $\left|\alpha\right\rangle$ ($\alpha \in \mathbb{R}$) along a quadrature angle ϕ coincides with the quadrature distribution of a coherent state $\left|\alpha\cos\phi\right\rangle$ measurement along $\phi = 0$ (the same holds for cat states). Finally, for single-photon added coherent states (SPACS) and squeezed states, the choice of $\phi = 0$ assures that the optimal quadrature angle is chosen to uncover the nonclassical features of the corresponding states. For squeezed states, measuring along $\phi = 0$ means that we measure the squeezed states along the optimal quadrature angle, i.e., the angle for which the quadrature distribution has the smallest (sub-shot-noise) variance. As discussed in Ch. 4, measuring squeezed states along other quadrature angles results in distributions that coincide with the ones produced by coherent states, or even broader distributions in the anti-squeezing regime.

As discussed in Ch. 4, we restrict the parameters of each state during the simulation to a range such that $P(|x > 8|) < 10^{-6}$. Furthermore, for squeezed states, we use random squeezing parameters $\xi \in [0.5, 1]$, where we use $\xi \geq 0.5$ because for small ξ, the distributions of squeezed and coherent state become to similar, resulting in a finite probability of a false positive nonclassicality classification of coherent states. Finally, for SPACS, the seed coherent amplitude is further restricted because, also here, for $|\alpha| > 3$, the distribution

State	Parameters
coherent (and mixtures)	$\alpha \in [-5, 5]$
thermal	$\bar{n} \in [0, 5]$
Fock	$n \in \{1, \ldots, 6\}$
squeezed-coherent	$\alpha \in [-5, 5], \xi \in [0.5, 1]$
SPACS	$\alpha \in [-3, 3]$

Table A.2: The parameter ranges of the different states that are used to simulate the training data. In the simulation, the parameters of each state are randomly sampled in the corresponding parameter range. α is the coherent amplitude, $\bar{n}$ is the mean photon number, n is the photon number, and ξ is the squeezing parameter.

becomes too similar to the one of a coherent state, see Ch. 4. The corresponding parameter ranges for the simulation of the different training states are summarized in Tab. A.2.

Appendix B

Measurement distributions in click-counting detection

In this appendix, we provide a detailed derivation of the multiplexed-click-counting measurement probabilities for the different classes of states that are used in the simulation of the training and test data of Sec. 4.4.

Calculation of click-counting distributions

The initial point of the calculation of the click-counting measurement distributions is the generalized binomial distribution [406] (that we recall from Sec. 4.4.1, see Eq. (4.16)),

$$p_k = \left\langle : \binom{N}{k} \left(e^{-\eta n/N}\right)^{N-k} \left(1 - e^{-\eta n/N}\right)^k : \right\rangle , \tag{B.1}$$

where p_k is the probability that the click-counting measurement results in k detector clicks, N is the number of measurement modes of the multiplexing scheme, η is the detection efficiency, and $n = a^\dagger a$ is the number operator of the incoming mode. In the following, we derive the explicit expression for p_k for the different classes of states that we have considered in Sec. 4.4.1. The results are summarized in Tab. B.1.

Coherent states

The measurement distribution for coherent states $|\alpha\rangle$ can be directly read from Eq. (B.1) because, due to the normal-ordering operation, one can insert $a \mapsto \alpha$, resulting in

$$p_k = \binom{N}{k} \left(e^{-\eta|\alpha|^2/N}\right)^{N-k} \left(1 - e^{-\eta|\alpha|^2/N}\right)^k . \tag{B.2}$$

State	Probability p_k
coherent	$\binom{N}{k}\left(e^{-\eta\lvert\alpha\rvert^2/N}\right)^{N-k}\left(1-e^{-\eta\lvert\alpha\rvert^2/N}\right)^{k}$
thermal	$\frac{1}{n_{\mathrm{th}}+1}\sum_{j=0}^{\infty}\left(\frac{n_{\mathrm{th}}}{n_{\mathrm{th}}+1}\right)^{j}\mathcal{D}^{\eta}_{k,j}$
Fock	$\mathcal{D}^{\eta}_{k,n}$, cf. Eq. (B.9)
squeezed	$\frac{1}{\cosh\lvert\xi\rvert}\sum_{n=0}^{\infty}\left(\frac{\tanh\lvert\xi\rvert}{2}\right)^{2n}\frac{(2n)!}{(n!)^2}\mathcal{D}^{\eta}_{k,2n}$
NPATS	$\frac{1}{(n_{\mathrm{th}}+1)n_{\mathrm{th}}^{n}}\sum_{j=n}^{\infty}\binom{j}{n}\left(\frac{n_{\mathrm{th}}}{n_{\mathrm{th}}+1}\right)^{j}\mathcal{D}^{\eta}_{k,j}$
even coherent	$\binom{N}{k}\sum_{j=0}^{k}\binom{k}{j}(-1)^j g(N-k+j,\alpha)$

Table B.1: Multiplexed click-counting probability p_k for the different states as a function of the state-specific parameters. N is the number of on-off detectors (i.e., the number of detection modes) in the multiplexing device and η is the detection efficiency of each detector.

Fock states

The calculation of the click counting statistics for a Fock state $|m\rangle$ is provided in Ref. [500]. We first factor out the operator in Eq. (B.1) to find

$$
\begin{aligned}
&:\binom{N}{k}\left(e^{-\eta n/N}\right)^{N-k}\left(1-e^{-\eta n/N}\right)^{k}: && \text{(B.3)}\\
=&\binom{N}{k}\sum_{j=0}^{k}\binom{k}{j}(-1)^{k-j}:\left(e^{-\eta n/N}\right)^{N-k}\left(e^{-\eta n/N}\right)^{k-j}: && \text{(B.4)}\\
=&\binom{N}{k}\sum_{j=0}^{k}\binom{k}{j}(-1)^{k-j}:\left(e^{-\eta n(1-j/N)}\right):. && \text{(B.5)}
\end{aligned}
$$

Next, using that any operator of the form $:e^{-\gamma n}:$ results in

$$
\begin{aligned}
\langle m|:e^{-\gamma n}:|m\rangle &= \langle m|\sum_{k=0}^{\infty}\frac{(-\gamma)^k}{k!}(a^\dagger)^k a^k|m\rangle && \text{(B.6)}\\
&= \sum_{k=0}^{m}\frac{(-\gamma)^k}{k!}\frac{m!}{(m-k)!} && \text{(B.7)}\\
&= (1-\gamma)^m, && \text{(B.8)}
\end{aligned}
$$

we obtain the click counting probabilities

$$
p_k \equiv \mathcal{D}^{\eta}_{k,m} = \binom{N}{k}\sum_{j=0}^{k}\binom{k}{j}(-1)^{k-j}\left(1-\eta\frac{N-j}{N}\right)^{m}, \qquad \text{(B.9)}
$$

where we have defined the symbol $\mathcal{D}^{\eta}_{k,m}$ that we will use below. Note that $\mathcal{D}^{\eta}_{k,m}$ corresponds to the symbol $\mathcal{D}^{1-\eta,\eta}_{k,m}$ in Ref. [500].

Squeezed vacuum states

For squeezed vacuum states $|\xi\rangle$, we use the state's expansion in the Fock basis given by [406]

$$|\xi\rangle = \frac{1}{\sqrt{\cosh \xi}} \sum_{n=0}^{\infty} \left(\frac{\tanh \xi}{2}\right)^n \frac{\sqrt{(2n)!}}{n!} |2n\rangle \,, \tag{B.10}$$

to obtain, by Eq. (B.9),

$$p_k = \frac{1}{\cosh |\xi|} \sum_{n=0}^{\infty} \left(\frac{\tanh |\xi|}{2}\right)^{2n} \frac{(2n)!}{(n!)^2} \mathcal{D}^{\eta}_{k,2n}. \tag{B.11}$$

Note that, in the simulations, we use a sufficiently large cut-off to approximate the infinite sum.

Thermal states and NPATS

Thermal states with average photon number n_{th} are defined as [498]

$$\rho_{\mathrm{th}} = \frac{1}{1+n_{\mathrm{th}}} \sum_{j=0}^{\infty} \left(\frac{n_{\mathrm{th}}}{1+n_{\mathrm{th}}}\right)^j |j\rangle \langle j| \,, \tag{B.12}$$

and n-photon-added thermal states (NPATS) are given as [498]

$$\rho_{\mathrm{th}+n} = \mathcal{N}(a^\dagger)^n \rho_{\mathrm{th}} a^n = \frac{1}{(1+n_{\mathrm{th}}) n_{\mathrm{th}}^n} \sum_{j=n}^{\infty} \binom{j}{n} \left(\frac{n_{\mathrm{th}}}{1+n_{\mathrm{th}}}\right)^j |j\rangle \langle j| \,, \tag{B.13}$$

where $\mathcal{N}$ is a normalization constant. Using Eq. (B.9), we find the click-counting probabilities for n-photon-added thermal states (for thermal states, insert $n = 0$) as

$$p_k = \frac{1}{(1+n_{\mathrm{th}}) n_{\mathrm{th}}^n} \sum_{j=n}^{\infty} \binom{j}{n} \left(\frac{n_{\mathrm{th}}}{1+n_{\mathrm{th}}}\right)^j D^{\eta}_{k,j}. \tag{B.14}$$

Again, in the simulation, we use a sufficiently high cut-off to approximate the infinite sum.

Even coherent states

Finally, the click-counting probabilities for even coherent states $|\alpha_+\rangle = \tilde{\mathcal{N}}(|\alpha\rangle + |-\alpha\rangle)$ ($\tilde{\mathcal{N}}$ is a normalization constant) can be derived directly from Eq. B.1. Defining [418]

$$g(\lambda, \alpha) = \langle \alpha_+| : e^{-\lambda n/N} : |\alpha_+\rangle = \frac{e^{-\lambda|\alpha|^2/N} + e^{-(\lambda/N-2)|\alpha|^2}}{1 + e^{-2|\alpha|^2}}, \tag{B.15}$$

we obtain

$$p_k = \binom{N}{k} \sum_{j=0}^{k} \binom{k}{j} (-1)^j g(N-k+j, \alpha). \tag{B.16}$$

State	Average number $\bar{n}$
coherent	$\lvert\alpha\rvert^2$
thermal	n_{th}
Fock	n
squeezed	$\sinh^2 \lvert\xi\rvert$
NPATS	$n_{\mathrm{th}}(n+1)+n$ [501]
even coherent	$\lvert\alpha\rvert^2 \frac{1-e^{-2\lvert\alpha\rvert^2}}{1+e^{-2\lvert\alpha\rvert^2}}$

Table B.2: Mean photon numbers $\bar{n}$ for the different states as a function of the state-specific parameters. N is the number of on-off detectors (i.e., the number of detection modes) in the multiplexing device and η is the quantum efficiency of each detector.

Parameter ranges of the simulation

As discussed in Sec. 4.4.2, we simulate all states from parameters that correspond to a mean photon number in the interval $\bar{n} \in [1, 16]$ (16 because we use a multiplexing into $N = 16$ measurement modes). The dependence of $\bar{n}$ on the state's parameters is shown in Tab. B.2.

Bibliography

[1] V. Gebhart, L. Pezzè and A. Smerzi. *Quantifying computational advantage of Grover's algorithm with the trace speed*. Sci. Rep. **11**, 1 (2021). (pages iii, 11, 29).

[2] V. Gebhart, L. Pezzè and A. Smerzi. *Genuine Multipartite Nonlocality with Causal-Diagram Postselection*. Phys. Rev. Lett. **127**, 140401 (2021). (pages iii, 11, 60, 61, 80).

[3] V. Gebhart, M. Bohmann, K. Weiher, N. Biagi, A. Zavatta, M. Bellini and E. Agudelo. *Identifying nonclassicality from experimental data using artificial neural networks*. Phys. Rev. Research **3**, 023229 (2021). (pages iii, 12, 102, 108, 110, 111, 112, 114, 115).

[4] V. Gebhart and M. Bohmann. *Neural-network approach for identifying nonclassicality from click-counting data*. Phys. Rev. Research **2**, 023150 (2020). (pages iii, 12, 115, 119).

[5] V. Gebhart, A. Smerzi and L. Pezzè. *Bayesian Quantum Multiphase Estimation Algorithm*. Phys. Rev. Applied **16**, 014035 (2021). (pages iii, 12, 142, 148, 149, 153, 154, 155, 156, 161, 163).

[6] J. P. Dowling and G. J. Milburn. *Quantum technology: the second quantum revolution*. Philos. Trans. R. Soc. London, Ser. A **361**, 1655 (2003). (page 1).

[7] A. Einstein, B. Podolsky and N. Rosen. *Can Quantum-Mechanical Description of Physical Reality Be Considered Complete?* Phys. Rev. **47**, 777 (1935). (pages 2, 9, 52).

[8] J. S. Bell. *On the Einstein Podolsky Rosen paradox*. Phys. Phys. Fiz. **1**, 195 (1964). (pages 2, 9, 44).

[9] J. S. Bell. *The theory of local beables*. Epistemol. Lett. **9**, 11 (1976). (pages 2, 9, 44).

[10] S. Wiesner. *Conjugate Coding*. SIGACT News **15**, 78–88 (1983). (page 2).

[11] C. H. Bennett and G. Brassard. *Quantum cryptography: Public key distribution and coin tossing*. Theor. Comput. Sci. **560**, 7 (2014). (page 2).

[12] A. K. Ekert. *Quantum cryptography based on Bell's theorem*. Phys. Rev. Lett. **67**, 661 (1991). (pages 2, 9, 23, 52).

[13] R. P. Feynman. *Simulating physics with computers*. Int. J. Theor. Phys. **21** (1982). (page 2).

[14] D. Deutsch. *Quantum theory, the Church–Turing principle and the universal quantum computer*. Proc. R. Soc. London, Ser. A **400**, 97 (1985). (page 2).

[15] E. Bernstein and U. Vazirani. *Quantum Complexity Theory*. SIAM J. Comput. **26**, 1411 (1997). (pages 2, 4, 15, 17, 18, 41).

[16] P. W. Shor. *Algorithms for quantum computation: discrete logarithms and factoring*. In *Proceedings of the 35th Annual Symposium on Foundations of Computer Science*, p. 124. IEEE, Santa Fe (1994). (pages 2, 5, 14, 15, 41, 136).

[17] C. M. Caves. *Quantum-mechanical noise in an interferometer*. Phys. Rev. D **23**, 1693 (1981). (pages 2, 130).

[18] N. Gisin, G. Ribordy, W. Tittel and H. Zbinden. *Quantum cryptography*. Rev. Mod. Phys. **74**, 145 (2002). (pages 2, 9).

[19] M. A. Nielsen and I. L. Chuang. *Quantum Computation and Quantum Information*. Cambridge University Press, Cambridge (2010). (pages 2, 3, 4, 5, 9, 16, 19, 21, 23, 25, 35, 134, 136, 137, 142, 143, 151, 159).

[20] C. L. Degen, F. Reinhard and P. Cappellaro. *Quantum sensing*. Rev. Mod. Phys. **89**, 035002 (2017). (page 2).

[21] L. Pezzè, A. Smerzi, M. K. Oberthaler, R. Schmied and P. Treutlein. *Quantum metrology with nonclassical states of atomic ensembles*. Rev. Mod. Phys. **90**, 035005 (2018). (pages 2, 4, 137).

[22] I. M. Georgescu, S. Ashhab and F. Nori. *Quantum simulation*. Rev. Mod. Phys. **86**, 153 (2014). (page 2).

[23] R. Horodecki, P. Horodecki, M. Horodecki and K. Horodecki. *Quantum entanglement*. Rev. Mod. Phys. **81**, 865 (2009). (pages 4, 7, 8, 15, 49, 91).

[24] N. Brunner, D. Cavalcanti, S. Pironio, V. Scarani and S. Wehner. *Bell nonlocality*. Rev. Mod. Phys. **86**, 419 (2014). (pages 4, 9, 44, 45, 46, 47, 52, 55, 66).

[25] C. H. Bennett and S. J. Wiesner. *Communication via one- and two-particle operators on Einstein-Podolsky-Rosen states*. Phys. Rev. Lett. **69**, 2881 (1992). (pages 4, 50).

[26] C. H. Bennett, G. Brassard, C. Crépeau, R. Jozsa, A. Peres and W. K. Wootters. *Teleporting an unknown quantum state via dual classical and Einstein-Podolsky-Rosen channels*. Phys. Rev. Lett. **70**, 1895 (1993). (pages 4, 50, 51, 90).

[27] L. Pezzé and A. Smerzi. *Entanglement, Nonlinear Dynamics, and the Heisenberg Limit*. Phys. Rev. Lett. **102**, 100401 (2009). (pages 4, 6, 7, 20, 38, 131, 141).

[28] P. Hyllus, W. Laskowski, R. Krischek, C. Schwemmer, W. Wieczorek, H. Weinfurter, L. Pezzé and A. Smerzi. *Fisher information and multiparticle entanglement*. Phys. Rev. A **85**, 022321 (2012). (pages 4, 7, 20, 32, 38, 131, 141).

[29] G. Tóth. *Multipartite entanglement and high-precision metrology*. Phys. Rev. A **85**, 022322 (2012). (pages 4, 7, 20, 38, 131, 141).

[30] G. Tóth and I. Apellaniz. *Quantum metrology from a quantum information science perspective*. J. Phys. A **47**, 424006 (2014). (pages 4, 137).

[31] R. Jozsa and N. Linden. *On the role of entanglement in quantum-computational speed-up*. Proc. R. Soc. London, Ser. A **459**, 2011 (2003). (pages 4, 16).

[32] V. Vedral. *The elusive source of quantum speedup*. Found. Phys. **40**, 1141 (2010). (pages 4, 6, 17, 19, 35, 41).

[33] A. Datta, S. T. Flammia and C. M. Caves. *Entanglement and the power of one qubit*. Phys. Rev. A **72**, 042316 (2005). (pages 4, 17, 18).

[34] L. K. Grover. *Quantum Mechanics Helps in Searching for a Needle in a Haystack*. Phys. Rev. Lett. **79**, 325 (1997). (pages 5, 7, 11, 23).

[35] S. Chakraborty, E. Fischer, A. Matsliah and R. de Wolf. *New Results on Quantum Property Testing*. In *Annual Conference on Foundations of Software Technology and Theoretical Computer Science (FSTTCS)*, p. 145. Schloss Dagstuhl – Leibniz-Zentrum fuer Informatik, Dagstuhl (2010). (page 5).

[36] V. Giovannetti, S. Lloyd and L. Maccone. *Quantum-Enhanced Measurements: Beating the Standard Quantum Limit*. Science **306**, 1330 (2004). (pages 6, 131).

[37] V. Giovannetti, S. Lloyd and L. Maccone. *Quantum Metrology*. Phys. Rev. Lett. **96**, 010401 (2006). (pages 6, 131, 141).

[38] B. L. Higgins, D. W. Berry, S. D. Bartlett, H. M. Wiseman and G. J. Pryde. *Entanglement-free Heisenberg-limited phase estimation*. Nature **450**, 393 (2007). (pages 6, 137, 138, 147, 148, 156, 163).

[39] J. P. Dowling. *Quantum optical metrology – the lowdown on high-N00N states*. Contemp. Phys. **49**, 125 (2008). (pages 6, 141).

[40] R. Cleve, A. Ekert, C. Macchiavello and M. Mosca. *Quantum algorithms revisited*. Proc. R. Soc. London, Ser. A **454**, 339 (1998). (pages 6, 132, 134, 135, 136, 142, 151).

[41] H. Ollivier and W. H. Zurek. *Quantum Discord: A Measure of the Quantumness of Correlations*. Phys. Rev. Lett. **88**, 017901 (2001). (pages 6, 18).

[42] L. Henderson and V. Vedral. *Classical, quantum and total correlations*. J. Phys. A **34**, 6899 (2001). (page 6).

[43] K. Modi, A. Brodutch, H. Cable, T. Paterek and V. Vedral. *The classical-quantum boundary for correlations: Discord and related measures*. Rev. Mod. Phys. **84**, 1655 (2012). (page 6).

[44] W. Vogel and J. Sperling. *Unified quantification of nonclassicality and entanglement*. Phys. Rev. A **89**, 052302 (2014). (pages 7, 95).

[45] N. Killoran, F. E. S. Steinhoff and M. B. Plenio. *Converting Nonclassicality into Entanglement*. Phys. Rev. Lett. **116**, 080402 (2016). (pages 7, 95).

[46] C. W. Helstrom. *Quantum detection and estimation theory*. Academic Press, New York (1976). (pages 7, 20, 130, 138).

[47] A. Holevo. *Probabilistic and Statistical Aspects of Quantum Theory*. North-Holland, Amsterdam (1982). (pages 7, 20, 130).

[48] W. K. Wootters. *Statistical distance and Hilbert space*. Phys. Rev. D **23**, 357 (1981). (pages 7, 19, 20).

[49] S. L. Braunstein and C. M. Caves. *Statistical distance and the geometry of quantum states*. Phys. Rev. Lett. **72**, 3439 (1994). (pages 7, 20).

[50] M. Gessner and A. Smerzi. *Statistical speed of quantum states: Generalized quantum Fisher information and Schatten speed*. Phys. Rev. A **97**, 022109 (2018). (pages 7, 19, 20, 21, 22, 31, 32, 38).

[51] M. Horodecki and J. Oppenheim. *(Quantumness in the context of) Resource Theories*. Int. J. Mod. Phys. B **27**, 1345019 (2013). (page 7).

[52] M. B. Plenio and S. Virmani. *An introduction to entanglement measures*. Quantum Info. Comput. **7**, 1 (2007). (page 7).

[53] C. H. Bennett, G. Brassard, S. Popescu, B. Schumacher, J. A. Smolin and W. K. Wootters. *Purification of Noisy Entanglement and Faithful Teleportation via Noisy Channels*. Phys. Rev. Lett. **76**, 722 (1996). (page 7).

[54] D. Schmid, T. C. Fraser, R. Kunjwal, A. B. Sainz, E. Wolfe and R. W. Spekkens. *Why standard entanglement theory is inappropriate for the study of Bell scenarios*. arXiv preprint arXiv:2004.09194 (2020). (pages 7, 8, 49, 50, 53, 87).

[55] J. Aberg. *Quantifying superposition*. arXiv preprint quant-ph/0612146 (2006). (page 8).

[56] T. Baumgratz, M. Cramer and M. B. Plenio. *Quantifying Coherence*. Phys. Rev. Lett. **113**, 140401 (2014). (page 8).

[57] A. Streltsov, G. Adesso and M. B. Plenio. *Colloquium: Quantum coherence as a resource*. Rev. Mod. Phys. **89**, 041003 (2017). (pages 8, 18, 21, 29, 38).

[58] G. Gour and R. W. Spekkens. *The resource theory of quantum reference frames: manipulations and monotones*. New J. Phys. **10**, 033023 (2008). (page 8).

[59] G. Gour, I. Marvian and R. W. Spekkens. *Measuring the quality of a quantum reference frame: The relative entropy of frameness*. Phys. Rev. A **80**, 012307 (2009). (page 8).

[60] M. Horodecki, P. Horodecki and J. Oppenheim. *Reversible transformations from pure to mixed states and the unique measure of information*. Phys. Rev. A **67**, 062104 (2003). (page 8).

[61] B. Yadin, F. C. Binder, J. Thompson, V. Narasimhachar, M. Gu and M. S. Kim. *Operational Resource Theory of Continuous-Variable Nonclassicality*. Phys. Rev. X **8**, 041038 (2018). (pages 8, 95).

[62] E. Chitambar and G. Gour. *Quantum resource theories*. Rev. Mod. Phys. **91**, 025001 (2019). (page 8).

[63] A. Peres. *Separability Criterion for Density Matrices*. Phys. Rev. Lett. **77**, 1413 (1996). (page 8).

[64] L. Gurvits. *Classical Deterministic Complexity of Edmonds' Problem and Quantum Entanglement*. In *Proceedings of the Thirty-Fifth Annual ACM Symposium on Theory of Computing*, p. 10. ACM, New York (2003). (pages 8, 46).

[65] F. Mintert, M. Kuś and A. Buchleitner. *Concurrence of Mixed Multipartite Quantum States*. Phys. Rev. Lett. **95**, 260502 (2005). (page 8).

[66] F. Mintert and A. Buchleitner. *Observable Entanglement Measure for Mixed Quantum States*. Phys. Rev. Lett. **98**, 140505 (2007). (page 8).

[67] C. Schmid, N. Kiesel, W. Wieczorek, H. Weinfurter, F. Mintert and A. Buchleitner. *Experimental Direct Observation of Mixed State Entanglement*. Phys. Rev. Lett. **101**, 260505 (2008). (page 8).

[68] F. G. S. L. Brandão. *Quantifying entanglement with witness operators*. Phys. Rev. A **72**, 022310 (2005). (page 8).

[69] H. M. Wiseman. *The two Bell's theorems of John Bell*. J. Phys. A **47**, 424001 (2014). (pages 9, 44, 45).

[70] E. Schrödinger. *Discussion of Probability Relations between Separated Systems*. Math. Proc. Camb. Philos. Soc. **31**, 555–563 (1935). (page 9).

[71] H. M. Wiseman, S. J. Jones and A. C. Doherty. *Steering, Entanglement, Nonlocality, and the Einstein-Podolsky-Rosen Paradox*. Phys. Rev. Lett. **98**, 140402 (2007). (pages 9, 10).

[72] R. Uola, A. C. S. Costa, H. C. Nguyen and O. Gühne. *Quantum steering*. Rev. Mod. Phys. **92**, 015001 (2020). (page 9).

[73] C. Branciard, E. G. Cavalcanti, S. P. Walborn, V. Scarani and H. M. Wiseman. *One-sided device-independent quantum key distribution: Security, feasibility, and the connection with steering*. Phys. Rev. A **85**, 010301 (2012). (page 9).

[74] R. F. Werner. *Quantum states with Einstein-Podolsky-Rosen correlations admitting a hidden-variable model*. Phys. Rev. A **40**, 4277 (1989). (page 9).

[75] J. Pearl. *Causality*. Cambridge University Press, Cambridge (2009). (pages 11, 62, 63, 64, 68, 69).

[76] B. Yurke and D. Stoler. *Einstein-Podolsky-Rosen effects from independent particle sources*. Phys. Rev. Lett. **68**, 1251 (1992). (pages 11, 79, 80, 82, 84, 87).

[77] S. Arora and B. Barak. *Computational Complexity: A Modern Approach*. Cambridge University Press, Cambridge (2009). (page 14).

[78] E. Tang. *A Quantum-Inspired Classical Algorithm for Recommendation Systems*. In *Proceedings of the 51st Annual ACM SIGACT Symposium on Theory of Computing*, p. 217. ACM, New York (2019). (page 14).

[79] L. Fortnow. *The Status of the P versus NP Problem*. Commun. ACM **52**, 78–86 (2009). (page 15).

[80] C. H. Bennett, E. Bernstein, G. Brassard and U. Vazirani. *Strengths and Weaknesses of Quantum Computing*. SIAM J. Comput. **26**, 1510 (1997). (pages 15, 28).

[81] C. Zalka. *Grover's quantum searching algorithm is optimal*. Phys. Rev. A **60**, 2746 (1999). (pages 15, 28).

[82] R. Raz and A. Tal. *Oracle Separation of BQP and PH*. In *Proceedings of the 51st Annual ACM SIGACT Symposium on Theory of Computing*, p. 13. ACM, New York (2019). (page 15).

[83] F. Arute, K. Arya, R. Babbush, D. Bacon, J. C. Bardin, R. Barends, R. Biswas, S. Boixo, F. G. Brandao, D. A. Buell et al. *Quantum supremacy using a programmable superconducting processor*. Nature **574**, 505 (2019). (pages 15, 35).

[84] H.-S. Zhong, H. Wang, Y.-H. Deng, M.-C. Chen, L.-C. Peng, Y.-H. Luo, J. Qin, D. Wu, X. Ding, Y. Hu, P. Hu, X.-Y. Yang, W.-J. Zhang, H. Li, Y. Li, X. Jiang, L. Gan, G. Yang, L. You, Z. Wang, L. Li, N.-L. Liu, C.-Y. Lu and J.-W. Pan. *Quantum computational advantage using photons*. Science **370**, 1460 (2020). (page 15).

[85] S. Aaronson and L. Chen. *Complexity-Theoretic Foundations of Quantum Supremacy Experiments*. In *32nd Computational Complexity Conference*, p. 22:1. Schloss Dagstuhl – Leibniz-Zentrum fuer Informatik, Dagstuhl (2017). (page 15).

[86] A. Bouland, B. Fefferman, C. Nirkhe and U. Vazirani. *On the complexity and verification of quantum random circuit sampling*. Nat. Phys. **15**, 159 (2019). (page 15).

[87] S. Aaronson and A. Arkhipov. *The Computational Complexity of Linear Optics*. In *Proceedings of the 43rd Annual Symposium on Theory of Computing*, p. 333. ACM, New York (2011). (page 15).

[88] N. Quesada, J. M. Arrazola and N. Killoran. *Gaussian boson sampling using threshold detectors*. Phys. Rev. A **98**, 062322 (2018). (page 15).

[89] R. Jozsa. *Entanglement and Quantum Computation*. arXiv preprint quant-ph/9707034 (1997). (page 15).

[90] D. Deutsch and R. Jozsa. *Rapid solution of problems by quantum computation*. Proc. R. Soc. London, Ser. A **439**, 553 (1992). (pages 15, 41).

[91] D. R. Simon. *On the Power of Quantum Computation*. SIAM J. Comput. **26**, 1474 (1997). (pages 15, 41).

[92] A. Ekert and R. Jozsa. *Quantum algorithms: entanglement-enhanced information processing*. Philos. Trans. R. Soc. A **356**, 1769 (1998). (pages 16, 38).

[93] N. Linden and S. Popescu. *Good Dynamics versus Bad Kinematics: Is Entanglement Needed for Quantum Computation?* Phys. Rev. Lett. **87**, 047901 (2001). (page 16).

[94] D. G. Cory, A. F. Fahmy and T. F. Havel. *Ensemble quantum computing by NMR spectroscopy*. Proc. Natl. Acad. Sci. U.S.A. **94**, 1634 (1997). (pages 16, 30).

[95] I. L. Chuang, L. M. Vandersypen, X. Zhou, D. W. Leung and S. Lloyd. *Experimental realization of a quantum algorithm*. Nature **393**, 143 (1998). (pages 16, 30).

[96] J. A. Jones, M. Mosca and R. H. Hansen. *Implementation of a quantum search algorithm on a quantum computer*. Nature **393**, 344 (1998). (pages 16, 30).

[97] S. Bravyi and A. Kitaev. *Universal quantum computation with ideal Clifford gates and noisy ancillas*. Phys. Rev. A **71**, 022316 (2005). (pages 16, 19, 95).

[98] G. Vidal. *Efficient Classical Simulation of Slightly Entangled Quantum Computations*. Phys. Rev. Lett. **91**, 147902 (2003). (page 16).

[99] R. Orús and J. I. Latorre. *Universality of entanglement and quantum-computation complexity*. Phys. Rev. A **69**, 052308 (2004). (page 16).

[100] E. Farhi, J. Goldstone, S. Gutmann and M. Sipser. *Quantum computation by adiabatic evolution*. arXiv preprint quant-ph/0001106 (2000). (page 16).

[101] E. Farhi, J. Goldstone, S. Gutmann, J. Lapan, A. Lundgren and D. Preda. *A Quantum Adiabatic Evolution Algorithm Applied to Random Instances of an NP-Complete Problem*. Science **292**, 472 (2001). (page 16).

[102] A. Wojcik and R. W. Chhajlany. *Quantum-classical correspondence in the oracle model of computation*. Int. J. Quantum Inf. **04**, 633 (2006). (page 17).

[103] D. A. Meyer. *Sophisticated Quantum Search Without Entanglement*. Phys. Rev. Lett. **85**, 2014 (2000). (pages 17, 23).

[104] D. Kenigsberg, T. Mor and G. Ratsaby. *Quantum advantage without entanglement*. Quantum Info. Comput. **6**, 606 (2006). (page 17).

[105] E. Biham, G. Brassard, D. Kenigsberg and T. Mor. *Quantum computing without entanglement*. Theor. Comput. Sci. **320**, 15 (2004). (page 17).

[106] E. Knill and R. Laflamme. *Power of One Bit of Quantum Information*. Phys. Rev. Lett. **81**, 5672 (1998). (pages 17, 18, 138).

[107] A. Datta, A. Shaji and C. M. Caves. *Quantum Discord and the Power of One Qubit*. Phys. Rev. Lett. **100**, 050502 (2008). (pages 17, 18).

[108] M. Van den Nest. *Universal Quantum Computation with Little Entanglement*. Phys. Rev. Lett. **110**, 060504 (2013). (page 17).

[109] R. Raussendorf and H. J. Briegel. *A One-Way Quantum Computer*. Phys. Rev. Lett. **86**, 5188 (2001). (pages 17, 53).

[110] D. Gross, S. T. Flammia and J. Eisert. *Most Quantum States Are Too Entangled To Be Useful As Computational Resources*. Phys. Rev. Lett. **102**, 190501 (2009). (page 17).

[111] D. Gottesman. *The Heisenberg representation of quantum computers*. In *22nd International Colloquium on Group Theoretical Methods in Physics*, p. 32. International Press, Cambridge (1998). (page 17).

[112] A. Ferraro, L. Aolita, D. Cavalcanti, F. M. Cucchietti and A. Acín. *Almost all quantum states have nonclassical correlations*. Phys. Rev. A **81**, 052318 (2010). (page 18).

[113] B. Dakić, V. Vedral and C. Brukner. *Necessary and Sufficient Condition for Nonzero Quantum Discord*. Phys. Rev. Lett. **105**, 190502 (2010). (page 18).

[114] M. Yoganathan and C. Cade. *The one clean qubit model without entanglement is classically simulable*. arXiv preprint arXiv:1907.08224 (2019). (page 18).

[115] M. Hillery. *Coherence as a resource in decision problems: The Deutsch-Jozsa algorithm and a variation*. Phys. Rev. A **93**, 012111 (2016). (page 18).

[116] J. Ma, B. Yadin, D. Girolami, V. Vedral and M. Gu. *Converting Coherence to Quantum Correlations*. Phys. Rev. Lett. **116**, 160407 (2016). (page 18).

[117] J. M. Matera, D. Egloff, N. Killoran and M. B. Plenio. *Coherent control of quantum systems as a resource theory*. Quantum Sci. Technol. **1**, 01LT01 (2016). (page 18).

[118] N. Anand and A. K. Pati. *Coherence and entanglement monogamy in the discrete analogue of analog Grover search*. arXiv preprint arXiv:1611.04542 (2016). (pages 18, 29, 38).

[119] H.-L. Shi, S.-Y. Liu, X.-H. Wang, W.-L. Yang, Z.-Y. Yang and H. Fan. *Coherence depletion in the Grover quantum search algorithm*. Phys. Rev. A **95**, 032307 (2017). (pages 18, 29, 38).

[120] M. Pan and D. Qiu. *Operator coherence dynamics in Grover's quantum search algorithm*. Phys. Rev. A **100**, 012349 (2019). (pages 18, 29).

[121] S. Bravyi, D. Gosset and R. König. *Quantum advantage with shallow circuits*. Science **362**, 308 (2018). (pages 18, 54).

[122] M. Howard, J. Wallman, V. Veitch and J. Emerson. *Contextuality supplies the 'magic' for quantum computation*. Nature **510**, 351 (2014). (page 19).

[123] S. Kochen and E. Specker. *The Problem of Hidden Variables in Quantum Mechanics*. J. Math. Mech. **17**, 59 (1967). (page 19).

[124] J. S. Bell. *On the Problem of Hidden Variables in Quantum Mechanics*. Rev. Mod. Phys. **38**, 447 (1966). (page 19).

[125] D. Gross. *Hudson's theorem for finite-dimensional quantum systems*. J. Math. Phys. **47**, 122107 (2006). (pages 19, 95).

[126] A. Mari and J. Eisert. *Positive Wigner Functions Render Classical Simulation of Quantum Computation Efficient*. Phys. Rev. Lett. **109**, 230503 (2012). (pages 19, 95).

[127] V. Veitch, C. Ferrie, D. Gross and J. Emerson. *Negative quasi-probability as a resource for quantum computation*. New J. Phys. **14**, 113011 (2012). (pages 19, 95).

[128] D. Stahlke. *Quantum interference as a resource for quantum speedup*. Phys. Rev. A **90**, 022302 (2014). (page 19).

[129] G. Chiribella, G. M. D'Ariano, P. Perinotti and B. Valiron. *Quantum computations without definite causal structure*. Phys. Rev. A **88**, 022318 (2013). (page 19).

[130] Y. Cai, H. N. Le and V. Scarani. *State complexity and quantum computation*. Annalen der Physik **527**, 684 (2015). (page 19).

[131] D. Petz. *Monotone metrics on matrix spaces*. Linear Algebra Its Appl. **244**, 81 (1996). (page 19).

[132] D. Spehner. *Quantum correlations and distinguishability of quantum states*. J. Math. Phys. **55**, 075211 (2014). (page 19).

[133] E. Hellinger. *Neue Begründung der Theorie quadratischer Formen von unendlichvielen Veränderlichen.* J. Reine Angew. Math. **1909**, 210 (1909). (page 20).

[134] L. Pezzè and A. Smerzi. *Quantum theory of phase estimation*. arXiv preprint arXiv:1411.5164 (2014). (pages 20, 40, 130, 132, 151).

[135] S. Szalay. *k-stretchability of entanglement, and the duality of k-separability and k-producibility*. Quantum **3**, 204 (2019). (page 21).

[136] Z. Ren, W. Li, A. Smerzi and M. Gessner. *Metrological Detection of Multipartite Entanglement from Young Diagrams*. Phys. Rev. Lett. **126**, 080502 (2021). (page 21).

[137] C. Fuchs and J. van de Graaf. *Cryptographic distinguishability measures for quantum-mechanical states*. IEEE Trans. Inf. Theor. **45**, 1216 (1999). (page 21).

[138] I. Marvian and R. W. Spekkens. *Extending Noether's theorem by quantifying the asymmetry of quantum states*. Nat. Commun. **5**, 1 (2014). (pages 21, 38).

[139] M. Oszmaniec, R. Augusiak, C. Gogolin, J. Kołodyński, A. Acín and M. Lewenstein. *Random Bosonic States for Robust Quantum Metrology*. Phys. Rev. X **6**, 041044 (2016). (page 21).

[140] M. Hayashi. *Quantum Information*. Springer, Berlin (2006). (page 21).

[141] H. Strobel, W. Muessel, D. Linnemann, T. Zibold, D. B. Hume, L. Pezzè, A. Smerzi and M. K. Oberthaler. *Fisher information and entanglement of non-Gaussian spin states*. Science **345**, 424 (2014). (pages 21, 130).

[142] L. Pezzè, Y. Li, W. Li and A. Smerzi. *Witnessing entanglement without entanglement witness operators*. Proc. Natl. Acad. Sci. U.S.A. **113**, 11459 (2016). (page 21).

[143] S. Lloyd. *Quantum search without entanglement*. Phys. Rev. A **61**, 010301 (1999). (pages 23, 38).

[144] J. Ahn, T. C. Weinacht and P. H. Bucksbaum. *Information Storage and Retrieval Through Quantum Phase*. Science **287**, 463 (2000). (page 23).

[145] N. Bhattacharya, H. B. van Linden van den Heuvell and R. J. C. Spreeuw. *Implementation of Quantum Search Algorithm using Classical Fourier Optics*. Phys. Rev. Lett. **88**, 137901 (2002). (page 23).

[146] C. M. Caves, I. H. Deutsch and R. Blume-Kohout. *Physical-resource requirements and the power of quantum computation*. J. Opt. B: Quantum Semiclass. Opt. **6**, S801 (2004). (page 23).

[147] G. Brassard, P. Høyer, M. Mosca and A. Tapp. *Quantum amplitude amplification and estimation*. In *Contemporary Mathematics Series Vol. 305*. ACM, Providence (2002). (page 23).

[148] G. Brassard, P. Høyer and A. Tapp. *Quantum counting*. In *Automata, Languages and Programming*, p. 820. Springer, Berlin (1998). (page 23).

[149] A. Gilyén, Y. Su, G. H. Low and N. Wiebe. *Quantum Singular Value Transformation and beyond: Exponential Improvements for Quantum Matrix Arithmetics*. In *Proceedings of the 51st Annual ACM SIGACT Symposium on Theory of Computing*, p. 193. ACM, New York (2019). (page 24).

[150] L. K. Grover. *A Framework for Fast Quantum Mechanical Algorithms*. In *Proceedings of the Thirtieth Annual ACM Symposium on Theory of Computing*, p. 53. ACM, New York (1998). (page 24).

[151] A. Ambainis. *Quantum Search Algorithms*. SIGACT News **35**, 22–35 (2004). (page 24).

[152] N. J. Cerf, L. K. Grover and C. P. Williams. *Nested quantum search and NP-hard problems*. Appl. Algebra Eng. Commun. Comput. **10**, 311 (2000). (page 24).

[153] A. Ambainis. *Quantum Walk Algorithm for Element Distinctness*. SIAM J. Comput. **37**, 210 (2007). (page 24).

[154] R. Babbush, J. R. McClean, M. Newman, C. Gidney, S. Boixo and H. Neven. *Focus beyond Quadratic Speedups for Error-Corrected Quantum Advantage*. PRX Quantum **2**, 010103 (2021). (page 24).

[155] M. Roget, S. Guillet, P. Arrighi and G. Di Molfetta. *Grover Search as a Naturally Occurring Phenomenon*. Phys. Rev. Lett. **124**, 180501 (2020). (page 24).

[156] A. Patel. *Quantum algorithms and the genetic code*. Pramana **56**, 367 (2001). (page 24).

[157] A. R. Brown. *Playing Pool with $|\psi\rangle$: from Bouncing Billiards to Quantum Search*. Quantum **4**, 357 (2020). (page 24).

[158] P. Kaye, R. Laflamme and M. Mosca. *An introduction to quantum computing*. Oxford University Press, Oxford (2007). (pages 25, 26).

[159] S. Braunstein and A. Pati. *Speed-up and entanglement in quantum searching*. Quantum Inf. Comput. **2**, 399 (2002). (pages 27, 29).

[160] A. Barenco, C. H. Bennett, R. Cleve, D. P. DiVincenzo, N. Margolus, P. Shor, T. Sleator, J. A. Smolin and H. Weinfurter. *Elementary gates for quantum computation*. Phys. Rev. A **52**, 3457 (1995). (page 28).

[161] D. Bruß and C. Macchiavello. *Multipartite entanglement in quantum algorithms*. Phys. Rev. A **83**, 052313 (2011). (page 28).

[162] D. A. Meyer and N. R. Wallach. *Global entanglement in multiparticle systems*. J. Math. Phys. **43**, 4273 (2002). (pages 28, 34).

[163] Y. Fang, D. Kaszlikowski, C. Chin, K. Tay, L. Kwek and C. Oh. *Entanglement in the Grover search algorithm*. Phys. Lett. A **345**, 265 (2005). (pages 28, 34).

[164] W. K. Wootters. *Entanglement of Formation of an Arbitrary State of Two Qubits*. Phys. Rev. Lett. **80**, 2245 (1998). (page 28).

[165] P. Rungta. *The quadratic speedup in Grover's search algorithm from the entanglement perspective*. Phys. Lett. A **373**, 2652 (2009). (pages 28, 34).

[166] M. Rossi, D. Bruß and C. Macchiavello. *Scale invariance of entanglement dynamics in Grover's quantum search algorithm*. Phys. Rev. A **87**, 022331 (2013). (pages 29, 34).

[167] T.-C. Wei and P. M. Goldbart. *Geometric measure of entanglement and applications to bipartite and multipartite quantum states*. Phys. Rev. A **68**, 042307 (2003). (page 29).

[168] O. Biham, M. A. Nielsen and T. J. Osborne. *Entanglement monotone derived from Grover's algorithm*. Phys. Rev. A **65**, 062312 (2002). (page 29).

[169] B. Pablo-Norman and M. Ruiz-Altaba. *Noise in Grover's quantum search algorithm*. Phys. Rev. A **61**, 012301 (1999). (page 29).

[170] G. L. Long, Y. S. Li, W. L. Zhang and C. C. Tu. *Dominant gate imperfection in Grover's quantum search algorithm*. Phys. Rev. A **61**, 042305 (2000). (page 29).

[171] I. Cohn, A. L. F. De Oliveira, E. Buksman and J. G. L. De Lacalle. *Grover's search with local and total depolarizing channel errors: Complexity analysis*. Int. J. Quantum Inf. **14**, 1650009 (2016). (pages 29, 36, 39, 40).

[172] H. Azuma. *Decoherence in Grover's quantum algorithm: Perturbative approach*. Phys. Rev. A **65**, 042311 (2002). (page 29).

[173] D. Shapira, S. Mozes and O. Biham. *Effect of unitary noise on Grover's quantum search algorithm*. Phys. Rev. A **67**, 042301 (2003). (pages 29, 39).

[174] N. Shenvi, K. R. Brown and K. B. Whaley. *Effects of a random noisy oracle on search algorithm complexity*. Phys. Rev. A **68**, 052313 (2003). (pages 29, 39).

[175] O. Regev and L. Schiff. *Impossibility of a quantum speed-up with a faulty oracle*. In *International Colloquium on Automata, Languages, and Programming*, p. 773. Springer, Berlin (2008). (pages 29, 39).

[176] K. Temme. *Runtime of unstructured search with a faulty Hamiltonian oracle*. Phys. Rev. A **90**, 022310 (2014). (pages 29, 39).

[177] D. Reitzner and M. Hillery. *Grover search under localized dephasing*. Phys. Rev. A **99**, 012339 (2019). (page 29).

[178] P. Vrana, D. Reeb, D. Reitzner and M. M. Wolf. *Fault-ignorant quantum search*. New J. Phys. **16**, 073033 (2014). (pages 29, 39).

[179] S. L. Braunstein, C. M. Caves, R. Jozsa, N. Linden, S. Popescu and R. Schack. *Separability of Very Noisy Mixed States and Implications for NMR Quantum Computing*. Phys. Rev. Lett. **83**, 1054 (1999). (page 30).

[180] E. Biham and D. Kenigsberg. *Grover's quantum search algorithm for an arbitrary initial mixed state*. Phys. Rev. A **66**, 062301 (2002). (page 35).

[181] A. Kay. *Degree of quantum correlation required to speed up a computation*. Phys. Rev. A **92**, 062329 (2015). (pages 35, 38).

[182] M. M. Wilde. *Quantum Information Theory*. Cambridge University Press, Cambridge (2013). (page 35).

[183] A. Peres and P. F. Scudo. *Unspeakable quantum information*. arXiv preprint quant-ph/0201017 (2002). (page 39).

[184] S. D. Bartlett, T. Rudolph and R. W. Spekkens. *Reference frames, superselection rules, and quantum information*. Rev. Mod. Phys. **79**, 555 (2007). (page 39).

[185] I. Marvian and R. W. Spekkens. *How to quantify coherence: Distinguishing speakable and unspeakable notions*. Phys. Rev. A **94**, 052324 (2016). (page 39).

[186] J.-Å. Larsson. *Loopholes in Bell inequality tests of local realism*. J. Phys. A **47**, 424003 (2014). (pages 44, 55, 59).

[187] B. Schumacher and M. Westmoreland. *Quantum Processes, Systems, and Information*. Cambridge University Press, Cambridge (2010). (page 44).

[188] T. Maudlin. *Quantum Non-Locality and Relativity*. Wiley, New York (1994). (page 44).

[189] H. Reichenbach. *The Direction of Time*. University of California Press, Berkeley (1991). (page 44).

[190] J. F. Clauser, M. A. Horne, A. Shimony and R. A. Holt. *Proposed Experiment to Test Local Hidden-Variable Theories*. Phys. Rev. Lett. **23**, 880 (1969). (page 46).

[191] S. Pironio. *Lifting Bell inequalities*. J. Math. Phys. **46**, 062112 (2005). (pages 46, 57).

[192] D. Avis, H. Imai, T. Ito and Y. Sasaki. *Deriving tight Bell inequalities for 2 parties with many 2-valued observables from facets of cut polytopes*. arXiv preprint quant-ph/0404014 (2004). (page 46).

[193] N. D. Mermin. *Extreme quantum entanglement in a superposition of macroscopically distinct states*. Phys. Rev. Lett. **65**, 1838 (1990). (page 47).

[194] R. Gallego, L. E. Würflinger, A. Acín and M. Navascués. *Operational Framework for Nonlocality*. Phys. Rev. Lett. **109**, 070401 (2012). (pages 47, 48, 68).

[195] G. Svetlichny. *Distinguishing three-body from two-body nonseparability by a Bell-type inequality*. Phys. Rev. D **35**, 3066 (1987). (pages 47, 68, 71).

[196] M. L. Almeida, D. Cavalcanti, V. Scarani and A. Acín. *Multipartite fully nonlocal quantum states*. Phys. Rev. A **81**, 052111 (2010). (pages 48, 68).

[197] J.-D. Bancal, J. Barrett, N. Gisin and S. Pironio. *Definitions of multipartite nonlocality*. Phys. Rev. A **88**, 014102 (2013). (pages 48, 68, 70).

[198] P. H. Eberhard and R. R. Ross. *Quantum field theory cannot provide faster-than-light communication*. Found. Phys. **2**, 127 (1989). (pages 48, 52).

[199] M. Seevinck and J. Uffink. *Sufficient conditions for three-particle entanglement and their tests in recent experiments*. Phys. Rev. A **65**, 012107 (2001). (pages 48, 49).

[200] J.-D. Bancal, N. Gisin, Y.-C. Liang and S. Pironio. *Device-Independent Witnesses of Genuine Multipartite Entanglement*. Phys. Rev. Lett. **106**, 250404 (2011). (page 49).

[201] D. Collins, N. Gisin, S. Popescu, D. Roberts and V. Scarani. *Bell-Type Inequalities to Detect True n-Body Nonseparability*. Phys. Rev. Lett. **88**, 170405 (2002). (pages 49, 54).

[202] D. Cavalcanti, M. L. Almeida, V. Scarani and A. Acin. *Quantum networks reveal quantum nonlocality*. Nat. Commun. **2**, 1 (2011). (page 49).

[203] P. Contreras-Tejada, C. Palazuelos and J. I. de Vicente. *Genuine Multipartite Nonlocality Is Intrinsic to Quantum Networks*. Phys. Rev. Lett. **126**, 040501 (2021). (page 49).

[204] M. Navascués, E. Wolfe, D. Rosset and A. Pozas-Kerstjens. *Genuine Network Multipartite Entanglement*. Phys. Rev. Lett. **125**, 240505 (2020). (page 49).

[205] L. Hardy. *Nonlocality for two particles without inequalities for almost all entangled states*. Phys. Rev. Lett. **71**, 1665 (1993). (page 50).

[206] P. H. Eberhard. *Background level and counter efficiencies required for a loophole-free Einstein-Podolsky-Rosen experiment*. Phys. Rev. A **47**, R747 (1993). (pages 50, 57).

[207] A. Acín, S. Massar and S. Pironio. *Randomness versus Nonlocality and Entanglement*. Phys. Rev. Lett. **108**, 100402 (2012). (page 50).

[208] E. Wolfe, A. Pozas-Kerstjens, M. Grinberg, D. Rosset, A. Acín and M. Navascués. *Quantum Inflation: A General Approach to Quantum Causal Compatibility*. Phys. Rev. X **11**, 021043 (2021). (pages 50, 51, 54, 69).

[209] T. Fritz. *Beyond Bell's theorem: correlation scenarios*. New J. Phys. **14**, 103001 (2012). (pages 51, 65, 84).

[210] C. Branciard, N. Gisin and S. Pironio. *Characterizing the Nonlocal Correlations Created via Entanglement Swapping*. Phys. Rev. Lett. **104**, 170401 (2010). (page 51).

[211] A. Pozas-Kerstjens, R. Rabelo, L. Rudnicki, R. Chaves, D. Cavalcanti, M. Navascués and A. Acín. *Bounding the Sets of Classical and Quantum Correlations in Networks*. Phys. Rev. Lett. **123**, 140503 (2019). (page 51).

[212] M.-O. Renou, E. Bäumer, S. Boreiri, N. Brunner, N. Gisin and S. Beigi. *Genuine Quantum Nonlocality in the Triangle Network*. Phys. Rev. Lett. **123**, 140401 (2019). (pages 51, 84).

[213] P. Abiuso, T. Kriváchy, E.-C. Boghiu, M.-O. Renou, A. Pozas-Kerstjens and A. Acín. *Quantum networks reveal single-photon nonlocality*. arXiv preprint arXiv:2108.01726 (2021). (pages 51, 84).

[214] M. Ringbauer, C. Giarmatzi, R. Chaves, F. Costa, A. G. White and A. Fedrizzi. *Experimental test of nonlocal causality*. Sci. Adv. **2**, e1600162 (2016). (pages 51, 65).

[215] R. Chaves, G. Carvacho, I. Agresti, V. Di Giulio, L. Aolita, S. Giacomini and F. Sciarrino. *Quantum violation of an instrumental test*. Nat. Phys. **14**, 291 (2018). (pages 51, 65).

[216] G. Carvacho, F. Andreoli, L. Santodonato, M. Bentivegna, R. Chaves and F. Sciarrino. *Experimental violation of local causality in a quantum network*. Nat. Comm. **8**, 1 (2017). (page 51).

[217] I. Agresti, D. Poderini, L. Guerini, M. Mancusi, G. Carvacho, L. Aolita, D. Cavalcanti, R. Chaves and F. Sciarrino. *Experimental device-independent certified randomness generation with an instrumental causal structure*. Commun. Phys. **3**, 1 (2020). (pages 51, 65).

[218] D. Poderini, I. Agresti, G. Marchese, E. Polino, T. Giordani, A. Suprano, M. Valeri, G. Milani, N. Spagnolo, G. Carvacho et al. *Experimental violation of n-locality in a star quantum network*. Nat. Commun. **11**, 1 (2020). (page 51).

[219] M.-O. Renou, D. Trillo, M. Weilenmann, L. P. Thinh, A. Tavakoli, N. Gisin, A. Acin and M. Navascues. *Quantum physics needs complex numbers*. arXiv preprint arXiv:2101.10873 (2021). (page 51).

[220] H. J. Kimble. *The quantum internet*. Nature **453**, 1023 (2008). (pages 51, 54).

[221] S. Wehner, D. Elkouss and R. Hanson. *Quantum internet: A vision for the road ahead*. Science **362** (2018). (pages 51, 55).

[222] A. Tavakoli, A. Pozas-Kerstjens, M.-X. Luo and M.-O. Renou. *Bell nonlocality in networks*. arXiv preprint arXiv:2104.10700 (2021). (page 51).

[223] R. Cleve and H. Buhrman. *Substituting quantum entanglement for communication*. Phys. Rev. A **56**, 1201 (1997). (page 52).

[224] D. M. Greenberger, M. A. Horne, A. Shimony and A. Zeilinger. *Bell's theorem without inequalities*. Am. J. Phys. **58**, 1131 (1990). (pages 52, 53, 79, 82, 83, 84).

[225] N. D. Mermin. *Quantum mysteries revisited*. Am. J. Phys. **58**, 731 (1990). (pages 52, 83, 84).

[226] C. Brukner, M. Żukowski, J.-W. Pan and A. Zeilinger. *Bell's Inequalities and Quantum Communication Complexity*. Phys. Rev. Lett. **92**, 127901 (2004). (page 52).

[227] T. S. Cubitt, D. Leung, W. Matthews and A. Winter. *Zero-Error Channel Capacity and Simulation Assisted by Non-Local Correlations*. IEEE Trans. Inf. Theor. **57**, 5509 (2011). (page 52).

[228] D. Ebler, S. Salek and G. Chiribella. *Enhanced Communication with the Assistance of Indefinite Causal Order*. Phys. Rev. Lett. **120**, 120502 (2018). (page 52).

[229] G. Chiribella, M. Banik, S. S. Bhattacharya, T. Guha, M. Alimuddin, A. Roy, S. Saha, S. Agrawal and G. Kar. *Indefinite causal order enables perfect quantum communication with zero capacity channels*. New J. Phys. **23**, 033039 (2021). (page 52).

[230] H. Buhrman, R. Cleve, S. Massar and R. de Wolf. *Nonlocality and communication complexity*. Rev. Mod. Phys. **82**, 665 (2010). (page 52).

[231] W. K. Wootters and W. H. Zurek. *A single quantum cannot be cloned*. Nature **299**, 802 (1982). (page 52).

[232] D. Mayers and A. C. Yao. *Quantum Cryptography with Imperfect Apparatus*. In *Proceedings of the 39th Annual Symposium on Foundations of Computer Science*, p. 503. IEEE, Palo Alto (1998). (page 53).

[233] S. Popescu and D. Rohrlich. *Which states violate Bell's inequality maximally?* Phys. Lett. A **169**, 411 (1992). (page 53).

[234] M. McKague, T. H. Yang and V. Scarani. *Robust self-testing of the singlet*. J. Phys. A **45**, 455304 (2012). (page 53).

[235] J. Barrett, L. Hardy and A. Kent. *No Signaling and Quantum Key Distribution*. Phys. Rev. Lett. **95**, 010503 (2005). (page 53).

[236] A. Acín, N. Gisin and L. Masanes. *From Bell's Theorem to Secure Quantum Key Distribution*. Phys. Rev. Lett. **97**, 120405 (2006). (page 53).

[237] A. Acín, N. Brunner, N. Gisin, S. Massar, S. Pironio and V. Scarani. *Device-Independent Security of Quantum Cryptography against Collective Attacks*. Phys. Rev. Lett. **98**, 230501 (2007). (page 53).

[238] U. Vazirani and T. Vidick. *Fully Device-Independent Quantum Key Distribution*. Phys. Rev. Lett. **113**, 140501 (2014). (page 53).

[239] S. Pironio, A. Acín, S. Massar, A. B. de La Giroday, D. N. Matsukevich, P. Maunz, S. Olmschenk, D. Hayes, L. Luo, T. A. Manning et al. *Random numbers certified by Bell's theorem*. Nature **464**, 1021 (2010). (page 53).

[240] B. W. Reichardt, F. Unger and U. Vazirani. *A Classical Leash for a Quantum System: Command of Quantum Systems via Rigidity of CHSH Games*. In *Proceedings of the 4th Conference on Innovations in Theoretical Computer Science*, p. 321. ACM, New York (2013). (page 53).

[241] U. Mahadev. *Classical Verification of Quantum Computations*. In *Proceedings of the 59th Annual Symposium on Foundations of Computer Science*, p. 259. IEEE, Paris (2018). (page 53).

[242] R. Raussendorf, D. E. Browne and H. J. Briegel. *Measurement-based quantum computation on cluster states*. Phys. Rev. A **68**, 022312 (2003). (page 53).

[243] M. Hillery, V. Bužek and A. Berthiaume. *Quantum secret sharing*. Phys. Rev. A **59**, 1829 (1999). (page 54).

[244] R. Cleve, D. Gottesman and H.-K. Lo. *How to Share a Quantum Secret*. Phys. Rev. Lett. **83**, 648 (1999). (page 54).

[245] A. Cabello. *Multiparty key distribution and secret sharing based on entanglement swapping*. arXiv preprint quant-ph/0009025 (2000). (page 54).

[246] J. Ribeiro, G. Murta and S. Wehner. *Fully device-independent conference key agreement*. Phys. Rev. A **97**, 022307 (2018). (page 54).

[247] G. Carrara, H. Kampermann, D. Bruß and G. Murta. *Genuine multipartite entanglement is not a precondition for secure conference key agreement*. Phys. Rev. Research **3**, 013264 (2021). (page 54).

[248] M. Epping, H. Kampermann, C. macchiavello and D. Bruß. *Multi-partite entanglement can speed up quantum key distribution in networks*. New J. Phys. **19**, 093012 (2017). (page 54).

[249] M. Pivoluska, M. Huber and M. Malik. *Layered quantum key distribution*. Phys. Rev. A **97**, 032312 (2018). (page 54).

[250] G. Murta, F. Grasselli, H. Kampermann and D. Bruß. *Quantum Conference Key Agreement: A Review*. Adv. Quantum Technol. **3**, 2000025 (2020). (page 54).

[251] M. Żukowski, A. Zeilinger, M. A. Horne and A. K. Ekert. *"Event-ready-detectors" Bell experiment via entanglement swapping*. Phys. Rev. Lett. **71**, 4287 (1993). (page 54).

[252] S. Bose, V. Vedral and P. L. Knight. *Multiparticle generalization of entanglement swapping*. Phys. Rev. A **57**, 822 (1998). (page 54).

[253] A. I. Lvovsky, B. C. Sanders and W. Tittel. *Optical quantum memory*. Nat. Photonics **3**, 706 (2009). (page 54).

[254] T. Northup and R. Blatt. *Quantum information transfer using photons*. Nat. Photonics **8**, 356 (2014). (page 54).

[255] A. Niezgoda and J. Chwedeńczuk. *Many-Body Nonlocality as a Resource for Quantum-Enhanced Metrology*. Phys. Rev. Lett. **126**, 210506 (2021). (page 55).

[256] A. Aspect. *Proposed experiment to test the nonseparability of quantum mechanics*. Phys. Rev. D **14**, 1944 (1976). (page 56).

[257] T. Scheidl, R. Ursin, J. Kofler, S. Ramelow, X.-S. Ma, T. Herbst, L. Ratschbacher, A. Fedrizzi, N. K. Langford, T. Jennewein and A. Zeilinger. *Violation of local realism with freedom of choice*. Proc. Natl. Acad. Sci. U.S.A. **107**, 19708 (2010). (page 56).

[258] M. J. W. Hall. *Local Deterministic Model of Singlet State Correlations Based on Relaxing Measurement Independence*. Phys. Rev. Lett. **105**, 250404 (2010). (page 56).

[259] J. Barrett and N. Gisin. *How Much Measurement Independence Is Needed to Demonstrate Nonlocality?* Phys. Rev. Lett. **106**, 100406 (2011). (page 56).

[260] S. J. Freedman and J. F. Clauser. *Experimental Test of Local Hidden-Variable Theories*. Phys. Rev. Lett. **28**, 938 (1972). (page 56).

[261] A. Aspect, J. Dalibard and G. Roger. *Experimental Test of Bell's Inequalities Using Time-Varying Analyzers*. Phys. Rev. Lett. **49**, 1804 (1982). (page 56).

[262] G. Weihs, T. Jennewein, C. Simon, H. Weinfurter and A. Zeilinger. *Violation of Bell's Inequality under Strict Einstein Locality Conditions*. Phys. Rev. Lett. **81**, 5039 (1998). (page 56).

[263] B. B. T. Collaboration et al. *Challenging local realism with human choices*. Nature **557**, 212 (2018). (page 56).

[264] J. Handsteiner, A. S. Friedman, D. Rauch, J. Gallicchio, B. Liu, H. Hosp, J. Kofler, D. Bricher, M. Fink, C. Leung, A. Mark, H. T. Nguyen, I. Sanders, F. Steinlechner, R. Ursin, S. Wengerowsky, A. H. Guth, D. I. Kaiser, T. Scheidl and A. Zeilinger. *Cosmic Bell Test: Measurement Settings from Milky Way Stars*. Phys. Rev. Lett. **118**, 060401 (2017). (page 56).

[265] D. Rauch, J. Handsteiner, A. Hochrainer, J. Gallicchio, A. S. Friedman, C. Leung, B. Liu, L. Bulla, S. Ecker, F. Steinlechner, R. Ursin, B. Hu, D. Leon, C. Benn, A. Ghedina, M. Cecconi, A. H. Guth, D. I. Kaiser, T. Scheidl and A. Zeilinger. *Cosmic Bell Test Using Random Measurement Settings from High-Redshift Quasars*. Phys. Rev. Lett. **121**, 080403 (2018). (page 56).

[266] P. M. Pearle. *Hidden-Variable Example Based upon Data Rejection*. Phys. Rev. D **2**, 1418 (1970). (pages 56, 57).

[267] J. F. Clauser and M. A. Horne. *Experimental consequences of objective local theories*. Phys. Rev. D **10**, 526 (1974). (pages 56, 57).

[268] D. W. Berry, H. Jeong, M. Stobińska and T. C. Ralph. *Fair-sampling assumption is not necessary for testing local realism*. Phys. Rev. A **81**, 012109 (2010). (page 57).

[269] D. Orsucci, J.-D. Bancal, N. Sangouard and P. Sekatski. *How post-selection affects device-independent claims under the fair sampling assumption*. Quantum **4**, 238 (2020). (page 57).

[270] D. S. Tasca, S. P. Walborn, F. Toscano and P. H. Souto Ribeiro. *Observation of tunable Popescu-Rohrlich correlations through postselection of a Gaussian state*. Phys. Rev. A **80**, 030101 (2009). (page 57).

[271] I. Gerhardt, Q. Liu, A. Lamas-Linares, J. Skaar, V. Scarani, V. Makarov and C. Kurtsiefer. *Experimentally Faking the Violation of Bell's Inequalities*. Phys. Rev. Lett. **107**, 170404 (2011). (page 57).

[272] L. Lydersen, C. Wiechers, C. Wittmann, D. Elser, J. Skaar and V. Makarov. *Hacking commercial quantum cryptography systems by tailored bright illumination*. Nat. Photonics **4**, 686 (2010). (page 57).

[273] F. Sciarrino, G. Vallone, A. Cabello and P. Mataloni. *Bell experiments with random destination sources*. Phys. Rev. A **83**, 032112 (2011). (pages 57, 59, 61, 84).

[274] N. D. Mermin. *The EPR Experiment—Thoughts about the "Loophole"*. Ann. N. Y. Acad. Sci. **480**, 422 (1986). (page 57).

[275] T. Vértesi, S. Pironio and N. Brunner. *Closing the Detection Loophole in Bell Experiments Using Qudits*. Phys. Rev. Lett. **104**, 060401 (2010). (page 57).

[276] H. Buhrman, P. Høyer, S. Massar and H. Röhrig. *Combinatorics and Quantum Nonlocality*. Phys. Rev. Lett. **91**, 047903 (2003). (page 57).

[277] J.-A. Larsson. *Bell's inequality and detector inefficiency*. Phys. Rev. A **57**, 3304 (1998). (page 57).

[278] M. A. Rowe, D. Kielpinski, V. Meyer, C. A. Sackett, W. M. Itano, C. Monroe and D. J. Wineland. *Experimental violation of a Bell's inequality with efficient detection*. Nature **409**, 791 (2001). (page 57).

[279] D. N. Matsukevich, P. Maunz, D. L. Moehring, S. Olmschenk and C. Monroe. *Bell Inequality Violation with Two Remote Atomic Qubits*. Phys. Rev. Lett. **100**, 150404 (2008). (page 57).

[280] M. Giustina, A. Mech, S. Ramelow, B. Wittmann, J. Kofler, J. Beyer, A. Lita, B. Calkins, T. Gerrits, S. W. Nam et al. *Bell violation using entangled photons without the fair-sampling assumption*. Nature **497**, 227 (2013). (page 57).

[281] B. G. Christensen, K. T. McCusker, J. B. Altepeter, B. Calkins, T. Gerrits, A. E. Lita, A. Miller, L. K. Shalm, Y. Zhang, S. W. Nam, N. Brunner, C. C. W. Lim, N. Gisin and P. G. Kwiat. *Detection-Loophole-Free Test of Quantum Nonlocality, and Applications*. Phys. Rev. Lett. **111**, 130406 (2013). (page 57).

[282] L. K. Shalm, E. Meyer-Scott, B. G. Christensen, P. Bierhorst, M. A. Wayne, M. J. Stevens, T. Gerrits, S. Glancy, D. R. Hamel, M. S. Allman, K. J. Coakley, S. D. Dyer, C. Hodge, A. E. Lita, V. B. Verma, C. Lambrocco, E. Tortorici, A. L. Migdall, Y. Zhang, D. R. Kumor, W. H. Farr, F. Marsili, M. D. Shaw, J. A. Stern, C. Abellán, W. Amaya, V. Pruneri, T. Jennewein, M. W. Mitchell, P. G. Kwiat, J. C. Bienfang, R. P. Mirin, E. Knill and S. W. Nam. *Strong Loophole-Free Test of Local Realism*. Phys. Rev. Lett. **115**, 250402 (2015). (page 57).

[283] M. Giustina, M. A. M. Versteegh, S. Wengerowsky, J. Handsteiner, A. Hochrainer, K. Phelan, F. Steinlechner, J. Kofler, J.-A. Larsson, C. Abellán, W. Amaya, V. Pruneri, M. W. Mitchell, J. Beyer, T. Gerrits, A. E. Lita, L. K. Shalm, S. W. Nam,

T. Scheidl, R. Ursin, B. Wittmann and A. Zeilinger. *Significant-Loophole-Free Test of Bell's Theorem with Entangled Photons*. Phys. Rev. Lett. **115**, 250401 (2015). (page 57).

[284] B. Hensen, H. Bernien, A. E. Dréau, A. Reiserer, N. Kalb, M. S. Blok, J. Ruitenberg, R. F. Vermeulen, R. N. Schouten, C. Abellán et al. *Loophole-free Bell inequality violation using electron spins separated by 1.3 kilometres*. Nature **526**, 682 (2015). (page 58).

[285] J. Berkson. *Limitations of the Application of Fourfold Table Analysis to Hospital Data*. Biometrics Bulletin **2**, 47 (1946). (page 58).

[286] G. J. Griffith, T. T. Morris, M. J. Tudball, A. Herbert, G. Mancano, L. Pike, G. C. Sharp, J. Sterne, T. M. Palmer, G. D. Smith et al. *Collider bias undermines our understanding of COVID-19 disease risk and severity*. Nat. Commun. **11**, 1 (2020). (page 58).

[287] P. Blasiak, E. Borsuk and M. Markiewicz. *On safe post-selection for Bell nonlocality: Causal diagram approach*. arXiv preprint arXiv:2012.07285 (2020). (pages 59, 67, 70, 73, 78).

[288] J.-Å. Larsson and R. D. Gill. *Bell's inequality and the coincidence-time loophole*. EPL **67**, 707 (2004). (page 59).

[289] J. D. Franson. *Bell inequality for position and time*. Phys. Rev. Lett. **62**, 2205 (1989). (pages 59, 87).

[290] S. Aerts, P. Kwiat, J.-A. Larsson and M. Z ⊠kowski. *Two-Photon Franson-Type Experiments and Local Realism*. Phys. Rev. Lett. **83**, 2872 (1999). (pages 59, 78).

[291] A. Cabello, A. Rossi, G. Vallone, F. De Martini and P. Mataloni. *Proposed Bell Experiment with Genuine Energy-Time Entanglement*. Phys. Rev. Lett. **102**, 040401 (2009). (pages 59, 78).

[292] G. Lima, G. Vallone, A. Chiuri, A. Cabello and P. Mataloni. *Experimental Bell-inequality violation without the postselection loophole*. Phys. Rev. A **81**, 040101 (2010). (page 59).

[293] A. Cuevas, G. Carvacho, G. Saavedra, J. Cariñe, W. Nogueira, M. Figueroa, A. Cabello, P. Mataloni, G. Lima and G. Xavier. *Long-distance distribution of genuine energy-time entanglement*. Nat. Commun. **4**, 1 (2013). (page 59).

[294] B. Yurke and D. Stoler. *Bell's-inequality experiments using independent-particle sources*. Phys. Rev. A **46**, 2229 (1992). (page 59).

[295] Y. Aharonov, D. Z. Albert and L. Vaidman. *How the result of a measurement of a component of the spin of a spin-1/2 particle can turn out to be 100*. Phys. Rev. Lett. **60**, 1351 (1988). (pages 60, 96).

[296] J. Dressel, M. Malik, F. M. Miatto, A. N. Jordan and R. W. Boyd. *Colloquium: Understanding quantum weak values: Basics and applications*. Rev. Mod. Phys. **86**, 307 (2014). (pages 60, 96).

[297] A. N. Jordan, J. Martínez-Rincón and J. C. Howell. *Technical Advantages for Weak-Value Amplification: When Less Is More*. Phys. Rev. X **4**, 011031 (2014). (pages 60, 96).

[298] J. Harris, R. W. Boyd and J. S. Lundeen. *Weak Value Amplification Can Outperform Conventional Measurement in the Presence of Detector Saturation*. Phys. Rev. Lett. **118**, 070802 (2017). (pages 60, 96).

[299] F. Piacentini, A. Avella, M. P. Levi, M. Gramegna, G. Brida, I. P. Degiovanni, E. Cohen, R. Lussana, F. Villa, A. Tosi, F. Zappa and M. Genovese. *Measuring Incompatible Observables by Exploiting Sequential Weak Values*. Phys. Rev. Lett. **117**, 170402 (2016). (pages 60, 96).

[300] C. Ferrie and J. Combes. *Weak Value Amplification is Suboptimal for Estimation and Detection*. Phys. Rev. Lett. **112**, 040406 (2014). (page 60).

[301] J. Combes, C. Ferrie, Z. Jiang and C. M. Caves. *Quantum limits on postselected, probabilistic quantum metrology*. Phys. Rev. A **89**, 052117 (2014). (page 60).

[302] D. R. Arvidsson-Shukur, N. Y. Halpern, H. V. Lepage, A. A. Lasek, C. H. Barnes and S. Lloyd. *Quantum advantage in postselected metrology*. Nat. Commun. **11**, 1 (2020). (page 60).

[303] T. Verma and J. Pearl. *Equivalence and Synthesis of Causal Models*. In *Proceedings of the Sixth Annual Conference on Uncertainty in Artificial Intelligence*, p. 255. Elsevier Science, USA (1990). (pages 63, 64).

[304] C. J. Wood and R. W. Spekkens. *The lesson of causal discovery algorithms for quantum correlations: causal explanations of Bell-inequality violations require fine-tuning*. New J. Phys. **17**, 033002 (2015). (pages 65, 68).

[305] R. Chaves, R. Kueng, J. B. Brask and D. Gross. *Unifying Framework for Relaxations of the Causal Assumptions in Bell's Theorem*. Phys. Rev. Lett. **114**, 140403 (2015). (page 65).

[306] R. Chaves, G. Moreno, E. Polino, D. Poderini, I. Agresti, A. Suprano, M. R. Barros, G. Carvacho, E. Wolfe, A. Canabarro, R. W. Spekkens and F. Sciarrino. *Causal networks and freedom of choice in Bell's theorem*. arXiv preprint arXiv:2105.05721 (2021). (page 65).

[307] J.-M. A. Allen, J. Barrett, D. C. Horsman, C. M. Lee and R. W. Spekkens. *Quantum Common Causes and Quantum Causal Models*. Phys. Rev. X **7**, 031021 (2017). (pages 68, 69).

[308] G. Vallone, P. Mataloni and A. Cabello. *Multiparty multilevel energy-time entanglement*. Phys. Rev. A **81**, 032105 (2010). (pages 73, 78, 84).

[309] M. Żukowski. *Violations of local realism in multiphoton interference experiments*. Phys. Rev. A **61**, 022109 (2000). (page 73).

[310] J.-W. Pan, Z.-B. Chen, C.-Y. Lu, H. Weinfurter, A. Zeilinger and M. Żukowski. *Multiphoton entanglement and interferometry*. Rev. Mod. Phys. **84**, 777 (2012). (page 74).

[311] M. Krenn, A. Hochrainer, M. Lahiri and A. Zeilinger. *Entanglement by Path Identity*. Phys. Rev. Lett. **118**, 080401 (2017). (page 74).

[312] C. K. Hong, Z. Y. Ou and L. Mandel. *Measurement of subpicosecond time intervals between two photons by interference*. Phys. Rev. Lett. **59**, 2044 (1987). (page 82).

[313] J.-W. Pan, D. Bouwmeester, M. Daniell, H. Weinfurter and A. Zeilinger. *Experimental test of quantum nonlocality in three-photon Greenberger–Horne–Zeilinger entanglement*. Nature **403**, 515 (2000). (page 83).

[314] J. Jogenfors and J.-Å. Larsson. *Energy-time entanglement, elements of reality, and local realism*. J. Phys. A **47**, 424032 (2014). (page 84).

[315] T. Kriváchy, Y. Cai, D. Cavalcanti, A. Tavakoli, N. Gisin and N. Brunner. *A neural network oracle for quantum nonlocality problems in networks*. npj Quantum Inf. **6**, 1 (2020). (pages 85, 101).

[316] P. Mitchell, S. Popescu and D. Roberts. *Conditions for the confirmation of three-particle nonlocality*. Phys. Rev. A **70**, 060101 (2004). (page 85).

[317] J. Kołodyński, A. Máttar, P. Skrzypczyk, E. Woodhead, D. Cavalcanti, K. Banaszek and A. Acín. *Device-independent quantum key distribution with single-photon sources*. Quantum **4**, 260 (2020). (page 86).

[318] W. Vogel and D.-G. Welsch. *Quantum optics*. Wiley, New York (2006). (pages 90, 91, 92, 93, 103, 104, 107, 108, 110, 116, 171, 173).

[319] H.-W. Lee. *Theory and application of the quantum phase-space distribution functions*. Phys. Rep. **259**, 147 (1995). (pages 90, 93).

[320] S. L. Braunstein and P. van Loock. *Quantum information with continuous variables*. Rev. Mod. Phys. **77**, 513 (2005). (pages 90, 95).

[321] S. L. Braunstein and H. J. Kimble. *Teleportation of Continuous Quantum Variables*. Phys. Rev. Lett. **80**, 869 (1998). (page 90).

[322] F. Benatti, R. Floreanini, F. Franchini and U. Marzolino. *Entanglement in indistinguishable particle systems*. Phys. Rep. **878**, 1 (2020). Entanglement in indistinguishable particle systems. (page 91).

[323] B. Morris, B. Yadin, M. Fadel, T. Zibold, P. Treutlein and G. Adesso. *Entanglement between Identical Particles Is a Useful and Consistent Resource*. Phys. Rev. X **10**, 041012 (2020). (page 91).

[324] M. Tichy, F. de Melo, M. Kuś, F. Mintert and A. Buchleitner. *Entanglement of identical particles and the detection process*. Fortschritte der Physik **61**, 225 (2013). (page 91).

[325] K. E. Cahill and R. J. Glauber. *Ordered Expansions in Boson Amplitude Operators*. Phys. Rev. **177**, 1857 (1969). (page 92).

[326] K. E. Cahill and R. J. Glauber. *Density Operators and Quasiprobability Distributions*. Phys. Rev. **177**, 1882 (1969). (page 92).

[327] G. S. Agarwal and E. Wolf. *Calculus for Functions of Noncommuting Operators and General Phase-Space Methods in Quantum Mechanics. I. Mapping Theorems and Ordering of Functions of Noncommuting Operators*. Phys. Rev. D **2**, 2161 (1970). (page 92).

[328] G. S. Agarwal and E. Wolf. *Calculus for Functions of Noncommuting Operators and General Phase-Space Methods in Quantum Mechanics. II. Quantum Mechanics in Phase Space*. Phys. Rev. D **2**, 2187 (1970). (page 92).

[329] R. J. Glauber. *Photon Correlations*. Phys. Rev. Lett. **10**, 84 (1963). (page 92).

[330] E. C. G. Sudarshan. *Equivalence of Semiclassical and Quantum Mechanical Descriptions of Statistical Light Beams*. Phys. Rev. Lett. **10**, 277 (1963). (page 92).

[331] E. Wigner. *On the Quantum Correction For Thermodynamic Equilibrium*. Phys. Rev. **40**, 749 (1932). (page 93).

[332] C. Ferrie. *Quasi-probability representations of quantum theory with applications to quantum information science*. Rep. Prog. Phys. **74**, 116001 (2011). (page 93).

[333] K. Husimi. *Some Formal Properties of the Density Matrix*. Proceedings of the Physico-Mathematical Society of Japan. 3rd Series **22**, 264 (1940). (page 93).

[334] E. Schrödinger. *Der stetige Übergang von der Mikro-zur Makromechanik*. Naturwissenschaften **14**, 664 (1926). (page 94).

[335] F. Shahandeh, A. P. Lund and T. C. Ralph. *Quantum correlations and global coherence in distributed quantum computing*. Phys. Rev. A **99**, 052303 (2019). (page 95).

[336] F. Shahandeh, A. P. Lund and T. C. Ralph. *Quantum Correlations in Nonlocal Boson Sampling*. Phys. Rev. Lett. **119**, 120502 (2017). (page 95).

[337] H. Kwon, K. C. Tan, T. Volkoff and H. Jeong. *Nonclassicality as a Quantifiable Resource for Quantum Metrology*. Phys. Rev. Lett. **122**, 040503 (2019). (page 95).

[338] C. Weedbrook, S. Pirandola, R. García-Patrón, N. J. Cerf, T. C. Ralph, J. H. Shapiro and S. Lloyd. *Gaussian quantum information*. Rev. Mod. Phys. **84**, 621 (2012). (page 95).

[339] H. Grote, K. Danzmann, K. L. Dooley, R. Schnabel, J. Slutsky and H. Vahlbruch. *First Long-Term Application of Squeezed States of Light in a Gravitational-Wave Observatory*. Phys. Rev. Lett. **110**, 181101 (2013). (page 95).

[340] M. Tse *et al.* *Quantum-Enhanced Advanced LIGO Detectors in the Era of Gravitational-Wave Astronomy*. Phys. Rev. Lett. **123**, 231107 (2019). (page 95).

[341] M. S. Kim, W. Son, V. Bužek and P. L. Knight. *Entanglement by a beam splitter: Nonclassicality as a prerequisite for entanglement*. Phys. Rev. A **65**, 032323 (2002). (page 95).

[342] S.-W. Lee, H. Jeong and D. Jaksch. *Testing quantum nonlocality by generalized quasiprobability functions*. Phys. Rev. A **80**, 022104 (2009). (page 95).

[343] R. W. Spekkens. *Negativity and Contextuality are Equivalent Notions of Nonclassicality*. Phys. Rev. Lett. **101**, 020401 (2008). (page 95).

[344] S. Rahimi-Keshari, M. Mehboudi, D. De Santis, D. Cavalcanti and A. Acín. *Verification of joint measurability using phase-space quasiprobability distributions*. Phys. Rev. A **104**, 042212 (2021). (page 96).

[345] J. G. Kirkwood. *Quantum Statistics of Almost Classical Assemblies*. Phys. Rev. **44**, 31 (1933). (page 96).

[346] P. A. M. Dirac. *On the Analogy Between Classical and Quantum Mechanics*. Rev. Mod. Phys. **17**, 195 (1945). (page 96).

[347] A. M. Steinberg. *Conditional probabilities in quantum theory and the tunneling-time controversy*. Phys. Rev. A **52**, 32 (1995). (page 96).

[348] J. S. Lundeen, B. Sutherland, A. Patel, C. Stewart and C. Bamber. *Direct measurement of the quantum wavefunction*. Nature **474**, 188 (2011). (page 96).

[349] J. S. Lundeen and C. Bamber. *Procedure for Direct Measurement of General Quantum States Using Weak Measurement*. Phys. Rev. Lett. **108**, 070402 (2012). (page 96).

[350] G. S. Thekkadath, L. Giner, Y. Chalich, M. J. Horton, J. Banker and J. S. Lundeen. *Direct Measurement of the Density Matrix of a Quantum System*. Phys. Rev. Lett. **117**, 120401 (2016). (page 96).

[351] N. Yunger Halpern, B. Swingle and J. Dressel. *Quasiprobability behind the out-of-time-ordered correlator*. Phys. Rev. A **97**, 042105 (2018). (page 96).

[352] V. Dunjko and H. J. Briegel. *Machine learning & artificial intelligence in the quantum domain: a review of recent progress*. Rep. Prog. Phys. **81**, 074001 (2018). (pages 96, 99).

[353] M. Schuld, I. Sinayskiy and F. Petruccione. *An introduction to quantum machine learning*. Contemp. Phys. **56**, 172 (2015). (page 96).

[354] J. Biamonte, P. Wittek, N. Pancotti, P. Rebentrost, N. Wiebe and S. Lloyd. *Quantum machine learning*. Nature **549**, 195 (2017). (page 96).

[355] M. A. Nielsen. *Neural networks and deep learning*. Determination Press, San Francisco (2015). (pages 96, 97, 98, 99, 106).

[356] P. Mehta, M. Bukov, C.-H. Wang, A. G. Day, C. Richardson, C. K. Fisher and D. J. Schwab. *A high-bias, low-variance introduction to machine learning for physicists*. Phys. Rep. **810**, 1 (2019). (pages 96, 99).

[357] K. Hornik, M. Stinchcombe and H. White. *Multilayer feedforward networks are universal approximators*. Neural Netw. **2**, 359 (1989). (page 97).

[358] P. Rebentrost, M. Mohseni and S. Lloyd. *Quantum Support Vector Machine for Big Data Classification*. Phys. Rev. Lett. **113**, 130503 (2014). (page 99).

[359] G. Torlai, G. Mazzola, J. Carrasquilla, M. Troyer, R. Melko and G. Carleo. *Neural-network quantum state tomography*. Nat. Phys. **14**, 447–450 (2018). (page 100).

[360] F. Verstraete, V. Murg and J. Cirac. *Matrix product states, projected entangled pair states, and variational renormalization group methods for quantum spin systems*. Adv. Phys. **57**, 143 (2008). (page 100).

[361] G. Carleo and M. Troyer. *Solving the quantum many-body problem with artificial neural networks*. Science **355**, 602 (2017). (page 100).

[362] E. S. Tiunov, V. V. T. (Vyborova), A. E. Ulanov, A. I. Lvovsky and A. K. Fedorov. *Experimental quantum homodyne tomography via machine learning*. Optica **7**, 448 (2020). (page 100).

[363] S. Ahmed, C. Sánchez Muñoz, F. Nori and A. F. Kockum. *Classification and reconstruction of optical quantum states with deep neural networks*. Phys. Rev. Research **3**, 033278 (2021). (page 100).

[364] A. M. Palmieri, E. Kovlakov, F. Bianchi, D. Yudin, S. Straupe, J. D. Biamonte and S. Kulik. *Experimental neural network enhanced quantum tomography*. npj Quantum Inf. **6**, 20 (2020). (page 100).

[365] A. Rocchetto, S. Aaronson, S. Severini, G. Carvacho, D. Poderini, I. Agresti, M. Bentivegna and F. Sciarrino. *Experimental learning of quantum states*. Sci. Adv. **5**, eaau1946 (2019). (page 100).

[366] E. van Nieuwenburg, Y.-H. Liu and S. Huber. *Learning phase transitions by confusion*. Nat. Phys. **13**, 435–439 (2017). (page 100).

[367] K. Kottmann, P. Huembeli, M. Lewenstein and A. Acín. *Unsupervised Phase Discovery with Deep Anomaly Detection*. Phys. Rev. Lett. **125**, 170603 (2020). (page 100).

[368] J. Carrasquilla and R. G. Melko. *Machine learning phases of matter*. Nat. Phys. **13**, 431–434 (2017). (page 100).

[369] S. Varsamopoulos, B. Criger and K. Bertels. *Decoding small surface codes with feedforward neural networks*. Quantum Sci. Technol. **3**, 015004 (2017). (page 100).

[370] H. P. Nautrup, N. Delfosse, V. Dunjko, H. J. Briegel and N. Friis. *Optimizing Quantum Error Correction Codes with Reinforcement Learning*. Quantum **3**, 215 (2019). (page 100).

[371] P. Baireuther, T. E. O'Brien, B. Tarasinski and C. W. J. Beenakker. *Machine-learning-assisted correction of correlated qubit errors in a topological code*. Quantum **2**, 48 (2018). (page 100).

[372] T. Fösel, P. Tighineanu, T. Weiss and F. Marquardt. *Reinforcement Learning with Neural Networks for Quantum Feedback*. Phys. Rev. X **8**, 031084 (2018). (page 100).

[373] M. Krenn, M. Malik, R. Fickler, R. Lapkiewicz and A. Zeilinger. *Automated Search for new Quantum Experiments*. Phys. Rev. Lett. **116**, 090405 (2016). (page 100).

[374] A. A. Melnikov, H. Poulsen Nautrup, M. Krenn, V. Dunjko, M. Tiersch, A. Zeilinger and H. J. Briegel. *Active learning machine learns to create new quantum experiments*. Proc. Natl. Acad. Sci. U.S.A. **115**, 1221 (2018). (page 100).

[375] M. Bukov, A. G. R. Day, D. Sels, P. Weinberg, A. Polkovnikov and P. Mehta. *Reinforcement Learning in Different Phases of Quantum Control*. Phys. Rev. X **8**, 031086 (2018). (page 101).

[376] K. K. Sabapathy, H. Qi, J. Izaac and C. Weedbrook. *Production of photonic universal quantum gates enhanced by machine learning*. Phys. Rev. A **100**, 012326 (2019). (page 101).

[377] N. Wiebe, C. Granade, C. Ferrie and D. G. Cory. *Hamiltonian Learning and Certification Using Quantum Resources*. Phys. Rev. Lett. **112**, 190501 (2014). (page 101).

[378] I. Agresti, N. Viggianiello, F. Flamini, N. Spagnolo, A. Crespi, R. Osellame, N. Wiebe and F. Sciarrino. *Pattern Recognition Techniques for Boson Sampling Validation*. Phys. Rev. X **9**, 011013 (2019). (page 101).

[379] D.-L. Deng. *Machine Learning Detection of Bell Nonlocality in Quantum Many-Body Systems*. Phys. Rev. Lett. **120**, 240402 (2018). (page 101).

[380] A. Canabarro, S. Brito and R. Chaves. *Machine Learning Nonlocal Correlations*. Phys. Rev. Lett. **122**, 200401 (2019). (page 101).

[381] A. Hentschel and B. C. Sanders. *Machine Learning for Precise Quantum Measurement*. Phys. Rev. Lett. **104**, 063603 (2010). (page 101).

[382] A. Lumino, E. Polino, A. S. Rab, G. Milani, N. Spagnolo, N. Wiebe and F. Sciarrino. *Experimental Phase Estimation Enhanced by Machine Learning*. Phys. Rev. Applied **10**, 044033 (2018). (page 101).

[383] G. Torlai, G. Mazzola, G. Carleo and A. Mezzacapo. *Precise measurement of quantum observables with neural-network estimators*. Phys. Rev. Research **2**, 022060 (2020). (page 101).

[384] V. Cimini, I. Gianani, N. Spagnolo, F. Leccese, F. Sciarrino and M. Barbieri. *Calibration of Quantum Sensors by Neural Networks*. Phys. Rev. Lett. **123**, 230502 (2019). (page 101).

[385] S. P. Nolan, A. Smerzi and L. Pezzè. *A machine learning approach to Bayesian parameter estimation*. arXiv preprint arXiv:2006.02369 (2020). (page 101).

[386] Y. Li, L. Pezzè, M. Gessner, Z. Ren, W. Li and A. Smerzi. *Frequentist and Bayesian Quantum Phase Estimation*. Entropy **20** (2018). (pages 101, 132, 151).

[387] S. P. Nolan, L. Pezzè and A. Smerzi. *Frequentist Parameter Estimation with Supervised Learning*. arXiv preprint arXiv:2105.12302 (2021). (page 101).

[388] V. Cimini, M. Barbieri, N. Treps, M. Walschaers and V. Parigi. *Neural Networks for Detecting Multimode Wigner Negativity*. Phys. Rev. Lett. **125**, 160504 (2020). (pages 101, 127).

[389] D. T. Smithey, M. Beck, M. G. Raymer and A. Faridani. *Measurement of the Wigner distribution and the density matrix of a light mode using optical homodyne tomography: Application to squeezed states and the vacuum*. Phys. Rev. Lett. **70**, 1244 (1993). (page 103).

[390] T. J. Dunn, I. A. Walmsley and S. Mukamel. *Experimental Determination of the Quantum-Mechanical State of a Molecular Vibrational Mode Using Fluorescence Tomography*. Phys. Rev. Lett. **74**, 884 (1995). (page 103).

[391] D. Leibfried, D. M. Meekhof, B. E. King, C. Monroe, W. M. Itano and D. J. Wineland. *Experimental Determination of the Motional Quantum State of a Trapped Atom*. Phys. Rev. Lett. **77**, 4281 (1996). (page 103).

[392] S. Deleglise, I. Dotsenko, C. Sayrin, J. Bernu, M. Brune, J.-M. Raimond and S. Haroche. *Reconstruction of non-classical cavity field states with snapshots of their decoherence*. Nature **455**, 510 (2008). (page 103).

[393] A. I. Lvovsky and M. G. Raymer. *Continuous-variable optical quantum-state tomography*. Rev. Mod. Phys. **81**, 299 (2009). (page 103).

[394] A. Mari, K. Kieling, B. M. Nielsen, E. S. Polzik and J. Eisert. *Directly Estimating Nonclassicality*. Phys. Rev. Lett. **106**, 010403 (2011). (page 103).

[395] M. Bohmann and E. Agudelo. *Phase-Space Inequalities Beyond Negativities*. Phys. Rev. Lett. **124**, 133601 (2020). (page 103).

[396] M. Bohmann, E. Agudelo and J. Sperling. *Probing nonclassicality with matrices of phase-space distributions*. Quantum **4**, 343 (2020). (page 103).

[397] N. Biagi, M. Bohmann, E. Agudelo, M. Bellini and A. Zavatta. *Experimental Certification of Nonclassicality via Phase-Space Inequalities*. Phys. Rev. Lett. **126**, 023605 (2021). (pages 104, 105).

[398] D. P. Kingma and J. Ba. *Adam: A method for stochastic optimization*. arXiv preprint arXiv:1412.6980 (2014). (pages 106, 119).

[399] E. Agudelo, J. Sperling, W. Vogel, S. Köhnke, M. Mraz and B. Hage. *Continuous sampling of the squeezed-state nonclassicality*. Phys. Rev. A **92**, 033837 (2015). (pages 107, 108).

[400] A. Zavatta, S. Viciani and M. Bellini. *Single-photon excitation of a coherent state: Catching the elementary step of stimulated light emission*. Phys. Rev. A **72**, 023820 (2005). (pages 110, 174).

[401] A. Zavatta, S. Viciani and M. Bellini. *Quantum-to-Classical Transition with Single-Photon-Added Coherent States of Light*. Science **306**, 660 (2004). (page 110).

[402] S. N. Filippov, V. I. Man'ko, A. S. Coelho, A. Zavatta and M. Bellini. *Single-photon-added coherent states: estimation of parameters and fidelity of the optical homodyne detection*. Phys. Scr. **T153**, 014025 (2013). (page 110).

[403] H. Paul, P. Törmä, T. Kiss and I. Jex. *Photon Chopping: New Way to Measure the Quantum State of Light*. Phys. Rev. Lett. **76**, 2464 (1996). (page 115).

[404] M. J. Fitch, B. C. Jacobs, T. B. Pittman and J. D. Franson. *Photon-number resolution using time-multiplexed single-photon detectors*. Phys. Rev. A **68**, 043814 (2003). (page 115).

[405] S. A. Castelletto, I. P. Degiovanni, V. Schettini and A. L. Migdall. *Reduced deadtime and higher rate photon-counting detection using a multiplexed detector array*. J. Mod. Opt. **54**, 337 (2007). (page 115).

[406] J. Sperling, W. Vogel and G. S. Agarwal. *True photocounting statistics of multiple on-off detectors*. Phys. Rev. A **85**, 023820 (2012). (pages 116, 117, 177, 179).

[407] L. Mandel. *Sub-Poissonian photon statistics in resonance fluorescence*. Opt. Lett. **4**, 205 (1979). (page 117).

[408] J. Sperling, W. Vogel and G. S. Agarwal. *Sub-Binomial Light*. Phys. Rev. Lett. **109**, 093601 (2012). (page 117).

[409] T. J. Bartley, G. Donati, X.-M. Jin, A. Datta, M. Barbieri and I. A. Walmsley. *Direct Observation of Sub-Binomial Light*. Phys. Rev. Lett. **110**, 173602 (2013). (page 117).

[410] J. Sperling, W. Vogel and G. S. Agarwal. *Correlation measurements with on-off detectors*. Phys. Rev. A **88**, 043821 (2013). (pages 117, 119).

[411] J. Sperling, M. Bohmann, W. Vogel, G. Harder, B. Brecht, V. Ansari and C. Silberhorn. *Uncovering Quantum Correlations with Time-Multiplexed Click Detection*. Phys. Rev. Lett. **115**, 023601 (2015). (pages 117, 120).

[412] J. Sperling, T. J. Bartley, G. Donati, M. Barbieri, X.-M. Jin, A. Datta, W. Vogel and I. A. Walmsley. *Quantum Correlations from the Conditional Statistics of Incomplete Data*. Phys. Rev. Lett. **117**, 083601 (2016). (page 117).

[413] M. Bohmann, R. Kruse, J. Sperling, C. Silberhorn and W. Vogel. *Probing free-space quantum channels with laboratory-based experiments*. Phys. Rev. A **95**, 063801 (2017). (page 117).

[414] A. Luis, J. Sperling and W. Vogel. *Nonclassicality Phase-Space Functions: More Insight with Fewer Detectors*. Phys. Rev. Lett. **114**, 103602 (2015). (page 117).

[415] M. Bohmann, J. Tiedau, T. Bartley, J. Sperling, C. Silberhorn and W. Vogel. *Incomplete Detection of Nonclassical Phase-Space Distributions*. Phys. Rev. Lett. **120**, 063607 (2018). (page 117).

[416] C. Lee, S. Ferrari, W. H. P. Pernice and C. Rockstuhl. *Sub-Poisson-binomial light*. Phys. Rev. A **94**, 053844 (2016). (page 117).

[417] M. Bohmann, L. Qi, W. Vogel and M. Chekhova. *Detection-device-independent verification of nonclassical light*. Phys. Rev. Research **1**, 033178 (2019). (page 117).

[418] T. Lipfert, J. Sperling and W. Vogel. *Homodyne detection with on-off detector systems*. Phys. Rev. A **92**, 053835 (2015). (pages 120, 179).

[419] R. Lynch. *The quantum phase problem: a critical review*. Phys. Rep. **256**, 367 (1995). (page 130).

[420] B. Yurke, S. L. McCall and J. R. Klauder. *SU(2) and SU(1,1) interferometers*. Phys. Rev. A **33**, 4033 (1986). (page 130).

[421] M. Kitagawa and M. Ueda. *Squeezed spin states*. Phys. Rev. A **47**, 5138 (1993). (page 130).

[422] D. J. Wineland, J. J. Bollinger, W. M. Itano and D. J. Heinzen. *Squeezed atomic states and projection noise in spectroscopy*. Phys. Rev. A **50**, 67 (1994). (page 130).

[423] T. Nagata, R. Okamoto, J. L. O'Brien, K. Sasaki and S. Takeuchi. *Beating the Standard Quantum Limit with Four-Entangled Photons*. Science **316**, 726 (2007). (page 130).

[424] M. Kacprowicz, R. Demkowicz-Dobrzański, W. Wasilewski, K. Banaszek and I. Walmsley. *Experimental quantum-enhanced estimation of a lossy phase shift*. Nat. Photonics **4**, 357 (2010). (page 130).

[425] C. Gross, T. Zibold, E. Nicklas, J. Esteve and M. K. Oberthaler. *Nonlinear atom interferometer surpasses classical precision limit*. Nature **464**, 1165 (2010). (page 130).

[426] C. R. Rao. *Information and the Accuracy Attainable in the Estimation of Statistical Parameters*. Springer, New York (1992). (page 130).

[427] H. Cramér. *Mathematical Methods of Statistics*. Princeton University Press, Princeton (1999). (page 130).

[428] M. W. Mitchell. *Metrology with entangled states*. In *Quantum Communications and Quantum Imaging III*, p. 263. SPIE (2005). (pages 132, 138).

[429] A. Y. Kitaev. *Quantum measurements and the Abelian stabilizer problem*. arXiv preprint quant-ph/9511026 (1995). (pages 132, 134, 136, 142, 150, 151).

[430] B. L. Higgins, D. W. Berry, S. D. Bartlett, M. W. Mitchell, H. M. Wiseman and G. J. Pryde. *Demonstrating Heisenberg-limited unambiguous phase estimation without adaptive measurements*. New J. Phys. **11**, 073023 (2009). (pages 133, 137, 138, 147, 148, 156, 157).

[431] D. S. Abrams and S. Lloyd. *Quantum Algorithm Providing Exponential Speed Increase for Finding Eigenvalues and Eigenvectors*. Phys. Rev. Lett. **83**, 5162 (1999). (pages 134, 137, 142, 151).

[432] A. Aspuru-Guzik, A. D. Dutoi, P. J. Love and M. Head-Gordon. *Simulated Quantum Computation of Molecular Energies*. Science **309**, 1704 (2005). (pages 134, 137, 142, 151).

[433] R. B. Griffiths and C.-S. Niu. *Semiclassical Fourier Transform for Quantum Computation*. Phys. Rev. Lett. **76**, 3228 (1996). (pages 136, 156).

[434] E. Martin-Lopez, A. Laing, T. Lawson, R. Alvarez, X.-Q. Zhou and J. L. O'brien. *Experimental realization of Shor's quantum factoring algorithm using qubit recycling*. Nat. Photonics **6**, 773 (2012). (page 137).

[435] T. Monz, D. Nigg, E. A. Martinez, M. F. Brandl, P. Schindler, R. Rines, S. X. Wang, I. L. Chuang and R. Blatt. *Realization of a scalable Shor algorithm*. Science **351**, 1068 (2016). (page 137).

[436] P. J. J. O'Malley, R. Babbush, I. D. Kivlichan, J. Romero, J. R. McClean, R. Barends, J. Kelly, P. Roushan, A. Tranter, N. Ding, B. Campbell, Y. Chen, Z. Chen, B. Chiaro, A. Dunsworth, A. G. Fowler, E. Jeffrey, E. Lucero, A. Megrant, J. Y. Mutus, M. Neeley, C. Neill, C. Quintana, D. Sank, A. Vainsencher, J. Wenner, T. C. White, P. V.

Coveney, P. J. Love, H. Neven, A. Aspuru-Guzik and J. M. Martinis. *Scalable Quantum Simulation of Molecular Energies*. Phys. Rev. X **6**, 031007 (2016). (page 137).

[437] K. Temme, T. J. Osborne, K. G. Vollbrecht, D. Poulin and F. Verstraete. *Quantum metropolis sampling*. Nature **471**, 87 (2011). (page 137).

[438] T. Rudolph and L. Grover. *Quantum Communication Complexity of Establishing a Shared Reference Frame*. Phys. Rev. Lett. **91**, 217905 (2003). (page 137).

[439] M. de Burgh and S. D. Bartlett. *Quantum methods for clock synchronization: Beating the standard quantum limit without entanglement*. Phys. Rev. A **72**, 042301 (2005). (page 137).

[440] P. Komar, E. M. Kessler, M. Bishof, L. Jiang, A. S. Sørensen, J. Ye and M. D. Lukin. *A quantum network of clocks*. Nat. Phys. **10**, 582 (2014). (pages 137, 140).

[441] E. M. Kessler, P. Kómár, M. Bishof, L. Jiang, A. S. Sørensen, J. Ye and M. D. Lukin. *Heisenberg-Limited Atom Clocks Based on Entangled Qubits*. Phys. Rev. Lett. **112**, 190403 (2014). (page 137).

[442] L. Pezzè and A. Smerzi. *Heisenberg-Limited Noisy Atomic Clock Using a Hybrid Coherent and Squeezed State Protocol*. Phys. Rev. Lett. **125**, 210503 (2020). (page 137).

[443] V. Giovannetti, S. Lloyd and L. Maccone. *Advances in quantum metrology*. Nat. Photonics **5**, 222 (2011). (page 137).

[444] D. W. Berry, B. L. Higgins, S. D. Bartlett, M. W. Mitchell, G. J. Pryde and H. M. Wiseman. *How to perform the most accurate possible phase measurements*. Phys. Rev. A **80**, 052114 (2009). (pages 137, 138, 147, 148, 156).

[445] L. Pezzé and A. Smerzi. *Sub shot-noise interferometric phase sensitivity with beryllium ions Schrödinger cat states*. EPL **78**, 30004 (2007). (pages 137, 139).

[446] S. Boixo and R. D. Somma. *Parameter estimation with mixed-state quantum computation*. Phys. Rev. A **77**, 052320 (2008). (page 138).

[447] L. Pezzè and A. Smerzi. *Quantum Phase Estimation Algorithm with Gaussian Spin States*. PRX Quantum **2**, 040301 (2021). (pages 138, 156).

[448] N. Wiebe and C. Granade. *Efficient Bayesian Phase Estimation*. Phys. Rev. Lett. **117**, 010503 (2016). (pages 138, 147, 152, 157, 158, 159, 160, 165).

[449] A. A. Berni, T. Gehring, B. M. Nielsen, V. Händchen, M. G. Paris and U. L. Andersen. *Ab initio quantum-enhanced optical phase estimation using real-time feedback control*. Nat. Photonics **9**, 577 (2015). (page 138).

[450] S. Daryanoosh, S. Slussarenko, D. W. Berry, H. M. Wiseman and G. J. Pryde. *Experimental optical phase measurement approaching the exact Heisenberg limit*. Nat. Commun. **9**, 1 (2018). (page 138).

[451] S. Paesani, A. A. Gentile, R. Santagati, J. Wang, N. Wiebe, D. P. Tew, J. L. O'Brien and M. G. Thompson. *Experimental Bayesian Quantum Phase Estimation on a Silicon Photonic Chip*. Phys. Rev. Lett. **118**, 100503 (2017). (pages 138, 157).

[452] R. Santagati, J. Wang, A. A. Gentile, S. Paesani, N. Wiebe, J. R. McClean, S. Morley-Short, P. J. Shadbolt, D. Bonneau, J. W. Silverstone, D. P. Tew, X. Zhou, J. L. O'Brien and M. G. Thompson. *Witnessing eigenstates for quantum simulation of Hamiltonian spectra*. Sci. Adv. **4**, eaap9646 (2018). (page 138).

[453] C. Bonato, M. S. Blok, H. T. Dinani, D. W. Berry, M. L. Markham, D. J. Twitchen and R. Hanson. *Optimized quantum sensing with a single electron spin using real-time adaptive measurements*. Nat. Nanotechnol. **11**, 247 (2016). (page 138).

[454] P. C. Humphreys, M. Barbieri, A. Datta and I. A. Walmsley. *Quantum Enhanced Multiple Phase Estimation*. Phys. Rev. Lett. **111**, 070403 (2013). (pages 139, 162, 165).

[455] W. Gorecki and R. Demkowicz-Dobrzanski. *Multiple-phase quantum interferometry–real and apparent gains of measuring all the phases simultaneously*. arXiv preprint arXiv:2107.10863 (2021). (page 139).

[456] T. Baumgratz and A. Datta. *Quantum Enhanced Estimation of a Multidimensional Field*. Phys. Rev. Lett. **116**, 030801 (2016). (pages 139, 140).

[457] M. D. Vidrighin, G. Donati, M. G. Genoni, X.-M. Jin, W. S. Kolthammer, M. Kim, A. Datta, M. Barbieri and I. A. Walmsley. *Joint estimation of phase and phase diffusion for quantum metrology*. Nat. Commun. **5**, 1 (2014). (page 139).

[458] P. J. D. Crowley, A. Datta, M. Barbieri and I. A. Walmsley. *Tradeoff in simultaneous quantum-limited phase and loss estimation in interferometry*. Phys. Rev. A **89**, 023845 (2014). (page 139).

[459] S. Ragy, M. Jarzyna and R. Demkowicz-Dobrzański. *Compatibility in multiparameter quantum metrology*. Phys. Rev. A **94**, 052108 (2016). (page 139).

[460] M. A. Ciampini, N. Spagnolo, C. Vitelli, L. Pezzè, A. Smerzi and F. Sciarrino. *Quantum-enhanced multiparameter estimation in multiarm interferometers*. Sci. Rep. **6**, 1 (2016). (pages 139, 141, 164).

[461] T. J. Proctor, P. A. Knott and J. A. Dunningham. *Multiparameter Estimation in Networked Quantum Sensors*. Phys. Rev. Lett. **120**, 080501 (2018). (pages 139, 140, 162, 165).

[462] W. Ge, K. Jacobs, Z. Eldredge, A. V. Gorshkov and M. Foss-Feig. *Distributed Quantum Metrology with Linear Networks and Separable Inputs*. Phys. Rev. Lett. **121**, 043604 (2018). (pages 139, 140, 162, 165).

[463] M. Gessner, L. Pezzè and A. Smerzi. *Sensitivity Bounds for Multiparameter Quantum Metrology*. Phys. Rev. Lett. **121**, 130503 (2018). (pages 139, 141, 162, 164, 165).

[464] A. Z. Goldberg, I. Gianani, M. Barbieri, F. Sciarrino, A. M. Steinberg and N. Spagnolo. *Multiphase estimation without a reference mode*. Phys. Rev. A **102**, 022230 (2020). (page 139).

[465] C. Oh, C. Lee, S. H. Lie and H. Jeong. *Optimal distributed quantum sensing using Gaussian states*. Phys. Rev. Research **2**, 023030 (2020). (page 139).

[466] M. Gessner, A. Smerzi and L. Pezzè. *Multiparameter squeezing for optimal quantum enhancements in sensor networks*. Nat. Commun. **11**, 1 (2020). (page 139).

[467] J. Rubio and J. Dunningham. *Bayesian multiparameter quantum metrology with limited data*. Phys. Rev. A **101**, 032114 (2020). (page 140).

[468] T. E. O'Brien, B. Tarasinski and B. M. Terhal. *Quantum phase estimation of multiple eigenvalues for small-scale (noisy) experiments*. New J. Phys. **21**, 023022 (2019). (pages 140, 147, 157, 165).

[469] R. D. Somma. *Quantum eigenvalue estimation via time series analysis*. New J. Phys. **21**, 123025 (2019). (page 140).

[470] M. I. Kolobov. *Quantum Imaging*. Springer, New York (2007). (page 140).

[471] F. Albarelli, M. Barbieri, M. Genoni and I. Gianani. *A perspective on multiparameter quantum metrology: From theoretical tools to applications in quantum imaging*. Phys. Lett. A **384**, 126311 (2020). (page 140).

[472] J. Řehaček, Z. Hradil, B. Stoklasa, M. Paúr, J. Grover, A. Krzic and L. L. Sánchez-Soto. *Multiparameter quantum metrology of incoherent point sources: Towards realistic superresolution*. Phys. Rev. A **96**, 062107 (2017). (page 140).

[473] C. Lupo and S. Pirandola. *Ultimate Precision Bound of Quantum and Subwavelength Imaging*. Phys. Rev. Lett. **117**, 190802 (2016). (page 140).

[474] S. Altenburg, M. Oszmaniec, S. Wölk and O. Gühne. *Estimation of gradients in quantum metrology*. Phys. Rev. A **96**, 042319 (2017). (page 140).

[475] I. Apellaniz, I. n. Urizar-Lanz, Z. Zimborás, P. Hyllus and G. Tóth. *Precision bounds for gradient magnetometry with atomic ensembles*. Phys. Rev. A **97**, 053603 (2018). (page 140).

[476] X. Guo, C. R. Breum, J. Borregaard, S. Izumi, M. V. Larsen, T. Gehring, M. Christandl, J. S. Neergaard-Nielsen and U. L. Andersen. *Distributed quantum sensing in a continuous-variable entangled network*. Nat. Phys. **16**, 281 (2020). (pages 140, 152, 163).

[477] M. A. Taylor and W. P. Bowen. *Quantum metrology and its application in biology*. Phys. Rep. **615**, 1 (2016). (page 140).

[478] V. Cimini, I. Gianani, L. Ruggiero, T. Gasperi, M. Sbroscia, E. Roccia, D. Tofani, F. Bruni, M. A. Ricci and M. Barbieri. *Quantum sensing for dynamical tracking of chemical processes*. Phys. Rev. A **99**, 053817 (2019). (page 140).

[479] E. Polino, M. Riva, M. Valeri, R. Silvestri, G. Corrielli, A. Crespi, N. Spagnolo, R. Osellame and F. Sciarrino. *Experimental multiphase estimation on a chip*. Optica **6**, 288 (2019). (pages 140, 152, 163, 165).

[480] M. Valeri, E. Polino, D. Poderini, I. Gianani, G. Corrielli, A. Crespi, R. Osellame, N. Spagnolo and F. Sciarrino. *Experimental adaptive Bayesian estimation of multiple phases with limited data*. npj Quantum Inf. **6**, 1 (2020). (pages 140, 147, 152, 163, 165).

[481] L.-Z. Liu, Y.-Z. Zhang, Z.-D. Li, R. Zhang, X.-F. Yin, Y.-Y. Fei, L. Li, N.-L. Liu, F. Xu, Y.-A. Chen et al. *Distributed quantum phase estimation with entangled photons*. Nat. Photonics **15**, 137 (2021). (pages 141, 152).

[482] S.-R. Zhao, Y.-Z. Zhang, W.-Z. Liu, J.-Y. Guan, W. Zhang, C.-L. Li, B. Bai, M.-H. Li, Y. Liu, L. You, J. Zhang, J. Fan, F. Xu, Q. Zhang and J.-W. Pan. *Field Demonstration of Distributed Quantum Sensing without Post-Selection*. Phys. Rev. X **11**, 031009 (2021). (pages 141, 152).

[483] D. Braun, G. Adesso, F. Benatti, R. Floreanini, U. Marzolino, M. W. Mitchell and S. Pirandola. *Quantum-enhanced measurements without entanglement*. Rev. Mod. Phys. **90**, 035006 (2018). (page 141).

[484] J. Liu and M. West. *Combined Parameter and State Estimation in Simulation-Based Filtering*. In *Sequential Monte Carlo Methods in Practice*, p. 197. Springer, New York (2001). (pages 147, 165).

[485] C. E. Granade, C. Ferrie, N. Wiebe and D. G. Cory. *Robust online Hamiltonian learning*. New J. Phys. **14**, 103013 (2012). (pages 147, 165).

[486] H. Niederreiter. *Random number generation and quasi-Monte Carlo methods*. SIAM, Philadelphia (1978). (page 149).

[487] D. W. Berry and H. M. Wiseman. *Optimal States and Almost Optimal Adaptive Measurements for Quantum Interferometry*. Phys. Rev. Lett. **85**, 5098 (2000). (page 151).

[488] Y. Xia, W. Li, W. Clark, D. Hart, Q. Zhuang and Z. Zhang. *Demonstration of a Reconfigurable Entangled Radio-Frequency Photonic Sensor Network*. Phys. Rev. Lett. **124**, 150502 (2020). (page 152).

[489] L. Zhang and K. W. C. Chan. *Scalable generation of multi-mode noon states for quantum multiple-phase estimation*. Sci. Rep. **8**, 1 (2018). (page 162).

[490] A. R. R. Carvalho, F. Mintert and A. Buchleitner. *Decoherence and Multipartite Entanglement*. Phys. Rev. Lett. **93**, 230501 (2004). (page 162).

[491] M. Bohmann, J. Sperling and W. Vogel. *Entanglement verification of noisy NOON states*. Phys. Rev. A **96**, 012321 (2017). (page 162).

[492] S. Hong, Y.-S. Kim, Y.-W. Cho, S.-W. Lee, H. Jung, S. Moon, S.-W. Han, H.-T. Lim et al. *Quantum enhanced multiple-phase estimation with multi-mode N00N states*. Nat. Commun. **12**, 1 (2021). (page 162).

[493] N. Spagnolo, L. Aparo, C. Vitelli, A. Crespi, R. Ramponi, R. Osellame, P. Mataloni and F. Sciarrino. *Quantum interferometry with three-dimensional geometry*. Sci. Rep. **2**, 1 (2012). (pages 163, 165).

[494] E. Polino, M. Valeri, N. Spagnolo and F. Sciarrino. *Photonic quantum metrology*. AVS Quantum Sci. **2**, 024703 (2020). (pages 163, 165).

[495] M. Żukowski, A. Zeilinger and M. A. Horne. *Realizable higher-dimensional two-particle entanglements via multiport beam splitters*. Phys. Rev. A **55**, 2564 (1997). (page 163).

[496] N. Spagnolo, C. Vitelli, L. Aparo, P. Mataloni, F. Sciarrino, A. Crespi, R. Ramponi and R. Osellame. *Three-photon bosonic coalescence in an integrated tritter*. Nat. Commun. **4**, 1 (2013). (page 163).

[497] D. Y. Vasylyev, A. A. Semenov and W. Vogel. *Toward Global Quantum Communication: Beam Wandering Preserves Nonclassicality*. Phys. Rev. Lett. **108**, 220501 (2012). (page 171).

[498] B. Kühn and W. Vogel. *Visualizing nonclassical effects in phase space*. Phys. Rev. A **90**, 033821 (2014). (pages 173, 179).

[499] A. P. Prudnikov, Y. A. Brychkov and O. I. Marichev. Integrals and series (1988). (page 173).

[500] J. Sperling, W. Vogel and G. S. Agarwal. *Quantum state engineering by click counting*. Phys. Rev. A **89**, 043829 (2014). (page 178).

[501] G. S. Agarwal and K. Tara. *Nonclassical character of states exhibiting no squeezing or sub-Poissonian statistics*. Phys. Rev. A **46**, 485 (1992). (page 180).