

Inferring properties of quantum systems with random measurements

DISSERTATION

zur Erlangung des Grades eines Doktors der Naturwissenschaften

vorgelegt von
Jan Lennart Bönsel

eingereicht bei der Naturwissenschaftlichen-Technischen Fakultät
der Universität Siegen
Siegen 2024

Gutachter:

Prof. Dr. Otfried Gühne

Prof. Dr. Mariami Gachechiladze

Prüfer:

Prof. Dr. Otfried Gühne

Prof. Dr. Mariami Gachechiladze

Prof. Dr. Christof Wunderlich

Jun.-Prof. Dr. Stefan Nimmrichter

Tag der mündlichen Prüfung: 30.10.2024

Abstract

How can quantum systems be probed by random measurements? And what are the advantages? With the advance of quantum technologies, larger and larger quantum systems have to be characterized. An important question in quantum information theory is therefore the development of efficient methods to infer properties of a quantum system. In this regard, random measurements are discussed in a variety of different contexts. The applications range from the estimation of the fidelity by sampling the Pauli operators at random to measurements in random directions for entanglement detection. Random measurements have also been used in the original formulation of shadow tomography.

In this thesis we first consider the evaluation of spin-squeezing inequalities by random measurements in Sec. 3. For this purpose we note that the spin-squeezing inequalities can also be retrieved from pair correlations. This opens the possibility to randomize the scheme. We show that spin-squeezing inequalities can be obtained from random pair correlations, i.e., correlations between random pairs of qubits. The spin-squeezing inequalities are nonlinear in the quantum state and thus we propose an approach to perform a statistical analysis of the nonlinear estimators. Our statistical analysis is not limited to spin-squeezing inequalities but can also be applied to other linear or nonlinear quantities.

In Sec. 4, we apply a similar randomized approach to Bell inequalities. Certain classes of Bell inequalities for multiqubit systems show the promising feature of an exponentially increasing violation of the bound in local theories. Whereas this makes the inequalities more robust against noise, it also comes with the caveat that the measurement resources increase exponentially. We show that it is not necessary to sample all measurement settings of the Bell inequality. A statistically significant violation of a Bell inequality can also be achieved by sampling fewer measurement settings at random. We further point out that for graph states, which are a specific subset of entangled multiqubit states, there are Bell inequalities known that exhibit an exponential nonlocality. As graph states can be readily adapted to the two-qubit connectivity of a quantum computer, they can be used to benchmark quantum computers by the produced nonlocality.

In Sec. 5 in turn, we consider measurements in random bases. We show that all invariants under local unitary transformations can be inferred from randomized measurements and give expressions for all invariants in the two qubit case. The method is implemented in an experiment and we include two applications. On the one hand, we derive the Bell violation that can be observed for the state in the experiment. On the other hand, we show that also the usefulness of the state in teleportation protocols can be assessed from the data.

Finally, in Sec. 6 we discuss a new formulation of shadow tomography. Whereas the original scheme uses unitaries that are sampled from a fixed set at random, we show that shadow tomography can also be formulated in terms of generalized measurements. This formulation puts the method in a new light. Especially, it shows that shadow tomography cannot only be implemented by randomization but also by introducing an ancilla system. The formulation in terms of generalized measurements in addition allows for a natural way to include noise and for the optimization of the measurements.

Zusammenfassung

Wie lassen sich Quantensysteme mit Hilfe von zufälligen Messungen erforschen? Und ergeben sich daraus Vorteile? Der Fortschritt der Quantentechnologien erfordert es, immer größere Quantensysteme zu charakterisieren. Eine drängende Frage in der Quanteninformationstheorie ist daher die Entwicklung effizienter Messmethoden. In diesem Zusammenhang wurden zufällige Messungen für eine Vielzahl unterschiedlicher Anwendungen diskutiert. Die Anwendungen reichen dabei von der Abschätzung der Fidelität eines Quantenzustands durch zufällige Messung der Pauli-Operatoren bis hin zu Messungen in zufälligen Richtungen zum Nachweis von Verschränkung. Zudem werden Zufallsmessungen auch in der ursprünglichen Formulierung von Shadow Tomography verwendet.

In dieser Arbeit betrachten wir in Abschnitt 3 zunächst die Auswertung von Spin-Squeezing-Ungleichungen durch zufällige Messungen. Dazu weisen wir darauf hin, dass die Spin-Squeezing-Ungleichungen auch aus Paarkorrelationen bestimmt werden können. Dies ermöglicht, das Schema zu randomisieren. Wir zeigen, dass Spin-Squeezing-Ungleichungen aus zufälligen Paarkorrelationen, d.h. Korrelationen zwischen zufälligen Paaren von Qubits, gewonnen werden können. Spin-Squeezing-Ungleichungen sind nicht linear im Quantenzustand und wir schlagen daher eine statistische Analyse vor, die auch bei nicht linearen Größen verwendet werden kann. Die statistische Analyse ist nicht auf Spin-Squeezing-Ungleichungen beschränkt, sondern kann auch auf andere lineare oder nichtlineare Größen angewendet werden.

In Abschnitt 4 wenden wir einen ähnlichen Ansatz auf Bell-Ungleichungen an. Bestimmte Klassen von Bell-Ungleichungen für Multiqubit-Systeme zeigen die vielversprechende Eigenschaft, dass die Verletzung der lokalen Schranke exponentiell mit der Anzahl der Qubits zunimmt. Dies macht die Ungleichungen zwar robuster gegen Rauschen, der experimentelle Test benötigt jedoch Messressourcen, die auch exponentiell ansteigen. Wir zeigen, dass es nicht notwendig ist, alle Messeinstellungen der Bell-Ungleichung zu messen. Eine statistisch signifikante Verletzung einer Bell-Ungleichung kann auch durch eine zufällige Auswahl weniger Messeinstellungen erreicht werden. Wir betonen, dass für Graphen-Zustände, die eine spezielle Untermenge von verschränkten Multiqubit-Zuständen darstellen, Bell-Ungleichungen mit einer exponentiellen Verletzung bekannt sind. Da Graphenzustände leicht an die Zwei-Qubit-Konnektivität eines Quantencomputers angepasst werden können, lassen sie sich für den Vergleich von Quantencomputern anhand der erzeugten Verletzung verwenden.

In Abschnitt 5 betrachten wir Messungen in zufälligen Messbasen. Wir zeigen, dass alle Invarianten unter lokal unitären Transformationen aus zufälligen Messungen abgeleitet werden können und geben die Ausdrücke aller Invarianten für ein System aus zwei Qubits an. Die Methode wird an einem System aus zwei Qubits experimentell durchgeführt. Als Anwendung leiten wir einerseits die mögliche Bell-Verletzung des Quantenzustands ab. Andererseits zeigen wir, dass mithilfe zufälliger Messungen abgeschätzt werden kann, ob der Quantenzustand sich für Teleportationsprotokolle eignet.

Schließlich diskutieren wir in Abschnitt 6 eine neue Formulierung der Shadow Tomography. Während die ursprüngliche Methode unitäre Transformationen verwendet, die zufällig aus einer festen Menge ausgewählt werden, zeigen wir, dass Shadow Tomography auch mit verallgemeinerten Messungen formuliert werden kann. Daraus ergeben sich neue Sichtweisen auf die Methode. Insbesondere zeigt die Formulierung mithilfe von verallgemeinerten Messungen, dass Shadow Tomography nicht nur durch zufällige Messungen sondern auch durch die Einführung eines Hilffsystems umgesetzt werden kann. Die Formulierung in Form von verallgemeinerten Messungen ermöglicht darüber hinaus die einfache Berücksichtigung von Rauschen und die Optimierung der Messungen.

Contents

Abstract	3
Zusammenfassung	4
Introduction	8
1 Quantum information theory	10
1.1 Entanglement	10
1.2 Graph states	11
1.3 Measurements	13
1.3.1 Projective measurements	14
1.3.2 Positive operator valued measures (POVMs)	14
1.4 Entanglement detection	16
1.4.1 Entanglement witnesses	16
1.4.2 Spin-squeezing inequalities	18
1.4.3 Bell inequalities	20
1.5 Randomized measurements	24
1.5.1 Motivation	24
1.5.2 Moments of the outcome distributions	25
1.5.3 Local unitary invariants	26
1.6 Shadow tomography	27
2 Statistical tools	30
2.1 Probability and random variables	30
2.2 Estimators	32
2.3 Hypothesis test	33
2.4 Concentration inequalities	35
3 Error estimation of different schemes to measure spin-squeezing inequalities	36
3.1 Introduction	36
3.2 Formulation as a hypothesis test	37
3.3 Three ways to measure spin-squeezing inequalities	38
3.3.1 Estimator based on the total spin	39
3.3.2 Estimator based on pair correlations	40
3.3.3 Estimator based on random pair correlations	41
3.4 Statistical analysis	43
3.4.1 Variances	44
3.4.2 Scaling of the variances	45
3.4.3 Statistical test	46
3.5 Discussion	47
4 Generating multipartite nonlocality to benchmark quantum computers	49
4.1 Introduction	49
4.2 Methods	51
4.2.1 Bell test as a hypothesis test	51
4.2.2 Random sampling	52
4.3 Number of measurement repetitions	53
4.4 Analysis of the Bell inequalities for the GHZ and LC state	54

4.5	Bell nonlocality as benchmark	55
4.5.1	Connectivities of current quantum computers	55
4.5.2	Noise	58
4.5.3	Simulation for an IBM quantum computer	59
4.6	Discussion	61
5	Complete characterization of quantum correlations by randomized measurements	63
5.1	Introduction	63
5.2	LU invariants from the moments of randomized measurements	64
5.3	Experimental setup	65
5.4	Applications	66
5.4.1	Detection of Bell nonlocality	66
5.4.2	Teleportation fidelity	67
5.4.3	Error bounds of the moments	67
5.4.4	Results for the Bell violation and teleportation fidelity	69
5.5	Conclusion	70
6	Optimizing shadow tomography with generalized measurements	72
6.1	Introduction	72
6.2	Shadow tomography formulated in terms of POVMs	73
6.2.1	Shadow tomography derived from the least square estimator	73
6.2.2	Relation to randomized projective measurements	75
6.2.3	Symmetry of generalized measurements and the computation of the classical shadows	76
6.2.4	Tensoring the shadow construction for many-body systems	77
6.2.5	Protocol of shadow tomography with generalized measurements	78
6.3	Statistical analysis: Shadow norm	78
6.4	Effects of noise in measurements	80
6.5	Inferring properties of the Ising model	81
6.6	Outlook	83
6.6.1	Optimization of POVMs for shadow tomography	83
6.6.2	Discussion	85
	Conclusion and outlook	86
	List of abbreviations	88
A	Additional calculations for Sec. 3	89
A.1	Unbiased estimators	89
A.1.1	Estimator based on pair correlations	89
A.1.2	Estimator based on random pair correlations	90
A.2	Derivation of the variances	91
A.2.1	Estimator based on the total spin	91
A.2.2	Estimator based on pair correlations	94
A.2.3	Estimator based on random pair correlations	97
A.3	Expressions for the singlet and Dicke state	100
A.3.1	Singlet state	100
A.3.2	Dicke states	101

B Additional calculations for Sec. 4	104
B.1 Unbiased estimators	104
B.1.1 Estimator in the infinite measurement limit	104
B.1.2 Estimator for finite repetitions	104
B.2 Hoeffding's inequality	104
B.2.1 Estimator in the infinite measurement limit	104
B.2.2 Estimator for finite repetitions	105
B.3 Preparation scheme for the LC state	105
List of Publications	109
Bibliography	109
Acknowledgments	120

Introduction

Since the formulation of quantum theory, it was recognized that it predicts peculiar effects that contradict the intuitive assumptions on a physical theory. For this reason, Einstein, Podolsky and Rosen were of the opinion that the theory is incomplete [5]. This led to the idea of **local hidden variable (LHV)** models to complete quantum mechanics to a local realistic theory. The reasoning was later formalized by Bell, which led to the formulation of Bell inequalities [6]. A violation of a Bell inequality contradicts the assumption of locality and is therefore nowadays referred to as nonlocality [7]. Intertwined with nonlocality is the effect of entanglement. The term entanglement was first used by Schrödinger [8] and refers to the effect that certain quantum states of composite systems do not allow for a separate description of the parts. It later turned out that entanglement has many applications and plays a crucial role, when it comes to the advantage of quantum technology over classical systems. Entanglement can for example be used to enhance the precision in metrology [9–11] or to generate secure keys in cryptography [12, 13] It moreover is a key resource in quantum computation [14, 15].

To make use of entanglement, however, requires good control over the quantum system. Not only does one have to be able to prepare and modify the quantum state. It is also important to characterize the system. For this purpose, the system has to be probed by measurements, e.g., to get insight into the state of the system [16] or to detect entanglement [17].

In recent years, random measurements have gotten more and more into focus. The term random measurements is used in various contexts. On the one hand, random measurements have been proposed to reduce the measurement resources. For example, the fidelity of a multiqubit state can be inferred by performing random Pauli measurements. This approach is known as direct fidelity estimation [18, 19]. Moreover, entanglement can be detected by measuring two-outcome observables at random [20, 21]. On the other hand, entanglement is invariant under local unitary transformations [17]. It is thus invariant under local basis change, and it has been shown that entanglement can be inferred from measurements in random bases that are sampled according to the Haar measure [22–25]. This approach is commonly referred to as randomized measurements. Finally, we note that random measurements are also used in the original formulation of shadow tomography [26]. The goal of shadow tomography is to estimate in principle any observable from the recorded data without reconstructing the density matrix. For this purpose, Ref. [26] proposed a scheme that applies specific unitaries at random before the measurement, which is equivalent to sampling from a set of observables that are tomographically complete.

In this thesis we discuss various aspects of random measurements. In doing so, we are concerned with the question which properties can be efficiently observed by random measurements. In Sec. 1, we introduce the basic concepts of quantum information theory that are used in this thesis. As one of the main goals is entanglement detection, we introduce entanglement witnesses, spin-squeezing inequalities and Bell inequalities. But we also give an introduction how entanglement can be verified from randomized measurements. Finally, we briefly describe the scheme of shadow tomography.

As we pose the question whether properties can be efficiently inferred from random measurements, we make use of statistical tools. For this purpose we introduce hypothesis testing in Sec. 2. The sample complexity can be assessed with the help of concentration inequalities.

In Sec. 3, we propose a scheme to evaluate spin-squeezing inequalities from random pair correlations. We furthermore assess the sample complexity of the approach. For this we discuss a statistical analysis of the nonlinear estimators for the spin-squeezing inequalities. We note that the analysis is not limited to spin-squeezing inequalities. It also allows gauging the statistics of other linear or nonlinear quantities.

In the same spirit, we dedicate Sec. 4 to the evaluation of multipartite Bell inequalities. We propose to sample the different terms of the Bell inequalities, i.e., the measurement settings, at random. We show that this can significantly reduce the sample complexity. Especially for multipartite Bell inequalities the sample complexity is an issue, as in many interesting cases the number of measurement settings increases exponentially with the number of parties.

Sec. 5 deals with randomized measurements, i.e., the measurements are performed in random local bases. It is thus only possible to infer properties that are invariant under **local unitary (LU)** transformations. We show how to express **LU** invariants in terms of the moments of randomized measurements and give the expressions for the complete set of **LU** invariants for two qubits. Finally, the two-qubit case is implemented in an experiment.

In Sec. 6, we formulate shadow tomography in terms of generalized measurements (**positive operator valued measure (POVM)**s). The focus in this section is not on how the tomographic data can be obtained from random measurements. This has already been established by the original scheme in [26]. Rather, the formulation in terms of **POVM**s allows for an implementation of shadow tomography without random measurements. The formulation thus simplifies the method of shadow tomography and allows naturally to take into account noise and to optimize the measurements.

1 Quantum information theory

We start with a brief introduction to the theory of quantum information [27, 28]. In doing so, we introduce the basic concepts that are needed in this thesis. First, we will discuss how quantum states are represented in quantum theory, which is directly related to the concept of entanglement. Afterward, we will introduce graph states, which are an important class of entangled multiqubit states. What follows is a description of measurements in quantum mechanics and an introduction how entanglement can be detected [17]. We close this section by a more detailed description of randomized measurements and shadow tomography.

1.1 Entanglement

In quantum mechanics, states are described by vectors $|\psi\rangle$ in a Hilbert space \mathcal{H} over the complex field \mathbb{C} . We denote a basis of \mathcal{H} by $\mathcal{B} = \{|i\rangle \mid i = 0, \dots, d-1\}$, where d is the dimension. A state $|\psi\rangle$ can be expanded in the basis \mathcal{B} , i.e.,

$$|\psi\rangle = \sum_{i=0}^{d-1} \alpha_i |i\rangle. \quad (1.1)$$

In the above expression, α_i is the projection of the state $|\psi\rangle$ on the basis state $|i\rangle$, i.e., $\alpha_i = \langle i|\psi\rangle$. A defining property of quantum mechanics is the way composite systems are described. Suppose two quantum systems A and B with bases $\mathcal{B}_A = \{|i\rangle_A \mid i = 0, \dots, d_A-1\}$ and $\mathcal{B}_B = \{|j\rangle_B \mid j = 0, \dots, d_B-1\}$. In case system A is in the state $|\psi\rangle_A = \sum_{i=0}^{d_A-1} \alpha_i |i\rangle_A$ and the state of system B is described by $|\phi\rangle_B = \sum_{j=0}^{d_B-1} \beta_j |j\rangle_B$, the state of the composite system is given by

$$|\psi\rangle_A \otimes |\phi\rangle_B := \sum_{i=0}^{d_A-1} \sum_{j=0}^{d_B-1} \alpha_i \beta_j |i\rangle_A \otimes |j\rangle_B. \quad (1.2)$$

In the above equation, \otimes denotes the tensor product of states in \mathcal{H}_A and \mathcal{H}_B . The composite Hilbert space is thus given by the tensor product space $\mathcal{H}_{AB} = \mathcal{H}_A \otimes \mathcal{H}_B$ that is spanned by the basis $\mathcal{B}_{AB} = \{|i\rangle_A \otimes |j\rangle_B \mid i = 0, \dots, d_A-1; j = 0, \dots, d_B-1\}$. We can see that the composite Hilbert space has dimension $\dim(\mathcal{H}_{AB}) = d_A \times d_B$. Composite systems are thus not described by the direct sum of the Hilbert spaces. Rather, the dimensions multiply.

It turns out, however, that not all states of the composite system can be described by states of the form in Eq. (1.2). This leads to the concept of entanglement.

Definition 1.1. A bipartite, pure state $|\psi\rangle \in \mathcal{H}_A \otimes \mathcal{H}_B$ is called separable if it can be written as a tensor product of states $|\phi\rangle_A \in \mathcal{H}_A$ and $|\varphi\rangle_B \in \mathcal{H}_B$, i.e.,

$$|\psi\rangle = |\phi\rangle_A \otimes |\varphi\rangle_B. \quad (1.3)$$

Otherwise, $|\psi\rangle$ is called entangled.

In case the knowledge about the quantum state is limited, the state can be described by a mixed state

$$\rho = \sum_i p_i |\psi_i\rangle\langle\psi_i|, \quad (1.4)$$

where the system is in state $|\psi_i\rangle$ with probability $p_i \geq 0$. In particular, it is $\sum_i p_i = 1$. Entanglement can also be defined for mixed states.

Definition 1.2. A bipartite mixed state ρ is called separable if it can be written as a convex combination of product states $\rho_i^A \otimes \rho_i^B$, i.e.,

$$\rho = \sum_i p_i \rho_i^A \otimes \rho_i^B. \quad (1.5)$$

Above ρ_i^A and ρ_i^B are states of system A and B respectively. Otherwise, the mixed state ρ is entangled.

From the above definition, we can conclude that the set of separable states is convex.

For more than two parties, entanglement becomes more complex. A fully separable state of n parties can be defined analogously to the bipartite case.

Definition 1.3. A mixed state ρ is called (fully) separable if it can be written as a convex combination of product states, i.e.,

$$\rho = \sum_i p_i \rho_i^{(1)} \otimes \rho_i^{(2)} \otimes \dots \otimes \rho_i^{(N)}, \quad (1.6)$$

where $\rho_i^{(k)}$ denotes the state of party k . Otherwise, at least some parts of the system are entangled.

From the above definition, we can conclude that multipartite entanglement can be further classified by identifying the number of separable parts. This concept is known as k -separability.

Definition 1.4. A multipartite, pure state $|\psi\rangle$ is called k -separable with $1 < k \leq N$ if the N parties can be divided into k parts P_1, \dots, P_k such that

$$|\psi\rangle = |\phi_1\rangle_{P_1} \otimes \dots \otimes |\phi_k\rangle_{P_k}, \quad (1.7)$$

where $|\phi_i\rangle_{P_i}$ is a state of part P_i . A mixed state ρ is correspondingly k -separable if it can be written as a convex combination of k -separable pure states. The pure states can be k -separable with respect to different partitions. In case a quantum state is neither k -separable for $k = 2, \dots, N$, it is N -partite entangled, which is referred to as **genuine multipartite entanglement (GME)**.

1.2 Graph states

Graph states are an important class of multiqubit states in quantum information theory [17, 29]. They are specifically useful as they describe a subset of entangled states.

Definition 1.5. Suppose $G = (V, E)$ is a graph with a set of n vertices V that are connected by a set of edges E . For each vertex i in G , we define a stabilizing operator g_i by

$$g_i := X_i \bigotimes_{j \in \mathcal{N}(i)} Z_j, \quad (1.8)$$

where $\mathcal{N}(i)$ is the neighborhood of vertex i , i.e., all vertices that are connected to i . In the above definition, X_i, Y_i and Z_i denote the Pauli matrices acting on qubit i . The graph state $|G\rangle$ that is associated with G is the common eigenstate of all stabilizing operators with eigenvalue $+1$, i.e.,

$$g_i |G\rangle = |G\rangle, \quad \text{for } i = 1, \dots, n. \quad (1.9)$$

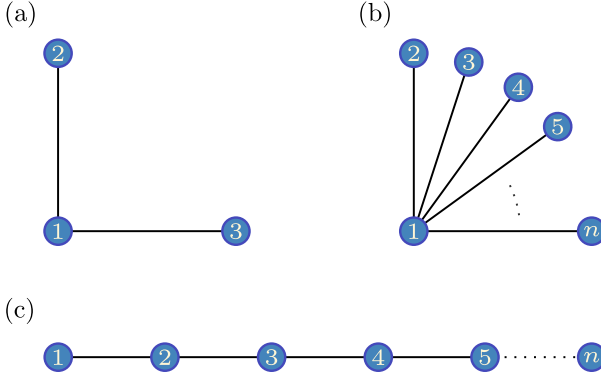


Figure 1.1: Three exemplary graph states: (a) For three qubits, the **GHZ** state and the **LC** state coincide. The star graph in (b) corresponds to the **GHZ** state of n qubits, whereas the line graph in (c) is associated to the **LC** state of n qubits. The figure is reprinted from [P4].

A few examples of graph states are shown in Fig. 1.1. Fig. 1.1 (a) is equivalent to the 3-qubit **Greenberger–Horne–Zeilinger (GHZ)**, whereas Fig. 1.1 (b) shows a graph for the n -qubit **GHZ** state. The graph state in Fig. 1.1 (c), in turn, is called **linear cluster (LC)** state. As a more detailed example we show that the graph state in Fig. 1.1 (b) is indeed equivalent to the n -qubit **GHZ** state.

Example 1.6. *As an example, we deduce the graph state that corresponds to the star graph in Fig. 1.1 (b). From Eq. (1.8), we get the stabilizers g_1, \dots, g_n , that are written in the left column below. By a local basis change for qubits $2, \dots, n$ that maps the eigenstates of the Pauli operator X to the eigenstates of Z , we can transform the stabilizers to $\tilde{g}_1, \dots, \tilde{g}_n$. The transformed stabilizers are shown in the right column.*

$$\begin{array}{ccc}
 g_1 = X_1 Z_2 Z_3 \dots Z_n & & \tilde{g}_1 = X_1 X_2 X_3 \dots X_n \\
 g_2 = Z_1 X_2 & \xrightarrow{\text{Local basis change}} & \tilde{g}_2 = Z_1 Z_2 \\
 g_3 = Z_1 X_3 & & \tilde{g}_3 = Z_1 Z_3 \\
 \vdots & & \vdots \\
 g_n = Z_1 X_n & & \tilde{g}_n = Z_1 Z_n
 \end{array}$$

For an eigenstate of $\tilde{g}_2, \dots, \tilde{g}_n$ the state of the qubits $2, \dots, n$ in the computational basis has to coincide with the state of the first qubit. The common eigenstate $|\tilde{G}\rangle$ is thus a superposition of the states $|00\dots 0\rangle$ and $|11\dots 1\rangle$. Since \tilde{g}_1 maps these states to each other, we see that both states have equal weights. We can thus conclude that the common eigenstate of the \tilde{g}_i is the **GHZ** state

$$|\tilde{G}\rangle = \frac{1}{\sqrt{2}}(|00\dots 0\rangle + |11\dots 1\rangle). \quad (1.10)$$

The graph state $|G\rangle$ that corresponds to the star graph in Fig. 1.1 (b) is thus up to local basis change equivalent to the **GHZ** state, i.e., they are **LU** equivalent.

Alternatively, the graph state $|G\rangle$ can also be constructed explicitly by the expression [29]

$$|G\rangle = \prod_{(i,j) \in E} \text{CZ}_{(i,j)} |+\rangle^{\otimes n}, \quad (1.11)$$

where E is the set of edges of the graph, $\text{CZ}_{(i,j)}$ is the controlled- Z gate acting on qubits i and j , and $|+\rangle = \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle)$.

Different graphs may lead to graph states that are connected by a **LU** transformation, which does not change the entanglement properties. A special class of **LU** transformations are local Clifford operations [29], that can be described by a graphical rule called local complementation.

Definition 1.7 (Local complementation). A local complementation LC_i in vertex i transforms a graph $G = (V, E)$ into a graph $G' = (V, E')$ by inverting the neighbourhood $\mathcal{N}(i)$ of vertex i . For two vertices $j, k \in \mathcal{N}(i)$, if $(j, k) \in E$ then $(j, k) \notin E'$ and vice versa. The set of vertices V is unchanged.

As an example, we show in Fig. 1.2 the graphs that are equivalent under local complementations to the 4-qubit star graph. We note that the neighborhood of the vertices 2, 3 and 4 only consists of vertex 1, i.e., $\mathcal{N}(i) = \{1\}$ for $i = 2, 3, 4$. As a result, a local complementation in these vertices leaves the graph unchanged. Only a local complementation in the center node 1 has a nontrivial effect. The neighborhood of the center node is $\mathcal{N}(1) = \{2, 3, 4\}$. As the vertices $\{2, 3, 4\}$ are not connected to each other, the local complementation LC_1 adds an edge between all the vertices $\{2, 3, 4\}$. The local complementation LC_1 thus results in the fully connected graph of four vertices in Fig. 1.2 (b). We note that the star graph can be retrieved from the fully connected graph by a local complementation in any vertex.

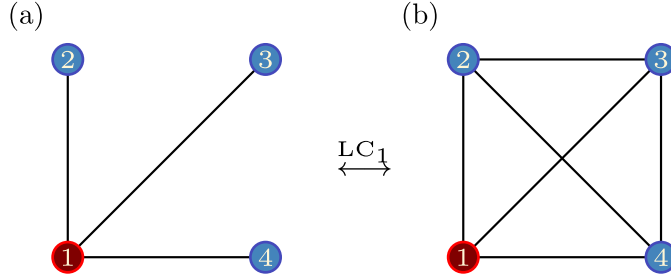


Figure 1.2: Local complementation of the 4-qubit star graph. The star graph in (a) is equivalent to the fully connected graph in (b). The graph (a) is transformed into the graph (b) and vice versa by a local complementation in vertex 1, i.e., LC_1 .

1.3 Measurements

So far, we have discussed the description of quantum states. We have seen that composite systems exhibit entangled states. The discussion was, however, limited to the theoretical aspects and detached from how properties of the states can be inferred by measurements. In this section, we therefore explain how measurements are described in quantum theory [28, 30].

Quantum mechanics postulates the measurement process that is often described as the collapse of the quantum state. The general form of a measurement is given by a set of K measurement operators A_k that fulfill the completeness relation [30]:

$$\sum_{k=1}^K A_k^\dagger A_k = \mathbb{1}. \quad (1.12)$$

Each of the operators A_k is associated to an outcome k of the measurement. In case the system is described by a state ρ , the probability to obtain result k is given by

$$p_k = \text{Tr} \left(A_k \rho A_k^\dagger \right). \quad (1.13)$$

Accordingly, the state collapses to

$$\rho_k = \frac{A_k \rho A_k^\dagger}{\text{Tr} \left(A_k \rho A_k^\dagger \right)}. \quad (1.14)$$

In the following, we will discuss two kinds of measurements that are commonly used: **projective measurements** and **positive operator valued measures (POVMs)**.

1.3.1 Projective measurements

The first type are **projective measurements**. In this case, the measurement operators A_k are orthogonal projectors, i.e., $A_k = P_k$. Orthogonal projectors are Hermitian operators with the property

$$P_k P_l = \delta_{kl} P_k. \quad (1.15)$$

Projective measurements are directly connected to observables and thus play an important role. An observable O in quantum mechanics is a Hermitian operator. As O is Hermitian, there exist a spectral decomposition [27]

$$O = \sum_i \lambda_i P_i, \quad (1.16)$$

where λ_i are the eigenvalues and P_i the projectors on the corresponding eigenspaces. The eigenvalues are identified with the possible measurement outcomes of the observable O and the projectors P_i as the measurement operators. The probability to obtain result λ_i is then given by Eq. (1.13), which takes the form

$$p_i = \text{Tr} \left(P_i \rho P_i^\dagger \right) = \text{Tr} (\rho P_i). \quad (1.17)$$

The expectation value of an observable is accordingly given by

$$\langle O \rangle = \sum_i p_i \lambda_i = \sum_i \text{Tr} (\rho P_i) \lambda_i = \text{Tr} (\rho O). \quad (1.18)$$

1.3.2 Positive operator valued measures (POVMs)

Another, more general type of measurement is a **POVM**. A **POVM** is a set of positive operators E_k that sum up to identity, i.e.,

$$\sum_{k=1}^K E_k = \mathbb{1} \quad \text{with} \quad E_k^\dagger = E_k, \quad E_k \geq 0 \quad \text{for} \quad k = 1, \dots, K. \quad (1.19)$$

The operators E_k are called effects and each effect is associated to a possible outcome of the **POVM**. The result k is observed with probability

$$p_k = \text{Tr} (\rho E_k). \quad (1.20)$$

We note, however, that a **POVM** does not fix the post-measurement state. The effects are related to the measurement operators by $E_k = A_k^\dagger A_k$. We can thus conclude that the effects E_k fix the measurement operators, but there is the freedom to choose a unitary, e.g., $A_k = U_k \sqrt{E_k}$.

As an example, we discuss two **POVMs** for a qubit. We note that for qubits, polyhedra inscribed in the Bloch sphere define **POVMs**. The vertices of the polyhedra correspond to the effects. We stress, however, that the effects have to be normalized to match the condition in Eq. (1.19). Fig. 1.3 shows two polyhedra that are commonly used: the tetrahedron and the octahedron. For the tetrahedron, the states that correspond to the Bloch vectors are given by

$$\begin{aligned} |t_1\rangle &= |0\rangle, & |t_2\rangle &= \frac{1}{\sqrt{3}} |0\rangle + \sqrt{\frac{2}{3}} |1\rangle, \\ |t_3\rangle &= \frac{1}{\sqrt{3}} |0\rangle + \sqrt{\frac{2}{3}} e^{2i\pi/3} |1\rangle, & |t_4\rangle &= \frac{1}{\sqrt{3}} |0\rangle + \sqrt{\frac{2}{3}} e^{4i\pi/3} |1\rangle. \end{aligned} \quad (1.21)$$

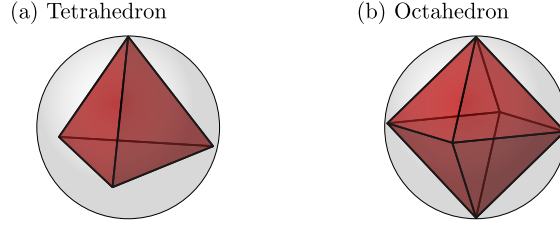


Figure 1.3: Examples of **POVMs** for a single qubit. For a qubit, **POVMs** can be described by polyhedra that are inscribed in the Bloch sphere. The effects are associated to the vertices of the polyhedra. The **POVM** described by the tetrahedron in (a) thus has four effects, whereas the octahedron in (b) defines a **POVM** with six effects. The octahedron describes the Pauli-**POVM**.

The set of effects is thus given by $E = \{\frac{1}{2} |t_1\rangle\langle t_1|, \frac{1}{2} |t_2\rangle\langle t_2|, \frac{1}{2} |t_3\rangle\langle t_3|, \frac{1}{2} |t_4\rangle\langle t_4|\}$.

The vertices of the octahedron can be identified with the eigenstates of the Pauli matrices, which we denote by $|x^\pm\rangle$ for the Pauli operator X . Correspondingly $|y^\pm\rangle$ and $|z^\pm\rangle$ refer to the eigenstates of Y and Z respectively. The normalized effects are then given by

$$E = \left\{ \frac{1}{3} |x^+\rangle\langle x^+|, \frac{1}{3} |x^-\rangle\langle x^-|, \frac{1}{3} |y^+\rangle\langle y^+|, \frac{1}{3} |y^-\rangle\langle y^-|, \frac{1}{3} |z^+\rangle\langle z^+|, \frac{1}{3} |z^-\rangle\langle z^-| \right\}. \quad (1.22)$$

For this reason the octahedron is associated to the Pauli-**POVM**.

To answer the question of how **POVMs** can be implemented, we first point out Naimark's theorem [28, 30].

Theorem 1.8 (Naimark's theorem). *For any **POVM** $\{E_k\}$ in a Hilbert space \mathcal{H} , there exists an isometry $A : \mathcal{H} \rightarrow \mathcal{H} \otimes \mathcal{H}_A$ to a larger Hilbert space $\mathcal{H} \otimes \mathcal{H}_A$ such that*

$$E_k = A^\dagger (\mathbb{1} \otimes |k\rangle\langle k|) A, \quad (1.23)$$

where $\{|k\rangle\}$ is a basis of the ancilla system \mathcal{H}_A .

Proof. The map $A : \mathcal{H} \rightarrow \mathcal{H} \otimes \mathcal{H}_A$, $A = \sum_k \sqrt{E_k} \otimes |k\rangle$ is an isometry as

$$A^\dagger A = \sum_{k,l} \sqrt{E_k} \sqrt{E_l} \langle k|l\rangle = \sum_k E_k = \mathbb{1}. \quad (1.24)$$

Moreover, A maps the projectors $\mathbb{1} \otimes |k\rangle\langle k|$ to the effects of the **POVM**, i.e.,

$$\begin{aligned} A^\dagger (\mathbb{1} \otimes |k\rangle\langle k|) A &= \left(\sum_m \sqrt{E_m} \otimes \langle m| \right) (\mathbb{1} \otimes |k\rangle\langle k|) \left(\sum_n \sqrt{E_n} \otimes |n\rangle \right) \\ &= \sum_m \sqrt{E_m} \sqrt{E_n} \langle m|k\rangle \langle k|n\rangle = E_k. \end{aligned} \quad (1.25)$$

□

Naimark's theorem provides a way to implement a **POVM**. For this the Hilbert space \mathcal{H} is extended by an ancilla system \mathcal{H}_A . The state $\rho \in \mathcal{H}$ is accordingly transformed to the dilated Hilbert space by $A\rho A^\dagger$. The projective measurement $\mathbb{1} \otimes |k\rangle\langle k|$ then yields the same statistics as the **POVM**, i.e.,

$$\text{Tr} ((\mathbb{1} \otimes |k\rangle\langle k|) A\rho A^\dagger) = \text{Tr} (A^\dagger (\mathbb{1} \otimes |k\rangle\langle k|) A\rho) = \text{Tr} (E_k \rho) = p_k. \quad (1.26)$$

We note that the projective measurement on the ancilla yields probabilistic results. In this regard, the **POVM** can be viewed as a randomized measurement, where the randomness is introduced by the ancilla system [30]. This raises the question whether a **POVM** can also be implemented by using a classical source of randomness. As an example, we consider the **Pauli-POVM** described by the octahedron in Fig. 1.3. The effects of this **POVM** consist of the eigenstates of the Pauli matrices. For this reason, we consider the projective measurement of the Pauli observables $M_\sigma = \{\sigma_x, \sigma_y, \sigma_z\}$. We moreover assume that the measurement implements the Pauli observables at random and with equal probability, i.e., $p_x = p_y = p_z = \frac{1}{3}$. With the knowledge about the performed measurement to possible outcomes are thus $\{x^\pm, y^\pm, z^\pm\}$. The probability to observe an outcome $k \in \{x^\pm, y^\pm, z^\pm\}$ is then given by the probability

$$p_k = p(\sigma(k))p(k|\sigma(k)) = \frac{1}{3} \text{Tr}(\rho |k\rangle\langle k|) = \text{Tr}\left(\rho \frac{1}{3} |k\rangle\langle k|\right) = \text{Tr}(\rho E_k), \quad (1.27)$$

where the Pauli observable $\sigma(k)$ that is associated to result k is chosen with probability $p(\sigma(k))$ and the conditional probability to subsequently obtain the outcome k is $p(k|\sigma(k))$. In the above equation, we have identified the effects of the **Pauli-POVM** that are given in Eq. (1.22). The probabilities coincide with the statistics of the **Pauli-POVM** and we can conclude that the **Pauli-POVM** can be simulated by classical randomization of projective measurements. We note, however, that not all **POVMs** can be simulated classically by randomization and post-processing [31].

1.4 Entanglement detection

We now elaborate on how to decide whether a state $|\psi\rangle$ is entangled or not. There are various methods to detect entanglement [17]. On the one hand there are criteria that require knowledge of the density matrix, like the criteria of the **positive partial transpose (PPT)** [32, 33] or the **computable cross norm or realignment (CCNR)** criterion [34, 35]. The **PPT** criterion needs insight in the eigenvalues of the partial transpose of the density matrix, whereas the **CCNR** criterion is formulated in terms of the singular values of the density matrix. As for large quantum systems it is infeasible to reconstruct the density matrix [16], alternative methods have to be used. In this section, we discuss methods that require only the measurement of specific observables. First, we will discuss entanglement witnesses. But we will also introduce spin-squeezing inequalities, that rely on the expectation values of the total angular momenta, and Bell inequalities, which for graph states involve the expectation values of stabilizer operators.

1.4.1 Entanglement witnesses

An observable whose expectation value is positive for all separable states is referred to as an entanglement witness. This is formalized in the definition below [17].

Definition 1.9. An entanglement witness \mathcal{W} is an observable that fulfills:

- (i) $\text{Tr}(\rho\mathcal{W}) \geq 0$ for all separable states,
- (ii) $\text{Tr}(\rho\mathcal{W}) < 0$ for at least one entangled state.

Measuring a negative expectation value $\langle \mathcal{W} \rangle_\rho < 0$ classifies the state ρ as entangled. From a geometric point of view, an entanglement witness is a hyperplane that splits the space of density matrices into two parts: the sector with $\text{Tr}(\rho\mathcal{W}) \geq 0$ and the one with $\text{Tr}(\rho\mathcal{W}) < 0$. This is depicted in Fig. 1.4. The subset of separable states is completely in the part with $\text{Tr}(\rho\mathcal{W}) \geq 0$. In contrast, the region of entangled state for which $\text{Tr}(\rho\mathcal{W}) < 0$ is detected as entangled by the

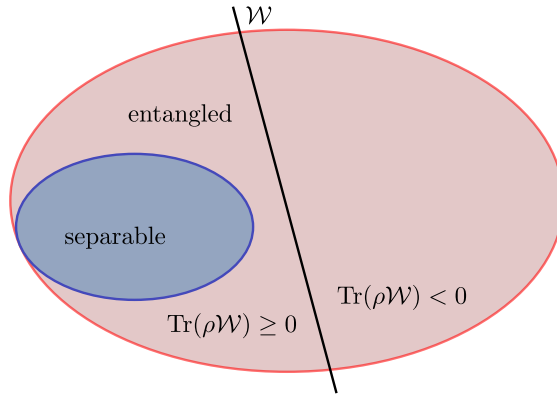


Figure 1.4: Schematic representation of an entanglement witness \mathcal{W} . \mathcal{W} splits the set of density matrices into two parts with $\text{Tr}(\rho\mathcal{W}) \geq 0$ and $\text{Tr}(\rho\mathcal{W}) < 0$, respectively. The subset of separable states are in the section with $\text{Tr}(\rho\mathcal{W}) \geq 0$ and all states in the part with $\text{Tr}(\rho\mathcal{W}) < 0$ are detected as entangled.

witness \mathcal{W} . The geometric picture moreover suggests that it is possible to find a hyperplane for any entangled state that separates it from the set of separable states, i.e., for any entangled state there exists a witness that detects the state as entangled. This was formally shown in [33].

As a first example, we discuss fidelity witnesses. Fidelity witnesses for an entangled state $|\psi\rangle$ have the form [17, 36]

$$\mathcal{W} = \alpha\mathbb{1} - |\psi\rangle\langle\psi|. \quad (1.28)$$

Depending on the different kinds of entanglement that should be witnessed, α has to be chosen accordingly. For example, if one likes to detect that a state is not fully separable, α has to be chosen as the maximal fidelity of a fully separable state with the target state $|\psi\rangle$, i.e., $\alpha = \max_{\rho \in \text{SEP}} \text{Tr}(\rho\mathcal{W})$, where SEP denotes the set of separable states. We note that this results in a witness that detects any entanglement in the state. In case, the witness should detect genuine n -partite entanglement, α has to be chosen as the maximal fidelity with the target state $|\psi\rangle$ that can be achieved with biseparable states, i.e., $\alpha = \max_{\rho \in \text{BISEP}} \text{Tr}(\rho\mathcal{W})$. BISEP refers to the set of biseparable, i.e., states that are 2-separable.

For a graph state $|G\rangle$ the maximal fidelity with a biseparable state is upper bounded by $\frac{1}{2}$ [37]. The witness

$$\mathcal{W} = \frac{1}{2}\mathbb{1} - |G\rangle\langle G| \quad (1.29)$$

thus detects GME [38]. We conclude this section by a brief description of how this witness can be measured. The projector on the graph state $|G\rangle$ can be expressed in terms of all stabilizers S of the stabilizer group \mathcal{S} :

$$|G\rangle\langle G| = \frac{1}{2^n} \sum_{S \in \mathcal{S}} S. \quad (1.30)$$

It is thus possible to evaluate the witness in Eq. (1.29) by measuring all stabilizers, which each constitute a string of Pauli operators. In this regard, the approach only relies on local measurements. The stabilizer group \mathcal{S} , however, consists of 2^n elements, where n is the number of qubits of the graph state. For this reason, it becomes quickly infeasible to measure all stabilizers with an increasing number of qubits n . A possible solution is to use a randomized measurement scheme. Ref. [18] introduces a method that is based on random sampling of Pauli strings to estimate the

fidelity. The Pauli strings are chosen according to a probability distribution that is determined by the target state. Especially, they show that for stabilizer states the total number of state samples is independent of the system size. Another approach is to find entanglement witnesses that require by construction less measurement resources. For example the stabilizer sum witness

$$\mathcal{W} = (n - 1)\mathbb{1} - \sum_{k=1}^n g_k \quad (1.31)$$

detects GME just from the measurement of the n generators g_k of the stabilizer group [36, 38].

1.4.2 Spin-squeezing inequalities

Another approach to detect entanglement uses spin-squeezing inequalities that are especially useful in n -partite systems with large n . The presentation in this section follows [P3]. Spin-squeezing was first introduced in the context of metrology. It was recognized that the precision of a measurement can be increased with the help of spin-squeezed states [9, 10, 39–42]. The conditions for spin-squeezing are commonly expressed in terms of the first and second moments of the angular momentum operator. In direction $\alpha = x, y, z$, the angular momentum operator of an n -qubit system is given by

$$J_\alpha = \frac{1}{2} \sum_{i=1}^n \sigma_\alpha^{(i)}, \quad (1.32)$$

where $\sigma_\alpha^{(i)}$ is the corresponding Pauli matrix for qubit i . In a simplified view, the variance in a direction orthogonal to the mean spin direction is reduced for a spin-squeezed state. This is reflected for instance by the spin-squeezing parameter ξ_R in Ref. [10]. A state that fulfills

$$\xi_R^2 = \frac{n(\Delta J_{\mathbf{n}_\perp})^2}{|\langle \mathbf{J} \rangle|^2} < 1 \quad (1.33)$$

is called spin-squeezed. In the above equation, $\langle \mathbf{J} \rangle = (\langle J_x \rangle, \langle J_y \rangle, \langle J_z \rangle)$ is the mean spin direction, i.e., the expectation value of the angular momentum vector. $(\Delta J_{\mathbf{n}_\perp})^2$ denotes the smallest variance of the spin in a direction \mathbf{n}_\perp orthogonal to the mean spin direction. This definition is similar to squeezed states of light, where the state also exhibits a reduced variance in a direction in phase space [43]. In Ref. [10], spin-squeezed states are used to improve the sensitivity in Ramsey spectroscopy. Experimentally, spin-squeezed states have been successfully prepared in atomic ensembles and especially in Bose-Einstein condensates [44–51].

The advantage of spin-squeezed states in metrology is due to quantum-mechanical correlations between the particles [9]. A connection to entanglement was first shown in Ref. [52]. All states that are fully separable fulfill the inequality

$$\frac{n(\Delta J_z)^2}{\langle J_x \rangle^2 + \langle J_y \rangle^2} \geq 1. \quad (1.34)$$

Thus, a violation implies that the state is entangled. After the formulation of the first spin-squeezing inequality, many other criteria have been found [53–57]. For large systems quantum state tomography quickly becomes infeasible [16]. Spin-squeezing inequalities, however, have the advantage that they also can be directly determined by measuring the first and second moment of the total angular momentum operator, which requires fewer measurements than quantum state tomography [58]. Spin-squeezing inequalities have thus been readily applied in experiments [52, 59–62]. For example, the original spin-squeezing inequality is introduced in Ref. [52] to verify entanglement in a Bose-Einstein condensate.

After the discovery of the original spin-squeezing inequality given by Eq. (1.34), refined spin-squeezing inequalities have been found. Here, we will focus on the optimal spin-squeezing inequalities formulated in Refs. [56, 57]. The aim of these inequalities is to detect entanglement and thus they differ from the spin-squeezing inequalities used in metrology. Though, they are of the same form and use both the first and second moments of the angular momentum operators given in Eq. (1.32):

$$\langle J_x^2 \rangle + \langle J_y^2 \rangle + \langle J_z^2 \rangle \leq \frac{n(n+2)}{4}, \quad (1.35a)$$

$$(\Delta J_x)^2 + (\Delta J_y)^2 + (\Delta J_z)^2 \geq \frac{n}{2}, \quad (1.35b)$$

$$\langle J_k^2 \rangle + \langle J_l^2 \rangle - \frac{n}{2} \leq (n-1)(\Delta J_m)^2, \quad (1.35c)$$

$$(n-1) [(\Delta J_k)^2 + (\Delta J_l)^2] \geq \langle J_m^2 \rangle + \frac{n(n-2)}{4}. \quad (1.35d)$$

The above inequalities are fulfilled by all fully separable states that are defined in Def. 1.3. Due to the limited information in terms of the first and second moments of the total angular momenta, the characterization of entanglement is not complete. The optimal spin-squeezing inequalities [56, 57] detect the maximal amount of entangled states that can be extracted from the first and second moments of the total angular momenta. In contrast to entanglement witnesses, however, spin-squeezing inequalities are usually non-linear in the quantum state as they involve second moments. In Eq. (1.35), (k, l, m) is a permutation of (x, y, z) . Whereas Eq. (1.35a) is valid for all quantum states, a violation of Eqs. (1.35b)–(1.35d) implies entanglement. For fixed mean spin $(\langle J_x \rangle, \langle J_y \rangle, \langle J_z \rangle)$, the inequalities in Eq. (1.35) define a polytope in the space of $(\langle J_x^2 \rangle, \langle J_y^2 \rangle, \langle J_z^2 \rangle)$ [57]. This is exemplary shown for $(\langle J_x \rangle, \langle J_y \rangle, \langle J_z \rangle) = (0, 0, 0)$ in Fig. 1.5. In the limit $N \rightarrow \infty$ but also for special cases of finite N , there exists a separable state for all points inside the polytope. In these cases the inequalities identify the maximal amount of entangled states that can be detected by the first and second moments of the angular momentum operators.

For example, Eq. (1.35b) detects many-body singlet states as entangled [57]. These states are eigenstates of the total angular momentum \vec{J} with eigenvalue zero, i.e.,

$$\begin{aligned} (\langle J_x \rangle, \langle J_y \rangle, \langle J_z \rangle) &= (0, 0, 0), \\ (\langle J_x^2 \rangle, \langle J_y^2 \rangle, \langle J_z^2 \rangle) &= (0, 0, 0). \end{aligned} \quad (1.36)$$

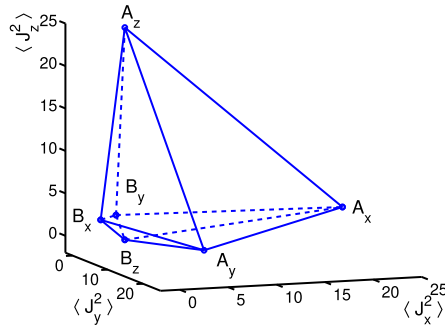


Figure 1.5: Polytope in the space of $(\langle J_x^2 \rangle, \langle J_y^2 \rangle, \langle J_z^2 \rangle)$ defined by the spin-squeezing inequalities in Eq. (1.35). For $(\langle J_x \rangle, \langle J_y \rangle, \langle J_z \rangle) = (0, 0, 0)$ all separable state lie within the blue polytope. The figure is reprinted from [57].

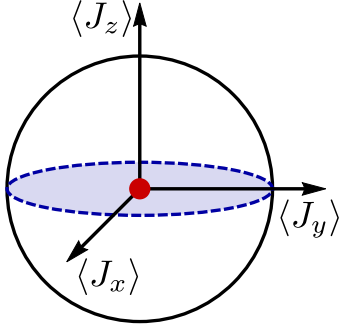


Figure 1.6: Visualization of the singlet state $|\Psi^-\rangle$ (red) and the Dicke state $|D_{N,N/2}\rangle$ (blue) on the collective Bloch sphere [63]. Singlet states are characterized by vanishing mean spin $\langle \vec{J} \rangle = 0$ and variances. Hence, the singlet state $|\Psi^-\rangle$ corresponds to the red dot at the origin. The Dicke state $|D_{N,N/2}\rangle$ is also at the origin, though it has a nonzero variance in the x - y plane that is shown by the blue shaded area. The figure is reprinted from [P3].

As the left-hand side of Eq. (1.35b) is non-negative, we see that Eq. (1.35b) is maximally violated by the many-body singlet states. To visualize the state, we can make use of the collective Bloch sphere [63]. In this representation the mean spin of the states is plotted. In addition, the variances $(\Delta J_x)^2$, $(\Delta J_y)^2$, $(\Delta J_z)^2$ are used to give the uncertainty. For many-body singlet states, both the mean spin and the variances are zero. Hence, many-body singlet states correspond to the red dot at the origin in Fig. 1.6. A specific example of a many-body singlet state for an even number of qubits N is given by

$$|\Psi^-\rangle = \bigotimes_{k=1}^{N/2} |\psi^-\rangle, \quad (1.37)$$

with the two-qubit singlet state $|\psi^-\rangle = \frac{1}{\sqrt{2}}(|01\rangle - |10\rangle)$. We show in App. A.3.1 that $|\Psi^-\rangle$ indeed fulfills the defining property in Eq. (1.36).

The inequality in Eq. (1.35c) is in turn maximally violated by the symmetric Dicke state $|D_{N,N/2}\rangle$ with $m = \frac{N}{2}$ excitations [57]. For m excitations, the symmetric Dicke state is defined as

$$|D_{N,m}\rangle = \binom{N}{m}^{-\frac{1}{2}} \sum_k P_k(|1_1, \dots, 1_m, 0_{m+1}, \dots, 0_N\rangle), \quad (1.38)$$

where the sum iterates over all distinct permutations P_k of the qubits and $\binom{N}{m}$ denotes the binomial coefficient. The first and second moments of the angular momenta of the Dicke states are

$$\begin{aligned} (\langle J_x \rangle, \langle J_y \rangle, \langle J_z \rangle) &= \left(0, 0, \frac{N}{2} - m \right), \\ (\langle J_x^2 \rangle, \langle J_y^2 \rangle, \langle J_z^2 \rangle) &= \left(\frac{N}{4} + \frac{m(N-m)}{2}, \frac{N}{4} + \frac{m(N-m)}{2}, \left[\frac{N}{2} - m \right]^2 \right). \end{aligned} \quad (1.39)$$

Hence, the Dicke state $|D_{N,N/2}\rangle$ with $m = \frac{N}{2}$ excitations is also located at the origin of the collective Bloch sphere. However, the variances $(\Delta J_x)^2$ and $(\Delta J_y)^2$ are nonzero. They are of the order $\mathcal{O}(N^2)$, which is of the same magnitude as the radius of the Bloch sphere. This is shown as the blue circle in Fig. 1.6.

Finally, many-body singlet states violate also the fourth spin-squeezing inequality (1.35d) [57].

1.4.3 Bell inequalities

Another approach to detect entanglement are Bell inequalities. Though, the primary goal of Bell inequalities is to exclude nonclassical correlations. In this section, we first consider the bipartite

case and discuss the CHSH inequality as an example. In doing so, we point out the relation to entanglement. Finally, we discuss multipartite Bell inequalities.

Bipartite case

Let us consider the bipartite Bell scenario in Fig. 1.7. The two parts are referred to as Alice (A) and Bob (B). The source S distributes an entangled pair of particles between Alice and Bob. Both, Alice and Bob can then choose their setting i and j to perform the corresponding measurement on the particles. The outcomes of the measurements are denoted a and b , respectively. In this section, we will consider the scenario that both parties can choose from two measurement settings, i.e., Alice can choose between the measurements A_1, A_2 and Bob between B_1 and B_2 . All observables are, moreover, dichotomic, i.e., they exhibit two possible outcomes $a, b \in \{-1, 1\}$. In the following we provide the argument for the CHSH inequality as presented in Ref. [64]. For the derivation, we need the following properties of a probability measure \mathbb{P} . Probability measures are introduced in Sec. 2.1.

Consider two events X and Y . We say X implies Y , i.e., $X \Rightarrow Y$, if $X \subseteq Y$. It thus holds

$$\begin{aligned} (X \Rightarrow Y) &\Rightarrow \mathbb{P}(X) \leq \mathbb{P}(Y), \\ \mathbb{P}(X \cup Y) &\leq \mathbb{P}(X) + \mathbb{P}(Y). \end{aligned} \tag{1.40}$$

Under the assumptions that the measurement setting i and outcome a do not influence the measurement B_j and vice versa (locality) the following implication for the observables A_i and B_j holds:

$$(A_1 = B_1) \cap (A_1 = B_2) \cap (A_2 = B_1) \Rightarrow (A_2 = B_2). \tag{1.41}$$

The CHSH inequality can be derived from the negation

$$(A_2 \neq B_2) \Rightarrow (A_1 \neq B_1) \cup (A_1 \neq B_2) \cup (A_2 \neq B_1). \tag{1.42}$$

From Eq. (1.40) follows for the probability

$$\mathbb{P}(A_2 \neq B_2) \leq \mathbb{P}(A_1 \neq B_1) + \mathbb{P}(A_1 \neq B_2) + \mathbb{P}(A_2 \neq B_1). \tag{1.43}$$

To arrive at the inequality in terms of expectation values, we note that $\mathbb{P}(A_i \neq B_j) = 1 - \mathbb{P}(A_i = B_j)$. The above inequality can thus be equivalently written as

$$-\mathbb{P}(A_2 = B_2) \leq 2 - \mathbb{P}(A_1 = B_1) - \mathbb{P}(A_1 = B_2) - \mathbb{P}(A_2 = B_1). \tag{1.44}$$

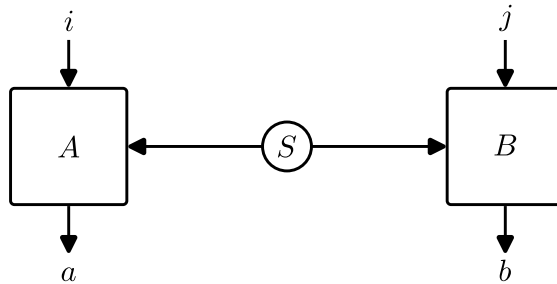


Figure 1.7: Bipartite Bell scenario. The source S prepares an entangled qubit pair and sends one qubit to each party, Alice (A) and Bob (B), respectively. Alice and Bob both can choose their measurement settings i and j and obtain the outcomes a and b .

The sum of Eqs. (1.43) and (1.44) thus yields

$$\begin{aligned} & [\mathbb{P}(A_1 = B_1) - \mathbb{P}(A_1 \neq B_1)] + [\mathbb{P}(A_1 = B_2) - \mathbb{P}(A_1 \neq B_2)] \\ & + [\mathbb{P}(A_2 = B_1) - \mathbb{P}(A_2 \neq B_1)] - [\mathbb{P}(A_2 = B_2) - \mathbb{P}(A_2 \neq B_2)] \leq 2. \end{aligned} \quad (1.45)$$

As A_i and B_j are dichotomic observables with outcomes $\{\pm 1\}$, we have $A_i B_j = 1$ for $A_i = B_j$ and $A_i B_j = -1$ for $A_i \neq B_j$. The expressions in square brackets are thus the expectation values, which yields the CHSH inequality

$$\langle A_1 B_1 \rangle + \langle A_1 B_2 \rangle + \langle A_2 B_1 \rangle - \langle A_2 B_2 \rangle \leq 2. \quad (1.46)$$

The CHSH inequality is derived under the assumption of locality. The idea of a local theory can be further formalized [7]. For this, one can define a variable λ that describes all information that can have a causal influence on both Alice and Bob. The measurements Alice and Bob perform on their system on the other hand are space-like separated and cannot influence each other under the locality assumption. The probabilities thus factorize in the sense

$$\mathbb{P}(ab|ij, \lambda) = \mathbb{P}(a|i, \lambda)\mathbb{P}(b|j, \lambda). \quad (1.47)$$

By integrating over the probability distribution $\mathbb{P}(\lambda)$ we can derive the probabilities to obtain results a, b when Alice and Bob have chosen the measurement settings i, j , i.e.,

$$\mathbb{P}(ab|ij) = \int d\mathbb{P}(\lambda)\mathbb{P}(a|i, \lambda)\mathbb{P}(b|j, \lambda). \quad (1.48)$$

A local theory that satisfies the above equation is also referred to as **LHV** model.

Finally, we can consider the probability distribution that is generated by a separable bipartite state $\rho^{AB} = \sum_k p_k \rho_k^A \otimes \rho_k^B$ defined in Def. 1.2. We consider projective measurements and denote the projectors for the outcome a, b of the observables A_i, B_j as $M_{a|i}$ and $M_{b|j}$, respectively. The probability distribution for a separable state is thus

$$\mathbb{P}(ab|ij) = \text{Tr}(\rho^{AB} M_{a|i} \otimes M_{b|j}) = \sum_k p_k \text{Tr}(\rho_k^A M_{a|i}) \text{Tr}(\rho_k^B M_{b|j}), \quad (1.49)$$

which with the identification $\mathbb{P}(a|i, \lambda) = \text{Tr}(\rho_\lambda^A M_{a|i})$ and $\mathbb{P}(b|j, \lambda) = \text{Tr}(\rho_\lambda^B M_{b|j})$ is of the form of a local theory in Eq. (1.48). We can conclude that the statistics of a separable state can be described by a local theory. Conversely, if the statistics cannot be written in the form of Eq. (1.48), e.g., if a violation of a Bell inequality is observed, the state must be entangled.

Multipartite case

The notion of a local theory can be extended to n parties. The parties can choose their measurement settings i_1, \dots, i_n and accordingly measure the observable A_{i_1}, \dots, A_{i_n} . We denote the outcomes of the measurements as a_1, \dots, a_n . The probability distribution in a local theory of n parties takes the form [65]

$$\mathbb{P}(a_1 \dots a_n | i_1 \dots i_n) = \int d\mathbb{P}(\lambda)\mathbb{P}(a_1 | i_1, \lambda) \dots \mathbb{P}(a_n | i_n, \lambda). \quad (1.50)$$

Ref. [65] derived a Bell inequality that is commonly referred to as Mermin's inequality:

$$\mathcal{B}^M = \frac{1}{2^i} [(X + iY)^{\otimes n} - (X - iY)^{\otimes n}]. \quad (1.51)$$

The classical bound that can be achieved by a **LHV** model is given by

$$\langle \mathcal{B}^M \rangle_{\text{LHV}} \leq C^M = \begin{cases} 2^{n/2}, & \text{for even } n \\ 2^{(n-1)/2}, & \text{for odd } n. \end{cases} \quad (1.52)$$

The maximal value of Mermin's inequality is achieved for the **GHZ** state with $Q^M = \langle \mathcal{B}^M \rangle_{\text{GHZ}} = 2^{n-1}$. Mermin's inequality thus exhibits a relative violation $D = Q^M/C^M$ that increases exponentially with n . This makes Mermin's inequality more resistant to noise. Let us for example consider the effect of depolarisation noise, i.e., the state is given by $\rho = p|\text{GHZ}\rangle\langle\text{GHZ}| + (1-p)\mathbb{1}/2^n$. From Eq. (1.51) it is apparent that Mermin's inequality only contains full correlation terms. As a result, the expectation value of the maximally mixed state is zero and we get

$$\langle \mathcal{B}^M \rangle_\rho = p \times Q^M. \quad (1.53)$$

To observe a violation it is thus sufficient to prepare the **GHZ** state with probability $p > C^M/Q^M = 2^{-\frac{n-2}{2}} (2^{-\frac{n-1}{2}})$ for even (odd) n . With increasing number of qubits, Mermin's inequality can thus tolerate an exponentially increasing amount of noise.

We now shift the discussion to Bell inequalities for graph states. We note that all graph states that contain at least one edge violate the Bell inequality that consists of all stabilizing operators [66], i.e.,

$$\mathcal{B} = \sum_{S \in \mathcal{S}} S. \quad (1.54)$$

In the above Bell operator, \mathcal{S} denotes the stabilizer group. We have seen above that the power of a Bell inequality can be measured in terms of the relative violation. In this sense, however, the Bell operator in Eq. (1.54) is not optimal for many graph states [67]. Instead Bell operators that only consist of a subset of the stabilizers often yield a higher relative violation. For example, also Mermin's inequality in Eq. (1.51) can be written in terms of the stabilizers of the **GHZ** state. The Bell operator

$$\mathcal{B}_n^{\text{GHZ}} = g_1 \prod_{i=2}^n (\mathbb{1} + g_i) \quad (1.55)$$

is up to local unitary transformations equivalent to Mermin's inequality in Eq. (1.51) [17]. We stress that $\mathcal{B}_n^{\text{GHZ}}$ only contains half of the stabilizers as opposed to all 2^{n-1} stabilizers in Eq. (1.54).

We have seen in Sec. 1.2 that for three qubits the **GHZ** and the **LC** state can be represented by the same graph state. Therefore, in the case of three qubits both states are equivalent up to local unitary transformations. Based on this, it is possible to derive a Bell inequality that is maximally violated by the **LC** state of three qubits. In terms of the stabilizers of the **LC** state, the Bell inequality reads [67–70]

$$\mathcal{B}_{n=3}^{\text{LC}} = (\mathbb{1} + g_1)g_2(\mathbb{1} + g_3). \quad (1.56)$$

It has been shown in [71] that the above Bell inequality can be generalized to **LC** states which contain a number of qubits that is a multiple of three. The generalized Bell inequality reads

$$\mathcal{B}_n^{\text{LC}} = \prod_{i=1}^{n/3} (\mathbb{1} + g_{3i-2})g_{3i-1}(\mathbb{1} + g_{3i}) \quad (1.57)$$

and the upper bound for **LHV** models is given by

$$\langle \mathcal{B}_n^{\text{LC}} \rangle_{\text{LHV}} \leq 2^{n/3} = C^{\text{LC}}. \quad (1.58)$$

As the Bell operator in Eq. (1.57) is expressed in terms of stabilizers that have a maximal eigenvalue of 1, we observe that the maximal violation is indeed obtained for the **LC** state with $Q^{\text{LC}} = \langle \mathcal{B}_n^{\text{LC}} \rangle_{\text{LC}} = 4^{n/3}$. We conclude that also the Bell inequality for the **LC** state in Eq. 1.57 allows for an exponential violation of the classical bound as $Q^{\text{LC}}/C^{\text{LC}} = 2^{n/3}$.

As a final remark, we point out that a violation of the Bell inequalities in this section does not imply genuine multipartite nonlocality. A violation of the Bell inequalities is a contradiction to the multipartite **LHV** model in Eq. (1.50). This model can already be violated in case two parties share nonclassical correlations. A generalization to genuine multipartite nonlocality is established by Svetlichny's inequality [72].

1.5 Randomized measurements

All entanglement criteria that we have introduced so far use explicit measurement settings. In this section we will discuss to what extent a quantum system can be characterized by randomized measurements. This approach relies on performing measurements on the system in random bases. We first motivate the detection of entanglement with randomized measurements and describe the advantages of the approach. Accordingly, we formalize the method. In doing so, we discuss that the outcome distribution of randomized measurements can be characterized by its moments. Finally, we introduce **LU** invariants and discuss the complete set of **LU** invariants for two qubits.

1.5.1 Motivation

To motivate the use of randomized measurements in the detection of entanglement, we first note that the Bell state $|\psi^-\rangle = 1/\sqrt{2}(|01\rangle - |10\rangle)$ is invariant under unitary transformations of the form $U \otimes U$, i.e.,

$$U \otimes U |\psi^-\rangle = e^{i\phi} |\psi^-\rangle \quad (1.59)$$

for some phase $\phi \in [-\pi, \pi)$. The measurement outcomes of the Bell state $|\psi^-\rangle$ are thus perfectly correlated in any basis as long as the basis coincides for both parties. In fact, it is a general property of entanglement to be invariant under **LU** transformations [17]. This poses the question, what we can learn in case the measurement bases of the two parties are chosen independently.

For this purpose, we consider the observable

$$\sigma_{\mathbf{u}_1}^{(A)} \otimes \sigma_{\mathbf{u}_2}^{(B)}. \quad (1.60)$$

$\sigma_{\mathbf{u}} = \mathbf{u} \cdot \boldsymbol{\sigma}$ denotes a rotated Pauli matrix with eigenbasis $|\mathbf{u}_{\pm}\rangle$ that corresponds to the points $\pm\mathbf{u}$ on the Bloch sphere. We choose $\mathbf{u}_1, \mathbf{u}_2$ at random according to the uniform distribution on the Bloch sphere. The observable in Eq. (1.60) thus corresponds to measurements in random bases. Fig. 1.8 shows two outcome distributions of the observable $\sigma_{\mathbf{u}_1}^{(A)} \otimes \sigma_{\mathbf{u}_2}^{(B)}$ for 10000 random directions \mathbf{u}_1 and \mathbf{u}_2 . The measurements are simulated for the product state $|00\rangle$ and the maximally entangled Bell state $|\psi^-\rangle$. The distributions of the two states in Fig. 1.8 differ significantly. This shows that the outcome distributions of randomized measurements can be used to characterize properties of the state at least to some extent. And it turns out that especially entanglement can be detected by randomized measurements [25].

We conclude this section by discussing the advantages of randomized measurements. On the one hand, randomized measurements do not require to align the measurement bases [23, 25]. This is especially relevant in case the measurements are performed on distant parties, when it is challenging to align the reference frames.

On the other hand, the approach is insensitive to specific kinds of noise [73–75]. In case randomized measurements are chosen according to the Haar measure, i.e., the measurement

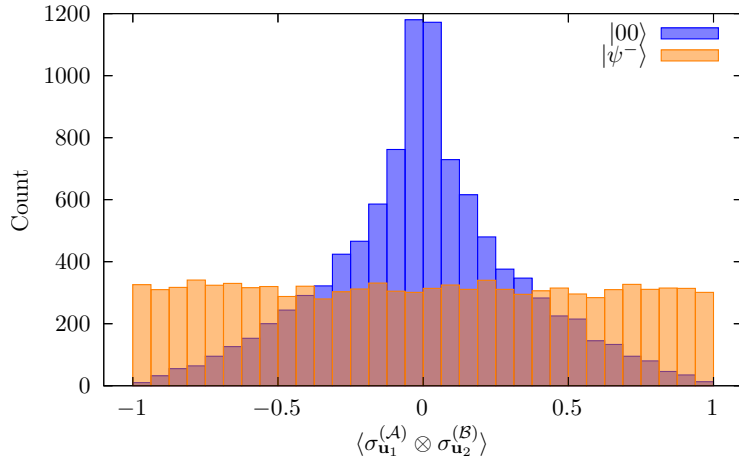


Figure 1.8: Binned outcome distribution of the observable $\sigma_{\mathbf{u}_1}^{(A)} \otimes \sigma_{\mathbf{u}_2}^{(B)}$ for 10000 random directions \mathbf{u}_1 and \mathbf{u}_2 . The distribution is shown for the product state $|00\rangle$ and the Bell state $|\psi^-\rangle$.

bases are sampled uniformly, the approach is insensitive to unitary noise [25]. This is useful as the effect of a noisy environment can often be described by random unitary rotations. This, however, does not change the sampling of the measurements.

Finally, randomized measurements allow to directly infer properties like entanglement [23, 24]. It is thus not necessary to reconstruct the density matrix of the system. As a result the method scales well to large system sizes.

1.5.2 Moments of the outcome distributions

In this section, we are interested in randomized measurements where the measurement on each part is rotated by a **LU** transformation. We assume a quantum system of n parts described by a density matrix ρ . The random sampling of the local measurement bases can equivalently be described by **LU** transformations of the quantum state, i.e., $\rho \mapsto (U_1 \otimes \dots \otimes U_n) \rho (U_1^\dagger \otimes \dots \otimes U_n^\dagger)$. Accordingly, an observable \mathcal{M} is measured. The outcome distribution can be described by its moments

$$\mathcal{R}_{\mathcal{M}}^{(t)}(\rho) := \int dU_1 \dots \int dU_n \text{Tr} \left[(U_1 \otimes \dots \otimes U_n) \rho (U_1^\dagger \otimes \dots \otimes U_n^\dagger) \mathcal{M} \right]^t. \quad (1.61)$$

Typically, the randomized measurement schemes are formulated for product observables [P2], i.e., observables of the form

$$\mathcal{M} = M_1 \otimes \dots \otimes M_n. \quad (1.62)$$

We note, however, that a general observable can be decomposed into product observables. The results for non-product observables can thus be derived from the data of product observables and classical postprocessing.

As an example, we illustrate how the state of two qubits can be characterized by the moments of the outcome distribution in Fig. 1.8. Fig. 1.8 shows the distribution of the observable $\sigma_{\mathbf{u}_1}^{(A)} \otimes \sigma_{\mathbf{u}_2}^{(B)}$ where the directions \mathbf{u}_1 and \mathbf{u}_2 are sampled uniformly on the Bloch sphere. We note that an arbitrary unitary U that acts on a qubit can be decomposed in a phase factor and a rotation by angle ϕ around the direction given by \mathbf{n} , i.e., $U = e^{i\alpha} R_{\mathbf{n}}(\phi)$ [76]. The rotation in turn is defined

as $R_{\mathbf{n}}(\phi) = \exp(-i\phi\mathbf{n} \cdot \boldsymbol{\sigma}/2)$. This yields that for a unitary U there exists a unit vector $\mathbf{u} \in \mathbb{R}^3$ such that $U^\dagger \sigma_z U = \mathbf{u} \cdot \boldsymbol{\sigma}$. The expectation value $\langle \sigma_{\mathbf{u}_1}^{(A)} \otimes \sigma_{\mathbf{u}_2}^{(B)} \rangle$ can thus be written as

$$\langle \sigma_{\mathbf{u}_1}^{(A)} \otimes \sigma_{\mathbf{u}_2}^{(B)} \rangle = \text{Tr} \left[(U_1 \otimes U_2) \rho (U_1^\dagger \otimes U_2^\dagger) \sigma_z \otimes \sigma_z \right]. \quad (1.63)$$

We thus see that the measurement $\sigma_{\mathbf{u}_1}^{(A)} \otimes \sigma_{\mathbf{u}_2}^{(B)}$ fits into the description with random unitaries. Accordingly, the second moment can also be written in terms of integrals over the Bloch sphere S^2 :

$$\mathcal{R}^{(2)}(\rho) = \frac{1}{(4\pi)^2} \int_{S^2} \int_{S^2} \langle \sigma_{\mathbf{u}_1}^{(A)} \otimes \sigma_{\mathbf{u}_2}^{(B)} \rangle^2 d\mathbf{u}_1 d\mathbf{u}_2. \quad (1.64)$$

It has been shown that for separable states it is $\mathcal{R}^{(2)} \leq 1/3^2$ [22]. The second moment thus yields the entanglement criterion

$$\mathcal{R}^{(2)}(\rho) > \frac{1}{3^2} \quad \Rightarrow \quad \rho \text{ is entangled.} \quad (1.65)$$

1.5.3 Local unitary invariants

The moments $\mathcal{R}_{\mathcal{M}}^{(t)}(\rho)$ are invariant under **LU** transformations. A function f of a quantum state ρ is invariant under **LU** transformations if

$$f \left[(U_1 \otimes \dots \otimes U_n) \rho (U_1^\dagger \otimes \dots \otimes U_n^\dagger) \right] = f[\rho]. \quad (1.66)$$

This poses the question of which quantities apart from the moments $\mathcal{R}_{\mathcal{M}}^{(t)}(\rho)$ are **LU** invariant. In the two-qubit case, all **LU** invariants have been classified, which are known as the Makhlin invariants [77]. The Makhlin invariants are formulated with the help of the Bloch decomposition for two qubits. A two-qubit state ρ can be written in the form

$$\rho = \frac{1}{4} \left(\mathbb{1} \otimes \mathbb{1} + \boldsymbol{\alpha} \cdot \boldsymbol{\sigma} \otimes \mathbb{1} + \mathbb{1} \otimes \boldsymbol{\beta} \cdot \boldsymbol{\sigma} + \sum_{i,j=1}^3 T_{ij} \sigma_i \otimes \sigma_j \right), \quad (1.67)$$

where $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are the Bloch vectors of the first and second qubit and T is the correlation matrix.

The Makhlin invariants can be expressed in terms of $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$ and T :

$$\begin{aligned} I_4 &= \alpha^2, & I_7 &= \beta^2, & I_2 &= \text{Tr}(TT^T), \\ I_{12} &= \boldsymbol{\alpha}^T T \boldsymbol{\beta}, & I_1 &= \det(T), \\ I_5 &= [\boldsymbol{\alpha}^T T]^2, & I_8 &= [T \boldsymbol{\beta}]^2, & I_3 &= \text{Tr}(TT^T T T^T), \\ I_{14} &= \text{Tr}[(\star \boldsymbol{\alpha}) T (\star \boldsymbol{\beta})^T T^T], \\ I_{13} &= \boldsymbol{\alpha}^T T T^T T \boldsymbol{\beta}, & I_6 &= [\boldsymbol{\alpha}^T T T^T]^2, & I_9 &= [T^T T \boldsymbol{\beta}]^2. \end{aligned} \quad (1.68)$$

The invariants are ordered such that the invariants of second order in the quantum state ρ are in the first row, the ones of degree three are in the second row, whereas the invariants in the third and fourth row are of degree three and four, respectively. The invariant I_{14} is expressed with the help of the Hodge star \star , which is defined as $(\star \boldsymbol{\alpha})_{ij} = \sum_k \epsilon_{ijk} \alpha_k$.

1.6 Shadow tomography

We now discuss the basic idea of shadow tomography. The term shadow tomography was introduced in Ref. [78] for the purpose to estimate the outcomes of observables without reconstructing the density matrix. As the name suggests the method falls into the category of tomography. It is thus not the aim to measure the observables directly. Rather, the observables should be estimated from the tomographic data. The method thus has to address the problem of how the recorded data can be efficiently stored. In this section, we will recap the scheme for shadow tomography that was proposed in Ref. [26]. The scheme is formulated for qubits and can be broken down into two separate steps: the data acquisition and the prediction of expectation values from the recorded data.

We will first discuss the data acquisition. The scheme assumes a system of n -qubits that is described by the density matrix ρ . As a first step, the state is rotated by a unitary U , i.e., $\rho \rightarrow U\rho U^\dagger$. The unitary U is chosen randomly from an ensemble \mathcal{U} . Common choices for the ensemble \mathcal{U} cover the group of global Clifford circuits $\text{Cl}(2^n)$ or local Clifford circuits $\text{Cl}(2)^{\otimes n}$, i.e., Pauli measurements. The rotated state is then measured in the computational basis $\{|b\rangle \mid b \in \{0, 1\}^n\}$. In case, where the ensemble \mathcal{U} contains a finite number of unitaries, the data of N repetitions of the experiment is thus of the form

$$(i_k, \hat{b}_k), \quad \text{for } k = 1, \dots, N, \quad (1.69)$$

where i_k denotes the index of the unitary. After the measurement, the random unitary transformation can be inverted by $U^\dagger |\hat{b}\rangle \langle \hat{b}| U$, where \hat{b} denotes the outcomes of the measurement in the computational basis.

To estimate expectation values from the recorded data, the average over both the outcomes and the unitaries of the final state is considered. This can be viewed as a quantum channel

$$\mathcal{M}(\rho) = \mathbb{E} \left[U^\dagger |\hat{b}\rangle \langle \hat{b}| U \right] = \mathbb{E}_U \left[\sum_{b \in \{0, 1\}^n} \mathbb{P}(\hat{b} = b) U^\dagger |b\rangle \langle b| U \right], \quad (1.70)$$

where \mathbb{E}_U denotes the average over the unitary ensemble and the probability to observe b in the computational basis reads $\mathbb{P}(\hat{b} = b) = \langle b| U\rho U^\dagger |b\rangle$. The unitary ensemble is tomographically complete if there are no two states $\sigma \neq \rho$ that yield the same statistics, i.e., there is a $U \in \mathcal{U}$ for which the outcome b has different probabilities $\langle b| U\sigma U^\dagger |b\rangle \neq \langle b| U\rho U^\dagger |b\rangle$. In this case, the channel \mathcal{M} can be inverted and

$$\hat{\rho} = \mathcal{M}^{-1}(U^\dagger |\hat{b}\rangle \langle \hat{b}| U) \quad (1.71)$$

is the estimate for the density operator ρ that is connected to the outcome \hat{b} for the unitary U . $\hat{\rho}$ yields in expectation the density operator ρ , i.e., $\mathbb{E}[\hat{\rho}] = \rho$. For special classes of the unitary ensemble, analytical expressions of the inverse \mathcal{M}^{-1} have been found. For example, for local Clifford circuits $\text{Cl}(2)^{\otimes n}$ the unitary transformation decomposes into a product state, i.e., $U|\hat{b}\rangle = \bigotimes_{j=1}^n U_j|\hat{b}_j\rangle$ and the inverse reads

$$\hat{\rho} = \bigotimes_{j=1}^n \left(3U_j^\dagger |\hat{b}_j\rangle \langle \hat{b}_j| U_j - \mathbb{1} \right). \quad (1.72)$$

On the one hand, the above expression decomposes into a product of single-qubit states. It is thus not necessary to construct the full density matrix if the goal is to estimate a product observable, i.e., $O = \bigotimes_{j=1}^n O_j$. On the other hand, Eq. (1.72) only contains Clifford operations that act on computational basis states. It can thus be efficiently represented in the stabilizer formalism [26, 79]. Each measurement yields a snapshot $\hat{\rho}$ that can be used to calculate the expectation value

of an observable O by $\text{Tr}(\hat{\rho}O)$. This corresponds to a single-shot estimate, which can be very noisy. For this reason, the classical shadow in Ref. [26] consists of N snapshots, i.e.,

$$S(\rho, N) = (\hat{\rho}_1, \dots, \hat{\rho}_N) \quad \text{with} \quad \hat{\rho}_k = \mathcal{M}^{-1}(U_k^\dagger |\hat{b}_k\rangle \langle \hat{b}_k| U_k). \quad (1.73)$$

Moreover, Ref. [26] uses the median of means estimation [80] to make the estimation more robust against outliers. For the median of means estimation, the single-shot estimates $(\hat{\rho}_1, \dots, \hat{\rho}_N)$ are divided into K groups of equal size. For each group, the average estimator $\hat{\rho}_{(k)}$ is calculated by

$$\hat{\rho}_{(k)} = \frac{1}{\lfloor N/K \rfloor} \sum_{l=(k-1)\lfloor N/K \rfloor + 1}^{k\lfloor N/K \rfloor} \hat{\rho}_l. \quad (1.74)$$

The expectation value of an observable O is then estimated by the median of the estimates of each group:

$$\hat{\delta}(N, K) = \text{median} [\text{Tr}(\hat{\rho}_{(1)}O), \dots, \text{Tr}(\hat{\rho}_{(K)}O)]. \quad (1.75)$$

What are the benefits of shadow tomography? To answer this question, Ref. [26] derived bounds on the sample complexity. For this purpose, the shadow norm is introduced that upper bounds the variance of an estimate. Let us consider an observable O and a single classical shadow $\hat{\rho}$. From the above discussion, the expectation value $\langle O \rangle$ is estimated by $\hat{\delta} = \text{Tr}(\hat{\rho}O)$. The variance $\text{Var}(\hat{\delta}) = \mathbb{E}[(\hat{\delta} - \mathbb{E}[\hat{\delta}])^2]$ is upper bounded by

$$\text{Var}(\hat{\delta}) \leq \left\| O - \frac{\text{Tr}(O)}{2^n} \mathbb{1} \right\|_{\text{shadow}}^2, \quad (1.76)$$

where the shadow norm is defined as

$$\|O\|_{\text{shadow}} := \max_{\sigma} \left(\mathbb{E}_U \sum_{b \in \{0,1\}^n} \langle b| U \sigma U^\dagger |b\rangle \langle b| U \mathcal{M}^{-1}(O) U^\dagger |b\rangle^2 \right)^{1/2}. \quad (1.77)$$

The shadow norm is derived by maximizing the variance over all states σ . In this sense, it describes the variance in the worst case scenario. The shadow norm is thus independent of the quantum state and only depends on the unitary ensemble. With the help of the shadow norm, Ref. [26] formulates the following theorem that assesses the sample complexity.

Theorem 1.10. *Fix a measurement primitive \mathcal{U} , a collection O_1, \dots, O_M of $2^n \times 2^n$ Hermitian matrices and accuracy parameters $\epsilon, \delta \in [0, 1]$. Set*

$$K = 2 \log(2M/\delta) \quad \text{and} \quad N = \frac{34}{\epsilon^2} \max_{1 \leq i \leq M} \left\| O_i - \frac{\text{Tr}(O_i)}{2^n} \mathbb{1} \right\|_{\text{shadow}}^2. \quad (1.78)$$

Then, a collection of NK independent classical shadows allow for accurately predicting all features via median of means prediction:

$$|\hat{\delta}_i(N, K) - \text{Tr}(\rho O_i)| \leq \epsilon \quad \text{for all} \quad 1 \leq i \leq M \quad (1.79)$$

with probability at least $1 - \delta$.

We conclude this section, by pointing out that the class of observables that can be efficiently predicted by shadow tomography depends on the unitary ensemble. To illustrate this, we note that for the example of Pauli measurements the shadow norm is upper bounded by

$$\left\| O - \frac{\text{Tr}(O)}{2^n} \mathbb{1} \right\|_{\text{shadow}}^2 \leq 4^{\text{locality}(O)} \|O\|_{\infty}^2, \quad (1.80)$$

where $\text{locality}(O)$ denotes the number of qubits on which O is acting non-trivially. We can thus conclude that shadow tomography based on Pauli measurements works best for observables that act only on few qubits.

2 Statistical tools

Statistical tools are necessary in quantum theory from two points of view. On the one hand, any physical theory has to be tested experimentally. The outcomes of experiments are, however, subjected to statistical fluctuations. To evaluate the results, it is thus necessary to perform a statistical analysis. On the other hand, quantum theory is also a statistical theory in itself. It only predicts the probabilities of the outcomes. For this reason, we will review some statistical methods in this section.

2.1 Probability and random variables

We start this section by introducing the basic terms in probability theory [81, 82]. The set of possible outcomes of an experiment is called sample space Ω . On the sample space Ω , a σ -algebra \mathcal{A} can be introduced, which formalizes the notation of an event. A σ -algebra is a set of subsets of Ω .

Definition 2.1 (σ -algebra). A set $\mathcal{A} \subseteq \mathcal{P}(\Omega)$ of the power set $\mathcal{P}(\Omega)$ of the sample space Ω is called σ -algebra if

- (i) $\emptyset \in \mathcal{A}$,
- (ii) if $A \in \mathcal{A}$ then $A^c \in \mathcal{A}$,
- (iii) $A_j \in \mathcal{A}$ for all $j \in \mathbb{N}$ implies $\bigcup_{j \in \mathbb{N}} A_j \in \mathcal{A}$.

In the above definition, $A^c = \Omega \setminus A$ denotes the complement. The elements in \mathcal{A} correspond to the events we would like to assign probabilities to. For this, a probability measure \mathbb{P} is defined.

Definition 2.2 (Probability measure). A function $\mathbb{P} : \mathcal{A} \rightarrow \mathbb{R}_+$ from a σ -algebra \mathcal{A} to the positive real numbers \mathbb{R}_+ is called probability measure or probability distribution if

- (i) for disjoint $A_j \in \mathcal{A}$ with $j \in \mathbb{N}$ it is $\mathbb{P}\left(\bigcup_{j \in \mathbb{N}} A_j\right) = \sum_{j \in \mathbb{N}} \mathbb{P}(A_j)$,
- (ii) $\mathbb{P}(\Omega) = 1$.

To put differently, a probability measure is additive for exclusive events and the probability for the certain event, i.e., to observe an outcome from the sample space Ω , is one. In this sense, a **POVM** gives rise to a probability measure for any quantum state ρ . **POVMs** have been introduced in Sec. 1.3.2. The triple $(\Omega, \mathcal{A}, \mathbb{P})$ is called a probability space. On a probability space, random variables can be defined.

Definition 2.3 (Random variable). Suppose $(\Omega, \mathcal{A}, \mathbb{P})$ is a probability space. A random variable X is a mapping

$$X : \Omega \rightarrow \Omega' \tag{2.1}$$

to another non-empty set Ω' with σ -algebra \mathcal{A}' , such that

$$\forall A \in \mathcal{A}' : X^{-1}(A) \in \mathcal{A}. \tag{2.2}$$

The inverse is defined as $X^{-1}(A') := \{\omega \in \Omega : X(\omega) \in A'\}$.

(a)	$s_z^{(2)} = 0$ $s_z^{(2)} = 1$		(b)	$s_z^{(2)} = 0$ $s_z^{(2)} = 1$		(c)	$s_z^{(2)} = 0$ $s_z^{(2)} = 1$	
$s_z^{(1)} = 0$	1/2	0	$s_z^{(1)} = 0$	1/2	0	$s_z^{(1)} = 0$	1/4	1/4
$s_z^{(1)} = 1$	0	1/2	$s_z^{(1)} = 1$	0	1/2	$s_z^{(1)} = 1$	1/4	1/4

Table 2.1: Probability distributions $\mathbb{P}(s_z^{(1)}, s_z^{(2)})$ to measure $s_z^{(i)} = 0, 1$ for qubit $i = 1, 2$. (a) shows the distribution for the Bell state $|\Phi^+\rangle$, (b) for the mixed state $\rho = \frac{1}{2}|00\rangle\langle 00| + \frac{1}{2}|11\rangle\langle 11|$ and (c) for the maximally mixed state $\frac{\mathbb{1}}{4}$.

An example of a random variable is the identity map $\text{Id} : \Omega \rightarrow \Omega$. In case $\Omega' = \mathbb{R}$, X is called a real random variable.

We have already noted that from a quantum state, probability distributions can be derived. Though, two quantum states can yield the same distribution. As an example we consider the pure state $|\psi\rangle = \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle)$ and the mixed state $\rho = \frac{1}{2}|0\rangle\langle 0| + \frac{1}{2}|1\rangle\langle 1|$. We denote the sample space of the Pauli operator Z as $\Omega = \{-1, 1\}$ and we use the power set $\mathcal{A} = \mathcal{P}(\Omega)$ as the set of events. The outcome of the measurement we associate to the random variable S . Then both states yield the probability distribution

$$\mathbb{P}_\psi(S = i) = |\langle i|\psi\rangle|^2 = \frac{1}{2} = \text{Tr}(\rho |i\rangle\langle i|) = \mathbb{P}_\rho(S = i) \quad (2.3)$$

for $i = 0, 1$.

Finally, we need the notion of statistical independence.

Definition 2.4. Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space. Two events $A, B \in \mathcal{A}$ are considered independent if

$$\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B). \quad (2.4)$$

Correspondingly, two random variable X, Y are independent if for all events A, B

$$\mathbb{P}(X \in A, Y \in B) = \mathbb{P}(X \in A)\mathbb{P}(Y \in B). \quad (2.5)$$

To illustrate this by an example, we consider the Bell state $|\Phi^+\rangle = \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle)$, the mixed state $\rho = \frac{1}{2}|00\rangle\langle 00| + \frac{1}{2}|11\rangle\langle 11|$ and the maximally mixed state $\frac{\mathbb{1}}{4}$ of two qubits. For a measurement of the observable $Z \otimes Z$, the sample space is $\Omega = \{00, 01, 10, 11\}$. The first digit is the outcome of the first qubit, which we denote by the random variable $s_z^{(1)}$. Correspondingly, we refer to the second digit as the random variable $s_z^{(2)}$ that describes the outcome of the second qubit. Tab. 2.1 shows the probability distributions that are derived from the quantum states. We note that the distribution of the Bell state $|\Phi^+\rangle$ is equal to the distribution of the mixed state ρ . Moreover, for this distribution the random variables $s_z^{(1)}$ and $s_z^{(2)}$ are not independent, as

$$\mathbb{P}_{\psi/\rho}(s_z^{(1)} = 0, s_z^{(2)} = 0) = \frac{1}{2} \neq \frac{1}{4} = \mathbb{P}_{\psi/\rho}(s_z^{(1)} = 0)\mathbb{P}_{\psi/\rho}(s_z^{(2)} = 0). \quad (2.6)$$

The probability distribution derived from the maximally mixed is shown in Tab. 2.1 (c) and it can be checked that for this distribution the random variables $s_z^{(1)}$ and $s_z^{(2)}$ are indeed independent.

To characterize a random variable with unknown probability distribution it is useful to consider the moments.

Definition 2.5 (Moments). The n -th moment of a random variable X is defined as

$$\mathbb{E}[X^n] := \int X^n d\mathbb{P} = \begin{cases} \sum_{x \in \Omega} x^n \mathbb{P}(x), & X \text{ is a discrete random variable} \\ \int x^n f(x) dx, & X \text{ is a continuous random variable.} \end{cases} \quad (2.7)$$

For $n = 1$, this defines the expectation value

$$\mu := \mathbb{E}[X] \quad (2.8)$$

of the random variable X . Instead of the second moment, it is more common to consider the variance that is defined as

$$\text{Var}(X) = (\Delta X)^2 := \mathbb{E}[(X - \mu)^2] = \mathbb{E}[X^2] - \mathbb{E}[X]^2. \quad (2.9)$$

As a final remark, we introduce the concept of conditional probability.

Definition 2.6. Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space. We consider two events $A, B \in \mathcal{A}$ with $\mathbb{P} > 0$. The probability

$$\mathbb{P}(A|B) := \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} \quad (2.10)$$

is called conditional probability.

$\mathbb{P}(A|B)$ describes the probability for outcome A if the event B is assumed to occur.

2.2 Estimators

In an experiment an observed quantity is always subjected to statistical fluctuations. On the one hand this can be due to noise. But also in an ideal case, quantum theory predicts that different outcomes occur with certain probabilities. It is thus not possible to directly access, e.g., the expectation value of an observable. Rather, it is necessary to estimate the parameter from the random outcomes of the experiment. In case a single value is estimated from the outcomes, this is known as point estimation [81].

Definition 2.7 (Point estimator). Suppose that X_1, \dots, X_n are **independent and identically distributed (iid)** random variables that describe the outcomes of an experiment. A function

$$\hat{\theta} = \hat{\theta}(X_1, \dots, X_n) \quad (2.11)$$

of the random outcomes X_1, \dots, X_n that estimates a parameter θ is called (point) estimator. We denote an estimator by a hat.

The biasedness of an estimator captures how much the expectation value of the estimator deviates from the true value.

Definition 2.8. The bias of an estimator is defined as

$$\text{bias}(\hat{\theta}) = \mathbb{E}[\hat{\theta}] - \theta, \quad (2.12)$$

where θ denotes the true value of the parameter. An estimator is called unbiased if $\text{bias}(\hat{\theta}) = 0$. In this case the estimator yields in expectation the true value, i.e., $\mathbb{E}[\hat{\theta}] = \theta$.

As two examples, we discuss the sample mean and the sample variance. These are used to estimate the mean and variance for a set of data points X_1, \dots, X_n . The sample mean is defined as

$$\hat{X} := \frac{1}{n} \sum_{j=1}^n X_j, \quad (2.13)$$

and it can be straightforwardly shown that the sample mean is an unbiased estimator $\mathbb{E}[\hat{X}] = \frac{1}{n} \sum_{j=1}^n \mathbb{E}[X_j] = \mu$. For the estimator to be unbiased it is thus necessary to have **iid** random variables, such that $\mathbb{E}[X_j] = \mu$ for all $j = 1, \dots, N$.

For the variance, it turns out that an unbiased estimator is given by [83]

$$\hat{S}^2 := \frac{1}{n-1} \sum_{j=1}^n (X_j - \hat{X})^2. \quad (2.14)$$

2.3 Hypothesis test

The previous section dealt with the estimation of a parameter θ from the random outcomes of an experiment. But, we did not answer the question how to interpret such an estimate. Often it is the case that we would like to verify a hypothesis that the parameter θ lies in the range Θ_0 . The complementary case is called alternative hypothesis. This is outlined below.

- Null hypothesis H_0 : $\theta \in \Theta_0$,
- Alternative hypothesis H_1 : $\theta \in \Theta_1$.

Θ_1 is the complement of the range Θ_0 , such that $\Theta_0 \cup \Theta_1$ covers the whole sample space of the parameter θ . A hypothesis test for a single parameter is usually left-sided, right-sided or centered. This is visualized in Fig. 2.1. For example, for an entanglement witnesses \mathcal{W} the null hypothesis is usually that the state is separable, i.e., $\langle \mathcal{W} \rangle \geq 0$. The alternative hypothesis is accordingly that $\langle \mathcal{W} \rangle < 0$, which implies that the state is entangled.

According to some decision criterion a rejection range Θ_R is defined, e.g., for a left-handed hypothesis test

$$\Theta_R = \{ \hat{\theta} < c \}. \quad (2.15)$$

We note that the criterion is formulated in terms of the estimator $\hat{\theta}$ for the parameter θ . For estimates smaller than some critical value c , the null hypothesis is dismissed. For example, for an entanglement witness, we accept the null hypothesis if $\langle \hat{\mathcal{W}} \rangle > c$ for some $c < 0$ and we dismiss H_0 if $\langle \hat{\mathcal{W}} \rangle < c$. Even though the bound for separable states is zero, the decision criterion is set to a negative value c to adjust the error probability.

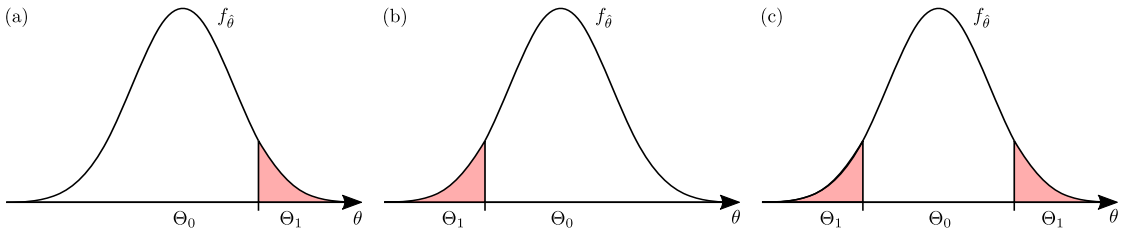


Figure 2.1: Common orientations of a hypothesis test for a single parameter. A one-dimensional hypothesis test is usually either (a) right-sided, (b) left-sided or (c) centered.

	Accept H_0	Reject H_0
H_0 is true		type I error
H_1 is true	type II error	

Table 2.2: Types of errors in a hypothesis test [81]. The case the null hypothesis H_0 is true but rejected is called type I error. Type II error in contrast refers to a falsely accepted H_0 .

For a hypothesis test there are two types of errors that originate from wrong classification. Either the null hypothesis H_0 is true but the test falsely dismisses H_0 . This is known as type I error. The other wrong classification can happen if H_0 is falsely accepted and H_1 is true. The types of errors are summarized in Tab. 2.2. The power of a hypothesis test can be characterized by the probabilities that the errors occur. The probability for type I error is denoted by

$$\alpha := \mathbb{P}(H_0 \text{ is rejected} \mid H_0 \text{ is true}), \quad (2.16)$$

whereas β refers to the probability of type II error:

$$\beta := \mathbb{P}(H_0 \text{ is accepted} \mid H_1 \text{ is true}). \quad (2.17)$$

α is thus the probability to falsely reject the null hypothesis H_0 and is known as the significance level of the hypothesis test. β in contrast denotes the probability to falsely accept H_0 . In this case, the alternative hypothesis H_1 is in fact true. The probability to rightfully discard H_0 is thus $\mathbb{P}(H_0 \text{ is rejected} \mid H_1 \text{ is true}) = 1 - \beta$. $1 - \beta$ is called the power of the test. In practice, hypothesis tests are designed to meet a given significance level of α . Common values are $\alpha = 0.05$ or smaller.

To design a test that has a given significance level α , however, requires evaluating the probability in Eq. (2.16). This in turn requires the knowledge or at least an assumption on the probability distribution of the estimator $\hat{\theta}$. Another possibility to assess the outcome of a hypothesis test is the p value.

Definition 2.9 (p value). The p value is defined as the probability for an outcome to be at least as extreme as the observed result θ_O given that the null hypotheses H_0 is true, i.e.,

$$p := \mathbb{P}(\hat{\theta} > \theta_O \mid H_0). \quad (2.18)$$

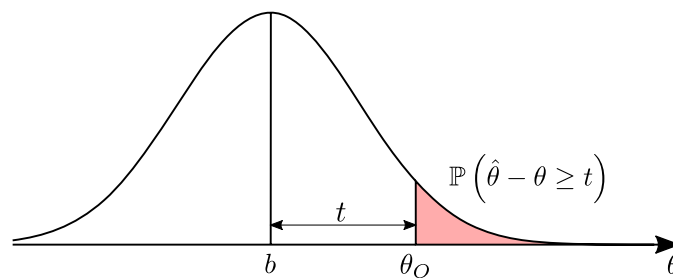


Figure 2.2: Upper bound of the p value for a right-sided hypothesis test. The aim is to check the null hypothesis $H_0 = \theta \in \Theta_0 = (-\infty, b]$, i.e., whether the parameter θ is smaller than some value b . In case H_0 is true, an unbiased estimator thus has to deviate from its mean by at least $t = \theta_O - b$ for a value θ_O to be observed. The corresponding probability $\mathbb{P}(\hat{\theta} - \theta \geq t)$ is shown in red.

In the above definition, the p value is defined for a right-sided hypothesis test. For a left-sided test the p value can be defined accordingly. To calculate the exact p value also requires knowledge about the probability distribution of the estimator $\hat{\theta}$. However, we can derive an upper bound for the p value. The derivation of the upper bound for a right-sided hypothesis test is visualized in Fig. 2.2. The goal is to verify the null hypothesis $H_0 = \theta \in \Theta_0 = (-\infty, b]$ that the parameter is in the range $(-\infty, b]$. In case H_0 is true, the mean of an unbiased estimator can thus take at most the value b . To observe a value θ_O it is thus necessary for the estimator to exceed its mean by at least $t = \theta_O - b$. This results in the upper bound

$$p \leq \mathbb{P}(\hat{\theta} - \theta \geq t). \quad (2.19)$$

2.4 Concentration inequalities

The right-hand side of Eq. (2.19) can in turn be upper bounded with concentration inequalities [81, 84]. Concentration inequalities bound the probability that a random variable X exceeds some value or deviates from their mean. Crucially, the inequalities do not require the full knowledge about the probability distribution of X . As a first example, let us consider a non-negative random variable X and $a > 0$. The expectation value of X can be lower bounded by

$$\begin{aligned} \mathbb{E}[X] &= \int_0^\infty x d\mathbb{P}_X(x) = \int_0^a x d\mathbb{P}_X(x) + \int_a^\infty x d\mathbb{P}_X(x) \\ &\geq \int_a^\infty x d\mathbb{P}_X(x) \geq a \int_a^\infty d\mathbb{P}_X(x) = a\mathbb{P}(X \geq a). \end{aligned} \quad (2.20)$$

This in turn yields an upper bound for the probability $\mathbb{P}(X \geq a)$ that the random variable X exceeds a value $a > 0$. The result is known as Markov's inequality

Theorem 2.10 (Markov's inequality). *Suppose X is a non-negative random variable and the expectation value $\mathbb{E}[X]$ exists. For any $a > 0$, it holds*

$$\mathbb{P}(X \geq a) \leq \frac{\mathbb{E}[X]}{a}. \quad (2.21)$$

We note that for Markov's inequality only the expectation value $\mathbb{E}[X]$ has to be known. In addition to Markov's inequality there is a variety of other concentration inequalities. In the remainder of this section, we will introduce two additional inequalities that are used later.

The first inequality is Cantelli's inequality [84]. Instead of the mean value, it relies on the variance of the estimator.

Theorem 2.11 (Cantelli's inequality). *Let X be a real random variable with mean $\mathbb{E}[X]$ and variance $\text{Var}(X)$. The probability that X exceeds its mean by at least $t \geq 0$ is upper bounded by*

$$\mathbb{P}(X - \mathbb{E}[X] \geq t) \leq \frac{\text{Var}(X)}{\text{Var}(X) + t^2}. \quad (2.22)$$

Another concentration inequality that does not rely on any moments of the probability distribution is Hoeffding's inequality [85]. For Hoeffding's inequality to be applicable, however, the random variable has to be a sum of independent random variables.

Theorem 2.12 (Hoeffding's inequality). *Suppose X_1, \dots, X_N are independent random variables with range $a_i \leq X_i \leq b_i$ for all $i = 1, \dots, N$. The probability that the sum $Y = \sum_{i=1}^N X_i$ exceeds its mean $\mathbb{E}[Y]$ by $t > 0$ is upper bounded by*

$$\mathbb{P}(Y - \mathbb{E}[Y] \geq t) \leq \exp\left(-\frac{2t^2}{\sum_{i=1}^N (b_i - a_i)^2}\right). \quad (2.23)$$

3 Error estimation of different schemes to measure spin-squeezing inequalities

In this chapter, we consider different measurement schemes for spin-squeezing inequalities. For this purpose, we formulate estimators that rely either on the total spin of a multiqubit system or on pair correlations. We show that the estimator in terms of pair correlations can be randomized in the sense that random pairs of qubits are measured. We further analyze the statistics of the estimator by deriving analytical expressions of their variances. The significance of an experimental result can thus be assessed with Cantelli's inequality. The presentation in this section is taken from publication [P3].

3.1 Introduction

We consider the problem of estimating spin-squeezing parameters from experimental data. For this purpose we discuss different schemes to measure the optimal spin-squeezing inequalities in Eq. (1.35) that have been introduced of Refs. [56, 57]. As a first approach, we consider an estimator based on the well-known sample mean and sample variance. This serves as a benchmark for the second and third approaches, which are formulated in terms of pair correlations. The second approach relies on the measurement of all pair correlations and single qubits. This in turn can be randomized. Instead of looking at all pair correlations and single qubits, the qubit pairs and single qubits are measured randomly. The approach to measure the qubits at random follows the methods in Refs. [18, 20, 21]. In Ref. [18] the fidelity of an N -qubit system is estimated. For this purpose, the Pauli measurements are performed at random according to some probability distribution that is determined by the target state. The authors show that this approach requires less resources than full tomography. Moreover, measurements from a set of two-outcome observables are drawn randomly in Refs. [20, 21] to verify entanglement.

The investigation of different schemes to test spin-squeezing inequalities is experimentally motivated. For different experimental setups, some of the approaches are more suitable than others. For example, in Bose-Einstein condensates the total angular momenta can usually be resolved by absorption images [47, 61, 86]. By shining a laser on the atoms whose frequency matches an internal transition, one of the spin states is pumped into an excited state. By imaging the intensity of the light that has passed through the atom cloud, the total number of atoms in each spin state can be inferred. As the atoms are indistinguishable and no measurements on single atoms are performed, an estimator that relies on the total angular momentum is appropriate. This is also the case for hot atomic vapors [87, 88]. In atomic vapors the total spin can be inferred from the Faraday effect. For this the vapor is irradiated by polarized light. The polarization rotates depending on the spin polarization and the changed polarization can be detected.

For distinguishable particles, we show that in addition to the total spin the spin-squeezing parameters can also be estimated from pair correlations. For example, ion traps usually allow the readout of the individual atoms by fluorescence measurements. For this, the ions are irradiated by a laser that couples the ions in the $|1\rangle$ state to an excited state [89]. In the process of decaying back to state $|1\rangle$, the ions emit a photon. The ions in state $|1\rangle$ can thus be identified by the photons they emit. The total spin can be determined accordingly in a postprocessing step by adding up the individual spins, although the simultaneous readout of the ions becomes challenging with increasing number of ions in the trap [90]. In this case we propose schemes that rely on pair correlations as an alternative. We note that the qubits have to be distinguishable to measure the pair correlations. We point out that also superconducting qubits are distinguishable and are read out individually. For this, the qubits are coupled to harmonic resonators [91]. Depending on the state of the qubit, the frequency of the resonator shifts, which can be detected

by a probe signal. The simultaneous readout, however, is a bit more affected by noise [92]. It might thus be advantageous to read out pair correlations. Finally, this also allows evaluation of spin-squeezing parameters for past experiments where only spin-spin correlations have been recorded [93].

Spin-squeezing parameters are usually formulated in terms of the first and second moments of angular momentum operators and thus constitute nonlinear quantities in the quantum state. To compare the different approaches, we discuss a statistical analysis that can be applied to nonlinear estimators. Consequently, our methods will be useful to estimate other nonlinear parameters (e.g., the purity or the Fisher information) as well.

As the spin-squeezing parameters can only be estimated from experimental data, we formulate the problem as a hypothesis test in Sec. 3.2. Thereafter, we will explain the three approaches to estimate the spin-squeezing parameters in detail in Sec. 3.3. Finally, we show how the statistics of the derived non-linear estimators can be analyzed in Sec. 3.4.

3.2 Formulation as a hypothesis test

To evaluate the spin-squeezing inequalities from experimental data, we define corresponding parameters that include the quantities to estimate:

$$\xi_a = \langle J_x^2 \rangle + \langle J_y^2 \rangle + \langle J_z^2 \rangle, \quad (3.1a)$$

$$\xi_b = (\Delta J_x)^2 + (\Delta J_y)^2 + (\Delta J_z)^2, \quad (3.1b)$$

$$\xi_c = \langle J_k^2 \rangle + \langle J_l^2 \rangle - (N-1)(\Delta J_m)^2, \quad (3.1c)$$

$$\xi_d = (N-1) [(\Delta J_k)^2 + (\Delta J_l)^2] - \langle J_m^2 \rangle. \quad (3.1d)$$

Then, the inequalities in Eq. (1.35) imply for separable states: $\xi_a \leq \frac{N(N+2)}{4}$, $\xi_b \geq \frac{N}{2}$, $\xi_c \leq \frac{N}{2}$, and $\xi_d \geq \frac{N(N-2)}{4}$. Although the spin-squeezing parameters, $\xi = \xi_u$ for $u = a, b, c, d$ in Eq. (3.1), contain terms that can be directly measured in an experiment, their exact values cannot be obtained from a finite number of measurement repetitions. In the following we consider how to estimate the spin-squeezing parameters in practice. For this purpose we focus on the spin-squeezing parameter in Eq. (3.1c). We note that for this parameter, the bound in Eq. (1.35c) is an upper bound. The hypothesis test is thus right-sided. The left-sided hypothesis test for spin-squeezing parameters that are lower bounded can be formulated analogously.

Let us begin by defining an estimator $\hat{\xi}$ for ξ (which we denote by a hat). An estimator $\hat{\xi}$ is a random variable according to some probability distribution, which can be created from experimental data. It is common to require that the estimator is unbiased, meaning that the expectation coincides with the target parameter value, i.e., $\mathbb{E}[\hat{\xi}] = \xi$. But due to the finite statistics, the estimator $\hat{\xi}$ exhibits fluctuations, and there are unavoidable errors in the estimation. The presence of such errors yields a finite probability for a violation of a spin-squeezing inequality, even though the actual quantum state is separable. For this reason we formulate the question of whether a state is entangled as a hypothesis test. The basics of a hypothesis test are presented in Sec. 2.3. We apply statistical methods described in Ref. [94], where the methods are used in the context of quantum state verification and fidelity estimation. Explicitly, we consider the hypotheses:

- **Null hypothesis \mathbf{H}_0 :** The quantum state is fully separable, i.e., $\rho = \sum_k p_k \rho_k^{(1)} \otimes \dots \otimes \rho_k^{(N)}$.
- **Alternative hypothesis \mathbf{H}_1 :** The quantum state is not fully separable, i.e., it is entangled.

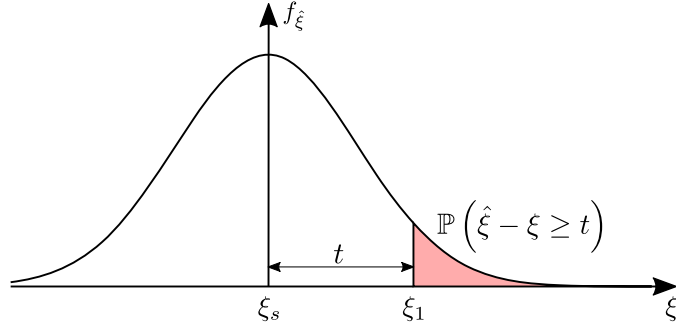


Figure 3.1: Upper bound of the p value. The plot shows an exemplary probability density function $f_{\hat{\xi}}$ of the estimator for a separable state with spin-squeezing parameter $\xi = \xi_s$. ξ_s denotes the extremal value that can be achieved by separable states. To observe an outcome ξ_1 , the estimator has to deviate at least by $t = \xi_1 - \xi_s$ from its mean. The probability $\mathbb{P}(\hat{\xi} - \xi \geq t)$ for this to happen corresponds to the red area. The figure is taken from [P3].

For the spin-squeezing parameter ξ_c the inequality (1.35c) is an upper bound. The hypothesis test is thus right-sided and the rejection criteria is of the form $\hat{\xi} > \varepsilon_c$ for some threshold ε_c . Note that ε_c does not necessarily correspond to the upper bound of the spin-squeezing parameter for separable states.

We are interested in the significance level α of an experimental result. This means that the probability to detect a state as entangled even though it was separable, i.e., the probability for *Type I* error, is at most α :

$$\mathbb{P}(\hat{\xi} > \varepsilon_c | H_0) \leq \alpha. \quad (3.2)$$

However, it is difficult to fix a threshold ε_c as the probability distribution of the estimator depends on the quantum state and is unknown. Rather, we use the p value to assess the significance of an experimental outcome ξ_1 . The p value denotes the probability that an outcome at least as extreme as ξ_1 is observed under the assumption that H_0 is true:

$$p = \mathbb{P}(\hat{\xi} \geq \xi_1 | H_0). \quad (3.3)$$

The p value depends on the specific separable state at hand. We can derive an upper bound of the p value by considering a state that saturates the separable bound, i.e., $\xi = \xi_s$. This is shown in Fig. 3.1. For a separable state the estimator has to deviate at least by $t = \xi_1 - \xi_s \geq 0$ from its mean, in case a violation is observed. As a result, we obtain the inequality

$$p \leq \mathbb{P}(\hat{\xi} - \xi \geq t). \quad (3.4)$$

The probability on the right-hand side can in turn be bounded with the help of concentration inequalities, e.g., Cantelli's inequality [84]. Concentration inequalities have been introduced in Sec. 2.4. These are large deviation bounds that typically involve the number of repetitions and thus connect the p value to the number of experimental runs. Finally, we say that a result with a certain p value has a confidence level of $\gamma = 1 - p$. As a result, we can determine the necessary number of repetitions to assure a given confidence level.

3.3 Three ways to measure spin-squeezing inequalities

In this section we are going to present the three measurement schemes to obtain the spin-squeezing parameters. As the expectation value is linear, we give the unbiased estimators for the

terms in the spin-squeezing parameters separately. We start with the scheme that uses total spin measurements. For this scheme we discuss the estimators for the expectation value $\langle J_\alpha \rangle$ and the variance $(\Delta J_\alpha)^2$. In contrast, for the approaches that are based on pair correlations we explain the estimators for $\langle J_\alpha^2 \rangle$ and $(\Delta J_\alpha)^2$, but also for $\langle J_\alpha \rangle$.

3.3.1 Estimator based on the total spin

The first approach relies on the measurement of the total spin. To evaluate the spin-squeezing parameters, the observables J_x, J_y , and J_z , that are defined in Eq. (1.32) are measured, i.e., the total spin in x, y , and z direction. In each direction $\alpha \in \{x, y, z\}$ the measurement is repeated K_{TS} times. We denote the measurement results of the k th repetition as $m_\alpha^{(k)}$. The possible outcomes are $m_\alpha^{(k)} \in \{-\frac{N}{2}, -\frac{N}{2} + 1, \dots, \frac{N}{2}\}$. This is depicted in Fig. 3.2. From the experimental data, we can infer the expectation value by the sample mean:

$$\widehat{\langle J_\alpha^2 \rangle}_{\text{TS}} = \sum_{k=1}^{K_{\text{TS}}} \frac{(m_\alpha^{(k)})^2}{K_{\text{TS}}}. \quad (3.5)$$

In the above estimator we used that the result m_α for a measurement of J_α implies the result m_α^2 for a measurement of J_α^2 .

Correspondingly, we can estimate the variance by the sample variance:

$$\widehat{(\Delta J_\alpha)^2}_{\text{TS}} = \frac{1}{K_{\text{TS}} - 1} \sum_{k=1}^{K_{\text{TS}}} \left(m_\alpha^{(k)} - \widehat{\langle J_\alpha \rangle}_{\text{TS}} \right)^2, \quad (3.6)$$

where $\widehat{\langle J_\alpha \rangle}_{\text{TS}} = \sum_{k=1}^{K_{\text{TS}}} \frac{m_\alpha^{(k)}}{K_{\text{TS}}}$ denotes the estimator for the expectation value $\langle J_\alpha \rangle$. Both the sample mean and the sample variance are unbiased estimators [83], i.e., it is $\mathbb{E}[\widehat{\langle J_\alpha^2 \rangle}_{\text{TS}}] = \langle J_\alpha^2 \rangle$ and $\mathbb{E}[\widehat{(\Delta J_\alpha)^2}_{\text{TS}}] = (\Delta J_\alpha)^2$. With these building blocks we can write down unbiased estimators for the spin-squeezing parameters in Eq. (3.1), e.g.,

$$(\hat{\xi}_c)_{\text{TS}} = \widehat{\langle J_x^2 \rangle}_{\text{TS}} + \widehat{\langle J_y^2 \rangle}_{\text{TS}} + (N - 1) \widehat{(\Delta J_z)^2}_{\text{TS}}. \quad (3.7)$$

We note that the three estimators on the right-hand side of the above equation rely on the outcomes of spin measurements in different directions. The data is thus obtained in different experimental runs and the estimators are statistically independent.

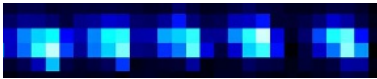
k						$m_\alpha^{(k)}$
1	$+\frac{1}{2}$	$-\frac{1}{2}$	$-\frac{1}{2}$	$+\frac{1}{2}$	$+\frac{1}{2}$	$+\frac{1}{2}$
2	$-\frac{1}{2}$	$-\frac{1}{2}$	$-\frac{1}{2}$	$+\frac{1}{2}$	$-\frac{1}{2}$	$+\frac{3}{2}$
			\vdots			
K_{TS}	$-\frac{1}{2}$	$+\frac{1}{2}$	$-\frac{1}{2}$	$-\frac{1}{2}$	$+\frac{1}{2}$	$-\frac{1}{2}$

Figure 3.2: Measurement scheme for the estimators $\widehat{\langle J_\alpha^2 \rangle}_{\text{TS}}$ and $\widehat{(\Delta J_\alpha)^2}_{\text{TS}}$. In each repetition k , the total spin of the system is measured. In an ion trap, this can be done by resonance fluorescence [89], which also gives access to the spin of the individual qubits. The figure includes an image of trapped $^{171}\text{Yb}^+$ ions, which is reprinted from [89]. The figure is taken from [P3].

Moreover, the total spin in each direction is measured K_{TS} times. The estimator in Eq. (3.7) thus requires in total $3K_{\text{TS}}$ state samples.

3.3.2 Estimator based on pair correlations

Instead of the total spin, an estimator can also be formulated in terms of pair correlations. This is motivated by the decomposition of the expectation value $\langle J_\alpha^2 \rangle$ in two-qubit correlations:

$$\langle J_\alpha^2 \rangle = \frac{N}{4} + \frac{1}{4} \sum_{i \neq j} \langle \sigma_\alpha^{(i)} \sigma_\alpha^{(j)} \rangle. \quad (3.8)$$

We can thus estimate the expectation value $\langle J_\alpha^2 \rangle$ by measuring the correlations of all distinct qubit pairs (AP). For this purpose, we propose to measure the two-qubit correlations $\langle \sigma_\alpha^{(P_1)} \sigma_\alpha^{(P_2)} \rangle$ K_{AP} times each. The above correlation has to be determined for every distinct pair, i.e., for all pairs $P = (P_1, P_2)$ of qubits P_1 and P_2 with $P_1 \neq P_2$. The corresponding estimator reads

$$\widehat{\langle J_\alpha^2 \rangle}_{\text{AP}} = \frac{N}{4} + \frac{1}{K_{\text{AP}}} \sum_P \sum_{k=1}^{K_{\text{AP}}} s_\alpha^{(P_1, k)} s_\alpha^{(P_2, k)}. \quad (3.9)$$

In the above equation, $s_\alpha^{(P_{1/2}, k)}$ denotes the spin in direction α of qubit $P_{1/2}$ in the k th measurement repetition of the pair P . This scheme is visualized in Fig. 3.3 (a). We show in App. A.1.1 that the above estimator is unbiased, i.e., $\mathbb{E}[\widehat{\langle J_\alpha^2 \rangle}_{\text{AP}}] = \langle J_\alpha^2 \rangle$.

To estimate the variances we propose two different schemes.

Scheme AP1. The first scheme uses the data as presented in Fig. 3.3 (a), i.e., it estimates the variance with the help of two-qubit correlations, although we assume that the measurement results for the individual qubits are captured. The estimator takes the form

$$\begin{aligned} (\widehat{\Delta J_\alpha})^2_{\text{AP}} &= \frac{N}{4} + \frac{1}{K_{\text{AP}}} \sum_P \sum_{k=1}^{K_{\text{AP}}} s_\alpha^{(P_1, k)} s_\alpha^{(P_2, k)} \\ &\quad - \frac{1}{K_{\text{AP}}(K_{\text{AP}} - 1)(N - 1)^2} \sum_{P, Q} \sum_{k \neq l}^{K_{\text{AP}}} s_\alpha^{(P_1, k)} s_\alpha^{(Q_2, l)}. \end{aligned} \quad (3.10)$$

With the estimators in Eqs. (3.9) and (3.10) we can compose estimators for the spin-squeezing parameters, e.g.,

$$(\hat{\xi}_c)_{\text{AP1}} = \widehat{\langle J_x^2 \rangle}_{\text{AP}} + \widehat{\langle J_y^2 \rangle}_{\text{AP}} + (N - 1) \widehat{(\Delta J_z)^2}_{\text{AP}}. \quad (3.11)$$

As $\mathbb{E}[(\widehat{\Delta J_\alpha})^2_{\text{AP}}] = (\Delta J_\alpha)^2$ (cf. App. A.1.1), $(\hat{\xi}_c)_{\text{AP1}}$ is also an unbiased estimator of ξ_c . Again, the three estimators on the right-hand side of Eq. (3.11) are obtained in different measurements and are thus statistically independent. For each direction all distinct pairs are measured K_{AP1} times, and hence the total number of state samples is $3N(N - 1)K_{\text{AP1}}$.

Scheme AP2. Alternatively, we can calculate the variance by estimating the expectation value $\langle J_\alpha \rangle^2$ separately, i.e., by $\widehat{\langle J_\alpha^2 \rangle}_{\text{AP}} - \widehat{\langle J_\alpha \rangle}_{\text{AP}}^2$. For this purpose we propose to measure the spin of one qubit in each experimental run. From the outcomes, the expectation value $\langle J_\alpha \rangle^2$ can be estimated by multiplying the results of two different experimental runs. This ensures that the two outcomes are statistically independent. To formulate the estimator, we measure all pairs (i, j) , where we allow $i = j$. We divide the number of repetitions into two groups. $\frac{K_{\text{AP}}}{2}$ of the

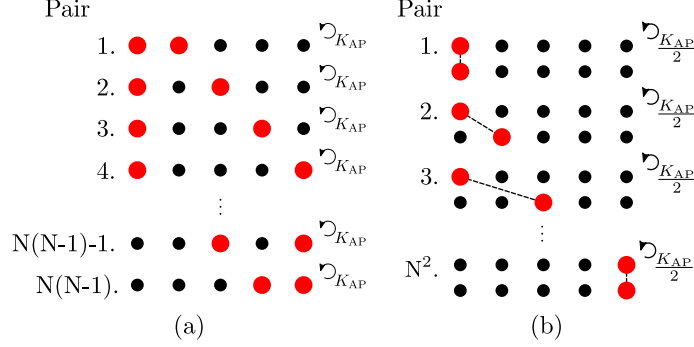


Figure 3.3: Measurement pattern for (a) $\widehat{\langle J_\alpha^2 \rangle}_{\text{AP}}$ in Eq. (3.9) as well as $\widehat{(\Delta J_\alpha)^2}_{\text{AP}}$ in Eq. (3.10) and (b) $\widehat{\langle J_\alpha \rangle}_{\text{AP}}^2$ in Eq. (3.12). In pattern (a) all $N(N-1)$ distinct pairs of qubits are measured K_{AP} -times. In contrast, in pattern (b) all N^2 pairs are measured, with each qubit observed only in $\frac{K_{\text{AP}}}{2}$ of the experimental runs to ensure statistical independence. The approach AP1 relies only on the measurement pattern (a), whereas for AP2 both the patterns (a) and (b) are used. The figure is reprinted from [P3].

times we measure the spin of qubit i and for the remaining repetitions we observe qubit j . This results in the following estimator:

$$\widehat{\langle J_\alpha \rangle}_{\text{AP}}^2 = \sum_{i,j=1}^N \frac{1}{\frac{K_{\text{AP}}}{2}} \sum_{k=1}^{\frac{K_{\text{AP}}}{2}} s_\alpha^{(i,2k)} s_\alpha^{(j,2k-1)}. \quad (3.12)$$

The measurement scheme is depicted in Fig. 3.3 (b). However, this comes at the expense that the correlations between the two estimators $\widehat{\langle J_\alpha^2 \rangle}_{\text{AP}}$ and $\widehat{\langle J_\alpha \rangle}_{\text{AP}}^2$ have to be taken into account or two independent datasets have to be obtained. In the following statistical analysis, we assume that two independent datasets are used. We show in App. A.1.1 that Eq. (3.12) is an unbiased estimator and thus

$$(\hat{\xi}_c)_{\text{AP2}} = \widehat{\langle J_x^2 \rangle}_{\text{AP}} + \widehat{\langle J_y^2 \rangle}_{\text{AP}} + (N-1) \left(\widehat{\langle J_z^2 \rangle}_{\text{AP}} - \widehat{\langle J_z \rangle}_{\text{AP}}^2 \right) \quad (3.13)$$

obeys $\mathbb{E}[(\hat{\xi}_c)_{\text{AP2}}] = \xi_c$. In case all estimators on the right-hand side of Eq. (3.13) are obtained from different datasets, they are independent. Since the estimator in Eq. (3.12) uses all pairs of qubits, the total number of state samples is $(4N-3)NK_{\text{AP2}}$.

3.3.3 Estimator based on random pair correlations

In the previous section we have formulated an estimator that relies on the measurement of all pair correlations and single qubits. Hence, the question arises whether the total number of measurements can be reduced by randomly choosing the pair correlations and qubits that are measured. This can be achieved by introducing additional random variables for the qubit indices (i, j) . For L_{RP} randomly chosen pairs, the estimator for $\langle J_\alpha^2 \rangle$ reads

$$\widehat{\langle J_\alpha^2 \rangle}_{\text{RP}} = \frac{N}{4} + \frac{N(N-1)}{K_{\text{RP}}L_{\text{RP}}} \sum_{l=1}^{L_{\text{RP}}} \sum_{k=1}^{K_{\text{RP}}} s_\alpha^{(I_l, k)} s_\alpha^{(J_l, k)}, \quad (3.14)$$

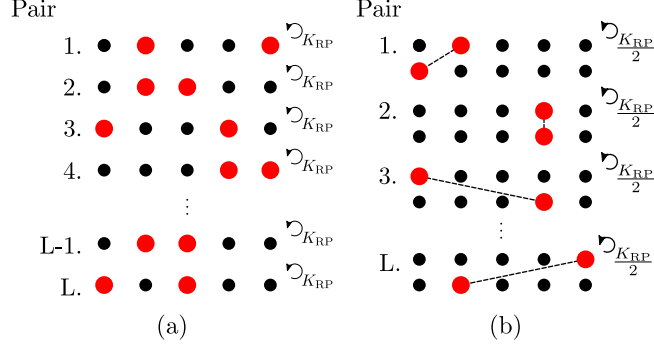


Figure 3.4: Measurement pattern for (a) $\widehat{\langle J_\alpha^2 \rangle}_{\text{RP}}$ in Eq. (3.14) and $\widehat{(\Delta J_\alpha)^2}_{\text{RP}}$ in Eq. (3.16) and for (b) $\widehat{\langle J_\alpha \rangle}_{\text{RP}}$ in Eq. (3.18). In pattern (a), L_{RP} random pair correlations are measured K_{RP} times each. Pattern (b) in turn uses also L_{RP} random pairs (i, j) , but with the possibility that $i = j$. In $\frac{K_{\text{RP}}}{2}$ of the repetitions qubit i is measured, whereas in the other repetitions qubit j is observed. The scheme RP1 is only based on the pattern (a), whereas RP2 relies on both patterns (a) and (b). The figure is reprinted from [P3].

where each pair is measured K_{RP} times. Similar to Eq. (3.9), $s_\alpha^{(\mathcal{I}_l, k)}$ and $s_\alpha^{(\mathcal{J}_l, k)}$ denote the spins in the k th measurement of the qubit pair $(\mathcal{I}_l, \mathcal{J}_l)$. However, in the above expression the indices \mathcal{I}_l and \mathcal{J}_l are random variables with $l \in \{1, \dots, L_{\text{RP}}\}$. As only distinct pairs are of interest, we use the probability distribution

$$\mathbb{P}(\mathcal{I}_l = i, \mathcal{J}_l = j) = \begin{cases} 1/[N(N-1)], & \text{for } i \neq j, \\ 0, & \text{for } i = j. \end{cases} \quad (3.15)$$

Also, the estimator in Eq. (3.14) is unbiased, as we show in App. A.1.2. The scheme is visualized in Fig. 3.4 (a).

Scheme RP1. From the data that is obtained by the pattern in Fig. 3.4 (a), we can also estimate the variance $(\Delta J_\alpha)^2$. Let us again denote the outcomes of the spin measurement in direction α for the k th repetition of the l th random pair $(\mathcal{I}_l, \mathcal{J}_l)$ by $s_\alpha^{(\mathcal{I}_l, k)}$ and $s_\alpha^{(\mathcal{J}_l, k)}$. Then, an unbiased estimator of $(\Delta J_\alpha)^2$ (cf. App. A.1.2) is given by

$$\begin{aligned} \widehat{(\Delta J_\alpha)^2}_{\text{RP}} &= \frac{N}{4} + \frac{N(N-1)}{L_{\text{RP}}K_{\text{RP}}} \sum_{l=1}^{L_{\text{RP}}} \sum_{k=1}^{K_{\text{RP}}} s_\alpha^{(\mathcal{I}_l, k)} s_\alpha^{(\mathcal{J}_l, k)} \\ &\quad - \frac{N^2}{L_{\text{RP}}(L_{\text{RP}}-1)K_{\text{RP}}^2} \sum_{l \neq m}^{L_{\text{RP}}} \sum_{k, q=1}^{K_{\text{RP}}} s_\alpha^{(\mathcal{I}_l, k)} s_\alpha^{(\mathcal{J}_m, q)}. \end{aligned} \quad (3.16)$$

The random variables $\mathcal{I}_l, \mathcal{J}_l$ obey the probability distribution in Eq. (3.15). Finally, we note that $\mathbb{E}[\widehat{(\Delta J_\alpha)^2}_{\text{RP}}] = (\Delta J_\alpha)^2$ as is shown in App. A.1.2. With the help of Eq. (3.14) and Eq. (3.16), unbiased estimators for the spin-squeezing parameters can be formulated, e.g.,

$$\widehat{(\xi_c)}_{\text{RP1}} = \widehat{\langle J_x^2 \rangle}_{\text{RP}} + \widehat{\langle J_y^2 \rangle}_{\text{RP}} + (N-1)\widehat{(\Delta J_z)^2}_{\text{RP}}. \quad (3.17)$$

Again, the estimators on the right-hand side of the above equation are statistically independent as they are obtained from different measurements. In total the estimator in Eq. (3.17) requires $3L_{\text{RP1}}K_{\text{RP1}}$ state samples.

Scheme RP2. Alternatively, we can estimate $\langle J_\alpha \rangle^2$ separately. For this purpose, we choose also L_{RP} random pairs $(\mathcal{I}_l, \mathcal{J}_l)$. However, in each experimental run only one qubit is measured. Thus for K_{RP} measurements of the pair $(\mathcal{I}_l, \mathcal{J}_l)$, we measure $\frac{K_{\text{RP}}}{2}$ times the spin of qubit \mathcal{I}_l and $\frac{K_{\text{RP}}}{2}$ times the spin of qubit \mathcal{J}_l . An estimator is retrieved by the product of the results for each pair $(\mathcal{I}_l, \mathcal{J}_l)$, i.e., $\langle J_\alpha \rangle^2$ can be obtained by the estimator

$$\widehat{\langle J_\alpha \rangle^2}_{\text{RP}} = \frac{2N^2}{K_{\text{RP}}L_{\text{RP}}} \sum_{l=1}^{L_{\text{RP}}} \sum_{k=1}^{\frac{K_{\text{RP}}}{2}} s_\alpha^{(\mathcal{I}_l, 2k)} s_\alpha^{(\mathcal{J}_l, 2k-1)}. \quad (3.18)$$

In the derivation of the above estimator, all pairs (i, j) have to be considered. Hence, we use the uniform probability distribution $\mathbb{P}(\mathcal{I}_l = i, \mathcal{J}_l = j) = \frac{1}{N^2}$ for all $i, j \in \{1, \dots, N\}$. Fig. 3.4 (b) shows a sketch of the estimator. In App. A.1.2 we prove that the estimator in Eq. (3.18) is unbiased and hence we can compose, for example, the unbiased estimator

$$(\hat{\xi}_c)_{\text{RP2}} = \widehat{\langle J_x^2 \rangle}_{\text{RP}} + \widehat{\langle J_y^2 \rangle}_{\text{RP}} + (N-1) \left(\widehat{\langle J_z^2 \rangle}_{\text{RP}} - \widehat{\langle J_z \rangle}_{\text{RP}}^2 \right). \quad (3.19)$$

The estimators for different spin directions are independent, as they have to be obtained from different datasets. Finally, we assume that also $\langle J_z^2 \rangle$ and $\langle J_z \rangle^2$ are estimated from different datasets, which assures that all estimators on the right-hand side of Eq. (3.19) are independent. This approach needs in total $4L_{\text{RP2}}K_{\text{RP2}}$ state samples.

3.4 Statistical analysis

As the estimators are formulated in terms of the random outcomes of the measurements, they are random variables themselves. Hence, they obey a probability distribution. This is exemplary shown for $(\hat{\xi}_c)_{\text{TS}}$ in Fig. 3.5. The following section therefore contains a statistical analysis of the estimators.

We start with the variances to get an insight on the spreading of the probability distribution. This results in state-dependent expressions for the variances. Accordingly, we use these results to derive probability bounds with Cantelli's inequality, which in turn can be used to assess the confidence level or the necessary number of measurements.

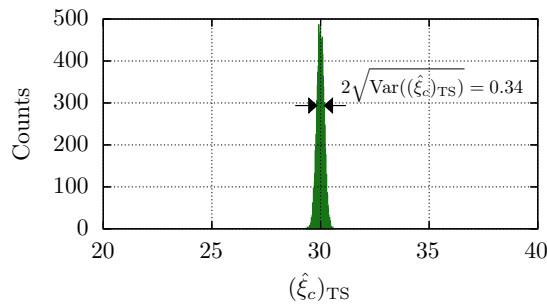


Figure 3.5: Probability distribution of the estimator $(\hat{\xi}_c)_{\text{TS}}$. The simulation has been performed for the 10-qubit Dicke state $|D_{10,5}\rangle$ defined in Eq. (1.38) with $K_{\text{TS}} = 7400$. The histogram contains 99 bins, but due to the small bin size of 0.02 they are not well resolved. The figure is reprinted from [P3].

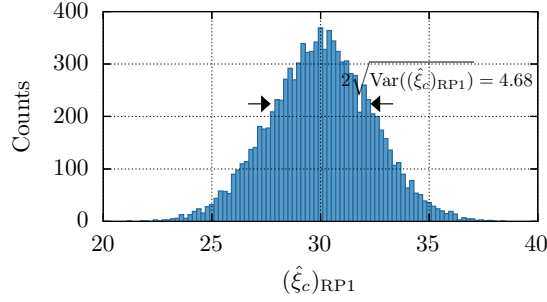


Figure 3.6: Probability distribution of the estimator $(\hat{\xi}_c)_{\text{RP1}}$. The simulation has been performed for the 10-qubit Dicke state $|D_{10,5}\rangle$. $L_{\text{RP1}} = 7400$ random pairs have been chosen with $K_{\text{RP1}} = 1$ repetition. The histogram consists of 99 bins with a size of 0.2. The figure is reprinted from [P3].

3.4.1 Variances

In App. A.2 we derive the variances of the estimators. Exemplarily, we discuss here the variance of the total spin estimator $(\hat{\xi}_c)_{\text{TS}}$ for the third spin-squeezing inequality in Eq. (1.35c):

$$\begin{aligned} \text{Var}[(\hat{\xi}_c)_{\text{TS}}] = & \frac{1}{K_{\text{TS}}} \left[(\Delta J_x^2)^2 + (\Delta J_y^2)^2 + (N-1)^2 (\Delta J_z^2)^2 \right. \\ & + (N-1)^2 \left[\frac{2}{K_{\text{TS}}-1} \langle J_z^2 \rangle^2 + 4 \frac{2K_{\text{TS}}-3}{K_{\text{TS}}-1} \langle J_z^2 \rangle \langle J_z \rangle^2 \right. \\ & \left. \left. - 4 \langle J_z^3 \rangle \langle J_z \rangle - 2 \frac{2K_{\text{TS}}-3}{K_{\text{TS}}-1} \langle J_z \rangle^4 \right] \right]. \end{aligned} \quad (3.20)$$

As expected, the variance decreases with the number of measurement repetitions K_{TS} . In Tab. 3.1 we show the variance of $(\hat{\xi}_c)_{\text{TS}}$ calculated with the analytic expression in Eq. (3.20) for $K_{\text{TS}} = 7400$. The total number of state samples is thus 22200. Indeed, the variance matches the simulation in Fig. 3.5 as $2 \times \text{Var}((\hat{\xi}_c)_{\text{TS}})^{1/2} = 0.3369 \approx 0.34$.

In the same manner, we derive in App. A.2 the variances of the estimators of schemes AP1 and AP2 as well as of schemes RP1 and RP2. In Tab. 3.1 the variances of the different estimators for the parameter ξ_c are shown. To compare the variances, we ensure that the total number of state samples is equal. For this reason we have chosen $K_{\text{AP2}} = 60$, $L_{\text{RP1}} = 7400$ with $K_{\text{RP1}} = 1$ and $L_{\text{RP2}} = 2775$ with $K_{\text{RP2}} = 2$. For scheme AP1, however, there is no integer K_{AP1} to match the total number of state preparations of 22200. For this reason, we choose $K_{\text{AP1}} = 82$ to obtain the closest number of state samples 22140.

The results in Tab. 3.1 show the smallest variance for the estimator $(\hat{\xi}_c)_{\text{TS}}$. $\text{Var}((\hat{\xi}_c)_{\text{TS}})$ is about two magnitudes smaller than the next bigger variances of schemes AP1 and RP1. We note that this appears reasonable, as in each experimental run only two qubits are measured in schemes AP1 and RP1. Therefore, in a hand-wavy sense, less information is extracted in each step. The variances of both schemes AP1 and RP1 are almost the same, in which the value is slightly larger for scheme AP1. This appears counterintuitive as the additional randomization in scheme RP1 is expected to introduce further uncertainty. We note, however, that this is due to the slightly fewer state samples used for scheme AP1. In case we increase the number of repetitions by 1, i.e., $K_{\text{AP1}} = 63$, we obtain $\text{Var}((\hat{\xi}_c)_{\text{AP1}}) = 5.5163$.

The fact that the variance of $(\hat{\xi}_c)_{\text{RP1}}$ is two orders larger than for $(\hat{\xi}_c)_{\text{TS}}$ is also revealed in

Estimator	$\text{Var}(\hat{\xi})$	Estimator	$\text{Var}(\hat{\xi})$
$(\hat{\xi}_c)_{\text{TS}}$	0.0284		
$(\hat{\xi}_c)_{\text{AP1}}$	5.5836	$(\hat{\xi}_c)_{\text{AP2}}$	24.5046
$(\hat{\xi}_c)_{\text{RP1}}$	5.5685	$(\hat{\xi}_c)_{\text{RP2}}$	25.6667

Table 3.1: Variances of the estimators for the Dicke state $|D_{10,5}\rangle$. The variances are obtained for $K_{\text{TS}} = 7400$, $K_{\text{AP1}} = 82$, and $K_{\text{AP2}} = 60$. For the randomized approaches we used $L_{\text{RP1}} = 7400$ with $K_{\text{RP1}} = 1$ and $L_{\text{RP2}} = 2775$ with $K_{\text{RP2}} = 2$. In this case the total number of measurements is 22200 for all schemes except AP1. For scheme AP1 the total number of state samples is slightly less: 22140.

Fig. 3.6. Fig. 3.6 shows the histogram of $(\hat{\xi}_c)_{\text{RP1}}$ for the Dicke state $|D_{10,5}\rangle$. From Tab. 3.1, we obtain $2 \times \text{Var}((\hat{\xi}_c)_{\text{RP1}})^{1/2} = 4.7195 \approx 4.68$. The deviation is attributed to the finite number of repetitions. The histogram in Fig. 3.6 is obtained from 10000 samples of $(\hat{\xi}_c)_{\text{RP1}}$.

Finally, the variances of schemes AP2 and RP2 are in turn almost an order larger than the variances of schemes AP1 and RP1. This seems plausible, as both schemes AP2 and RP2 rely also on measurements on single qubits, and thus less information is revealed from each state sample as compared to schemes AP1 and RP1. In detail, the variance $\text{Var}((\hat{\xi}_c)_{\text{RP2}})$ is slightly larger than the variance $\text{Var}((\hat{\xi}_c)_{\text{AP2}})$, which we attribute to the additional randomness in scheme RP2.

3.4.2 Scaling of the variances

Next we will compare the estimators by the scaling of the variances for specific states that violate the spin-squeezing inequalities. On the one hand, we use the many-body singlet state presented in Eq. (1.37) for the spin-squeezing inequalities in Eqs. (1.35b) and (1.35d). Many-body singlet states maximally violate Eq. (1.35b) and also show a violation of Eq. (1.35d). On the other hand, the spin-squeezing inequality (1.35c) is maximally violated by the Dicke state $|D_{N,N/2}\rangle$ defined in Eq. (1.38). Thus, we analyze the variances of ξ_c with the help of the Dicke state.

For the many-body singlet state in Eq. (1.37), we show the expressions for the variances $\text{Var}(\hat{\xi}_b)$ and $\text{Var}(\hat{\xi}_d)$ in Tab. 3.2. We observe that for the total spin estimator both $\text{Var}((\hat{\xi}_b)_{\text{TS}}) = 0$ and $\text{Var}((\hat{\xi}_d)_{\text{TS}}) = 0$. This is due to the properties in Eq. (1.36). Moreover, Tab. 3.2 shows that both $\text{Var}((\hat{\xi}_b)_{\text{AP1}})$ and $\text{Var}((\hat{\xi}_b)_{\text{AP2}})$ scale as $\mathcal{O}(N^2)$, whereas $\text{Var}((\hat{\xi}_b)_{\text{RP1}})$ and $\text{Var}((\hat{\xi}_b)_{\text{RP2}})$ scale as $\mathcal{O}(N^4)$. As expected, the variances of the estimator that use random pair correlations scale worse with N . We note that the scaling differs in a factor of N^2 , which corresponds to the order of qubit pairs. Alike, we obtain that the variances for the spin-squeezing parameter ξ_d scale as $\text{Var}((\hat{\xi}_d)_{\text{AP1}}), \text{Var}((\hat{\xi}_d)_{\text{AP2}}) \sim \mathcal{O}(N^4)$, and $\text{Var}((\hat{\xi}_d)_{\text{RP1}}), \text{Var}((\hat{\xi}_d)_{\text{RP2}}) \sim \mathcal{O}(N^6)$. We note that the result differs to that of ξ_b by a factor of N^2 . This is due to the additional factor of $N - 1$ in the parameter ξ_d .

Finally, for the variance of ξ_c we consider the Dicke state $|D_{N,N/2}\rangle$. The results in Tab. 3.2 show that for this case the total spin estimator has a nonzero variance that scales as $\text{Var}((\hat{\xi}_c)_{\text{TS}}) \sim \mathcal{O}(N^4)$. For the Dicke state, also the variances of the estimator AP1 and AP2 show the same scaling in N , i.e., $\text{Var}((\hat{\xi}_c)_{\text{AP1}}), \text{Var}((\hat{\xi}_c)_{\text{AP2}}) \sim \mathcal{O}(N^4)$. However, we note that each pair has to be measured K_{AP} times. Thus, in case the variance is considered as a function of the total number of experimental runs, the scaling is $\mathcal{O}(N^6)$. As for the other parameters, we observe that the variances of the randomized approaches RP1 and RP2 are two orders larger than for schemes AP1 and AP2. We obtain $\text{Var}((\hat{\xi}_c)_{\text{RP1}}), \text{Var}((\hat{\xi}_c)_{\text{RP2}}) \sim \mathcal{O}(N^6)$.

	ξ_b	ξ_c	ξ_d
TS	0	$\frac{N^4+4N^3-4N^2-16N}{64K_{\text{TS}}}$	0
AP1	$\frac{3N(K_{\text{AP1}}\mathcal{O}(N^5)-\mathcal{O}(N^5))}{16(K_{\text{AP1}}-1)K_{\text{AP1}}(N-1)^4}$	$\frac{N(K_{\text{AP1}}\mathcal{O}(N^5)-\mathcal{O}(N^5))}{32(K_{\text{AP1}}-1)K_{\text{AP1}}(N-1)^2}$	$\frac{N(K_{\text{AP1}}\mathcal{O}(N^5)-\mathcal{O}(N^5))}{16(K_{\text{AP1}}-1)K_{\text{AP1}}(N-1)^2}$
AP2	$\frac{9N^2-6N}{16K_{\text{AP2}}}$	$\frac{6N^5-20N^4+25N^3-16N^2+4N}{32K_{\text{AP2}}(N-1)}$	$\frac{6N^4-16N^3+15N^2-6N}{16K_{\text{AP2}}}$
RP1	$\frac{3N^3(L_{\text{RP1}}\mathcal{O}(N^3)+\mathcal{O}(N^2))}{16(L_{\text{RP1}}-1)L_{\text{RP1}}(N-1)^2}$	$\frac{N^2(L_{\text{RP1}}\mathcal{O}(N^4)+\mathcal{O}(N^3))}{32(L_{\text{RP1}}-1)L_{\text{RP1}}}$	$\frac{N^3(L_{\text{RP1}}\mathcal{O}(N^3)+\mathcal{O}(N^2))}{16(L_{\text{RP1}}-1)L_{\text{RP1}}}$
RP2	$\frac{9N^4-6N^3}{16K_{\text{RP2}}L_{\text{RP2}}}$	$\frac{6N^6-16N^5+17N^4-12N^3+4N^2}{32K_{\text{RP2}}L_{\text{RP2}}}$	$\frac{6N^6-16N^5+15N^4-6N^3}{16K_{\text{RP2}}L_{\text{RP2}}}$

Table 3.2: Expressions for the variances for specific states. The inequality Eq. (1.35b) is maximally violated by many-body singlet states in Eq. (1.37). For this reason, we show the variance for the many-body singlet states. Correspondingly, we evaluate the variance of $\hat{\xi}_c$ for the Dicke state $|D_{N,N/2}\rangle$, as this state violates Eq. (1.35c) the most. Finally, the many-body singlet states also violate Eq. (1.35d), and the variance for these states is shown in the third column. We note that the variances for scheme RP1 are given for the case $K_{\text{RP1}} = 1$.

3.4.3 Statistical test

With the help of the variances, we can now make a statement on the p value and thus on the significance of an experimental result. For this purpose we use Cantelli's inequality [84] that has been introduced in Sec. 2.4. Cantelli's inequality is a bound on the probability that a real-valued random variable $\hat{\xi}$ exceeds its mean value by an amount t , i.e.,

$$\mathbb{P}\left(\hat{\xi} - \mathbb{E}(\hat{\xi}) \geq t\right) \leq \frac{\text{Var}(\hat{\xi})}{\text{Var}(\hat{\xi}) + t^2}. \quad (3.21)$$

As we are focused on unbiased estimators, Cantelli's inequality bounds the probability that the estimator deviates from the actual value $\xi = \mathbb{E}[\hat{\xi}]$. However, the variances depend on the quantum state. As a result, we have to take the upper bound of the variance over all quantum states.

In this section, however, we use a different approach. We consider that the target state is the Dicke state $|D_{N,N/2}\rangle$. In addition, we assume that only depolarization noise affects the state preparation, i.e., the prepared state has the form

$$\rho = p |D_{N,N/2}\rangle\langle D_{N,N/2}| + (1-p) \frac{\mathbb{1}}{2^N}, \quad (3.22)$$

with $p \in [0, 1]$. For $N = 10$ qubits, the variances are plotted in Fig. 3.7. We can observe that for the whole range of p , the variances of $(\hat{\xi}_c)_{\text{AP1}}$ and $(\hat{\xi}_c)_{\text{RP1}}$ are at least two orders of magnitude larger than the variance of $(\hat{\xi}_c)_{\text{TS}}$. The variances of $(\hat{\xi}_c)_{\text{AP2}}$ and $(\hat{\xi}_c)_{\text{RP2}}$ are in turn three orders larger.

Moreover, Fig. 3.7 shows that the variances take the smallest value for the pure Dicke state, i.e., for $p = 1$. What is more, the spin-squeezing inequality in Eq. (1.35c) detects the mixture ρ as entangled for $p > p^* = \frac{N-1}{2N-1}$. For $N = 10$, the critical value is $p^* = 0.47$. As a result, all variances take the maximum in the region of separable states.

To assess the number of state preparations, we can make use of Cantelli's inequality in Eq. (3.21). For this purpose we use the assumption that the state is always described by the mixture in Eq. (3.22). Moreover, we use that for any $t > 0$, the bound in Cantelli's inequality is monotonically increasing in the variance. For this reason we can take the maximum of the variance for the mixture ρ . To verify a violation of t by at least a confidence level of $\gamma = 1 - p$,

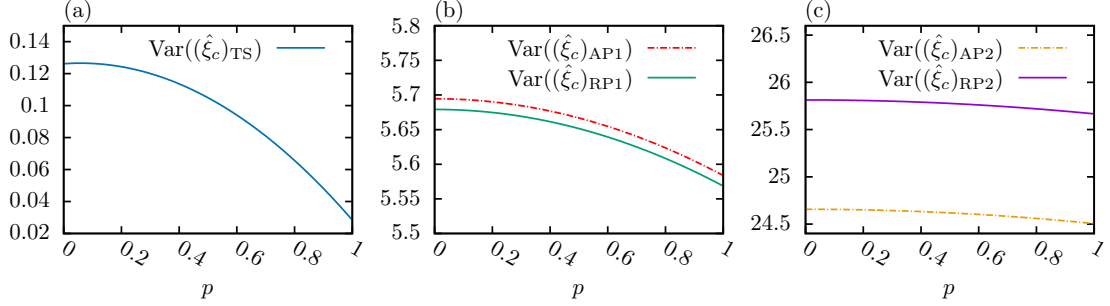


Figure 3.7: Variances of the estimators $(\hat{\xi}_c)_{\text{TS}}$, $(\hat{\xi}_c)_{\text{AP1}}$, $(\hat{\xi}_c)_{\text{AP2}}$, $(\hat{\xi}_c)_{\text{RP1}}$ and $(\hat{\xi}_c)_{\text{RP2}}$ for the Dicke state of $N = 10$ qubits $|D_{10,5}\rangle$ mixed with depolarization noise, i.e., $\rho = p|D_{10,5}\rangle\langle D_{10,5}| + (1-p)\mathbb{1}/2^N$. The variances are obtained for $K_{\text{TS}} = 7400$, $K_{\text{AP1}} = 82$, $K_{\text{AP2}} = 60$, $L_{\text{RP1}} = 7400$ with $K_{\text{RP1}} = 1$ and $L_{\text{RP2}} = 2775$ with $K_{\text{RP2}} = 2$. The figure is reprinted from [P3].

we can rearrange Cantelli's inequality to obtain a lower bound for the necessary number of state samples. To be precise, Cantelli's inequality yields the number of repetitions K_{TS} , K_{AP1} , K_{AP2} and L_{RP1} under the assumption that $K_{\text{RP1}} = 1$. In addition, we used Cantelli's inequality to obtain the product $K_{\text{RP2}}L_{\text{RP2}}$ such that the observed violation of t has a confidence of γ . In Fig. 3.8 we show the total number of state preparations, i.e., we plot $3K_{\text{TS}}$, $3K_{\text{AP1}}N(N-1)$, $K_{\text{AP2}}N(4N-3)$, $3K_{\text{RP1}}L_{\text{RP1}}$ and $4K_{\text{RP2}}L_{\text{RP2}}$.

The figure shows that $(\hat{\xi}_c)_{\text{TS}}$ requires the least number of state samples. Moreover, the number of state samples that are necessary in scheme TS scales favorably with N compared to the other schemes. It is interesting to note that the number of required state samples for schemes AP1, AP2, RP1, and RP2 obeys the same scaling with N . For schemes AP1 and RP1, almost the same number of state samples are needed. This appears reasonable, as the variances of both schemes do not differ much, as is shown in Fig. 3.7 (b). More samples are needed for schemes AP2 and RP2. However, Fig. 3.7 (c) shows that also the variances of schemes AP2 and RP2 do not differ by much. Schemes AP2 and RP2 thus require almost the same number of state samples, whereas slightly more state samples are needed for scheme RP2.

3.5 Discussion

We have discussed different approaches to estimate spin-squeezing inequalities from experimental data. On the one hand, this includes the straightforward estimation from measurements of the total angular momenta. On the other hand, we have introduced different schemes to estimate spin-squeezing inequalities from pair correlations. In doing so we have shown that it is also possible to sample the pair correlations at random. By providing different schemes to estimate spin-squeezing inequalities, it is possible to choose the most suitable approach for a given experimental setup.

As the second important point of the paper, we have performed a rigorous error analysis of the schemes. The spin-squeezing inequalities are nonlinear in the quantum state and thus are the estimators. We therefore provide an error analysis based on the variances of the estimators and Cantelli's inequality that can be applied to generic nonlinear estimators. Finally, we apply the error analysis to derive the necessary number of measurements to verify a violation of a spin-squeezing inequality with a given confidence level.

Our methods and results will not only be useful for experimental studies on spin-squeezing,

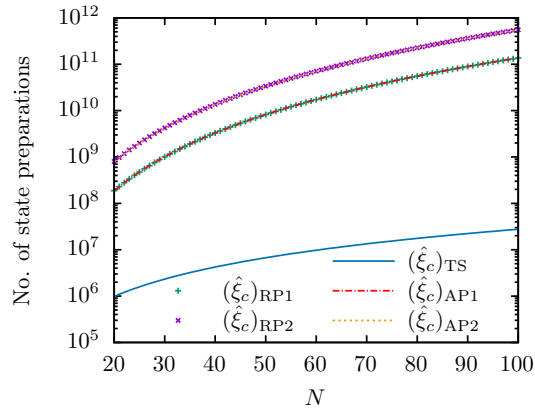


Figure 3.8: Number of state preparations necessary to verify a violation of Eq. (1.35c) by $t = 0.1 \times \frac{N}{2}$ with a significance level of $\gamma = 0.95$. The figure is reprinted from [P3].

they also lead to further research questions worthy of study. For instance, one may extend the presented methods to other nonlinear figures of merit, such as the purity or entropies of a quantum state. Furthermore, it would be interesting to study the concept of randomly picked subsets of particles also in the context of other methods of quantum state analysis, such as shadow tomography.

4 Generating multipartite nonlocality to benchmark quantum computers

In the previous section, we have discussed how spin-squeezing inequalities can be evaluated by measuring random pair correlation. We now apply a similar idea to evaluate Bell inequalities. Multipartite Bell inequalities often have the problem that they require a number of measurement settings that increases exponentially with the number of qubits. In this section, we show that the Bell inequalities can also be evaluated by sampling the terms of the Bell operators at random. This reduces the measurement resources. For the presentation, we follow publication [P4].

4.1 Introduction

The term “ n -partite nonlocality” refers to correlations between n parties that cannot be explained by any local realistic model [7, 95]. It can be detected by a violation of a multipartite Bell inequality, which shows that at least one of the assumptions of a local realistic model is false [6, 96]. The experimental test of n -partite nonlocality is, however, challenging. The first and main reason is that it is very difficult to have physical systems with n parts that can be prepared in a genuinely n -partite entangled quantum state [65] and on which specific local measurements can be performed on each of the n individual parts. From this perspective, quantum computers offer a unique chance to realistically go to a large n and test quantum theory. Quantum computers have dozens, hundreds, or even thousands of qubits which can, in principle, be prepared in arbitrary quantum states and then measured individually.

We propose a method to produce and certify n -qubit Bell nonlocality with unprecedented large n . Quantum mechanics predicts nonlocality and that the violation increases exponentially with n . The main motivation is thus to experimentally test this prediction for large n , i.e., in the “macroscopic” limit. Specifically, in our case, the aim is observing nonlocality produced by “a superposition of macroscopically distinct states” [65]. In this respect, this work is in the line of recent results showing that quantum computers can produce correlations that are impossible in other platforms [97] or used for many-body simulation of fermionic systems [98].

The second motivation is to use the observed Bell violation as a benchmark to compare different quantum computers. Since, as far as we know, n -partite nonlocality is a phenomenon specific to quantum theory, one can think of using it to quantify the “quantumness” of the device that has produced it. This can be achieved by the fraction $D_n = Q_n/C_n$ of the maximal quantum value Q_n and the classical bound C_n [65–67, 99]. For our use-case, we stress that D_n is related to the resistance of the violation to depolarization noise. Moreover, D_n is associated to the detection efficiency that is required to classically simulate the quantum nonlocality [67].

A variety of benchmarks have been proposed to test the quality of quantum computers, i.e., the quantum volume or the cross-entropy benchmark [91, 100, 101]. Still, no universally accepted standard has been established. Current approaches do not fulfill the ideal requirements to be independent of the noise model and the hardware: to not be tied to one algorithm but still being predictive and scalable in practice [101]. The phenomenon of n -partite nonlocality is promising in this regard, as a Bell violation has an interpretation independent of the hardware and the specific noise. Observing a violation of a Bell inequality requires specific quantum states and measurements. Consequently, Bell inequalities can be used to certify both measurements and states [102], which makes them attractive for benchmarking. In contrast, entanglement witnesses certify a quantum state given some well-characterized measurements, which requires additional assumptions. In addition, a Bell test can be carried out, in principle, for all pure entangled states [103]. In this sense, using Bell inequalities for benchmarking does not rely on a specific algorithm to prepare a certain quantum state. The fraction D_n grows with system size, and we

show that this facilitates the scaling of the Bell test to many qubits. Finally, an observed Bell violation can be used to lower bound the fidelity, which allows one to predict the quality of other computations.

Experimentally, violations of n -partite Bell inequalities have been observed in a variety of physical systems that are promising for quantum computing [104–114]. So far, however, mostly systems with a relatively small number of parties, n , have been considered. In an ion trap, violations of up to $n = 14$ have been verified [104], whereas nonlocality has also been shown for $n = 6$ with photons [105]. $n = 3$ nonlocality has been verified on a **nuclear magnetic resonance (NMR)** quantum simulator [114]. Finally, nonlocality has also been investigated in atomic ensembles [115] and in optical lattices [106]. The largest number of parties has been achieved with superconducting qubits with up to $n = 57$ [112]. To limit the experimental resources, however, this reference considered Bell inequalities that do not show an exponential violation in n .

The question remains why exponentially increasing nonlocality has not been verified with quantum computers before. There are, fundamentally, three challenges:

(i) Typically, quantum computers can apply two-qubit gates only on some specific pairs of qubits. Hereafter, we will refer to the map that specifies these pairs as the connectivity of the quantum computer. A consequence of this limitation is that not all connectivities allow us to equally well prepare an n -qubit **GHZ** state [116], which was the default option in [104, 107].

(ii) Quantum computers with larger n are typically more affected by noise and decoherence. Therefore, the larger n is, the harder it becomes to prepare the target state and to observe a violation of a Bell inequality.

(iii) The number of different combinations of local measurements (contexts) needed to test a Bell inequality increases with n . For example, in the case that there are two measurements per qubit, the number of contexts scales exponentially in n and, typically, so does the number of terms needed to test the Bell inequality. Therefore, measuring all contexts becomes infeasible for large n .

In this work, we show how to overcome or alleviate each of these three challenges. First, we will show that there is a natural solution to problem (i). For a given connectivity, we can focus on the graph states that are compatible with the connectivity graph. Graph states are a specific set of pure entangled states [29] and can be prepared by applying controlled-Z (CZ) gates on adjacent qubits in the graph. Thus, if we assume that CZ gates can be performed between connected qubits, graph states that correspond to subgraphs of the two-qubit connectivity can be readily prepared.

For n -qubit graph states, there are some general methods to obtain n -partite Bell inequalities [66–68, 70, 71]. However, it is, in general, a hard task to find the optimal one (in the sense of resistance to noise of the violation); the optimal n -partite Bell inequalities in terms of the stabilizers have been identified for some graph states [67]. In particular, the optimal n -partite Bell inequalities associated to the **GHZ** and **LC** state are known [67, 71]. These states correspond to the extreme cases of connectivity: On the one side, the **GHZ** state is easy to prepare when all qubits can be coupled to each other. On the other side, we consider the **LC** state that can be conveniently prepared on a quantum computer with minimal connectivity, i.e., the connectivity graph is a line. Cluster states have, in addition, the advantage to be more resistant to decoherence [117]. Quantum theory predicts that for the **GHZ** and **LC** states, the ratio D_n can be made arbitrarily large by increasing the number of qubits [65, 71, 118]. This means that in theory, the resistance to noise of multipartite nonlocality *grows exponentially* with the number of particles. This helps to overcome (ii).

In addition, the main aim of this work is to introduce a general method to address problem (iii). For this purpose, we discuss how the expectation value of a Bell operator can be estimated

from the measurements of only a few terms. The terms are chosen at random and thus the method falls into line with previous randomized measurement approaches, e.g., direct fidelity estimation [18, 19] or few-copy entanglement verification [119]. As this method is not restricted to a specific Bell inequality, it can be applied to the Bell operator that is most appropriate for the experimental set-up taking into account the feasible interactions.

We note, however, that quantum computers usually are not suited for a loophole-free Bell test. For example, ions are typically in the same trap and superconducting qubits on the same chip. It is thus not possible to rule out communication. However, in principle, the interactions can be tuned to minimize the crosstalk between the qubits. Hereafter, we will refer to this assumption as the no-crosstalk assumption and we will make it on the belief that quantum computers are the only way to investigate n -partite nonlocality with large n .

The proposed method uses graph states that have been introduced in Sec. 1.2. Moreover, we have seen in Sec. 1.4.3 that there are Bell inequalities for graph states that show a relative violation that increases exponentially with the number of qubits. In Sec. 4.2, we will explain the method to measure the Bell inequalities. As we propose to evaluate the Bell inequalities by random sampling in Sec. 4.2.2, we first formulate the Bell test as a hypothesis test in Sec. 4.2.1. After we discuss the sample complexity in Sec. 4.3, we will apply the method exemplarily to the Bell inequalities of the GHZ and LC states in Sec. 4.4. These Bell inequalities cover the extreme cases of connectivities in quantum computers. Finally, we use this to benchmark actual architectures in Sec. 4.5 and include a simulation for the IBM Eagle quantum processor in Sec. 4.5.3.

4.2 Methods

The Bell operators in Sec. 1.4.3 rely on a number of measurement settings that scales exponentially in the number of qubits n , i.e., there are, in principle, an exponential number of terms to measure. The number of observables thus quickly becomes infeasible. In this section, we introduce a general method to evaluate a Bell inequality by sampling random terms. The terms of the Bell inequality are picked according to a uniform probability distribution. This approach stands in line with previous schemes that use randomization to reduce the number of measurements, e.g., direct fidelity estimation [18, 19] or few-copy entanglement detection [21]. Finally, there exist different methods to assess the statistical strength of Bell tests [64, 120–122]. We gauge the significance of a violation with the help of the p value. For this purpose, we start by formulating a Bell test as a hypothesis test.

4.2.1 Bell test as a hypothesis test

The task is to evaluate a general Bell inequality with classical bound C , i.e., to check the inequality

$$\langle \mathcal{B} \rangle \leq C. \quad (4.1)$$

The expectation value on the right-hand side, however, cannot be inferred exactly in an experiment. Rather, the expectation value has to be estimated from multiple experimental repetitions. For this purpose, it is useful to consider an unbiased estimator $\langle \hat{\mathcal{B}} \rangle$, which we denote by a hat. An estimator $\langle \hat{\mathcal{B}} \rangle$ is a function of the experimental data and it is unbiased in case it reproduces the actual value in expectation, i.e., $\mathbb{E}[\langle \hat{\mathcal{B}} \rangle] = \langle \mathcal{B} \rangle$.

Due to the finite statistics, however, the estimate $\langle \hat{\mathcal{B}} \rangle$ fluctuates. There is thus a nonzero probability to observe a violation of the Bell inequality, although the actual state does not violate it. To quantify the probability that an observed violation is only due to statistical fluctuations, we formulate the Bell test as a hypothesis test. The hypotheses are as follows:

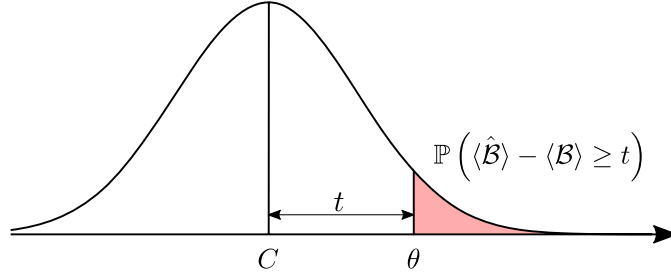


Figure 4.1: Upper bound of the p value. To observe a value of $\theta > C$ even though the state obeys the classical bound, the estimator $\langle \hat{\mathcal{B}} \rangle$ has to exceed its mean by at least t . The figure is taken from [P4].

- **Null hypothesis H_0 :** The measurement outcomes can be described by a LHV model.
- **Alternative hypothesis H_1 :** The Bell inequality is violated.

To gauge whether the observed data is in contradiction with the null hypothesis H_0 , we look at the p value. The p value is defined as the probability to observe an estimate at least as large as some value θ in case H_0 is true, i.e.,

$$p = \mathbb{P}(\langle \hat{\mathcal{B}} \rangle > \theta \mid H_0). \quad (4.2)$$

But the p value is hard to calculate since the probability depends on the probability distribution of the estimator $\langle \hat{\mathcal{B}} \rangle$, which is unknown. We can, however, upper bound the p value. A local theory can, at most, reach a value of $\langle \mathcal{B} \rangle = \mathbb{E}[\langle \hat{\mathcal{B}} \rangle] = C$. Thus, in case H_0 is true, the estimator $\langle \hat{\mathcal{B}} \rangle$ has to exceed its mean by at least $t = \theta - C$ if a violation $\langle \hat{\mathcal{B}} \rangle = \theta > C$ is observed. This is sketched in Fig. 4.1. We consequently obtain the upper bound

$$p \leq \mathbb{P}(\langle \hat{\mathcal{B}} \rangle - \langle \mathcal{B} \rangle \geq t). \quad (4.3)$$

In the following, we use Hoeffding's inequality [85] to upper bound the right-hand side of Eq. (4.3). Hoeffding's inequality is a large deviation bound that typically involves the number of repetitions. This allows us to connect the number of repetitions to the p value. Finally, we say that an observed result has a confidence level of $\gamma = 1 - p$.

4.2.2 Random sampling

So far, we have noted that there are known multipartite Bell inequalities that adapt to the architecture of the quantum computer and show a growing resistance to white noise. This allows one to overcome problems (i) and (ii). We have also seen, however, that the number of terms grows exponentially, which makes it experimentally infeasible to measure all terms. In this section, we describe a scheme to overcome this problem and which allows one to evaluate the Bell inequalities.

A general Bell operator \mathcal{B} can be written as a sum of observables:

$$\mathcal{B} = \sum_{j=1}^M B_j. \quad (4.4)$$

We note that for the Bell inequalities in Sec. 1.4.3, the number of terms equals the quantum bound, i.e., $M = Q_n$. We propose to estimate the expectation value of the Bell operator by

measuring the expectation value of L randomly chosen terms. To analyze how many terms L have to be sampled, we first assume that the expectation values can be inferred directly, i.e., we consider the limit of infinite measurements. With this simplification, the estimator reads

$$\langle \hat{\mathcal{B}} \rangle_\infty = \frac{M}{L} \sum_{l=1}^L \langle B_{J_l} \rangle. \quad (4.5)$$

We note that in the above expression, J_1, \dots, J_L are independent random variables with possible outcomes in the range $\{1, \dots, M\}$. We assume that all outcomes are equally likely, i.e., $\mathbb{P}(J_l = j) = \frac{1}{M}$ for all $j \in \{1, \dots, M\}$. In App. B.1.1, we show that the estimator is unbiased, i.e., $\mathbb{E}[\langle \hat{\mathcal{B}} \rangle_\infty] = \langle \mathcal{B} \rangle$. To show that

To assess the significance of an observed violation, we consider the p value. We upper bound the p value with the help of Eq. (4.3). The probability on the right-hand side of Eq. (4.3) can be bounded with the help of concentration inequalities. Here, we consider Hoeffding's inequality, which yields

$$p \leq \mathbb{P}(\langle \hat{\mathcal{B}} \rangle_\infty - \langle \mathcal{B} \rangle \geq t) \leq \exp\left(-\frac{t^2}{2M^2}L\right). \quad (4.6)$$

The details of Hoeffding's inequality are presented in App. B.2.1. This result can be used to derive a lower bound on the necessary L . To reach a confidence level of $\gamma = 1 - p$,

$$L \geq \left\lceil -\frac{2M^2}{t^2} \ln(1 - \gamma) \right\rceil \quad (4.7)$$

random terms of the Bell operator have to be sampled.

4.3 Number of measurement repetitions

We now take into account that the expectation values cannot be inferred directly. Rather, the measurement of each chosen term B_j has to be repeated multiple times. In this section, we are interested in how many repetitions are necessary. For this purpose, we adjust the estimator in Eq. (4.5) to include the measurement repetitions. We assume that every term is measured K times. This yields the estimator

$$\langle \hat{\mathcal{B}} \rangle = \frac{M}{KL} \sum_{l=1}^L \sum_{k=1}^K b_{J_l}^{(k)}. \quad (4.8)$$

In the above estimator, $b_j^{(k)}$ denotes the measurement outcome of the term B_j in the k th repetition. Thus, we have $b_j^{(k)} \in \{\pm 1\}$. As in Eq. (4.5), J_1, \dots, J_L denote independent random variables that uniformly distributed take values in the range $\{1, \dots, M\}$. We also show in App. B.1.2 that the estimator in Eq. (4.8) is unbiased, i.e., $\mathbb{E}[\langle \hat{\mathcal{B}} \rangle] = \langle \mathcal{B} \rangle$. Finally, we can again use Eq. (2.19) and Hoeffding's inequality to bound the p value,

$$p \leq \mathbb{P}(\langle \hat{\mathcal{B}} \rangle - \langle \mathcal{B} \rangle \geq t) \leq \exp\left(-\frac{t^2}{2M^2}KL\right). \quad (4.9)$$

The detailed evaluation of Hoeffding's inequality is shown in App. B.2.2. However, we have already derived a lower bound for L . We thus obtain a lower bound on K from Eq. (4.9):

$$K \geq \frac{1}{L} \left\lceil -\frac{2M^2}{t^2} \ln(1 - \gamma) \right\rceil. \quad (4.10)$$

Since $L \geq \lceil -\frac{2M^2}{t^2} \ln(1 - \gamma) \rceil$, we observe that $K \geq 1$. As a result, we can conclude that it is sufficient to measure each term only once.

4.4 Analysis of the Bell inequalities for the GHZ and LC state

In this section, we are going to apply the method to the Bell inequalities for graph states. In particular, we will look at the Bell inequalities for the **GHZ** and the **LC** states. As described in Sec. 4.1, these Bell inequalities are promising to detect a large violation.

For this purpose, we express the observed expectation value as a fraction α of the quantum bound, i.e.,

$$\langle \hat{\mathcal{B}} \rangle = \alpha Q_n. \quad (4.11)$$

As we have seen in Sec. 1.4.3, to observe a violation of the Bell inequality we have to have $\alpha > (C_n/Q_n) \xrightarrow{n \rightarrow \infty} 0$. In case a violation is observed, it is $t = \alpha Q_n - C_n$. The number of random terms that are necessary to ensure a confidence level γ is given by Eq. (4.7) that takes the form

$$L \geq \left\lceil -\frac{2}{(\alpha - D_n^{-1})^2} \ln(1 - \gamma) \right\rceil. \quad (4.12)$$

We note that for the Bell inequalities in Sec. 1.4.3, the number of terms equals the quantum bound, i.e., $M = Q_n$. As D_n^{-1} converges to zero in the limit $n \rightarrow \infty$, we can conclude that L converges against some constant value for fixed α . This is shown for the Bell inequalities associated to the **GHZ** and **LC** states in Fig. 4.2.

For both the **GHZ** state and the **LC** state, we have assumed an observed violation of $\langle \hat{\mathcal{B}} \rangle = \alpha Q_n$ with $\alpha = 0.6$. This moreover implies that the assumed violation increases with the number of qubits n . We thus stress that Fig. 4.2 is only valid in case at least a value $\langle \hat{\mathcal{B}} \rangle = \alpha Q_n$ is observed and does not show the generic scaling of L with n . As an example, we can make the following observation:

Observation 1. *In case a violation $\langle \hat{\mathcal{B}} \rangle = \alpha Q_n$ of the Bell inequality associated to the GHZ state or the LC state for $n = 81$ qubits has been observed by sampling $L = 80$ random terms, the result has a confidence level of $\gamma = 5\sigma$.*

We note that for small n , L exceeds the number of contexts of the Bell operator. This, however, is not a contradiction as we do not exclude that a term is sampled multiple times.

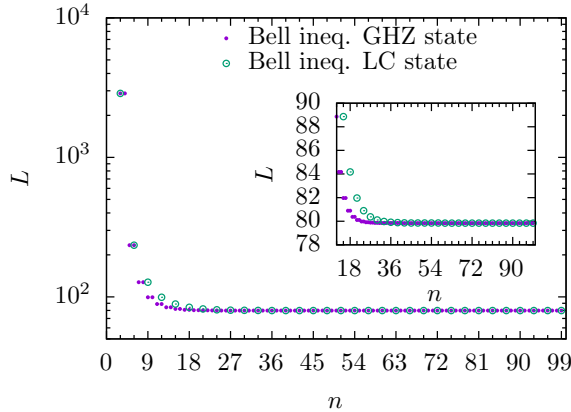


Figure 4.2: Necessary number of random observables, L , which are measured $K = 1$ times each, such that an observed violation of at least $\langle \hat{\mathcal{B}} \rangle = \alpha Q_n$ for $\alpha = 0.6$ has a confidence of $\gamma = 5\sigma$. The figure is taken from [P4].

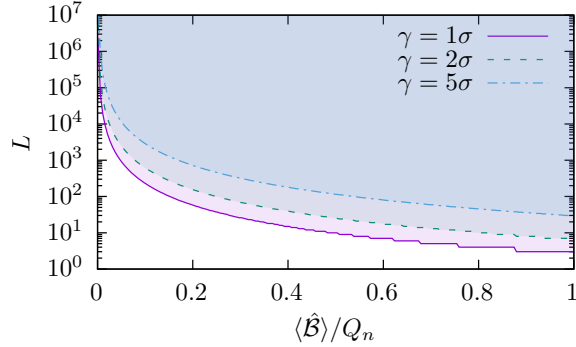


Figure 4.3: Confidence levels γ for an observed violation $\langle \hat{B} \rangle$ with L random terms ($K = 1$) in case the classical bound C_n is negligible compared to the observed violation $\langle \hat{B} \rangle$. The figure is taken from [P4].

In addition, Eq. (4.12) shows that with increasing L a decreasing violation $\langle \hat{B} \rangle = \alpha Q_n$ with $\alpha \sim \mathcal{O}(L^{-1/2})$ can be verified.

Finally, we see in Fig. 4.2 that already for relatively small n the classical bound C_n becomes negligible compared to the observed violation $0.6 \times Q_n$. Therefore, it does not have an effect on L for large n . We thus show the contours of the confidence levels $\gamma = 1\sigma, 2\sigma$ and 5σ in the limit $n \rightarrow \infty$ for a given observation $\langle \hat{B} \rangle$ with L random terms in Fig. 4.3.

4.5 Bell nonlocality as benchmark

We now discuss Bell nonlocality as a benchmark for quantum computers. We will investigate the Bell violations that can be achieved on current quantum computers. For this purpose, we first consider the connectivity graphs of different quantum computers and discuss what Bell inequalities can be used. Afterward, we include noise to assess realistic violations.

4.5.1 Connectivities of current quantum computers

We start by having a look at the connectivity graphs of different quantum computers and the Bell inequalities that can be evaluated on the different architectures. In Fig. 4.4, we show the connectivity graphs of a few current quantum computers. The first connectivity in Fig. 4.4 (a) is the star graph of five qubits, i.e., one central qubit connected to four other qubits. This layout is used, e.g., in the Starmon-5 quantum processor [123]. Fig. 4.4 (b) shows the connectivity graph that is used by IBM’s Falcon processor [124]. The ion trap quantum computer in Ref. [125] has 20 qubits that can all be coupled. The corresponding connectivity graph is shown in Fig. 4.4 (c). We also include the connectivity graphs of Google’s Sycamore processor in Fig. 4.4 (d) [91] that has 53 qubits and IBM’s Eagle processor [124] that has 127 qubits in Fig. 4.4 (e).

As the optimal Bell inequality that is associated to the connectivity graph is, in general, hard to determine, we will focus on the Bell inequalities for the GHZ and the LC states. In the following, we are interested in the largest GHZ and LC states that can be natively prepared on the different layouts in Fig. 4.4, i.e., that can be prepared by the available two-qubit gates only.

Observation 2. *On a quantum computer of n qubits, it is always possible to prepare a GHZ state of all n qubits with a circuit of $\mathcal{O}(n)$ depth.*

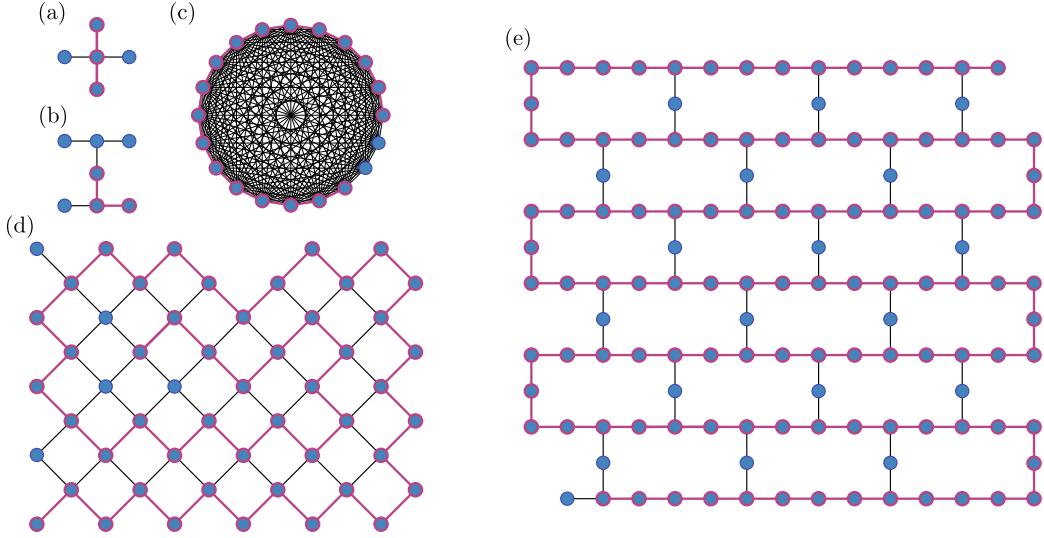


Figure 4.4: Connectivities of different quantum computers. The connectivity graph in (a) is used, for example, by the Starmon-5 quantum processor [123], whereas IBM’s Falcon processor [124] is based on layout (b). (c) The connectivity of the ion trap quantum computer in Ref. [125]. (d) The connectivity graph of Google’s Sycamore processor [91]; (e) IBM’s Eagle processor `ibm_brisbane` [124]. The figure is reprinted from [P4].

Proof. The connectivity graph of quantum computers is usually connected. Thus, by the following steps, a **GHZ** state of all n qubits can be prepared. The steps are illustrated in Fig. 4.5 for the architecture in Fig. 4.4 (b).

1. Prepare a star graph with center at the qubit with the largest connectivity [Fig. 4.5 (b)].
2. By performing two local complementations, the center of the star graph can be shifted to any node of the graph. Thus, the center can be moved to a vertex with still uncoupled neighbors [Figs. 4.5 (c) and 4.5 (d)].
3. By applying a CZ gate between the center node and the uncoupled neighbors, the adjacent qubits can be added to the **GHZ** state [Fig. 4.5 (e)].
4. Step 2 and 3 can be repeated until all qubits are coupled.

This procedure requires at least $n - 1$ consecutive CZ gates. In the worst case, there are two local complementations needed between all CZ gates. Combined with the initial Hadamard gates, the circuit depth is $3n + 1$. \square

We point out that the **GHZ** state can also be prepared in logarithmic step complexity [126, 127], depending on the connectivity.

For the **LC** state, in contrast, we make the following observation.

Observation 3. *Assume that the connectivity graph of a quantum computer is connected. Then, a linear cluster state containing all n qubits can be prepared with a circuit depth of $\mathcal{O}(n)$. In practice, however, it is often beneficial to prepare the **LC** state that is associated to the longest simple path in the connectivity graph [19]. The corresponding circuit has a constant depth of three, independent of the number of qubits.*

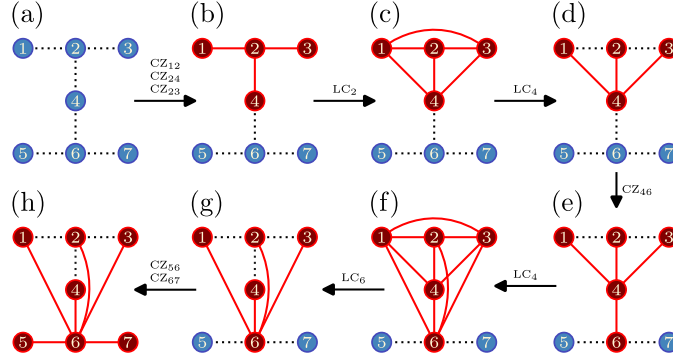


Figure 4.5: Preparation scheme for the GHZ state on the 7-qubit interaction topology that is, for example, used by IBM’s Falcon processor. The dotted lines indicate the physical CZ gates, whereas the graph state is drawn in red. The figure is reprinted from [P4].

Proof. We show in App. B.3 that it is always possible to prepare a LC state that contains all qubits in linear circuit depth. The longest simple path, in contrast, can be generated by first preparing all qubits in the $|+\rangle$ state, i.e., by applying Hadamard gates. Afterwards, every second CZ gate can be performed in parallel. The circuit depth is thus three, independent of the length of the simple path. The circuit is shown for the 6-qubit LC state in Fig. 4.6 (a). \square

As we only know the Bell inequality for the LC state with a number of qubits that is a multiple of three, we search for the longest path of length divisible by three. The longest paths that fulfill this restriction are drawn in pink in Fig. 4.4. The connectivity graphs in Figs. 4.4 (a) and 4.4 (b) allow one to prepare a 3-qubit LC state. The largest simple path on the 20 qubit full-connectivity graph in Fig. 4.4 (c) with length divisible by three has length 18. The quantum computer in Fig. 4.4 (c) thus allows one to check the Bell inequality for the 18 qubit LC state. To find the longest path is an NP-complete problem [128]. We can thus not verify if the marked paths for the layouts in Figs. 4.4 (d) and 4.4 (e) are indeed the longest paths. In Fig. 4.4 (d), we have identified a 48 qubit path as longest simple path. The layout in Fig. 4.4 (e) in turn allows the preparation of the 108 qubit LC state.

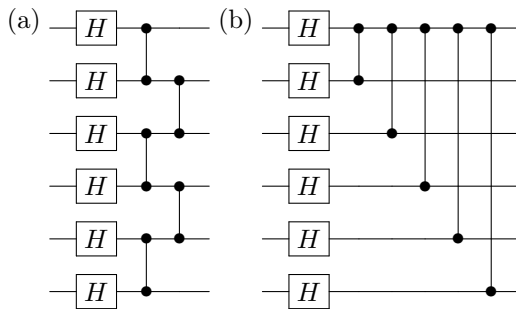


Figure 4.6: Quantum circuits to prepare (a) the LC state and (b) the GHZ state for $n = 6$ qubits. The circuits make use of the Hadamard gate denoted by H and CZ operations that are represented by the connected dots. The figure is reprinted from [P4].

	Single-qubit gate	Two-qubit gate	Readout
IBM Eagle	2.417×10^{-4}	7.409×10^{-3}	1.390×10^{-2}
Google Sycamore			
isolated	1.5×10^{-3}	3.6×10^{-3}	3.1×10^{-2}
simultaneous	1.6×10^{-3}	6.2×10^{-3}	3.8×10^{-2}

Table 4.1: Average error rates of IBM’s Eagle processor `ibm_brisbane` calibrated on 27.02.2024 18:32:14 [124] and Google’s Sycamore processor [91]. Google gives the error rates for the cases that the gates are performed isolated or simultaneously on all qubits.

4.5.2 Noise

In the following section, we will discuss the effect of noise. The goal is to roughly estimate the violation that can be realistically observed on quantum computers. Commonly, the errors of quantum computers are specified in terms of the error rates for single-qubit gates, two-qubit gates, and readout. The error rates for IBM’s Eagle processor `ibm_brisbane` and Google’s Sycamore processor are shown in Tab. 4.1. The error rate for the single-qubit gates is typically about an order smaller than the other error rates. We thus neglect the single-qubit errors. The readout error on the other side can be mitigated by classical postprocessing [129, 130]. We thus focus on the error in the two-qubit gates. For this purpose, we will consider a simple depolarisation noise model.

In the depolarization noise model, we assume that an error results, on average, in a maximally mixed state, i.e., the circuit for the graph state $|G\rangle$ prepares the mixture

$$\rho = p |G\rangle\langle G| + (1-p) \frac{\mathbb{1}}{2^n}. \quad (4.13)$$

The probability that no error in the two-qubit gates occurs is $p = (1 - p_2)^{N_2}$, where p_2 is the error rate and N_2 the number of the two-qubit gates. The observed violation is thus

$$\langle \mathcal{B} \rangle = p \times Q_n. \quad (4.14)$$

In Fig. 4.6, we show the circuits to prepare the **LC** and **GHZ** states in the case of $n = 6$ qubits. For the **LC** state, the CZ gates can be performed simultaneously and thus the circuit exhibits a constant depth of three, independent of the number of qubits. For the **GHZ** state, however, the preparation with CZ gates requires the gates to be consecutively applied. The circuit depth thus grows with the number of qubits. The number of CZ gates, however, is equal.

	Violation $\langle \mathcal{B} \rangle / Q$	L for $\gamma = 5\sigma$
Google Sycamore		
LC state ($n = 48$)	0.6913	41
GHZ state ($n = 53$)	0.7656	34
IBM <code>ibm_brisbane</code>		
LC state ($n = 108$)	0.4396	102
GHZ state ($n = 127$)	0.3800	136

Table 4.2: Results for the depolarization noise model that takes the 2-qubit gate error into account. For the noise data of Google’s Sycamore [91] and IBM’s Eagle processor [124], we show the violations in terms of the quantum bound Q . The last column shows the number of random terms, L , that have to be sampled to verify the violation with a confidence of 5σ .

The results in Tab. 4.2 show that the simple noise model predicts a violation of more than $0.6 \times Q$ for Google’s Sycamore processor and a bit less for `ibm_brisbane`. We note, however, that we considered more qubits on the IBM machine. A violation can thus be verified on both machines with $L \sim \mathcal{O}(10) - \mathcal{O}(10^2)$. This is of the same order as in Sec. 4.4. However, we stress that the considered noise model does not take the 1-qubit gate errors and readout errors into account. Real quantum computers, moreover, do not implement all gates natively. IBM’s Eagle processor, for example, does not support the CZ gate. Rather, it has to be composed of the available gates. The circuit in practice thus contains more gates and possibly exhibits a larger depth. For these reasons, the noise model overestimates the violation. However, it is still interesting to note that the noise model predicts the number of samples L to increase exponentially in n . This is shown in Fig. 4.7.

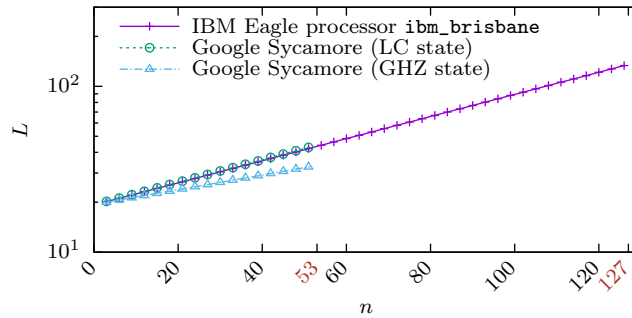


Figure 4.7: Number of terms, L , that have to be sampled to verify the violation predicted by the depolarization noise model with a confidence level of $\gamma = 5\sigma$. The figure is taken from [P4].

4.5.3 Simulation for an IBM quantum computer

Finally, we simulate the Bell inequalities of the **LC** and the **GHZ** states for the IBM Eagle quantum processor. For this purpose, we use the Qiskit AerSimulator [131] with the noise data of the quantum computer `ibm_brisbane` available at [124]. In Fig. 4.6, we show the ideal circuits to prepare the **LC** and the **GHZ** states for $n = 6$ qubits. The advantage of the **LC** state is that it can be prepared by a circuit of constant depth of three, whereas the step complexity for the **GHZ** state increases with the number of qubits, n . We note, however, that IBM’s Eagle processor does not implement the Hadamard and CZ gates natively. Rather, the gates have to be composed in terms of the available gate set. In practice, the circuits thus involve more gates and exhibit a larger depth.

After the preparation, L random terms of the corresponding Bell inequality are measured. The measurement of each random term is not repeated, i.e., $K = 1$. Fig. 4.9 shows the average expectation values of the Bell inequalities for the **LC** and **GHZ** states of up to $n = 24$ qubits. We have chosen $L = 800$ random terms and the average is taken over 10 repetitions. Moreover, the number of qubits is a multiple of three as only in this case is a good Bell inequality for the **LC** state known. Fig. 4.9 shows that for both states, the simulation predicts a Bell violation that increases exponentially with n . The **LC** state, however, shows a slightly higher relative violation compared to the **GHZ** state. This can be seen in Fig. 4.10 (a). The plot in Fig. 4.10 (a) displays the observed violation as a fraction of the total number of terms in the Bell inequality,

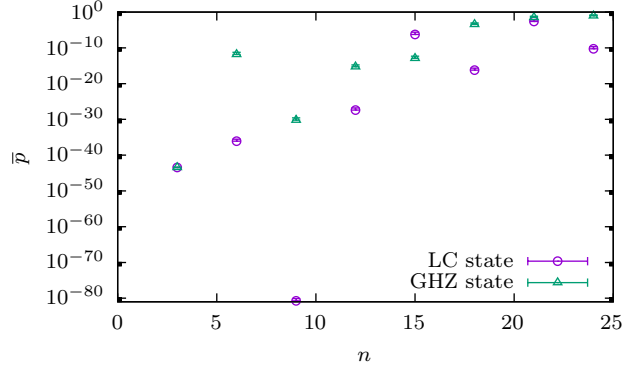


Figure 4.8: Average p value of the results for the **LC** and **GHZ** states in Fig. 4.9. The average is taken over the p values of the 10 repetitions. The error bars denote the standard deviation. We plot only the top error bars as we are interested in the uncertainty to larger p values. The figure is reprinted from [P4].

i.e., $t^2/M^2 = ((\hat{\mathcal{B}}) - C_n)^2/M^2$. We fit the data points to an exponential and obtain

$$\begin{aligned} \left(\frac{t^2}{M^2}\right)_{\text{LC}}(n) &= \exp(-0.10n - 0.65), \\ \left(\frac{t^2}{M^2}\right)_{\text{GHZ}}(n) &= \exp(-0.17n - 0.54). \end{aligned} \quad (4.15)$$

This affirms that the relative violation of the **GHZ** state decreases faster with n compared to the **LC** state. We attribute this to the larger circuit depth that is required for the **GHZ** state. The preparation of the **GHZ** state is thus more affected by noise. The smaller relative violation is also the reason for the larger p values for the **GHZ** state, which we present in Fig. 4.8. Except for the cases of $n = 3, 15$, the p values of the **LC** state are smaller, which implies a higher significance

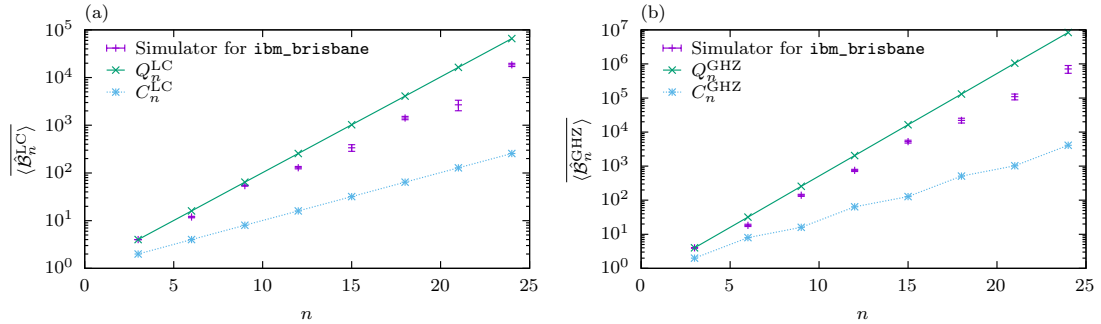


Figure 4.9: Simulation for the IBM Eagle quantum processor. The simulation uses the error rates of the real device `ibm_brisbane` [124] and the violation is estimated by measuring $L = 800$ random terms of the Bell inequality $K = 1$ times each. In (a) the average expectation value of the Bell inequality \mathcal{B}^{LC} is shown for the **LC** state, whereas (b) shows the violation of \mathcal{B}^{GHZ} for the **GHZ** state. In both cases, the expectation values are averaged over 10 repetitions and the error bars show the standard deviation. The figure is reprinted from [P4].

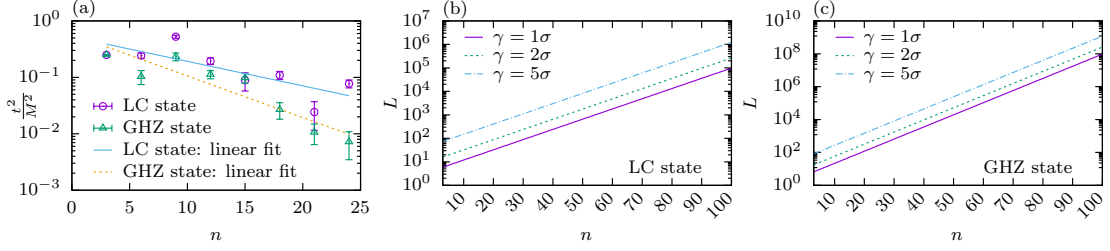


Figure 4.10: (a) Observed violation in relation to the number of terms in the Bell inequality, i.e., $t^2/M^2 = (\langle \hat{\mathcal{B}} \rangle - C_n)^2/M^2$. The data points are fitted by an exponential. From the fitted function t^2/M^2 , we estimate the number of random terms, L , that have to be sampled to reach a confidence of $\gamma = 1\sigma, 2\sigma, 5\sigma$. (b) The results for the **LC** state; (c) the predictions for the **GHZ** state. The figure is reprinted from [P4].

of the observed violation.

Finally, we use Eq. (4.15) to extrapolate the violation to larger n . Hoeffding's inequality in Eq. (4.9) yields the following for $K = 1$ and a target value for the confidence γ :

$$L(n, \gamma) \geq \left\lceil -\frac{2}{(t^2/M^2)(n)} \ln(1 - \gamma) \right\rceil. \quad (4.16)$$

The number of necessary sampled terms, L , is shown in Fig. 4.10 (b) for the **LC** state and in Fig. 4.10 (c) for the **GHZ** state. We first point out the different scaling of L compared to Fig. 4.2. Whereas in Fig. 4.2 L decreases with n until a plateau is reached, Fig. 4.10 suggests that L increases exponentially with n . This, however, can be attributed to fact that Fig. 4.2 has been obtained with the assumption that the relative violation stays constant with the number of qubits, n . But Fig. 4.10 (a) indicates that this assumption is not true for a real quantum computer. However, we still obtain that the number of terms, L , that have to be sampled is much smaller than the total number of terms, M .

As an example, we discuss L for the case of $n = 81$ qubits and a target confidence of $\gamma = 5\sigma$:

$$\begin{aligned} L_{\text{LC}}(n = 81, \gamma = 5\sigma) &= 182813, \\ L_{\text{GHZ}}(n = 81, \gamma = 5\sigma) &= 47170845. \end{aligned} \quad (4.17)$$

The above numbers are much smaller than $M_{\text{LC}} \approx 10^{16}$ and $M_{\text{GHZ}} \approx 10^{24}$, but larger than the values predicted in Sec. 4.4. As previous, we attribute this to the unrealistic assumption that, independent of n , a violation $\langle \mathcal{B}_n \rangle = 0.6 \times Q_n$ can be observed. For $n = 24$ qubits the simulation yields $\langle \hat{\mathcal{B}}_n^{\text{LC}} \rangle / Q_n^{\text{LC}} = 0.28$ for the **LC** state and $\langle \hat{\mathcal{B}}_n^{\text{GHZ}} \rangle / Q_n^{\text{GHZ}} = 0.09$ for the **GHZ** state, which is already much smaller than 0.6 in both cases.

4.6 Discussion

We have demonstrated a method to test n -partite Bell nonlocality on quantum computers. Quantum computers often have restricted two-qubit connectivity. We have thus pointed out that graph states are a natural choice of nonlocal states that can be readily prepared if the graph is a subgraph of the connectivity graph. Moreover, for certain graph states, good Bell inequalities are known. These Bell inequalities allow for an exponential violation of the classical bound, but in turn also typically require an exponential number of measurements. On the one hand, the exponential violation makes them increasingly robust to noise. On the other hand, it is impossible

to measure all terms in an experiment. We have solved this problem by proposing a method in the manner of randomized measurements, e.g., direct fidelity estimation [18, 19] or few-copy entanglement detection [119]. By sampling the terms of the Bell operator at random, the number of measurements can be drastically reduced. The violation can, however, still be verified with high significance. We have gauged the significance of a result with Hoeffding’s inequality. It thus depends on the violation that is observed. To assess the usefulness of the method on real devices, we have first used a simple depolarization noise model to estimate realistic violations. Finally, we have simulated the method for the IBM Eagle quantum processor. As expected with increasing accuracy of the noise model, the predicted violation shrinks. However, also, the simulator of the IBM processor predicts the number of terms that have to be sampled to be much smaller than the total number of terms in the Bell inequalities.

Our method will hence be useful to verify Bell violations in quantum systems of many qubits. This includes current quantum computers in the NISQ regime, e.g., the quantum computers accessible at IBM Quantum [124]. In addition, the observed Bell violation can be used to benchmark and compare different quantum computers. The Bell violation can be interpreted as a measure for the nonclassical correlations that can be produced. The preparation of the associated state depends on the connectivity of the quantum computer. We thus can benchmark the nonclassical correlations for states that require different levels of two-qubit connectivity. Finally, we stress that our method is not restricted to qubits and can be readily applied to Bell inequalities with higher local dimension.

Furthermore, the method could also be refined. For example, as the Bell inequalities only include stabilizers of the graph state, all observables commute. It might thus be feasible to find a (possibly very complicated) **POVM** to simultaneously measure all of the terms.

Moreover, it could be interesting to analyze the Bell inequalities with other statistical methods. Instead of the p value, one might look at the Kullback-Leibler divergence that has been used to assess the statistical strength of Bell inequalities for few parties [64].

The relation to other benchmarks, e.g., the quantum volume [132, 133] or the layer fidelity [134], is also yet to be explored. In particular, the layer fidelity can be measured by benchmarking a linear string of qubits of the quantum computer.

5 Complete characterization of quantum correlations by randomized measurements

In this section, we are going to discuss how **LU** invariants can be evaluated with the help of randomized measurements. In particular, for two-qubits, we derive explicit expressions for the Makhlin invariants, that have been introduced in Sec. 1.5.3. Moreover, we note that the eigenvalues of a density matrix are invariant under unitary transformations and we show that they can be derived from the **LU** invariants. As an application we discuss how Bell nonlocality and the teleportation fidelity can be obtained from randomized measurements. The section is based on [P2], and some subsection have been taken on.

5.1 Introduction

We have discussed in Sec. 1.5 that entanglement is invariant under local unitary transformations. Entanglement is thus independent of the choice of the local bases. Moreover, this suggests that entanglement can be characterized by **LU** invariants. We are mainly concerned with a bipartite system of two qubits that is described by a density operator ρ_{AB} . From the definition in Sec. 1.5.3, a function f of the density matrix is **LU** invariant if $f(\rho_{AB}) = f((U_A \otimes U_B)\rho_{AB}(U_A^\dagger \otimes U_B^\dagger))$. As a result, the function value can be inferred by sampling the function with random unitaries that are chosen according to the Haar measure, i.e.,

$$f(\rho_{AB}) = \iint dU_A dU_B f(U_A \otimes U_B \rho_{AB} U_A^\dagger \otimes U_B^\dagger). \quad (5.1)$$

Especially, f is **LU** invariant if it can be expanded in terms of expectation values of **LU** invariant observables \mathcal{M}_i , i.e.,

$$f(\rho_{AB}) = \sum_{\vec{t}} c_{\vec{t}} \langle \mathcal{M}_1 \rangle^{t_1} \langle \mathcal{M}_2 \rangle^{t_2} \dots, \quad (5.2)$$

where $\vec{t} = (t_1, t_2, \dots)$ denotes the vector of the indices t_i . As the observables are **LU** invariant they can be obtained from randomized measurements. The expectation values can be identified with the moments of the outcome distribution:

$$f(\rho_{AB}) = \sum_{\vec{t}} c_{\vec{t}} \mathcal{R}_{\mathcal{M}_1}^{(t_1)}(\rho_{AB}) \mathcal{R}_{\mathcal{M}_2}^{(t_2)}(\rho_{AB}) \dots \quad (5.3)$$

For a bipartite observable the t th moment in Eq. (1.61) takes the form

$$\mathcal{R}_{\mathcal{M}}^{(t)}(\rho) := \iint dU_A dU_B \{ \text{tr}[(U_A \otimes U_B)\rho(U_A^\dagger \otimes U_B^\dagger)\mathcal{M}] \}^t. \quad (5.4)$$

In the literature, randomized measurements have been investigated for product observables in Refs. [22, 25, 73, 135, 136]. As the moments are obtained from randomized measurements that are sampled according to the Haar measure, the method does not require a shared reference frame. In this sense, randomized measurements require less control over the experiment. This simplifies the experimental implementation. Moreover, we have discussed in Sec. 1.5 that randomized measurements are insensitive to local unitary noise [25]. However, it has to be guaranteed that the measurements are indeed sampled according to the Haar measure in an experiment. As an alternative, the moments can also be inferred from unitary t -designs [137, 138]. Unitary t -designs in turn require the implementation of specific unitaries. The original scheme of shadow tomography in Ref. [26] is also formulated in terms of random unitaries that are sampled from

a fixed finite set. We note, however, that the aim of shadow tomography is to infer arbitrary properties of the quantum state, whereas from the moments in Eq. (5.4) only LU invariant quantities can be derived.

In this work we extend the standard schemes and consider also randomized measurements with non-product observables. We point out that non-product observables can always be written as a sum of product observables. The moments for non-product observables can hence be obtained from the data of multiple product observables.

The moments of the outcome distribution are the quantities that can be directly measured. Whereas in principle any (polynomial) LU invariant can be expanded in terms of the moments, so far only a subset of these moments have been utilized for specific tasks like entanglement detection [23, 24, 139, 140] or fidelity estimation [18, 140].

In this work we present a general framework to express LU invariants in terms of the moments of randomized measurements. For finite dimensional systems of local dimension d , the set of polynomial LU invariants is finitely generated [141, 142]. This simplifies the discussion as it is sufficient to consider the generators. For the case of two qubits and special classes of higher-dimensional systems, the complete set of generators has been characterized [77, 143]. In the first part of the section, we reflect on how all two-qubit invariants can be expressed in terms of the moments of randomized measurements. Based on this connection all invariants can be inferred from randomized measurements. We explain by the example of the Kempe invariant in three qubits [144, 145] that the approach can also be applied to higher-dimensional systems. Finally, we test our method by an experiment with two qubits. We show that from the obtained invariants, Bell nonlocality can be verified and the usefulness of the state for teleportation can be assessed. Especially, we discuss the statistical analysis of the results.

5.2 LU invariants from the moments of randomized measurements

The goal is to experimentally evaluate the Makhlin invariants that have been introduced in Sec. 1.5.3. For this purpose we reproduce the expressions of the Makhlin invariants in terms of the moments of randomized measurements [P2, 146].

Let us now state explicitly how to measure the LU invariants in a randomized measurement scheme. To that end, recall that in order to observe the moments in Eq. (5.4), one has to choose an appropriate observable. Here, we show how to choose it in order to obtain the invariants I_1 , I_2 and I_3 .

As a first example, we explore the moments for the choice $\mathcal{M} = Z \otimes Z$. Note that choosing any other combination of Pauli matrices yields the same results, as they are related by local unitary rotations. For this choice, the first moment vanishes and we obtain as the second moment

$$\mathcal{R}_{Z \otimes Z}^{(2)} = \frac{1}{9} \text{tr}(TT^T) = \frac{1}{9} I_2, \quad (5.5)$$

where the occurring integrals can be solved using Weingarten calculus [P2, 147].

Next, $t = 3$ yields again zero (as well as any odd moment). For $t = 4$, we obtain

$$\begin{aligned} \mathcal{R}_{Z \otimes Z}^{(4)} &= \frac{1}{75} [2 \text{tr}(TT^T TT^T) + \text{tr}(TT^T)^2] \\ &= \frac{1}{75} [2I_3 + I_2^2], \end{aligned} \quad (5.6)$$

giving access to the invariant I_3 .

Finally, as $I_1 = \det(T)$ flips sign under partial transposition, we consider the non-product observable $\mathcal{M}_{\det} = \sum_{i=1}^3 \sigma_i \otimes \sigma_i$ and $t = 3$. Note that even though the observable is non-product, the moments can still be obtained by local measurements, as the expectation value can

be obtained from the three measurements $X \otimes X$, $Y \otimes Y$, $Z \otimes Z$ for a fixed choice of unitaries. The corresponding moment yields

$$\mathcal{R}_{\mathcal{M}_{\text{det}}}^{(3)}(\rho) = \det(T) = I_1. \quad (5.7)$$

Similarly, we can find appropriate observables to express also the remaining Makhlin invariants in terms of moments of randomized measurements [P2].

5.3 Experimental setup

We experimentally verify the functionality of the proposed randomized measurement method for the two-qubit case using polarization-entangled photon pairs. A schematic of our experimental setup is shown in Fig. 5.1 (a). The **entangled photon source (EPS)** generates signal and idler photon pairs via four-wave mixing in a **dispersion shifted fiber (DSF)** [148]. The DSF is pumped with a 50 MHz pulsed fiber laser centered at 1552.52 nm and is arranged in a Sagnac loop with a **polarizing beam splitter (PBS)** to entangle the signal and idler in polarization. The photons are spectrally demultiplexed into 100 GHz-spaced channels on the **International Telecommunication Union (ITU)** grid after the Sagnac loop, resulting in photons with a temporal duration of about 15 ps [149, 150]. For the experiment described here, ITU channels 27 (1555.75 nm) and 35 (1549.32 nm) are used. The source is tunable and typically outputs $\mu = 0.001 - 0.1$ pairs per pump pulse. Each photon is detected with gated InGaAs detectors with detection efficiencies of $\eta \sim 20\%$ and dark count probabilities of $\sim 4 \times 10^{-5}$ per gate [151, 152].

Given the polarization-entangled state generated by our source, we must implement random local unitary rotations in the form of random polarization state rotations. Therefore, we utilize the scrambling function of automated polarization controllers in order to apply random polarization rotations (for the remainder of the paper, a polarization controller operating in scrambling mode will be referred to as a polarization scrambler).

After verifying that the polarization scramblers can be used to apply sufficiently-random unitaries, we measured unbiased estimators (see [P2] App. C) for the I_1 , I_2 , and I_3 invariants via Eqs. (5.5)-(5.7). Each polarization scrambler was set to rotate incident light to a random polarization state (therefore, acting as a random unitary), and coincidences were measured in different bases. To that end, we define for each of the two parties $i = 1, 2$ the local bases $\{|H\rangle_i, |V\rangle_i\}$ of horizontally and vertically polarized light, $\{|D\rangle_i, |A\rangle_i\}$ of diagonal and anti-diagonal polarized light and $\{|L\rangle_i, |R\rangle_i\}$ of left circular and right circular polarized light. Note that while we associate these bases to polarization states, the unitary invariance of the measured quantities allows us to choose any local bases, as long as they are rotated by $\frac{\pi}{2}$ on the Bloch sphere w.r.t. each other. In particular, the bases for measuring photons 1 and 2 do not need to be aligned, i.e. $|H\rangle_1$ and $|H\rangle_2$ do not need to be equivalent on their respective Bloch spheres. We then measured in each combination of these local bases repeatedly for $M = 200$ different settings of the polarization scramblers, i.e. 200 different random unitaries were applied, where for each of these settings, $K \approx 1500$ repetitions were used to measure the expectation value.

The method to estimate I_1 , I_2 , and I_3 from finite measurement results is discussed in detail in [P2] App. C. Although we collect measurement results in the $\{|H\rangle_i, |V\rangle_i\}$, $\{|D\rangle_i, |A\rangle_i\}$ and $\{|L\rangle_i, |R\rangle_i\}$ bases described above, the estimators for I_2 and I_3 only require projective measurements in a single joint basis. Therefore, those estimators are calculated using only a subset of the data, for example, the results for $|H\rangle_1|H\rangle_2$, $|H\rangle_1|V\rangle_2$, $|V\rangle_1|H\rangle_2$, and $|V\rangle_1|V\rangle_2$. On the other hand, the estimator for I_1 requires projective measurements in all three of the measured joint bases. After calculating the invariants, the experiment described above was repeated 25 different times to allow for a statistical analysis of the results.

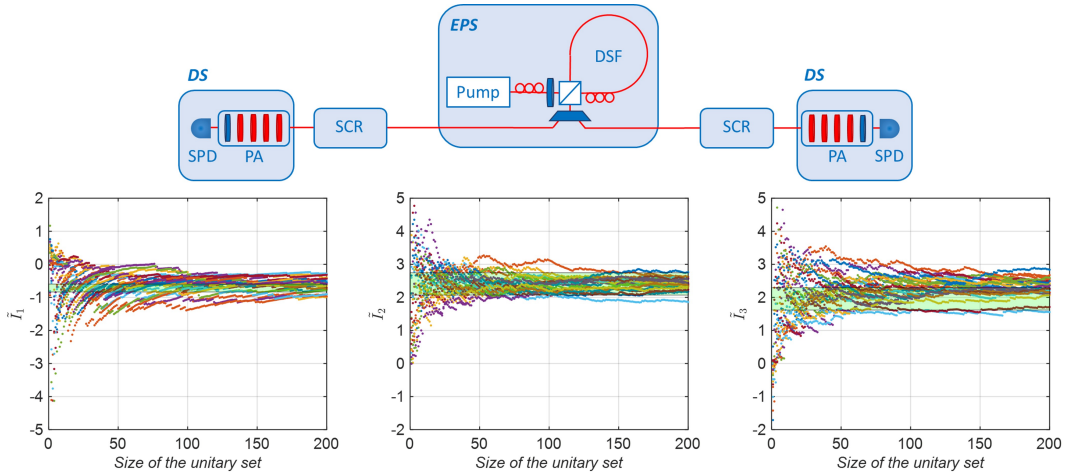


Figure 5.1: (a) Schematic diagram of the experimental setup for performing the randomized measurement protocol with polarization-entangled photon pairs. DS: detector station. DSF: dispersion-shifted fiber. EPS: entangled photon source. PA: polarization analyzer. SCR: polarization scrambler. SPD: single photon detector. (b, c, d) Experimentally determined unbiased estimators for the invariants I_1 , I_2 , and I_3 for 25 different runs consisting of measurements with 200 different random unitaries. The black lines show the expected values of each parameter calculated from the density matrix of the experimental system. The figure has been taken from [P2].

The experimentally determined estimators of the I_1 , I_2 , and I_3 invariants for all 25 runs (each run is shown in a different color) are shown in Fig. 5.1 (b), (c), and (d), respectively. The green band in all plots corresponds to the expected value of each invariant to allow for comparison with our method. The band represents the mean value plus or minus the standard deviation of each invariant calculated by performing quantum state tomography many times to characterize the state output by the EPS. A more-detailed description of how these expected values are calculated is found in App. E.3 in the Supplemental Material [P2]. The experimentally determined invariants converge near the expected values, therefore validating our randomized measurement protocol.

5.4 Applications

Finally, we show that the moments I_1 , I_2 and I_3 can be used to obtain the maximal Bell violation that can be achieved with the prepared quantum state. Moreover, the invariants can be used to assess the usefulness of the state in quantum teleportation. We first reproduce how the Bell violation and the teleportation fidelity can be derived from the moments [P2, 146]. We, afterward, discuss how the error can be analyzed, and give the results with the corresponding error bounds.

5.4.1 Detection of Bell nonlocality

The most straight-forward application is the evaluation of $I_2 = \text{tr}(TT^T)$, also known as the two-body sector length [153]. A quantum state is entangled if $I_2 > 1$, and the maximal value is $I_2 = 3$ for Bell states. Note that additional knowledge of $I_3 = \text{tr}(TT^T TT^T)$ allows for the detection of many more entangled states [24]. Combined knowledge of $I_1 = \det(T)$, $I_2 = \text{tr}(TT^T)$ and

$I_3 = \text{tr}(TT^T TT^T)$ is useful for completely determining if a state's nonlocality can be detected by a CHSH-like inequality: Given the observable [154]

$$\mathcal{B} = \sum_{i,j=1}^3 [a_i(c_j + d_j) + b_i(c_j - d_j)]\sigma_i \otimes \sigma_j, \quad (5.8)$$

where \vec{a} , \vec{b} , \vec{c} and \vec{d} are real, normalized vectors, its expectation value is bounded by 2 for local states. For a given quantum state, the maximum expectation value that one can observe by varying the vectors that define the observable is given by $2\sqrt{\lambda_1^2 + \lambda_2^2}$, where λ_1 and λ_2 are the two largest singular values of the correlation matrix T [155]. Thus, the quantity

$$\text{CHSH}(\rho) = 2\sqrt{\lambda_1^2 + \lambda_2^2} - 2 \quad (5.9)$$

measures the observable violation.

As the squares of the singular values of T coincide with the eigenvalues of TT^T , we can obtain them by measuring the coefficients of the characteristic polynomial

$$p_T(x) = x^3 - \text{tr}(TT^T)x^2 - \frac{1}{2} [\text{tr}(TT^T TT^T) - \text{tr}(TT^T)^2]x - \det(T)^2, \quad (5.10)$$

which are LU invariants, and calculating its roots. However, some care is needed to properly transfer statistical errors from finite statistics of the invariants to the roots of this polynomial. We explain the data analysis methods in Sec. 5.4.3.

5.4.2 Teleportation fidelity

As a second figure of merit, we can decide whether a given two-qubit state is useful in a teleportation protocol. There, the maximal fidelity f_{\max} of the teleported state is given by [156]

$$f_{\max} = \frac{F_{\max}d + 1}{d + 1}, \quad (5.11)$$

where in our case $d = 2$ and F_{\max} is the maximal overlap of the distributed state with the maximally entangled state $|\phi^+\rangle = \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle)$ under local operations and classical communication. As LU rotations constitute a subset of these, we can lower bound F_{\max} by optimizing over LUs instead, yielding [157]

$$F_{\max} \geq F_{\max}^U := \frac{1}{4} \max\{1 - \lambda_1 - \lambda_2 - \lambda_3, 1 - \lambda_1 + \lambda_2 + \lambda_3, 1 + \lambda_1 - \lambda_2 + \lambda_3, 1 + \lambda_1 + \lambda_2 - \lambda_3\}. \quad (5.12)$$

By examining the invariants I_1 , I_2 and I_3 , we can minimize F_{\max}^U over all singular values λ_i which are compatible with the observed data, giving a lower bound on the teleportation fidelity of the prepared quantum state.

5.4.3 Error bounds of the moments

In order to derive confidence levels for the quantities that we derive, we apply Hoeffding's inequality (see also Sec. 2.4), stating that for a set of n statistically independent random variables

$\{X_1, \dots, X_n\}$, where each $X_i \in [a_i, b_i]$, the probability of observing their sum deviating more than δ from the mean value is bounded by [85]

$$\mathbb{P}\left(\left|\sum_i X_i - \mathbb{E}[\sum_i X_i]\right| \geq \delta\right) \leq 2 \exp\left(-\frac{2\delta^2}{\sum_i (b_i - a_i)^2}\right). \quad (5.13)$$

We will use this inequality for independent, subsequent runs of the same inequality, i.e., $a_i \equiv a$, $b_i \equiv b$ for all i , and are interested in the mean, which demands a rescaling of $X_i \rightarrow X_i/n$. By fixing the right-hand side to a desired error probability $1 - \gamma$, where γ denotes the confidence, we obtain the following upper bound for the deviation δ :

$$\delta = \frac{b - a}{\sqrt{2n}} \sqrt{\ln\left(\frac{2}{1 - \gamma}\right)}. \quad (5.14)$$

This allows us, for any given estimator of an experimental quantity which yields numbers in the range $[a, b]$, to derive the maximal deviation of the true mean value given a certain confidence level of γ . We will use this to bound the **LU** invariants $\text{tr}(TT^T)$, $\text{tr}(TT^T TT^T)$ and $\det(T)$, needed to determine the roots of the characteristic polynomial in Eq. (5.10) in the main text.

What is left is to translate the confidence region of these coefficients into confidence regions of the roots. For that, we first give a naive bound on the confidence level that all of the measured quantities lie within their individual confidence levels.

To that end, consider first two random variables x and y , the experimental values of which being with confidence level γ in the regions $[x_0, x_1]$ and $[y_0, y_1]$, respectively. This is depicted in Fig. 5.2, where the confidence intervals of the two variables split the graph into 9 different regions, where the symbols a, b, \dots, i denote the probability to find a pair of measurement results in the corresponding region. We have $d + e + f = b + e + h = \gamma$ and $a + b + c + g + h + i = a + c + d + f + g + i = 1 - \gamma$. The probability to find an experimental value outside of at least one of the confidence intervals is given by $a + b + c + d + f + g + h + i = 2(1 - \gamma) - (a + c + g + i) \leq 2(1 - \gamma)$. Therefore, the probability to have measurement results within both confidence bands is at least $1 - 2(1 - \gamma)$. The same argument can be applied to more than two variables, $x^{(1)}, \dots, x^{(n)}$, where

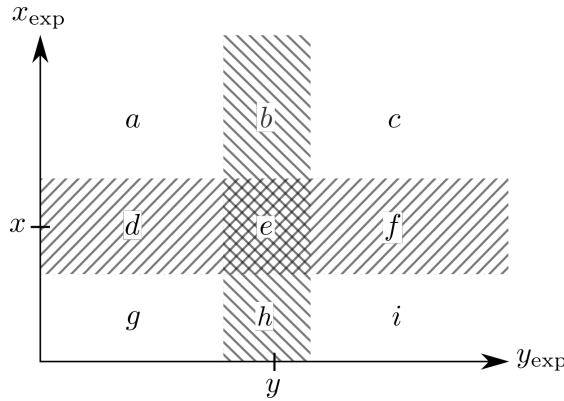


Figure 5.2: Illustration of the estimate in the text for two random variables: The probability to find experimental values inside of region e is lower-bounded by the expression in Eq. (5.15). The figure is taken from [P2].

each variable $x^{(i)}$ lies in the interval $[x_0^{(i)}, x_1^{(i)}]$ with confidence γ , yielding

$$\mathbb{P}(\forall i : x_0^{(i)} \leq x^{(i)} \leq x_1^{(i)}) \geq 1 - n(1 - \gamma). \quad (5.15)$$

Finally, we determine the expression in Eq. (5.9) for each choice of coefficients in their corresponding confidence regions to obtain a range of violations, in which the true violation lies with confidence level $1 - 3(1 - \gamma)$.

5.4.4 Results for the Bell violation and teleportation fidelity

In order to certify the achieved violation of the CHSH inequality using Eq. (5.9), we first have to determine appropriate confidence intervals for the three LU invariants $\text{tr}(TT^T)$, $\det(T)$ and $\text{tr}(TT^T TT^T)$, respectively. To do so, we follow two approaches, the first of which assumes that the coarse-graining of all 25×200 data points in 25 groups is enough to justify that the resulting data be normally distributed. Calculating the mean values and standard deviations based on this assumption yields

$$\det(T) = -0.62 \pm 0.15, \quad (5.16)$$

$$\text{tr}(TT^T) = 2.41 \pm 0.15, \quad (5.17)$$

$$\text{tr}(TT^T TT^T) = 2.21 \pm 0.21, \quad (5.18)$$

with 3σ , i.e., 99.73% confidence levels.

Alternatively, we can drop the assumption of normally distributed data and use the Hoeffding inequality to determine appropriate error bounds [85]. For instance, the sector length $\text{tr}(TT^T)$ can be directly expressed as a sample mean $\sum_{i=1}^M X_i/M$ of the squared correlation function $X_i = \langle U_A^\dagger Z U_A \otimes U_B^\dagger Z U_B \rangle^2 = 9(p_{00}^{(i)} - p_{01}^{(i)} - p_{10}^{(i)} + p_{11}^{(i)})^2$, with $X_i \in [0, 9]$. Applying the Hoeffding inequality to this case allows us to assign, with confidence γ , the following two-sided error bound:

$$\delta = \frac{9}{\sqrt{2M}} \sqrt{\ln\left(\frac{2}{1-\gamma}\right)}, \quad (5.19)$$

which for $M = 25 \times 200 = 5000$ and $\gamma = 0.9973$ (3σ) leads to

$$\text{tr}(TT^T) = 2.41 \pm 0.24. \quad (5.20)$$

Using the Hoeffding inequality for the determinant, we obtain

$$\det(T) = -0.62 \pm 1.09, \quad (5.21)$$

which is significantly worse compared to the Gaussian estimate. Lastly, for $\text{tr}(TT^T TT^T)$, we cannot apply the same arithmetic, as this quantity does not originate directly from a sample average over the runs but instead also involves the square of the respective sector length $\text{tr}(TT^T)^2$. As a workaround, we exploit the insight that the range of physically allowed values of the quantity $\text{tr}(TT^T TT^T)$ is constrained by the value of $\text{tr}(TT^T)$. Thus, we can derive a region of compatible values of $\text{tr}(TT^T TT^T)$ from the confidence region of $\text{tr}(TT^T)$, i.e. $\text{tr}(TT^T) = 2.41 \pm 0.24$. Following this procedure, we obtain

$$\text{tr}(TT^T TT^T) = 2.00 \pm 0.42. \quad (5.22)$$

We now have two sets of results including 3σ confidence levels for the three invariants under consideration. We can thus proceed to calculate the roots of the characteristic polynomial Eq. (5.10), and, respectively, the achievable violation of the CHSH inequality. In order to determine the best permissible value of the latter, we scan the whole range of allowed values of the three invariants and calculate the corresponding roots and CHSH violation for each of them. Finally, we use the largest violation which is still compatible with the observed data. The confidence of this violation is then given by $1 - 3(1 - \gamma) = 0.991 \approx 2.6\sigma$, as detailed in Sec. 5.4.3.

Following the above procedure, we obtain the following CHSH violations:

$$\text{CHSH}_{\text{Gauss}} \geq 0.46, \quad (5.23)$$

$$\text{CHSH}_{\text{Hoeff}} \geq 0.40. \quad (5.24)$$

Note that the maximal observable value is given by $2\sqrt{2} - 2 \approx 0.83$. Similarly, by requiring a higher confidence level of 5σ for the invariants, or equivalently $\gamma = 0.9999994$, we obtain

$$\text{CHSH}_{\text{Gauss}} \geq 0.42, \quad (5.25)$$

$$\text{CHSH}_{\text{Hoeff}} \geq 0.34, \quad (5.26)$$

with confidence $1 - 3(1 - \gamma) = 0.999998 \approx 4.7\sigma$. Using either method, our results clearly show that the randomized measurement protocol successfully determines that the state output by our **EPS** has the potential to violate a CHSH inequality.

The same error analysis can be used to obtain the lower bound F_{max}^U in Eq. (5.12) in the main text from the confidence intervals of the **LU** invariants. For a confidence level of 3σ of the **LU** invariants, we obtain a lower bound of at least:

$$(F_{\text{max}}^U)_{\text{Gauss}} = 0.88, \quad (5.27)$$

$$(F_{\text{max}}^U)_{\text{Hoeff}} = 0.85, \quad (5.28)$$

with a confidence level of $1 - 3(1 - \gamma) = 0.991 \approx 2.6\sigma$. Similarly, we obtain the lower bounds

$$(F_{\text{max}}^U)_{\text{Gauss}} = 0.86, \quad (5.29)$$

$$(F_{\text{max}}^U)_{\text{Hoeff}} = 0.60, \quad (5.30)$$

for a confidence level of 5σ of the **LU** invariants. The confidence level of the bounds is $1 - 3(1 - \gamma) = 0.999998 \approx 4.7\sigma$.

5.5 Conclusion

In this section, we have shown that all **LU** invariants can be directly inferred from randomized measurements of appropriately chosen **LU** observables. For the special case of two qubits, we have derived explicit expressions of the **LU** invariants in terms of the moments of randomized measurements.

We have demonstrated the introduced method by an experiment with entangled photon pairs. The experiment has given access to the **LU** invariants I_1 , I_2 , and I_3 . The invariants in turn are connected to different properties of the quantum state. As applications, we have calculated from the invariants the maximal Bell violation that could be observed from the prepared state. Moreover, we have assessed the usefulness of the state for quantum teleportation. To back the experimental findings up we have considered the confidence levels of the results by a statistical analysis. Finally, we point out that as a byproduct of the investigation, we have introduced a

method to verify the degree of randomness of the implemented unitary transformation. This method is described in detail in [P2].

We emphasize the simplicity of the presented scheme which allows inferring several important properties of the underlying quantum state through a number of randomly assorted measurements. For this reason, it will be an interesting direction of future research to extend the present explicit constructions for two-qubit states also to higher-dimensional systems, which likely will find ample applications in quantum communication tasks. Also, it would be desirable to extend our approach to the characterization of nonlocal quantum channels and multiparticle quantum correlations such as multi-setting Bell nonlocality or spin-squeezing.

6 Optimizing shadow tomography with generalized measurements

In the previous sections we have discussed how specific properties of quantum systems can be inferred. This included the direct measurement of spin-squeezing inequalities, Bell inequalities and LU invariants. This section in contrast discusses shadow tomography. The idea of quantum state tomography is to reconstruct the quantum state. Accordingly, any observable can be estimated. The aim of shadow tomography, however, is not to reconstruct the density matrix but to directly estimate expectation values from the tomographic data. This section is adapted from publication [P1].

6.1 Introduction

For quantum technology it is crucial to have precise control over isolated quantum systems. This refers to both the preparation and the manipulation as well as the readout of quantum states. To characterize a quantum system it is necessary to read out the relevant information. For this purpose, it is necessary to perform measurements on the system that ideally allow the estimation of arbitrary observables. A common approach is to characterize the state of the system by quantum state tomography [16]. The key idea of quantum state tomography is to perform a sufficiently large number of different measurements such that the density operator of the state can be reconstructed [158–162].

For large systems, however, it is impossible to reconstruct the density operator as the dimension increases exponentially with the number of particles. In practice, it turns out though that often not the entire information of the density operator is used later on [78]. This is for example the case if one is only interested in some properties, e.g., the expectation value of certain observables or the entropy. Ref. [78] thus has theoretically proposed shadow tomography as a method to avoid reconstructing the density operator and directly estimating the observables from the tomographic data. The practical scheme that was thereafter introduced in [26] has attracted a lot of attention.

Here, we will only briefly sketch the idea of the protocol. A more detailed description is given in Sec. 1.6. Whereas quantum state tomography is only considered useful when enough statistics is gathered such that the prediction are accurate, state estimators can already be applied to arbitrary sparse data [163]. This fact is, for example, used in data science and machine learning [164, 165]. From a single data point a noisy estimate can be derived. The noise can be reduced by averaging over all the data points. Even though the estimated state from sparse data can be very noisy and far away from the actual state, the noisy estimate can accurately predict certain observables [26, 78]. For the method to be computationally advantageous, however, it is crucial that the expectation values and other properties can be calculated without reconstructing the density operator [26]. Shadow tomography scales more easily to large systems than full tomography.

To collect data, Ref. [26] proposed to first apply unitaries that are randomly drawn from a specified set. The system is afterward measured in the computational basis. Equivalently, this method can be seen as choosing a random measurement from a chosen set. The technique has been applied in various scenarios that include energy estimation [166, 167], entanglement detection [140, 168], metrology [169], analyzing scrambled data [170], and quantum chaos [171]. Moreover, there have been additional improvements of the performance of the scheme [74, 140, 172–175] and the method has been generalized to channel shadow tomography [176, 177].

In this section, we formulate shadow tomography in terms of POVMs. This theoretical framework contains the randomization of unitaries as a special case and is thus a generalization

of the original scheme in [26]. Moreover, **POVMs** allow analyzing unavoidable errors in realistic quantum measurements [91, 178] that can not be easily described by projective measurements. We note that Ref. [179] has also proposed a procedure for shadow tomography with **POVMs**. The method in [179], however, uses the original construction of classical shadows in Ref. [26] by manually synthesizing the post-measurement states for **POVMs**. On the contrary, here we show that classical shadows for **POVMs** can be derived straightforwardly from the least-square estimator [163], which requires no further assumptions on the post-measurement states and contains that for ideal measurements as a special case. In this sense, our proposed framework is more general and at the same time simpler than randomization of unitaries.

6.2 Shadow tomography formulated in terms of POVMs

We are now going to discuss the main results of publication [P1], that has also been presented in [180]. In Sec. 1.6, we have described the implementation of shadow tomography that has been proposed in Ref. [26]. In this implementation, the state is measured after random unitaries from a fixed ensemble have been applied. This is equivalent to measurements in random bases. We have described in Sec. 1.3.2 that some **POVMs** can be implemented by random measurements. This suggests that shadow tomography can be formulated with **POVMs**.

6.2.1 Shadow tomography derived from the least square estimator

We consider a quantum system with dimension D that is composed of qubits. As described in Sec. 1.3.2, a **POVM** E is a set of effects $E = \{E_1, E_2, \dots, E_N\}$ that are positive and sum up to identity, i.e., $\sum_{k=1}^N E_k = \mathbb{1}$. Each effect is associated to an outcome k that is observed with probability $p_k = \text{Tr}(\rho E_k)$. A **POVM** thus maps a density operator ρ to a probability distribution, i.e., it defines the map $\Phi_E : M_D \rightarrow \mathbb{R}$,

$$\Phi_E(\rho) = (\text{Tr}(\rho E_k))_{k=1}^N, \quad (6.1)$$

where M_D denotes the set of density matrices in dimension D .

In contrast to **POVMs**, observables in quantum mechanics give rise to projective measurements. Projective measurements are a special class of **POVMs** whose effects consist of the D projectors that project on the eigenspaces of the observables. For example, the measurement of the Pauli operator σ_x corresponds to the **POVM** with the effects $\{|x^+\rangle\langle x^+|, |x^-\rangle\langle x^-|\}$. In turn, measuring the three Pauli observables σ_x, σ_y and σ_z at random with equal probability $1/3$ can be described by the **POVM** with effects $\{|x^\pm\rangle\langle x^\pm|, |y^\pm\rangle\langle y^\pm|, |z^\pm\rangle\langle z^\pm|\}$. This is explained in Sec. 1.3.2. The effects of the Pauli-**POVM** form an octahedron on the Bloch sphere, thus it is also referred to as octahedron **POVM**. The same construction can be used for other polytopes. A few examples of polytopes that are used to construct **POVMs** are shown in Fig. 6.1.

The goal in tomography is to estimate the density operator ρ from the observed measurement outcomes, i.e., from the frequencies the different outcomes occur. This can be described formally by an estimator χ , which is a map that transfers a probability distribution to a density operator, i.e., $\chi : \mathbb{R}^N \rightarrow M_D$. We require the estimator to be unbiased and linear. This means that in expectation, i.e., for the exact probabilities of the **POVM**, χ yields the correct quantum state ρ . As a result, $(\chi \circ \Phi_E) = \mathbb{1}_{M_D}$ acts as the identity over M_D .

A **POVM** is informationally complete if the density matrix can be uniquely derived from the statistics. If the **POVM** has just enough effects to form a basis for the operator space, it is not overcomplete. In this case Φ_E is invertible and $\chi = \Phi_E^{-1}$. An overcomplete **POVM** on the other hand does not uniquely define χ . It is thus reasonable to consider the estimate that minimizes

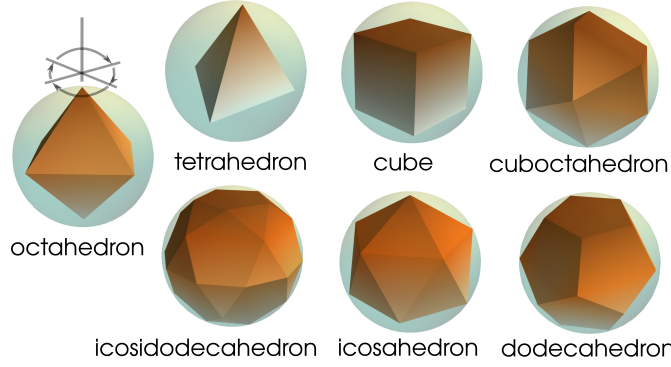


Figure 6.1: Examples of polytopes on the Bloch sphere that are used to define **POVMs**: octahedron ($N = 6$), tetrahedron ($N = 4$), cube ($N = 8$), cuboctahedron ($N = 12$), icosahedron ($N = 12$), dodecahedron ($N = 20$), icosidodecahedron ($N = 30$). The figure is reprinted from [P1].

the sum of squares, i.e.,

$$\chi_{\text{LS}}(\vec{p}) = \arg \min_{\tau} L(\tau), \quad \text{where} \quad L(\tau) = \sum_{i=1}^N [\text{tr}(\tau E_i) - p_i]^2. \quad (6.2)$$

The estimator χ_{LS} is called the least square estimator. L is a quadratic function and thus only takes a single minimum. At the minimum, the variation $\delta L = L(\tau + \delta\tau) - L(\tau)$ vanished up to linear order in $\delta\tau$. From the first order expansion in $\delta\tau$,

$$\delta L = 2 \sum_{i=1}^N [\text{tr}(\tau E_i) - p_i] \text{tr}(E_i \delta\tau), \quad (6.3)$$

it follows that $\delta L = 0$ for all choices of $\delta\tau$ if

$$\sum_{i=1}^N [\text{tr}(\tau E_i) - p_i] E_i = 0. \quad (6.4)$$

To rewrite the above equation in terms of the map Φ_E , we note that the adjoint $\Phi_E^\dagger : \mathbb{R}^N \rightarrow M_D$ is defined by

$$\langle \Phi_E(\rho), \mathbf{v} \rangle_{\mathbb{R}^N} \stackrel{!}{=} \langle \rho, \Phi_E^\dagger(\mathbf{v}) \rangle_{M_D}, \quad \text{for all} \quad \mathbf{v} \in \mathbb{R}^N \text{ and } \rho \in M_D, \quad (6.5)$$

where $\langle \cdot, \cdot \rangle_{\mathbb{R}^N}$ and $\langle \cdot, \cdot \rangle_{M_D}$ denote the inner products on \mathbb{R}^N and M_D , respectively. From the above equation, we get

$$\Phi_E^\dagger(\mathbf{v}) = \sum_{k=1}^N v_k E_k. \quad (6.6)$$

Eq. (6.4) can thus be written as

$$\Phi_E^\dagger[\Phi_E(\tau) - \vec{p}] = 0. \quad (6.7)$$

This yields the following expression for the least square estimator

$$\chi_{\text{LS}} = (\Phi_E^\dagger \Phi_E)^{-1} \Phi_E^\dagger. \quad (6.8)$$

We note that the solution is also valid in case Φ_E is invertible, $\chi_{\text{LS}} = \Phi_E^{-1}$. It is important to point out that the estimator χ_{LS} is indeed linear. We can thus easily check that χ_{LS} is an unbiased estimator. In expectation, the observed probabilities match the theoretical ones, i.e., $\Phi_E(\rho)$, and we get

$$\chi_{\text{LS}}(\Phi_E(\rho)) = (\Phi_E^\dagger \Phi_E)^{-1} \Phi_E^\dagger(\Phi_E(\rho)) = (\Phi_E^\dagger \Phi_E)^{-1}(\Phi_E^\dagger \Phi_E)(\rho) = \rho. \quad (6.9)$$

The estimator thus yields the correct ρ if the exact measurement statistics are known.

As the estimator is linear, it commutes with taking the average over the data. It is thus possible to associate an estimate to each data point. To clarify this, we denote the outcome k of a single measurement by the elementary statistics vector $\mathbf{q}_k = (\delta_{ki})_{i=1}^N \in \mathbb{R}^N$. By averaging over the complete dataset $\{\mathbf{q}_i\}_{i=1}^M$, the probabilities of the outcomes can be estimated, i.e.,

$$\hat{\rho} = \frac{1}{M} \sum_{i=1}^M \mathbf{q}_{k_i}. \quad (6.10)$$

The estimator χ_{LS} then splits into the average over the estimates for the single data points:

$$\chi_{\text{LS}}(\hat{\rho}) = \frac{1}{M} \sum_{i=1}^M \chi_{\text{LS}}(\mathbf{q}_{k_i}). \quad (6.11)$$

We identify the estimates that are associated to the elementary outcomes as the classical shadows, i.e., $\hat{\rho}_k := \chi_{\text{LS}}(\mathbf{q}_k)$. To simplify the notation, we note that for an informationally complete **POVM** the map $\Phi_E^\dagger \Phi_E$ is invertible, and we define

$$C_E(\rho) := \Phi_E^\dagger \Phi_E(\rho) = \sum_{k=1}^N \text{Tr}(\rho E_k) E_k. \quad (6.12)$$

We further note that for an elementary statistics vector the adjoint map in Eq. (6.6) takes the form $\Phi_E^\dagger(\mathbf{q}_k) = E_k$. As a result the classical shadows can be written as

$$\hat{\rho}_k := \chi_{\text{LS}}(\mathbf{q}_k) = C_E^{-1}(E_k). \quad (6.13)$$

Similar to the classical shadows in Ref. [26], the average over the shadows of an infinite number of measurements yields the state ρ , i.e., $\lim_{M \rightarrow \infty} \frac{1}{M} \sum_{i=1}^M \hat{\rho}_{k_i} = \rho$. The convergence alone, though, does not imply that χ_{LS} is a good estimator. Rather, this is suggested by the fact that χ_{LS} is the least square estimator. We also stress that the estimate $\hat{\rho}$ is not necessarily a valid density operator as it is not guaranteed that $\text{Tr}(\hat{\rho}) = 1$. This only holds if all effects have the same trace.

Finally, we note that it is essential for shadow tomography that the estimator is linear. This enables the estimation of observables from single data points, which, however, can be very noisy. A low number of samples, $M \ll D$, can still be sufficient to estimate many linear observables [26, 78].

6.2.2 Relation to randomized projective measurements

The definition of the classical shadows in Eq. (6.13) resembles the one given in Ref. [26], that we have discussed in Sec. 1.6. This poses the question of how the two definitions are related.

The scheme in Ref. [26] applies unitaries that are randomly chosen from a fixed set \mathcal{U} before a projective measurement in the computational basis is performed. In case the unitary U is applied,

and the system is measured to be in the computational basis state $|b\rangle$ with $b \in \{0, 1\}^n$, the post-measurement state corresponds to the state $U^\dagger|\hat{b}\rangle\langle\hat{b}|U$ before the unitary has been applied. The procedure can thus be seen as a projective measurement in the basis $\{U^\dagger|b\rangle \mid b = 1, \dots, D\}$. For a finite unitary ensemble \mathcal{U} of size $|\mathcal{U}|$, the measurement can be identified as a **POVM** with effects

$$\left\{ \frac{1}{|\mathcal{U}|} U^\dagger|\hat{b}\rangle\langle\hat{b}|U \mid U \in \mathcal{U}, b \in \{0, 1\}^n \right\}. \quad (6.14)$$

The randomized measurement approach can thus be viewed as an implementation of the above **POVM** [27, 31].

The formulation in terms of **POVMs** opens several interesting perspectives. We first note that the description of the randomized measurement approach is much more complicated than the formulation of the corresponding **POVM**. In Sec. 1.6, we have discussed that a common choice of the unitary ensemble is the group of local Clifford circuits $\text{Cl}(2)^{\otimes n}$. The Clifford group $\text{Cl}(2)$ for a single qubit has 24 elements [181]. The local Clifford circuits implement Pauli measurements that can be described by the octahedron **POVM** that has 6 outcomes corresponding to the effects $1/3 \times \{|x^\pm\rangle\langle x^\pm|, |y^\pm\rangle\langle y^\pm|, |z^\pm\rangle\langle z^\pm|\}$. There are thus more parameters needed for the randomized implementation, which makes it more complex and harder to optimize. Ultimately, it seems intractable to determine the optimal unitary ensemble. We show, however, that it is possible to optimize the **POVM**.

Furthermore, we have discussed in Sec. 1.3.2 that the octahedron **POVM** can also be implemented by measuring the Pauli observables at random. The same **POVM** can thus be implemented by different sets of unitaries. But it is not possible to simulate all **POVMs** by randomized measurements [31, 182].

Finally, we note that **POVMs** can also be implemented by coupling the system to an ancilla system by means of Naimark's theorem 1.8. The formulation with **POVMs** thus paves the way to implement shadow tomography in a different way.

6.2.3 Symmetry of generalized measurements and the computation of the classical shadows

In the previous section, we have derived the classical shadows from the least square estimates, that are associated to the elementary outcomes. To apply Eq. (6.13), we need to invert the map C_E . In this section, we show how the inverse map C_E^{-1} can be found from the symmetry of the **POVM**. A **POVM** $E = (E_1, E_2, \dots, E_N)$ is called symmetric if there is a subgroup G of the permutation group over $(1, 2, \dots, N)$ and a representation $U : G \rightarrow U(D)$ such that $E_{g(k)} = U_g E_k U_{g^{-1}}$ [183]. If E is symmetric under the action of G , it follows that $\Phi_E : M_D \rightarrow \mathbb{R}^N$ is covariant under G , i.e., $\Phi_E(U_g \rho U_{g^{-1}}) = g[\Phi_E(\rho)]$. The action of g on the probability distribution is defined as $(g[\mathbf{p}])_k = \mathbf{p}_{g^{-1}(k)}$. The adjoint map Φ_E^\dagger is accordingly covariant, i.e., $\Phi_E^\dagger[g(\mathbf{p})] = U_g \Phi^\dagger(\mathbf{p}) U_{g^{-1}}$ for all distributions \mathbf{p} . As a result, the map $C_E = \Phi_E^\dagger \Phi_E$ and its inverse C_E^{-1} are covariant, too. For all hermitian operators X and $g \in G$, it is $C_E^{-1}(U_g X U_{g^{-1}}) = U_g C_E^{-1}(X) U_{g^{-1}}$.

A **POVM** E is called uniform if there is a permutation G that acts transitively on the outcomes, i.e., for any two effects E_i, E_j there is a group element g such that $E_j = U_g E_i U_{g^{-1}}$. It directly follows that $\text{Tr}(E_i) = \text{Tr}(E_j) =: \alpha$ for all $i, j \in \{1, \dots, N\}$. The stabilizer subgroup G_k over k is the subgroup of G that leaves k invariant [183]. Since G_k leaves k invariant, E_k commutes with all representations of the group $U(G_k)$. As C_E^{-1} is covariant, this also holds for the classical shadows $\hat{\rho}_k = C_E^{-1}(E_k)$. E is called rigidly symmetric [183] if the representation U restricted to any stabilizer subgroup G_k over k has exactly two irreducible representations. An operator commuting with $U(G_k)$ can thus only be a linear combination of E_k and $\mathbb{1}$. As a result

d	ST	N	Comments	a	b
2	8	6	Octahedron	9	-1
		8	Cube	12	-1
		12	Cuboctahedron	18	-1
	16	12	Icosahedron	18	-1
		20	Dodecahedron	30	-1
		30	Icosidodecahedron	45	-1
3	24	21		28	-1
	25	12	csMUB	16	-1
	27	45		60	-1
		60		80	-1
4	28	12	Real MUB	9	$-\frac{1}{2}$
	29	20	csMUB	25	-1
		40		50	-1
		80		100	-1
	30	300		225	$-\frac{1}{2}$
	31	60		75	-1
		480		600	-1

Table 6.1: Parameters for the inverse of the measurement channels for rigidly symmetric POVMs constructed in Ref. [183] with the same naming convention of the symmetry groups. Each row corresponds to a single POVM, which is decomposed to several projective measurements in Ref. [183]. The decomposition can be used to simulate the POVM by randomized measurements.

if E is rigidly symmetric, then the classical shadows are of the form

$$\hat{\rho}_k = aE_k + b\mathbb{1}. \quad (6.15)$$

What is left is to determine the parameters a and b . For this purpose, we notice that $E_k = C_E(\hat{\rho}_k) = a \sum_{l=1}^N \text{Tr}(E_k E_l) E_l + b\alpha\mathbb{1}$. We can now consider the trace of this equation and also the trace of the equation multiplied from both sides with E_k . This yields

$$\alpha = a\alpha^2 + b\alpha D, \quad (6.16)$$

$$\beta = a\gamma + b\alpha^2, \quad (6.17)$$

with $\beta = \text{Tr}(E_k^2)$, $\gamma = \sum_{l=1}^N \text{Tr}(E_k E_l)^2$, both independent of k as E is uniform.

We can thus express the coefficients a and b in terms of properties of the considered POVM:

$$a = (D\beta - \alpha^2)/(D\gamma - \alpha^3), \quad (6.18)$$

$$b = (\gamma - \alpha\beta)/(D\gamma - \alpha^3). \quad (6.19)$$

Various rigidly symmetric measurements are investigated in Ref. [183] and we give the corresponding parameters in Tab. 6.1.

6.2.4 Tensoring the shadow construction for many-body systems

The idea of shadow tomography is to avoid reconstructing the density matrix such that also properties of large systems can be inferred. For this goal it is important to efficiently store and process the classical shadows. For this purpose, Ref. [26] used the stabilizer formalism [79]. In

this section, we discuss that for a **POVM** that is composed of generalized measurements for the single qubits, also the classical shadows are product states.

We consider a system with n qubits, such that the total dimension is $D = 2^n$. Moreover, we choose (identical or not identical) **POVMs** $\{E^{(1)}, E^{(2)}, \dots, E^{(n)}\}$ for each of the qubits, that are described by a collection of N_i effects $E^{(i)} = \{E_k^{(i)}\}_{k=1}^{N_i}$ each. The effects of the **POVM** E^{tot} on the whole system can thus be labeled by a string of outcomes $\mathbf{k} = \{k^{(1)}, k^{(2)}, \dots, k^{(n)}\}$. The corresponding effect is

$$E_{\mathbf{k}}^{\text{tot}} = E_{k^{(1)}}^{(1)} \otimes E_{k^{(2)}}^{(2)} \otimes \dots \otimes E_{k^{(n)}}^{(n)}. \quad (6.20)$$

We note that the combined **POVM** is rigidly symmetric, too, and it thus follows that also the classical shadows can be written as a tensor product:

$$\hat{\rho}_{\mathbf{k}}^{\text{tot}} = \hat{\rho}_{k^{(1)}}^{(1)} \otimes \hat{\rho}_{k^{(2)}}^{(2)} \otimes \dots \otimes \hat{\rho}_{k^{(n)}}^{(n)}, \quad (6.21)$$

where $\hat{\rho}_{k^{(i)}}^{(i)}$ being the classical shadow corresponding to the measurement $E^{(i)}$ on the i th qubit.

Observables that also split into a product of single-qubit observables can thus be readily estimated without computing the classical shadows of the full system. This would quickly become impossible as the full classical shadows are of size $D \times D$. For a product observable $X = X^{(1)} \otimes X^{(2)} \otimes \dots \otimes X^{(n)}$ a single outcome \mathbf{k} gives rise to a single estimate

$$\langle \hat{X} \rangle = \text{Tr}[\hat{\rho}_{k^{(1)}}^{(1)} X^{(1)}] \text{Tr}[\hat{\rho}_{k^{(2)}}^{(2)} X^{(2)}] \dots \text{Tr}[\hat{\rho}_{k^{(n)}}^{(n)} X^{(n)}]. \quad (6.22)$$

The final estimate of $\langle X \rangle$ is as usual obtained by averaging over all data points. It is thus not necessary to reconstruct the large density matrix of the full system.

6.2.5 Protocol of shadow tomography with generalized measurements

We will end this section by an overview of shadow tomography with a **POVM** $E = \{E_i\}_{i=1}^N$. The aim is to estimate a set of observables $\mathcal{X} = \{X_i\}_{i=1}^m$.

1. For measurement $E = \{E_i\}_{i=1}^N$, the classical shadows $\{\hat{\rho}_k\}_{k=1}^N$ are determined by Eqs. (6.13) or (6.15).
2. The data is collected by performing the **POVM** E on the desired quantum state. In total, the measurement is repeated M times such that a collection $\{k_j\}_{j=1}^M$ of M outcomes is recorded.
- 3a. The mean values of $\{X_i\}_{i=1}^m$ are estimated by $\langle X_i \rangle \approx 1/M \sum_{j=1}^M \text{Tr}(\hat{\rho}_{k_j} X_i)$.

We note that the procedure can also be used to estimate polynomial functions of ρ [26]. For example, the purity $\text{Tr}(\rho^2)$ can be estimated by adjusting the last step to

- 3b. The purity $\text{Tr}(\rho^2)$ is estimated by $\text{Tr}(\rho^2) \approx 1/[M(M-1)] \sum_{j_1 \neq j_2} \text{Tr}(\hat{\rho}_{j_1} \hat{\rho}_{j_2})$.

6.3 Statistical analysis: Shadow norm

After we have discussed the formulation of shadow tomography in terms of **POVMs**, this section is concerned with the statistical analysis. The classical shadows in Eq. (6.13) serve as the tomographic data in shadow tomography. They have to be efficiently stored and are later on used to estimate observables. Each classical shadow gives rise to a single estimate

$$\hat{x}_k = \text{Tr}(\hat{\rho}_k X). \quad (6.23)$$

Since the average of the classical shadows $1/M \sum_{i=1}^M \hat{\rho}_{k_i}$ converges to the actual state ρ , i.e., $\lim_{M \rightarrow \infty} 1/M \sum_{i=1}^M \hat{\rho}_{k_i} = \rho$, the average $1/M \sum_{i=1}^M \hat{x}_{k_i}$ converges to $\langle X \rangle$. We note that the asymptotic rate of convergence is related to the variance of the estimator. This can be seen for example from Cantelli's inequality in Thm. 2.11. For this reason, we will look at the variance $\text{Var}(\hat{x}_k)$ of a single data point \hat{x}_k . The variance is given by

$$\text{Var}(\hat{x}_k) = \mathbb{E}[\hat{x}_k^2] - \mathbb{E}[\hat{x}_k]^2 = \sum_{k=1}^N \mathbb{P}(k) \hat{x}_k^2 - \langle X \rangle^2 = \sum_{k=1}^N \text{Tr}(\rho E_k) \hat{x}_k^2 - \langle X \rangle^2, \quad (6.24)$$

where we have used that $\mathbb{P}(k) = \text{Tr}(\rho E_k)$ and $\mathbb{E}[\hat{x}_k] = \langle X \rangle$. By neglecting the second term and maximizing the the variance over all states, we find the upper bound

$$\begin{aligned} \text{Var}(\hat{x}_k) &\leq \sum_{k=1}^N \text{Tr}(\rho E_k) \hat{x}_k^2 = \sum_{k=1}^N \text{Tr}(\rho E_k) \text{Tr}(\rho_k X)^2 \leq \max_{\rho} \text{Tr} \left(\rho \sum_{k=1}^N \text{Tr}(\rho_k X)^2 E_k \right) \\ &= \lambda_{\max} \left\{ \sum_{k=1}^N \text{Tr}(\rho_k X)^2 E_k \right\} =: \|X\|_E^2, \end{aligned} \quad (6.25)$$

where $\lambda_{\max}\{\cdot\}$ denotes the maximal eigenvalue of the operator. $\|X\|_E$ is the shadow norm of the operator X . To show the equivalence to the definition of the shadow norm in Ref. [26], we use Eq. (6.14). The original scheme uses randomized projective measurements that correspond to a POVM with effects $\{E_{U,b} = 1/|\mathcal{U}| \times U^\dagger |b\rangle \langle b| U \mid U \in \mathcal{U}, b \in \{0,1\}^n\}$. In case the ensemble \mathcal{U} is finite, Eq. (6.25) can thus be written as

$$\|X\|_E^2 = \max_{\rho} \text{Tr} \left(\rho \sum_{k=1}^N \text{Tr}(\rho_k X)^2 E_k \right) = \max_{\rho} \sum_{U \in \mathcal{U}} \sum_{b \in \{0,1\}^n} \text{Tr} [C_E^{-1}(E_{U,b}) X]^2 \text{Tr}(\rho E_{U,b}). \quad (6.26)$$

The measurement channel \mathcal{M} in [26] is related to C_E , by $\mathcal{M}(\rho) = |\mathcal{U}| C_E(\rho)$. The inverse of C_E is thus $C_E^{-1}(\rho) = |\mathcal{U}| \mathcal{M}^{-1}(\rho)$ and we get by inserting the effects:

$$\begin{aligned} \|X\|_E^2 &= \max_{\rho} \sum_{U \in \mathcal{U}} \sum_{b \in \{0,1\}^n} \text{Tr} \left[|\mathcal{U}| \mathcal{M}^{-1} \left(\frac{1}{|\mathcal{U}|} U^\dagger |b\rangle \langle b| U \right) X \right]^2 \text{Tr} \left(\rho \frac{1}{|\mathcal{U}|} U^\dagger |b\rangle \langle b| U \right) \\ &= \max_{\rho} \frac{1}{|\mathcal{U}|} \sum_{U \in \mathcal{U}} \sum_{b \in \{0,1\}^n} \text{Tr}(\rho U^\dagger |b\rangle \langle b| U) \text{Tr} [\mathcal{M}^{-1}(U^\dagger |b\rangle \langle b| U) X]^2. \end{aligned} \quad (6.27)$$

The first sum can be identified with the average over the unitary ensemble. Moreover, we note that \mathcal{M}^{-1} is self-adjoint [26], i.e., $\text{Tr}(\mathcal{M}^{-1}(X)Y) = \text{Tr}(X\mathcal{M}^{-1}(Y))$. We can thus perform the traces and arrive at the definition of the shadow norm in Ref. [26]:

$$\|X\|_E^2 = \max_{\rho} \mathbb{E}_U \sum_{b \in \{0,1\}^n} \langle b| U \rho U^\dagger |b\rangle \langle b| U \mathcal{M}^{-1}(X) U^\dagger |b\rangle^2. \quad (6.28)$$

From the classical shadows not only one observable can be estimated but also a set of observables \mathcal{X} . We assume that all observables have the same physical unit. The shadow norm of \mathcal{X} can then be defined as the maximal shadow norm,

$$\kappa_E^2(\mathcal{X}) = \max\{\|X\|_E^2 : X \in \mathcal{X}\}. \quad (6.29)$$

As the shadow norm is an upper bound on the variance, the smaller κ_E^2 , the better is the accuracy of the estimates. The shadow norm is derived for the worst case scenario. We note, however, that the target state could be far away from the worst case scenario. It thus might be also informative to consider the variance in the average case.

6.4 Effects of noise in measurements

In realistic experiments not only the state of the system, but also the measurements are subjected to noise. Errors in the measurements stem from different sources, e.g., the setup of the measurement might be affected by noise or there is noise in the readout signal [91, 178].

If the implementation of the measurement goes wrong with probability p , and in this case yields a random result, we can describe the noisy measurement by a depolarization noise model. This can be readily formulated as a **POVM** with effects $\{p\mathbb{1}/N + (1-p)E_k\}_{k=1}^N$. The impact of depolarization noise on the shadow norm is shown in Fig. 6.2 (a) for classical shadows constructed from the tetrahedron and octahedron **POVM**. As observables, we have chosen 128 projectors on pure states that have been randomly sampled according to the Haar measure. Fig. 6.2 (b) shows that for small failure probability p the shadow norm increases only slightly. For larger p , however, the figure bears that κ_E^2 increases exponentially with p . What is not shown in the figure is that in fact κ_E^2 diverges for $p \rightarrow 1$. This is expected as in this case the measurements yield completely random outcomes and no information can be extracted.

The second case refers to noise in the readout signal [184, 185]. This can lead to misclassification of the outcomes and is thus called readout error. This kind of error is for example caused by decoherence during the measurement process or by inability to perfectly distinguish the states $|0\rangle$ and $|1\rangle$. We note, however, that readout error can be mitigated by classical postprocessing [129, 130, 184, 185].

In the following, we describe a model that assumes the readout error to affect the qubits independently, i.e., the model does not take crosstalk during the measurement into account. The readout noise is assumed to be uncorrelated and can be written as a tensor product. It is thus sufficient to consider the effect on a single qubit. We denote the probability that outcome 0 in the computational basis is misread as 1 by q_+ , whereas the opposite case that outcome 1 is misclassified as 0 occurs with probability q_- . The effect on one qubit can thus be written in terms of the matrix

$$A = \begin{pmatrix} 1 - q_+ & q_- \\ q_+ & 1 - q_- \end{pmatrix}, \quad (6.30)$$

where A_{ij} is the probability that outcome j is interpreted as outcome i . A 2-outcome measurement $E = (E_0, E_1)$ is therefore modified to a new measurement \tilde{E} with effects $\tilde{E}_0 = A_{00}E_0 + A_{01}E_1$ and $\tilde{E}_1 = A_{10}E_0 + A_{11}E_1$. We have discussed in Sec. 1.3.2 that the Pauli-**POVM** can be simulated by randomization of the Pauli observables. In this case, each Pauli observable is separately affected by readout errors. This in turn can be described by the **POVM** with effects $\{1/3 \times [(1 - q_{\pm}) |t^{\pm}\rangle\langle t^{\pm}| + q_{\mp} |t^{\mp}\rangle\langle t^{\mp}|], t = x, y, z\}$. This shows that the model for readout error depends on the implementation of the **POVM**. In case the **POVM** is implemented by coupling the system to n ancilla qubits, also the ancilla qubits are affected by readout noise. In principle, the noise acts differently on the qubits. The readout error can thus be described by the tensor product $A = A_1 \otimes \dots \otimes A_n$, where A_k describes the readout error for the ancilla qubit k . The effects of the noisy **POVM**, however, depend on the exact implementation by the ancilla system, such that no generic form can be given.

For the analysis of readout error it is convenient to define the average error rate $\bar{q} = (q_+ + q_-)/2$. The difference in the error rates can, accordingly, be described by $\epsilon = (q_+ - q_-)/(q_+ + q_-)$.

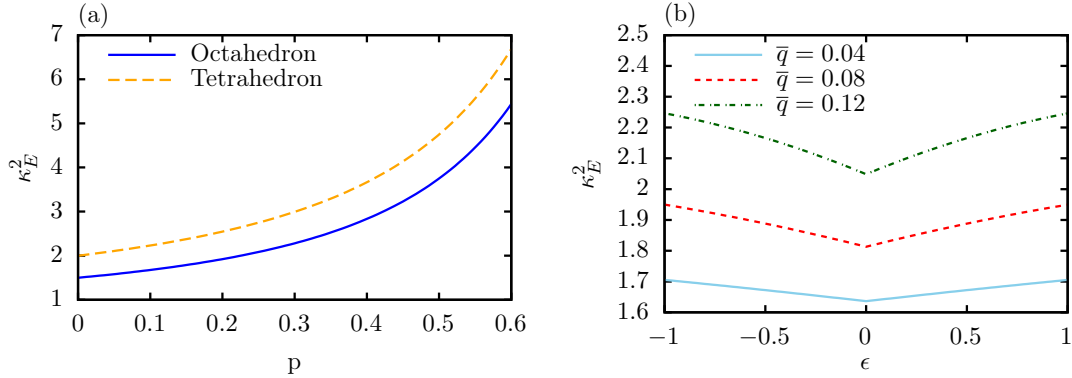


Figure 6.2: Shadow norm κ_E^2 in the presence of depolarization noise (a) and simple readout error noise (b). The maximal shadow norm is computed for 128 projections on pure states that are sampled randomly according to the Haar measure. For analogous results see Ref. [P1, 180].

Analogous to depolarization noise, we assess the quality of shadow tomography in the presence of readout error by the shadow norm κ_E^2 . For this purpose we have chosen $|\mathcal{X}| = 128$ projections of rank one that are sampled randomly according to the Haar measure as observables. Fig. 6.2 (b) shows the shadow norm as a function of ϵ for the average error rates of $\bar{q} = 0.04, 0.08$ and 0.12 . The classical shadows are obtained with the octahedron POVM. For small average error \bar{q} , the shadow norm is only weakly dependent on ϵ . We note that readout noise has the same effect as depolarization noise if $\epsilon = 0$. In this sense, Fig. 6.2 (b) confirms that a small amount of noise has limited effect on the κ_E^2 and thus on the quality of shadow tomography.

6.5 Inferring properties of the Ising model

We now investigate the Ising model with the help of shadow tomography. In particular, we are interested in the deviations from the exact results and the convergence of the estimates.

The Hamiltonian of the transverse Ising model is [186, 187]

$$H = -J \sum_{\langle ij \rangle} Z_i Z_j - h \sum_i X_i \quad (6.31)$$

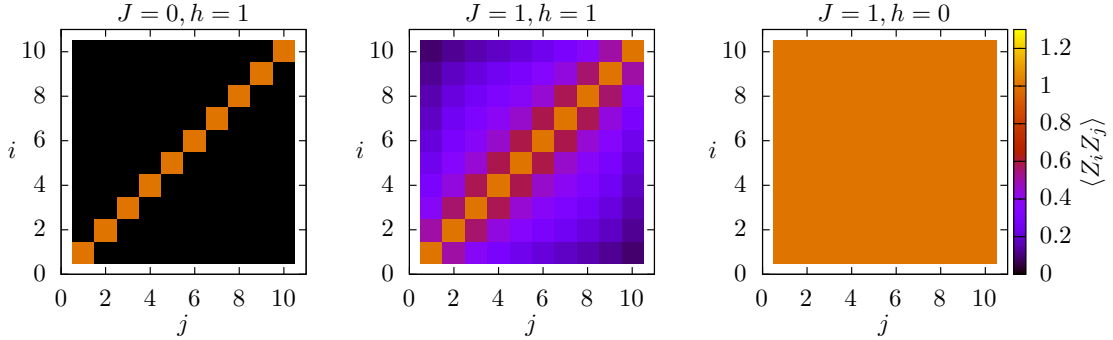
and describes distinguishable spin- $\frac{1}{2}$ particles that are distributed on a lattice. Z_i denotes the Pauli Z matrix for spin i and $\langle ij \rangle$ are the pairs of adjacent spins. The interaction is thus determined by the spin in z direction and the strength of the interaction is described by J . In addition the spins are subjected to a magnetic field of strength h that couples to the spins by the Pauli X matrix. To get a first insight into the properties of the model, we will discuss the phases the ground state |GS) of the model exhibits. For this purpose we will look at the limiting cases.

In the limit $J > 0$ and $h = 0$, the Hamiltonian only consists of the first term, i.e., $H = -J \sum_{\langle ij \rangle} Z_i Z_j$. In this case the eigenstates are common eigenstates of Z_i for all i . The energy is minimized in case all spin are oriented in the same direction. Based on this, the ground state is degenerate and the subspace of lowest energy is spanned by the two states

$$|00 \dots 0\rangle \quad \text{and} \quad |11 \dots 1\rangle. \quad (6.32)$$

This phase is known as the ordered phase [187].

(a) Exact diagonalization



(b) Shadow tomography

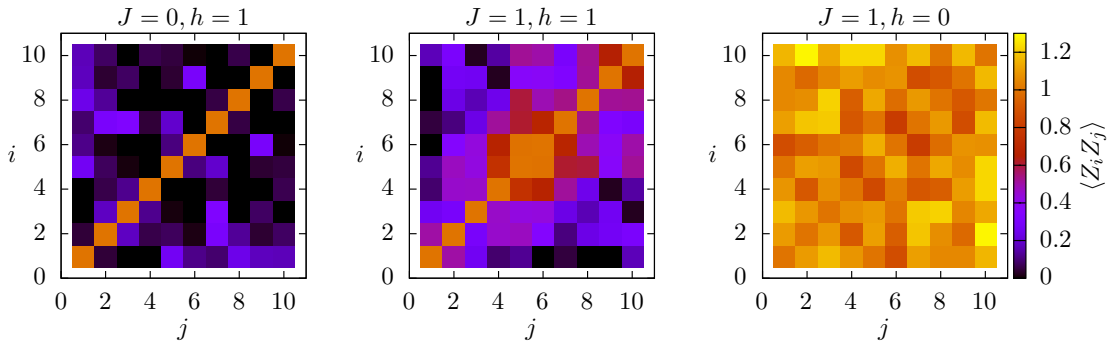


Figure 6.3: Correlation matrices $\langle Z_i Z_j \rangle$ for the transverse Ising model of $n = 10$ spin- $\frac{1}{2}$ particles. (a) shows the exact results that are obtained by diagonalization, whereas (b) depicts the estimates from shadow tomography. The estimates in (b) have been calculated from 1000 classical shadows that have been constructed by the tetrahedron **POVM**.

In contrast, in the limit $J = 0$ and $h > 0$ the Hamiltonian consists only of the second term $H = -h \sum_i X_i$ such that the eigenstates are common eigenstates of all X_i . The ground state is given by

$$|+\dots+\rangle, \quad (6.33)$$

where $|+\rangle$ is the eigenstate of Pauli X for eigenvalue $+1$.

We now investigate the correlations $\langle Z_i Z_j \rangle$ with the help of shadow tomography. For this purpose we consider the Ising model of $n = 10$ spins that are distributed on a linear chain. We generate 1000 classical shadows with the tetrahedron **POVM**. The tetrahedron **POVM** is defined for a single qubit and the composite measurement can, accordingly, be constructed by the tensor product (Sec. 6.2.4). The results are shown in Fig. 6.3. Panel (a) in Fig. 6.3 shows the results that are obtained by exact diagonalization, whereas in (b) the estimates from shadow tomography are shown. We can observe that shadow tomography predicts the features of the correlation matrix correctly. The data is, however, subjected to statistical fluctuations. For 1000 classical shadows, deviations from the exact value of roughly 30% can be observed.

In the next step, we therefore study the convergence of shadow tomography. For this purpose, we consider an Ising chain of three qubits with the parameters $J = 1$ and $h = 1$. As observables we choose either the set of Pauli observables $\{\sigma_x^{(i)} | i = 1, 2, 3\}$, $\{\sigma_y^{(i)} | i = 1, 2, 3\}$ or

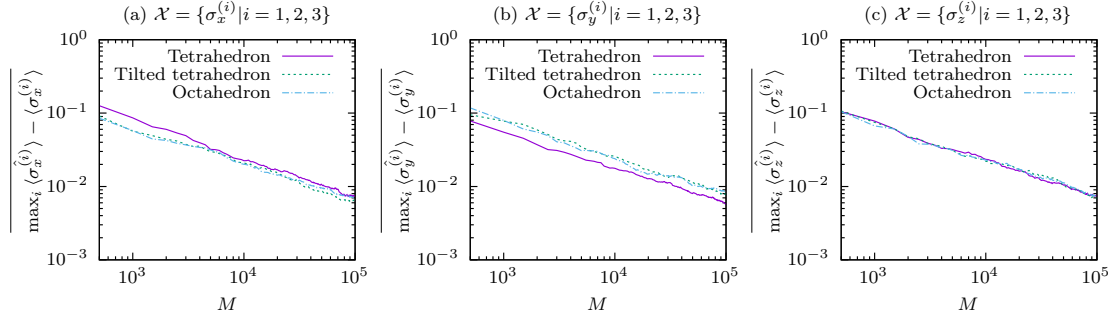


Figure 6.4: Maximal error as a function of the number of classical shadows M . For the tetrahedron, tilted tetrahedron and the octahedron **POVM**, we have estimated the observables (a) $\{\sigma_x^{(i)} | i = 1, 2, 3\}$, (b) $\{\sigma_y^{(i)} | i = 1, 2, 3\}$ and (c) $\{\sigma_z^{(i)} | i = 1, 2, 3\}$. The maximal error is averaged over 50 repetitions. For the calculation, the parameters $J = 1$ and $h = 1$ have been used.

$\{\sigma_z^{(i)} | i = 1, 2, 3\}$. The expectation values are estimated from M classical shadows, which are constructed with either the tetrahedron, the tiled tetrahedron or the octahedron **POVM**. The tiled tetrahedron measurement corresponds to the violet tetrahedron in Fig. 6.5 (b). In Fig. 6.4, we plot the maximal deviation in the set of observables. The maximal deviation is averaged over 50 repetitions. Fig. 6.4 reveals that in all simulations the maximal error decreases polynomially with the number of classical shadows M . Fig. 6.4 (a) shows the results for the observables $\{\sigma_x^{(i)} | i = 1, 2, 3\}$. The results show that the tetrahedron **POVM** yields slightly larger errors than the tilted tetrahedron and the octahedron measurement. The reverse result is obtained for the observables $\{\sigma_y^{(i)} | i = 1, 2, 3\}$. Fig. 6.4 (b) shows that the tetrahedron **POVM** is least affected by error. For the observables $\{\sigma_z^{(i)} | i = 1, 2, 3\}$ in turn Fig. 6.4 (c) shows no difference between the different **POVMs**. It is interesting to note that for the chosen observables, the tilted tetrahedron **POVM** yields the same magnitude of errors as the octahedron measurement. We see that it depends on the **POVM** which observables can be estimated well from the classical shadows.

6.6 Outlook

To conclude this section we discuss a direct application of the formulation of shadow tomography in terms of **POVMs**: In case the observables that should be estimated are known, the formulation allows the optimization of the **POVM**. Finally, we end this section by discussing the further prospects.

6.6.1 Optimization of **POVMs** for shadow tomography

In this section we discuss the optimization of shadow tomography that has been presented in Ref. [P1]. We assume that the set of observables \mathcal{X} that should be estimated from the classical shadows, are known in advance. The goal is to find the **POVM** E^* that minimizes the shadow norm:

$$E^* = \arg \min_E \kappa_E^2(\mathcal{X}). \quad (6.34)$$

We have noted in Sec. 6.2.2 that the optimization in case of randomized measurements is cumbersome. The set of generalized measurements is convex. This simplifies the optimization. We mention, however, that the shadow norm is not convex in the **POVM** E . We thus have performed the optimization with the help of simulated annealing [P1].

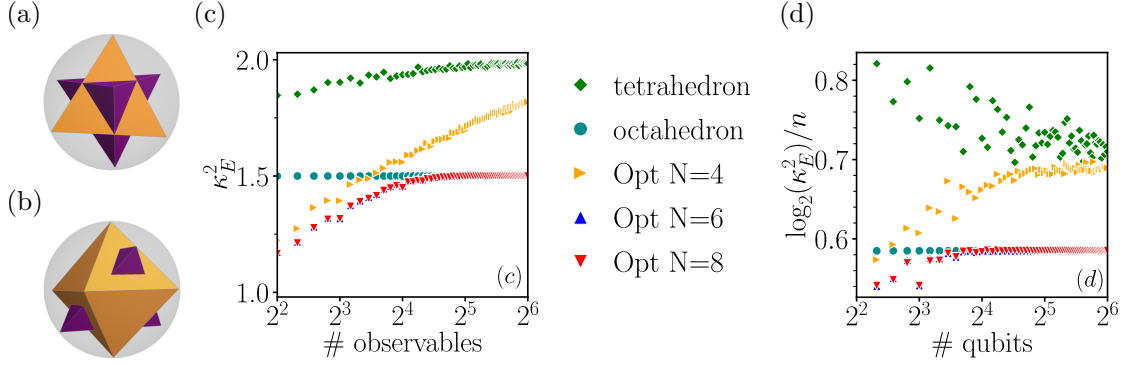


Figure 6.5: Targeted observables and optimal generalized measurements. (a) For the observables that correspond to the four projections defined by the orange tetrahedron, the measurement corresponding to the invert tetrahedron **POVM** (violet) is optimal. (b) For observables corresponding to eigenprojections of the Pauli observables σ_x , σ_y and σ_z (orange octahedron), the violet tetrahedron **POVM** is optimal. (c) Optimal shadow norms given by the optimizer (labeled **Opt** with the number of measurement outcomes) as a function of the number of single-qubit projection observables randomly distributed according to the Haar measure. (d) Similarly, optimal shadow norms given by the optimizer as a function of the number of qubits. The observables are tensor products of single-qubit projections distributed according to the Haar measure. In (c) and (d), the shadow norm for the tetrahedron and octahedron **POVM** are also shown. The figure is taken from [P1].

As a first example, we investigate the case of a single qubit. Specifically, we consider the following observables \mathcal{X} : (a) the 4 projections that correspond to the orange tetrahedron in Fig. 6.5 (a), (b) the projections onto the eigenstates of the Pauli observables and (c) random projections that are sampled according to the Haar measure.

(a): For the 4 projections that correspond to the orange tetrahedron in Fig. 6.5 (a), we first try classical shadows that have been constructed with the tetrahedron **POVM**. In this case the squared shadow norm κ_E^2 is 2. The octahedron measurement yields an improved shadow norm of $3/2$. The optimization, however, results in the tetrahedron **POVM** that is plotted in violet in Fig. 6.5 (a). Compared to the orange tetrahedron, the violet one is centrally inverted and yields a squared shadow norm of $\kappa_E^2 = 1$.

(b) Next, we consider the projections onto the eigenstates of the Pauli observables. The projections correspond to the orange octahedron in Fig. 6.5 (b). The classical shadows obtained with the octahedron measurement itself give $\kappa_E^2 = 3/2$. Interestingly, the optimization yields $\kappa_E^2 = 3/2$ also with the tetrahedron **POVM** of 4 outcomes that is shown in violet in Fig. 6.5 (b).

(c) Finally, we consider random projections that are sampled according to the Haar measure on the Bloch sphere. In Fig. 6.5 (c), we show the shadow norm that has been found with the optimizer as a function of the number of random observables. For small number of observables ($|\mathcal{X}| \lesssim 15$), there is always a **POVM** with $N = 4, 6$ and 8 outcomes that yields a smaller shadow norm than the standard tetrahedron ($N = 4$) or the octahedron measurements ($N = 6$). In case the number of outcomes is $N = 6$ or 8 , the squared shadow norm converges to $\kappa_E^2 = 3/2$ for an increasing number of observables. We point out that this is the value of the octahedron measurement. This highlights that the octahedron **POVM** is special. For the octahedron **POVM** the squared shadow norm of any projection is $\kappa_E^2 = 3/2$. We show in Ref. [P1] that if all observables are projections onto pure states and the effects of the **POVM** have equal trace, the

octahedron measurement is optimal.

Next, we will consider how the **POVM** can be optimized for many-body systems. The number of parameters that are necessary to describe a **POVM** increase exponentially with the number of qubits. To simplify and make the optimization feasible, we assume that **POVM** that factorizes as a tensor product over the qubits as discussed in Sec. 6.2.4. Moreover, we assume that, $E^{(1)} = E^{(2)} = \dots = E^{(n)}$. This assumption is reasonable in case there is no preference among the qubits. Under these assumptions, the complexity of the optimization only increases linear in the number of qubits and observables. For an explicit calculation, we consider a system of up to $n = 64$ qubits. Moreover, we choose $|\mathcal{X}| = n$ product observables, i.e., the observables can also be written as a tensor product of single qubit observables.

The single qubit observables in turn are again projections onto pure states that are sampled at random according to the Haar measure. Our optimization confirms this expectation that the optimal product measurement are similar to the ones discussed for a single qubit. Fig. 6.5(d) shows that for small number of qubits ($n \lesssim 10$), the optimal **POVM** with $N = 6$ and $N = 8$ effects yields significantly lower shadow norms compared to the tetrahedron or octahedron measurements. In addition, the optimal shadow norm converges to the one given by the octahedron measurement with increasing number of qubits. This again shows that the octahedron **POVM** is special.

6.6.2 Discussion

We end this section by a summary and outlook. Whereas shadow tomography was originally formulated in terms of randomized measurements [26], we have shown that shadow tomography can also be based on **POVMs**. In doing so, we have derived the classical shadows as the least squares estimator. Moreover, we have discussed the class of rigidly symmetric **POVMs** for which analytical expressions for the classical shadows can be derived. We have further demonstrated how the observables can be estimated without reconstructing the full density matrix. This constitutes the main advantage of shadow tomography compared to full quantum state tomography and requires an efficient representation of the classical shadows. For this purpose, we show that it is convenient to consider **POVMs** that are composed as a tensor product of single qubit measurements. We further demonstrate the advantage of the formulation in terms of **POVMs** by showing that noise can be readily taken into account and that the **POVM** can be optimized in case the observables to estimate are known.

The formulation in terms of **POVMs** also opens various perspectives for future research. For example can a more realistic analysis of noise be performed for experiments based on shadow tomography [188, 189]. Moreover, it would be interesting to extend the method the channel tomography or to incorporate the technique of derandomization [172]. We have seen that the octahedron **POVM** takes a special role in qubit systems. It could be investigated if there are also **POVMs** that are optimal for a large class of observables in higher dimensions.

Finally, it would also be interesting to work out whether quantum state tomography benefits from the shadow tomography and vice versa. In particular, statistical methods that have been developed in the context of full quantum state tomography [190] could improve the statistical analysis of shadow tomography.

Conclusion and outlook

In this thesis, we have considered how properties of quantum systems can be obtained from random measurements. For this purpose, we have used random measurements in various forms.

In Sec. 3 and 4, we have discussed how random measurements can help to extract certain properties of quantum states. We have discussed in Sec. 3 that spin-squeezing inequalities can be evaluated by random pair correlations. In this case randomness is used in the choice of the qubit pairs that are sampled. An important result of this section is also the statistical analysis. As the spin-squeezing inequalities are nonlinear in the quantum state, also the estimates are nonlinear in the data. We thus have proposed a statistical analysis that can be applied to nonlinear quantities. The statistical analysis is, however, not limited to spin-squeezing inequalities. It can also be used for other nonlinear properties. For example, the method would also be of interest for the statistical analysis of the purity or of entropies. Finally, we point out that random sampling of the qubits could also be used in other contexts. This is especially promising for averaged quantities like average two-qubit correlations that cannot be obtained from collective measurements.

We have used a similar approach in Sec. 4. Multipartite Bell inequalities often exhibit a number of measurement settings that increases exponentially with the number of parties. It thus quickly becomes infeasible to implement all measurement settings. As a solution we propose to sample the measurement settings at random. With a similar statistical analysis as in Sec. 3, we have shown that by random sampling fewer measurements are necessary. To demonstrate the method, we have performed a simulation for an IBM quantum computer. We have furthermore pointed out that a Bell violation can be used to benchmark quantum computers. There are various directions for further research. Bell violations on a quantum computer have to be taken with a grain of salt, as the qubits are not space-like separated. Communication between the qubits thus cannot be ruled out and therefore quantum computers exhibit the locality loophole. But to close the locality loophole, also the random sampling has to be adapted. In the proposed scheme the measurement settings are chosen globally. To guarantee the independence of the measurement settings, the settings have to be chosen locally. Multipartite Bell inequalities, however, typically do not contain all combinations of local measurement settings. In fact, the incorrect settings can simply be discarded in a postprocessing step, but this comes at the expense of an additional sampling overhead that has to be investigated. Maybe it is also possible to sample the measurement settings more efficiently. It might further be interesting to analyze the statistics of the Bell inequalities with the relative entropy [64] and explore the connection of nonlocality to other benchmarks, as the quantum volume [132, 133] or the layer fidelity [134].

In Sec. 5, we have used measurements in random bases, which can be described by applying random unitaries before the measurement. The unitaries are sampled according to the Haar measure, which makes the sampling invariant under LU transformations. We have shown that all LU invariants of a quantum system can be inferred from randomized measurements. We further have given explicit expressions for the LU invariants of a two-qubit system and implement the method with photons. As applications, we have derived from the obtained invariants the maximal violation of a Bell inequality that could be observed for the state. In addition, we have discussed whether the state could be used for quantum teleportation. As a direct generalization, it would be interesting to express also LU invariants of higher dimensional systems in terms of the moments of randomized measurements. Another research direction is to investigate whether Bell nonlocality in multipartite systems can be efficiently detected with randomized measurements. In Sec. 4, we have proposed a method to evaluate Bell inequalities by sampling the terms at random. The problem of sampling the measurement settings locally could potentially be solved by randomized measurements.

Finally, in Sec. 6 we have formulated shadow tomography in terms of **POVMs**. This has led to new insights into the method of shadow tomography. On the one hand, we were able to better characterize the measurements for which the measurement channel can be analytically inverted. To be precise, we have derived the classical shadows for rigidly symmetric **POVMs**. The main idea of shadow tomography is to find a method that allows estimating arbitrary observables from the recorded data and scales well to high-dimensional systems. We have pointed out that, for this purpose, it is constructive to consider measurements that are composed as the tensor product of single-qubit **POVMs**. We, moreover, have shown that the new formulation allows for a natural adaptation to noise and that an optimization of the **POVM** can be performed. Ultimately, the formulation in terms of **POVMs** also puts the implementation in a new light. Whereas the original scheme in [26] uses random unitaries, our formulation allows for an implementation without randomization. Rather than by randomization, **POVMs** can also be implemented by an ancilla system. This also paves the way for future research. Even though the defining property of shadow tomography is to avoid the reconstruction of the density matrix, the formulation in terms of **POVMs** shows striking similarity to ordinary quantum state tomography. It thus might be interesting to investigate whether the statistical tools that have been developed in the context of quantum state tomography can be used for shadow tomography [190]. Moreover, it would be interesting to consider shadow tomography in the presence of more realistic noise and to generalize the formalism to channel tomography.

List of abbreviations

- CCNR** computable cross norm or realignment. 16
- DSF** dispersion shifted fiber. 65, 66
- EPS** entangled photon source. 65, 66, 70
- GHZ** Greenberger–Horne–Zeilinger. 12, 23, 50, 51, 54–61
- GME** genuine multipartite entanglement. 11, 17, 18
- iid** independent and identically distributed. 32, 33
- ITU** International Telecommunication Union. 65
- LC** linear cluster. 12, 23, 24, 50, 51, 54–61
- LHV** local hidden variable. 8, 22–24, 52
- LU** local unitary. 9, 12, 24–26, 63, 64, 67–70, 72, 86
- NMR** nuclear magnetic resonance. 50
- PBS** polarizing beam splitter. 65
- POVM** positive operator valued measure. 9, 14–16, 30, 62, 72–85, 87
- PPT** positive partial transpose. 16

A Additional calculations for Sec. 3

A.1 Unbiased estimators

In this appendix we will prove that the estimators in Sec. 3.3 are unbiased. For the total spin estimator, $\langle J_\alpha^2 \rangle_{\text{TS}}$ is the sample mean and $(\widehat{\Delta J_\alpha})^2_{\text{TS}}$ is the sample variance. Hence, the estimators are known to be unbiased [83]. We will therefore focus on the estimators that rely on pair correlations.

A.1.1 Estimator based on pair correlations

As the term $s_\alpha^{(P_1,k)} s_\alpha^{(P_2,k)}$ is the product of the spins of the qubit pair $P = (P_1, P_2)$ in the same measurement repetition k , we have for $i \in \{1, 2\}$

$$\mathbb{E} \left[s_\alpha^{(P_i,k)} \right] = \frac{1}{2} \langle \sigma_\alpha^{(P_i)} \rangle \quad \text{and} \quad \mathbb{E} \left[s_\alpha^{(P_1,k)} s_\alpha^{(P_2,k)} \right] = \frac{1}{4} \langle \sigma_\alpha^{(P_1)} \sigma_\alpha^{(P_2)} \rangle. \quad (\text{A.1})$$

As a result, we obtain

$$\mathbb{E} \left[\widehat{\langle J_\alpha^2 \rangle}_{\text{AP}} \right] = \frac{N}{4} + \sum_P \frac{1}{K_{\text{AP}}} \sum_{k=1}^{K_{\text{AP}}} \mathbb{E} \left[s_\alpha^{(P_1,k)} s_\alpha^{(P_2,k)} \right] = \frac{N}{4} + \frac{1}{4} \sum_P \langle \sigma_\alpha^{(P_1)} \sigma_\alpha^{(P_2)} \rangle = \langle J_\alpha^2 \rangle, \quad (\text{A.2})$$

where we have used that

$$\frac{1}{4} \sum_P \langle \sigma_\alpha^{(P_1)} \sigma_\alpha^{(P_2)} \rangle = \frac{1}{4} \sum_{P_1 \neq P_2} \langle \sigma_\alpha^{(P_1)} \sigma_\alpha^{(P_2)} \rangle = \frac{1}{4} \sum_{P_1, P_2} \langle \sigma_\alpha^{(P_1)} \sigma_\alpha^{(P_2)} \rangle - \frac{1}{4} \sum_{P_1 = P_2} \langle (\sigma_\alpha^{(P_1)})^2 \rangle = \langle J_\alpha^2 \rangle - \frac{N}{4}. \quad (\text{A.3})$$

Next, we will show that the estimator for the variance in Eq. (3.10) is unbiased.

$$\begin{aligned} \mathbb{E} \left[(\widehat{\Delta J_\alpha})^2 \right] &= \frac{N}{4} + \frac{1}{K_{\text{AP}}} \sum_P \sum_k \mathbb{E} \left[s_\alpha^{(P_1,k)} s_\alpha^{(P_2,k)} \right] \\ &\quad - \frac{1}{K_{\text{AP}}(K_{\text{AP}} - 1)(N - 1)^2} \sum_{P, Q} \sum_{k \neq l} \mathbb{E} \left[s_\alpha^{(P_1,k)} s_\alpha^{(Q_2,l)} \right] \\ &= \frac{N}{4} + \frac{1}{K_{\text{AP}}} \sum_P \sum_k \frac{1}{4} \langle \sigma_\alpha^{(P_1)} \sigma_\alpha^{(P_2)} \rangle \\ &\quad - \frac{1}{K_{\text{AP}}(K_{\text{AP}} - 1)(N - 1)^2} \sum_{P, Q} \sum_{k \neq l} \mathbb{E} \left[s_\alpha^{(P_1,k)} \right] \mathbb{E} \left[s_\alpha^{(Q_2,l)} \right] \\ &= \frac{N}{4} + \sum_P \frac{1}{4} \langle \sigma_\alpha^{(P_1)} \sigma_\alpha^{(P_2)} \rangle - \frac{1}{K_{\text{AP}}(K_{\text{AP}} - 1)(N - 1)^2} \sum_{P, Q} \sum_{k \neq l} \frac{1}{2} \langle \sigma_\alpha^{(P_1)} \rangle \frac{1}{2} \langle \sigma_\alpha^{(Q_2)} \rangle \\ &= \langle J_\alpha^2 \rangle - \frac{1}{(N - 1)^2} \left(\sum_P \frac{1}{2} \langle \sigma_\alpha^{(P_1)} \rangle \right) \left(\sum_Q \frac{1}{2} \langle \sigma_\alpha^{(Q_2)} \rangle \right) = \langle J_\alpha^2 \rangle - \langle J_\alpha \rangle^2 = (\Delta J_\alpha)^2 \end{aligned} \quad (\text{A.4})$$

In the above calculation, we have used that

$$\sum_P \frac{1}{2} \langle \sigma_\alpha^{(P_1)} \rangle = \sum_{P_1 \neq P_2} \frac{1}{2} \langle \sigma_\alpha^{(P_1)} \rangle = (N - 1) \left\langle \sum_{P_1} \frac{1}{2} \sigma_\alpha^{(P_1)} \right\rangle = (N - 1) \langle J_\alpha \rangle. \quad (\text{A.5})$$

To show that the estimator of $\langle J_\alpha \rangle^2$ is unbiased, we use that the qubits of each pair (i, j) are measured in different experimental runs. Hence, the outcomes are statistically independent: $\mathbb{E}[s_\alpha^{(i,2k)} s_\alpha^{(j,2k-1)}] = \mathbb{E}[s_\alpha^{(i,2k)}] \mathbb{E}[s_\alpha^{(j,2k-1)}] = \frac{1}{4} \langle \sigma_\alpha^{(i)} \rangle \langle \sigma_\alpha^{(j)} \rangle$. Thus, the estimator is unbiased:

$$\begin{aligned} \mathbb{E} \left[\widehat{\langle J_\alpha \rangle}_{\text{AP}}^2 \right] &= \sum_{i,j=1}^N \frac{1}{\frac{K_{\text{AP}}}{2}} \sum_{k=1}^{\frac{K_{\text{AP}}}{2}} \mathbb{E}[s_\alpha^{(i,2k)} s_\alpha^{(j,2k-1)}] = \sum_{i,j=1}^N \frac{1}{4} \langle \sigma_\alpha^{(i)} \rangle \langle \sigma_\alpha^{(j)} \rangle = \left\langle \frac{1}{2} \sum_{i=1}^N \sigma_\alpha^{(i)} \right\rangle \left\langle \frac{1}{2} \sum_{j=1}^N \sigma_\alpha^{(j)} \right\rangle \\ &= \langle J_\alpha \rangle^2. \end{aligned} \tag{A.6}$$

A.1.2 Estimator based on random pair correlations

We now want to calculate the expectation values of the estimators that make use of random pair correlations. For this, we have to first specify how the expectation value has to be understood. As both the indices of the qubits and the outcomes are random variables, we can calculate the expectation value by the law of iterated expectations. In case \mathcal{I} is a random variable that takes the values $\mathcal{I} = 1, \dots, N$ with uniform probability, the expectation value of the spin of qubit \mathcal{I} evaluates to

$$\begin{aligned} \mathbb{E} \left[s_\alpha^{(\mathcal{I},k)} \right] &= \mathbb{E}_{\mathcal{I}} \left[\mathbb{E}_s \left(s_\alpha^{(\mathcal{I},k)} \mid \mathcal{I} \right) \right] = \mathbb{E}_{\mathcal{I}} \left[\frac{1}{2} \langle \sigma_\alpha^{(\mathcal{I})} \rangle \right] = \sum_{i=1}^N \mathbb{P}(\mathcal{I} = i) \frac{1}{2} \langle \sigma_\alpha^{(i)} \rangle = \frac{1}{N} \sum_{i=1}^N \frac{1}{2} \langle \sigma_\alpha^{(i)} \rangle \\ &= \frac{1}{N} \langle J_\alpha \rangle. \end{aligned} \tag{A.7}$$

In the above equation, we used the uniform probability distribution $\mathbb{P}(\mathcal{I} = i) = \frac{1}{N}$. We note that this is indeed the marginal distribution in case we sample uniformly all distinct pairs with the probability distribution in Eq. (3.15), i.e., $\mathbb{P}(\mathcal{I} = i) = \sum_{j=1}^N \mathbb{P}(\mathcal{I} = i, \mathcal{J} = j) = \sum_{j \neq i}^N \frac{1}{N(N-1)} = \frac{1}{N} = \mathbb{P}(\mathcal{J} = j)$. Similarly, we can show that

$$\begin{aligned} \mathbb{E} \left[s_\alpha^{(\mathcal{I},k)} s_\alpha^{(\mathcal{J},k)} \right] &= \mathbb{E}_{\mathcal{I}, \mathcal{J}} \left[\mathbb{E}_s \left(s_\alpha^{(\mathcal{I},k)} s_\alpha^{(\mathcal{J},k)} \mid \mathcal{I}, \mathcal{J} \right) \right] = \mathbb{E}_{\mathcal{I}, \mathcal{J}} \left[\frac{1}{4} \langle \sigma_\alpha^{(\mathcal{I})} \sigma_\alpha^{(\mathcal{J})} \rangle \right] \\ &= \sum_{i \neq j}^N \mathbb{P}(\mathcal{I} = i, \mathcal{J} = j) \frac{1}{4} \langle \sigma_\alpha^{(i)} \sigma_\alpha^{(j)} \rangle \\ &= \sum_{i \neq j}^N \frac{1}{N(N-1)} \frac{1}{4} \langle \sigma_\alpha^{(i)} \sigma_\alpha^{(j)} \rangle \\ &= \frac{1}{N(N-1)} \left(\langle J_\alpha^2 \rangle - \frac{N}{4} \right), \end{aligned} \tag{A.8}$$

where we have used in the third step that only distinct pairs are considered, i.e., $\mathbb{P}(\mathcal{I} = i, \mathcal{J} = i) = 0$. We can thus evaluate the expectation value:

$$\begin{aligned} \mathbb{E} \left[\widehat{\langle J_\alpha^2 \rangle}_{\text{RP}} \right] &= \frac{N}{4} + \frac{N(N-1)}{K_{\text{RP}} L_{\text{RP}}} \sum_{l=1}^{L_{\text{RP}}} \sum_{k=1}^{K_{\text{RP}}} \mathbb{E} \left[s_\alpha^{(\mathcal{I},k)} s_\alpha^{(\mathcal{J},k)} \right] \\ &= \frac{N}{4} + \frac{N(N-1)}{K_{\text{RP}} L_{\text{RP}}} \sum_{l=1}^{L_{\text{RP}}} \sum_{k=1}^{K_{\text{RP}}} \frac{1}{N(N-1)} \left(\langle J_\alpha^2 \rangle - \frac{N}{4} \right) = \langle J_\alpha^2 \rangle. \end{aligned} \tag{A.9}$$

To calculate the expectation value of the estimator $(\widehat{\Delta J_\alpha})_{\text{RP}}^2$, we use Eq. (A.8) and apply that the product in the last term of the estimator only involves independent random variables:

$$\begin{aligned}
& \mathbb{E} \left[(\widehat{\Delta J_\alpha})_{\text{RP}}^2 \right] \\
&= \frac{N}{4} + \frac{N(N-1)}{L_{\text{RP}} K_{\text{RP}}} \sum_{l=1}^{L_{\text{RP}}} \sum_{k=1}^{K_{\text{RP}}} \mathbb{E} \left[s_\alpha^{(\mathcal{I}_l, k)} s_\alpha^{(\mathcal{J}_l, k)} \right] - \frac{N^2}{L_{\text{RP}}(L_{\text{RP}}-1)K_{\text{RP}}^2} \sum_{l \neq m}^{L_{\text{RP}}} \sum_{k, q=1}^{K_{\text{RP}}} \mathbb{E} \left[s_\alpha^{(\mathcal{I}_l, k)} \right] \mathbb{E} \left[s_\alpha^{(\mathcal{J}_m, q)} \right] \\
&= \frac{N}{4} + \frac{N(N-1)}{L_{\text{RP}} K_{\text{RP}}} \sum_{l=1}^{L_{\text{RP}}} \sum_{k=1}^{K_{\text{RP}}} \frac{1}{N(N-1)} \left(\langle J_\alpha^2 \rangle - \frac{N}{4} \right) - \frac{N^2}{L_{\text{RP}}(L_{\text{RP}}-1)K_{\text{RP}}^2} \sum_{l \neq m}^{L_{\text{RP}}} \sum_{k, q=1}^{K_{\text{RP}}} \frac{1}{N^2} \langle J_\alpha \rangle^2 \\
&= \langle J_\alpha^2 \rangle - \langle J_\alpha \rangle^2 = (\Delta J_\alpha)^2.
\end{aligned} \tag{A.10}$$

Similarly, we can show for the estimator of $\langle J_\alpha \rangle^2$ for scheme RP2 that

$$\mathbb{E} \left[(\widehat{J_\alpha})_{\text{RP}}^2 \right] = \frac{2N^2}{K_{\text{RP}} L_{\text{RP}}} \sum_{l=1}^{L_{\text{RP}}} \sum_{k=1}^{\frac{K_{\text{RP}}}{2}} \mathbb{E} [s_\alpha^{(\mathcal{I}_l, 2k)}] \mathbb{E} [s_\alpha^{(\mathcal{J}_l, 2k-1)}] = \frac{2N^2}{K_{\text{RP}} L_{\text{RP}}} \sum_{l=1}^{L_{\text{RP}}} \sum_{k=1}^{\frac{K_{\text{RP}}}{2}} \frac{1}{N^2} \langle J_\alpha \rangle^2 = \langle J_\alpha \rangle^2. \tag{A.11}$$

Here, we used again that the outcomes of each pair (i, j) are obtained from different experimental runs and are thus independent. Moreover, all pairs are considered with uniform probability, such that also the random variables \mathcal{I}, \mathcal{J} are independent.

A.2 Derivation of the variances

A.2.1 Estimator based on the total spin

The variances of the estimators can be derived from the variances of the parts. Thus, we are going to derive the variances $\text{Var}(\langle J_\alpha^2 \rangle_{\text{TS}}) = \mathbb{E}[(\widehat{J_\alpha^2})_{\text{TS}}^2] - (\mathbb{E}[\widehat{J_\alpha^2}])^2$ and $\text{Var}((\widehat{\Delta J_\alpha})_{\text{TS}}^2) = \mathbb{E}[(\widehat{(\Delta J_\alpha)^2})_{\text{TS}}^2] - \mathbb{E}[(\widehat{\Delta J_\alpha})_{\text{TS}}^2]^2$.

For a clear arrangement of the calculation, we will apply a graphical representation. We use the first expression to explain the representation.

$$\begin{aligned}
\mathbb{E} \left[(\widehat{J_\alpha^2})_{\text{TS}}^2 \right] &= \mathbb{E} \left[\left(\frac{\sum_{k=1}^{K_{\text{TS}}} (m_\alpha^{(k)})^2}{K_{\text{TS}}} \right)^2 \right] = \frac{1}{K_{\text{TS}}^2} \sum_{k_1, k_2=1}^{K_{\text{TS}}} \mathbb{E} \left[(m_\alpha^{(k_1)})^2 (m_\alpha^{(k_2)})^2 \right] \\
&= \frac{1}{K_{\text{TS}}^2} \sum_{k_1 \neq k_2=1}^{K_{\text{TS}}} \mathbb{E} \left[(m_\alpha^{(k_1)})^2 (m_\alpha^{(k_2)})^2 \right] + \frac{1}{K_{\text{TS}}^2} \sum_{k_1=k_2=1}^{K_{\text{TS}}} \mathbb{E} \left[(m_\alpha^{(k_1)})^4 \right] \\
&\quad \begin{array}{ccc} k_1 & k_2 & \\ \bullet & \bullet & \\ & & + \end{array} \quad \begin{array}{ccc} k_1 & k_2 & \\ \bullet & \bullet & \\ & & \underbrace{\hspace{1cm}} \end{array} \\
&= \frac{K_{\text{TS}} - 1}{K_{\text{TS}}} \langle J_\alpha^2 \rangle^2 + \frac{1}{K_{\text{TS}}} \langle J_\alpha^4 \rangle
\end{aligned} \tag{A.12}$$

In the above derivation, we consider the terms with $k_1 \neq k_2$ separately, as the outcomes are obtained in different experimental runs and thus $\mathbb{E}[(m_\alpha^{(k_1)})^2 (m_\alpha^{(k_2)})^2] = \mathbb{E}[(m_\alpha^{(k_1)})^2] \mathbb{E}[(m_\alpha^{(k_2)})^2] = \langle J_\alpha^2 \rangle^2$. For terms with more indices it becomes tedious to write down all possibilities for the

indices to coincide. Therefore, we will use the graphical representation instead. Two indices are connected if and only if they coincide. With Eq. (A.12) we can evaluate the variance

$$\begin{aligned} \text{Var} \left(\widehat{\langle J_\alpha^2 \rangle}_{\text{TS}} \right) &= \mathbb{E} \left[\left(\widehat{\langle J_\alpha^2 \rangle}_{\text{TS}} \right)^2 \right] - \left(\mathbb{E} \left[\widehat{\langle J_\alpha^2 \rangle}_{\text{TS}} \right] \right)^2 = \frac{K_{\text{TS}} - 1}{K_{\text{TS}}} \langle J_\alpha^2 \rangle^2 + \frac{1}{K_{\text{TS}}} \langle J_\alpha^4 \rangle - \langle J_\alpha^2 \rangle^2 \\ &= \frac{1}{K_{\text{TS}}} \left(\langle J_\alpha^4 \rangle - \langle J_\alpha^2 \rangle^2 \right) = \frac{1}{K_{\text{TS}}} \left(\Delta J_\alpha^2 \right)^2. \end{aligned} \quad (\text{A.13})$$

Next, we will derive the variance of $\left(\widehat{\Delta J_\alpha} \right)_{\text{TS}}^2$.

$$\begin{aligned} \mathbb{E} \left\{ \left[\left(\widehat{\Delta J_\alpha} \right)_{\text{TS}}^2 \right]^2 \right\} &= \mathbb{E} \left\{ \left[\frac{1}{K_{\text{TS}} - 1} \sum_{k_1=1}^{K_{\text{TS}}} \left(m_\alpha^{(k_1)} - \sum_{k_2=1}^K \frac{m_\alpha^{(k_2)}}{K_{\text{TS}}} \right)^2 \right]^2 \right\} \\ &= \frac{1}{(K_{\text{TS}} - 1)^2} \left\{ \sum_{k_1, k_2=1}^{K_{\text{TS}}} \mathbb{E} \left[\left(m_\alpha^{(k_1)} \right)^2 \left(m_\alpha^{(k_2)} \right)^2 \right] - \frac{2}{K_{\text{TS}}} \sum_{k_1, k_2, k_3=1}^{K_{\text{TS}}} \mathbb{E} \left[\left(m_\alpha^{(k_1)} \right)^2 m_\alpha^{(k_2)} m_\alpha^{(k_3)} \right] \right. \\ &\quad \left. + \frac{1}{K_{\text{TS}}^2} \sum_{k_1, k_2, k_3, k_4=1}^{K_{\text{TS}}} \mathbb{E} \left[m_\alpha^{(k_1)} m_\alpha^{(k_2)} m_\alpha^{(k_3)} m_\alpha^{(k_4)} \right] \right\}. \end{aligned} \quad (\text{A.14})$$

To evaluate the expression above, we apply the graphical representation to the sums. We obtain

$$\sum_{k_1, k_2=1}^{K_{\text{TS}}} \mathbb{E} \left[\left(m_\alpha^{(k_1)} \right)^2 \left(m_\alpha^{(k_2)} \right)^2 \right] = \begin{array}{c} k_1 \quad k_2 \\ \bullet \quad \bullet \end{array} + \begin{array}{c} k_1 \quad k_2 \\ \bullet \quad \bullet \\ \text{---} \end{array} = K_{\text{TS}}(K_{\text{TS}} - 1) \langle J_\alpha^2 \rangle^2 + K_{\text{TS}} \langle J_\alpha^4 \rangle, \quad (\text{A.15})$$

$$\sum_{k_1, k_2, k_3=1}^K \mathbb{E} \left[\left(m_\alpha^{(k_1)} \right)^2 m_\alpha^{(k_2)} m_\alpha^{(k_3)} \right] \quad (\text{A.16})$$

$$= \begin{array}{c} k_1 \quad k_1 \quad k_2 \quad k_3 \\ \bullet \quad \bullet \quad \bullet \quad \bullet \end{array} + \begin{array}{c} k_1 \quad k_1 \quad k_2 \quad k_3 \\ \bullet \quad \bullet \quad \bullet \quad \bullet \\ \text{---} \end{array} + \begin{array}{c} k_1 \quad k_1 \quad k_2 \quad k_3 \\ \bullet \quad \bullet \quad \bullet \quad \bullet \\ \text{---} \end{array} + \begin{array}{c} k_1 \quad k_1 \quad k_2 \quad k_3 \\ \bullet \quad \bullet \quad \bullet \quad \bullet \\ \text{---} \end{array} + \begin{array}{c} k_1 \quad k_1 \quad k_2 \quad k_3 \\ \bullet \quad \bullet \quad \bullet \quad \bullet \\ \text{---} \end{array} \quad (\text{A.17})$$

$$= K_{\text{TS}}(K_{\text{TS}} - 1)(K_{\text{TS}} - 2) \langle J_\alpha^2 \rangle \langle J_\alpha \rangle^2 + K_{\text{TS}}(K_{\text{TS}} - 1) \langle J_\alpha^2 \rangle^2 \quad (\text{A.18})$$

$$+ 2K_{\text{TS}}(K_{\text{TS}} - 1) \langle J_\alpha^3 \rangle \langle J_\alpha \rangle + K_{\text{TS}} \langle J_\alpha^4 \rangle \quad (\text{A.19})$$

and

$$\begin{aligned} \sum_{k_1, k_2, k_3, k_4=1}^K \mathbb{E} \left[m_\alpha^{(k_1)} m_\alpha^{(k_2)} m_\alpha^{(k_3)} m_\alpha^{(k_4)} \right] &= \begin{array}{c} k_1 \quad k_2 \quad k_3 \quad k_4 \\ \bullet \quad \bullet \quad \bullet \quad \bullet \end{array} \\ &+ \begin{array}{c} k_1 \quad k_2 \quad k_3 \quad k_4 \\ \bullet \quad \bullet \quad \bullet \quad \bullet \\ \text{---} \end{array} + \begin{array}{c} k_1 \quad k_2 \quad k_3 \quad k_4 \\ \bullet \quad \bullet \quad \bullet \quad \bullet \\ \text{---} \end{array} + \begin{array}{c} k_1 \quad k_2 \quad k_3 \quad k_4 \\ \bullet \quad \bullet \quad \bullet \quad \bullet \\ \text{---} \end{array} + \begin{array}{c} k_1 \quad k_2 \quad k_3 \quad k_4 \\ \bullet \quad \bullet \quad \bullet \quad \bullet \\ \text{---} \end{array} + \begin{array}{c} k_1 \quad k_2 \quad k_3 \quad k_4 \\ \bullet \quad \bullet \quad \bullet \quad \bullet \\ \text{---} \end{array} \\ &+ \begin{array}{c} k_1 \quad k_2 \quad k_3 \quad k_4 \\ \bullet \quad \bullet \quad \bullet \quad \bullet \\ \text{---} \end{array} + \begin{array}{c} k_1 \quad k_2 \quad k_3 \quad k_4 \\ \bullet \quad \bullet \quad \bullet \quad \bullet \\ \text{---} \end{array} + \begin{array}{c} k_1 \quad k_2 \quad k_3 \quad k_4 \\ \bullet \quad \bullet \quad \bullet \quad \bullet \\ \text{---} \end{array} \\ &+ \begin{array}{c} k_1 \quad k_2 \quad k_3 \quad k_4 \\ \bullet \quad \bullet \quad \bullet \quad \bullet \\ \text{---} \end{array} + \begin{array}{c} k_1 \quad k_2 \quad k_3 \quad k_4 \\ \bullet \quad \bullet \quad \bullet \quad \bullet \\ \text{---} \end{array} + \begin{array}{c} k_1 \quad k_2 \quad k_3 \quad k_4 \\ \bullet \quad \bullet \quad \bullet \quad \bullet \\ \text{---} \end{array} + \begin{array}{c} k_1 \quad k_2 \quad k_3 \quad k_4 \\ \bullet \quad \bullet \quad \bullet \quad \bullet \\ \text{---} \end{array} + \begin{array}{c} k_1 \quad k_2 \quad k_3 \quad k_4 \\ \bullet \quad \bullet \quad \bullet \quad \bullet \\ \text{---} \end{array} \\ &= K_{\text{TS}}(K_{\text{TS}} - 1)(K_{\text{TS}} - 2)(K_{\text{TS}} - 3) \langle J_\alpha \rangle^4 + 6K_{\text{TS}}(K_{\text{TS}} - 1)(K_{\text{TS}} - 2) \langle J_\alpha^2 \rangle \langle J_\alpha \rangle^2 \\ &\quad + 3K_{\text{TS}}(K_{\text{TS}} - 1) \langle J_\alpha^2 \rangle^2 + 4K_{\text{TS}}(K_{\text{TS}} - 1) \langle J_\alpha^3 \rangle \langle J_\alpha \rangle + K_{\text{TS}} \langle J_\alpha^4 \rangle. \end{aligned} \quad (\text{A.20})$$

This finally yields

$$\begin{aligned}
\mathbb{E} \left[\left(\widehat{(\Delta J_\alpha)^2}_{\text{TS}} \right)^2 \right] &= \frac{1}{(K_{\text{TS}} - 1)^2} \left[(K_{\text{TS}}(K_{\text{TS}} - 1) \langle J_\alpha^2 \rangle^2 + K_{\text{TS}} \langle J_\alpha^4 \rangle \right. \\
&\quad - \frac{2}{K_{\text{TS}}} \left(K_{\text{TS}}(K_{\text{TS}} - 1)(K_{\text{TS}} - 2) \langle J_\alpha^2 \rangle \langle J_\alpha \rangle^2 + K_{\text{TS}}(K_{\text{TS}} - 1) \langle J_\alpha^2 \rangle^2 \right. \\
&\quad \quad \left. \left. + 2K_{\text{TS}}(K_{\text{TS}} - 1) \langle J_\alpha^3 \rangle \langle J_\alpha \rangle + K_{\text{TS}} \langle J_\alpha^4 \rangle \right) \right. \\
&\quad \left. + \frac{1}{K_{\text{TS}}^2} \left(K_{\text{TS}}(K_{\text{TS}} - 1)(K_{\text{TS}} - 2)(K_{\text{TS}} - 3) \langle J_\alpha \rangle^4 + 6K_{\text{TS}}(K_{\text{TS}} - 1)(K_{\text{TS}} - 2) \langle J_\alpha^2 \rangle \langle J_\alpha \rangle^2 \right. \right. \\
&\quad \quad \left. \left. + 3K_{\text{TS}}(K_{\text{TS}} - 1) \langle J_\alpha^2 \rangle^2 + 4K_{\text{TS}}(K_{\text{TS}} - 1) \langle J_\alpha^3 \rangle \langle J_\alpha \rangle + K_{\text{TS}} \langle J_\alpha^4 \rangle \right) \right] \\
&= \frac{1}{K_{\text{TS}}} \langle J_\alpha^4 \rangle - \frac{4}{K_{\text{TS}}} \langle J_\alpha^3 \rangle \langle J_\alpha \rangle + \frac{(K_{\text{TS}} - 1)^2 + 2}{K_{\text{TS}}(K_{\text{TS}} - 1)} \langle J_\alpha^2 \rangle^2 - 2 \frac{(K_{\text{TS}} - 2)(K_{\text{TS}} - 3)}{K_{\text{TS}}(K_{\text{TS}} - 1)} \langle J_\alpha^2 \rangle \langle J_\alpha \rangle^2 \\
&\quad + \frac{(K_{\text{TS}} - 2)(K_{\text{TS}} - 3)}{K_{\text{TS}}(K_{\text{TS}} - 1)} \langle J_\alpha \rangle^4.
\end{aligned} \tag{A.21}$$

As a result, we arrive at the variance

$$\begin{aligned}
\text{Var} \left(\widehat{(\Delta J_\alpha)^2}_{\text{TS}} \right) &= \mathbb{E} \left[\left(\widehat{(\Delta J_\alpha)^2}_{\text{TS}} \right)^2 \right] - \left(\mathbb{E} \left[\widehat{(\Delta J_\alpha)^2}_{\text{TS}} \right] \right)^2 \\
&= \frac{1}{K_{\text{TS}}} \langle J_\alpha^4 \rangle - \frac{4}{K_{\text{TS}}} \langle J_\alpha^3 \rangle \langle J_\alpha \rangle + \frac{(K_{\text{TS}} - 1)^2 + 2}{K_{\text{TS}}(K_{\text{TS}} - 1)} \langle J_\alpha^2 \rangle^2 - 2 \frac{(K_{\text{TS}} - 2)(K_{\text{TS}} - 3)}{K_{\text{TS}}(K_{\text{TS}} - 1)} \langle J_\alpha^2 \rangle \langle J_\alpha \rangle^2 \\
&\quad + \frac{(K_{\text{TS}} - 2)(K_{\text{TS}} - 3)}{K_{\text{TS}}(K_{\text{TS}} - 1)} \langle J_\alpha \rangle^4 - \left(\Delta J_\alpha^2 \right)^2 \\
&= \frac{1}{K_{\text{TS}}} \langle J_\alpha^4 \rangle - \langle J_\alpha^2 \rangle^2 - \frac{4}{K_{\text{TS}}} \langle J_\alpha^3 \rangle \langle J_\alpha \rangle + \frac{(K_{\text{TS}} - 1)^2 + 2}{K_{\text{TS}}(K_{\text{TS}} - 1)} \langle J_\alpha^2 \rangle^2 \\
&\quad - 2 \frac{(K_{\text{TS}} - 2)(K_{\text{TS}} - 3) - K_{\text{TS}}(K_{\text{TS}} - 1)}{K_{\text{TS}}(K_{\text{TS}} - 1)} \langle J_\alpha^2 \rangle \langle J_\alpha \rangle^2 \\
&\quad + \frac{(K_{\text{TS}} - 2)(K_{\text{TS}} - 3) - K_{\text{TS}}(K_{\text{TS}} - 1)}{K_{\text{TS}}(K_{\text{TS}} - 1)} \langle J_\alpha \rangle^4 \\
&= \frac{1}{K_{\text{TS}}} (\Delta J_\alpha^2)^2 - \frac{K_{\text{TS}} - 1}{K_{\text{TS}}} \langle J_\alpha^2 \rangle^2 - \frac{4}{K_{\text{TS}}} \langle J_\alpha^3 \rangle \langle J_\alpha \rangle + \frac{(K_{\text{TS}} - 1)^2 + 2}{K_{\text{TS}}(K_{\text{TS}} - 1)} \langle J_\alpha^2 \rangle^2 \\
&\quad + 4 \frac{2K_{\text{TS}} - 3}{K_{\text{TS}}(K_{\text{TS}} - 1)} \langle J_\alpha^2 \rangle \langle J_\alpha \rangle^2 - 2 \frac{2K_{\text{TS}} - 3}{K_{\text{TS}}(K_{\text{TS}} - 1)} \langle J_\alpha \rangle^4 \\
&= \frac{1}{K_{\text{TS}}} (\Delta J_\alpha^2)^2 - \frac{4}{K_{\text{TS}}} \langle J_\alpha^3 \rangle \langle J_\alpha \rangle + \frac{2}{K_{\text{TS}}(K_{\text{TS}} - 1)} \langle J_\alpha^2 \rangle^2 + 4 \frac{2K_{\text{TS}} - 3}{K_{\text{TS}}(K_{\text{TS}} - 1)} \langle J_\alpha^2 \rangle \langle J_\alpha \rangle^2 \\
&\quad - 2 \frac{2K_{\text{TS}} - 3}{K_{\text{TS}}(K_{\text{TS}} - 1)} \langle J_\alpha \rangle^4.
\end{aligned} \tag{A.22}$$

We note that the above expression coincides with [83]. As the measurements in x , y , and z direction are obtained in independent experimental runs, the different estimators are statistically independent. We can thus write the variance as the sum of the variances of the individual terms. Thereby, we can derive the variances of the spin-squeezing parameters from the above expressions.

As an example, we derive the variance of $(\hat{\xi}_c)_{\text{TS}}$ that is given in the main text in Eq. (3.20):

$$\begin{aligned}
\text{Var} \left((\hat{\xi}_c)_{\text{TS}} \right) &= \text{Var} \left(\langle \widehat{J_x^2} \rangle_{\text{TS}} + \langle \widehat{J_y^2} \rangle_{\text{TS}} - (N-1) \langle \widehat{(\Delta J_z)^2} \rangle_{\text{TS}} \right) \\
&= \text{Var} \left(\langle \widehat{J_x^2} \rangle_{\text{TS}} \right) + \text{Var} \left(\langle \widehat{J_y^2} \rangle_{\text{TS}} \right) + \text{Var} \left(-(N-1) \langle \widehat{(\Delta J_z)^2} \rangle_{\text{TS}} \right) \\
&= \text{Var} \left(\langle \widehat{J_x^2} \rangle_{\text{TS}} \right) + \text{Var} \left(\langle \widehat{J_y^2} \rangle_{\text{TS}} \right) + (N-1)^2 \text{Var} \left(\langle \widehat{(\Delta J_z)^2} \rangle_{\text{TS}} \right) \\
&= \frac{1}{K_{\text{TS}}} \left[(\Delta J_x^2)^2 + (\Delta J_y^2)^2 + (N-1)^2 (\Delta J_z^2)^2 \right. \\
&\quad \left. + (N-1)^2 \left(-4 \langle J_\alpha^3 \rangle \langle J_\alpha \rangle + \frac{2}{K_{\text{TS}}-1} \langle J_\alpha^2 \rangle^2 + 4 \frac{2K_{\text{TS}}-3}{K_{\text{TS}}-1} \langle J_\alpha^2 \rangle \langle J_\alpha \rangle^2 - 2 \frac{2K_{\text{TS}}-3}{K_{\text{TS}}-1} \langle J_\alpha \rangle^4 \right) \right].
\end{aligned} \tag{A.23}$$

In the above calculation, we used that the estimators $\langle \widehat{J_x^2} \rangle_{\text{TS}}^2$, $\langle \widehat{J_y^2} \rangle_{\text{TS}}^2$, and $\langle \widehat{(\Delta J_z)^2} \rangle_{\text{TS}}$ are statistically independent.

A.2.2 Estimator based on pair correlations

We are going to derive the variances of $\langle \widehat{J_\alpha^2} \rangle_{\text{AP}}$, $\langle \widehat{(\Delta J_\alpha)^2} \rangle_{\text{AP}}$, and $\langle \widehat{J_\alpha} \rangle_{\text{AP}}^2$. As a result, this allows us to obtain the variances of the spin-squeezing parameters. We start with the variance of the estimator $\langle \widehat{J_\alpha^2} \rangle_{\text{AP}}$ in Eq. (3.9). As the constant term does not contribute and the second term is a sum of independent random variables, we obtain

$$\text{Var} \left(\langle \widehat{J_\alpha^2} \rangle_{\text{AP}} \right) = \text{Var} \left(\frac{N}{4} + \frac{1}{K_{\text{AP}}} \sum_P \sum_{k=1}^{K_{\text{AP}}} s_\alpha^{(P_1,k)} s_\alpha^{(P_2,k)} \right) = \frac{1}{K_{\text{AP}}^2} \sum_P \sum_{k=1}^{K_{\text{AP}}} \text{Var} \left(s_\alpha^{(P_1,k)} s_\alpha^{(P_2,k)} \right). \tag{A.24}$$

With the help of Eq. (A.1) and

$$\mathbb{E} \left[\left(s_\alpha^{(P_1,k)} s_\alpha^{(P_2,k)} \right)^2 \right] = \frac{1}{16} \langle (\sigma_\alpha^{(P_1)})^2 (\sigma_\alpha^{(P_2)})^2 \rangle = \frac{1}{16}, \tag{A.25}$$

we can evaluate the variances of the individual terms:

$$\text{Var} \left(s_\alpha^{(P_1,k)} s_\alpha^{(P_2,k)} \right) = \mathbb{E} \left[\left(s_\alpha^{(P_1,k)} s_\alpha^{(P_2,k)} \right)^2 \right] - \mathbb{E} \left[s_\alpha^{(P_1,k)} s_\alpha^{(P_2,k)} \right]^2 = \frac{1}{16} \left(1 - \langle \sigma_\alpha^{(P_1)} \sigma_\alpha^{(P_2)} \rangle^2 \right). \tag{A.26}$$

As a result, we obtain

$$\text{Var} \left(\langle \widehat{J_\alpha^2} \rangle_{\text{AP}} \right) = \frac{1}{K_{\text{AP}}^2} \sum_P \sum_{k=1}^{K_{\text{AP}}} \frac{1}{16} \left(1 - \langle \sigma_\alpha^{(P_1)} \sigma_\alpha^{(P_2)} \rangle^2 \right) = \frac{1}{16 K_{\text{AP}}} \left(N(N-1) - \sum_{i \neq j} \langle \sigma_\alpha^{(i)} \sigma_\alpha^{(j)} \rangle^2 \right), \tag{A.27}$$

where we made it explicit that the sum over all pairs $P = (i, j)$ only contains distinct pairs $i \neq j$.

To determine the variance of the estimator $\langle \widehat{(\Delta J_\alpha)^2} \rangle_{\text{AP}}$ in Eq. (3.10), we use Bienaymé's identity

[191]:

$$\begin{aligned}
& \text{Var} \left((\widehat{\Delta J_\alpha})^2_{\text{AP}} \right) \\
&= \text{Var} \left(\frac{N}{4} + \frac{1}{K_{\text{AP}}} \sum_P \sum_{k=1}^{K_{\text{AP}}} s_\alpha^{(P_1,k)} s_\alpha^{(P_2,k)} - \frac{1}{K_{\text{AP}}(K_{\text{AP}}-1)(N-1)^2} \sum_{P,Q} \sum_{k \neq l}^{K_{\text{AP}}} s_\alpha^{(P_1,k)} s_\alpha^{(Q_2,l)} \right) \\
&= \frac{1}{K_{\text{AP}}^2} \sum_P \sum_{k=1}^{K_{\text{AP}}} \text{Var} \left(s_\alpha^{(P_1,k)} s_\alpha^{(P_2,k)} \right) + \frac{1}{K_{\text{AP}}^2 (K_{\text{AP}}-1)^2 (N-1)^4} \text{Var} \left(\sum_{P,Q} \sum_{k \neq l}^{K_{\text{AP}}} s_\alpha^{(P_1,k)} s_\alpha^{(Q_2,l)} \right) \\
&\quad - \frac{2}{K_{\text{AP}}^2 (K_{\text{AP}}-1)(N-1)^2} \text{Cov} \left(\sum_P \sum_{k=1}^{K_{\text{AP}}} s_\alpha^{(P_1,k)} s_\alpha^{(P_2,k)}, \sum_{P,Q} \sum_{k \neq l}^{K_{\text{AP}}} s_\alpha^{(P_1,k)} s_\alpha^{(Q_2,l)} \right).
\end{aligned} \tag{A.28}$$

In the above expression, we applied that the constant term does not contribute to the variance. Moreover, the random variables in the first sum are statistically independent, such that the variance of the sum is just the sum of the variances. The variance can be obtained by plugging in the expressions

$$\begin{aligned}
& \text{Var} \left(s_\alpha^{(P_1,k)} s_\alpha^{(P_2,k)} \right) = \frac{1}{16} \left(1 - \langle \sigma_\alpha^{(P_1)} \sigma_\alpha^{(P_2)} \rangle^2 \right), \\
& \text{Var} \left(\sum_{P,Q} \sum_{k \neq l}^{K_{\text{AP}}} s_\alpha^{(P_1,k)} s_\alpha^{(Q_2,l)} \right) \\
&= K_{\text{AP}}(K_{\text{AP}}-1)(K_{\text{AP}}-2)(K_{\text{AP}}-3)(N-1)^4 \langle J_\alpha \rangle^4 \\
&\quad + K_{\text{AP}}(K_{\text{AP}}-1)(K_{\text{AP}}-2)(N-1)^2 \langle J_\alpha \rangle^2 \left[2(N-1) \left((N-1) \langle J_\alpha \rangle^2 + \frac{N}{4} - \sum_{i=1}^N \frac{1}{4} \langle \sigma_\alpha^{(i)} \rangle^2 \right) \right. \\
&\quad \quad \left. + 2 \left(\langle J_\alpha^2 \rangle + N(N-2) \langle J_\alpha \rangle^2 - \frac{N}{4} + \sum_{i=1}^N \frac{1}{4} \langle \sigma_\alpha^{(i)} \rangle^2 \right) \right] \\
&\quad + K_{\text{AP}}(K_{\text{AP}}-1) \left[(N-1)^2 \left((N-1) \langle J_\alpha \rangle^2 + \frac{N}{4} - \sum_{i=1}^N \frac{1}{4} \langle \sigma_\alpha^{(i)} \rangle^2 \right)^2 \right. \\
&\quad \quad \left. + \left(\langle J_\alpha^2 \rangle + N(N-2) \langle J_\alpha \rangle^2 - \frac{N}{4} + \sum_{i=1}^N \frac{1}{4} \langle \sigma_\alpha^{(i)} \rangle^2 \right)^2 \right] \\
&\quad - K_{\text{AP}}^2 (K_{\text{AP}}-1)^2 (N-1)^4 \langle J_\alpha \rangle^4,
\end{aligned} \tag{A.29}$$

and

$$\begin{aligned}
& \text{Cov} \left(\sum_P \sum_{k=1}^{K_{\text{AP}}} s_{\alpha}^{(P_1,k)} s_{\alpha}^{(P_2,k)}, \sum_{P,Q} \sum_{k \neq l}^{K_{\text{AP}}} s_{\alpha}^{(P_1,k)} s_{\alpha}^{(Q_2,l)} \right) \\
&= K_{\text{AP}}(K_{\text{AP}} - 1)(K_{\text{AP}} - 2) \left(\langle J_{\alpha}^2 \rangle - \frac{N}{4} \right) (N - 1)^2 \langle J_{\alpha} \rangle^2 \\
&+ K_{\text{AP}}(K_{\text{AP}} - 1)(N - 1) \langle J_{\alpha} \rangle \\
&\quad \left[\frac{N - 1}{2} \langle J_{\alpha} \rangle + 2(N - 1) \langle J_{\alpha} \rangle \left(\langle J_{\alpha}^2 \rangle - \frac{N}{4} \right) - \sum_{i \neq j} \frac{1}{8} \langle \sigma_{\alpha}^{(i)} \sigma_{\alpha}^{(j)} \rangle \left(\langle \sigma_{\alpha}^{(i)} \rangle + \langle \sigma_{\alpha}^{(j)} \rangle \right) \right] \\
&- K_{\text{AP}}^2(K_{\text{AP}} - 1)(N - 1)^2 \langle J_{\alpha} \rangle^2 \left(\langle J_{\alpha}^2 \rangle - \frac{N}{4} \right). \tag{A.30}
\end{aligned}$$

Finally, we can combine Eq. (A.27) and Eq. (A.28) to obtain the variance of the spin-squeezing parameters, e.g.,

$$\text{Var} \left((\hat{\xi}_c)_{\text{AP}1} \right) = \text{Var} \left(\widehat{\langle J_x^2 \rangle}_{\text{AP}} \right) + \text{Var} \left(\widehat{\langle J_y^2 \rangle}_{\text{AP}} \right) + (N - 1)^2 \text{Var} \left(\widehat{\langle \Delta J_z \rangle}_{\text{AP}}^2 \right). \tag{A.31}$$

However, due to the size of the equation, we omit an explicit expression.

For the variance of $\widehat{\langle J_{\alpha} \rangle}_{\text{AP}}^2$, we compute

$$\text{Var} \left(\widehat{\langle J_{\alpha} \rangle}_{\text{AP}}^2 \right) = \text{Var} \left(\sum_{i,j=1}^N \frac{1}{K_{\text{AP}}} \sum_{k=1}^{\frac{K_{\text{AP}}}{2}} s_{\alpha}^{(i,2k)} s_{\alpha}^{(j,2k-1)} \right) = \frac{4}{K_{\text{AP}}^2} \sum_{i,j=1}^N \sum_{k=1}^{\frac{K_{\text{AP}}}{2}} \text{Var} \left(s_{\alpha}^{(i,2k)} s_{\alpha}^{(j,2k-1)} \right). \tag{A.32}$$

As we have $(s_{\alpha}^{(i,k)})^2 = \frac{1}{4}$, we obtain

$$\begin{aligned}
\text{Var} \left(s_{\alpha}^{(i,2k)} s_{\alpha}^{(j,2k-1)} \right) &= \mathbb{E} \left[\left(s_{\alpha}^{(i,2k)} s_{\alpha}^{(j,2k-1)} \right)^2 \right] - \mathbb{E} \left[s_{\alpha}^{(i,2k)} s_{\alpha}^{(j,2k-1)} \right]^2 \\
&= \frac{1}{16} - \left(\mathbb{E} \left[s_{\alpha}^{(i,2k)} \right] \mathbb{E} \left[s_{\alpha}^{(j,2k-1)} \right] \right)^2 = \frac{1}{16} \left(1 - \langle \sigma_{\alpha}^{(i)} \rangle^2 \langle \sigma_{\alpha}^{(j)} \rangle^2 \right). \tag{A.33}
\end{aligned}$$

The variance of $\widehat{\langle J_{\alpha} \rangle}_{\text{AP}}^2$ thus takes the form

$$\text{Var} \left(\widehat{\langle J_{\alpha} \rangle}_{\text{AP}}^2 \right) = \frac{4}{K_{\text{AP}}^2} \sum_{i,j=1}^N \sum_{k=1}^{\frac{K_{\text{AP}}}{2}} \frac{1}{16} \left(1 - \langle \sigma_{\alpha}^{(i)} \rangle^2 \langle \sigma_{\alpha}^{(j)} \rangle^2 \right) = \frac{1}{8K_{\text{AP}}} \left(N^2 - \sum_{i,j=1}^N \langle \sigma_{\alpha}^{(i)} \rangle^2 \langle \sigma_{\alpha}^{(j)} \rangle^2 \right). \tag{A.34}$$

With the assumption that the terms $\langle J_z^2 \rangle$ and $\langle J_z \rangle^2$ are calculated from the data of different experimental runs, we have $\mathbb{E}[\widehat{\langle J_z^2 \rangle} \widehat{\langle J_z \rangle^2}] = \mathbb{E}[\widehat{\langle J_z^2 \rangle}] \mathbb{E}[\widehat{\langle J_z \rangle^2}]$. Hence, the variance of the spin-squeezing estimators can be deduced from the variances derived in this section. For the estimator

$(\hat{\xi}_c)_{\text{AP}2}$, we obtain

$$\begin{aligned}
\text{Var} \left((\hat{\xi}_c)_{\text{AP}2} \right) &= \text{Var} \left(\widehat{\langle J_x^2 \rangle}_{\text{AP}} \right) + \text{Var} \left(\widehat{\langle J_y^2 \rangle}_{\text{AP}} \right) + (N-1)^2 \left[\text{Var} \left(\widehat{\langle J_z^2 \rangle}_{\text{AP}} \right) + \text{Var} \left(\widehat{\langle J_z \rangle}_{\text{AP}}^2 \right) \right] \\
&= \frac{1}{16K_{\text{AP}}} \left[2N(N-1) - \sum_{i \neq j} \left(\langle \sigma_x^{(i)} \sigma_x^{(j)} \rangle^2 + \langle \sigma_y^{(i)} \sigma_y^{(j)} \rangle^2 \right) \right] \\
&\quad + \frac{(N-1)^2}{8K_{\text{AP}}} \left[\frac{N(N-1)}{2} - \frac{1}{2} \sum_{i \neq j} \langle \sigma_z^{(i)} \sigma_z^{(j)} \rangle^2 + N^2 - \sum_{i,j} \langle \sigma_z^{(i)} \rangle^2 \langle \sigma_z^{(j)} \rangle^2 \right].
\end{aligned} \tag{A.35}$$

A.2.3 Estimator based on random pair correlations

We now consider the estimation using random pair correlations. First, we will derive the variance of the estimator $\widehat{\langle J_\alpha^2 \rangle}_{\text{RP}}$ in Eq. (3.14). For this purpose we use that the constant term in the estimator does not contribute to the variance. Moreover, all terms in the second sum are independent random variables, and thus we can write the variance as the sum of the variances of the individual terms, i.e.,

$$\begin{aligned}
\text{Var} \left(\widehat{\langle J_\alpha^2 \rangle}_{\text{RP}} \right) &= \text{Var} \left(\frac{N}{4} + \frac{N(N-1)}{K_{\text{RP}}L_{\text{RP}}} \sum_{l=1}^{L_{\text{RP}}} \sum_{k=1}^{K_{\text{RP}}} s_\alpha^{(\mathcal{I}_l, k)} s_\alpha^{(\mathcal{J}_l, k)} \right) \\
&= \frac{N^2(N-1)^2}{K_{\text{RP}}^2 L_{\text{RP}}^2} \sum_{l=1}^{L_{\text{RP}}} \sum_{k=1}^{K_{\text{RP}}} \text{Var} \left(s_\alpha^{(\mathcal{I}_l, k)} s_\alpha^{(\mathcal{J}_l, k)} \right).
\end{aligned} \tag{A.36}$$

To calculate the variances of the individual terms, we can make use of Eq. (A.8) and that $(s_\alpha^{(\mathcal{I}_l, k)})^2 = \frac{1}{4}$:

$$\begin{aligned}
\text{Var} \left(s_\alpha^{(\mathcal{I}_l, k)} s_\alpha^{(\mathcal{J}_l, k)} \right) &= \mathbb{E} \left[\left(s_\alpha^{(\mathcal{I}_l, k)} s_\alpha^{(\mathcal{J}_l, k)} \right)^2 \right] - \left(\mathbb{E} \left[s_\alpha^{(\mathcal{I}_l, k)} s_\alpha^{(\mathcal{J}_l, k)} \right] \right)^2 \\
&= \frac{1}{16} - \frac{1}{N^2(N-1)^2} \left(\langle J_\alpha^2 \rangle - \frac{N}{4} \right)^2.
\end{aligned} \tag{A.37}$$

Thus, the variance takes the form

$$\begin{aligned}
\text{Var} \left(\widehat{\langle J_\alpha^2 \rangle}_{\text{RP}} \right) &= \frac{N^2(N-1)^2}{K_{\text{RP}}^2 L_{\text{RP}}^2} \sum_{l=1}^{L_{\text{RP}}} \sum_{k=1}^{K_{\text{RP}}} \left[\frac{1}{16} - \frac{1}{N^2(N-1)^2} \left(\langle J_\alpha^2 \rangle - \frac{N}{4} \right)^2 \right] \\
&= \frac{1}{K_{\text{RP}}L_{\text{RP}}} \left(\frac{N^3(N-2)}{16} - \langle J_\alpha^2 \rangle^2 + \frac{N}{2} \langle J_\alpha^2 \rangle \right).
\end{aligned} \tag{A.38}$$

To calculate the variance of the estimator $\widehat{(\Delta J_\alpha)^2}_{\text{RP}}$, we restrict ourselves to the case of one repetition for each random pair, i.e., $K_{\text{RP}} = 1$. Moreover, we use that the first constant term in

Eq. (3.16) does not contribute to the variance:

$$\begin{aligned}
\text{Var} \left(\widehat{(\Delta J_\alpha)^2}_{\text{RP}} \right) &= \text{Var} \left(\frac{N}{4} + \frac{N(N-1)}{L_{\text{RP}}} \sum_{l=1}^{L_{\text{RP}}} s_\alpha^{(\mathcal{I}_l)} s_\alpha^{(\mathcal{J}_l)} - \frac{N^2}{L_{\text{RP}}(L_{\text{RP}}-1)} \sum_{l \neq m}^{L_{\text{RP}}} s_\alpha^{(\mathcal{I}_l)} s_\alpha^{(\mathcal{J}_m)} \right) \\
&= \frac{N^2(N-1)^2}{L_{\text{RP}}^2} \sum_{l=1}^{L_{\text{RP}}} \text{Var} \left(s_\alpha^{(\mathcal{I}_l)} s_\alpha^{(\mathcal{J}_l)} \right) + \frac{N^4}{L_{\text{RP}}^2(L_{\text{RP}}-1)^2} \text{Var} \left(\sum_{l \neq m}^{L_{\text{RP}}} s_\alpha^{(\mathcal{I}_l)} s_\alpha^{(\mathcal{J}_m)} \right) \\
&\quad - 2 \frac{N^3(N-1)}{L_{\text{RP}}^2(L_{\text{RP}}-1)} \text{Cov} \left(\sum_{l=1}^{L_{\text{RP}}} s_\alpha^{(\mathcal{I}_l)} s_\alpha^{(\mathcal{J}_l)}, \sum_{l \neq m}^{L_{\text{RP}}} s_\alpha^{(\mathcal{I}_l)} s_\alpha^{(\mathcal{J}_m)} \right).
\end{aligned} \tag{A.39}$$

In the above expression we used Bienaymé's identity. In addition, we applied that the random variables $s_\alpha^{(\mathcal{I}_l)} s_\alpha^{(\mathcal{J}_l)}$ in the first sum are statistically independent. The separate terms evaluate to

$$\begin{aligned}
\text{Var} \left(s_\alpha^{(\mathcal{I}_l)} s_\alpha^{(\mathcal{J}_l)} \right) &= \frac{1}{16} - \left[\frac{1}{N(N-1)} \left(\langle J_\alpha^2 \rangle - \frac{N}{4} \right) \right]^2, \\
\text{Var} \left(\sum_{l \neq m}^{L_{\text{RP}}} s_\alpha^{(\mathcal{I}_l)} s_\alpha^{(\mathcal{J}_m)} \right) &= \frac{L_{\text{RP}}}{16} \left(-\frac{32 \langle J_\alpha \rangle^4 (L_{\text{RP}}-1)(2L_{\text{RP}}-3)}{N^4} \right. \\
&\quad + \frac{8 \langle J_\alpha \rangle^2 (L_{\text{RP}}-2)(L_{\text{RP}}-1)(4 \langle J_\alpha^2 \rangle + (N-2)N)}{(N-1)N^3} \\
&\quad \left. + \frac{(L_{\text{RP}}-1)(N-4 \langle J_\alpha^2 \rangle)^2}{(N-1)^2 N^2} + L_{\text{RP}} - 1 \right), \\
\text{Cov} \left(\sum_{l=1}^{L_{\text{RP}}} s_\alpha^{(\mathcal{I}_l)} s_\alpha^{(\mathcal{J}_l)}, \sum_{l \neq m}^{L_{\text{RP}}} s_\alpha^{(\mathcal{I}_l)} s_\alpha^{(\mathcal{J}_m)} \right) &= \frac{L_{\text{RP}}(L_{\text{RP}}-1)(L_{\text{RP}}-2)}{N(N-1)} \left(\langle J_\alpha^2 \rangle - \frac{N}{4} \right) \frac{1}{N^2} \langle J_\alpha \rangle^2 \\
&\quad + \frac{L_{\text{RP}}(L_{\text{RP}}-1)}{2N^2} \langle J_\alpha \rangle^2 \\
&\quad - \frac{L_{\text{RP}}^2(L_{\text{RP}}-1)}{N^3(N-1)} \left(\langle J_\alpha^2 \rangle - \frac{N}{4} \right) \langle J_\alpha \rangle^2.
\end{aligned} \tag{A.40}$$

Plugging the expressions in Eq. (A.39) yields

$$\begin{aligned}
&\text{Var} \left(\widehat{(\Delta J_\alpha)^2}_{\text{RP}} \right) \\
&= \frac{1}{16(L_{\text{RP}}-1)L_{\text{RP}}(N-1)^2} \left[-32 \langle J_\alpha \rangle^4 (2L_{\text{RP}}-3)(N-1)^2 \right. \\
&\quad - 8 \langle J_\alpha \rangle^2 (N-1) (4 \langle J_\alpha^2 \rangle (-3L_{\text{RP}}N + 2L_{\text{RP}} + 4N - 2) + N^2(L_{\text{RP}}N - 2)) \\
&\quad - 16 \langle J_\alpha^2 \rangle^2 (L_{\text{RP}}(N-1)^2 - 2N(N-1) - 1) + 8 \langle J_\alpha^2 \rangle N (L_{\text{RP}}(N-1)^2 - 2N(N-1) - 1) \\
&\quad \left. + N^3 (L_{\text{RP}}(N-2)(N-1)^2 + N(2N-3) + 2) \right].
\end{aligned} \tag{A.41}$$

With the help of Eqs. (A.38) and (A.41), we can derive the variances for scheme RP1 in case

$K_{\text{RP1}} = 1$, e.g.,

$$\begin{aligned}
& \text{Var} \left((\hat{\xi}_c)_{\text{RP1}} \right) = \text{Var} \left(\widehat{\langle J_x^2 \rangle}_{\text{RP1}} \right) + \text{Var} \left(\widehat{\langle J_y^2 \rangle}_{\text{RP1}} \right) + (N-1)^2 \text{Var} \left(\widehat{(\Delta J_z)^2}_{\text{RP1}} \right) \\
&= \frac{1}{L_{\text{RP}}} \left[\frac{N^3(N-2)}{8} - (\langle J_x^2 \rangle^2 + \langle J_y^2 \rangle^2) + \frac{N}{2} (\langle J_x^2 \rangle + \langle J_y^2 \rangle) \right] \\
&+ \frac{1}{16(L_{\text{RP}}-1)L_{\text{RP}}(N-1)^2} \left[-32\langle J_\alpha \rangle^4(2L_{\text{RP}}-3)(N-1)^2 \right. \\
&\quad - 8\langle J_\alpha \rangle^2(N-1)(4\langle J_\alpha^2 \rangle(-3L_{\text{RP}}N+2L_{\text{RP}}+4N-2) + N^2(L_{\text{RP}}N-2)) \\
&\quad - 16\langle J_\alpha^2 \rangle^2(L_{\text{RP}}(N-1)^2 - 2N(N-1) - 1) + 8\langle J_\alpha^2 \rangle N(L_{\text{RP}}(N-1)^2 - 2N(N-1) - 1) \\
&\quad \left. + N^3(L_{\text{RP}}(N-2)(N-1)^2 + N(2N-3) + 2) \right]. \tag{A.42}
\end{aligned}$$

Similarly, we obtain the variance of the estimator $\widehat{\langle J_\alpha \rangle}_{\text{RP}}^2$ in Eq. (3.18) as the individual terms of the sum are independent:

$$\begin{aligned}
\text{Var} \left(\widehat{\langle J_\alpha \rangle}_{\text{RP}}^2 \right) &= \text{Var} \left(\frac{2N^2}{K_{\text{RP}}L_{\text{RP}}} \sum_{l=1}^{L_{\text{RP}}} \sum_{k=1}^{\frac{K_{\text{RP}}}{2}} s_\alpha^{(\mathcal{I}_l, 2k)} s_\alpha^{(\mathcal{J}_l, 2k-1)} \right) \\
&= \frac{4N^4}{K_{\text{RP}}^2 L_{\text{RP}}^2} \sum_{l=1}^{L_{\text{RP}}} \sum_{k=1}^{\frac{K_{\text{RP}}}{2}} \text{Var} \left(s_\alpha^{(\mathcal{I}_l, 2k)} s_\alpha^{(\mathcal{J}_l, 2k-1)} \right). \tag{A.43}
\end{aligned}$$

The variance of the individual terms yields

$$\begin{aligned}
\text{Var} \left(s_\alpha^{(\mathcal{I}_l, 2k)} s_\alpha^{(\mathcal{J}_l, 2k-1)} \right) &= \mathbb{E} \left[\left(s_\alpha^{(\mathcal{I}_l, 2k)} s_\alpha^{(\mathcal{J}_l, 2k-1)} \right)^2 \right] - \left(\mathbb{E} \left[s_\alpha^{(\mathcal{I}_l, 2k)} s_\alpha^{(\mathcal{J}_l, 2k-1)} \right] \right)^2 \\
&= \frac{1}{16} - \left(\mathbb{E} \left[s_\alpha^{(\mathcal{I}_l, 2k)} \right] \mathbb{E} \left[s_\alpha^{(\mathcal{J}_l, 2k-1)} \right] \right)^2 = \frac{1}{16} - \frac{1}{N^4} \langle J_\alpha \rangle^4. \tag{A.44}
\end{aligned}$$

In the above equation, we have used again that $(s_\alpha^{(\mathcal{I}_l, k)})^2 = \frac{1}{4}$. In addition, $s_\alpha^{(\mathcal{I}_l, 2k)}$ and $s_\alpha^{(\mathcal{J}_l, 2k-1)}$ are obtained in different experimental runs and are thus independent. Finally, we have applied Eq. (A.7). As a result, we end up with

$$\begin{aligned}
\text{Var} \left(\widehat{\langle J_\alpha \rangle}_{\text{RP}}^2 \right) &= \frac{4N^4}{K_{\text{RP}}^2 L_{\text{RP}}^2} \sum_{l=1}^{L_{\text{RP}}} \sum_{k=1}^{\frac{K_{\text{RP}}}{2}} \left(\frac{1}{16} - \frac{1}{N^4} \langle J_\alpha \rangle^4 \right) = \frac{2N^4}{K_{\text{RP}}L_{\text{RP}}} \left(\frac{1}{16} - \frac{1}{N^4} \langle J_\alpha \rangle^4 \right) \\
&= \frac{1}{K_{\text{RP}}L_{\text{RP}}} \left(\frac{N^4}{8} - 2\langle J_\alpha \rangle^4 \right). \tag{A.45}
\end{aligned}$$

In case $\langle J_z^2 \rangle$ and $\langle J_z \rangle^2$ are estimated from different datasets, we can derive the variance of

$(\hat{\xi}_c)_{\text{RP2}}$:

$$\begin{aligned}
\text{Var} \left((\hat{\xi}_c)_{\text{RP2}} \right) &= \text{Var} \left(\widehat{\langle J_x^2 \rangle}_{\text{RP}} \right) + \text{Var} \left(\widehat{\langle J_y^2 \rangle}_{\text{RP}} \right) + (N-1)^2 \left(\text{Var} \left(\widehat{\langle J_z^2 \rangle}_{\text{RP}} \right) + \text{Var} \left(\widehat{\langle J_x \rangle}_{\text{RP}}^2 \right) \right) \\
&= \frac{1}{K_{\text{RP2}} L_{\text{RP2}}} \left[\frac{N^3(N-2)}{8} - (\langle J_x^2 \rangle + \langle J_y^2 \rangle) + \frac{N}{2} (\langle J_x \rangle + \langle J_y \rangle) \right. \\
&\quad \left. + (N-1)^2 \left(\frac{N^3(N-2)}{16} - \langle J_z^2 \rangle + \frac{N}{2} \langle J_z \rangle + \frac{N^4}{8} - 2\langle J_z \rangle^4 \right) \right] \\
&= \frac{1}{K_{\text{RP2}} L_{\text{RP2}}} \left[\frac{N^3(N-2)}{8} + \frac{N}{2} (\langle J_x^2 \rangle + \langle J_y^2 \rangle + \langle J_z^2 \rangle) - (\langle J_x \rangle^2 + \langle J_y \rangle^2 + \langle J_z \rangle^2) \right. \\
&\quad \left. + N(N-2) \left(\frac{N}{2} \langle J_z \rangle - \langle J_z \rangle^2 \right) + (N-1)^2 \left(\frac{N^3(N-2)}{16} + \frac{N^4}{8} - 2\langle J_z \rangle^4 \right) \right]. \tag{A.46}
\end{aligned}$$

A.3 Expressions for the singlet and Dicke state

In this Appendix we derive the expressions of the variances for the singlet state in Eq. (1.36) and the Dicke state in Eq. (1.38). For this purpose we evaluate all expectation values that appear in the variances of the different schemes. Specifically, we give explicit expressions for Eq. (3.20), Eq. (A.31), Eq. (A.35), Eq. (A.42), and Eq. (A.46).

A.3.1 Singlet state

We start with the singlet state in Eq. (1.37), i.e.,

$$|\Psi^-\rangle = \bigotimes_{k=1}^{N/2} |\psi^-\rangle, \tag{A.47}$$

with the two-qubit singlet state $|\psi^-\rangle = \frac{1}{\sqrt{2}}(|01\rangle - |10\rangle)$. The state $|\Psi^-\rangle$ is indeed a many-body singlet state, as it is an eigenstate of the total angular momentum J_α :

$$\begin{aligned}
J_\alpha |\Psi^-\rangle &= \left(\frac{1}{2} \sum_{i=1}^N \sigma_\alpha^{(i)} \right) \bigotimes_{k=1}^{N/2} |\psi^-\rangle \\
&= \frac{1}{2} \sum_{i=1}^{N/2} |\psi^-\rangle \otimes \dots \otimes |\psi^-\rangle \otimes \underbrace{\left(\sigma_\alpha^{(2i-1)} + \sigma_\alpha^{(2i)} \right)}_{=0} |\psi^-\rangle \otimes |\psi^-\rangle \otimes \dots \otimes |\psi^-\rangle = 0. \tag{A.48}
\end{aligned}$$

Thus we obtain that for $|\Psi^-\rangle$ all moments of the angular momentum are zero, i.e., for all $n \in \mathbb{N}$,

$$\langle J_\alpha^n \rangle = 0. \tag{A.49}$$

In particular, this also shows the defining property of many-body singlet states in Eq. (1.36). In addition, we have that $\sigma_x \otimes \sigma_x |\psi^-\rangle = -|\psi^-\rangle$, $\sigma_y \otimes \sigma_y |\psi^-\rangle = |\psi^-\rangle$ and $\sigma_z \otimes \sigma_z |\psi^-\rangle = -|\psi^-\rangle$. Therefore, the expectation value

$$\langle \Psi^- | \sigma_\alpha^{(i)} \sigma_\alpha^{(j)} | \Psi^- \rangle = \begin{cases} \pm 1, & \text{in case the reduced state is } \rho_{ij} = |\psi^-\rangle \langle \psi^-| \\ 0, & \text{otherwise.} \end{cases} \tag{A.50}$$

In the above equation, we have used that for the two-qubit singlet state holds, $\langle \psi^- | \sigma_\alpha \otimes \mathbb{1} | \psi^- \rangle = \langle \psi^- | \mathbb{1} \otimes \sigma_\alpha | \psi^- \rangle = 0$. As a result, we obtain the following for the singlet state $|\Psi^-\rangle$:

$$\sum_{i \neq j} \langle \sigma_\alpha^{(i)} \sigma_\alpha^{(j)} \rangle^2 = \sum_{\substack{i \neq j \\ \rho_{ij} = |\psi^-\rangle \langle \psi^-|}} 1 = N, \quad (\text{A.51})$$

as $|\Psi^-\rangle$ is composed of $\frac{N}{2}$ two-qubit singlet states $|\psi^-\rangle$ and each pair is counted twice. Moreover, we obtain

$$\sum_{i,j=1}^N \langle \sigma_\alpha^{(i)} \rangle^2 \langle \sigma_\alpha^{(j)} \rangle^2 = 0. \quad (\text{A.52})$$

With these expressions we can evaluate the variances of the estimators $\hat{\xi}_b$ and $\hat{\xi}_d$, which results in

$$\begin{aligned} \text{Var} \left((\hat{\xi}_b)_{\text{TS}} \right) &= 0, \\ \text{Var} \left((\hat{\xi}_b)_{\text{AP1}} \right) &= \frac{3N (K_{\text{AP1}}(N-2)(N-1)^4 - N^5 + 6N^4 - 13N^3 + 14N^2 - 7N + 2)}{16(K_{\text{AP1}} - 1)K_{\text{AP1}}(N-1)^4}, \\ \text{Var} \left((\hat{\xi}_b)_{\text{AP2}} \right) &= \frac{3N(3N-2)}{16K_{\text{AP2}}}, \\ \text{Var} \left((\hat{\xi}_b)_{\text{RP1}} \right) &= \frac{3N^3 (L_{\text{RP1}}(N-2)(N-1)^2 + 2N^2 - 3N + 2)}{16(L_{\text{RP1}} - 1)L_{\text{RP1}}(N-1)^2}, \\ \text{Var} \left((\hat{\xi}_b)_{\text{RP2}} \right) &= \frac{3N^3(3N-2)}{16K_{\text{RP2}}L_{\text{RP2}}}. \end{aligned} \quad (\text{A.53})$$

and

$$\begin{aligned} \text{Var} \left((\hat{\xi}_d)_{\text{TS}} \right) &= 0, \\ \text{Var} \left((\hat{\xi}_d)_{\text{AP1}} \right) &= \frac{N (K_{\text{AP1}}(N-1)^2 (2N^3 - 8N^2 + 11N - 6) - 2N^5 + 12N^4 - 27N^3 + 32N^2 - 19N + 6)}{16(K_{\text{AP1}} - 1)K_{\text{AP1}}(N-1)^2}, \\ \text{Var} \left((\hat{\xi}_d)_{\text{AP2}} \right) &= \frac{N (6N^3 - 16N^2 + 15N - 6)}{16K_{\text{AP2}}}, \\ \text{Var} \left((\hat{\xi}_d)_{\text{RP1}} \right) &= \frac{N^3 (L_{\text{RP1}} (2N^3 - 8N^2 + 11N - 6) + 4N^2 - 7N + 6)}{16(L_{\text{RP1}} - 1)L_{\text{RP1}}}, \\ \text{Var} \left((\hat{\xi}_d)_{\text{RP2}} \right) &= \frac{N^3 (6N^3 - 16N^2 + 15N - 6)}{16K_{\text{RP2}}L_{\text{RP2}}}. \end{aligned} \quad (\text{A.54})$$

A.3.2 Dicke states

The first and second moments of the Dicke states are [57]

$$\begin{aligned} (\langle J_x \rangle, \langle J_y \rangle, \langle J_z \rangle) &= (0, 0, \frac{N}{2} - m), \\ (\langle J_x^2 \rangle, \langle J_y^2 \rangle, \langle J_z^2 \rangle) &= \left[\frac{N}{4} + \frac{m(N-m)}{2}, \frac{N}{4} + \frac{m(N-m)}{2}, \left(\frac{N}{2} - m \right)^2 \right]. \end{aligned} \quad (\text{A.55})$$

Moreover, the Dicke states $|D_{N,m}\rangle$ in Eq. (1.38) are eigenstates of J_z :

$$J_z |D_{N,m}\rangle = \binom{N}{m}^{-\frac{1}{2}} \sum_k \underbrace{J_z P_k(|1_1, \dots, 1_m, 0_{m+1}, \dots, 0_N\rangle)}_{=[(N-m)-m]/2P_k(|1_1, \dots, 1_m, 0_{m+1}, \dots, 0_N\rangle)} = \left[\frac{N}{2} - m\right] |D_{N,m}\rangle. \quad (\text{A.56})$$

Therefore, we have for the Dicke states

$$\langle J_z^n \rangle = \left(\frac{N}{2} - m\right)^n. \quad (\text{A.57})$$

To evaluate the variances of the estimators, we need the fourth moments of J_x and J_y . Therefore, we calculate for $\alpha = x, y$ and $i \neq j$

$$\begin{aligned} & \langle D_{N,m} | \sigma_\alpha^{(i)} \sigma_\alpha^{(j)} | D_{N,m} \rangle \\ &= \left[\binom{N}{m}^{-\frac{1}{2}} \sum_k P_k(|1_1, \dots, 1_m, 0_{m+1}, \dots, 0_N\rangle) \right] \sigma_\alpha^{(i)} \sigma_\alpha^{(j)} \\ & \quad \left[\binom{N}{m}^{-\frac{1}{2}} \sum_l P_l(|1_1, \dots, 1_m, 0_{m+1}, \dots, 0_N\rangle) \right] \\ &= \binom{N}{m}^{-1} \sum_{k,l} \langle s_{P_k(1)} | s_{P_l(1)} \rangle \times \dots \times \langle s_{P_k(i)} | \sigma_\alpha^{(i)} | s_{P_l(i)} \rangle \times \dots \times \langle s_{P_k(j)} | \sigma_\alpha^{(j)} | s_{P_l(j)} \rangle \times \dots \\ & \quad \times \langle s_{P_k(N)} | s_{P_l(N)} \rangle, \end{aligned} \quad (\text{A.58})$$

where $s_i \in \{0, 1\}$ is the state of qubit i . For the terms to be nonzero, we have $|s_{P_k(n)}\rangle = |s_{P_l(n)}\rangle$ for $n \neq i, j$ and $|s_{P_k(n)}\rangle \neq |s_{P_l(n)}\rangle$ for $n = i, j$, since σ_x and σ_y are off-diagonal. As a result, for $n \neq i, j$ the distinct permutations P_k and P_l coincide. Hence, both $(s_{P_k(i)}, s_{P_k(j)})$ and $(s_{P_l(i)}, s_{P_l(j)})$ have to be a permutation of $(0, 1)$. Moreover, fixing the permutation $(s_{P_k(i)}, s_{P_k(j)})$ determines the permutation $(s_{P_l(i)}, s_{P_l(j)})$, such that the matrix elements are non-zero. There are two permutations $(s_{P_k(i)}, s_{P_k(j)})$ of $(0, 1)$ and $\binom{N-2}{m-1}$ permutations to distribute the remaining states. We thus obtain

$$\langle D_{N,m} | \sigma_\alpha^{(i)} \sigma_\alpha^{(j)} | D_{N,m} \rangle = \binom{N}{m}^{-1} \times \binom{N-2}{m-1} \times 2 = \frac{2m(N-m)}{N(N-1)}. \quad (\text{A.59})$$

Similarly, we obtain for $\alpha = x, y$ and distinct i, j, k, l

$$\langle D_{N,m} | \sigma_\alpha^{(i)} \sigma_\alpha^{(j)} \sigma_\alpha^{(k)} \sigma_\alpha^{(l)} | D_{N,m} \rangle = \frac{6m(m-1)(N-m-1)(N-m)}{N(N-1)(N-2)(N-3)}. \quad (\text{A.60})$$

With these expressions we can determine the fourth moments for $\alpha = x, y$:

$$\begin{aligned}
\langle J_\alpha^4 \rangle &= \frac{1}{16} \sum_{i,j,k,l} \langle \sigma_\alpha^{(i)} \sigma_\alpha^{(j)} \sigma_\alpha^{(k)} \sigma_\alpha^{(l)} \rangle \\
&= \frac{1}{16} \left[\sum_{i \neq j \neq k \neq l} \langle \sigma_\alpha^{(i)} \sigma_\alpha^{(j)} \sigma_\alpha^{(k)} \sigma_\alpha^{(l)} \rangle + 6 \sum_{i=j \neq k \neq l} \langle (\sigma_\alpha^{(i)})^2 \sigma_\alpha^{(k)} \sigma_\alpha^{(l)} \rangle \right. \\
&\quad \left. + 3 \sum_{i=j \neq k=l} \langle (\sigma_\alpha^{(i)})^2 (\sigma_\alpha^{(k)})^2 \rangle + 4 \sum_{i=j=k \neq l} \langle (\sigma_\alpha^{(i)})^3 \sigma_\alpha^{(l)} \rangle + \sum_{i=j=k=l} \langle (\sigma_\alpha^{(i)})^4 \rangle \right] \quad (\text{A.61}) \\
&= \frac{1}{16} \left[\sum_{i \neq j \neq k \neq l} \langle \sigma_\alpha^{(i)} \sigma_\alpha^{(j)} \sigma_\alpha^{(k)} \sigma_\alpha^{(l)} \rangle + 2(3N-4) \sum_{i \neq j} \langle \sigma_\alpha^{(i)} \sigma_\alpha^{(j)} \rangle + 3N^2 - 2N \right] \\
&= \frac{1}{16} [N(3N-2) + 4(3N-4)m(N-m) + 6m(m-1)(N-m-1)(N-m)].
\end{aligned}$$

With these expressions in turn we obtain the variances

$$\begin{aligned}
\text{Var} \left((\hat{\xi}_c)_{\text{TS}} \right) &= \frac{N(N^3 + 4N^2 - 4N - 16)}{64K_{\text{TS}}}, \\
\text{Var} \left((\hat{\xi}_c)_{\text{AP1}} \right) &= \frac{N(K_{\text{AP1}}(2N^5 - 10N^4 + 21N^3 - 25N^2 + 16N - 4) - 2N^5 + 10N^4 - 19N^3 + 21N^2 - 12N + 4)}{32(K_{\text{AP1}} - 1)K_{\text{AP1}}(N-1)^2}, \\
\text{Var} \left((\hat{\xi}_c)_{\text{AP2}} \right) &= \frac{N(6N^4 - 20N^3 + 25N^2 - 16N + 4)}{32K_{\text{AP2}}(N-1)}, \\
\text{Var} \left((\hat{\xi}_c)_{\text{RP1}} \right) &= \frac{N^2(L_{\text{RP1}}(2N^4 - 8N^3 + 13N^2 - 12N + 4) + 4N^3 - 9N^2 + 12N - 4)}{32(L_{\text{RP1}} - 1)L_{\text{RP1}}}, \\
\text{Var} \left((\hat{\xi}_c)_{\text{RP2}} \right) &= \frac{N^2(6N^4 - 16N^3 + 17N^2 - 12N + 4)}{32K_{\text{RP2}}L_{\text{RP2}}}. \quad (\text{A.62})
\end{aligned}$$

B Additional calculations for Sec. 4

B.1 Unbiased estimators

In this Appendix we show that the estimators used in the main text are unbiased.

B.1.1 Estimator in the infinite measurement limit

First, we assume that the expectation values can be inferred directly, i.e., that we can repeat the measurement of the operator infinite times. In this case, the estimator is given by Eq. (4.5). The expectation value has to be calculated with respect to the random variables J_l . With $\mathbb{E}[\langle B_{J_l} \rangle] = \sum_{j=1}^M p(J_l = j) \langle B_j \rangle = \frac{1}{M} \sum_{j=1}^M \langle B_j \rangle$, we obtain

$$\mathbb{E}[\langle \hat{\mathcal{B}} \rangle_\infty] = \frac{M}{L} \sum_{l=1}^L \mathbb{E}[\langle B_{J_l} \rangle] = \frac{M}{L} \sum_{l=1}^L \frac{1}{M} \sum_{j=1}^M \langle B_j \rangle = \sum_{j=1}^M \langle B_j \rangle = \langle \mathcal{B} \rangle. \quad (\text{B.1})$$

B.1.2 Estimator for finite repetitions

To calculate the expectation value of the estimator in Eq. (4.8), we note that both the measurement outcomes b_j and the index J of the terms are random variables. Hence, the expectation value of the estimator has to be taken over both the measurement outcomes as well as the random picking, i.e., over J . To evaluate the expectation value, we can thus make use of the law of iterated expectation. That is,

$$\mathbb{E}[\dots] = \mathbb{E}_J \left\{ \mathbb{E}_{b_{J_l}} [\dots | J_l = J] \right\}. \quad (\text{B.2})$$

This results in

$$\begin{aligned} \mathbb{E}[\langle \hat{\mathcal{B}} \rangle] &= \frac{M}{KL} \sum_{l=1}^L \sum_{k=1}^K \mathbb{E}_J \left\{ \underbrace{\mathbb{E}_{b_{J_l}} [b_{J_l}^{(k)} | J_l = J]}_{=\langle B_J \rangle} \right\} \\ &= \frac{M}{KL} \sum_{l=1}^L \sum_{k=1}^K \mathbb{E}_J [\langle B_J \rangle] \\ &= \frac{M}{KL} \sum_{l=1}^L \sum_{k=1}^K \underbrace{\sum_{j=1}^M p(j) \langle B_j \rangle}_{=\sum_{j=1}^M \frac{1}{M} \langle B_j \rangle} \\ &= \sum_{j=1}^M \langle B_j \rangle = \langle \mathcal{B} \rangle. \end{aligned} \quad (\text{B.3})$$

B.2 Hoeffding's inequality

B.2.1 Estimator in the infinite measurement limit

The estimator in Eq. (4.5) can be written as a sum of random variables as follows:

$$\langle \hat{\mathcal{B}} \rangle_\infty = \sum_{l=1}^L \underbrace{\frac{M}{L} \langle B_{J_l} \rangle}_{=: X_l}. \quad (\text{B.4})$$

Since each term B_j in the Bell operator is a tensor product of Pauli operators, $\langle B_j \rangle \in [-1, 1]$ and thus $-\frac{M}{L} = a_l \leq X_l \leq b_l = \frac{M}{L}$. Moreover, the bounded random variables X_l are independent, as they are obtained from different experimental runs. We can thus use Hoeffding's inequality [85], which states that

$$\begin{aligned} \mathbb{P}(\langle \hat{\mathcal{B}} \rangle_\infty - \langle \mathcal{B} \rangle \geq t) &\leq \exp\left(-\frac{2t^2}{\sum_{l=1}^L (b_l - a_l)^2}\right) \\ &= \exp\left(-\frac{2t^2}{\sum_{l=1}^L \left(\frac{2M}{L}\right)^2}\right) \\ &= \exp\left(-\frac{t^2}{2M^2}L\right). \end{aligned} \tag{B.5}$$

B.2.2 Estimator for finite repetitions

As in the previous section, we can apply Hoeffding's inequality to the estimator in Eq. (4.8). Also the estimator in Eq. (4.8) is a sum of independent random variables,

$$\langle \hat{\mathcal{B}} \rangle = \sum_{l=1}^L \sum_{k=1}^K \underbrace{\frac{M}{KL} b_{J_l}^{(k)}}_{=: Y_{kl}}. \tag{B.6}$$

Since the outcomes $b_{J_l}^{(k)}$ are obtained from different experimental runs, they are independent and thus are the random variables Y_{kl} . In addition, the outcomes can only take the values $b_{J_l}^{(k)} \in \{-1, 1\}$. Therefore, we have that the random variables Y_{kl} are bounded as $-\frac{M}{KL} = a_{kl} \leq Y_{kl} \leq b_{kl} = \frac{M}{KL}$. Finally, we get from Hoeffding's inequality,

$$\begin{aligned} \mathbb{P}(\langle \hat{\mathcal{B}} \rangle - \langle \mathcal{B} \rangle \geq t) &\leq \exp\left(-\frac{2t^2}{\sum_{l=1}^L \sum_{k=1}^K (b_{kl} - a_{kl})^2}\right) \\ &= \exp\left(-\frac{2t^2}{\sum_{l=1}^L \sum_{k=1}^K \left(\frac{2M}{KL}\right)^2}\right) \\ &= \exp\left(-\frac{t^2}{2M^2}KL\right). \end{aligned} \tag{B.7}$$

B.3 Preparation scheme for the LC state

We discuss a scheme to prepare a LC state with all qubits of an n -qubit quantum computer. This can be done by preparing all qubits in the $|+\rangle$ state and then applying CZ gates between some of them. A problem can be that the connectivity of the quantum computer does not allow one to perform a specific gate between qubits i and j directly. The following lemma shows that this is not a fundamental problem.

Lemma B.1. *Consider a qubit array with a connected connectivity graph, where a CZ gate should be applied to two qubits for graph state generation from the state $|+\rangle^{\otimes n}$. This can be achieved by a sequence of CZ gates between adjacent qubits (in the sense of the connectivity graph) and local complementations.*

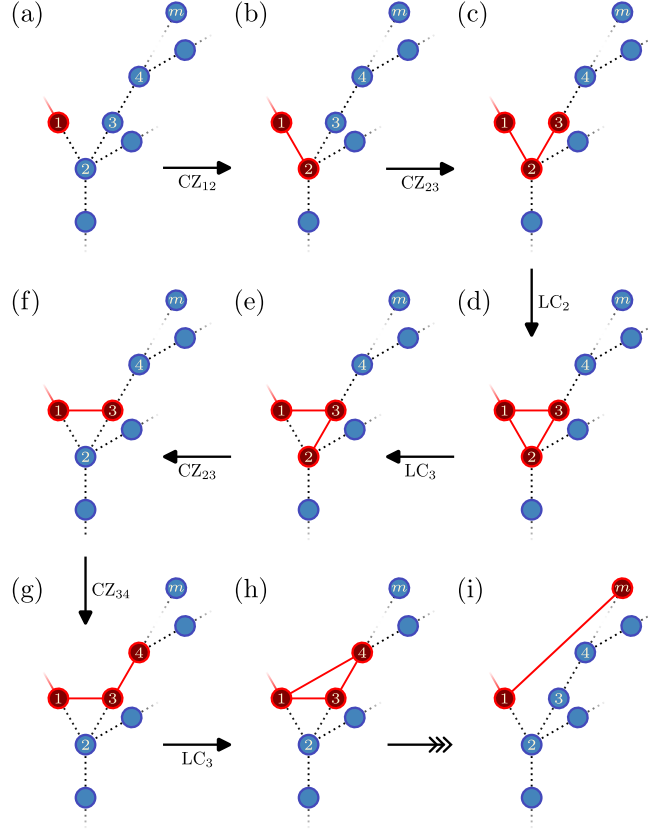


Figure B.1: Scheme to decompose the CZ gate between arbitrary qubits 1 and m into a sequence of CZ gates and local complementations. The dotted lines denote the CZ gates between adjacent qubits and local complementations. The red lines indicate the graph state. Qubit 1 can already be coupled to different qubits, whereas the qubits $2, \dots, m$ have to be uncoupled. The figure is reprinted from [P4].

Proof. We give an explicit construction that is visualized in Fig. B.1. The initial state is shown in Fig. B.1 (a). We would like to perform a CZ gate between qubits 1 and m , i.e., CZ_{1m} . The interaction topology, however, does not allow a direct coupling. Rather, the qubits 1 and m are connected by the path of qubits $1, 2, \dots, m$. Here, all the qubits $1, 2, \dots, m$ should be in the $|+\rangle$ state; in particular, it is important that no CZ gate has been applied yet to the qubits $2, \dots, m$. In this case, we can apply the following scheme:

- (1) Connect qubit 2 by performing the CZ_{12} gate to generate the first $\text{PAIR}(1, 2)$ [Fig. B.1 (b)].
- (2) While $k < m$, transform $\text{PAIR}(1, k) \rightarrow \text{PAIR}(1, k + 1)$ by the following:
 - (a) Couple qubit k and $k + 1$ by $CZ_{k, k+1}$ [Fig. B.1 (c)].
 - (b) Couple qubit 1 and $k+1$ by a local complementation on qubit k , i.e., LC_k [Fig. B.1 (d)].
 - (c) Cancel the CZ gate between qubits 1 and k by performing LC_{k+1} [Fig. B.1 (e)].
 - (d) Cancel the CZ gate between qubits k and $k + 1$ by $CZ_{k, k+1}$ [Fig. B.1 (f)].

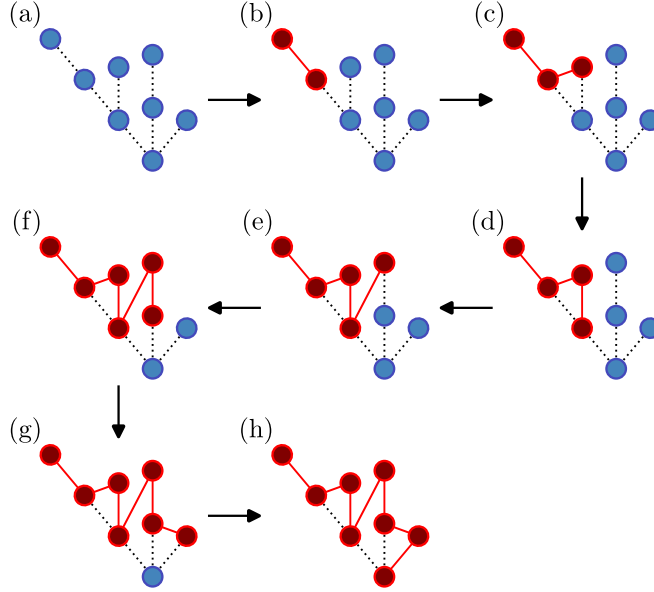


Figure B.2: Illustration of the scheme in Observation 4 for an exemplary two-qubit connectivity of eight qubits. The dotted lines denote the CZ gates that can be performed, whereas the red lines indicate the graph state. The figure is taken from [P4].

This allows one to decompose the CZ_{1m} gate into a sequence with circuit depth $3(m - 2) + 1$. We note that step (2) is only necessary for $m > 1$ and requires three steps as the local complementations in (b) and (c) can be combined. \square

Lemma B.1 can be used to construct a linear cluster state on an arbitrary interaction topology.

Observation 4. *On a quantum computer of n qubits, it is possible to prepare an n -qubit LC state with $\mathcal{O}(n)$ circuit depth.*

Proof. The connectivity of a quantum computer is a connected graph G . It thus has a spanning tree, i.e., a tree graph that covers all vertices of G . A tree graph in turn can be covered by a LC state by the following steps. At the start, all qubits are assumed to be prepared in the state $|+\rangle^{\otimes n}$.

- (1) We start at a leaf and successively couple the adjacent qubits in the direction of the root by CZ operations.
- (2) At a branch-off, check whether the other branch has already been covered. If all other branches have already been covered, we continue step (1) in the direction of the root. Otherwise, Lemma B.1 can be used to couple the last qubit to a leaf in the uncovered branch. From there, we can continue again with step (1).

The scheme is shown for an exemplary two-qubit connectivity in Fig. B.2. To investigate the circuit depth, we note that steps (1) and (2) are executed alternately. We thus count the number of steps for each run. k_i denotes the number of steps for the i th execution of step (1), whereas l_i stands for the steps required for the i th execution of step (2). In step (1), adjacent qubits are consecutively coupled by CZ gates. We thus have $\sum_i k_i < n$. In each step (2), a qubit at

distance m_i is coupled and, from Lemma B.1, we know that $l_i = 3(m_i - 2) + 1$. As each branch is only passed once, we have $\sum_i (m_i - 2) \leq n$. Moreover, there are less than n branch-offs, i.e., $\sum_i 1 \leq n$. Therefore, we obtain $\sum_i l_i = \sum_i 3(m_i - 2) + 1 \leq 4n$. The final circuit depth of the scheme is thus upper bounded by $\sum_i (k_i + l_i) \leq n + 4n = 5n$. \square

List of Publications

- [P1] H. Chau Nguyen, **Jan Lennart Bönsel**, Jonathan Steinberg, and Otfried Gühne. “Optimizing Shadow Tomography with Generalized Measurements”. *Phys. Rev. Lett.* 129 (2022), 220502. DOI: [10.1103/PhysRevLett.129.220502](https://doi.org/10.1103/PhysRevLett.129.220502).
- [P2] Nikolai Wyderka, Andreas Ketterer, Satoya Imai, **Jan Lennart Bönsel**, Daniel E. Jones, Brian T. Kirby, Xiao-Dong Yu, and Otfried Gühne. “Complete Characterization of Quantum Correlations by Randomized Measurements”. *Phys. Rev. Lett.* 131 (2023), 090201. DOI: [10.1103/PhysRevLett.131.090201](https://doi.org/10.1103/PhysRevLett.131.090201).
- [P3] **Jan Lennart Bönsel**, Satoya Imai, Ye-Chao Liu, and Otfried Gühne. “Error estimation of different schemes to measure spin-squeezing inequalities”. *Phys. Rev. A* 110 (2024), 022410. DOI: [10.1103/PhysRevA.110.022410](https://doi.org/10.1103/PhysRevA.110.022410).
- [P4] **Jan Lennart Bönsel**, Otfried Gühne, and Adán Cabello. “Generating multipartite non-locality to benchmark quantum computers”. *Phys. Rev. A* 111 (2025), 012207. DOI: [10.1103/PhysRevA.111.012207](https://doi.org/10.1103/PhysRevA.111.012207).

Bibliography

- [5] A. Einstein, B. Podolsky, and N. Rosen. “Can Quantum-Mechanical Description of Physical Reality Be Considered Complete?”. *Phys. Rev.* 47 (1935), 777–780. DOI: [10.1103/PhysRev.47.777](https://doi.org/10.1103/PhysRev.47.777).
- [6] J. S. Bell. “On the Einstein Podolsky Rosen paradox”. *Phys. Phys. Fiz.* 1 (1964), 195–200. DOI: [10.1103/PhysicsPhysiqueFizika.1.195](https://doi.org/10.1103/PhysicsPhysiqueFizika.1.195).
- [7] N. Brunner et al. “Bell nonlocality”. *Rev. Mod. Phys.* 86 (2014), 419–478. DOI: [10.1103/RevModPhys.86.419](https://doi.org/10.1103/RevModPhys.86.419).
- [8] E. Schrödinger. “Die gegenwärtige Situation in der Quantenmechanik”. *Naturwissenschaften* 23 (1935), 823–828. DOI: [10.1007/BF01491914](https://doi.org/10.1007/BF01491914).
- [9] M. Kitagawa and M. Ueda. “Squeezed spin states”. *Phys. Rev. A* 47 (1993), 5138–5143. DOI: [10.1103/PhysRevA.47.5138](https://doi.org/10.1103/PhysRevA.47.5138).
- [10] D. J. Wineland et al. “Spin squeezing and reduced quantum noise in spectroscopy”. *Phys. Rev. A* 46 (1992), R6797–R6800. DOI: [10.1103/PhysRevA.46.R6797](https://doi.org/10.1103/PhysRevA.46.R6797).
- [11] L. Pezzé and A. Smerzi. “Entanglement, Nonlinear Dynamics, and the Heisenberg Limit”. *Phys. Rev. Lett.* 102 (2009), 100401. DOI: [10.1103/PhysRevLett.102.100401](https://doi.org/10.1103/PhysRevLett.102.100401).
- [12] C. H. Bennett and G. Brassard. “Quantum cryptography: Public key distribution and coin tossing”. *Theor. Comput. Sci.* 560 (2014), 7–11. DOI: <https://doi.org/10.1016/j.tcs.2014.05.025>.
- [13] A. K. Ekert. “Quantum cryptography based on Bell’s theorem”. *Phys. Rev. Lett.* 67 (1991), 661–663. DOI: [10.1103/PhysRevLett.67.661](https://doi.org/10.1103/PhysRevLett.67.661).
- [14] P. W. Shor. “Polynomial-Time Algorithms for Prime Factorization and Discrete Logarithms on a Quantum Computer”. *SIAM J. Comput.* 26 (1997), 1484–1509. DOI: [10.1137/S0097539795293172](https://doi.org/10.1137/S0097539795293172).
- [15] A. Ekert and R. Jozsa. “Quantum computation and Shor’s factoring algorithm”. *Rev. Mod. Phys.* 68 (1996), 733–753. DOI: [10.1103/RevModPhys.68.733](https://doi.org/10.1103/RevModPhys.68.733).
- [16] A. Anshu and S. Arunachalam. “A survey on the complexity of learning quantum states”. *Nat. Rev. Phys.* 6 (2023), 59–69. DOI: [10.1038/s42254-023-00662-4](https://doi.org/10.1038/s42254-023-00662-4).

- [17] O. Gühne and G. Tóth. “Entanglement detection”. *Phys. Rep.* 474 (2009), 1–75. DOI: [10.1016/j.physrep.2009.02.004](https://doi.org/10.1016/j.physrep.2009.02.004).
- [18] S. T. Flammia and Y.-K. Liu. “Direct Fidelity Estimation from Few Pauli Measurements”. *Phys. Rev. Lett.* 106 (2011), 230501. DOI: [10.1103/PhysRevLett.106.230501](https://doi.org/10.1103/PhysRevLett.106.230501).
- [19] S. Cao et al. “Generation of genuine entanglement up to 51 superconducting qubits”. *Nature* 619 (2023), 738–742. DOI: [10.1038/s41586-023-06195-1](https://doi.org/10.1038/s41586-023-06195-1).
- [20] A. Dimić and B. Dakić. “Single-copy entanglement detection”. *npj Quantum Inf.* 4 (2018), 11. DOI: [10.1038/s41534-017-0055-x](https://doi.org/10.1038/s41534-017-0055-x).
- [21] V. Saggio et al. “Experimental few-copy multipartite entanglement detection”. *Nat. Phys.* 15 (2019), 935–940. DOI: [10.1038/s41567-019-0550-4](https://doi.org/10.1038/s41567-019-0550-4).
- [22] M. C. Tran et al. “Quantum entanglement from random measurements”. *Phys. Rev. A* 92 (2015), 050301. DOI: [10.1103/PhysRevA.92.050301](https://doi.org/10.1103/PhysRevA.92.050301).
- [23] A. Ketterer, N. Wyderka, and O. Gühne. “Characterizing Multipartite Entanglement with Moments of Random Correlations”. *Phys. Rev. Lett.* 122 (2019), 120505. DOI: [10.1103/PhysRevLett.122.120505](https://doi.org/10.1103/PhysRevLett.122.120505).
- [24] S. Imai et al. “Bound Entanglement from Randomized Measurements”. *Phys. Rev. Lett.* 126 (2021), 150501. DOI: [10.1103/PhysRevLett.126.150501](https://doi.org/10.1103/PhysRevLett.126.150501).
- [25] P. Cieřliński et al. “Analysing quantum systems with randomised measurements”. *Phys. Rep.* 1095 (2024), 1–48. DOI: [10.1016/j.physrep.2024.09.009](https://doi.org/10.1016/j.physrep.2024.09.009).
- [26] H.-Y. Huang, R. Kueng, and J. Preskill. “Predicting many properties of a quantum system from very few measurements”. *Nat. Phys.* 16 (2020), 1050–1057. DOI: [10.1038/s41567-020-0932-7](https://doi.org/10.1038/s41567-020-0932-7).
- [27] T. Heinosaari and M. Ziman. *The Mathematical Language of Quantum Theory: From Uncertainty to Entanglement*. Cambridge University Press, 2011. DOI: [10.1017/CB09781139031103](https://doi.org/10.1017/CB09781139031103).
- [28] J. Watrous. *The Theory of Quantum Information*. Cambridge University Press, 2018. DOI: [10.1017/9781316848142](https://doi.org/10.1017/9781316848142).
- [29] M. Hein et al. “Entanglement in graph states and its applications”. *Proc. Int. Sch. Phys. "Enrico Fermi"* 162 (2006), 115–218. DOI: [10.3254/978-1-61499-018-5-115](https://doi.org/10.3254/978-1-61499-018-5-115).
- [30] I. Bengtsson and K. Życzkowski. *Geometry of Quantum States: An Introduction to Quantum Entanglement*. 2nd ed. Cambridge University Press, 2017. DOI: [10.1017/CB09780511535048](https://doi.org/10.1017/CB09780511535048).
- [31] M. Oszmaniec et al. “Simulating Positive-Operator-Valued Measures with Projective Measurements”. *Phys. Rev. Lett.* 119 (2017), 190501. DOI: [10.1103/PhysRevLett.119.190501](https://doi.org/10.1103/PhysRevLett.119.190501).
- [32] A. Peres. “Separability Criterion for Density Matrices”. *Phys. Rev. Lett.* 77 (1996), 1413–1415. DOI: [10.1103/PhysRevLett.77.1413](https://doi.org/10.1103/PhysRevLett.77.1413).
- [33] M. Horodecki, P. Horodecki, and R. Horodecki. “Separability of mixed states: necessary and sufficient conditions”. *Phys. Lett. A* 223 (1996), 1–8. DOI: [https://doi.org/10.1016/S0375-9601\(96\)00706-2](https://doi.org/10.1016/S0375-9601(96)00706-2).
- [34] K. Chen and L. Wu. “A matrix realignment method for recognizing entanglement”. *Quantum Inf. Comput.* 3 (2003), 193–202. DOI: [10.26421/QIC3.3-1](https://doi.org/10.26421/QIC3.3-1).
- [35] O. Rudolph. “Further Results on the Cross Norm Criterion for Separability”. *Quantum Inf. Process.* 4 (2005), 219–239. DOI: [10.1007/s11128-005-5664-1](https://doi.org/10.1007/s11128-005-5664-1).

- [36] G. Tóth and O. Gühne. “Detecting Genuine Multipartite Entanglement with Two Local Measurements”. *Phys. Rev. Lett.* 94 (2005), 060501. DOI: [10.1103/PhysRevLett.94.060501](https://doi.org/10.1103/PhysRevLett.94.060501).
- [37] Y. Zhou et al. “Detecting multipartite entanglement structure with minimal resources”. *npj Quantum Inf.* 5 (2019), 83. DOI: [10.1038/s41534-019-0200-9](https://doi.org/10.1038/s41534-019-0200-9).
- [38] R. Zander and C. K.-U. Becker. “Benchmarking Multipartite Entanglement Generation with Graph States”. *Adv. Quantum Technol.* 8 (2025), 2400239. DOI: [10.1002/qute.202400239](https://doi.org/10.1002/qute.202400239).
- [39] D. J. Wineland et al. “Squeezed atomic states and projection noise in spectroscopy”. *Phys. Rev. A* 50 (1994), 67–88. DOI: [10.1103/PhysRevA.50.67](https://doi.org/10.1103/PhysRevA.50.67).
- [40] E. S. Polzik. “The squeeze goes on”. *Nature* 453 (2008), 45–46. DOI: [10.1038/453045a](https://doi.org/10.1038/453045a).
- [41] G. Tóth and I. Apellaniz. “Quantum metrology from a quantum information science perspective”. *J. Phys. A: Math. Theor.* 47 (2014), 424006. DOI: [10.1088/1751-8113/47/42/424006](https://doi.org/10.1088/1751-8113/47/42/424006).
- [42] J. Ma et al. “Quantum spin squeezing”. *Phys. Rep.* 509 (2011), 89–165. DOI: [10.1016/j.physrep.2011.08.003](https://doi.org/10.1016/j.physrep.2011.08.003).
- [43] D. F. Walls. “Squeezed states of light”. *Nature* 306 (1983), 141–146. DOI: [10.1038/306141a0](https://doi.org/10.1038/306141a0).
- [44] W. Wasilewski et al. “Quantum Noise Limited and Entanglement-Assisted Magnetometry”. *Phys. Rev. Lett.* 104 (2010), 133601. DOI: [10.1103/PhysRevLett.104.133601](https://doi.org/10.1103/PhysRevLett.104.133601).
- [45] T. Fernholz et al. “Spin Squeezing of Atomic Ensembles via Nuclear-Electronic Spin Entanglement”. *Phys. Rev. Lett.* 101 (2008), 073601. DOI: [10.1103/PhysRevLett.101.073601](https://doi.org/10.1103/PhysRevLett.101.073601).
- [46] K. Hammerer, A. S. Sørensen, and E. S. Polzik. “Quantum interface between light and atomic ensembles”. *Rev. Mod. Phys.* 82 (2010), 1041–1093. DOI: [10.1103/RevModPhys.82.1041](https://doi.org/10.1103/RevModPhys.82.1041).
- [47] M. F. Riedel et al. “Atom-chip-based generation of entanglement for quantum metrology”. *Nature* 464 (2010), 1170–1173. DOI: [10.1038/nature08988](https://doi.org/10.1038/nature08988).
- [48] C. F. Ockeloen et al. “Quantum Metrology with a Scanning Probe Atom Interferometer”. *Phys. Rev. Lett.* 111 (2013), 143001. DOI: [10.1103/PhysRevLett.111.143001](https://doi.org/10.1103/PhysRevLett.111.143001).
- [49] W. Muessel et al. “Scalable Spin Squeezing for Quantum-Enhanced Magnetometry with Bose-Einstein Condensates”. *Phys. Rev. Lett.* 113 (2014), 103004. DOI: [10.1103/PhysRevLett.113.103004](https://doi.org/10.1103/PhysRevLett.113.103004).
- [50] J. A. Hines et al. “Spin Squeezing by Rydberg Dressing in an Array of Atomic Ensembles”. *Phys. Rev. Lett.* 131 (2023), 063401. DOI: [10.1103/PhysRevLett.131.063401](https://doi.org/10.1103/PhysRevLett.131.063401).
- [51] C. Gross. “Spin squeezing, entanglement and quantum metrology with Bose-Einstein condensates”. *J. Phys. B: At. Mol. Opt. Phys.* 45 (2012), 103001. DOI: [10.1088/0953-4075/45/10/103001](https://doi.org/10.1088/0953-4075/45/10/103001).
- [52] A. Sørensen et al. “Many-particle entanglement with Bose-Einstein condensates”. *Nature* 409 (2001), 63–66. DOI: [10.1038/35051038](https://doi.org/10.1038/35051038).
- [53] G. Tóth. “Entanglement detection in optical lattices of bosonic atoms with collective measurements”. *Phys. Rev. A* 69 (2004), 052327. DOI: [10.1103/PhysRevA.69.052327](https://doi.org/10.1103/PhysRevA.69.052327).

- [54] J. K. Korbicz, J. I. Cirac, and M. Lewenstein. “Spin Squeezing Inequalities and Entanglement of N Qubit States”. *Phys. Rev. Lett.* 95 (2005), 120502. DOI: [10.1103/PhysRevLett.95.120502](https://doi.org/10.1103/PhysRevLett.95.120502).
- [55] G. Tóth. “Detection of multipartite entanglement in the vicinity of symmetric Dicke states”. *J. Opt. Soc. Am. B* 24 (2007), 275–282. DOI: [10.1364/josab.24.000275](https://doi.org/10.1364/josab.24.000275).
- [56] G. Tóth et al. “Optimal Spin Squeezing Inequalities Detect Bound Entanglement in Spin Models”. *Phys. Rev. Lett.* 99 (2007), 250405. DOI: [10.1103/PhysRevLett.99.250405](https://doi.org/10.1103/PhysRevLett.99.250405).
- [57] G. Tóth et al. “Spin squeezing and entanglement”. *Phys. Rev. A* 79 (2009), 042334. DOI: [10.1103/PhysRevA.79.042334](https://doi.org/10.1103/PhysRevA.79.042334).
- [58] J. K. Korbicz et al. “Generalized spin-squeezing inequalities in N -qubit systems: Theory and experiment”. *Phys. Rev. A* 74 (2006), 052319. DOI: [10.1103/PhysRevA.74.052319](https://doi.org/10.1103/PhysRevA.74.052319).
- [59] B. Julsgaard, A. Kozhekin, and E. S. Polzik. “Experimental long-lived entanglement of two macroscopic objects”. *Nature* 413 (2001), 400–403. DOI: [10.1038/35096524](https://doi.org/10.1038/35096524).
- [60] J. Estève et al. “Squeezing and entanglement in a Bose-Einstein condensate”. *Nature* 455 (2008), 1216–1219. DOI: [10.1038/nature07332](https://doi.org/10.1038/nature07332).
- [61] M. Fadel et al. “Entanglement Quantification in Atomic Ensembles”. *Phys. Rev. Lett.* 127 (2021), 010401. DOI: [10.1103/PhysRevLett.127.010401](https://doi.org/10.1103/PhysRevLett.127.010401).
- [62] L. Dellantonio et al. “Multipartite entanglement detection with nonsymmetric probing”. *Phys. Rev. A* 95 (2017), 040301(R). DOI: [10.1103/PhysRevA.95.040301](https://doi.org/10.1103/PhysRevA.95.040301).
- [63] N. Friis et al. “Entanglement certification from theory to experiment”. *Nat. Rev. Phys.* 1 (2019), 72–87. DOI: [10.1038/s42254-018-0003-5](https://doi.org/10.1038/s42254-018-0003-5).
- [64] W. van Dam, R. Gill, and P. Grunwald. “The Statistical Strength of Nonlocality Proofs”. *IEEE Trans. Inf. Theory* 51 (2005), 2812–2835. DOI: [10.1109/TIT.2005.851738](https://doi.org/10.1109/TIT.2005.851738).
- [65] N. D. Mermin. “Extreme quantum entanglement in a superposition of macroscopically distinct states”. *Phys. Rev. Lett.* 65 (1990), 1838–1840. DOI: [10.1103/PhysRevLett.65.1838](https://doi.org/10.1103/PhysRevLett.65.1838).
- [66] O. Gühne et al. “Bell Inequalities for Graph States”. *Phys. Rev. Lett.* 95 (2005), 120405. DOI: [10.1103/PhysRevLett.95.120405](https://doi.org/10.1103/PhysRevLett.95.120405).
- [67] A. Cabello, O. Gühne, and D. Rodríguez. “Mermin inequalities for perfect correlations”. *Phys. Rev. A* 77 (2008), 062106. DOI: [10.1103/PhysRevA.77.062106](https://doi.org/10.1103/PhysRevA.77.062106).
- [68] V. Scarani et al. “Nonlocality of cluster states of qubits”. *Phys. Rev. A* 71 (2005), 042325. DOI: [10.1103/PhysRevA.71.042325](https://doi.org/10.1103/PhysRevA.71.042325).
- [69] A. Cabello. “Stronger Two-Observer All-Versus-Nothing Violation of Local Realism”. *Phys. Rev. Lett.* 95 (2005), 210401. DOI: [10.1103/PhysRevLett.95.210401](https://doi.org/10.1103/PhysRevLett.95.210401).
- [70] G. Tóth, O. Gühne, and H. J. Briegel. “Two-setting Bell inequalities for graph states”. *Phys. Rev. A* 73 (2006), 022303. DOI: [10.1103/PhysRevA.73.022303](https://doi.org/10.1103/PhysRevA.73.022303).
- [71] O. Gühne and A. Cabello. “Generalized Ardehali-Bell inequalities for graph states”. *Phys. Rev. A* 77 (2008), 032108. DOI: [10.1103/PhysRevA.77.032108](https://doi.org/10.1103/PhysRevA.77.032108).
- [72] G. Svetlichny. “Distinguishing three-body from two-body nonseparability by a Bell-type inequality”. *Phys. Rev. D* 35 (1987), 3066–3069. DOI: [10.1103/PhysRevD.35.3066](https://doi.org/10.1103/PhysRevD.35.3066).
- [73] A. Elben et al. “The randomized measurement toolbox”. *Nat. Rev. Phys.* 5 (2023), 9–24. DOI: [10.1038/s42254-022-00535-2](https://doi.org/10.1038/s42254-022-00535-2).

- [74] S. Chen et al. “Robust Shadow Estimation”. *PRX Quantum* 2 (2021), 030348. DOI: [10.1103/PRXQuantum.2.030348](https://doi.org/10.1103/PRXQuantum.2.030348).
- [75] D. E. Koh and S. Grewal. “Classical Shadows With Noise”. *Quantum* 6 (2022), 776. DOI: [10.22331/q-2022-08-16-776](https://doi.org/10.22331/q-2022-08-16-776).
- [76] M. A. Nielsen and I. L. Chuang. *Quantum Computation and Quantum Information: 10th Anniversary Edition*. Cambridge University Press, 2010. DOI: [10.1017/CB09780511976667](https://doi.org/10.1017/CB09780511976667).
- [77] Y. Makhlin. “Nonlocal properties of two-qubit gates and mixed states, and the optimization of quantum computations”. *Quantum Inf. Process.* 1 (2002), 243–252. DOI: [10.1023/A:1022144002391](https://doi.org/10.1023/A:1022144002391).
- [78] S. Aaronson. “Shadow tomography of quantum states”. *Proceedings of the 50th Annual ACM SIGACT Symposium on Theory of Computing*. STOC 2018. Los Angeles, CA, USA: Association for Computing Machinery, 2018, 325–338. DOI: [10.1145/3188745.3188802](https://doi.org/10.1145/3188745.3188802).
- [79] D. Gottesman. “Stabilizer Codes and Quantum Error Correction”. PhD Thesis. California Institute of Technology, 1997. DOI: [10.48550/arXiv.quant-ph/9705052](https://doi.org/10.48550/arXiv.quant-ph/9705052).
- [80] A. Nemirovsky and D. Yudin. *Problem Complexity and Method Efficiency in Optimization*. John Wiley & Sons Ltd, 1983.
- [81] L. Wasserman. *All of Statistics: A Concise Course in Statistical Inference*. Springer Texts in Statistics. Springer New York, 2004. DOI: [10.1007/978-0-387-21736-9](https://doi.org/10.1007/978-0-387-21736-9).
- [82] H.-P. Scheffler. “Stochastik I”. Lecture notes. 2002. URL: https://www.uni-siegen.de/fb6/src/scheffler/lehre/stochastik1_20021128.pdf.
- [83] B. O’Neill. “Some Useful Moment Results in Sampling Problems”. *Am. Stat.* 68 (2014), 282–296. DOI: [10.1080/00031305.2014.966589](https://doi.org/10.1080/00031305.2014.966589).
- [84] B. K. Ghosh. “Probability Inequalities Related to Markov’s Theorem”. *Am. Stat.* 56 (2002), 186–190. DOI: [10.1198/000313002119](https://doi.org/10.1198/000313002119).
- [85] W. Hoeffding. “Probability Inequalities for Sums of Bounded Random Variables”. *J. Am. Stat. Assoc.* 58 (1963), 13–30. DOI: [10.2307/2282952](https://doi.org/10.2307/2282952).
- [86] R. Schmied et al. “Bell correlations in a Bose-Einstein condensate”. *Science* 352 (2016), 441–444. DOI: [10.1126/science.aad8665](https://doi.org/10.1126/science.aad8665).
- [87] J. Kong et al. “Measurement-induced, spatially-extended entanglement in a hot, strongly-interacting atomic system”. *Nat. Commun.* 11 (2020), 2415. DOI: [10.1038/s41467-020-15899-1](https://doi.org/10.1038/s41467-020-15899-1).
- [88] K. Mouloudakis et al. “Interspecies spin-noise correlations in hot atomic vapors”. *Phys. Rev. A* 108 (2023), 052822. DOI: [10.1103/PhysRevA.108.052822](https://doi.org/10.1103/PhysRevA.108.052822).
- [89] C. Piltz et al. “A trapped-ion-based quantum byte with 10⁻⁵ next-neighbour cross-talk”. *Nat. Commun.* 5 (2014), 4679. DOI: [10.1038/ncomms5679](https://doi.org/10.1038/ncomms5679).
- [90] C. D. Bruzewicz et al. “Trapped-ion quantum computing: Progress and challenges”. *Appl. Phys. Rev.* 6 (2019), 021314. DOI: [10.1063/1.5088164](https://doi.org/10.1063/1.5088164).
- [91] F. Arute et al. “Quantum supremacy using a programmable superconducting processor”. *Nature* 574 (2019), 505–510. DOI: [10.1038/s41586-019-1666-5](https://doi.org/10.1038/s41586-019-1666-5).
- [92] M. AbuGhanem and H. Eleuch. “Full quantum tomography study of Google’s Sycamore gate on IBM’s quantum computers”. *EPJ Quantum Technol.* 11 (2024), 36. DOI: [10.1140/epjqt/s40507-024-00248-8](https://doi.org/10.1140/epjqt/s40507-024-00248-8).

- [93] Q.-X. Mei et al. “Experimental Realization of the Rabi-Hubbard Model with Trapped Ions”. *Phys. Rev. Lett.* 128 (2022), 160504. DOI: [10.1103/PhysRevLett.128.160504](https://doi.org/10.1103/PhysRevLett.128.160504).
- [94] X.-D. Yu, J. Shang, and O. Gühne. “Statistical Methods for Quantum State Verification and Fidelity Estimation”. *Adv. Quantum Technol.* 5 (2022), 2100126. DOI: <https://doi.org/10.1002/qute.202100126>.
- [95] J.-D. Bancal et al. “Definitions of multipartite nonlocality”. *Phys. Rev. A* 88 (2013), 014102. DOI: [10.1103/PhysRevA.88.014102](https://doi.org/10.1103/PhysRevA.88.014102).
- [96] C. Vieira, R. Ramanathan, and A. Cabello. “Test of the physical significance of Bell nonlocality”. 2024. DOI: [10.48550/arXiv.2402.00801](https://doi.org/10.48550/arXiv.2402.00801).
- [97] M.-C. Chen et al. “Ruling Out Real-Valued Standard Formalism of Quantum Theory”. *Phys. Rev. Lett.* 128 (2022), 040403. DOI: [10.1103/PhysRevLett.128.040403](https://doi.org/10.1103/PhysRevLett.128.040403).
- [98] D. Jafferis et al. “Traversable wormhole dynamics on a quantum processor”. *Nature* 612 (2022), 51–55. DOI: [10.1038/s41586-022-05424-3](https://doi.org/10.1038/s41586-022-05424-3).
- [99] R. F. Werner and M. M. Wolf. “All-multipartite Bell-correlation inequalities for two dichotomic observables per site”. *Phys. Rev. A* 64 (2001), 032112. DOI: [10.1103/PhysRevA.64.032112](https://doi.org/10.1103/PhysRevA.64.032112).
- [100] J. Eisert et al. “Quantum certification and benchmarking”. *Nat. Rev. Phys.* 2 (2020), 382–390. DOI: [10.1038/s42254-020-0186-4](https://doi.org/10.1038/s42254-020-0186-4).
- [101] J. Frank et al. “Heuristic-free Verification-inspired Quantum Benchmarking”. 2024. DOI: [10.48550/arXiv.2404.10739](https://doi.org/10.48550/arXiv.2404.10739).
- [102] I. Šupić and J. Bowles. “Self-testing of quantum systems: a review”. *Quantum* 4 (2020), 337. DOI: [10.22331/q-2020-09-30-337](https://doi.org/10.22331/q-2020-09-30-337).
- [103] S. Popescu and D. Rohrlich. “Generic quantum nonlocality”. *Phys. Lett. A* 166 (1992), 293–297. DOI: [10.1016/0375-9601\(92\)90711-T](https://doi.org/10.1016/0375-9601(92)90711-T).
- [104] B. P. Lanyon et al. “Experimental Violation of Multipartite Bell Inequalities with Trapped Ions”. *Phys. Rev. Lett.* 112 (2014), 100403. DOI: [10.1103/PhysRevLett.112.100403](https://doi.org/10.1103/PhysRevLett.112.100403).
- [105] C. Zhang et al. “Experimental Greenberger-Horne-Zeilinger-Type Six-Photon Quantum Nonlocality”. *Phys. Rev. Lett.* 115 (2015), 260402. DOI: [10.1103/PhysRevLett.115.260402](https://doi.org/10.1103/PhysRevLett.115.260402).
- [106] S. Pelisson, L. Pezzè, and A. Smerzi. “Nonlocality with ultracold atoms in a lattice”. *Phys. Rev. A* 93 (2016), 022115. DOI: [10.1103/PhysRevA.93.022115](https://doi.org/10.1103/PhysRevA.93.022115).
- [107] D. Alsina and J. I. Latorre. “Experimental test of Mermin inequalities on a five-qubit quantum computer”. *Phys. Rev. A* 94 (2016), 012314. DOI: [10.1103/PhysRevA.94.012314](https://doi.org/10.1103/PhysRevA.94.012314).
- [108] M. Swain et al. “Experimental demonstration of the violations of Mermin’s and Svetlichny’s inequalities for W and GHZ states”. *Quantum Inf. Process.* 18 (2019), 218. DOI: [10.1007/s11128-019-2331-5](https://doi.org/10.1007/s11128-019-2331-5).
- [109] D. González, D. F. de la Pradilla, and G. González. “Revisiting the Experimental Test of Mermin’s Inequalities at IBMQ”. *Int. J. Theor. Phys.* 59 (2020), 3756–3768. DOI: [10.1007/s10773-020-04629-4](https://doi.org/10.1007/s10773-020-04629-4).
- [110] E. Bäumer, N. Gisin, and A. Tavakoli. “Demonstrating the power of quantum computers, certification of highly entangled measurements and scalable quantum nonlocality”. *npj Quantum Inf.* 7 (2021), 117. DOI: [10.1038/s41534-021-00450-x](https://doi.org/10.1038/s41534-021-00450-x).

- [111] G. Amouzou et al. “Entanglement and nonlocality of four-qubit connected hypergraph states”. *Int. J. Quantum Inf.* 20 (2022), 2250001. DOI: [10.1142/S0219749922500010](https://doi.org/10.1142/S0219749922500010).
- [112] B. Yang et al. “Testing Scalable Bell Inequalities for Quantum Graph States on IBM Quantum Devices”. *IEEE J. Emerg. Sel. Top. Circuits Syst.* 12 (2022), 638–647. DOI: [10.1109/JETCAS.2022.3201730](https://doi.org/10.1109/JETCAS.2022.3201730).
- [113] H. de Boutray et al. “Mermin polynomials for non-locality and entanglement detection in Grover’s algorithm and Quantum Fourier Transform”. *Quantum Inf. Process.* 20 (2021), 91. DOI: [10.1007/s11128-020-02976-z](https://doi.org/10.1007/s11128-020-02976-z).
- [114] D. Singh et al. “Experimental construction of a symmetric three-qubit entangled state and its utility in testing the violation of a Bell inequality on an NMR quantum simulator”. *EPL* 140 (2022), 68001. DOI: [10.1209/0295-5075/acab7e](https://doi.org/10.1209/0295-5075/acab7e).
- [115] N. J. Engelsen et al. “Bell Correlations in Spin-Squeezed States of 500 000 Atoms”. *Phys. Rev. Lett.* 118 (2017), 140401. DOI: [10.1103/PhysRevLett.118.140401](https://doi.org/10.1103/PhysRevLett.118.140401).
- [116] D. M. Greenberger, M. A. Horne, and A. Zeilinger. “Going Beyond Bell’s Theorem”. *Bell’s Theorem, Quantum Theory and Conceptions of the Universe*. Ed. by M. Kafatos. Dordrecht: Springer Netherlands, 1989, 69–72. DOI: [10.1007/978-94-017-0849-4_10](https://doi.org/10.1007/978-94-017-0849-4_10).
- [117] W. Dür and H.-J. Briegel. “Stability of Macroscopic Entanglement under Decoherence”. *Phys. Rev. Lett.* 92 (2004), 180403. DOI: [10.1103/PhysRevLett.92.180403](https://doi.org/10.1103/PhysRevLett.92.180403).
- [118] M. Ardehali. “Bell inequalities with a magnitude of violation that grows exponentially with the number of particles”. *Phys. Rev. A* 46 (1992), 5375–5378. DOI: [10.1103/PhysRevA.46.5375](https://doi.org/10.1103/PhysRevA.46.5375).
- [119] V. Saggio et al. “Experimental few-copy multipartite entanglement detection”. *Nat. Phys.* 15 (2019), 935–940. DOI: [10.1038/s41567-019-0550-4](https://doi.org/10.1038/s41567-019-0550-4).
- [120] A. Peres. “Bayesian Analysis of Bell Inequalities”. *Fortschr. Phys.* 48 (2000), 531–535. DOI: [10.1002/\(SICI\)1521-3978\(200005\)48:5/7<531::AID-PROP531>3.0.CO;2-%23](https://doi.org/10.1002/(SICI)1521-3978(200005)48:5/7<531::AID-PROP531>3.0.CO;2-%23).
- [121] Y. Zhang, E. Knill, and S. Glancy. “Statistical strength of experiments to reject local realism with photon pairs and inefficient detectors”. *Phys. Rev. A* 81 (2010), 032117. DOI: [10.1103/PhysRevA.81.032117](https://doi.org/10.1103/PhysRevA.81.032117).
- [122] D. Elkouss and S. Wehner. “(Nearly) optimal P values for all Bell inequalities”. *npj Quantum Inf.* 2 (2016), 16026. DOI: [10.1038/npjqi.2016.26](https://doi.org/10.1038/npjqi.2016.26).
- [123] QuTech. “Quantum Inspire Starmon-5 Fact Sheet”. [Accessed: 17.11.2023]. 2020. URL: <https://qutech.nl/wp-content/uploads/2020/04/3.-Technical-Fact-Sheet-Quantum-Inspire-Starmon-5.pdf>.
- [124] IBM Quantum. [Accessed: 17.11.2023]. 2021. URL: <https://quantum.ibm.com/>.
- [125] N. Friis et al. “Observation of Entangled States of a Fully Controlled 20-Qubit System”. *Phys. Rev. X* 8 (2018), 021012. DOI: [10.1103/PhysRevX.8.021012](https://doi.org/10.1103/PhysRevX.8.021012).
- [126] D. Cruz et al. “Efficient Quantum Algorithms for GHZ and W States, and Implementation on the IBM Quantum Computer”. *Adv. Quantum Technol.* 2 (2019), 1900015. DOI: <https://doi.org/10.1002/qute.201900015>.
- [127] N. Yu and T.-C. Wei. “Learning marginals suffices!” 2023. DOI: [10.48550/arXiv.2303.08938](https://doi.org/10.48550/arXiv.2303.08938).
- [128] A. Schrijver. *Combinatorial Optimization: Polyhedra and Efficiency*. Algorithms and Combinatorics 24. Springer Berlin, Heidelberg, 2002. URL: <https://link.springer.com/book/9783540443896>.

- [129] F. B. Maciejewski, Z. Zimborás, and M. Oszmaniec. “Mitigation of readout noise in near-term quantum devices by classical post-processing based on detector tomography”. *Quantum* 4 (2020), 257. DOI: [10.22331/q-2020-04-24-257](https://doi.org/10.22331/q-2020-04-24-257).
- [130] Z. Cai et al. “Quantum error mitigation”. *Rev. Mod. Phys.* 95 (2023), 045005. DOI: [10.1103/RevModPhys.95.045005](https://doi.org/10.1103/RevModPhys.95.045005).
- [131] A. Javadi-Abhari et al. “Quantum computing with Qiskit”. 2024. DOI: [10.48550/arXiv.2405.08810](https://doi.org/10.48550/arXiv.2405.08810).
- [132] N. Moll et al. “Quantum optimization using variational algorithms on near-term quantum devices”. *Quantum Sci. Technol.* 3 (2018), 030503. DOI: [10.1088/2058-9565/aab822](https://doi.org/10.1088/2058-9565/aab822).
- [133] C. H. Baldwin et al. “Re-examining the quantum volume test: Ideal distributions, compiler optimizations, confidence intervals, and scalable resource estimations”. *Quantum* 6 (2022), 707. DOI: [10.22331/q-2022-05-09-707](https://doi.org/10.22331/q-2022-05-09-707).
- [134] D. C. McKay et al. “Benchmarking Quantum Processor Performance at Scale”. 2023. DOI: [10.48550/arXiv.2311.05933](https://doi.org/10.48550/arXiv.2311.05933).
- [135] L. Knips et al. “Multipartite entanglement analysis from random correlations”. *npj Quantum Inf.* 6 (2020), 51. DOI: [10.1038/s41534-020-0281-5](https://doi.org/10.1038/s41534-020-0281-5).
- [136] L. Knips. “A Moment for Random Measurements”. *Quantum Views* 4 (2020), 47. DOI: [10.22331/qv-2020-11-19-47](https://doi.org/10.22331/qv-2020-11-19-47).
- [137] D. Gross, K. Audenaert, and J. Eisert. “Evenly distributed unitaries: On the structure of unitary designs”. *J. Math. Phys.* 48 (2007), 052104. DOI: [10.1063/1.2716992](https://doi.org/10.1063/1.2716992).
- [138] A. Ketterer, N. Wyderka, and O. Gühne. “Entanglement characterization using quantum designs”. *Quantum* 4 (2020), 325. DOI: [10.22331/q-2020-09-16-325](https://doi.org/10.22331/q-2020-09-16-325).
- [139] T. Brydges et al. “Probing Rényi entanglement entropy via randomized measurements”. *Science* 364 (2019), 260–263. DOI: [10.1126/science.aau4963](https://doi.org/10.1126/science.aau4963).
- [140] A. Elben et al. “Mixed-State Entanglement from Local Randomized Measurements”. *Phys. Rev. Lett.* 125 (2020), 200501. DOI: [10.1103/PhysRevLett.125.200501](https://doi.org/10.1103/PhysRevLett.125.200501).
- [141] M. Grassl, M. Rötteler, and T. Beth. “Computing local invariants of quantum-bit systems”. *Phys. Rev. A* 58 (1998), 1833–1839. DOI: [10.1103/PhysRevA.58.1833](https://doi.org/10.1103/PhysRevA.58.1833).
- [142] T. A. Springer. *Invariant theory*. Lecture Notes in Mathematics 585. Springer Berlin, Heidelberg, 2006. DOI: [10.1007/BFb0095644](https://doi.org/10.1007/BFb0095644).
- [143] B.-Z. Sun, S.-M. Fei, and Z.-X. Wang. “On Local Unitary Equivalence of Two and Three-qubit States”. *Sci. Rep.* 7 (2017), 4869. DOI: [10.1038/s41598-017-04717-2](https://doi.org/10.1038/s41598-017-04717-2).
- [144] J. Kempe. “Multiparticle entanglement and its applications to cryptography”. *Phys. Rev. A* 60 (1999), 910–916. DOI: [10.1103/PhysRevA.60.910](https://doi.org/10.1103/PhysRevA.60.910).
- [145] H. Barnum and N. Linden. “Monotones and invariants for multi-particle quantum states”. *J. Phys. A: Math. Gen.* 34 (2001), 6787–6805. DOI: [10.1088/0305-4470/34/35/305](https://doi.org/10.1088/0305-4470/34/35/305).
- [146] S. Imai. “Randomized measurements as a tool in quantum information processing”. PhD thesis. Universität Siegen, 2023. DOI: [10.25819/ubsi/10422](https://doi.org/10.25819/ubsi/10422).
- [147] B. Collins, S. Matsumoto, and J. Novak. “The Weingarten Calculus”. *Not. Amer. Math. Soc.* 69 (2022), 734–745. DOI: [10.1090/noti2474](https://doi.org/10.1090/noti2474).
- [148] M. Fiorentino et al. “All-fiber photon-pair source for quantum communications”. *IEEE Photonics Technol. Lett.* 14 (2002), 983–985. DOI: [10.1109/LPT.2002.1012406](https://doi.org/10.1109/LPT.2002.1012406).

- [149] S. X. Wang and G. S. Kanter. “Robust Multiwavelength All-Fiber Source of Polarization-Entangled Photons With Built-In Analyzer Alignment Signal”. *IEEE J. Sel. Top. Quantum Electron.* 15 (2009), 1733–1740. DOI: [10.1109/JSTQE.2009.2022278](https://doi.org/10.1109/JSTQE.2009.2022278).
- [150] NuCrypt. “Quantum optical instrumentation”. <http://nucrypt.net/quantum-optical-instrumentation.html>. [Accessed: 11.10.2022].
- [151] D. E. Jones, B. T. Kirby, and M. Brodsky. “Joint Characterization of Two Single Photon Detectors with a Fiber-based Source of Entangled Photon Pairs”. *Frontiers in Opt. Opt. Soc. of America*. 2017, JW4A.37. DOI: [10.1364/FIO.2017.JW4A.37](https://doi.org/10.1364/FIO.2017.JW4A.37).
- [152] D. E. Jones, B. T. Kirby, and M. Brodsky. “In-situ calibration of fiber-optics entangled photon distribution system”. *IEEE Photonics Soc. Summer Topical Meeting Ser.* 2017, 123–124. DOI: [10.1109/PHOSST.2017.8012681](https://doi.org/10.1109/PHOSST.2017.8012681).
- [153] N. Wyderka and O. Gühne. “Characterizing quantum states via sector lengths”. *J. Phys. A: Math. Theor.* 53 (2020), 345302. DOI: [10.1088/1751-8121/ab7f0a](https://doi.org/10.1088/1751-8121/ab7f0a).
- [154] F. Verstraete and M. M. Wolf. “Entanglement versus Bell Violations and Their Behavior under Local Filtering Operations”. *Phys. Rev. Lett.* 89 (2002), 170401. DOI: [10.1103/PhysRevLett.89.170401](https://doi.org/10.1103/PhysRevLett.89.170401).
- [155] R. Horodecki, P. Horodecki, and M. Horodecki. “Violating Bell inequality by mixed spin-12 states: necessary and sufficient condition”. *Phys. Lett. A* 200 (1995), 340–344. DOI: [10.1016/0375-9601\(95\)00214-N](https://doi.org/10.1016/0375-9601(95)00214-N).
- [156] M. Horodecki, P. Horodecki, and R. Horodecki. “General teleportation channel, singlet fraction, and quasidistillation”. *Phys. Rev. A* 60 (1999), 1888. DOI: [10.1103/PhysRevA.60.1888](https://doi.org/10.1103/PhysRevA.60.1888).
- [157] O. Gühne, Y. Mao, and X.-D. Yu. “Geometry of Faithful Entanglement”. *Phys. Rev. Lett.* 126 (2021), 140503. DOI: [10.1103/PhysRevLett.126.140503](https://doi.org/10.1103/PhysRevLett.126.140503).
- [158] D. T. Smithey et al. “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 (1993), 1244–1247. DOI: [10.1103/PhysRevLett.70.1244](https://doi.org/10.1103/PhysRevLett.70.1244).
- [159] D. F. V. James et al. “Measurement of qubits”. *Phys. Rev. A* 64 (2001), 052312. DOI: [10.1103/PhysRevA.64.052312](https://doi.org/10.1103/PhysRevA.64.052312).
- [160] H. Häffner et al. “Scalable multiparticle entanglement of trapped ions”. *Nature* 438 (2005), 643–646. DOI: [10.1038/nature04279](https://doi.org/10.1038/nature04279).
- [161] C. Schwemmer et al. “Systematic Errors in Current Quantum State Tomography Tools”. *Phys. Rev. Lett.* 114 (2015), 080403. DOI: [10.1103/PhysRevLett.114.080403](https://doi.org/10.1103/PhysRevLett.114.080403).
- [162] M. Paris and J. Řeháček. *Quantum State Estimation*. Lecture Notes in Physics 649. Springer Berlin, Heidelberg, 2004. DOI: [10.1007/b98673](https://doi.org/10.1007/b98673).
- [163] M. Guță et al. “Fast state tomography with optimal error bounds”. *J. Phys. A: Math. Theor.* 53 (2020), 204001. DOI: [10.1088/1751-8121/ab8111](https://doi.org/10.1088/1751-8121/ab8111).
- [164] P. Mehta et al. “A high-bias, low-variance introduction to Machine Learning for physicists”. *Phys. Rep.* 810 (2019), 1–124. DOI: [10.1016/j.physrep.2019.03.001](https://doi.org/10.1016/j.physrep.2019.03.001).
- [165] C. M. Bishop. *Pattern Recognition and Machine Learning*. Springer New York, 2006. URL: <https://link.springer.com/book/9780387310732>.
- [166] C. Hadfield. “Adaptive Pauli Shadows for Energy Estimation” (2021). DOI: [10.48550/arXiv.2105.12207](https://doi.org/10.48550/arXiv.2105.12207).

- [167] C. Hadfield et al. “Measurements of Quantum Hamiltonians with Locally-Biased Classical Shadows”. *Commun. Math. Phys.* 391 (2022), 951–967. DOI: [10.1007/s00220-022-04343-8](https://doi.org/10.1007/s00220-022-04343-8).
- [168] A. Neven et al. “Symmetry-resolved entanglement detection using partial transpose moments”. *npj Quantum Inf.* 7 (2021), 152. DOI: [10.1038/s41534-021-00487-y](https://doi.org/10.1038/s41534-021-00487-y).
- [169] A. Rath et al. “Quantum Fisher Information from Randomized Measurements”. *Phys. Rev. Lett.* 127 (2021), 260501. DOI: [10.1103/PhysRevLett.127.260501](https://doi.org/10.1103/PhysRevLett.127.260501).
- [170] R. J. Garcia, Y. Zhou, and A. Jaffe. “Quantum scrambling with classical shadows”. *Phys. Rev. Res.* 3 (2021), 033155. DOI: [10.1103/PhysRevResearch.3.033155](https://doi.org/10.1103/PhysRevResearch.3.033155).
- [171] L. K. Joshi et al. “Probing Many-Body Quantum Chaos with Quantum Simulators”. *Phys. Rev. X* 12 (2022), 011018. DOI: [10.1103/PhysRevX.12.011018](https://doi.org/10.1103/PhysRevX.12.011018).
- [172] H.-Y. Huang, R. Kueng, and J. Preskill. “Efficient Estimation of Pauli Observables by Derandomization”. *Phys. Rev. Lett.* 127 (2021), 030503. DOI: [10.1103/PhysRevLett.127.030503](https://doi.org/10.1103/PhysRevLett.127.030503).
- [173] T. Zhang et al. “Experimental Quantum State Measurement with Classical Shadows”. *Phys. Rev. Lett.* 127 (2021), 200501. DOI: [10.1103/PhysRevLett.127.200501](https://doi.org/10.1103/PhysRevLett.127.200501).
- [174] H.-Y. Hu and Y.-Z. You. “Hamiltonian-driven shadow tomography of quantum states”. *Phys. Rev. Res.* 4 (2022), 013054. DOI: [10.1103/PhysRevResearch.4.013054](https://doi.org/10.1103/PhysRevResearch.4.013054).
- [175] H.-Y. Hu, S. Choi, and Y.-Z. You. “Classical shadow tomography with locally scrambled quantum dynamics”. *Phys. Rev. Res.* 5 (2023), 023027. DOI: [10.1103/PhysRevResearch.5.023027](https://doi.org/10.1103/PhysRevResearch.5.023027).
- [176] R. Levy, D. Luo, and B. K. Clark. “Classical shadows for quantum process tomography on near-term quantum computers”. *Phys. Rev. Res.* 6 (2024), 013029. DOI: [10.1103/PhysRevResearch.6.013029](https://doi.org/10.1103/PhysRevResearch.6.013029).
- [177] J. Helsen et al. “Shadow estimation of gate-set properties from random sequences”. *Nat. Commun.* 14 (2023), 5039. DOI: [10.1038/s41467-023-39382-9](https://doi.org/10.1038/s41467-023-39382-9).
- [178] Y. Chen et al. “Detector tomography on IBM quantum computers and mitigation of an imperfect measurement”. *Phys. Rev. A* 100 (2019), 052315. DOI: [10.1103/PhysRevA.100.052315](https://doi.org/10.1103/PhysRevA.100.052315).
- [179] A. Acharya, S. Saha, and A. M. Sengupta. “Shadow tomography based on informationally complete positive operator-valued measure”. *Phys. Rev. A* 104 (2021), 052418. DOI: [10.1103/PhysRevA.104.052418](https://doi.org/10.1103/PhysRevA.104.052418).
- [180] J. Steinberg. “Mathematical structures in quantum information theory: tensors, correlations and state estimation”. PhD thesis. Universität Siegen, 2023. DOI: [10.25819/ubsi/10377](https://doi.org/10.25819/ubsi/10377).
- [181] E. Van Den Berg. “A simple method for sampling random Clifford operators”. *2021 IEEE International Conference on Quantum Computing and Engineering (QCE)*. 2021, 54–59. DOI: [10.1109/QCE52317.2021.00021](https://doi.org/10.1109/QCE52317.2021.00021).
- [182] G. M. D’Ariano, P. L. Presti, and P. Perinotti. “Classical randomness in quantum measurements”. *J. Phys. A: Math. Gen.* 38 (2005), 5979–5991. DOI: [10.1088/0305-4470/38/26/010](https://doi.org/10.1088/0305-4470/38/26/010).
- [183] H. C. Nguyen et al. “Symmetries between measurements in quantum mechanics” (2020). DOI: [10.48550/arXiv.2003.12553](https://doi.org/10.48550/arXiv.2003.12553).

- [184] S. Bravyi et al. “Mitigating measurement errors in multiqubit experiments”. *Phys. Rev. A* 103 (2021), 042605. DOI: [10.1103/PhysRevA.103.042605](https://doi.org/10.1103/PhysRevA.103.042605).
- [185] R. Hicks et al. “Active readout-error mitigation”. *Phys. Rev. A* 105 (2022), 012419. DOI: [10.1103/PhysRevA.105.012419](https://doi.org/10.1103/PhysRevA.105.012419).
- [186] R. B. Stinchcombe. “Ising model in a transverse field. I. Basic theory”. *J. Phys. C: Solid State Phys.* 6 (1973), 2459. DOI: [10.1088/0022-3719/6/15/009](https://doi.org/10.1088/0022-3719/6/15/009).
- [187] S. Suzuki, J.-i. Inoue, and B. K. Chakrabarti. *Quantum Ising phases and transitions in transverse Ising models*. Lecture Notes in Physics 862. Springer Berlin, Heidelberg, 2012. DOI: [10.1007/978-3-642-33039-1](https://doi.org/10.1007/978-3-642-33039-1).
- [188] L. E. Fischer et al. “Ancilla-free implementation of generalized measurements for qubits embedded in a qudit space”. *Phys. Rev. Res.* 4 (2022), 033027. DOI: [10.1103/PhysRevResearch.4.033027](https://doi.org/10.1103/PhysRevResearch.4.033027).
- [189] R. Stricker et al. “Experimental Single-Setting Quantum State Tomography”. *PRX Quantum* 3 (2022), 040310. DOI: [10.1103/PRXQuantum.3.040310](https://doi.org/10.1103/PRXQuantum.3.040310).
- [190] C. de Gois and M. Kleinmann. “User-friendly confidence regions for quantum state tomography”. *Phys. Rev. A* 109 (2024), 062417. DOI: [10.1103/PhysRevA.109.062417](https://doi.org/10.1103/PhysRevA.109.062417).
- [191] A. Klenke. *Probability Theory: A Comprehensive Course*. 3rd ed. Universitext. Springer Cham, 2020. DOI: [10.1007/978-3-030-56402-5](https://doi.org/10.1007/978-3-030-56402-5).

Acknowledgments

Above all, I would like to thank my supervisor, Otfried Gühne, for the opportunity to do my PhD in his group. I'm deeply grateful for his support during my PhD. In particular, I'd like to thank him that he introduced me to the research in quantum information theory and that he was always available for discussions. I'm also grateful for his support finding a research topic at the beginning of my PhD and applying for a scholarship.

I'd also like to thank Mariami Gachechiladze for the discussions and that she refereed my thesis.

Moreover, I'd like to thank all my collaborators. In particular, I thank Satoya Imai and Ye-Chao Liu for the collaboration on my first project and the discussions on spin-squeezing and randomized measurements. Special thanks also go to Chau Nguyen for the discussions on shadow tomography and Nikolai Wyderka for his guidance in the project on LU invariants. I'm also very grateful for the discussions with Adán Cabello and his great hospitality during my research visit in Seville. My time in Seville was also greatly enriched by Henrique, Sebastián, Gwendal, and Maëlle.

I also would like to thank all members of the TQO group. I really appreciate the great social cohesion. In particular, I'd like to thank Lina Vandr e and Kiara Hansenne but also Fynn Otto for being great office mates. I really appreciate their friendship and the non-scientific side projects. I, moreover, enjoyed the conversations with Leonardo Santos over dinner. In addition, I also would like to thank Daniela Lehmann for always willing to help with bureaucracy.

Finally, I'd like to thank Satoya Imai, Chau Nguyen, Lina Vandr e, Yi Li, Ties Ohst, and Matthias Kleinmann for proofreading parts of my thesis.

Last, I would like to thank the people from acrobatics, from the ADFC, and all my friends. I'm also very grateful for the support of my parents and sister. Spending time with them was always a refreshing break.